═══REVIEW COMMONS MANUSCRIPT═══

IMPORTANT:

- Manuscripts submitted to Review Commons are peer reviewed in a journal-agnostic way.
- Upon transfer of the peer reviewed preprint to a journal, the referee reports will be available in full to the handling editor.
- The identity of the referees will NOT be communicated to the authors unless the reviewers choose to sign their report.
- The identity of the referee will be confidentially disclosed to any affiliate journals to which the manuscript is transferred.

GUIDELINES:

- For reviewers: https://www.reviewcommons.org/reviewers
- For authors: https://www.reviewcommons.org/authors

CONTACT:

The Review Commons office can be contacted directly at: office@reviewcommons.org

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 highly expressed in cranial neural crest cells (CNCCs) and associated with syndromes featuring craniofacial anomalies, plays a critical role in this early regulatory process. Using *SALL4*-haploinsufficient human iPSC models that recapitulate clinical haploinsufficiency, we show that SALL4 directly interacts with the BAF chromatin remodeling complex at the neuroectodermal stage, specifically when neural crest lineage specific gene expression is induced from the neural plate border. The SALL4-BAF interaction at this stage is essential for the recruitment of BAF to CNCC-induction enhancers. 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

INTRODUCTION

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

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

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

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

Sall4 remains expressed in mouse craniofacial regions until the E9.5 stage, including the developing frontonasal process as well as the mandibular and maxillary arches[18]. These craniofacial structures originate from cranial neural crest cells (CNCCs), a subpopulation of neural crest cells that migrate dorsolaterally to give rise to the cartilage and bones of the face and anterior skull, as well as cranial neurons and glia[21–25]. 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 [26]. Many of these distinctive features are classic indicators of impaired development in cranial neural crest cells. *SALL4*-associated phenotypes often include Duane anomaly, a congenital strabismus resulting from abnormal oculomotor nerve innervation[26], and sensorineural hearing loss[27], but also upper limb anomalies (Duane-radial ray syndrome or DRRS) and, in some cases, renal anomalies.

Together, these findings suggest that SALL4 may play a critical role in CNCC induction, specification and differentiation. However, its involvement in this developmental process has not yet been investigated. Previous studies have also suggested that SALL4 interacts directly with chromatin remodeling complexes. For example, several studies in mice have identified Sall4 as an interactor of the NuRD chromatin remodeling complex, in both embryonic cells and various developmental processes[15,28–31]. Additionally, recent mass-spectrometry experiments conducted by our group in early stages of human iPSC-to-CNCC differentiation, and specifically at the neuroectodermal stage, revealed that SALL4 interacts with the BAF chromatin remodeling complex during this developmental process[32].

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 cranial neural crest cell specification. Given that human SALL4-associated developmental syndromes result from *SALL4* haploinsufficiency, we generated *SALL4* heterozygous human iPSC lines using CRISPR-Cas9-mediated deletion (*SALL4*-het-KO). With this approach, we demonstrate that, at early stages of CNCC specification, SALL4 is required for the recruitment of the BAF complex to open the chromatin to enhancers essential for CNCC fate. In the absence of SALL4, cells progress normally through the neuroectodermal stage but fail to specify into CNCCs, as the lineage-specific enhancers remain repressed. Finally, using *in silico* approaches, we predict that SALL4A isoform directly interacts with the BAF subunit DPF2.

RESULTS

SALL4 interacts with the BAF complex

Cranial neural crest cells can be generated *in vitro* starting from human induced pluripotent stem cells (hiPSCs). This protocol is approximately two weeks long, and it first goes through a neuroectoderm intermediate [32–36]. In our previous work[32], mass spectrometry (IP-MS) conducted at early stages of the iPSC-to-CNCC differentiation protocol, and specifically Day-5, which corresponds to the neuroectodermal stage, revealed that SALL4 interacts with the BAF chromatin remodeling complex in this developmental cell population. Since this was the first report of a SALL4-BAF interaction, we performed a co-immunoprecipitation (co-IP) at Day-5 to validate it. Specifically, we performed pull-down of ARID1B, which is the only ARID1 subunit active and incorporated in BAF at this stage [32], and blotted for SALL4. This co-IP confirmed that ARID1B interacts with both SALL4 isoforms at this time-point (Fig. 1A).

The BAF chromatin remodeling complex is a multi-subunit assembly composed of a dozen components, which are differentially incorporated in a tissue-specific manner[37]. 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[38]. While our mass-spec and co-IP experiments support the interaction between SALL4 and the canonical BAF (ARID1B is not incorporated in the 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[39] (https://alphafoldserver.com/). 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, initially, the longer SALL4 isoform, SALL4A. The output was statistically significant, with a 0.45 predicted template modeling (pTM) score and 0.47 interface predicted template modeling (ipTM) score (Supplementary Figs. S1A, B). Through this approach, AlphaFold3 predicted a direct interaction between DPF2, a broadly expressed BAF subunit, and the zinc-fingers 5 and 6 (ZFC3) and 7 (ZFC4) of SALL4A (Figs. 1B, C, D; Supplementary Figs. S1A, B; Supplementary Table 1). *DPF2* is a member of the *BAF45* gene family, which also includes *PHF10*, *DPF1*, and *DPF3*. We were able to rule out its paralogues based on their distinct expression patterns, incorporation into specific BAF complexes, and previous reports that BAF45 subunits are not functionally redundant[40]. In particular, PHF10 is exclusively associated with the pBAF complex, which does not incorporate ARID1B[41], while DPF1 and DPF3 are specifically incorporated into the neuron-specific nBAF complex[42]. Moreover, we compared the gene expression of each *DPF* paralogue

134 at both Day-0 and Day-5 and saw no significant difference in expression between *SALL4*-WT and
135 *SALL4*-het-KO for any of them (Figs. 1E, F).
136 This prediction was particularly interesting given that DPF2 deletion has been linked to early
137 neural ectoderm differentiation in embryonic stem cells (ESCs)[40]. DPF2 exists in two isoforms,
138 resulting from alternative splicing of exon 7 and are spliced in a tissue-specific and
139 developmental-stage-specific manner[43]. DPF2-L is not alternatively spliced and is expressed in
140 ESCs, while DPF2-S is expressed in neurons. Interestingly, both isoforms are expressed in neural
141 stem cells. Notably, of the two main SALL4 isoforms (SALL4A and SALL4B), SALL4B lacks the
142 portion of exon 2 that contains ZFC3. This prompted us to explore whether the SALL4-DPF2
143 interaction could be isoform-specific. To address this, we reanalysed the interactions in
144 AlphaFold3, incorporating SALL4B to the same BAF subunits as before. To ensure biological
145 relevance, we first excluded two of the five models that did not exhibit interactions with the AT-
146 rich DNA sequence, as SALL4 is expected to bind DNA for its function. Among the remaining
147 three models, we observed that SALL4B interacted with DPF2, but unlike the SALL4A isoform,
148 its interaction was not domain specific. Instead, the predicted interactions for SALL4B were
149 characterized by annotated disordered regions and appeared to be largely driven by a non-
150 specific, zinc-finger-to-zinc-finger association (Supplementary Figs. S1C, D; Supplementary
151 Table 2), with no particular domain-specific interactions identified. This pattern may be due to the
152 tendency of AlphaFold3 learning algorithm to associate similar domains, such as zinc-fingers,
153 based on their predicted functional specificity rather than spatial conformations. These findings
154 suggest that, during neural ectoderm specification, SALL4 might associate with BAF in an
155 isoform-dependent manner.
156 To accurately interpret our AlphaFold3 predictions, we examined two key confidence metrics: the
157 Predicted Aligned Error (PAE), which assesses the reliability of the relative positioning of residues
158 within the structure, and the Predicted Local Distance Difference Test (pLDDT), which provides a
159 per-residue confidence score for the model (Supplementary Table 1). Our SALL4-DPF2 contact
160 analysis revealed 394 atomic interactions (center-to-center distance ≤ 8Å) involving 139 residues
161 (Fig. 1D and Supplementary Table 1). Encouragingly, the pLDDT values were highest in known
162 functional domains involved in the interaction, particularly ZF5, ZF6, and ZF7 in SALL4A, as well
163 as the PHD1 and PHD2 domains in DPF2, reinforcing the reliability of these predicted structures.
164 Consistently, our model also aligned closely with experimentally resolved structures at 3 Å
165 resolution available for the BAF base module (Supplementary Fig. 1E). To refine our analysis, we
166 focused on residues where at least half of their atomic interactions had a pLDDT > 65 (70 ± 10%)
167 and that were annotated as part of a functional domain, ensuring that only higher-confidence and

168 potentially functional interactions were considered. As a result, ZF7 was excluded from further
169 analysis (Supplementary Table 1).
170 In summary, our experiments and analyses confirmed that SALL4 interacts with BAF during early
171 stages of human CNCC specification, and predicted that this interaction might be mediated by
172 the BAF subunit DPF2 through the ZFC3.
173
174 **Generation of a *SALL4*-haploinsufficient human iPSC lines**
175 Our experiments so far have revealed that SALL4 interacts with the BAF complex at early stages
176 of CNCC specification, and specifically at the neuroectodermal stage. To investigate the scope of
177 the SALL4-BAF interaction during this developmental process, we generated a *SALL4*
178 haploinsufficient iPSC line (hereafter *SALL4*-het-KO; Fig. 2A). We opted for a heterozygous KO
179 of *SALL4* as this recapitulates the human *SALL4* syndromes, which are all caused by *SALL4*
180 haploinsufficient genetic variants[27,44,45]. Two different CRISPR/Cas9 clones were generated and
181 used as biological replicates. Specifically, the CRISPR-KO was generated by introducing a 11 bp
182 (clone C4) deletion and a 19 bp (clone F1) deletion, respectively (Supplementary Fig. S2A), both
183 within the 3'-terminus of exon 1 of *SALL4*. This is a conserved region across both *SALL4* isoforms
184 that lacks functionally significant domains. Therefore, both *SALL4* isoforms (*SALL4A* and
185 *SALL4B*) were equally impacted by the CRISPR deletions.
186 SALL4 immunoblot and immunofluorescence assays confirmed that SALL4 protein expression
187 was significantly reduced in the *SALL4*-het-KO iPSCs compared to the unedited wild-type line
188 (*SALL4*-WT; Figs. 2B, C; Supplementary Fig. S2B, C). We noted that SALL4A protein expression
189 in the *SALL4*-het-KO line was significantly less than the expected 50% of the wild-type,
190 suggesting potential negative feedback mechanisms triggered by the reduced SALL4 levels,
191 possibly due to the preferentially homodimeric nature of this protein. On the other hand, SALL4B
192 protein was barely detectable in both *SALL4*-WT and *SALL4*-het-KO conditions (Fig. 2B).
193 In contrast, at the gene expression level we did not observed significant differences in SALL4
194 expression between *SALL4*-WT and *SALL4*-het-KO conditions, for any of the isoforms (Fig. 2D).
195 This may reflect autoregulatory mechanisms, where reduced SALL4 protein levels in the knockout
196 trigger compensatory transcriptional upregulation of the gene which however does not result in
197 increased protein level because of nonsense-mediated decay or other post-transcriptional
198 regulatory mechanisms.
199 Next, we investigated whether other SALL paralogs (*SALL1*, *SALL2*, *SALL3*) could compensate
200 for the loss of SALL4. First, we investigated their expression at mRNA level level (Fig. 2E) and
201 found that *SALL1* was the only paralog significantly upregulated in the *SALL4*-het-KO cell line

202 (Fig. 2E). SALL1 upregulation was confirmed also at protein level (Fig. 2F), indicating a potential
203 compensation in pluripotent cells.
204 Importantly, *SALL4*-het-KO iPSCs remained viable and exhibited hallmarks iPSC morphology,
205 including a high nucleus-to-cytoplasm ratio and growth in compact colonies with well-defined
206 borders (Supplementary Fig. S2D). Furthermore, these cells show no significant difference in
207 growth rate compared to their isogenic controls (Fig. 2G).
208
209 **Decreased SALL4 expression does not affect the stem cell identity of the iPSCs**
210 Given the debated role of SALL4 in maintaining pluripotency, we investigated whether *SALL4*-
211 het-KO impacted iPSC stem cell identity. Thus, we examined the protein levels of the canonical
212 pluripotency transcription factors NANOG and OCT4 and found that they remained unaltered in
213 the *SALL4*-het-KO cells relative to the *SALL4*-WT control (Fig. 3A). Similarly, the levels of the
214 pluripotency surface markers TRA-1-60-R and SSEA4 were also not significantly affected by
215 decreased SALL4 expression (Fig. 3B; Supplementary Fig. S4A). Moreover, we examined the
216 transcriptome of the iPSCs using bulk RNA-seq and identified only 114 differentially expressed
217 genes in *SALL4*-het-KO cells relative to the *SALL4*-WT control (FDR < 0.05; Fig. 3C and
218 Supplementary Table 3). Notably, canonical pluripotency markers, such as *POU5F1, NANOG,*
219 *SOX2, KLF4,* and *MYC*, were not significantly differentially expressed (Fig. 3C and
220 Supplementary Table 3). Consistent with this, gene ontology analysis confirmed that neither up-
221 (Supplementary Fig. S3A) or downregulated (Supplementary Fig. S3B) genes were enriched for
222 pluripotency-associated terms. Finally, by performing tri-lineage differentiation we confirmed that
223 SALL4 depletion did not impact the ability of the *SALL4*-het-KO cells to successfully differentiate
224 into all three of the germ layers (Supplementary Fig. S3C).
225 Together, these findings suggest that *SALL4* haploinsufficiency did not affect IPSC stem cell
226 identity and self-renewal, possibly as a result of SALL1-mediated compensation.
227
228 **SALL4 relocates to CNCC developmental enhancers upon CNCC induction cues**
229 After establishing that *SALL4*-het-KO did not affect the stemness of the iPSCs, we differentiated
230 the iPSCs into CNCCs using the same aforementioned protocol[32–36] (Fig. 4A). Since the SALL4-
231 BAF interaction was detected at Day-5 of this protocol[32], corresponding to neuroectodermal stage,
232 we performed CUT&RUN to profile SALL4 binding in *SALL4*-WT cells at the iPSC stage and at
233 Day-5 of differentiation. We performed two SALL4 CUT&RUN replicates per timepoint, and only
234 retained replicated peaks in each timepoint (FDR < 0.05). With this approach, we identified
235 comparable amounts of SALL4-bound genomic sites at Day-0 and Day-5 (29,088 and 32,707,

236 replicated peaks, respectively, FDR < 0.05; Fig. 4B). Further analysis of these sites indicated that
237 approximately 20% of the Day-0 SALL4 peaks were unique to the iPSC stage and not detected
238 at Day-5 (hereafter SALL4 Day-0 specific peaks; n = 6,114; Fig. 4B, 4C, left). On the other hand,
239 ~50% of the replicated peaks detected at Day-5 were exclusive of this developmental stage and
240 were not detected at Day-0 (hereafter SALL4 Day-5-specific peaks, n = 17,137; Fig. 4B and 4C,
241 right). Finally, a subset of ~15,000 sites remained bound by SALL4 at both time points. Together,
242 this data indicate that, upon neural crest differentiation cues, there is a major shift in SALL4
243 binding, resulting in the relocation of this transcription factor at ~17,000 previously unbound
244 regions.
245 To characterize the scope of SALL4 relocation, we analyzed the positional distribution of SALL4
246 peaks by associating each peak with its nearest transcription start site (TSS), identifying 4,362
247 genes at Day-0 and 6,861 genes at Day-5 (Supplementary Table 4). Consistent with SALL4's
248 established role as an enhancer-binding transcription factor, 87% of SALL4-bound regions at Day-
249 0 were located >1 kb from the nearest TSS (n = 3,816), suggesting these represent putative
250 enhancers. This proportion increased substantially to 97% (n = 6,644) at Day-5, demonstrating
251 SALL4's preferential binding to distal regulatory regions (Fig. 4D).
252 To investigate the scope of the newly acquired SALL4 Day-5 peaks, we first looked at the nearest
253 genes. GO terms enriched in the “Day-5 specific” genes included pathways associated with neural
254 crest development, such as neural tube development, face development, cardiac chamber
255 development, and cell-cell adhesion mediated by cadherin (Fig. 4E; FDR < 0.05).
256 To further characterise the SALL4 bound elements at Day-0 and Day-5, we performed sequence-
257 based computational motif analysis. At Day-0, motifs associated with core pluripotency
258 transcription factors such as OCT4, NANOG, and SOX2 did not rank among the top 20 enriched
259 sequences, further supporting that SALL4 may not be essential for maintaining pluripotency in
260 human cells (Supplementary Fig. S5A).
261 On the other hand, in addition to general enhancer activator factors (AP-1, TEAD), the Day-5
262 specific SALL4 peaks were enriched for many CNCC signature factors, including SOX9, TFAP2A,
263 TWIST1/2, SNAI1, NR2F1/2, TCF4/12 and several others (Fig. 4F). Moreover, neuroectodermal
264 transcription factors, such as ZIC2/3 were also enriched. Additionally, *de novo* motif analysis in
265 iPSCs revealed enrichment for the CTCF-like motifs (Supplementary Fig. S5B), while at the Day-
266 5 there was enrichment for enhancer activators TEADS/ETS and AP-1 (Supplementary Fig. S5C).
267 Finally, AT-rich motifs resembling the SALL4 annotated motif were found only in a small minority
268 of peaks in iPSCs (Supplementary Fig. S5B). This might suggest that SALL4 binding at the CNCC

enhancers is not dictated by recognition of its own motif, and possibly mediated by other lineage-specific transcription factors.

In summary, our data so far indicate that SALL4 relocates to CNCC developmental enhancers upon differentiation cues.

Chromatin accessibility at the CNCC enhancers is SALL4 dependent

Given that we demonstrated that SALL4 interacts with BAF at early stages of CNCC specification, we reasoned that SALL4 binding at the CNCC developmental enhancers could be necessary for their chromatin activation. Thus, we profiled chromatin accessibility by performing ATAC-seq at Day-0 and Day-5 in both *SALL4*-WT and *SALL4*-het-KO conditions (two biological replicates per timepoint per condition, only peaks present in both replicates were considered; FDR < 0.05). The number of ATAC-seq replicated peaks identified in *SALL4*-WT condition was comparable between Day-0 (61,660 peaks; FDR < 0.05) and Day-5 (58,375 peaks; FDR < 0.05; Fig. 5A). Next, we specifically investigated the chromatin state of the 17,137 genomic sites that gained SALL4 binding at Day-5, and found that SALL4 binding at these sites correlates with a marked increase in chromatin accessibility compared to Day-0 (Fig. 5B). Strikingly, the accessibility at these regions was significantly attenuated in *SALL4*-het-KO relative to *SALL4*-WT (Fig. 5C).

Overall, these findings suggest that, in response to differentiation signals, SALL4 relocates to CNCC developmental enhancers, which subsequently become activated through chromatin opening, which is SALL4 dependent.

Motif analysis of regions that lost accessibility in *SALL4*-het-KO cells (n = 1,649; Fig. 5A) showed again enrichment for CNCC factors such as ZIC2, ZEB1, ZEB2, and SNAIL, along with BMP signalling factors such as the SMAD paralogs, and enhancer activators such as AP-1 (Fig. 5D). On the other hand, regions aberrantly accessible in the *SALL4*-het-KO cells (n = 3,460; Fig. 5A) showed enrichment for pluripotency factors, CTCF, and neural plate markers such as the ZIC paralogs (Fig. 5E).

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

Since SALL4 binding at the CNCC enhancers is necessary for chromatin opening, we sought to determine whether the SALL4-bound developmental enhancers were also occupied by BAF, and if BAF binding at these sites could be SALL4 dependent. We performed ChIP-seq for BRG1 (SMARCA4), which is the only catalytic subunit of BAF active at this time point, in *SALL4*-WT cells at Day-5. We found that BRG1 consistently binds the same genomic sites as SALL4,

supporting the idea that these factors interact and co-occupy regulatory elements during early differentiation (Fig. 6A).

To assess whether BRG1 binding depends on SALL4, we performed ChIP-seq for BRG1 in both *SALL4*-WT and *SALL4*-het-KO cells at Day-5 (two biological replicates, one per clone). In *SALL4*-WT cells, we detected 32,833 replicated BRG1 peaks (FDR < 0.05), as opposed to only 14,116 peaks in *SALL4*-het-KO cells. This suggests that SALL4 loss triggered widespread loss of BAF from chromatin. We thus focused on the BRG1 peaks that were lost in *SALL4*-het-KO cells and found that these sites were largely not bound by SALL4 at Day-0, but they gained SALL4 binding at Day-5 (Figs. 6B–D). As expected, loss of SALL4 and BRG1 from these sites led to marked decrease of chromatin accessibility (Fig. 6E). Moreover, we explored the identity of some of these putative enhacers by confirming that these overlapped with some of those reported on databases such as VISTA Enhancer Browser[46] and GeneHancer[47] (see examples in Fig. 6F). Overall, these experiments support that SALL4 is required to recruit the BAF complex at CNCC enhancers.

To further confirm that these regions are enhancers, we leveraged publicly available H3K4me1 and H3K4me3 ChIP-seq datasets generated using a comparable protocol at our Day-5 equivalent [36,48]. Moreover, we performed ChIP-seq for H3K27ac at Day-5, Day-11 and Day-14, roughly corresponding to CNCC-induction, early CNCC, late CNCC stages. This analysis revealed that nearly all the sites bound by SALL4 and BRG1 are marked by the enhancer mark H3K4me1, and only a small subset of them has the promoter mark H3K4me3 (Supplementary Fig. S6A). Moreover, nearly all of these sites have the active enhancer mark H3K27ac at Day-5, and the acetylation levels progressively increase from CNCC induction to CNCC stage, supporting that these are *bona fide* CNCC enhancers that increase in activation after SALL4 binding (Supplementary Figs. S6B, C)

While a large portion of BRG1 peaks was lost in the absence of SALL4, we also identified a subset of novel BRG1 peaks that were absent in *SALL4*-WT cells but gained in the *SALL4*-het-KO cells. These BRG1 peaks gained in the *SALL4*-het-KO cells were associated with genes linked to neurogenesis and nervous system development (Supplementary Fig. S7A and Supplementary Table 4). In contrast, the genes located near BRG1 peaks lost in the *SALL4*-het-KO cells were involved in embryonic organ development, epithelial differentiation, and cell migration processes (Supplementary Fig. S7B and Supplementary Table 4).

Together, these results suggest that SALL4 is a key regulator of BAF complex recruitment to CNCC-specific enhancers, ensuring their accessibility during differentiation. In the absence of SALL4, BRG1 fails to localize to these sites and is instead partially redirected to neurodevelopmental enhancers, leading to a misallocation of chromatin remodeling activity,

337 potentially leading to a shift towards default neural differentiation at the expense of CNCC
338 induction and specification.
339
340 ***SALL4*-het-KO cells are not able to exit the neural ectoderm stage and to induce the cranial**
341 **neural crest fate**
342 Our data thus far indicate that SALL4 relocates to CNCC enhancers as early as the
343 neuroectodermal stage, recruiting BAF to these regulatory elements for chromatin activation.
344 Next, we investigated the functional consequences of SALL4 loss by generating CNCCs using
345 *SALL4*-WT and *SALL4*-het-KO cells. As expected, the *SALL4*-WT CNCCs expressed the
346 signature CNCC markers at both the mRNA (Fig. 7A) and protein (Fig. 7B and Supplementary
347 Fig. S4B) level, while *SALL4*-het-KO failed to induce these genes (Fig. 7A). This was further
348 supported by RNA-seq conducted at Day-14, which identified 1,106 genes differentially expressed
349 in *SALL4*-het-KO CNCCs relative to *SALL4*-WT (FDR < 0.05; Fig. 7C and Supplementary Table
350 3). This was in sharp contrast with the small number of genes differentially expressed at the iPSC
351 stage (n = 114; Fig. 3C and Supplementary Table 3) and at Day-5 (n = 170; Supplementary Fig.
352 S7C; Supplementary Table 3).
353 Of the 1,106 genes differentially expressed at Day-14, ~60% (n = 659) were downregulated while
354 the remaining were upregulated (Fig. 7C). Canonical CNCC markers such as *TFAP2A*, *NR2F1*,
355 and *SOX9* were significantly downregulated in *SALL4*-het-KO cells. Additionally, genes
356 associated with neural plate border, early neural crest induction and epithelial-to-mesenchymal
357 transition (EMT) were also downregulated (*MSX1*, *GBX2*, *SNAI2*, *SOX5*, *ZEB1/2*; Fig. 7D, F).
358 In line with these findings, gene ontology analysis revealed that pathways associated with face
359 development, autonomic nervous system, gastrulation, and regulation of morphogenesis were all
360 downregulated (Fig. 7D). The 447 upregulated genes were instead involved in cell regulatory
361 pathways, especially vesicle-mediated transport of inorganic substances (Fig. 7E).
362 As aforementioned, genes associated with EMT were also dysregulated. The process of EMT
363 requires a switch in cadherin expression where E-cadherin (*CDH1*) must be downregulated, while
364 N-cadherin (*CDH2*) must be upregulated. Related to this, our RNA-seq data revealed that this
365 cadherin switch did not occur in the differentiating *SALL4*-het-KO iPSCs, suggesting that the cells
366 remained in an epithelial state. This was further corroborated by the upregulation of several
367 epithelial genes in the differentiating *SALL4*-het-KO cells, including *EpCAM*, *OTX2*, *CDH1*, and
368 *OVOL2* (Fig. 7F). The latter is involved in maintaining epithelial lineages by repressing the
369 expression of the EMT gene *ZEB1*, which was downregulated in the *SALL4*-het-KO condition[49,50].

370 In summary, our data indicate that CNCC specification failed in *SALL4*-het-KO conditions, as
371 reflected by the large number of differentially expressed genes. On the other hand, the
372 transcriptomes of the iPSCs and of the neuroectodermal cells (i.e. Day-5) were only modestly
373 affected by SALL4 loss, suggesting that SALL4 may not be essential for neuroectoderm
374 formation, while it is essential for CNCC induction and specification.
375 To investigate this further, we systematically examined the expression of a panel of stage-specific
376 marker genes at Day-0, Day-5 and Day-14 (Supplementary Fig. S8A–D) including core
377 pluripotency genes (*POU5F1* and *NANOG*), neural plate border specifiers (*ZIC1, GBX2, MSX1/2,*
378 *PAX3/7, GATA2/3*), CNCC specification markers (*SNAI1/2, SOX9, TFAP2A*), and post-migratory
379 CNCC markers (*SOX5, SOX6, SOX10*). In *SALL4*-WT cells, pluripotency markers exhibited a
380 progressive and expected downregulation as differentiation advanced from Day-0 to Day-14. In
381 *SALL4*-het-KO cells, a similar decline in pluripotency markers was observed between Day-0 and
382 Day-5, but instead of continuing to decrease, these markers remained similarly expressed through
383 Day-14 (Supplementary Fig. S8A). Despite this, the two conditions showed similar trends,
384 confirming that decreased levels of SALL4 did not significantly impact pluripotency markers
385 expression during exit from pluripotency. Similarly, no significant difference was observed in the
386 expression of neural plate border specifiers (Supplementary Fig. S8B), suggesting that SALL4 is
387 not required for human neuroectoderm specification. Conversely, CNCC specification markers
388 and post-migratory CNCC markers all increased in expression after Day-5 in *SALL4*-WT cells,
389 while the SALL4-depleted cells failed to maintain a comparable increase in expression
390 (Supplementary Fig. S8C, D).
391 In summary, *SALL4*-het-KO cells are competent to generate neuroectoderm-like cells (Day-5) but
392 fail to induce the neural crest lineage (Day-14).
393
394 **SALL4A but not SALL4B is necessary for CNCC specification**
395 Given previous reports suggesting distinct roles for the two SALL4 isoforms in ESCs[12], and our
396 prediction that the interaction between SALL4 and DPF2 may be isoform-specific, we investigated
397 whether CNCC differentiation exhibits preferential dependence on a particular SALL4 isoform.
398 Timecourse of mRNA expression in *SALL4*-WT cells revealed that *SALL4B* is consistently lowly
399 expressed throughout differentiation, whereas *SALL4A* expression increases upon differentiation
400 cues, peaking around Day-5 and then declining over time (Supplementary Figs. S9A, B). On the
401 other hand, the paralog SALL1 is expressed at low levels in iPSCs and barely detectable
402 throughout the entire differentiation protocol (Supplementary Fig. S9C).

403 To investigate potential isoform-specific requirements for CNCC development, we selectively
404 degraded SALL4A while preserving SALL4B expression. To achieve this, we treated cells with
405 pomalidomide, a thalidomide derivative that selectively degrades SALL4A but not SALL4B[51,52].
406 This compound functions by binding to cereblon (CRBN), a substrate receptor of the CRL4-CRBN
407 E3 ubiquitin ligase complex (CUL4-DDB1-RBX1-CRBN), thereby altering its substrate specificity
408 to selectively recruit SALL4A for ubiquitin-mediated degradation[53,54] (Fig. 8A).
409 To test the effects of SALL4A degradation, we differentiated *SALL4*-WT iPSCs into CNCCs using
410 the aforementioned protocol. Since pomalidomide is unstable in aqueous buffers and requires
411 regular replenishment, the cells were treated with either DMSO (control) or 5 μM pomalidomide
412 (5 μM POM) every 24 hours (Fig. 8B). To confirm consistent degradation of SALL4A throughout
413 the differentiation process, we performed immunoblotting at key time points (Day-5 and Day-14).
414 As expected, SALL4A was successfully degraded, while SALL4B expression remained intact (Fig.
415 8C). Importantly, pomalidomide degrades several proteins beyond SALL4, and its protein
416 degradation activity is typically cell-type specific[55]. Two of the known pomalidomide targets have
417 suggested functions in neural crest development. These are ARID2[33,56] and ZMYM2[57,58].
418 Crucially, immunoblot analysis for these two proteins performed at Day 5 following treatment with
419 5 μM pomalidomide revealed that the levels of both ARID2 and ZMYM2 remained unchanged
420 compared to the DMSO control (Supplementary Figs. S9D, E), further supporting that
421 pomalidomide affinity for its targets is cell-type specific, as previously suggested [33,56]. This control
422 experiments allowed to rule out potential off-target effects caused by unwanted degradation of
423 one of these two proteins. The other known pomalidomide targets, including IKAROS, AIOLOS,
424 PLZF and ZNFX1 have no established contribution to neural crest specification and in fact they
425 are not highly expressed in our system.
426 Next, we assessed the downstream effects of SALL4A degradation by performing RT-qPCR and
427 immunofluorescence for key neural crest markers, such as *TFAP2A*, *NR2F1*, and *SOX9* (Figs.
428 8D, E), and EMT markers, namely vimentin (VIM) and EpCAM (Fig. 8F). Degradation of SALL4A
429 recapitulated what we previously observed in *SALL4*-het-KO cells. Specifically, the cells treated
430 with pomalidomide failed to differentiate successfully into CNCCs despite the persistent
431 expression of SALL4B.
432 Taken together, our results highlight a specific and essential role for SALL4A in orchestrating
433 chromatin dynamics during CNCC specification and suggest that SALL4B might function in a
434 different regulatory capacity rather than acting as a compensatory factor for SALL4A.

435

436

437 **DISCUSSION**

438 During embryonic development, SALL4 plays a critical role in forming major anatomical
439 structures, including the heart, limbs, cerebral midline, and pituitary gland[54]. Correspondingly,
440 mutations in the haploinsufficient gene *SALL4* are associated with congenital syndromes
441 characterized by malformations in these structures. These syndromes also frequently include
442 distinct craniofacial anomalies, such as epicanthal folds, hypertelorism, a depressed nasal bridge,
443 and hemifacial microsomia[59–61], all of which are hallmark features of disrupted cranial neural crest
444 cell development.

445 Notably, at the murine embryonic stage E9.5, Sall4 expression persists within the frontonasal
446 prominence, which is the most anterior embryonic region[18], implying a potential role for SALL4 in
447 craniofacial development. Given the limited existing knowledge about SALL4 function in this
448 context, we utilized a heterozygous *SALL4* knockout human induced pluripotent stem cell (iPSC)
449 line to investigate its role during cranial neural crest cell (CNCC) specification. Although SALL4
450 expression begins as early as the 2-cell embryonic stage, its role in maintaining pluripotency
451 remains debated. Some studies indicate that SALL4 cooperates closely with key pluripotency
452 factors such as OCT4 and NANOG to sustain stem cell identity[13,14]. Others, however, report that
453 SALL4 has only limited involvement within pluripotency regulatory networks[15–17]. Our findings
454 align with the latter hypothesis, demonstrating that decreased expression of SALL4 does not
455 compromise iPSC identity, possibly due to compensatory mechanisms involving its paralogue,
456 SALL1.

457 Using CUT&RUN, we observed that at early stages of CNCC induction, SALL4 relocates to
458 enhancers of genes involved in CNCC formation, and that SALL4 binding at these regulatory
459 elements is required for their activation. In humans, SALL4 exists in two major isoforms generated
460 by alternative splicing: SALL4A (1,053 amino acids), which retains all eight zinc-finger motifs, and
461 SALL4B (617 amino acids), which contains a partially spliced exon 2. Rao et al.[12] proposed distinct
462 functional roles for these isoforms in murine ESCs, demonstrating that SALL4A homodimers
463 localize predominantly to enhancers of developmental genes, whereas pluripotency-associated
464 enhancers are occupied either by heterodimers comprising both isoforms or by SALL4B alone. In
465 mouse ESCs, approximately 10% of the total Sall4 protein, and primarily the Sall4A isoform, is
466 incorporated into the NuRD complex [15,29]. Here, we identify and characterize a novel interaction
467 between SALL4 and the BAF complex, predicted to be mediated specifically through the SALL4A
468 isoform and the BAF subunit DPF2. Furthermore, we show that upon receiving CNCC induction
469 cues, SALL4 and BAF co-localize to enhancers of genes involved in CNCC specification to prime
470 and activate them, with BAF recruitment being SALL4-dependent.

471 Interestingly, DPF2 and SALL4 exhibit notable functional parallels. Similar to SALL4, DPF2
472 occupancy undergoes significant repositioning during the transition from pluripotency toward early
473 neural commitment in mouse embryoid bodies [62]. Although the role of DPF2 in CNCCs has not
474 been directly investigated, its early recruitment to neuronal enhancers suggests a potential
475 involvement in neuroectodermal differentiation and the subsequent lineage commitment of
476 neuroectoderm-derived cells, including CNCCs. Moreover, genomic sequences bound by DPF2
477 in mouse neuroectodermal cells are AT-rich, resembling the preferred binding motif of
478 SALL4[43,63,64]. It is important to note that AlphaFold3 core is built on deep learning rather than
479 reinforcement learning. As such, it is trained on large datasets of experimentally determined
480 protein structures, such as those in the Protein Data Bank (PDB). For SALL4, experimentally
481 resolved structures are limited to the ZFC4 domain bound to DNA and ~60 amino acids of ZF2 in
482 complex with cereblon. This scarcity of complete empirical structures means that AlphaFold
483 predictions for SALL4, particularly for regions outside these known fragments, may have lower
484 pIDDT scores than well-characterized proteins. Consequently, the absence of more
485 comprehensive experimental SALL4 data could significantly affect the reliability of the proposed
486 SALL4–DPF2 interaction model. Additionally, AlphaFold 5,000-token limit per prediction poses a
487 constraint when modeling large multi-subunit complexes such as BAF. Fortunately, the BAF
488 complex has been extensively characterized, and our predicted model match experimentally
489 resolved structures at 3 Å resolution (Supplementary Fig. S1E). Further high-resolution structural
490 proteomic analysis, such as Cryo-EM and X-ray crystallography are required to confirm a direct
491 protein-protein interaction between SALL4 and DPF2.
492 Our study demonstrated that while cranial neural crest cell induction and specification fail in the
493 absence of SALL4, early neuroectodermal commitment and neural plate border identity remain
494 unaffected. Our findings demonstrate that, during the earliest stages of CNCC specification,
495 SALL4 occupies CNCC-associated enhancers and recruits the BAF complex. Together, these
496 observations support previous studies that proposed that neural crest cell fate commitment is
497 initiated well before overt gene expression changes occur [65–67]. This finding is further supported
498 by clinical observations. In fact, phenotypes arising from *SALL4* haploinsufficiency affect many
499 neural crest derivatives, including craniofacial structures and peripheral neurons. In contrast,
500 phenotypes implicating defects in neural ectoderm or neurogenesis are exceptionally rare[68,69] and
501 usually associated with dysregulation of potassium channels[54]. In the rare cases of intellectual
502 disability reported, such phenotypes were attributed to broader deletions on chromosome
503 20q13.13-q13.2 rather than *SALL4* loss alone[68].

504 In the absence of SALL4, after two weeks of differentiation the cells acquire an aberrant fate,
505 expressing a mix of pluripotency genes (suggesting lack of cell differentiation), and neuronal
506 genes (suggesting derailed cell fate). This indicates that most cells safely reach the
507 neuroectodermal stage but fail to go beyond that, while some of the cells simply do not
508 differentiate or regress back to pluripotency. However, further experiments will be necessary to
509 further support this model.
510 Interestingly, only a small minority of the CNCC enhancers bound by SALL4 has the AT-rich motif
511 that is thought to be recognized by SALL4 [63,64]. This suggests that SALL4 might bind these
512 enhancers in cooperation with other lineage-specific transcription factors.
513 SALL1 also recognizes AT-rich motifs, but similar to SALL4 it also binds unspecifically, likely in
514 cooperation with other factors. Since SALL1 is upregulated in the *SALL4*-het-KO cells, it might
515 possible that some of the sites associated with a gain of BRG1 upon loss of SALL4 could be
516 associated with SALL1 being upregulated and hijacking BAF recruitment to distinct sites.
517 SALL4 was recently identified as a direct target of thalidomide[70], a drug widely prescribed to
518 pregnant women in the late 1950s to alleviate morning sickness. The subsequent alarming rise in
519 births of children with congenital anomalies, including limb malformations (phocomelia), facial
520 abnormalities such as inner ear defects, ocular motility issues, and facial nerve palsy, as well as
521 kidney and heart defects, led to the identification of thalidomide as the causative agent. However,
522 the molecular mechanism by which thalidomide induced these developmental defects remained
523 elusive until nearly six decades after the drug's withdrawal from the market. Building on earlier
524 findings that thalidomide and its derivatives bind to cereblon (CRBN), Matyskiela et al.
525 demonstrated that SALL4 is recruited to the CRL4–CRBN E3 ubiquitin ligase complex following
526 thalidomide-induced ubiquitination, resulting in its proteasomal degradation[53]. More recently,
527 research demonstrated that only the SALL4A isoform, but not SALL4B, is susceptible to
528 degradation by immunomodulatory imide drugs such as thalidomide and its derivative
529 pomalidomide[52]. Our data suggests that SALL4 role in CNCC specification is isoform-specific.
530 Herein, we show that pharmacological degradation of the SALL4A isoform using pomalidomide
531 replicated the Day-14 phenotype observed in *SALL4*-het-KO cells. These results indicate that
532 while SALL4B may contribute to other aspects of cellular identity or differentiation, it is insufficient
533 to drive CNCC fate commitment in the absence of SALL4A. In this context, our findings might
534 contribute to shed light on the birth malformations caused by thalidomide in the 1960s, potentially
535 revealing some of the underlying mechanisms. However, while we ensured that known
536 pomalidomide targets important for CNCC developments were not degraded in our system, we
537 cannot rule out degradation of previously uncharacterized and unknown pomalidomide targets

with key roles in CNCC specification. While highly unlikely, this is an important limitation of our study, and therefore the findings of our pomalidomide experiments) should be taken with caution.

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.

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Marco Trizzino (m.trizzino@imperial.ac.uk).

ACKNOWLEDGMENTS

We thank the Facility for Imaging by Light Microscopy (FILM) at Imperial College London for providing access to instrumentation and technical support, which is part-supported by funding from the Wellcome Trust (grant 104931/Z/14/Z) and BBSRC (grant BB/L015129/1). We also thank the Genomics Facility at The Wistar Institute for the Next Generation Sequencing support. For this work, MT was funded by the Mathers Foundation by Biotechnology and Biological Sciences Research Council (BBSRC, grant BB/Y000854/1).

We thank Prof. Brian Hendrich (University of Cambridge) and Dr. Andria Koulle (Imperial College London) for insightful discussion of the data and of early versions of the manuscript.

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.

DECLARATION OF INTERESTS

Authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Document S1. Supplemental figures

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

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

Table S3. List of TPMs and DEGs (RNA-seq) for Day-0, Day-5, and Day14.

Table S4. List of nearest genes to SALL4 CUT&RUN peaks and BRG1 ChIP-seq peaks.

572 **FIGURE TITLES AND LEGENDS**

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

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

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

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

582 (D) Zoomed view of SALL4 (pink) and DPF2 (purple) interacting surfaces. Highlighted are the
583 functional domains involved in the predicted interaction (ZF6 and ZF7 of SALL4 and PHD1 and
584 PHD2 of DPF2). In red, contacts predicted by AlphaFold[39].

585 (E) Box plot displaying expression (TPMs) of the four members of the DPF protein family (*DPF1*,
586 DPF2, *DPF3, and PHF10*) in iPSCs (n = 2). Differences between conditions were assessed using
587 unpaired student's t-test.

588 (F) Box plot displaying expression (TPMs) of the four members of the DPF protein family (*DPF1*,
589 DPF2, *DPF3, and PHF10*) in neuroectodermal-like cells (Day-5) (n = 3). Differences between
590 conditions were assessed using unpaired student's t-test.

591

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

593 (A) Visual representation of the CRISPR/Cas9 approach used on the *SALL4* gene model. The
594 star represents the exact location of the deletion generated by Cas9. Sanger sequencing
595 confirming the 11-bp deletion in one allele of *SALL4* is displayed. Made with BioRender.com.

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

598 (C) Bar plots showing densitometry analysis of SALL4A and SALL4B expression in WT and
599 *SALL4*-het-KO iPSC lines from immunoblot assay.

600 (D) Bar plots showing RNA expression in TPMs for SALL4A and SALL4B in WT and *SALL4*-het-
601 KO iPSC lines. Individual biological replicates (n = 2) are displayed as circles. Statistical
602 significance was assessed using unpaired student's t-test. Error bars display standard error of
603 the mean.

604 (E) Bar plots displaying expression TPMs of all *SALL* paralogues (*SALL1*, *SALL2*, *SALL3, SALL4*)
605 in *SALL4*-WT and *SALL4*-het-KO iPSCs. Individual biological replicates (n = 2) are displayed as

606 circles. Statistical significance was assessed using unpaired student's t-test. Error bars display
607 standard error of the mean.
608 (F) Immunoblot for SALL1 in *SALL4*-WT and *SALL4*-het-KO iPSC lines. GAPDH was used as a
609 loading control.
610 (G) Growth curve comparing *SALL4*-WT with *SALL4*-het-KO iPSC lines. N = 2 biologically
611 independent experiments. Error bars display standard error of the mean.
612
613 **Figure 3. Stem cell identity is unaffected by SALL4 loss**
614 (A) (Left) Immunofluorescence labeling quantifying the expression of the canonical pluripotency
615 markers OCT4 (top) and NANOG (bottom) in the two iPSC lines. Scale bar: 100 μm. (Right)
616 Corrected Total Cell Fluorescence (CTCF) values were used to quantify and compare markers
617 expression (n= 30). Error bars display standard error of the mean).
618 (B) Flow cytometry quantifying the expression of surface markers for pluripotency (TRA-60-1-R
619 and SSEA-4).
620 (C) Volcano plot of genes differentially expressed in *SALL4*-het-KO relative to *SALL4*-WT in
621 iPSCs (Day-0). Blue dots (n = 50) represent downregulated genes with p-adj < 0.05 and
622 log2FoldChange < -1.5. Purple dots (n = 63) represent upregulated genes with p-adj < 0.05 and
623 log2FoldChange > 1.5.
624
625 **Figure 4. SALL4 relocates to CNCC enhancers to induce chromatin accessibility**
626 (A) (Bottom) Schematic representation of the iPSC-to-CNCC differentiation protocol. The
627 experimental timeline shows in vitro differentiation of iPSCs into cranial neural crest cells
628 (CNCCs) via a neuroectodermal-like stage. Cells were cultured under self-renewing pluripotency
629 conditions until Day 0, at which point they were transitioned to N2B27 medium supplemented with
630 EGF, FGF, and insulin. Around Day 7–8, the medium was further modified by replacing insulin
631 with bovine serum albumin (BSA). Key analyses were performed at Day-0, Day-5, and Day-14.
632 (Top) Immunofluorescence labeling quantifying the expression of the canonical stage-specific
633 markers (iPSCs: OCT4; neuroectoderm: OTX2; CNCCs: SOX9) showing ratio of cells
634 successfully undergoing stage-specific specifications during CNCC generation. Scale bar: 100
635 μm.
636 (B) Diagram showing the number of regions bound by SALL4 uniquely at Day-0 ("Day-0 specific";
637 n = 6,114), at Day-5 ("Day-5 specific"; n = 17,137), and regions that maintain SALL4 binding at
638 both time points ("shared", n = 15,570).

639 (C) Heatmaps showing the distribution of SALL4 binding at Day-0 and Day-5 across Day-0
640 specific (left) and Day-5 specific (right) binding regions. CUT&RUN signal intensity was plotted
641 across ±2 kb regions centered on peak summits. Rows represent individual peaks ranked by
642 signal intensity.
643 (D) Genomic distribution of SALL4 binding sites relative to the nearest transcription start site
644 (TSS). Peaks are categorized by distance into six bins: <1 kb, 1–3 kb, 3–5 kb, 5–10 kb, 10–100
645 kb, and >100 kb. Peaks located at >1kb from the nearest TSS were classified as putative
646 enhancers. Percentages of peaks for each bin are represented by bars to the left (downstream)
647 or right (upstream) of the TSS.
648 (E) Gene ontology analysis of downregulated nearest genes to Day-5 specific SALL4-bound
649 regions as determined using WebGestalt[71] over-representation pathway analysis, using affinity
650 propagation as parameter for redundancy removal (FDR < 0.05).
651 (F) Table of motifs enriched in Day-5 specific SALL4-bound regions as determined using
652 HOMER[72].
653
654 **Figure 5. Chromatin accessibility at the CNCC enhancers is SALL4 dependent**
655 (A) Venn diagram showing chromatin accessibility in *SALL4*-WT (ATAC-seq peaks n = 58,375)
656 and *SALL4*-het-KO (n = 60,186). It also depicts the number of regions uniquely accessible in
657 *SALL4*-WT ("Sites losing accessibility in KO"; n = 1,649), in *SALL4*-het-KO ("Sites gaining
658 accessibility in KO"; n = 3,460), and regions that remain accessible in both conditions ("Sites
659 maintaining accessibility in WT and KO", n = 60,186).
660 (B) Heatmaps showing ATAC-seq signals at Day-0 (left) and Day-5 (right), at Day-5 specific
661 SALL4 CUT&RUN peaks. Signal intensity was plotted across ±2 kb regions centered on peak
662 summits. Rows represent individual peaks ranked by signal intensity.
663 (C) Average profiles showing ATAC-seq signal at Day 5-specific SALL4 CUT&RUN peaks (n =
664 17,137) in *SALL4*-WT and *SALL4*-het-KO cells in two distinct biological replicates.
665 (D) Table of motifs enriched in regions gaining or maintaining accessibility at Day-5 in *SALL4*-het-
666 KO regions as determined using HOMER[72].
667 (E) Table of motifs enriched in regions losing accessibility at Day-5 in *SALL4*-het-KO regions as
668 determined using HOMER[72].
669
670 **Figure 6. SALL4 recruits BAF to CNCC enhancers.**

671 (A) Heatmap showing BRG1 ChIP-seq signal at the genomic sites bound by SALL4 in WT cells
672 at Day-5 (n = 17,137). Signal intensity was plotted across ± 2 kb regions centered on peak
673 summits. Rows represent individual peaks ranked by signal intensity.
674 (B) Venn diagram showing peaks bound by SALL4 uniquely at Day-5 and that lose accessibility
675 in *SALL4*-het-KO.
676 (C) Heatmap and average profiles show BRG1 ChIP-seq signal at Day-5 at the BRG1 peaks lost
677 in *SALL4*-het-KO cells (n = 13,505). Signal intensity was plotted across ±2 kb regions centered
678 on peak summits. Rows represent individual peaks ranked by signal intensity.
679 (D) Heatmap and average profiles show SALL4 CUT&RUN signal at the genomic sites that lose
680 BRG1 in *SALL4*-het-KO cells at Day-0 and Day-5 (n = 13,505). Signal intensity was plotted across
681 ±2 kb regions centered on peak summits. Rows represent individual peaks ranked by signal
682 intensity.
683 (E) Heatmaps and average profile show ATAC-seq signal at the genomic sites that lose BRG1 in
684 *SALL4*-het-KO cells (n = 13,505). Signal intensity was plotted across ±2 kb regions centered on
685 peak summits. Rows represent individual peaks ranked by signal intensity.
686 (F) Two representative examples of enhancers bound by SALL4 at Day 5 that lose BRG1 co-
687 occupancy and chromatin accessibility in the *SALL4*-het-KO cells. Both correspond to ChIP-seq
688 peaks annotated as enhancer regions, as reported in GeneHancer (left) and VISTA Enhancer
689 Browser (right).
690
691 **Figure 7. CNCC induction and specification are impaired in *SALL4*-het-KO.**
692 (A) Bar plot displaying relative mRNA expression of canonical CNCC markers in *SALL4*-WT and
693 *SALL4*-het-KO cells at Day-14 (n = 2). Statistical significance was assessed using unpaired
694 student's t-test. 18S rRNA was used to normalize.
695 (B) Flow cytometry for the CNCC surface marker CD99 in *SALL4*-WT and *SALL4*-het-KO CNCCs
696 at Day-14.
697 (C) Volcano plot of genes differentially expressed in *SALL4*-het-KO relative to *SALL4*-WT in
698 CNCCs (Day-14). Blue dots (n= 659) represent downregulated genes with p-adj < 0.05 and
699 log2FoldChange < -1.5. Purple dots (n= 447) represent upregulated genes with p-adj < 0.05 and
700 log2FoldChange > 1.5.
701 (D) and (E) GO term enrichment for upregulated (D) and downregulated (E) genes in *SALL4*-het-
702 KO cells determined using WebGestalt[71] over-representation pathway analysis, using affinity
703 propagation as parameter for redundancy removal (FDR < 0.05).

704 (F) Heatmap displaying the expression of canonical epithelial-to-mesenchymal transition (EMT)
705 markers in *SALL4*-WT and *SALL4*-het-KO cells (Day-14).
706
707 **Figure 8. SALL4A is the only SALL4 isoform necessary for CNCC specification**

708 (A) Schematic of pomalidomide-mediated ubiquitin ligation. Ub, ubiquitin. POM, pomalidomide.
709 CRBN, DDB1, Cul4 and Rbx1 are subunits of CRL4/CRBN E3 ubiquitin ligase.
710 (B) Schematic representation of pomalidomide treatment during iPSC-to-CNCC differentiation
711 protocol.
712 (C) Timecourse immunoblot for SALL4 in *SALL4*-WT at Day-5 and Day-14 after daily treatment
713 with DMSO (control) or 5μM pomalidomide (5μM POM). GAPDH was used as a loading control.
714 (D) Box plot displaying the relative mRNA expression of canonical CNCC markers (SOX9,
715 TFAP2A, and NR2F1) in *SALL4*-WT at Day-14 treated with 5μM pomalidomide or DMSO (control)
716 daily for 14 days. Individual biological replicates (n = 2) are displayed as circles. Statistical
717 significance was assessed using unpaired student's t-test. Error bars display standard error of
718 the mean.
719 (E) Immunofluorescence labeling quantifying the expression of the canonical CNCC markers
720 SOX9 and TFAP2A in *SALL4*-WT treated with 5μM pomalidomide or DMSO (control) daily for 14
721 days. Scale bar: 100 μm.
722 (F) Immunofluorescence labeling quantifying the expression of the canonical EMT markers
723 EpCAM and VIM in *SALL4*-WT treated with 5uM pomalidomide or DMSO (control) daily for 14
724 days. Scale bar: 100 μm.
725

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

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

763 embryonic development by the transcriptional regulation of Pou5f1. Nat Cell Biol *8*, 1114–
764 1123. https://doi.org/10.1038/ncb1481.

765 14. Yang, J. (2018). SALL4 as a transcriptional and epigenetic regulator in normal and
766 leukemic hematopoiesis. Preprint at BioMed Central Ltd., https://doi.org/10.1186/s40364-
767 017-0115-6 https://doi.org/10.1186/s40364-017-0115-6.

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

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

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

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

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

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

791 21. Dash, S., and Trainor, P.A. (2020). The development, patterning and evolution of neural
792 crest cell differentiation into cartilage and bone. Preprint at Elsevier Inc.,
793 https://doi.org/10.1016/j.bone.2020.115409 https://doi.org/10.1016/j.bone.2020.115409.

794 22. Szabó, A., Szabó, S., and Mayor, R. (2018). Mechanisms of Neural Crest Migration.
795 https://doi.org/10.1146/annurev-genet-120417.

796 23. Milet, C., and Monsoro-Burq, A.H. (2012). Neural crest induction at the neural plate border
797 in vertebrates. Preprint at Academic Press Inc.,
798 https://doi.org/10.1016/j.ydbio.2012.01.013 https://doi.org/10.1016/j.ydbio.2012.01.013.

24. Roth, D.M., Bayona, F., Baddam, P., and Graf, D. (2021). Craniofacial Development: Neural Crest in Molecular Embryology. Preprint at Springer, https://doi.org/10.1007/s12105-021-01301-z https://doi.org/10.1007/s12105-021-01301-z.

25. Pla, P., and Monsoro-Burq, A.H. (2018). The neural border: Induction, specification and maturation of the territory that generates neural crest cells. Dev Biol *444*, S36–S46. https://doi.org/10.1016/J.YDBIO.2018.05.018.

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

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

28. 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*. https://doi.org/10.1371/journal.pone.0005577.

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

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

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

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

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

34. Barnada, S.M., Giner de Gracia, A., Morenilla-Palao, C., López-Cascales, M.T., Scopa, C., Waltrich, F.J., Mikkers, H.M.M., Cicardi, M.E., Karlin, J., Trotti, D., et al. (2024). ARID1A-BAF coordinates ZIC2 genomic occupancy for epithelial-to-mesenchymal transition in

838 cranial neural crest specification. The American Journal of Human Genetics.
839 https://doi.org/10.1016/j.ajhg.2024.07.022.

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

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

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

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

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

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

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

865 42. Lessard, J., Wu, J.I., Ranish, J.A., Wan, M., Winslow, M.M., Staahl, B.T., Wu, H.,
866 Aebersold, R., Graef, I.A., and Crabtree, G.R. (2007). An Essential Switch in Subunit
867 Composition of a Chromatin Remodeling Complex during Neural Development. Neuron *55*,
868 201–215. https://doi.org/10.1016/j.neuron.2007.06.019.

869 43. Nazim, M., Lin, C.H., Feng, A.C., Xiao, W., Yeom, K.H., Li, M., Daly, A.E., Tan, X., Vu, H.,
870 Ernst, J., et al. (2024). Alternative splicing of a chromatin modifier alters the transcriptional
871 regulatory programs of stem cell maintenance and neuronal differentiation. Cell Stem Cell
872 *31*, 754-771.e6. https://doi.org/10.1016/j.stem.2024.04.001.

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

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

46. Kosicki, M., Baltoumas, F.A., Kelman, G., Boverhof, J., Ong, Y., Cook, L.E., Dickel, D.E., Pavlopoulos, G.A., Pennacchio, L.A., and Visel, A. (2025). VISTA Enhancer browser: an updated database of tissue-specific developmental enhancers. Nucleic Acids Res *53*, D324–D330. https://doi.org/10.1093/nar/gkae940.

47. Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Stein, T.I., Rosen, N., Kohn, A., Twik, M., Safran, M., et al. (2017). GeneHancer: Genome-wide integration of enhancers and target genes in GeneCards. Database *2017*. https://doi.org/10.1093/database/bax028.

48. Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–285. https://doi.org/10.1038/nature09692.

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

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

51. 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. Preprint, https://doi.org/10.1101/2024.01.03.574032 https://doi.org/10.1101/2024.01.03.574032.

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

53. 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. https://doi.org/10.1038/s41589-018-0129-x.

54. Kodytková, A., Dušátková, P., Anne Amaratunga, S., Plachý, L., Pruhová, Š., and Lebl, J. (2023). Integrative role of the sall4 gene: From thalidomide embryopathy to genetic defects of the upper limb, internal organs, cerebral midline, and pituitary. Preprint at S. Karger AG, https://doi.org/10.1159/000531452 https://doi.org/10.1159/000531452.

55. Sievers, Q.L., Petzold, G., Bunker, R.D., Renneville, A., Słabicki, M., Liddicoat, B.J., Abdulrahman, W., Mikkelsen, T., Ebert, B.L., and Thomä, N.H. (2018). Defining the human

916 C2H2 zinc finger degrome targeted by thalidomide analogs through CRBN. Science (1979)
917 *362*. https://doi.org/10.1126/science.aat0572.

918 56. Yamamoto, J., Suwa, T., Murase, Y., Tateno, S., Mizutome, H., Asatsuma-Okumura, T.,
919 Shimizu, N., Kishi, T., Momose, S., Kizaki, M., et al. (2020). ARID2 is a pomalidomide-
920 dependent CRL4CRBN substrate in multiple myeloma cells. Nat Chem Biol *16*, 1208–
921 1217. https://doi.org/10.1038/s41589-020-0645-3.

922 57. Jourdeuil, K., Neilson, K.M., Tavares, A.L.P., and Moody, S.A. (2025). Zmym2 Alters
923 Expression of Key Craniofacial Genes. Preprint at John Wiley and Sons Inc,
924 https://doi.org/10.1002/dvg.70018 https://doi.org/10.1002/dvg.70018.

925 58. Renneville, A., Gasser, J.A., Grinshpun, D.E., Jean Beltran, P.M., Udeshi, N.D.,
926 Matyskiela, M.E., Clayton, T., McConkey, M., Viswanathan, K., Tepper, A., et al. (2021).
927 Avadomide Induces Degradation of ZMYM2 Fusion Oncoproteins in Hematologic
928 Malignancies. Blood Cancer Discov *2*, 251–265. https://doi.org/10.1158/2643-3230.BCD-
929 20-0105.

930 59. Miller, M.T. (1991). Thalidomide embryopathy: a model for the study of congenital
931 incomitant horizontal strabismus. Transaction of the American Ophtalmological Society,
932 623–674.

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

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

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

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

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

953 65. Stuhlmiller, T.J., and García-Castro, M.I. (2012). FGF/MAPK signaling is required in the
954 gastrula epiblast for avian neural crest induction. Development *139*, 289–300.
955 https://doi.org/10.1242/dev.070276.

956 66. Basch, M.L., Bronner-Fraser, M., and García-Castro, M.I. (2006). Specification of the
957 neural crest occurs during gastrulation and requires Pax7. Nature *441*, 218–222.
958 https://doi.org/10.1038/nature04684.

959 67. Patthey, C., Gunhaga, L., and Edlund, T. (2008). Early development of the central and
960 peripheral nervous systems is coordinated by Wnt and BMP signals. PLoS One *3*.
961 https://doi.org/10.1371/journal.pone.0001625.

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

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

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

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

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

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

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

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

989 76. Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S.,
990 Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-

991 sequencing data analysis. Nucleic Acids Res *44*, W160–W165.
992 https://doi.org/10.1093/NAR/GKW257.

993 77. Herbreteau, G., Airaud, F., Pierre-Noël, E., Vallée, A., Bézieau, S., Théoleyre, S., Blons,
994 H., Garinet, S., and Denis, M.G. (2021). MEM: An algorithm for the reliable detection of
995 microsatellite instability (MSI) on a small NGS panel in colorectal cancer. Cancers (Basel)
996 *13*. https://doi.org/10.3390/cancers13164203.

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

1000 79. Polit, L., Kerdivel, G., Gregoricchio, S., Esposito, M., Guillouf, C., and Boeva, V. (2021).
1001 CHIPIN: ChIP-seq inter-sample normalization based on signal invariance across
1002 transcriptionally constant genes. BMC Bioinformatics *22*. https://doi.org/10.1186/s12859-
1003 021-04320-3.

1004
1005

1006 **METHODS**

1007 **KEY RESOURCES TABLE**

1008

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

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

TFAP2A	5'-GCCTCTCGCTCCTCAGCTCC-3'	5'-CGTTGGCAGCTTTACGTCTCCC-3'
NR2F1	5'-ATCGTGCTGTTCACGTCAGA-3'	5'-GCTCCTCACGTACTCCTCCA-3'
SOX9	5'-GTACCCGCACTTGCACAAC-3'	5'-TCTCFCTCTCGTTCAGAAGTC-3'
Software and algorithms		
AlphaFold3	https://alphafoldserver.com/	SCR_025454
TrimGalore!	https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/	SCR_011847
FlowJo Software version 10.9		SCR_008520
BEDTools		SCR_006646
Deeptools		SCR_016366
R Studio version 4.4.3		
GraphPad Prism 10.2.3	http://www.graphpad.com/faq/viewfaq.cfm?faq=1362	SCR_002798
DESeq2		SCR_015687

METHOD DETAILS

AlphaFold3 structure prediction

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

1021 **Human iPSC culture**

1022 Wild-type iPSC lines were obtained from the iPSC Core of the University of Pennsylvania (Control
1023 line-1: SV20 line, male, age 43). iPSC lines were cultured in feeder-free, serum-free mTeSR Plus
1024 medium (STEMCELL Technologies) with 1% Penicillin-Streptomycin (5,000 U/mL; Gibco
1025 #15140122) on tissue culture plates coated overnight with Geltrex™ LDEV-Free hESC-qualified
1026 Reduced Growth Factor Basement Membrane Matrix (Fisher-Scientific). Cells were passaged
1027 ~1:8 at 60-80% confluency using 0.5mM EDTA to create 50-200 cell clusters.

1028

1029 **CRISPR/Cas9-mediated SALL4 knock-out**

1030 The heterozygous *SALL4* knock-out in the human SV20 iPSC line was generated via
1031 CRISPR/Cas9 gene-editing by the biotech company Synthego. Specifically, two clones were
1032 used. In both clones Cas9 targeted the first exon of the *SALL4* gene in the established SV20 iPSC
1033 line producing an 11-base deletion (c.230delGGGCCGCATCT; clone C4) and a 19-base deletion
1034 (c.del226GCTGGGGCCGCATCT; clone F1) respectively, which in both cases generated a
1035 premature stop codon. The two clones were used as biological replicates in all experiments.

1036

1037 **Trilineage differentiation**

1038 iPSCs were differentiated into the three germ layers (endoderm, ectoderm and mesoderm) using
1039 the Human Pluripotent Stem Cell Functional Identification Kit (R&D Systems, SC027B). The
1040 expression of relevant markers was then assessed using immunofluorescence with 10 ug/ml of
1041 antibodies provided in the kit (Goat anti-human SOX17, Goat anti-human OTX2 and, Goat anti-
1042 human TBXT).

1043

1044 **iPSC-to-CNCC differentiation**

1045 All iPSC lines were differentiated into CNCCs as previously described (Prescott et al., 2015).
1046 When iPS cells reached ~60% confluence, they were dissociated and treated with CNCC
1047 differentiation media (1:1 Neurobasal medium/D-MEM F-12 medium (Invitrogen), 0.5× B-27
1048 supplement with Vitamin A (50× stock, Gibco, #17504044), 0.5× N-2 supplement (100× stock,
1049 Gibco; #17502001), 20 ng/ml human FGF (Gibco), 20 ng/ml recombinant EGF (Sigma-Aldrich),
1050 5 µg/ml bovine insulin (Sigma-Aldrich) and 1× Glutamax-I supplement (100× stock, Invitrogen).
1051 Medium (2 ml) was changed every other day. Around day 7-8 of differentiations, cells were
1052 transitioned to CNCC early maintenance media (1:1 Neurobasal medium/D-MEM F-12 medium
1053 (Invitrogen), 0.5× B-27 supplement with Vitamin A (50× stock, Invitrogen), 0.5× N-2 supplement
1054 (100× stock, Invitrogen), 20 ng/ml bFGF (Biolegend), 20 ng/ml EGF (Sigma-Aldrich), 1 mg/ml

1055 bovine serum albumin, serum replacement grade (Gemini Bio-Products # 700-104 P) and 1×
1056 Glutamax-I supplement (100× stock, Invitrogen). Differentiating cells were passaged ~1:6 with
1057 0.5mM EDTA or Accutase Cell Detachment Solution (Stemcell Technologies; #07920), plated in
1058 a fresh, Geltrex-coated 6-well plates. After passaging, cells were treated with 0.05% ROCK-
1059 inhibitor (Y27632; Stemcell Technologies) for 24 hours.
1060

1061 **Real-time quantitative polymerase chain reaction (RT-qPCR)**

1062 Cells were harvested as single cells, and the total RNA was extracted using Monarch® Total RNA
1063 Miniprep Kit (New England Biolabs) following manufacturer's instructions. 600ng of total RNA was
1064 used to synthesize cDNA using Maxima First Strand cDNA Synthesis Kit for RT-qPCR (K1641,
1065 Thermo Fisher). For each RT-qPCR reaction, 15 ng of cDNA was used with 0.1 µM of each primer,
1066 10 µL of PowerUp™ SYBR™ Green Master Mix (Applied Biosystems) in a final volume of 20 µl.
1067 Samples were analyzed with CFX Connect Real-Time PCR Detection System (Bio-Rad) using
1068 the following thermal cycling parameters: 3 min at 95 °C, followed by 40 cycles of 10 s at 95 °C,
1069 20 s at 63 °C followed by 30 s at 72 °C. Each sample was run in triplicate. 18 S rRNA was used
1070 as normalizer.
1071

1072 **Nuclear Fractionation and Co-immunoprecipitation**

1073 Cells were washed twice with ice-cold PBS, scraped, and collected by centrifugation at 10,000 ×
1074 g for 1 minute. Pellets were resuspended in 500 µL of Buffer A (10mM HEPES pH7.9, 1.5mM
1075 $MgCl_2$, 10mM KCl, 1mM PMSF, 0.5mM DTT, protease inhibitor cocktail) and cytoplasmic fractions
1076 were obtained by incubating the cells on ice for 15 minutes, adding of 0.05% NP-40, followed by
1077 centrifugation at 17,000 × g for 10 minutes at 4 °C. The resulting pellet was gently washed with
1078 Buffer A and resuspended in Buffer C (10mM HEPES pH7.9, 12.5% glycerol, 0.75mM $MgCl_2$,
1079 10mM KCl, 0.1mM EDTA, 400mM NaCl, 1mM PMSF, 0.5mM DTT, protease inhibitor cocktail).
1080 Samples were incubated at 4 °C with shaking at 3,400 rpm for 1 hour, with vortexing every 10
1081 minutes. Nuclear fractions were collected by centrifugation at 17,000 × g for 10 minutes at 4 °C.
1082 For co-immunoprecipitation (co-IP), 25 µL of Dynabeads™ Protein A (ThermoFisher, Cat#
1083 10002D) were washed twice with PBS containing 0.1% Tween-20 (PBS-T), resuspended in 1 mL
1084 of PBS-T, and incubated with 2 µg of the appropriate antibody for 1 hour at 4 °C with rotation.
1085 Bead-antibody complexes were washed twice with PBS-T, resuspended in 125 µL PBS-T, and
1086 incubated with 100 µg of nuclear extract for 3 hours at 4 °C with rotation. Beads were then washed
1087 five times with Buffer C and eluted in 25 µL of Buffer C supplemented with 4.1 µL of 6× reducing
1088 Laemmli SDS sample buffer (ThermoFisher Scientific). Eluates and 5% input controls (nuclear

1089 extract diluted 1:10 in PBS and mixed with 1:6 volume of 6× Laemmli buffer) were boiled for 5
1090 minutes prior to SDS-PAGE. For Western blotting, 20 µL of each sample were loaded as
1091 described below.
1092
1093 **Western Blot**
1094 For each condition, cells were harvested and lysed with cold radioimmunoprecipitation assay
1095 buffer (Pierce™ RIPA buffer, Thermo Fisher) with Pierce™ Protease Inhibitor Mini Tablets, EDTA-
1096 free (Thermo Fisher). After centrifugation at 4°C at 16,000x g for 10 minutes, the total extracted
1097 protein was quantified using Pierce™ BCA Protein Assay Kit, following manufacturer's
1098 indications. Thirty micrograms of total protein were loaded in Novex WedgeWell 4–20% Tris-
1099 Glycine Gel (Invitrogen) and separated through gel electrophoresis (SDS–PAGE) Tris-Glycine-
1100 SDS buffer (Invitrogen). The proteins were then transferred to nitrocellulose membrane (Thermo
1101 Fisher Scientific) and incubated for one hour in Intercept® (PBS) Blocking Buffer (Licor). After
1102 blocking, membranes were incubated overnight at 4°C (while rocking) in suitable antibody diluted
1103 in blocking buffer and 0.2% Tween-20, at recommended concentrations. Membranes were
1104 washed in 0.1% PBS-T before and after being incubated for one hour in the appropriate secondary
1105 antibody (1:15,000 dilution). The antibodies were then visualized using Odyssey XF Imaging
1106 System (Licor) or ChemiDocTM MP Imaging System and imaged with Image Studio Software
1107 (Licor) or Fiji[73].
1108
1109 **Flow Cytometry analysis of surface markers**
1110 To obtain a single-cell suspension for flow cytometry analysis, cells were treated with Accutase
1111 for 5 min at 37°C. Cells were then washed with cold phosphate-buffered saline (PBS)-2% fetal
1112 bovine serum (FBS) and live cells were counted. 1×10^6 cells/condition were resuspended in
1113 100 µL PBS-2% FBS and stained. For pluripotency evaluation, 4 µl of the respective antibodies
1114 were used. For analysis of differentiation, 2 µl of the respective antibodies were used. Cells were
1115 incubated for 15 min on ice and protected from light, before transferring them into FACS tubes
1116 containing an additional 300 µL PBS-2% FBS. Flow cytometry data were acquired using
1117 NovoCyte Penteon Flow Cytometer Systems 5 Lasers and analyzed with FlowJo Software version
1118 10.9.
1119
1120 **Immunofluorescence**
1121 Cells were passaged onto Geltrex-coated coverslips and fixed at required confluence or time-
1122 point. In detail, cells were incubated in 4% PFA at 37°C for 15 minutes, followed by three 2-minute

washes in 1× PBS. For permeabilization, cells were treated with 0.1% (v/v) Triton X-100 (T8787, Sigma-Aldrich) at room temperature for 10 minutes before three 2-minute washes in 1× PBS + 0.1% (v/v) Tween 20 and a 1-hour blocking upon 10% Donkey serum (D9663, Sigma-Aldrich) at room temperature. Cells were incubated with corresponding PBS-diluted primary antibodies protected from direct light at 4°C overnight. After three 5-minute washes with 1× PBS + 0.1% Tween 20, cells were incubated for 1 hour in PBS-diluted secondary antibodies and 15 minutes with DAPI (1:5,000; 422801, BioLegend). Cells were then washed three times with 1× PBS + 0.1% Tween for 5 minutes and once with 1× PBS. Coverslips were then fetched and mounted onto slides using Fluorescence Mounting Medium (S3023, Agilent Technologies). Slides were imaged using confocal microscope (LSM780, Zeiss). The intensity of fluorescent signal was measured using Corrected Total Cell Fluorescence (CTCF) using Image J. Briefly, for each cell, a region of interest (ROI) was drawn around the cell boundary, and the Integrated Density (sum of all pixel intensity values within the ROI), area of ROI (total number of pixels contained in the ROI), and the mean background fluorescence (average pixel intensity measured from a background region of the image lacking fluorescent signal) were measured.

The corrected fluorescence intensity was then calculated as:

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

RNA-seq

Total RNA was extracted using Monarch® Total RNA Miniprep Kit (New England Biolabs, T2010S) according to the manufacturer's instructions. RNA was quantified using NanoDrop 2000 UV Visible Spectrophotometer, while the RNA integrity was checked on Tapestation 2200 (Agilent). Only samples with RIN value above 8.0 were used for transcriptome analysis. RNA libraries were prepared using 1 µg of total RNA input using NEBNext® Poly(A) mRNA Magnetic Isolation Module, NEBNext® UltraTM II Directional RNA Library Prep Kit for Illumina® according to the manufacturer's instructions (New England Biolabs).

RNA-sequencing analysis

After removing the adapters with TrimGalore! (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) , Kallisto was used to quantify reads mapping to each gene. Differential gene expression levels were determined using DESeq2[74]. All statistical analyses were performed using the latest versions of BEDTools[75], deepTools[76], R Studio (version 4.4.3; Trophy Case), and GraphPad Prism software version 10.2.3 (http://www.graphpad.com/faq/viewfaq.cfm?faq=1362; GraphPad, San Diego, CA).

1157 **CUT&RUN**

1158 For each condition, 1x 10^5 cells were processed using CUT&RUN Assay Kit #86652 (Cell
1159 Signaling Technology). Briefly, each sample was resuspended in Concanavalin A Beads and
1160 treated with an Antibody Binding Buffer and the suitable antibody. After the antibody incubation,
1161 each reaction underwent a step of pAG-MNase Enzyme binding, and subsequent DNA Digestion
1162 and diffusion. The diffused DNA was then purified using Phenol/Chloroform Extraction and
1163 Ethanol Precipitation. The purified DNA was quantified using QuantiFluor dsDNA System, using
1164 the ONE dsDNA protocol (Promega). Libraries for sequencing were prepared using NEBNext
1165 Ultra II DNA Library Prep Kit for Illumina (New England BioLabs) and the libraries size distribution
1166 was assessed through automated electrophoresis using the Agilent 2200 Tapestation System
1167 (Agilent Technology).

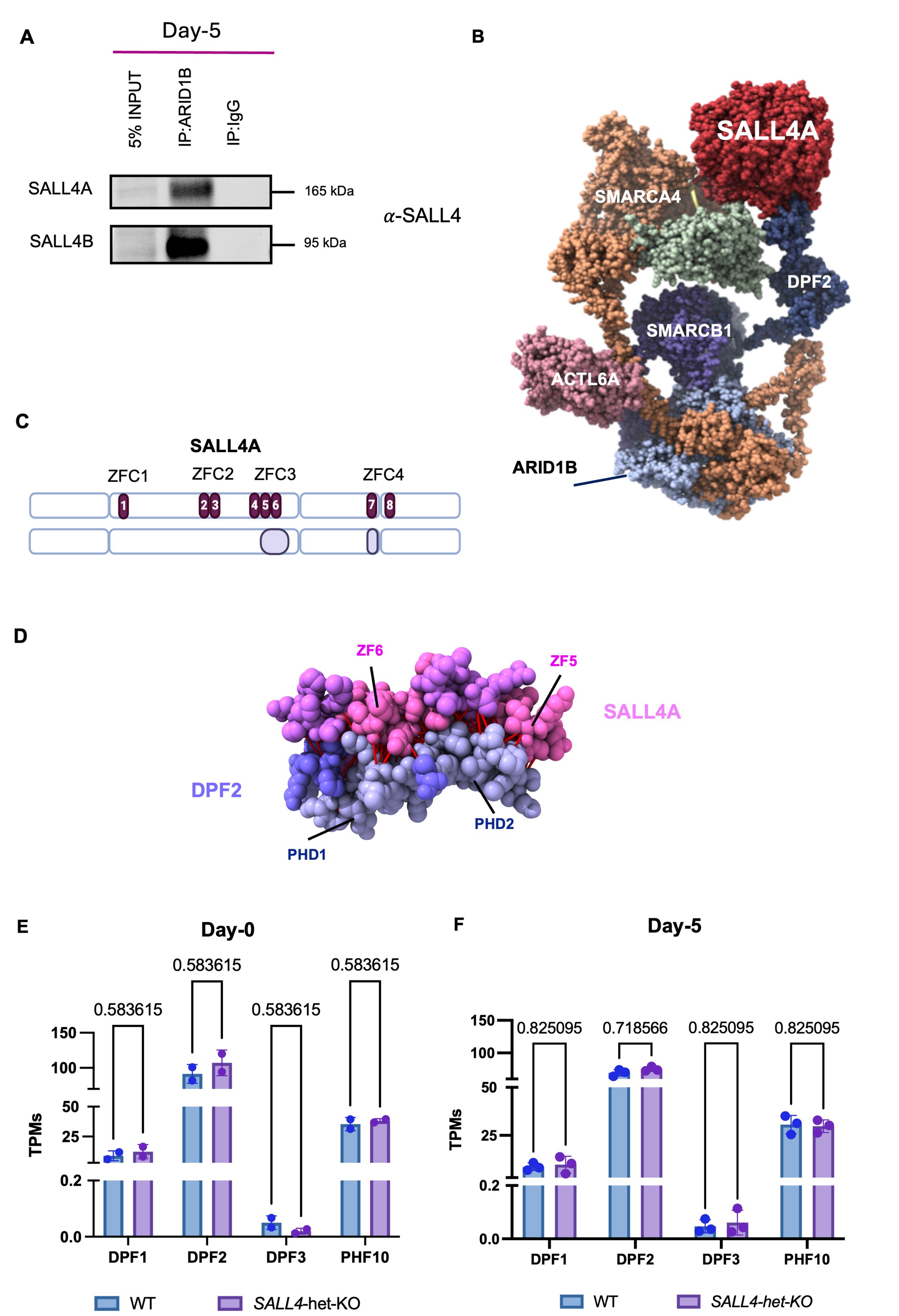
1168

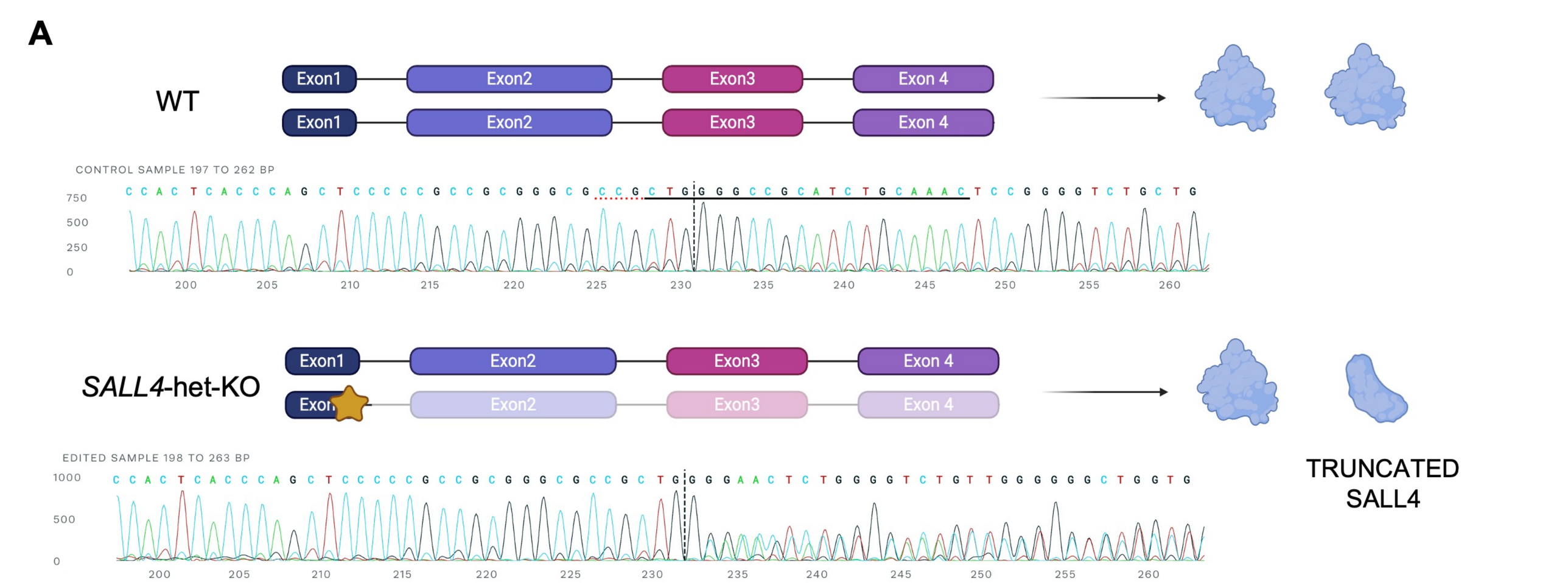
1169 **ChIP-seq**

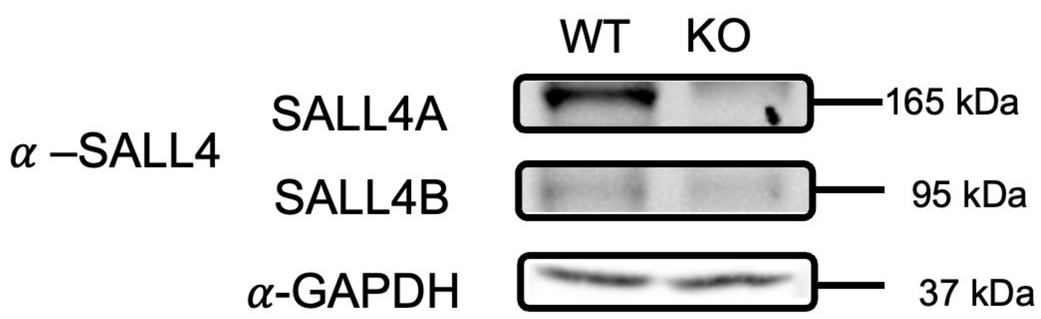
1170 All samples from different conditions were processed together to prevent batch effects.
1171 Approximately 15 million cells were double cross-linked with 2mM disuccinimidyl glutarate (DSG)
1172 for 20 minutes followed by 1% formaldehyde for 10 minutes at room temperature, quenched with
1173 125 mM glycine, harvested, and washed twice with 1X PBS. The fixed cell pellet was resuspended
1174 in ChIP lysis buffer (150 mM NaCl, 1% Triton X-100, 0.7% SDS, 500 μM DTT, 10 mM Tris-HCl, 5
1175 mM EDTA) and chromatin was sheared to an average length of 200–900 base-pairs, using a
1176 Covaris S220 Ultrasonicator at 5% duty factor for 7 minutes for 3 times. The chromatin lysate was
1177 diluted with SDS-free ChIP lysis buffer. 10μg of antibody was used for BRG1. The antibody was
1178 added to at least 5μg of sonicated chromatin along with Dynabeads Protein G magnetic beads
1179 (Invitrogen) and incubated with rotation at 4°C overnight. The beads were washed twice with each
1180 of the following buffers: Mixed Micelle Buffer (150 mM NaCl, 1% Triton X- 100, 0.2% SDS, 20 mM
1181 Tris-HCl, 5 mM EDTA, 65% sucrose), Buffer 200 (200 mM NaCl, 1% Triton X-100, 0.1% sodium
1182 deoxycholate, 25 mM HEPES, 10 mM Tris-HCl, 1 mM EDTA), LiCl detergent wash (250 mM LiCl,
1183 0.5% sodium deoxycholate, 0.5% NP-40, 10 mM Tris-HCl, 1 mM EDTA) and a final wash was
1184 performed with cold 0.1X TE. Finally, beads were resuspended in 1X TE containing 1% SDS and
1185 incubated at 65°C for 10 min to elute immunocomplexes. The elution was repeated twice, and the
1186 samples were incubated overnight at 65°C to reverse cross-linking, along with the input (5% of
1187 the starting material). The DNA was digested with 0.5 mg/ml Proteinase K for 3 hours at 55°C
1188 and then purified using the ChIP DNA Clean & Concentrator kit (Zymo) and quantified with QUBIT.
1189 Barcoded libraries were made with NEBNext Ultra II DNA Library Prep Kit for Illumina) using

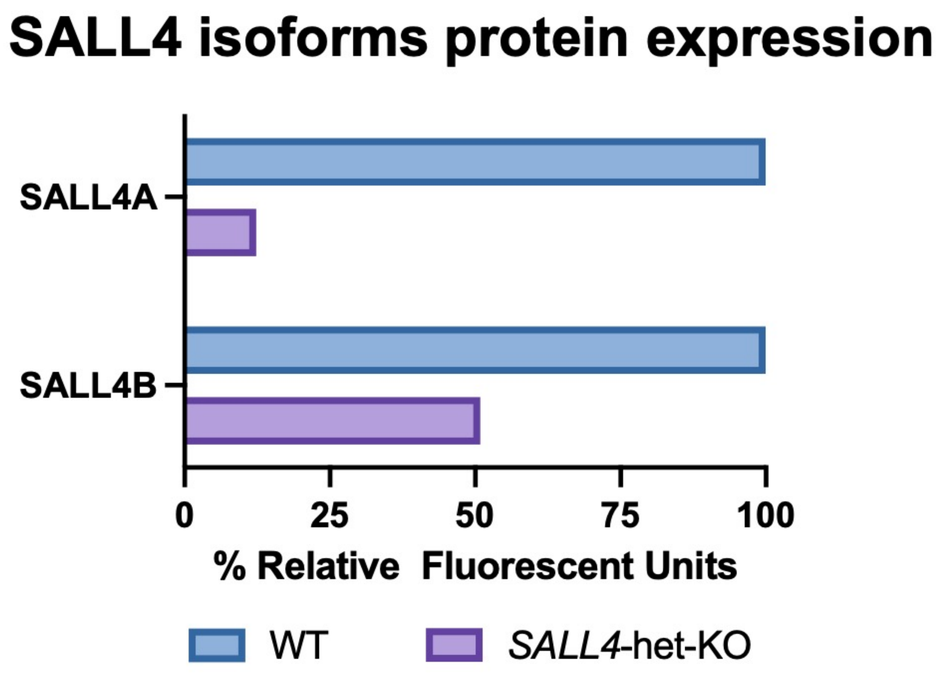
1190 NEBNext Multiplex Oligos Dual Index Primers for Illumina (New England BioLabs) and sequenced
1191 on NovaSeq X Plus instrument (Illumina).
1192
1193 **ATAC-seq**
1194 One hundred thousand cells/condition were harvested by enzymatic dissociation using
1195 StemPro™ Accutase™ (Thermo Fisher Scientific). Each cell pellet was processed using ATAC-
1196 Seq Kit (Active Motif), following the manufacturer's directions. Briefly, the fresh cell pellets were
1197 washed in 1X PBS and lysate in ATAC Lysis Buffer. Each lysate underwent a process of
1198 tagmentation and purification, before being amplified through PCR, using the following thermal
1199 cycler parameters: 72°C for 5 min, 98°C for 30 sec, followed by 10 cycles of 98°C for 10 sec,
1200 63°C for 30 sec, 72°C for 1 min. Each PCR reaction was cleaned up using SPRI beads (1.2X of
1201 the sample volume). The Agilent 2200 Tapestation System (Agilent Technologies) was used as
1202 quality control tool before performing next generation sequencing (NGS).
1203
1204 **Bioinformatics analysis of CHIP-seq, ATAC-seq and CUT&RUN data**
1205 Adapters were removed with TrimGalore!
1206 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), the sequences were aligned
1207 to the reference hg19, using Burrows-Wheeler Alignment tool, with the MEM algorithm[77]. Uniquely
1208 mapping aligned reads were filtered based on mapping quality (MAPQ > 10) to restrict our
1209 analysis to higher quality and uniquely mapped reads, and PCR duplicates were removed.
1210 MACS2[78] was used to call peaks using the default parameters at 5% FDR.
1211
1212 **Motif analysis**
1213 BED files for the regions of interest were produced using BEDTools[75]. Motif analysis was
1214 performed using HOMER[72]. Shuffled input sequences were used as background. E-values <
1215 0.001 were used as a threshold for significance.
1216
1217 **QUANTIFICATION AND STATISTICAL ANALYSIS**
1218 RT-qPCR and image quantification data were analyzed using Prism software version 10.2.3
1219 (http://www.graphpad.com/faq/viewfaq.cfm?faq=1362; GraphPad, San Diego, CA). Data is
1220 represented as mean ± standard error of mean (SEM). GO enrichment analysis was performed
1221 using Webgestalt 2019[71]. All statistical analyses were performed using the latest versions of
1222 BEDTools[75], Deeptools[76], R version 4.4.3 (Throphy Case; https://www.R-project.org/), and Prism
1223 software version 10.2.3 (http://www.graphpad.com/faq/viewfaq.cfm?faq=1362; GraphPad, San

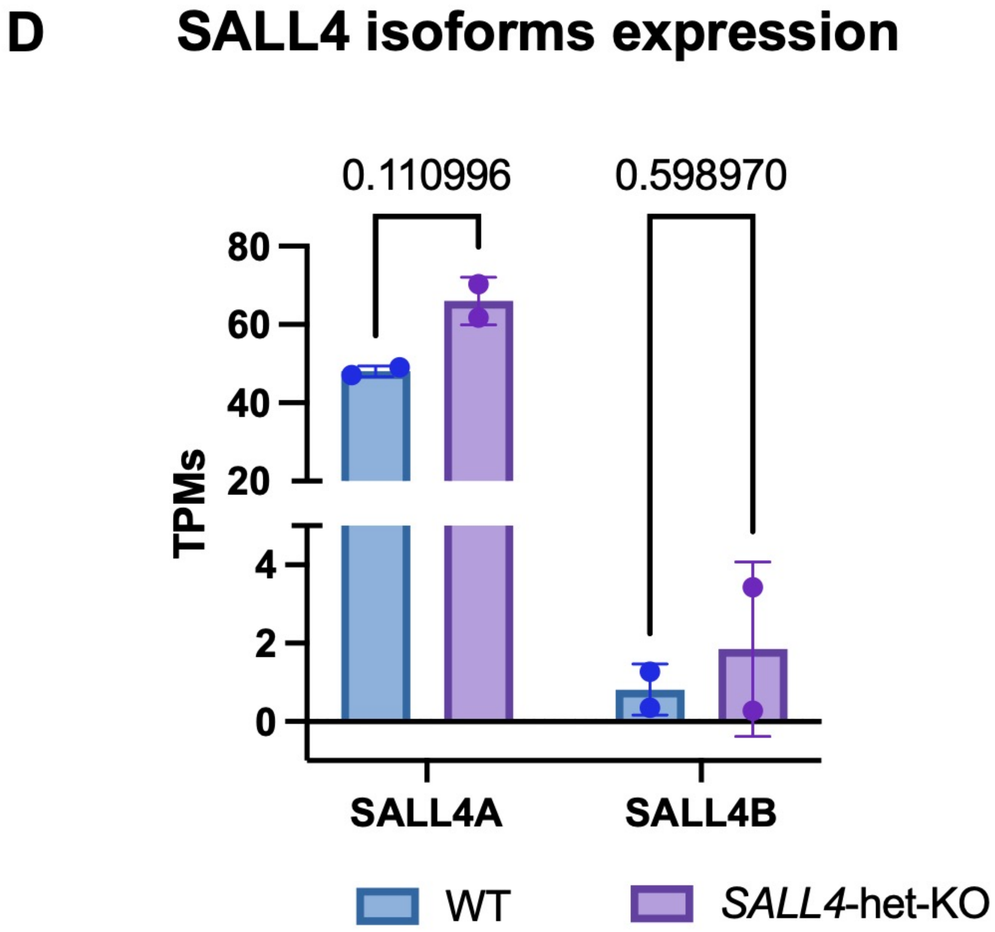
1224 Diego, CA). In particular, Bigwigs were generated with Deeptools and normalized by sequencing
1225 depth using the same software, and upper-quantile normalized using CHIPIN[79].

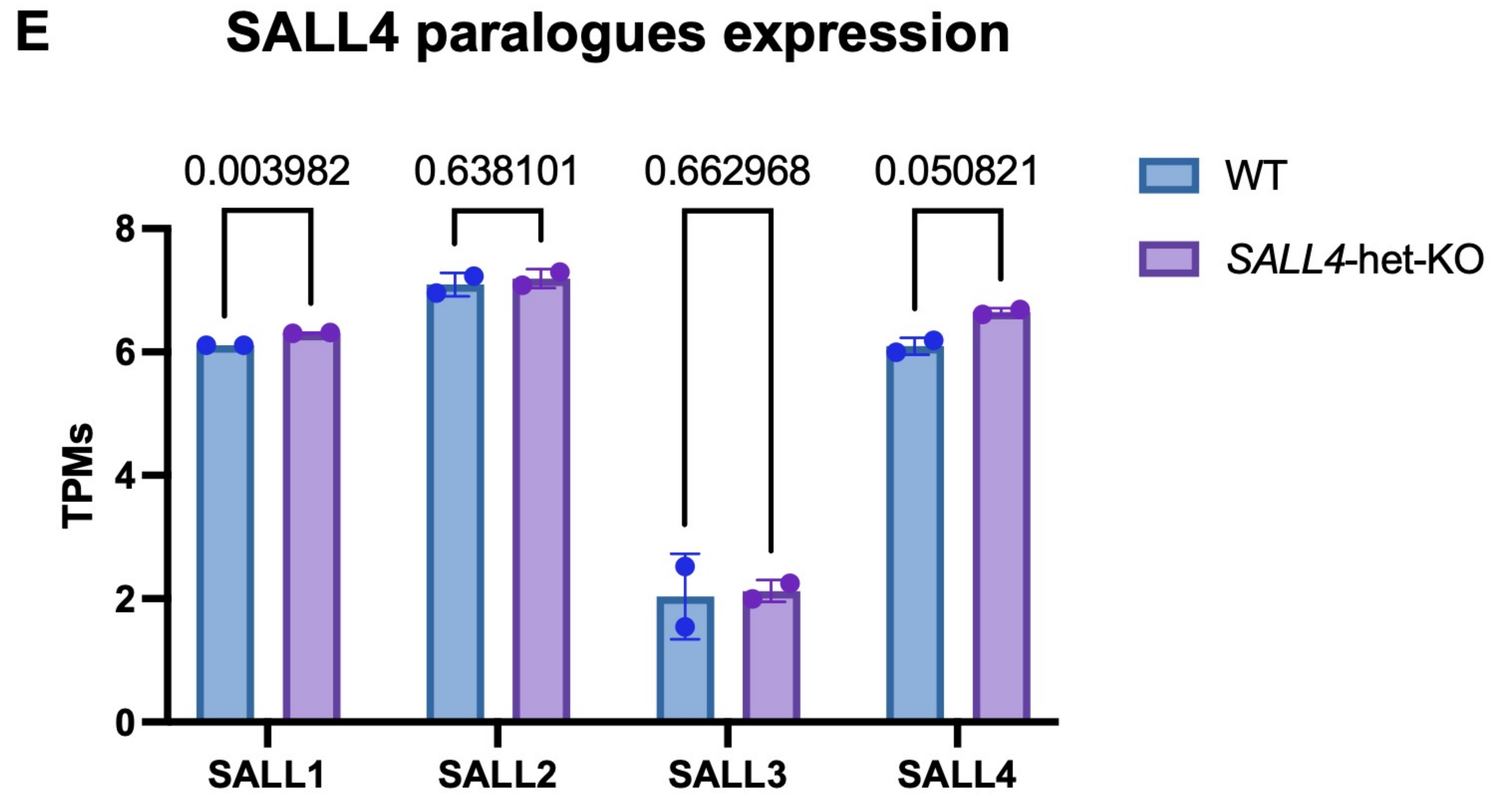

A
Day-5
5% INPUT
IP:ARID1B
IP:IgG
SALL4A
SALL4B
165 kDa
95 kDa
α-SALL4
B
SALL4A
SMARCA4
DPF2
SMARCB1
ACTL6A
ARID1B
C
SALL4A
ZFC1
ZFC2
ZFC3
ZFC4
1
2 3
4 5 6
7
8
D
ZF6
ZF5
SALL4A
DPF2
PHD2
PHD1
E
Day-0
0.583615
0.583615
0.583615
TPMs
150
100
50
25
0.2
0.0
DPF1
DPF2
DPF3
PHF10
WT
SALL4-het-KO
F
Day-5
0.825095
0.718566
0.825095
0.825095
TPMs
150
100
50
25
0.2
0.0
DPF1
DPF2
DPF3
PHF10
WT
SALL4-het-KO

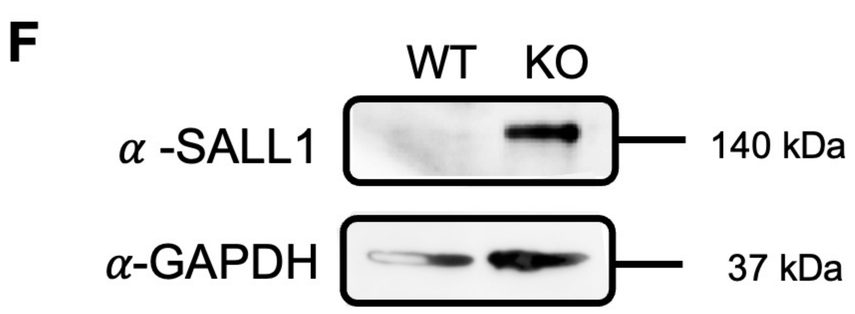

A
WT
Exon1
Exon2
Exon3
Exon 4
CONTROL SAMPLE 197 TO 262 BP
SALL4-het-KO
EDITED SAMPLE 198 TO 263 BP
TRUNCATED SALL4

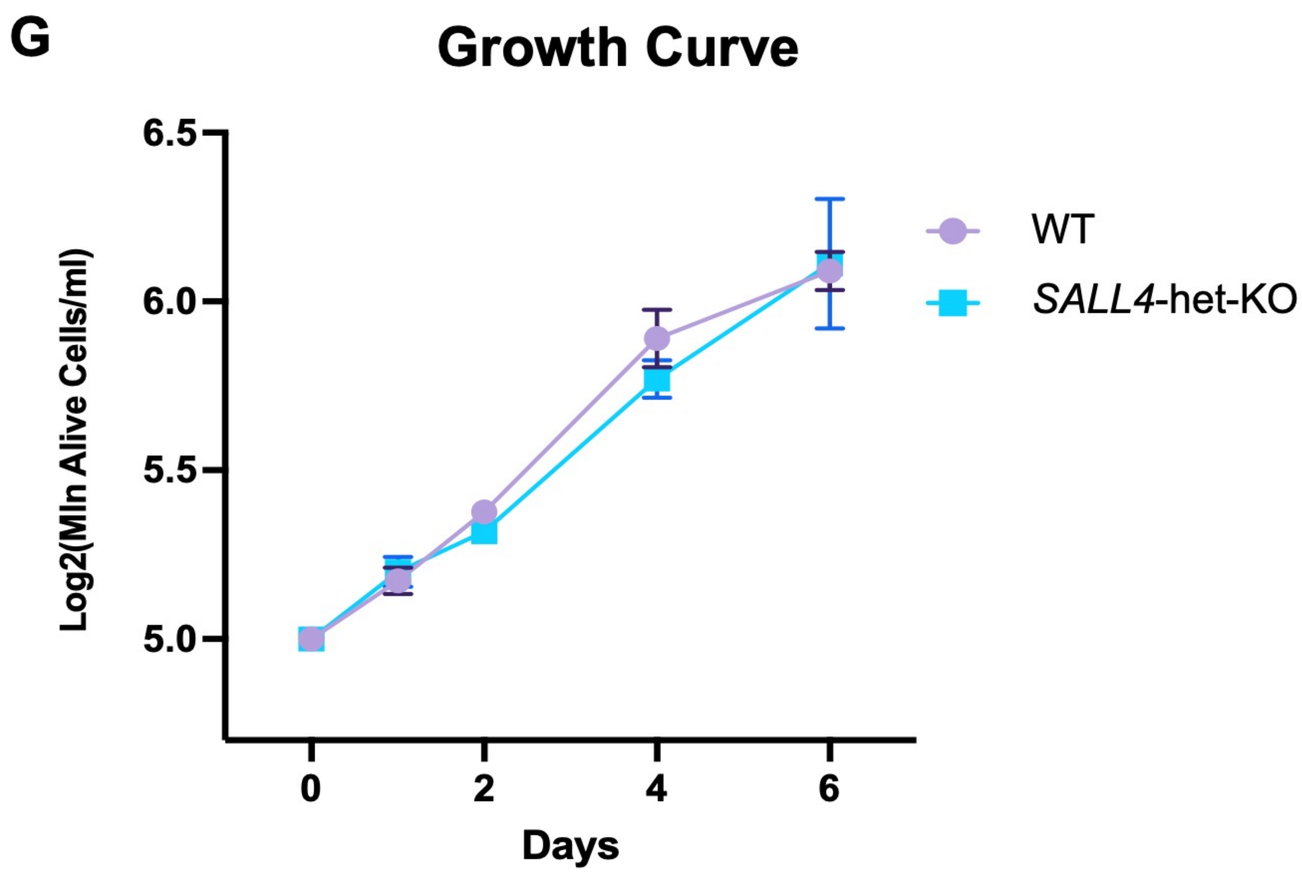

B
WT
KO
α –SALL4
SALL4A
SALL4B
α-GAPDH
165 kDa
95 kDa
37 kDa

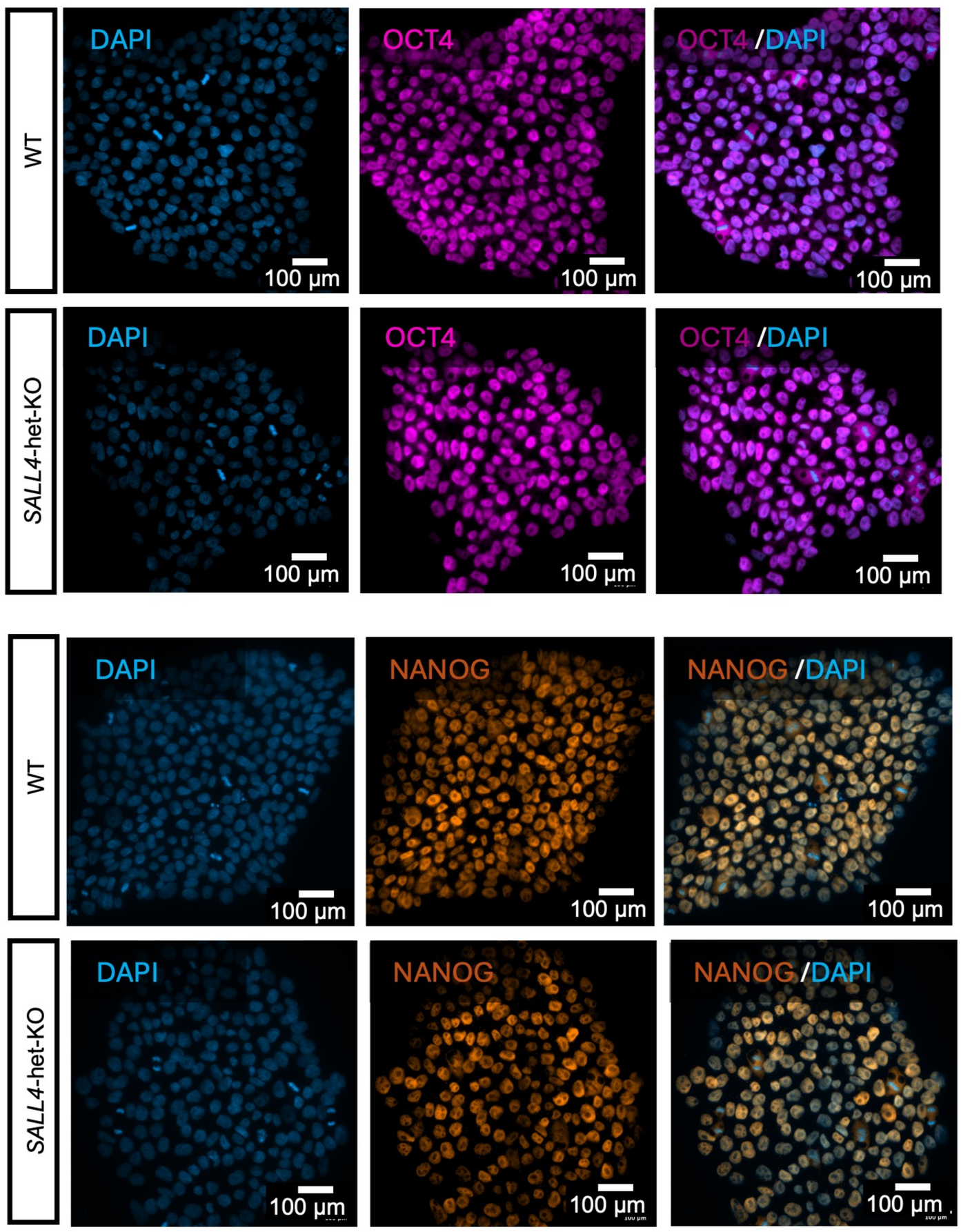

C
SALL4 isoforms protein expression
SALL4A
SALL4B
0
25
50
75
100
% Relative Fluorescent Units
WT
SALL4-het-KO

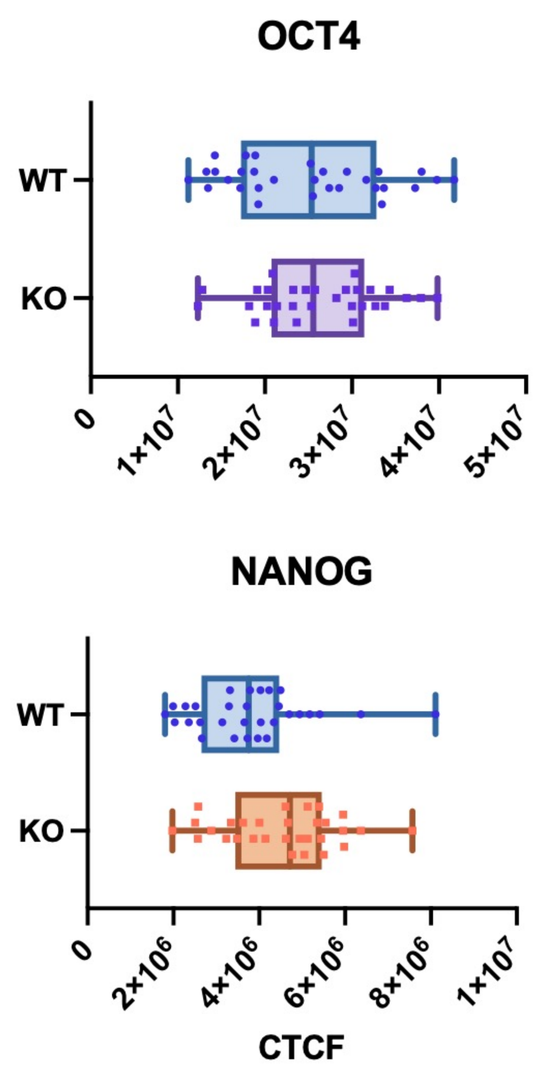

D
SALL4 isoforms expression
0.110996
0.598970
TPMs
80
60
40
20
4
2
0
SALL4A
SALL4B
WT
SALL4-het-KO

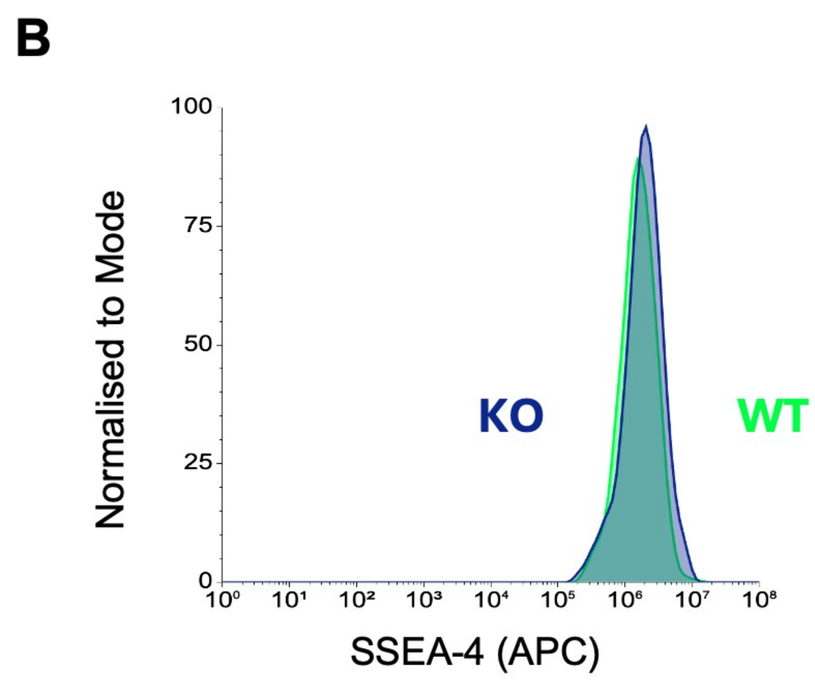

E
SALL4 paralogues expression
0.003982
0.638101
0.662968
0.050821
TPMs
8
6
4
2
0
SALL1
SALL2
SALL3
SALL4
WT
SALL4-het-KO

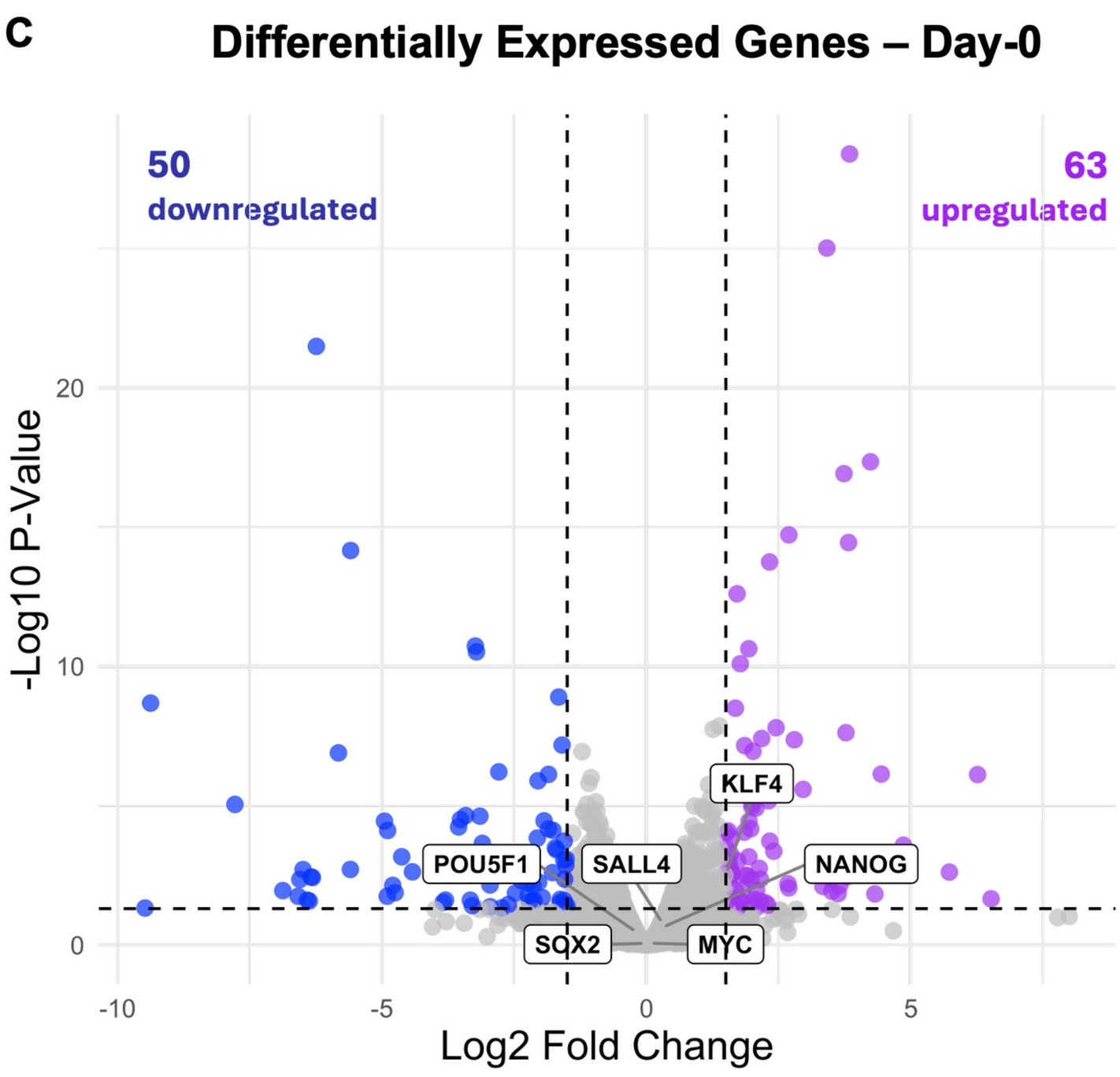

F
WT
KO
α -SALL1
α-GAPDH
140 kDa
37 kDa

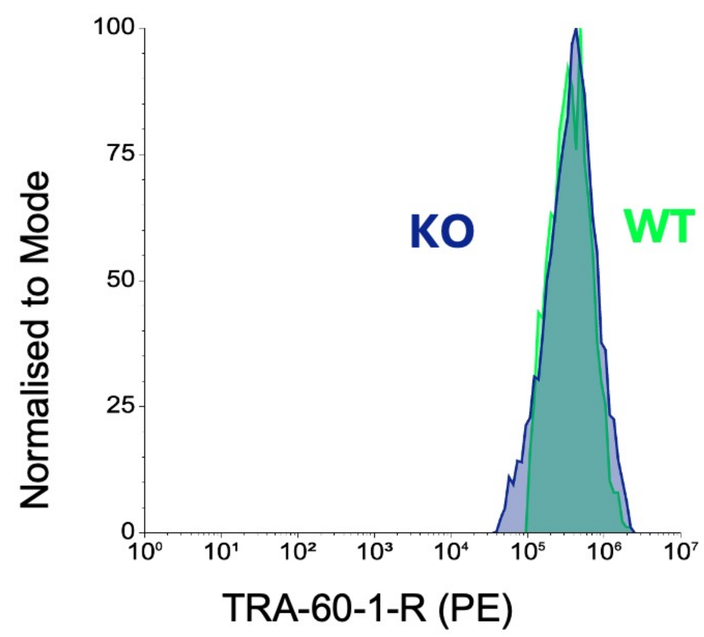

G
Growth Curve
Log2(Mln Alive Cells/ml)
6.5
6.0
5.5
5.0
0
2
4
6
Days
WT
SALL4-het-KO

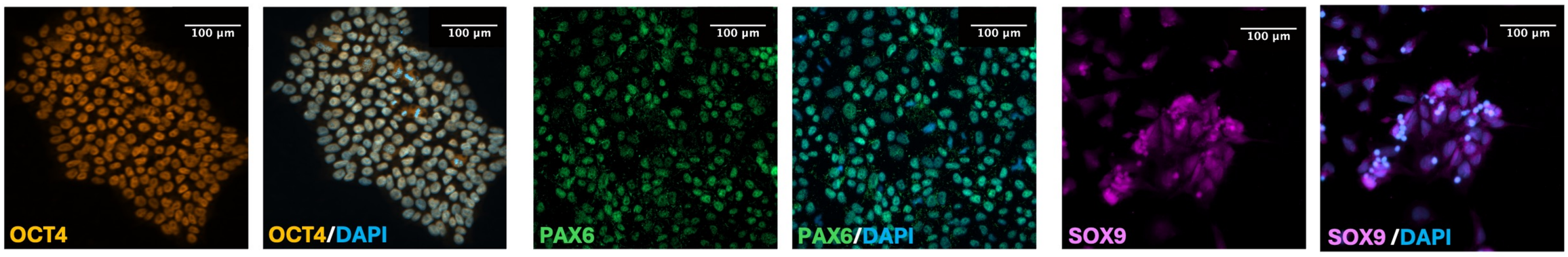

A
WT
SALL4-het-KO
DAPI
OCT4
OCT4 /DAPI
NANOG
NANOG /DAPI
100 µm

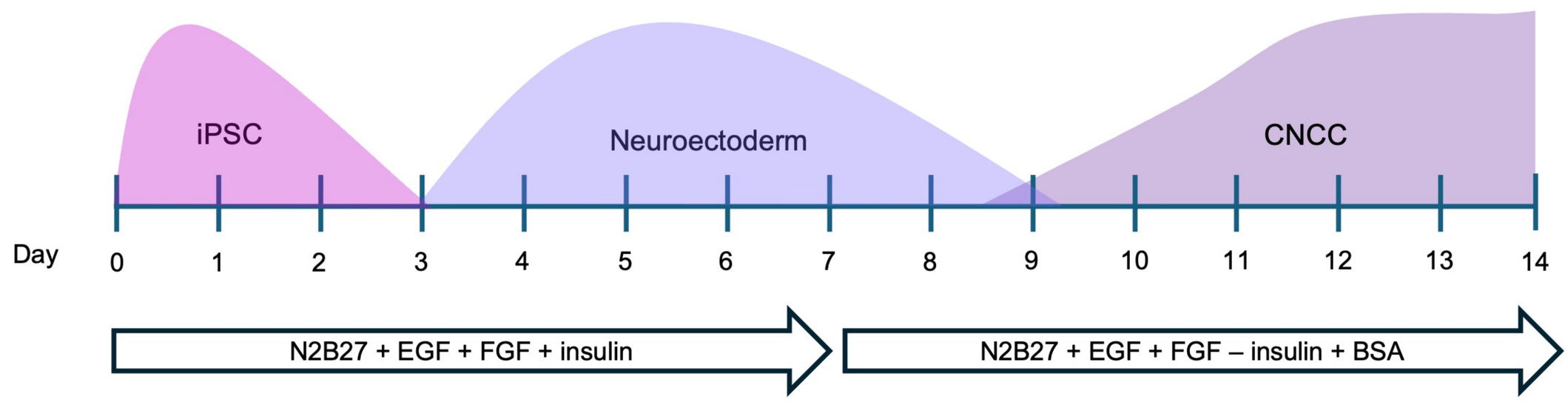

OCT4
NANOG
WT
KO
0
1×10^7
2×10^7
3×10^7
4×10^7
5×10^7
2×10^6
4×10^6
6×10^6
8×10^6
1×10^7
CTCF

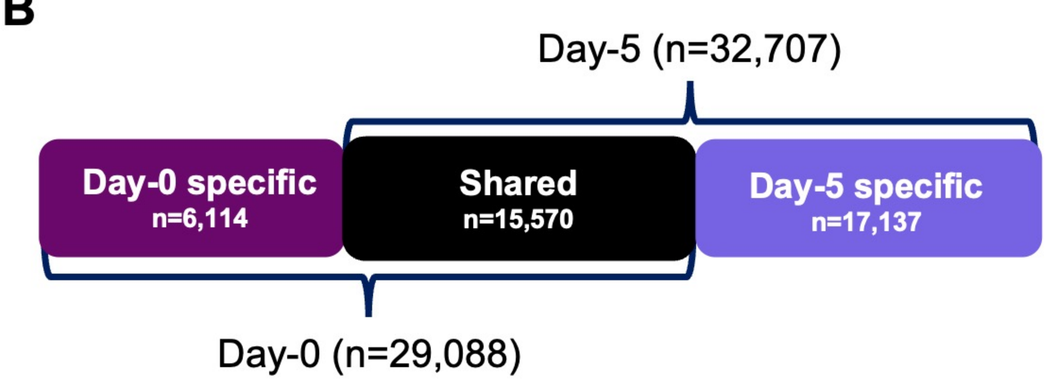

B
Normalised to Mode
KO
WT
SSEA-4 (APC)

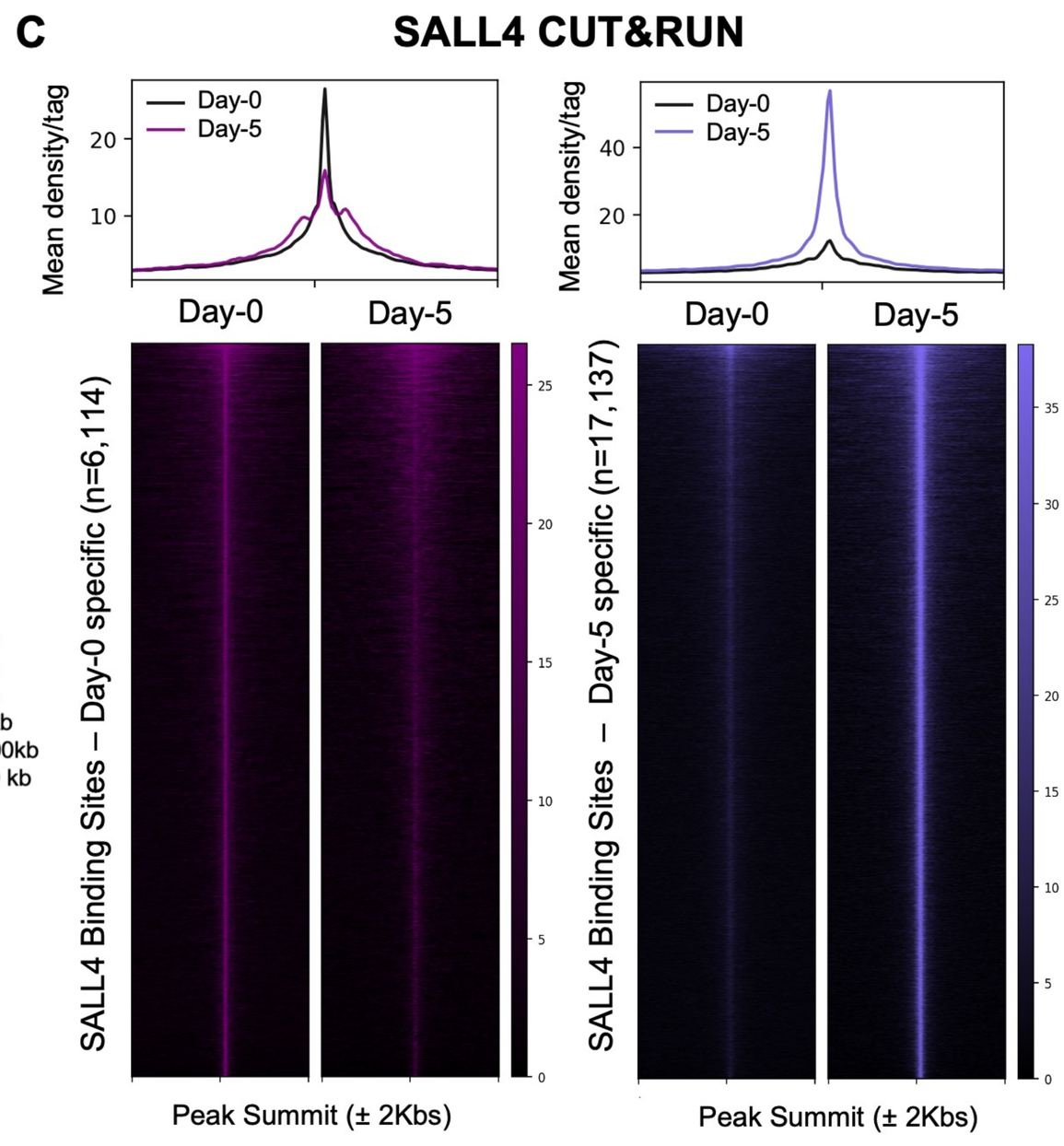

C
Differentially Expressed Genes – Day-0
50
downregulated
63
upregulated
-Log10 P-Value
Log2 Fold Change
KLF4
POU5F1
SALL4
NANOG
SOX2
MYC

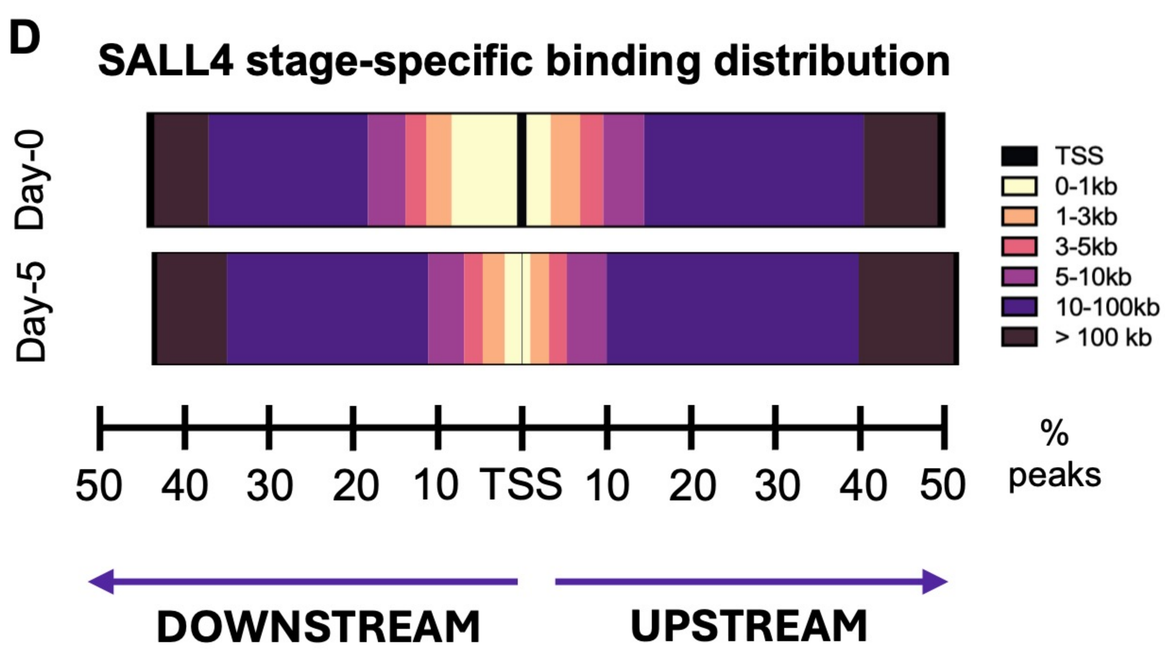

Normalised to Mode
KO
WT
TRA-60-1-R (PE)

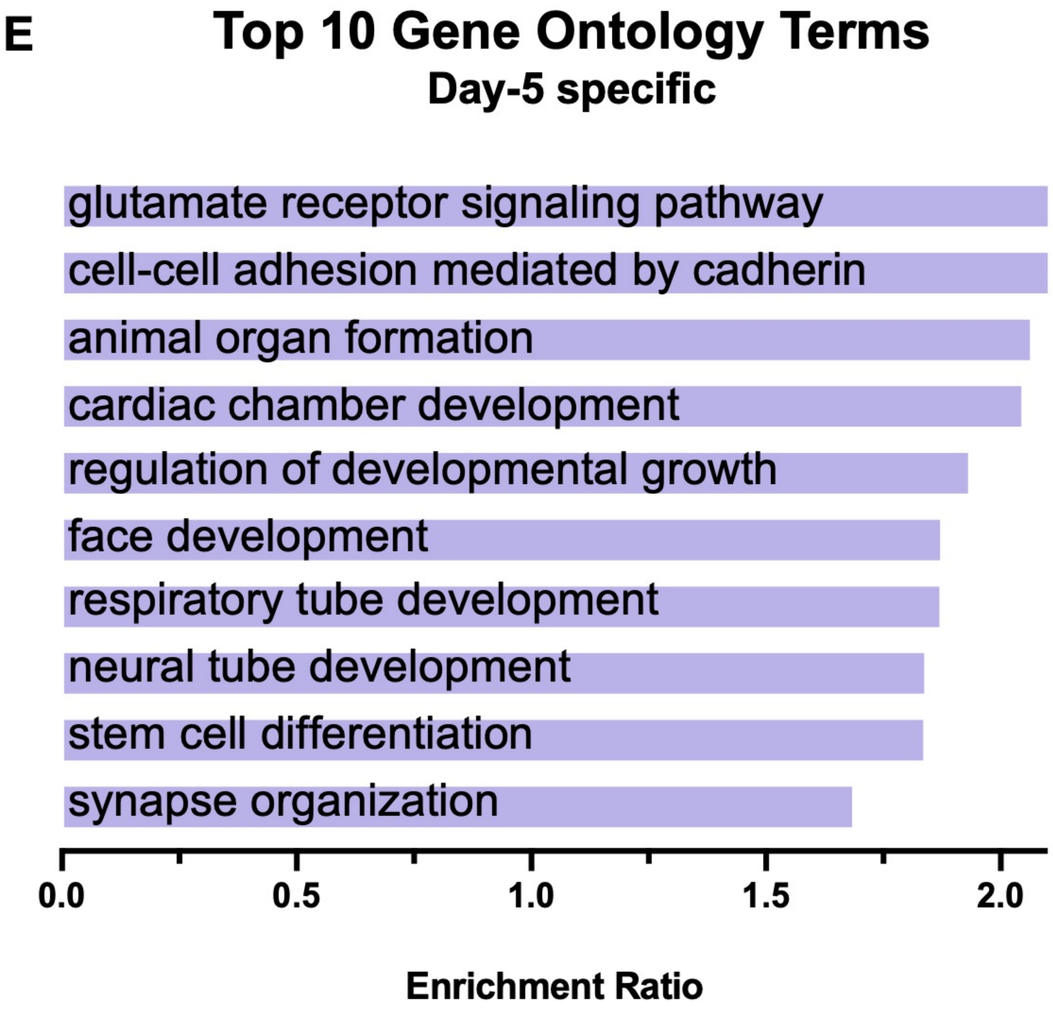










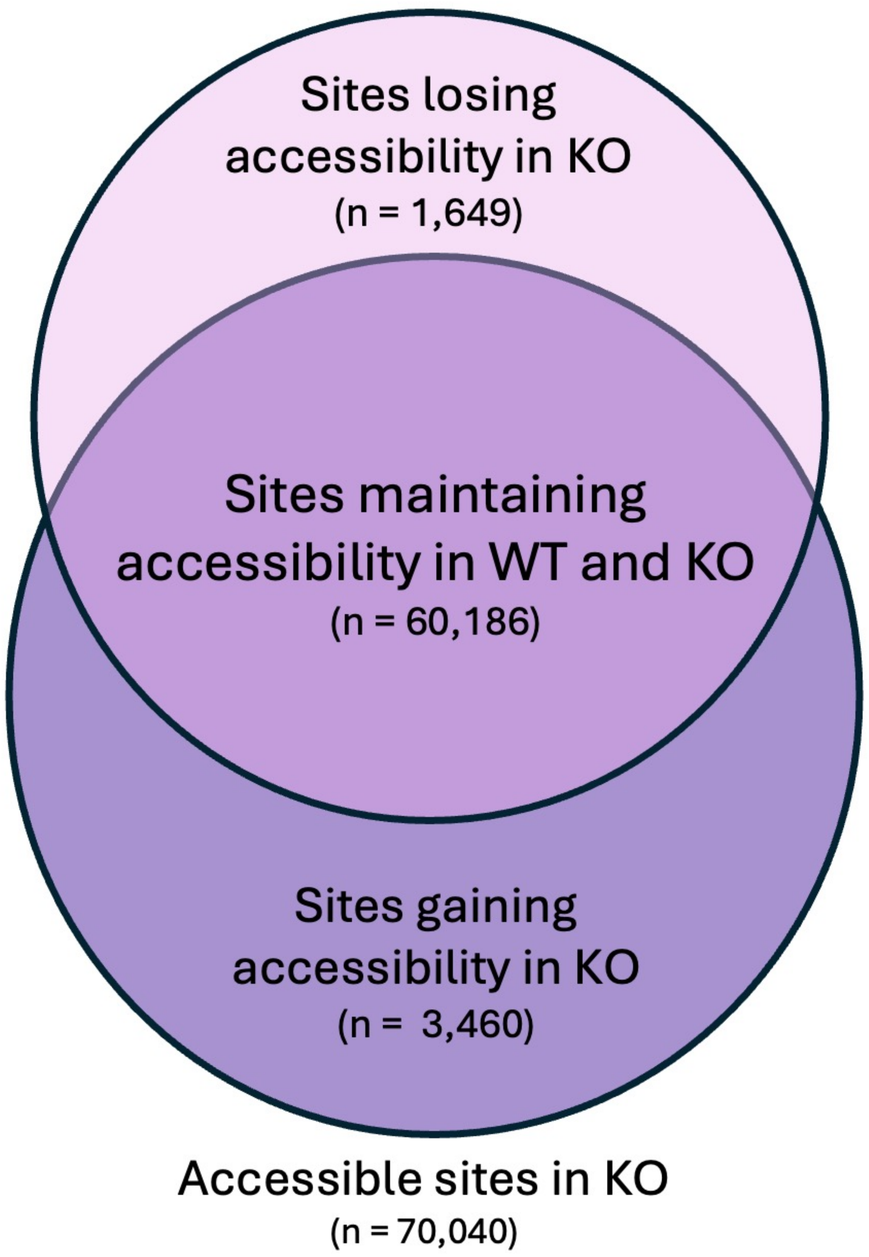


F

SALL4 CUT&RUN Day-5 specific motifs			
MOTIF #	Motif Consensus		p-value
380		TEAD	1e-1238
13		JUN/FOS/AP1 FOS/FRA	1e-969
256		OCT4-SOX2	1e-933
343		SOX9 SOX10	1e-328
409		ZIC NANOG	1e-329
391		TWIST TCF	1e-96

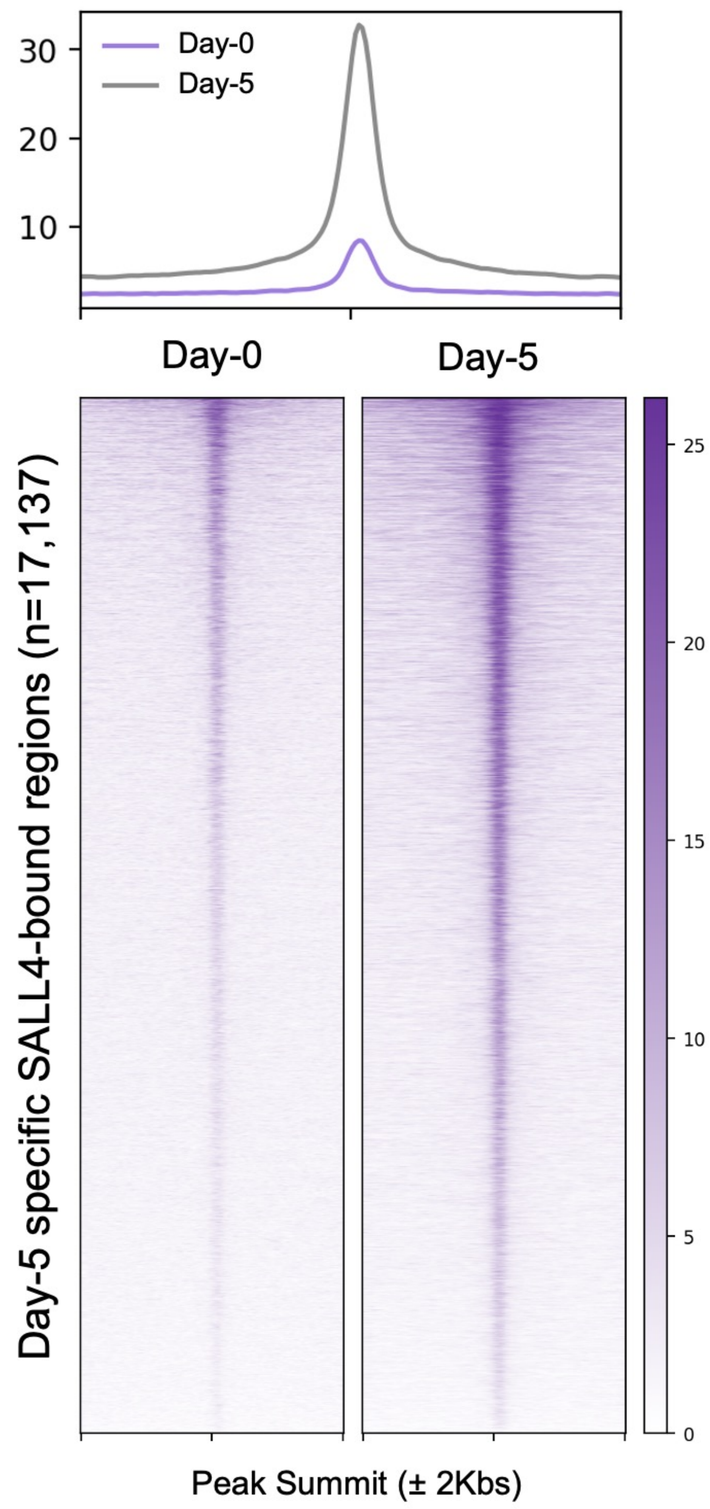
A

ATAC-seq peaks at Day 5



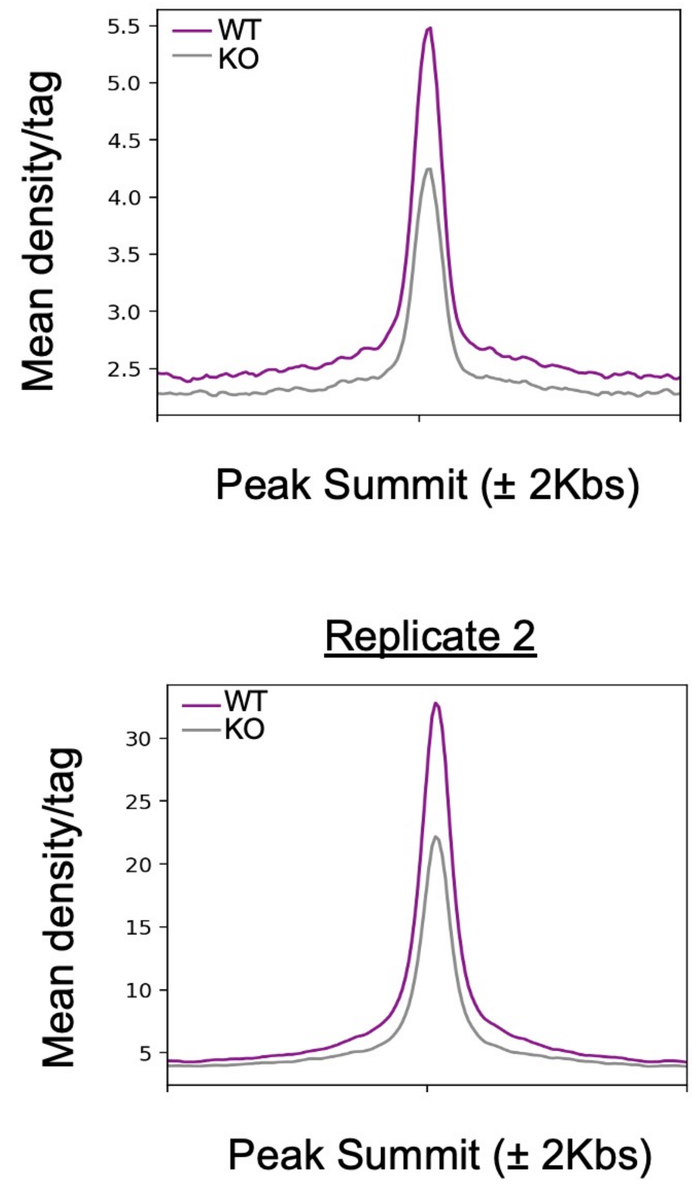
B

ATAC-seq



C

ATAC-seq at SALL4 Day-5 specific regions



D

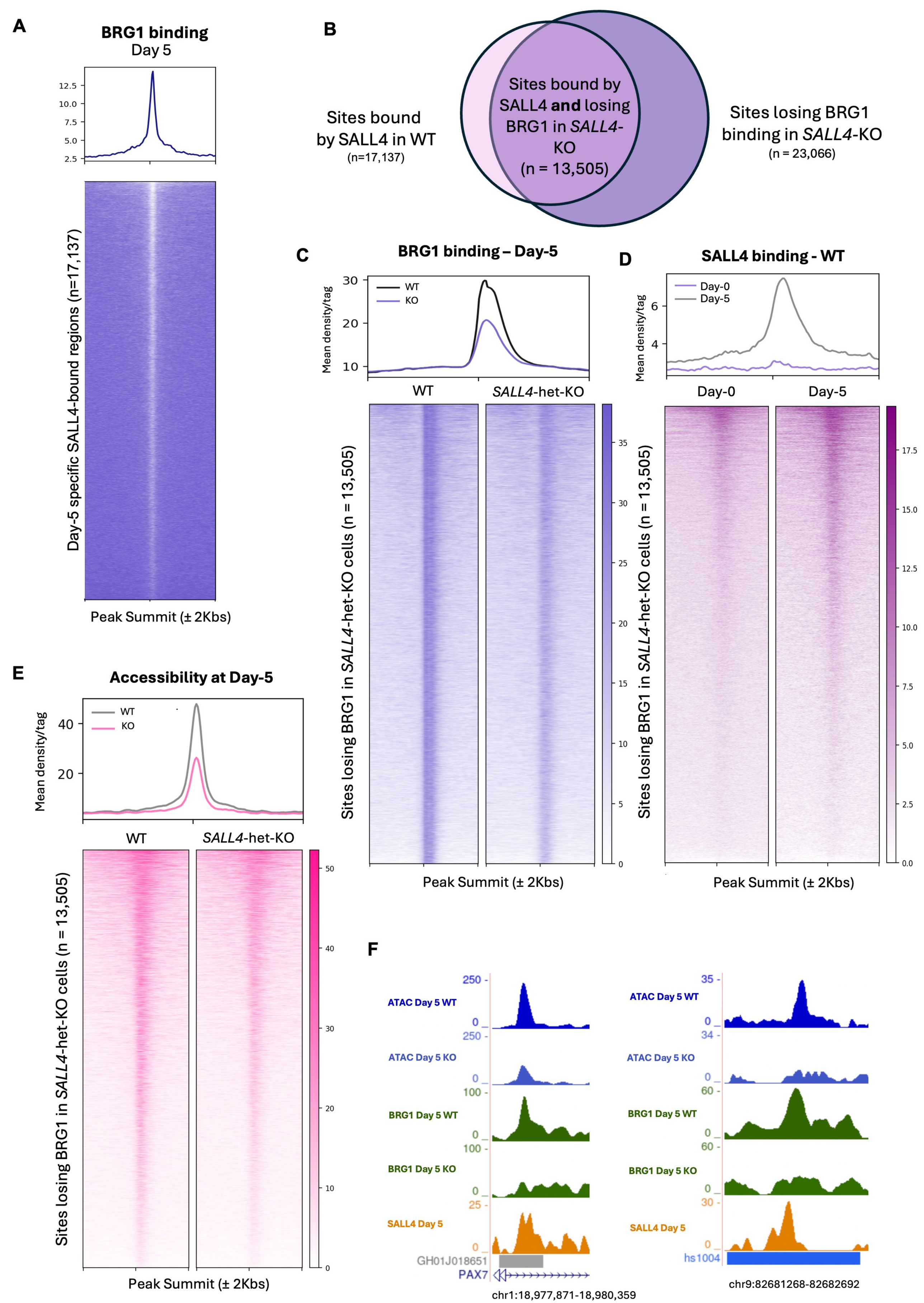
Gaining accessibility in KO Day 5 motifs

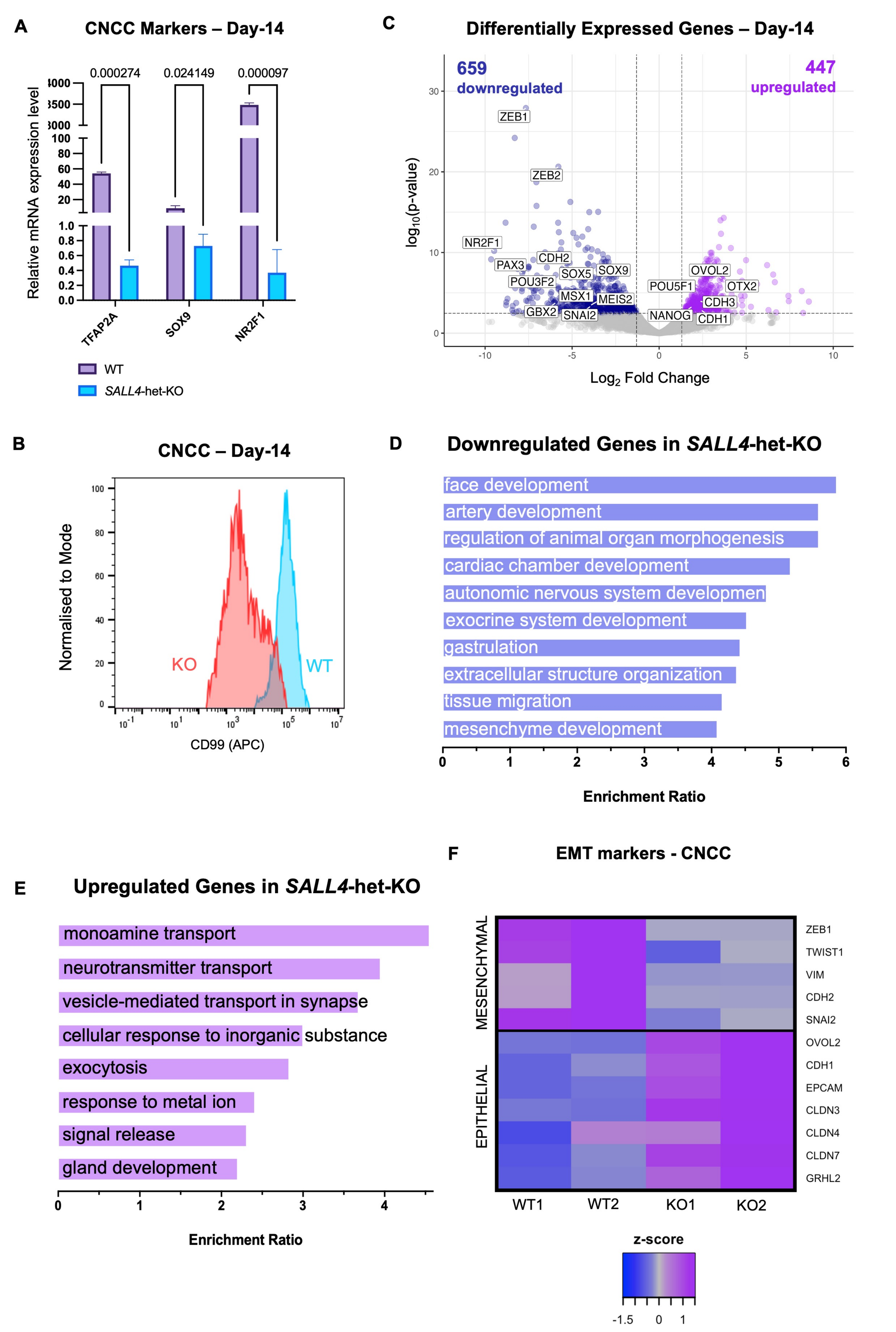
MOTIF #	Motif Consensus		p-value
51		CTCF BORIS	$1e^{-270}$
256		POU5F1/SOX2/ NANOG POU2F3 (epidermis) POU3F1 (nervous s.) POU2F2 (metastases)	$1e^{-103}$
410 (8%)		ZIC3 ZIC2 NANOG	$1e^{-84}$

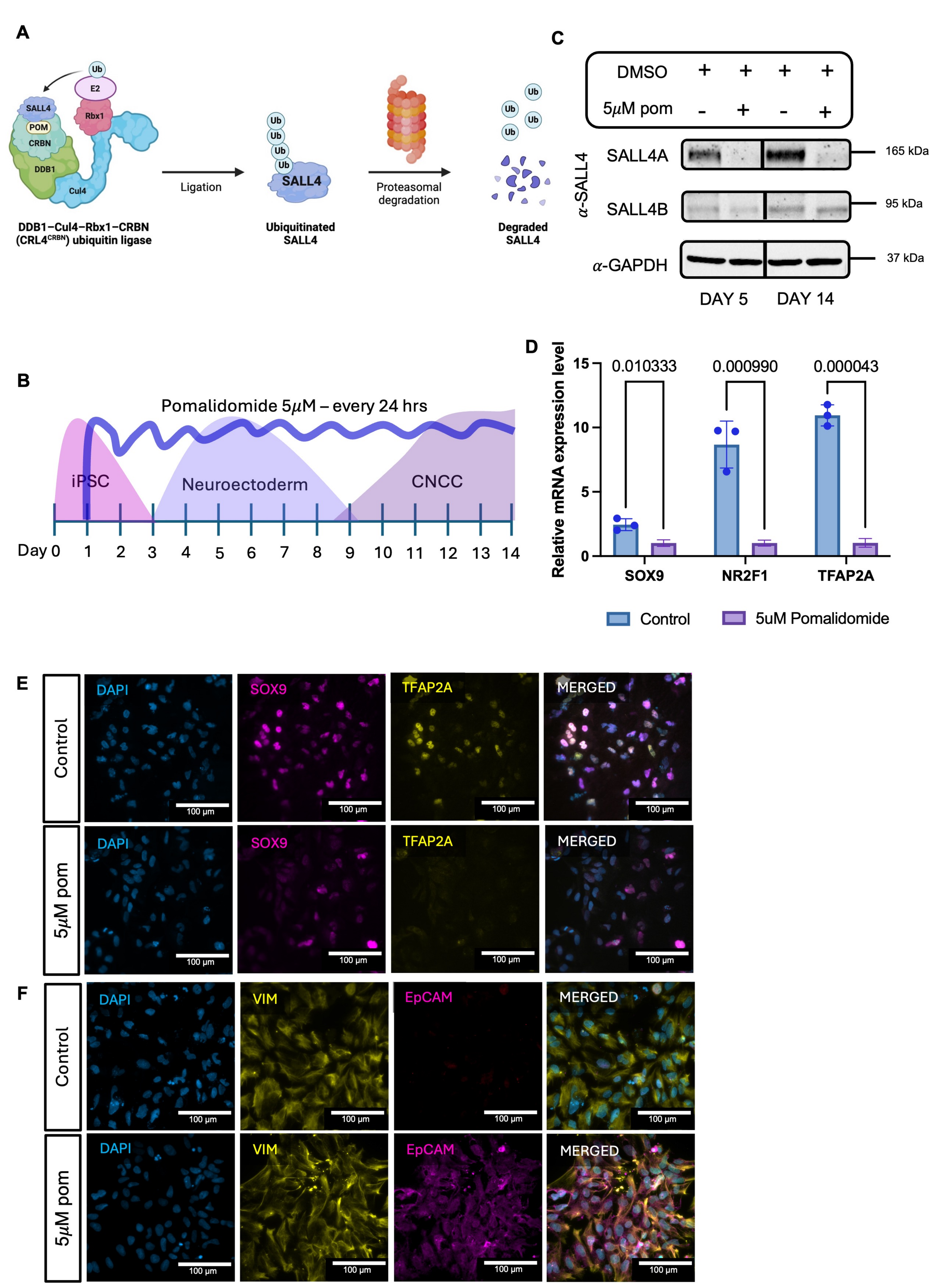
E

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	$1e^{-37}$


A
BRG1 binding
Day 5
12.5
10.0
7.5
5.0
2.5
Day-5 specific SALL4-bound regions (n=17,137)
Peak Summit (± 2Kbs)
B
Sites bound by SALL4 in WT
(n=17,137)
Sites bound by SALL4 and losing BRG1 in SALL4-KO
(n = 13,505)
Sites losing BRG1 binding in SALL4-KO
(n = 23,066)
C
BRG1 binding – Day-5
WT
KO
Mean density/tag
30
20
10
WT
SALL4-het-KO
Sites losing BRG1 in SALL4-het-KO cells (n = 13,505)
Peak Summit (± 2Kbs)
D
SALL4 binding - WT
Day-0
Day-5
Mean density/tag
6
4
Day-0
Day-5
Sites losing BRG1 in SALL4-het-KO cells (n = 13,505)
Peak Summit (± 2Kbs)
E
Accessibility at Day-5
WT
KO
Mean density/tag
40
20
WT
SALL4-het-KO
Sites losing BRG1 in SALL4-het-KO cells (n = 13,505)
Peak Summit (± 2Kbs)
F
ATAC Day 5 WT
ATAC Day 5 KO
BRG1 Day 5 WT
BRG1 Day 5 KO
SALL4 Day 5
GH01J018651
PAX7
chr1:18,977,871-18,980,359
hs1004
chr9:82681268-82682692


A
CNCC Markers – Day-14
0.000274
0.024149
0.000097
Relative mRNA expression level
TFAP2A
SOX9
NR2F1
WT
SALL4-het-KO
C
Differentially Expressed Genes – Day-14
659
downregulated
447
upregulated
ZEB1
ZEB2
NR2F1
PAX3
CDH2
SOX5
SOX9
POU3F2
MSX1
MEIS2
GBX2
SNAI2
OVOL2
POU5F1
OTX2
CDH3
NANOG
CDH1
log10(p-value)
Log2 Fold Change
B
CNCC – Day-14
Normalised to Mode
KO
WT
CD99 (APC)
D
Downregulated Genes in SALL4-het-KO
face development
artery development
regulation of animal organ morphogenesis
cardiac chamber development
autonomic nervous system developmen
exocrine system development
gastrulation
extracellular structure organization
tissue migration
mesenchyme development
Enrichment Ratio
E
Upregulated Genes in SALL4-het-KO
monoamine transport
neurotransmitter transport
vesicle-mediated transport in synapse
cellular response to inorganic substance
exocytosis
response to metal ion
signal release
gland development
Enrichment Ratio
F
EMT markers - CNCC
MESENCHYMAL
EPITHELIAL
ZEB1
TWIST1
VIM
CDH2
SNAI2
OVOL2
CDH1
EPCAM
CLDN3
CLDN4
CLDN7
GRHL2
WT1
WT2
KO1
KO2
z-score


A
Ub
E2
SALL4
Rbx1
POM
CRBN
DDB1
Cul4
DDB1–Cul4–Rbx1–CRBN
(CRL4CRBN) ubiquitin ligase
Ligation
Ub
SALL4
Ubiquitinated
SALL4
Proteasomal
degradation
Degraded
SALL4
C
DMSO + + + +
5μM pom - + - +
α-SALL4
SALL4A
SALL4B
α-GAPDH
165 kDa
95 kDa
37 kDa
DAY 5
DAY 14
B
Pomalidomide 5μM – every 24 hrs
iPSC
Neuroectoderm
CNCC
Day 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
D
Relative mRNA expression level
0.010333
0.000990
0.000043
SOX9
NR2F1
TFAP2A
Control
5uM Pomalidomide
E
Control
5μM pom
DAPI
SOX9
TFAP2A
MERGED
100 μm
F
VIM
EpCAM