Inability to switch from ARID1A-BAF to ARID1B-BAF impairs exit from pluripotency and commitment towards neural crest formation in *ARID1B*-related neurodevelopmental disorders

Luca Pagliaroli[1,#], Patrizia Porazzi[2,#], Alyxandra T. Curtis[1], Chiara Scopa[1], Harald M.M. Mikkers[3], Christian Freund[4], Lucia Daxinger[5], Sandra Deliard[6], Sarah A. Welsh[6], Sarah Offley[6], Connor A. Ott[1], Bruno Calabretta[2], Samantha A. Brugmann[7], Gijs W.E. Santen*[,8], and Marco Trizzino*[,1,9]

= Co-First authors
***** = Co-Senior authors
1) Department of Biochemistry and Molecular Biology, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA
2) Department of Cancer Biology, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA
3) Department of Cell & Chemical Biology, Leiden University Medical Center, Leiden, The Netherlands
4) LUMC hiPSC Hotel, Dept. Anatomy & Embryology, Leiden University Medical Center, Leiden, The Netherlands
5) Department of Human Genetics, Leiden University Medical Center (LUMC), Leiden 2300, RC, The Netherlands
6) Gene Expression and Regulation Program, The Wistar Institute, Philadelphia, PA
7) Division of Developmental Biology, Department of Pediatrics Cincinnati Children's Hospital Medical Center, Cincinnati, OH
8) Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands
9) Corresponding author: Marco Trizzino, Department of Biochemistry and Molecular Biology, Sidney Kimmel Medical College, Thomas Jefferson University, 233 S 10th Street, BLSB 826, Philadelphia, PA, 19104. E-mail: marco.trizzino@jefferson.edu

Abstract

The BAF complex modulates chromatin accessibility. Specific BAF configurations have functional consequences, and subunit switches are essential for cell differentiation. *ARID1B* and its paralog *ARID1A* encode for mutually exclusive BAF subunits. *De novo ARID1B* haploinsufficient mutations cause a neurodevelopmental disorder spectrum, including Coffin-Siris syndrome, which is characterized by neurological and craniofacial features. Here, we reprogrammed *ARID1B*$^{+/-}$ Coffin-Siris patient-derived skin fibroblasts into iPSCs and modeled cranial neural crest cell (CNCC) formation. We discovered that ARID1B is active only during the first stage of this process, coinciding with neuroectoderm specification, where it is part of a lineage-specific BAF configuration (ARID1B-BAF), which includes SMARCA4 and nine additional subunits. ARID1B-BAF acts as a gatekeeper, ensuring exit from pluripotency and lineage commitment, by attenuating *NANOG*, *SOX2* and thousands of enhancers directly regulated by these two pluripotency factors at the iPSC stage.

In iPSCs, these enhancers are maintained active by an ARID1A-containing BAF. At the onset of differentiation, cells transition from ARID1A-BAF to ARID1B-BAF, eliciting attenuation of the NANOG/SOX2 networks, and triggering pluripotency exit. Coffin-Siris patient cells fail to perform the ARID1A/ARID1B switch and maintain ARID1A-BAF at pluripotency enhancers throughout all stages of CNCC formation. This leads to a persistent and aberrant SOX2 and NANOG activity, which impairs CNCC formation. In fact, despite showing the typical neural crest signature (TFAP2A^{+}, SOX9^{+}), *ARID1B*-haploinsufficient CNCCs are also NANOG/OCT4-positive, in stark contrast with the *ARID1B*-wt CNCCs, which are NANOG/OCT4-negative.

These findings suggest a connection between *ARID1B* mutations, neuroectoderm formation, and a pathogenic mechanism for Coffin-Siris syndrome.

Keywords: BAF, ARID1B, Coffin-Siris, pluripotency enhancers, NANOG, SOX2, neural crest, neuroectoderm

Introduction

Cell fate commitment is a complex process that requires timely regulation of developmental genes. This phenomenon is mediated by the concerted activity of transcription factors (TFs) and chromatin regulators that modulate the interaction between *cis*-regulatory elements (enhancers, promoters) and RNA Polymerase II to promote gene expression. In this framework, a key role is played by the Brg1/Brm associated factor (BAF) chromatin-remodeling complex. BAF leverages ATP to modulate nucleosome positioning and chromatin accessibility genome-wide[1]. Different configurations of BAF, with context specific functions, have been described, and switches between subunits have been reported to be linked to specific developmental stages[2,3]. All known canonical BAF configurations require the presence of a subunit containing an AT-rich DNA binding domain (ARID). Namely, in the BAF complex this function is carried out by two mutually exclusive subunits: ARID1A and ARID1B[4-6]. Previous studies in mouse embryonic stem cells (mESCs) have identified an ESC-specific configuration of BAF which regulates pluripotency and self-renewal of embryonic stem cells (esBAF)[4-6]. Importantly, the esBAF exclusively incorporates ARID1A and not ARID1B. One of these studies also identified a non-canonical version of BAF (gBAF), which did not contain an ARID subunit and was involved in pluripotency maintenance of mESCs [4].
De novo haploinsufficient mutations in the *ARID1B* gene cause a spectrum of neurodevelopmental disorders, ranging from Coffin Siris syndrome to non-syndromic intellectual disability[7-12]. Coffin-Siris syndrome is associated with intellectual disability, specific craniofacial features, growth impairment, feeding difficulties and congenital anomalies such as heart and kidney defects[13]. Although other BAF components may also be mutated in this syndrome, approximately 75% of mutations are in *ARID1B* [11,14,15].
In addition to Coffin-Siris, genome-wide sequencing in unselected cohorts of patients with intellectual disability (ID) shows that *ARID1B* is always in the top-5 of causative genes, explaining about 1% of all ID cases[9,16]. Whereas several studies utilizing murine models recapitulate the neurological phenotypes typical of the *ARID1B*-associated syndromes[17-20], the molecular function of ARID1B in cell fate commitment during human development is still poorly understood.
Several hallmark features of *ARID1B* haploinsufficient individuals, including severe craniofacial, cardiac and digestive system abnormalities, suggest impaired neural crest cell migration as a pathological etiology[12]. Further, *ARID1B* is one of the most commonly mutated genes in neuroblastoma, a pediatric tumor of neural crest origin[21]. Thus, neural crest formation, migration, and differentiation represent suitable models to study the consequences of *ARID1B* mutations. To specifically address the molecular consequences of *ARID1B* haploinsufficient mutations in neural crest formation and development, we reprogrammed skin fibroblasts of two unrelated *ARID1B*$^{+/-}$ Coffin-Siris patients into induced Pluripotent Stem Cells (iPSCs) and used these patient-derived iPSCs to specifically model formation of cranial neural crest cell (CNCC), a multipotent cell population that forms through a neuroectodermal sphere intermediate that eventually give rise to migratory CNCCs.
Herein, we report the discovery of a lineage specific BAF configuration, containing ARID1B, SMARCA4 and seven additional subunits (ARID1B-BAF). In line with findings indicating that the esBAF and the gBAF do not contain ARID1B[4-6], we demonstrate that *ARID1B* mutations do not affect self-renewal and pluripotency of human iPSCs, as pluripotency is conveyed via binding of an ARID1A-containing BAF to pluripotency-associated enhancers of the SOX2 and NANOG networks. On the other hand, we show that ARID1B-BAF is required for lineage specification and exit from pluripotency. In fact, ARID1B-BAF is only transiently active during early stages of iPSC-to-CNCC differentiation, and specifically during the formation of the neuroectodermal spheres, where it replaces ARID1A-BAF at the SOX2/NANOG enhancers and elicits their repression.
Importantly, we demonstrate that *ARID1B*$^{+/-}$ cells from Coffin-Siris patients are unable to switch from ARID1A-BAF to ARID1B-BAF at the onset of neuroectoderm formation, and instead maintain ARID1A-BAF at pluripotency enhancers throughout the entire differentiation process. Failure to replace ARID1A with

ARID1B leads to defective exit from pluripotency and impaired cranial neural crest formation. These findings provide evidence for a direct connection between *ARID1B* mutations and a pathogenic mechanism for ARID1B-associated neurodevelopmental syndromes.

Results

Coffin-Siris patient-derived iPSCs are pluripotent and proliferate normally

To investigate the function of ARID1B in craniofacial development, we obtained skin fibroblasts from two unrelated *ARID1B*$^{+/-}$ Coffin-Siris Syndrome patients (hereafter Patient-19 and Patient-26; Fig. 1a,b), one male and one female, both carrying previously identified *de novo ARID1B* mutations. In detail, Patient-19 presented a nonsense mutation (c.3223C>T;p.Arg1075*; Fig. 1b), while Patient-26 had a frameshift mutation (c.2598del;Tyr867Thrfs*47; Fig. 1b) [10,14]. In both cases, a premature STOP codon was generated (Fig. 1b).

The fibroblasts were reprogrammed into iPSCs by the LUMC hiPSC Hotel (Leiden University). Patient-derived iPSCs exhibited regular morphology (Fig. 1c) and expressed pluripotency genes, as shown by both immunofluorescence (Fig. 1d) and RT-qPCR (Fig. 1e). Further, patient-derived iPSCs grew at the same rate as an *ARID1B*$^{+/+}$ control line (Control line-1; Fig. 1f).

Importantly, the aberrant STOP codon introduced by the mutations was located either upstream (Patient-26) or inside (Patient-19) the AT-Rich Interactive Domain (ARID) (Fig. 1b), which is required for ARID1B's interaction with chromatin[22]. Moreover, in both patients, the new STOP codon was localized upstream of the Nuclear Localization Signal (NLS, Fig. 1b), suggesting that the gene product arising from the mutated allele would not be able to reach the nucleus or access the chromatin, even in the unlikely case that the transcript escaped non-sense mediated mRNA decay[23]. To test this, we performed cellular fractionation in patient and control iPSCs and conducted an ARID1B western blot on the chromatin fraction with an antibody raised against a peptide in the N-terminus of ARID1B, upstream of the mutated regions (sc-32762). Consistent with our hypothesis, the immunoblot on the chromatin fraction revealed a significantly reduced amount of ARID1B protein in both patient samples relative to the control *ARID1B*$^{+/+}$ iPSC line (Control line-1; Supplementary Figure S1a). ARID1B was reduced by approximately 60% and 80% in Patients 19 and 26, respectively. ARID1B was not detected in the cytoplasmic or nuclear fraction of any cell lines (Supplementary Fig S1a). Thus, *ARID1B* haploinsufficient iPSCs remained pluripotent, did not exhibit any growth defects, but did display significantly less chromatin-bound ARID1B than control iPSCs.

CNCC formation is impaired in Coffin-Siris patient-derived iPSCs

Utilizing published methods for iPSC-to-CNCC differentiation[24] we generated CNCCs from control iPSCs, in 14 days (Fig. 2a,b). A time-course western blot conducted during the differentiation of Control line-1 revealed that ARID1B protein was robustly expressed during the first week of differentiation, peaking between days 5 and 7, with some expected variability between biological replicates (Fig. 2c; see Fig. 7 for a second replicate). After day 7; however, ARID1B protein was markedly downregulated (Fig. 2c). The window of robust ARID1B expression (day-1 to day-7) coincided with the differentiation of the iPSCs into neuroectodermal spheres, suggesting that this BAF subunit may have a role in neuroectoderm specification.

Next, we induced CNCC differentiation in two Coffin-Siris patient lines and compared them to Control iPSC-derived CNCCs using flow cytometry to measure multiple pluripotency (SSEA-4, TRA-1-60-R) and CNCC (CD10, CD99) surface markers. Cells were sampled at day-zero (iPSCs), day-5 (neuroectoderm), and day-14 (CNCC). Notably, CNCC formation was impaired in both patient-derived lines, as evident by a sizable cell population that was double-positive for pluripotency surface markers even after 14 days of differentiation (Fig. 2d; Supplementary Fig. S2). This double positive population comprised 4.5% and

19.5% of the cells in the two patient lines, respectively (Fig. 2d). In line with this, a large fraction of patient cells showed significantly lower expression of CNCC surface markers even after 14 days of differentiation relative to the control line (Fig. 2e).

To further characterize patient-derived CNCCs, we performed immunofluorescence for pluripotency (OCT4, NANOG) and neural crest (SOX9) markers, in control and patient lines at day-14. Interestingly, CNCCs derived from patient iPSCs displayed a gene expression signature distinct from that of control CNCCs (Fig. 3). In fact, nearly all patient derived CNCCs were positive for SOX9, NANOG (Patients 19 and 26) and OCT4 (Patient 26 only; Fig. 3). This was in stark contrast to the control line, in which SOX9-positive cells were almost all OCT4- and NANOG-negative (Fig. 3; See Supplementary Fig. S3 for quantifications and p-values). Together, these data suggested that while *ARID1B* is dispensable for pluripotency, haploinsufficiency of this gene severely impaired CNCC formation, resulting in CNCCs expressing key pluripotency genes.

Chromatin accessibility is dysregulated in differentiating Coffin-Siris patient-derived lines

We used Next-Generation Sequencing to investigate why *ARID1B* haploinsufficient Coffin-Siris iPSCs did not successfully differentiate into CNCCs. Given that ARID1B protein levels in control cells reach a peak between days 5 and 7, samples were taken at day-5 to perform genomic analyses. Experiments were conducted with two biological replicates per condition (two control lines, two patient lines). For each condition, a male and a female were included to avoid sex-specific confounding effects. Technical replicates were also used for each biological replicate. To avoid batch effects, all the biological replicates and conditions were processed together.

Since ARID1B is a component of the BAF chromatin-remodeling complex, we profiled chromatin accessibility with ATAC-seq. Overall, at day 5, 29,758 ATAC-seq peaks were identified across all replicates and conditions (patients and controls; FDR <0.05; Fig. 4a). Conversely, 5,540 peaks were specific to the patient iPSCs (i.e. replicated in all patient's iPSC replicates and not detected in any of the controls; hereafter patient-specific ATAC-seq regions; Fig. 4a,b; Supplementary File S1). Finally, only 578 peaks were specific to the controls (hereafter control-specific ATAC-seq regions; Fig. 4a,c; Supplementary File S1). We therefore focused on the 5,540 patient-specific ATAC-seq regions because they represented 91% (5,540/6,118) of all regions with differential chromatin accessibility between patient and control lines.

At day 0, ATAC-seq performed in iPSCs revealed that the 5,540 regions were highly accessible, with no significant differences between patient and control lines (Supplementary Fig. S4a). By day 5, this dramatically shifted and 5,511 of the 5,540 regions (99.4%) were called as peaks exclusive to the patient lines. These data suggested that these were regions highly accessible in iPSCs and repressed by day 5 of iPSC-to-CNCC differentiation. Such repression is impaired by *ARID1B*-haploinsufficiency, indicating that chromatin accessibility in the 5,540 genomic sites may be directly regulated by an ARID1B-containing BAF during exit from pluripotency and neuroectoderm specification. Thus, we investigated ARID1B's binding at these regions both in day 0 (iPSCs) and in day 5 cells. At day 0, these regions were not bound by ARID1B (Supplementary Fig. S4b). This was expected, given that that ARID1B is not a component of any of the BAF configurations predominant in iPSCs and ESCs (esBAF and gBAF[4-6]), and likely explains why differences in chromatin accessibility at these regions between patient and control lines at day 0 was not observed.

On the other hand, at the day 5, the 5,540 regions were bound by ARID1B in both control lines, while the binding was almost entirely lost in both patient lines (Fig. 4d). This loss of binding correlated with lack of chromatin repression in both patient lines. Together, these findings indicated that gain of ARID1B binding at these 5,540 genomic sites in early stages of iPSC-to-CNCC differentiation may be required for repression of the pluripotency program.

The ARID1B-BAF attenuates thousands of enhancers at the onset of CNCC differentiation

To determine the nature of the 5,540 genomic regions, we associated a gene to each region based on the distance from the nearest Transcription Start Site (TSS). Overall, 87.5% of the ARID1B ChIP-seq peaks were located >10 Kbs from the nearest TSS and may represent putative enhancers, while the remaining 12.5% are likely promoters. ChIP-seq time-course for H3K27ac in control cells revealed that many of these regulatory regions were enriched for H3K27ac in iPSCs, progressively lost this enrichment during differentiation, and by day 5 had very reduced H3K27ac signal (Fig. 4e). The gradual decrease in H3K27ac mirrored the steady increase in ARID1B expression detected during the early stages of iPSC-to-CNCC differentiation, and specifically during the formation of the neuroectodermal spheres (Fig. 2c). Consistent with this, the differentiating cells from both patients had significantly higher levels of H3K27ac in these regions, relative to the two control lines at day 5 (Wilcoxon's Rank Sum Test $p < 2.2 \times 10^{-16}$ in all the patient vs control pairwise comparisons; Fig. 4f, g).

We further investigated the fate of these regions over the course of iPSC-to-CNCC differentiation. We found that while these cis-regulatory elements were still largely active in both patient lines at the day-7, they were inactive in the control line (Fig. 4g). Interestingly, at day 9 of differentiation, these regions were inactivated (i.e., no H3K27ac signal) in Patient-19 line, while remained active (i.e., persistent H3K27ac signal) in the Patient-26 line (Fig. 4g).

Based on the high H3K27ac signal that the 5,540 patient-specific regions display at the iPSC stage (Fig. 4e), we surmised that these sites could represent cis-regulatory elements important for pluripotency. In line with this hypothesis, DNA-motif analysis on the 5,540 regions revealed that they were enriched for the binding sites of multiple pluripotency factors, including SOX2 and NANOG (Fig. 4h; Supplementary File S2).

To ensure that the molecular phenotypes observed were directly caused by the *ARID1B* mutations, and not by co-occurring mutations in other genes coincidentally shared by both (unrelated) patients we employed shRNAs to knock-down ARID1B in the Control Line-1. We were able to obtain a partial knock-down of ARID1B at the iPSC stage (shRNA-1; Supplementary Fig. 4c), which represented a suitable model for *ARID1B* haploinsufficiency. ARID1B-KD iPSCs were put through the CNCC differentiation protocol and collected at day 5, and profiled via ATAC-seq and ChIP-seq for H3K27ac. Notably, both sequencing experiments perfectly recapitulated our findings in the patient-derived lines. Upon ARID1B-KD, we detected significantly increased chromatin accessibility and H3K27ac signal in the 5,540 patients-specific regions relative to the same iPSC line transduced with a control shRNA (Wilcoxon's Rank Sum Test $p < 2.2 \times 10^{-16}$; Supplementary Fig. S4d, e).

Together, these data indicate that ARID1B-BAF modulates the chromatin accessibility of a specific set of ~4,900 pluripotency enhancers and ~600 promoters that are highly active in iPSCs, moderately active at the onset of neuroectoderm formation, and inactive by day 7 and for the remaining course of CNCC formation (Fig. 4e). These data suggested that impaired attenuation of these cis-regulatory elements in the *ARID1B* haploinsufficient cells, subsequently hampers the entire differentiation process towards a CNCC fate.

"Pluripotency" and "Exit from Pluripotency" genes are dysregulated in differentiating patient lines

Impaired attenuation of ~4,900 pluripotency-relevant enhancers and ~600 promoters could have a profound effect on gene expression levels. Indeed, RNA-seq conducted at day 5 identified 2,356 differentially expressed genes, 1,685 of which were downregulated, and 671 upregulated (FDR <5%; 5a). As expected, *ARID1B* was one of the top downregulated genes in patient CNCCs (Fig. 5a). In stark contrast, only 54 genes were identified as differentially expressed when we performed RNA-seq at the iPSC stage (FDR <5%). This suggested that ARID1B is important for lineage commitment, again mirroring the progressive increase in the ARID1B protein level observed during early stages of differentiation (Fig. 2c). The small number of differentially expressed genes identified at the iPSC stage was again consistent with the finding that esBAF and gBAF do not include ARID1B[4-6].

Notably, 598/2,356 (25.4%) of the genes differentially expressed at day 5 also represented the nearest gene to one of the 5,540 pluripotency enhancers and promoters aberrantly active in the Coffin-Siris patient cells at the same time point (Supplementary File S3). These results suggested that over a quarter of differentially expressed genes were under the direct control of ARID1B-BAF throughout modulation of chromatin accessibility at associated enhancers and promoters. As expected, when we compared these 598 genes against the entire set of 2,356 differentially expressed genes, we found that the 598 genes exhibited enrichment for genes upregulated in patient cells (Fisher's Exact Test $p < 0.0001$). Ingenuity Pathway Analysis on the 598 genes identified five of the top canonical pathways as associated with either pluripotency or exit from pluripotency, as well as Wnt-β catenin signaling pathway[25,26] (Fig. 5b; Supplementary File S4).

In accordance with the ATAC-seq data, SOX2 was detected among the top upstream regulators (Fig. 5c), and three of the most important pluripotency factors, *NANOG, SOX2* and *POU5F1* (OCT4), were highly expressed in the patient lines at day 5 (Fig. 5d).

Both the "*Role of NANOG in Mammalian Embryonic Stem Cell Pluripotency"* and the "*PPARα/RXRα Activation"* pathways were enriched in the 598 genes (Fig. 5b). Namely, the genes belonging to the former pathway were all upregulated, while those belonging to the latter were downregulated (Fig. 5d). These two pathways caught our attention because they are thought to antagonize each other. More specifically, NANOG blocks the differentiation of pluripotent cells and establishes the pluripotent state during somatic cell reprogramming. Conversely, the PPARα/RXRα pathway is activated at the onset of differentiation to promote exit from pluripotency[27]. The activation of PPARα/RXRα contributes to the repression of the *NANOG* network to allow efficient exit from the undifferentiated stage[27-29]. Consistent with this, PPARα-inhibitors have been employed to improve iPSC reprogramming[27].

These gene expression data were in agreement with the immunofluorescence findings and with the perpetual *NANOG* and *OCT4* expression in patient derived CNCCs at day 14 of differentiation (Fig. 3).

Finally, we looked for potential overlap between our set of 2,356 differentially expressed genes at day 5 of iPSC differentiation toward CNCC and a set of genes identified as differentially expressed in a recent RNA-seq study which compared cerebellar tissue of *ARID1B*$^{+/-}$ and *ARID1B-wt* mice[30]. Notably, 1,297 of the 2,356 genes (55%) found as differentially expressed in our study were also found as differentially expressed in the KO-mouse model dataset. This overlap (55%) was significantly higher than expected by chance (Fisher's Exact Test $p < 2.2 \times 10^{-16}$) and suggested that the pathways regulated by ARID1B are important for both craniofacial and brain development.

Taken together, our RNA-seq data suggested that differentiating *ARID1B*$^{+/-}$ patient-derived lines exhibited a persistent upregulation of multiple pluripotency factors and associated gene networks, along with downregulation of genes responsible for exit from pluripotency. Dysregulation of these gene networks subsequently impaired neuroectoderm specification and CNCC formation.

Aberrant SOX2 and NANOG activity in the *ARID1B* haploinsufficient patient cells

Our experiments indicate that the *ARID1B* haploinsufficient cells fail to attenuate thousands of pluripotency enhancers and promoters enriched for SOX2 and NANOG binding sites (Fig. 4a–h). Further, at day-5 of iPSC-to-CNCC differentiation, the expression of *SOX2* and *NANOG* is significantly higher in the patient derived cells than in the controls, and the gene regulatory networks associated with these pluripotency factors are also upregulated (Fig. 5b–d). Moreover, both patient-derived CNCCs (day-14) exhibit aberrant expression of NANOG (Fig. 3).

Given these findings, we set out to investigate the binding profile of SOX2 and NANOG in patient and control lines by ChIP-seq at CNCC day-5. Our spike-in normalized SOX2 ChIP-seq revealed that 3,284/5,540 (59.7%) patient-specific ATAC-seq peaks exhibit significantly higher SOX2 binding in patients relative to control lines (Fig. 6a). In line with this, the chromatin at these regions is accessible in the patient lines but not in the control lines (Fig. 6b). SOX2 is a pioneer factor that can bind condensed nucleosomes to open

the chromatin and allow binding of other factors[31]. As demonstrated by previous studies in mouse embryonic stem cells (mESCs), SOX2 and other pluripotency pioneer factors (e.g. OCT4) require the BAF complex to perform their pioneer activity[6,31,32]. Our findings indicate that, in control conditions, the ARID1B-BAF complex likely antagonizes the cooperation between other BAF configurations and SOX2, counter-acting the pioneer activity of the latter as soon as cell differentiation is induced. Further, we identified an additional set of 497 SOX2 peaks specific to the patient lines, which did not exhibit changes in chromatin accessibility. Moreover, we also identified 1,146 SOX2 peaks exclusive of the control lines (Supplementary File S5). Importantly, these control-specific SOX2 peaks were located in proximity to genes associated with neural crest differentiation, including *TFAP2A*, *PAX6*, *PAX7*, *WNT4*, *ENO1*, *C8B*, and *SERBP1* among others. These findings are consistent with two recent studies which suggested that SOX2-chromatin interactions are rewired upon differentiation cues[33,34]. Such rewiring appears impaired in *ARID1B*-haploinsufficient cells, which aberrantly maintain SOX2 at pluripotency-associated enhancers, and at the same time fail to reposition this transcription factor at the developmental enhancers.

Next, we profiled NANOG at day-5 of differentiation. For this transcription factor, the spike-in normalized ChIP-seq revealed 4,538 peaks unique to the patients (Supplementary File S6). However, in this case, only 219 (4.8%) of the patient-specific NANOG peaks overlapped a patient-specific ATAC-seq peak. We thus interrogated our ATAC-seq data to determine the state of chromatin accessibility at the 4,538 patient-specific NANOG peaks, and overall found no significant changes in accessibility in these regions between the patients and the control lines (Fig. 5c). We note that nearly a quarter of the patient-specific NANOG peaks were found in regions of repressed chromatin (Fig. 6c,e), in line with recent studies which suggested that NANOG can bind repressed chromatin like other pioneer pluripotency factors[35,36].

Despite no changes in chromatin accessibility, the NANOG ChIP-seq signal at the 4,538 patient-specific NANOG peaks was significantly higher in the patient than in the control lines (Wilcoxon's Rank Sum Test $p < 2.2 \times 10^{-16}$ in all the patient vs control pairwise comparisons; Fig. 6d,e). We hypothesized that the increased NANOG binding detected in the patients' cells (Fig. 6d) could reflect increased *NANOG* expression (Fig. 6d). In fact, several elegant studies in embryonic stem cells have demonstrated that changes in *NANOG* dosage mark the transition from an undifferentiated state (high *NANOG*), to a state with differentiation potential (low *NANOG*)[37-39]. Importantly, it has been shown that SOX2 and OCT4 bind a cis-regulatory element in the promoter region of *NANOG*, likely modulating its expression[40,41]. Thus, we examined this cis-regulatory element in detail. As expected, at day-5 of iPSC-to-CNCC differentiation, the chromatin accessibility at the promoter-proximal element is significantly higher in the two patient lines than in the two controls (Student's T-Test p=0.0065; Fig. 6f). Accordingly, increased chromatin accessibility correlates with increased SOX2 binding on the cis-regulatory element (Fig. 6f), perhaps explaining the higher *NANOG* gene expression reported in patient-derived lines at day-5. Lastly, our shRNA experiments also confirmed these findings, demonstrating that the knock-down of ARID1B in the Control line-1 line correlates with a sizeable increase in accessibility at the *NANOG* cis-regulatory element (Fig. 6g), thus suggesting that ARID1B-BAF directly modulates *NANOG* expression dosage at the onset of differentiation.

In sum, the *ARID1B* haploinsufficient lines exhibit persistent activity of two key pluripotency factors (SOX2, NANOG) in the early and late stages of CNCC formation. The aberrant activity of SOX2 and NANOG likely leads to impaired lineage commitment and inefficient CNCC formation.

A switch from ARID1A-BAF to ARID1B-BAF is necessary for exit from pluripotency

We next wanted to elucidate how the BAF complex compensated for *ARID1B* haploinsufficiency. As mentioned above, ARID1A and ARID1B represent the only two subunits of the BAF harboring an ARID domain, which is leveraged by the complex to interact with the chromatin[22]. A third ARID subunit (ARID2) is exclusive to a different configuration of the complex (pBAF). *ARID2* mutations have been shown to cause a neurodevelopmental disorder that does not fully recapitulate the Coffin-Siris syndrome phenotype,

336 although there is some overlap[42]. Compensatory mechanisms between ARID1A and ARID1B were recently
337 demonstrated in ovarian cancer[43]. Thus, we hypothesized that *ARID1B* haploinsufficient patient-derived
338 cells may compensate for partial loss of ARID1B with ARID1A. To test this, we first assessed ARID1A protein
339 levels in *ARID1B*-wt cells during CNCC formation and found that ARID1A expression was complementary
340 to ARID1B (Fig. 7a). In agreement with the specific composition of the esBAF, which requires ARID1A[5,6],
341 human iPSCs expressed high levels of ARID1A and relatively low ARID1B (Fig. 7a). On the other hand, at
342 the initiation of iPSC-to-CNCC differentiation (day 1) ARID1B was upregulated while ARID1A was strongly
343 repressed to levels barely detectable (Fig. 7a). ARID1B remained the only active ARID1 subunit between
344 days 1-5 (i.e., during the formation of the neuroectodermal cells; Fig. 7a). Finally, at day 7, ARID1B was
345 abruptly downregulated and ARID1A was restored at high levels at day 9. This latter switch corresponded
346 to the beginning of the differentiation of the neuroectodermal spheres into migratory CNCCs (Fig. 7a).
347 Together, these data suggested that throughout the course of CNCC differentiation, multiple switches
348 between ARID1A and ARID1B occur, and that these two ARID1 subunits regulate specific developmental
349 stages during CNCC formation.
350 Next, we profiled ARID1A protein levels during iPSC-to-CNCC differentiation in patient derived lines. Time-
351 course immunoblotting revealed that the temporary decommissioning of ARID1A during neuroectoderm
352 specification (~days 1-7) failed to occur in both of the patient cell lines (Fig. 7b). In fact, in the patient
353 lines, ARID1A protein levels were maintained throughout the course of iPSC differentiation (days 1-11)
354 (Fig. 7b).
355 In addition to perpetual expression, there was an approximate 5-fold increase in ARID1A expression in
356 patient cell lines, relative to controls at day 5 (Patient 19: 4.9-fold; Patient 26: 5.1-fold; Fig. 7c), a time
357 point which exhibited robust ARID1B expression in control cells. Together, these data suggested that the
358 patient cells compensated for the partial loss of ARID1B by maintaining aberrantly high ARID1A levels
359 throughout the differentiation process.
360 A recent study conducted on liver cells demonstrated that ARID1A- and ARID1B-containing BAF may have
361 antagonistic functions in the transcriptional regulation of specific genes, with ARID1B acting prevalently
362 as a repressor of enhancer elements, as opposed to the ARID1A, which mostly functions as an activator[44].
363 Hence, we hypothesized that the perpetual and robust ARID1A protein levels detected in patient-derived
364 cells during iPSC-to-CNCC differentiation, and specifically during neuroectoderm specification, may result
365 in prolonged activity of pluripotency enhancers. Consistent with the recent studies which reported that
366 the (ARID1A-containing) esBAF regulates pluripotency genes in iPSCs and ESCs[4-6], ARID1A ChIP-seq
367 performed at the iPSC stage revealed that the 5,540 pluripotency enhancers and promoters were bound
368 by ARID1A in all the four lines (Fig. 7d). Conversely, the same experiment conducted at day 5 of
369 differentiation revealed that while ARID1A-binding was lost at the 5,540 pluripotency elements in control
370 cell lines, it was maintained in patient cell lines (Fig. 7d). In summary, these data indicated that the status
371 of a set of ~5,500 pluripotency enhancers and promoters was regulated by ARID1A at the iPSC stage and
372 by ARID1B during exit from pluripotency and neuroectodermal lineage commitment. Importantly,
373 *ARID1B*-haploinsufficiency in patient cell lines triggered a compensatory mechanism which resulted in
374 persistent binding of ARID1A at pluripotency enhancers throughout differentiation.
375
376 **The ARID1B-BAF complex exclusively incorporates SMARCA4 as an ATPase subunit**
377 Finally, we addressed the composition of the ARID1B-BAF complex at day 5 of iPSC-to-CNCC
378 differentiation. To do so, we performed immunoprecipitation of endogenous ARID1B followed by mass-
379 spectrometry (IP-MS). In control cells, ARID1B coeluted with a total of ten additional BAF subunits
380 (hereafter ARID1B-BAF; Supplementary Fig. S5a). In mammals, BAF complexes incorporate two widely
381 interchangeable and mutually exclusive ATPase subunits, SMARCA2, and SMARCA4. Remarkably,
382 SMARCA4 was the only ATPase subunit identified as coeluting with ARID1B in our IP-MS, while zero
383 peptides of SMARCA2 were detected (Supplementary Fig. S5a). This suggests that ARID1B-BAF selectively

incorporates only SMARCA4 as a catalytic subunit, while it does not tolerate the incorporation of SMARCA2.

We repeated this experiment in the patient-lines to determine if perpetually active ARID1A replaced ARID1B in the ARID1B-BAF or if instead it was part of a completely different BAF configuration. We thus performed ARID1A IP-MS at day 5 in patient lines. Notably, in the two patient lines ARID1A co-eluted with all the other subunits of the ARID1B-BAF (Supplementary Fig. S5b). In summary, the perpetual presence of ARID1A-BAF in patient cells correlated with a very specific (and previously uncharacterized) configuration of the BAF complex (ARID1B-BAF), which was active during exit from pluripotency and neuroectoderm formation.

Intriguingly, the transcription factor SALL4 also coeluted with ARID1B at day 5, suggesting a possible interaction with the complex. Like ARID1B, SALL4 is also dispensable for the maintenance of the pluripotency networks, while it is essential for lineage commitment in early mammalian development, during which it targets sites with binding motifs also recognized by SOX2, OCT4 and NANOG[45-48]. SALL4 was previously shown to interact with the NuRD repressive complex[46], while interactions with BAF have been largely unexplored. It was recently demonstrated that this transcription factor has affinity for AT-rich regions[49], thus providing further support to the ARID1B-SALL4 interaction. *SALL4* mutations are also associated with developmental syndromes, including Okihiro syndrome, Holt-Oram syndrome, and Townes-Brocks Syndrome[50]. Notably, the *SALL4* gene is downregulated in the Coffin-Siris patients at CNCC day-5 but not in undifferentiated iPSCs, suggesting a possible feedback mechanism between *ARID1B* and *SALL4* during lineage commitment. Future studies will be necessary to support the speculation that SALL4 serves as an intermediator for ARID1B-BAF recruitment at the pluripotency enhancers.

Discussion

ARID1B is a member of the evolutionarily conserved SWI/SNF (BAF) chromatin remodeler[22,51]. *De novo* haploinsufficient mutations in the *ARID1B* gene cause severe neurodevelopmental disorders which affect both physical and cognitive development. In this study, we investigated the role of Coffin-Siris-associated *ARID1B* mutations in the context of craniofacial development and report the discovery of a novel function of the BAF complex: attenuation of the gene expression program associated with pluripotency maintenance upon differentiation cues. We found that this repressive function is performed at pluripotency enhancers and promoters by a specific and novel BAF complex configuration (ARID1B-BAF), which is composed of 9 subunits, with the enzymatic activity seemingly carried out exclusively by SMARCA4.

As a consequence of the *ARID1B* mutations, Coffin-Siris patient cells fail to repress the pluripotency elements. This subsequently elicits aberrant SOX2 activity genome-wide, which in turn leads to the upregulation of multiple pluripotency genes, including *NANOG* and its associated gene network, and to the downregulation of the genes responsible for coordinating the exit from pluripotency (PPARα/RXRα pathway). We demonstrated that these pluripotency enhancers are normally maintained in an active state by ARID1A-BAF at the iPSC stage, and subsequently repressed by the ARID1B-BAF throughout neuroectoderm formation, which is the first stage of CNCC development.

A switch between ARID1A-BAF and ARID1B-BAF upon differentiation cues is hence necessary for commitment towards the neuroectodermal lineage. We additionally report that a second switch from ARID1B-BAF to ARID1A-BAF takes place later, when the neuroectodermal spheres differentiate into CNCCs. This suggests that ARID1B is likely the dominant ARID1 subunit during neuroectoderm specification, while ARID1A regulates the following stage, during which the neuroectodermal spheres differentiate into migratory CNCCs. It is worth noting that the neuroectoderm not only gives rise to all regions of the brain and central nervous system (hindbrain, midbrain, forebrain, spinal cord and motor neurons), but also gives rise to the neural crest cells, which emanate out from the dorsal aspect of the neural tube. Therefore, we speculate that the dysregulation of neuroectoderm specification caused by

ARID1B mutations may underlie both the cognitive impairment and craniofacial abnormalities that are typical of Coffin-Siris syndrome. This model is further supported by a significant overlap of differentially expressed genes identified in our study (neural crest formation) and in a recently published *ARID1B*-KO mouse model (brain tissue)[30].

Other studies have previously suggested that switches between SWI/SNF subunits play important roles in cell fate determination. For example, a switch between the two catalytic subunits SMARCA4 and SMARCA2 mediates the activation of human IFNγ-activated genes[52]. Similarly, a gain of the subunit BAF53a in the neuron-specific BAF (nBAF) is required to control cell cycle exit in developing neurons[2,3]. With our study, we discovered a novel, binary switch between BAF subunits (ARID1A/ARID1B), critical for the exit from pluripotency. Importantly, a balance between pro-self-renewal and pro-differentiation signals is pivotal for the determination of stem cell fate[53]. We demonstrate that such balance is lost in Coffin-Siris patients, whose cells are unable to perform the ARID1A/ARID1B switch at the pluripotency enhancers at the onset of differentiation. This switch is essential to successfully complete CNCC differentiation.

Pluripotency is orchestrated by a transcription factor network that needs to be extinguished in an orderly manner to enable lineage commitment and differentiation[53-55]. We find that ARID1B-BAF plays an essential role in this process, by means of repressive activity at pluripotency enhancers for the SOX2, NANOG, and OCT4 networks. Similarly, an association between SOX3 and the SMARCA2 ATPase subunit of BAF was recently suggested in a study of neural development in the Nicolaides-Baraitser syndrome[56]. It is worth noting that Coffin-Siris and Nicolaides-Baraitser syndromes share many physical and neurological phenotypes[56-58].

The BAF complex is predominantly considered as a transcriptional activator, which balances the Polycomb Repressor Complexes (PRC1, PRC2) in the modulation of gene expression[7,59]. Despite this widely held belief, repressive activity for BAF has been reported. For instance, a study conducted on hepatocellular carcinoma cell lines uncovered that ARID1A-containing BAF activates and represses roughly equal numbers of genes, while ARID1B-containing BAF was found to primarily repress enhancer activity[44]. Our experiments corroborate these findings, supporting an enhancer-repressor function for ARID1B-BAF. We demonstrate that the repressive activity of ARID1B-BAF is specific to a set of ~4,900 enhancers and ~600 promoters, enriched for SOX2 and NANOG binding sites. In control cells, these cis-regulatory elements are highly active at the iPSC stage, moderately active in the first four days of differentiation, and finally repressed by day 5, a time point at which we observed the peak of ARID1B protein expression. Coffin-Siris patient-derived iPSC lines exhibit aberrant chromatin accessibility at these cis-regulatory elements for many days after the onset of iPSC-to-CNCC differentiation, enforcing a perpetual pluripotency signature which persists even after two weeks of differentiation.

Patient-26-derived cells display the most extreme cellular and molecular phenotype, with a large population of cells remaining pluripotent at day-14 of differentiation, likely as a consequence of higher *SOX2, OCT4* and *NANOG* expression and activity. Cells derived from this patient also show the highest levels of ARID1A binding at these enhancers at day 5 of CNCC differentiation. Although it is difficult to formally compare disease severity since there are no accepted severity scales for Coffin-Siris syndrome, it is worth noting that Patient-26 had a more severe disease process than Patient-19. For example, Patient-26 was not able to speak at 7 years, whilst Patient-19 started speaking at 4 years. Additionally, Patient-26 was affected by pyloric stenosis, a congenital anomaly in the digestive tract thought to be associated with impaired migration of the enteric neural crest. We consider it unlikely that the difference is caused solely by the mutations in *ARID1B*, since both patients show comparable reduction in ARID1B protein levels. We speculate that additional genetic factors may cooperate with *ARID1B* haploinsufficiency to determine the clinical severity of the syndrome. However, additional experiments with a larger set of patient-derived cell lines would be required to support this model. Furthermore, it would also be important to investigate other differentiation lineages to elucidate whether the ARID1A/ARID1B switch is only important for neural

crest differentiation or if, instead, it represents a more widespread mechanism utilized by stem cells to exit the pluripotent state and undergo lineage commitment.

Finally, further investigations will be necessary to elucidate the mechanism(s) responsible for the repressive activity of ARID1B-BAF. Recent studies have demonstrated that the function of BAF (including ARID1A-BAF) as transcriptional activator is mediated by the AP-1 transcription factors[56,60,61]. On the other hand, little is known of potential co-factors mediating the repressive function of ARID1B-BAF.

Acknowledgements

The authors are deeply grateful to the patients and their families, who donated the samples for this research. The authors thank the physicians and technicians who collected and cultured the skin samples. This work was funded by the G. Harold and Leila Y. Mathers Foundation (MT), and the Gisela Thier Fellowship (GWES). The authors thank the Genomic Facilities at Thomas Jefferson University and at the Wistar Institute for the sequencing effort. The authors are grateful to Kelly Vonk (Leiden University) for the help with experimental procedures in the iPSC generation. An invaluable support was provided to the authors by the Stem Cell and Regenerative Neuroscience Center at Thomas Jefferson University, and in particular by Dr. Elizabeth Kropf. Lastly, MT and GWES are thankful to Dr. Samantha Vergano (Children's Hospital of the King Daughters) for making this collaboration and this work possible.

Author contributions: MT, GWES and LP designed the project. GWES recruited the patients and obtained the skin fibroblasts. HMMM and CF reprogrammed the patient fibroblasts into iPSCs and assessed their quality. LD performed initial iPSC characterization experiments. LP performed most of the experiments, with crucial contributions from PP. ATC, CS, CAO, and SAW contributed to specific experiments (flowcytometry, immunoblots, mass-spectrometry). BC provided intellectual contribution and financial support to PP. SAB provided helpful discussion and editing. MT, LP, PP and SD analyzed the data. MT, GWES, PP, and LP wrote the manuscript, which was read and approved by all the authors.

Data availability: The original genome-wide data generate for this paper are deposited in the GEO database (accession number GSE169654).

Figure legends

Figure 1 – iPSCs derived from Coffin-Siris patients are pluripotent and proliferate normally. (a) Study system: iPSCs were derived from skin fibroblasts of two unrelated Coffin-Siris patients. The iPSCs were used in this study to generate Cranial Neural Crest Cells (CNCCs) and perform genomic experiments to investigate the effect of *ARID1B* mutations. **(b)** Graphical illustration of the *ARID1B* haploinsufficient mutations affecting the two studied patients. The numbers in the gene model refer to *ARID1B's* isoform NM_020732.3. **(c)** Colony of iPSCs derived from Patient-19 showing typical iPSC morphology. **(d, e)** Immunofluorescence and rt-qPCR quantifying the expression of the key pluripotency markers in iPSCs derived from Control and Patient lines. P-values are from Student's T-test (* = $p < 0.05$; ** $p < 0.005$; *** = $p < 0.0005$). **(f)** Growth curve comparing an *ARID1B*-wt Control iPSC Line with the two patient lines. The patient cells do not exhibit growth impairment. P-values are from Student's T-test (* = $p < 0.05$; ** $p < 0.005$; *** = $p < 0.0005$).

Figure 2 – CNCC differentiation is impaired in the patient cells. (a, b) CNCC differentiation was optimized using an *ARID1B*-wt Control Line. After 14 days, the cells exhibited the classic CNCC morphology and expressed the CNCC markers. P-values are from Student's T-test (* = $p < 0.05$; ** $p < 0.005$; *** = $p < 0.0005$). **(c)** Time-course immunoblot conducted using Control Line-1 during CNCC differentiation shows that ARID1B is active in the first 7 days of the differentiation, with a peak of activity between day-5 and

day-7. The ARID1B protein level strongly decreases after day-7. **(d, e)** Flowcytometry quantifying expression of surface markers for pluripotency and CNCC differentiation in control line-1 and in the two patient lines. A large cell population is still pluripotent in both patients after 14 days (d). The patient lines also show reduced expression of CNCC surface markers after 14 days of differentiation relative to an *ARID1B*-wt Control Line at the same time point (e).

Figure 3 – Aberrant NANOG and OCT4 activity in the patient-derived CNCC. Immunofluorescence for SOX9, NANOG, and OCT4 performed at the day-14 of iPSC-to-CNCC differentiation in Control Line-1, Patient-19 and Patient-26 cells. The patient cells are aberrantly NANOG-positive, and Patient-26 is also OCT4-positive.

Figure 4 – **Chromatin remodeling at pluripotency enhancers is dysregulated in the patient cells. (a)** At CNCC day-5, 29,578 ATAC-seq peaks are shared between patient and control lines. 5,540 peaks are specific of the patients. 578 peaks are specific of the controls. **(b)** UCSC Genome Browser example of a PATIENT-SPECIFIC ATAC-seq peak. **(c)** UCSC Genome Browser example of a CONTROL-SPECIFIC ATAC-seq peak. **(d)** ARID1B ChIP-seq heatmaps (Four lines; CNCC Day-5) centered on the PATIENT-SPECIFIC ATAC-seq peaks **(e)** Heatmaps of H3K27ac ChIP-seq time-course at the 5,540 PATIENT-SPECIFIC ATAC-seq peaks (Control Line-1). **(f)** H3K27ac ChIP-seq average profiles centered on the PATIENT-SPECIFIC ATAC-seq regions (CNCC Day-5). Statistical significance assessed with Wilcoxon's Rank Sum test ($p < 2.2 \times 10^{-16}$ in all the patient vs control comparisons). **(g)** Heatmaps of H3K27ac ChIP-seq signal at Days 5, 7, and 9 at the 5,540 PATIENT-SPECIFIC ATAC-seq peaks (Control Line-1, Patient-19, Patient-26). **(h)** Motif analysis at the PATIENT-SPECIFIC ATAC-seq regions revealed enrichment for the binding motif of multiple pluripotency factors.

Figure 5 – "**Pluripotency" and "Exit from Pluripotency" genes are dysregulated in differentiating patient CNCCs. (a)** RNA-seq volcano plot shows the differentially expressed genes between patient and control lines at CNCC Day-5. *ARID1B* is one of the top downregulated genes. **(b)** Top canonical pathways (IPA analysis) enriched in the set of 598 differentially expressed genes that also represent the closest gene to a PATIENT-SPECIFIC ATAC-seq peak. **(c)** Top upstream regulators (IPA analysis) enriched in the same set of 598 genes used for panel b. **(d)** RNA-seq heatmap displaying expression patterns at CNCC Day-5 for pluripotency genes, for genes of the NANOG network, and for genes associated to exit from pluripotency (PPARα/RXRα activation pathway).

Figure 6 – Aberrant SOX2 and NANOG activity in the patient cells at CNCC Day-5. (a) SOX2 ChIP-seq average profile for 3,284 patient-specific ATAC-seq peaks showing patient-specific SOX2 signal (spike-in normalized; CNCC Day-5). Statistical significance assessed with Wilcoxon's Rank Sum test ($p < 2.2 \times 10^{-16}$ in all the patient vs control comparisons). **(b)** ATAC-seq heatmaps at the 3,284 peaks shown in Fig. 5a reveal that these regions display increased chromatin accessibility in the patients relative to the two control lines. **(c)** ATAC-seq heatmaps at 4,538 patient-specific NANOG peaks display no changes in accessibility between patient and control lines. **(d)** NANOG ChIP-seq heatmaps at 4,538 patient-specific NANOG peaks (spike-in normalized; CNCC Day-5). **(e)** Example of patient-specific NANOG peak in a region with no chromatin accessibility (CNCC Day-5). **(f)** At CNCC Day-5, a cis-regulatory element in the promoter region of *NANOG* is more accessible in the patients than in the control lines. The same element also displays higher SOX2 binding in the patients than in the controls. **(g)** Knock-down of ARID1B from Control Line-1 also elicits an increase in chromatin accessibility at the cis-regulatory element in the promoter region of *NANOG* (CNCC Day-5).

Figure 7 – A switch between ARID1A-BAF and ARID1B-BAF is required for a successful exit from pluripotency. (a) Time-course immunoblot conducted using Control Line-1 during CNCC differentiation

shows that ARID1A is active at the iPSC stage, and abruptly downregulated at day-1 of differentiation. ARID1A protein level is upregulated again after day-7, mirroring ARID1B's downregulation at the same time point. **(b)** ARID1A time-course immunoblot in the two patient lines. **(c)** ARID1A immunoblot: both patient lines display aberrantly high ARID1A's protein level at CNCC Day-5. **(d)** ChIP-seq for ARID1A in the two control lines and the two patient lines at iPSC stage (day-0) and at CNCC Day-5. Heatmaps are centered on the 5,540 pluripotency enhancers.

MATERIALS AND METHODS

Human iPSC culture

Control iPSC lines were obtained from the iPSC Core of the University of Pennsylvania (Control line-1: SV20 line, male, age 43) and from the Coriell Institute for Medical Research (Camden, NJ. Control line-2: GM23716, female, age 16).

Skin fibroblasts from the two pediatric Coffin-Siris patients (one teenager one young adult) were obtained by the team of Dr. Gijs Santen at Leiden University. Patient 19 is a female, while Patient 26 is a male. The fibroblasts were reprogrammed into iPSCs with the polycistronic lentiviral vector LV.RRL.PPT.SF.hOKSM.idTomato.-preFRT by LUMC human iPSC Hotel as described elsewhere[62,63].

Multiple clones per line were derived. For each clone, pluripotency was assessed by immunofluorescence microscopy using antibodies against NANOG, OCT3/4, SSEA4 and Tra-1-81 under maintenance conditions and antibodies against (TUBB3, AFP and CD31) after spontaneous differentiation into the 3 germ layers as described elsewhere[62]. Clones with proper pluripotent characteristics were selected for downstream usage. Karyotyping by G binding was assessed for all the four lines by the Leiden University Medical Center and Short Tandem Repeat (STR) profiling was performed by the Leiden University Medical Center and and then replicated by the Stem Cell and Regenerative Neuroscience Center at Thomas Jefferson University. The iPSC lines were expanded in feeder-free, serum-free mTeSR™1 medium (STEMCELL Technologies). Cells were passaged ~1:10 at 80% confluency using ReLeSR (STEMCELL Technologies) and small cell clusters (50–200 cells) were subsequently plated on tissue culture dishes coated overnight with Geltrex™ LDEV-Free hESC-qualified Reduced Growth Factor Basement Membrane Matrix (Fisher-Scientific).

CNCC Differentiation

The iPSC lines were differentiated into CNCC as previously described[24]. Briefly, iPSCs were treated with CNCC Derivation media: 1:1 Neurobasal medium/D-MEM F-12 medium (Invitrogen), 0.5× B-27 supplement with Vitamin A (50× stock, Invitrogen), 0.5× N-2 supplement (100× stock, Invitrogen), 20 ng/ml bFGF (Biolegend), 20 ng/ml EGF (Sigma-Aldrich), 5 µg/ml bovine insulin (Sigma-Aldrich) and 1× Glutamax-I supplement (100× stock, Invitrogen). Medium (3ml) was changed every day. Three days after the appearance of the migratory CNCC, cells were detached using accutase and placed into geltrex-coated plates. The early migratory CNCCs were then transitioned to CNCC early maintenance media: 1:1 Neurobasal medium/D-MEM F-12 medium (Invitrogen), 0.5× B-27 supplement with Vitamin A (50× stock, Invitrogen), 0.5× N-2 supplement (100× stock, Invitrogen), 20 ng/ml bFGF (Biolegend), 20 ng/ml EGF (Sigma-Aldrich), 1 mg/ml bovine serum albumin, serum replacement grade (Gemini Bio-Products # 700-104P) and 1× Glutamax-I supplement (100× stock, Invitrogen).

ARID1B Knock-down

To make concentrated lentivirus, HEK293T cells were transfected with a pLenti plasmid in which we cloned an shRNA for *ARID1B* (GPP Web Portal: TRCN0000107361). iPSCs were lentivirally transduced by incubating the cells with concentrated virus overnight at 37 C. The next morning the media was changed, and 2 mg/ml puromycin (InvivoGen) were added 24h after infection. After 72 hours, the iPSCs that

survived the selection were then differentiated in CNCC using the above described protocol, and collected at Day-5 for the genomic experiments. The cells were kept under puromycin selection for the entire duration of the differentiation. The knock-down efficiency was quantified via western blot.

Flow cytometry analysis of surface markers

To obtain a single cell suspension for flow cytometry analysis, control and patient cells were treated with Accutase for 5 minutes. Cells were then washed with cold PBS-2% FBS and live cells were counted. 1×10^6 cells/condition were resuspended in 100 µL PBS-2% FBS and stained. For pluripotency evaluation, 4 µl of the respective antibodies were used: APC anti-human SSEA-4 antibody (Biolegend, #330417) and PE anti-human TRA-1-60-R antibody (Biolegend, #330609). For analysis of differentiation, 2 µl of the respective antibodies were used: FITC anti-human CD10 (Miltenyi Biotec, #130-124-262) and APC anti-human CD99 (Miltenyi Biotec, #130-121-096). Cells were incubated for 15 min on ice and protected from light, before transferring them into FACS tubes containing additional 300 µL PBS-2% FBS. Flow cytometry data were acquired using a BD LSR II flow cytometer and analyzed with FlowJo Software version 10.7.

Western Blot

For total lysate, cells were harvested and washed three times in 1X PBS and lysed in RIPA buffer (50mM Tris-HCl pH7.5, 150mM NaCl, 1% Igepal, 0.5% sodium deoxycholate, 0.1% SDS, 500uM DTT) with proteases inhibitors. Twenty µg of whole cell lysate were loaded in Novex WedgeWell 4-20% Tris-Glycine Gel (Invitrogen) and separated through gel electrophoresis (SDS-PAGE) Tris-Glycine-SDS buffer (Invitrogen). The proteins were then transferred to ImmunBlot PVDF membranes (ThermoFisher) for antibody probing. Membranes were incubated with 10% BSA in TBST for 30 minutes at room temperature (RT), then incubated for variable times with the suitable antibodies diluted in 5% BSA in 1X TBST, washed with TBST and incubated with a dilution of 1:10000 of secondary antibody for one hour at RT. The antibody was then visualized using Super Signal West Dura Extended Duration Substrat (ThermoFisher) and imaged with Amersham Imager 680.

Cell fractionation

5×10^6cells/condition were collected and suspended in E1 buffer (50mM HEPES-KOH, 140mM NaCl, 1mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, 1mM DTT, 1X Proteinase Inhibitor) followed by a centrifugation step of 1100 g at 4°C for 2min. The cytoplasmic fraction was collected in a fresh tube. Cells were washed two more times with E1 buffer. Pellet was subsequently suspended in E2 buffer (10mM Tris-HCl, 200mM NaCl, 1mM EDTA, 0.5mM EGTA, 1X Proteinase Inhibitor) followed by a centrifugation step of 1100 g at 4°C for 2 min. Nuclear fraction was collected in a fresh tube. Cells were washed two more times with E2 buffer. After the third wash, pellet was suspended in E3 buffer (500mM Tris-HCl, 500mM NaCl, 1X Proteinase Inhibitor) and sonicated for 15 sec (5 sec ON/ 5 sec OFF). Cytoplasmic, nuclear and chromatin fraction were centrifuge at 16000 g for 10min at 4°C.

Antibodies

ARID1B ChIP-Seq: Abcam ab57461. ARID1B western blot: Santa-Cruz sc-32762 and Abcam ab57461. ARID1A ChIP-Seq: GeneTex GTX129433. ARID1A western blot: Cell Signaling Technologies 12354S. Beta-Actin western blot: Cell Signaling Technologies 8457P. SOX2 ChIP-Seq: Active Motif 39843. NANOG ChIP-Seq: R&D Systems AF1997. H3K27ac ChIP-Seq: Abcam ab4729. GAPDH western blot: Cell Signaling Technologies 5174T. CD10 Flow Cytometry: Miltenyi Biotech 130-124-262. CD99 Flow Cytometry: Miltenyi Biotech 130-121-086. SSEA4 Flow Cytometry: Biolegend 330417. TRA-1-60-R Flow Cytometry: Biolegend 330609. IgG ChIP-qPCR: Cell Signaling Technologies 2729S. Cell Signaling HRP-conjugated anti-rabbit (7074S) and anti-mouse (7076S) were used as secondary antibodies in western blot. Spike-in Antibody: Active Motif 61686. Spike-in Chromatin: Active Motif 53083. Antibodies used in immunofluorescence:

Anti-Mouse OCT4 (STEMCELL TECHNOLOGIES, 60059, 1:200); Rabbit Monoclonal Anti-Sox9 (abcam, ab185230,1:250); pPolyclonal Goat Anti-Nanog (R&D System, AF1997-SP,1:20); donkey anti-goat IgG (H+L) Alexa 488 (Jackson ImmunoResearch, 705-545-003, 1:500); donkey anti-mouse IgG (H+L) Alexa 647 (Jackson ImmunoResearch, 715-605-150, 1:500); Donkey anti-rabbit IgG (H+L) Cy3 (Jackson ImmunoResearch, 11-165-152, 1:500). , Monoclonal mouse anti-human TRA-1-60 Antibody (Millipore MAB4360C3,1:100), Monoclonal mouse Anti-Stage-Specific Embryonic Antigen-4 Antibody (Millipore, MAB4304,1:100), Polyclonal Goat Anti- Human/Mouse Oct-3/4 Antibody (R&D System, AF1759,1:20).

Immunofluorescence

Upon fixation (4% PFA for 10 minutes), cells were permeabilized in blocking solution (0.1% Triton X-100, re PBS, 5% normal donkey serum) and then incubated with the antibody of interest. The total number of cells in each field was determined by counterstaining cell nuclei with 4,6-diamidine-2-phenylindole dihydrochloride (DAPI; Sigma-Aldrich; 50 mg/ml in PBS for 15 min at RT). Immunostained cells were mounted in Aqua-Poly/Mount (Polysciences) and analyzed at epi-fluorescent or confocal microscopy, using a Nikon A1R+. Images were captured with a ×40 objectives and a pinhole of 1.2 Airy unit. Analyses were performed in sequential scanning mode to rule out cross-bleeding between channels.

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

Cells were lysed in Tri-reagent and RNA was extracted using the Direct-zol RNA MiniPrep kit (Zymo research). 600ng of template RNA was retrotranscribed into cDNA using RevertAid first strand cDNA synthesis kit (Thermo Scientific) according to manufacturer directions. 15ng of cDNA were used for each real-time quantitative PCR reaction with 0.1 µM of each primer, 10 µL of PowerUp™ SYBR™ Green Master Mix (Applied Biosystems) in a final volume of 20 µl, using QuantStudio 3 Real-Time PCR System (Applied Biosystem). Thermal cycling parameters were set as following: 3 minutes at 95°C, followed by 40 cycles of 10 s at 95°C, 20 s at 63°C followed by 30 s at 72°C. Each sample was run in triplicate. 18S rRNA was used as normalizer. Primer sequences are reported in Supplementary Table S1.

ChIP-Seq and ChiP-qPCR

Samples from different conditions were processed together to prevent batch effects.

For SOX2, NANOG and H3K27ac, for each replicate, 10 million cells were cross-linked with 1% formaldehyde for 5 min at room temperature, quenched with 125mM glycine, harvested and washed twice with 1× PBS. The pellet was resuspended in ChIP lysis buffer (150 mM NaCl, 1% Triton X-100, 0,7% SDS, 500 µM DTT, 10 mM Tris-HCl, 5 mM EDTA) and chromatin was sheared to an average length of 200–500 bp, using a Covaris S220 Ultrasonicator. The chromatin lysate was diluted with SDS-free ChIP lysis buffer. For ChIP-seq, 10 µg of antibody (3 µg for H3K27ac) was added to 5 µg of sonicated chromatin along with Dynabeads Protein A magnetic beads (Invitrogen) and incubated at 4 °C overnight. For SOX2 and NANOG ChIP-seq, 10 ng of spike-in Drosophila chromatin (Active Motif) was added to each sample with 2 µg spike-in antibody (Active Motif). On day 2, beads were washed twice with each of the following buffers: Mixed Micelle Buffer (150 mM NaCl, 1% Triton X-100, 0.2% SDS, 20 mM Tris-HCl, 5 mM EDTA, 65% sucrose), Buffer 500 (500 mM NaCl, 1% Triton X-100, 0.1% Na deoxycholate, 25 mM HEPES, 10 mM Tris-HCl, 1 mM EDTA), LiCl/detergent wash (250 mM LiCl, 0.5% Na deoxycholate, 0.5% NP-40, 10 mM Tris-HCl, 1 mM EDTA) and a final wash was performed with 1× TE. Finally, beads were resuspended in 1× TE containing 1% SDS and incubated at 65 °C for 10 min to elute immunocomplexes. Elution was repeated twice, and the samples were further incubated overnight at 65 °C to reverse cross-linking, along with the untreated input (5% of the starting material). On day 3, after treatment with 0.5 mg/ml Proteinase K for 1h at 65 °C, DNA was purified with Zymo ChIP DNA Clear Concentrator kit and quantified with QUBIT.

For ARID1A and ARID1B ChIP-Seq, 10 million cells were cross-linked with EGS (150 mM) for 30min at room temperature followed by a second cross-link with 1% formaldehyde for 15 min at room temperature. The

formaldehyde was quenched with by adding glycine (0.125M) for 10 min at room temperature. Cells were washed twice with 1× PBS. Pellet was resuspended in buffer LB1 (50 mM Hepes-KOH, 140 mM NaCl, 1 mM EDTA, 10% Glycerol, 0.5% NP-40, 0.255 Triton X-100), incubated 10 min at 4 °C followed by a centrifugation step of 600g for 5 min at 4 °C. Pellet was suspended in buffer LB2 (10 mM Tris-HCl, 20 mM NaCl, 1 mM EDTA, 0.5 mM EGTA) incubated 10 min at 4 °C followed by a centrifugation step of 600g for 5 min at 4 °C. Cells were then resuspended in buffer LB3 (10 mM Tris-HCl, 200 mM NaCl, 1mM EDTA, 0.5 mM EGTA, 0.1% Na-DOC, 0.5% N-laurosylsarcosine) incubated 10 min at 4 °C followed by a centrifugation step of 600g for 5min at 4 °C. Pellet was suspended in LB3 and chromatin was sheared to an average length of 200–500 bp, using a Covaris S220 Ultrasonicator. For each sample, 15 ug of sonicated chromatin was incubated at 4 °C overnight along with Dynabeads Protein G conjugated with 10ug of antibody. On day 2, beads were washed once with each of the following buffers: WB1 (50 mM Tris-HCl, 150 mM NaCl, 0.15 SDS, 0.1% Na-DOC, 1% Triton X-100, 1 mM EDTA), WB2 (50 mM Tris-HCl, 500 mM NaCl, 0.15 SDS, 0.1% Na-DOC, 1% Triton X-100, 1 mM EDTA), WB3 (10 mM Tris-HCl, 250 mM LiCL, 0.55 NP-40. 0.55 Na-DOC, 1 mM EDTA), TE Buffer (10 mM Tris-HCl, 1mM EDTA). Finally, beads were resuspended in EB (10 mM tris-HCl, 0.55 SDS, 300 mM NaCl, 5mM EDTA) and incubated at 65 °C for 30 min to elute immunocomplexes. Elution was repeated twice, and the samples were further incubated overnight at 65 °C to reverse cross-linking, along with the untreated input (5% of the starting material). On day 3, after treatment with 0.5 mg/ml Proteinase K for 1h at 65 °C.

For all ChIP-seq experiments, barcoded libraries were made with NEB ULTRA II DNA Library Prep Kit for Illumina, and sequenced on Illumina NextSeq 500, producing 75bp SE reads.

For ChIP-qPCR, on day 1 the sonicated lysate was aliquot into single immunoprecipitations of 2.5×10^6 cells each. A specific antibody or a total rabbit IgG control was added to the lysate along with Protein A magnetic beads (Invitrogen) and incubated at 4 °C overnight. On day3, ChIP eluates and input were assayed by real-time quantitative PCR in a 20 µl reaction with the following: 0.4 µM of each primer, 10 µl of PowerUp SYBR Green (Applied Biosystems), and 5 µl of template DNA (corresponding to 1/40 of the elution material) using the fast program on QuantStudio qPCR machine (Applied Biosystems). Thermal cycling parameters were: 20sec at 95 °C, followed by 40 cycles of 1sec at 95°C, 20sec at 60°C.

ChIP-seq Analyses

After removing the adapters, the sequences were aligned to the reference hg19, using Burrows Wheeler Alignment tool (BWA), with the MEM algorithm[64] . Aligned reads were filtered based on mapping quality (MAPQ > 10) to restrict our analysis to higher quality and likely uniquely mapped reads, and PCR duplicates were removed. We called peaks for each individual using MACS2[65] (H3K27ac) or Homer[66], at 5% FDR, with default parameters.

RNA-Seq

Cells were lysed in Tri-reagent (Zymo research) and total RNA was extracted using Quick-RNA Miniprep kit (Zymo research) according to the manufacturer's instructions. RNA was further quantified using DeNovix DS-11 Spectrophotometer while the RNA integrity was checked on Bioanalyzer 2100 (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® and NEBNext® UltraTM II DNA Library Prep Kit for Illumina® according to the manufacturer's instructions (New England Biolabs).

RNA-Seq Analyses

Reads were aligned to hg19 using STAR v2.5[67], in 2-pass mode with the following parameters: --quantMode TranscriptomeSAM --outFilterMultimapNmax 10 - -outFilterMismatchNmax 10 --outFilterMismatchNoverLmax 0.3 --alignIntronMin 21 -- alignIntronMax 0 --alignMatesGapMax 0 --

alignSJoverhangMin 5 --runThreadN 12 -- twopassMode Basic --twopass1readsN 60000000 --sjdbOverhang 100. We filtered bam files based on alignment quality (q = 10) using Samtools v0.1.19[64]. We used the latest annotations obtained from Ensembl to build reference indexes for the STAR alignment. Kallisto[68] was used to count reads mapping to each gene. RSEM[69] was instead used to obtain FPKM (Fragments Per Kilobase of exon per Million fragments mapped). We analyzed differential gene expression levels with DESeq2[70], with the following model: design = ~condition, where condition indicates either CTRL or Patients.

ATAC-Seq

For ATAC-Seq experiments, 50,000 cells per condition were processed as described in the original ATAC-seq protocol paper[71]. ATAC-seq data were processed with the same pipeline described for ChIP-seq, with one modification: all mapped reads were offset by +4 bp for the forward-strand and -5 bp for the reverse-strand. After peak calling (MACS2), peaks replicated in all 4 lines (hereafter consensus peaks) were used for downstream analyses.

Nuclear extract, IP and LC-MS/MS

After collection, cells were washed twice with ice cold PBS before resuspension in co-IP buffer (20mM Tris pH 7.9, 100mM NaCl, 0.1% NP-40, 0.5mM DTT, protease inhibitors), and rotated for 5 minutes at 4°C. After spinning down at 2000rpm for 10 minutes, the nuclear pellet was resuspended in buffer C (20mM Tris pH 8.0, 1.5mM MgCl2, 0.42M NaCl, 25% glycerol, 0.2mM EDTA, 0.5mM DTT, protease inhibitors), dounce homogenized (with B pestle), and incubated at 4°C for 30 minutes. The extract was centrifuged at 12,000rpm for 30 minutes, and the supernatant was kept as nuclear extract. The nuclear extract was dialyzed overnight in BC80 (20mM Tris pH 8.0, 80mM KCl, 0.2mM EDTA, 10% glycerol, 1mM B-mercaptoethanol, 0.2mM phenylmethylsulfonyl fluoride (PMSF)), cleared, and stored at -80°C. For the IP, 1.5mg of nuclear extract was incubated for 3 hours at 4°C with 6µg ARID1B antibody and 50µL of Dynabeads Protein A, and the control IP was performed with 0.75mg of nuclear extract and 25µL of Dynabeads Protein A. Beads were washed three times with co-IP buffer, followed by a final wash with 0.05% NP-40 in PBS. Elution was performed by agitation in 0.1M glycine pH 3.0 for one minute, and 1M Tris base pH 11.0 was added to neutralize the pH of the eluate. Eluates were prepared for SDS-PAGE and run on a Novex WedgeWell 10% Tris-Glycine Gel (Invitrogen) with Tris-Glycine-SDS buffer (Bio-Rad), at 110V for 10 minutes. The gel was stained with Colloidal Blue staining kit (Invitrogen), and further processed at the proteomics facility at the Wistar Institute. Briefly, the gel lanes were excised, reduced with TCEP, alkylated with iodoacetamide, and digested with trypsin. Tryptic digests were analyzed using LC-MS/MS (a standard 90 minute LC gradient on the Thermo Q Exactive HF mass spectrometer). MS/MS spectra were searched with full tryptic specificity against the UniProt human database (10/02/2020) using MaxQuant 1.6.17.0, and also searched for the common protein N-terminal acetylation, Asn deamidation, and Met oxidation. Protein and peptide false discovery rate was set at 1%.

Statistical and genomic analyses

All statistical analyses were performed using R v3.3.1. BEDtools v2.27.1[72] was used for genomic analyses. Pathway analysis was performed with Ingenuity Pathway Analysis Suite (QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis). Motif analyses were performed using the Meme-Suite[73], and specifically with the Meme-ChIP application. Fasta files of the regions of interest were produced using BEDTools v2.27.1. Shuffled input sequences were used as background. E-values < 0.001 were used as threshold for significance[73].

References

1. Wang, W. *et al.* Purification and biochemical heterogeneity of the mammalian SWI-SNF complex. *EMBO J* **15**, 5370-82 (1996).
2. Braun, S.M.G. *et al.* BAF subunit switching regulates chromatin accessibility to control cell cycle exit in the developing mammalian cortex. *Genes Dev* **35**, 335-353 (2021).
3. Lessard, J. *et al.* An essential switch in subunit composition of a chromatin remodeling complex during neural development. *Neuron* **55**, 201-15 (2007).
4. Gatchalian, J. *et al.* A non-canonical BRD9-containing BAF chromatin remodeling complex regulates naive pluripotency in mouse embryonic stem cells. *Nat Commun* **9**, 5139 (2018).
5. Ho, L. *et al.* An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proc Natl Acad Sci U S A* **106**, 5187-91 (2009).
6. Ho, L. *et al.* An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proc Natl Acad Sci U S A* **106**, 5181-6 (2009).
7. Alfert, A., Moreno, N. & Kerl, K. The BAF complex in development and disease. *Epigenetics Chromatin* **12**, 19 (2019).
8. Halgren, C. *et al.* Corpus callosum abnormalities, intellectual disability, speech impairment, and autism in patients with haploinsufficiency of ARID1B. *Clin Genet* **82**, 248-55 (2012).
9. Hoyer, J. *et al.* Haploinsufficiency of ARID1B, a member of the SWI/SNF-a chromatin-remodeling complex, is a frequent cause of intellectual disability. *Am J Hum Genet* **90**, 565-72 (2012).
10. Santen, G.W. *et al.* Mutations in SWI/SNF chromatin remodeling complex gene ARID1B cause Coffin-Siris syndrome. *Nat Genet* **44**, 379-80 (2012).
11. Tsurusaki, Y. *et al.* Coffin-Siris syndrome is a SWI/SNF complex disorder. *Clin Genet* **85**, 548-54 (2014).
12. van der Sluijs, P.J. *et al.* The ARID1B spectrum in 143 patients: from nonsyndromic intellectual disability to Coffin-Siris syndrome. *Genet Med* **21**, 1295-1307 (2019).
13. Vergano, S.A., van der Sluijs, P.J. & Santen, G. ARID1B-Related Disorder. in *GeneReviews((R))* (eds. Adam, M.P. *et al.*) (Seattle (WA), 1993).
14. Santen, G.W. *et al.* Coffin-Siris syndrome and the BAF complex: genotype-phenotype study in 63 patients. *Hum Mutat* **34**, 1519-28 (2013).
15. Wieczorek, D. *et al.* A comprehensive molecular study on Coffin-Siris and Nicolaides-Baraitser syndromes identifies a broad molecular and clinical spectrum converging on altered chromatin remodeling. *Hum Mol Genet* **22**, 5121-35 (2013).
16. Wright, C.F. *et al.* Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet* **385**, 1305-14 (2015).
17. Jung, E.M. *et al.* Arid1b haploinsufficiency disrupts cortical interneuron development and mouse behavior. *Nat Neurosci* **20**, 1694-1707 (2017).
18. Ka, M., Chopra, D.A., Dravid, S.M. & Kim, W.Y. Essential Roles for ARID1B in Dendritic Arborization and Spine Morphology of Developing Pyramidal Neurons. *J Neurosci* **36**, 2723-42 (2016).
19. Shibutani, M. *et al.* Arid1b Haploinsufficiency Causes Abnormal Brain Gene Expression and Autism-Related Behaviors in Mice. *Int J Mol Sci* **18**(2017).

20. Smith, A.L., Jung, E.M., Jeon, B.T. & Kim, W.Y. Arid1b haploinsufficiency in parvalbumin- or somatostatin-expressing interneurons leads to distinct ASD-like and ID-like behavior. *Sci Rep* **10**, 7834 (2020).
21. Sausen, M. *et al.* Integrated genomic analyses identify ARID1A and ARID1B alterations in the childhood cancer neuroblastoma. *Nat Genet* **45**, 12-7 (2013).
22. Mashtalir, N. *et al.* Modular Organization and Assembly of SWI/SNF Family Chromatin Remodeling Complexes. *Cell* **175**, 1272-1288 e20 (2018).
23. Inoue, K. *et al.* Molecular mechanism for distinct neurological phenotypes conveyed by allelic truncating mutations. *Nat Genet* **36**, 361-9 (2004).
24. Prescott, S.L. *et al.* Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* **163**, 68-83 (2015).
25. Liu, X. *et al.* De Novo ARID1B mutations cause growth delay associated with aberrant Wnt/beta-catenin signaling. *Hum Mutat* **41**, 1012-1024 (2020).
26. Vasileiou, G. *et al.* Chromatin-Remodeling-Factor ARID1B Represses Wnt/beta-Catenin Signaling. *Am J Hum Genet* **97**, 445-56 (2015).
27. Lee, J., Lee, J. & Cho, Y.S. Peroxisome Proliferator-Activated Receptor alpha Agonist and Its Target Nanog Cooperate to Induce Pluripotency. *J Clin Med* **7**(2018).
28. Mullen, E.M., Gu, P. & Cooney, A.J. Nuclear Receptors in Regulation of Mouse ES Cell Pluripotency and Differentiation. *PPAR Res* **2007**, 61563 (2007).
29. Rajasingh, J. & Bright, J.J. 15-Deoxy-delta12,14-prostaglandin J2 regulates leukemia inhibitory factor signaling through JAK-STAT pathway in mouse embryonic stem cells. *Exp Cell Res* **312**, 2538-46 (2006).
30. Ellegood, J. *et al.* Neuroanatomy and behavior in mice with a haploinsufficiency of AT-rich interactive domain 1B (ARID1B) throughout development. *Mol Autism* **12**, 25 (2021).
31. Dodonova, S.O., Zhu, F., Dienemann, C., Taipale, J. & Cramer, P. Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. *Nature* **580**, 669-672 (2020).
32. King, H.W. & Klose, R.J. The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. *Elife* **6**(2017).
33. Blassberg, R. *et al.* Sox2 levels configure the WNT response of epiblast progenitors responsible for vertebrate body formation. 2020.12.29.424684 (2020).
34. Bunina, D. *et al.* Genomic Rewiring of SOX2 Chromatin Interaction Network during Differentiation of ESCs to Postmitotic Neurons. *Cell Syst* **10**, 480-494 e8 (2020).
35. Heurtier, V. *et al.* The molecular logic of Nanog-induced self-renewal in mouse embryonic stem cells. *Nat Commun* **10**, 1109 (2019).
36. Novo, C.L. *et al.* The pluripotency factor Nanog regulates pericentromeric heterochromatin organization in mouse embryonic stem cells. *Genes Dev* **30**, 1101-15 (2016).
37. De Kumar, B. *et al.* Dynamic regulation of Nanog and stem cell-signaling pathways by Hoxa1 during early neuro-ectodermal differentiation of ES cells. *Proc Natl Acad Sci U S A* **114**, 5838-5845 (2017).
38. Navarro, P. *et al.* OCT4/SOX2-independent Nanog autorepression modulates heterogeneous Nanog gene expression in mouse ES cells. *EMBO J* **31**, 4547-62 (2012).

39. Silva, J. *et al.* Nanog is the gateway to the pluripotent ground state. *Cell* **138**, 722-37 (2009).
40. Gagliardi, A. *et al.* A direct physical interaction between Nanog and Sox2 regulates embryonic stem cell self-renewal. *EMBO J* **32**, 2231-47 (2013).
41. Rodda, D.J. *et al.* Transcriptional regulation of nanog by OCT4 and SOX2. *J Biol Chem* **280**, 24731-7 (2005).
42. Shang, L. *et al.* Mutations in ARID2 are associated with intellectual disabilities. *Neurogenetics* **16**, 307-14 (2015).
43. Trizzino, M. *et al.* The Tumor Suppressor ARID1A Controls Global Transcription via Pausing of RNA Polymerase II. *Cell Rep* **23**, 3933-3945 (2018).
44. Raab, J.R., Resnick, S. & Magnuson, T. Genome-Wide Transcriptional Regulation Mediated by Biochemically Distinct SWI/SNF Complexes. *PLoS Genet* **11**, e1005748 (2015).
45. Lu, J. *et al.* Stem cell factor SALL4 represses the transcriptions of PTEN and SALL1 through an epigenetic repressor complex. *PLoS One* **4**, e5577 (2009).
46. Miller, A. *et al.* Sall4 controls differentiation of pluripotent cells independently of the Nucleosome Remodelling and Deacetylation (NuRD) complex. *Development* **143**, 3074-84 (2016).
47. Tatetsu, H. *et al.* SALL4, the missing link between stem cells, development and cancer. *Gene* **584**, 111-9 (2016).
48. Yang, J. *et al.* Genome-wide analysis reveals Sall4 to be a major regulator of pluripotency in murine-embryonic stem cells. *Proc Natl Acad Sci U S A* **105**, 19756-61 (2008).
49. Pantier, R. *et al.* SALL4 controls cell fate in response to DNA base composition. *Mol Cell* **81**, 845-858 e8 (2021).
50. Kohlhase, J. SALL4-Related Disorders. in *GeneReviews((R))* (eds. Adam, M.P. *et al.*) (Seattle (WA), 1993).
51. Pagliaroli, L. & Trizzino, M. The Evolutionary Conserved SWI/SNF Subunits ARID1A and ARID1B Are Key Modulators of Pluripotency and Cell-Fate Determination. *Front Cell Dev Biol* **9**, 643361 (2021).
52. Zhang, Y. *et al.* A switch from hBrm to Brg1 at IFNgamma-activated sequences mediates the activation of human genes. *Cell Res* **20**, 1345-60 (2010).
53. Betschinger, J. *et al.* Exit from pluripotency is gated by intracellular redistribution of the bHLH transcription factor Tfe3. *Cell* **153**, 335-47 (2013).
54. Jaenisch, R. & Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132**, 567-82 (2008).
55. Takahashi, K. & Yamanaka, S. A decade of transcription factor-mediated reprogramming to pluripotency. *Nat Rev Mol Cell Biol* **17**, 183-93 (2016).
56. Gao, F. *et al.* Heterozygous Mutations in SMARCA2 Reprogram the Enhancer Landscape by Global Retargeting of SMARCA4. *Mol Cell* **75**, 891-904 e7 (2019).
57. Ejaz, R., Babul-Hirji, R. & Chitayat, D. The evolving features of Nicolaides-Baraitser syndrome - a clinical report of a 20-year follow-up. *Clin Case Rep* **4**, 351-5 (2016).
58. Wolff, D. *et al.* In-Frame Deletion and Missense Mutations of the C-Terminal Helicase Domain of SMARCA2 in Three Patients with Nicolaides-Baraitser Syndrome. *Mol Syndromol* **2**, 237-244 (2012).

59. Kadoch, C. & Crabtree, G.R. Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Sci Adv* **1**, e1500447 (2015).
60. Mathur, R. *et al.* ARID1A loss impairs enhancer-mediated gene regulation and drives colon cancer in mice. *Nat Genet* **49**, 296-302 (2017).
61. Vierbuchen, T. *et al.* AP-1 Transcription Factors and the BAF Complex Mediate Signal-Dependent Enhancer Selection. *Mol Cell* **68**, 1067-1082 e12 (2017).
62. Warlich, E. *et al.* Lentiviral vector design and imaging approaches to visualize the early stages of cellular reprogramming. *Mol Ther* **19**, 782-9 (2011).
63. Chen, J. *et al.* Rational optimization of reprogramming culture conditions for the generation of induced pluripotent stem cells with ultra-high efficiency and fast kinetics. *Cell Res* **21**, 884-94 (2011).
64. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).
65. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
66. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-89 (2010).
67. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
68. Bray, N.L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525-7 (2016).
69. Li, B. & Dewey, C.N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
70. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
71. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. & Greenleaf, W.J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-8 (2013).
72. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-2 (2010).
73. Bailey, T.L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**, W202-8 (2009).

FIGURE 1

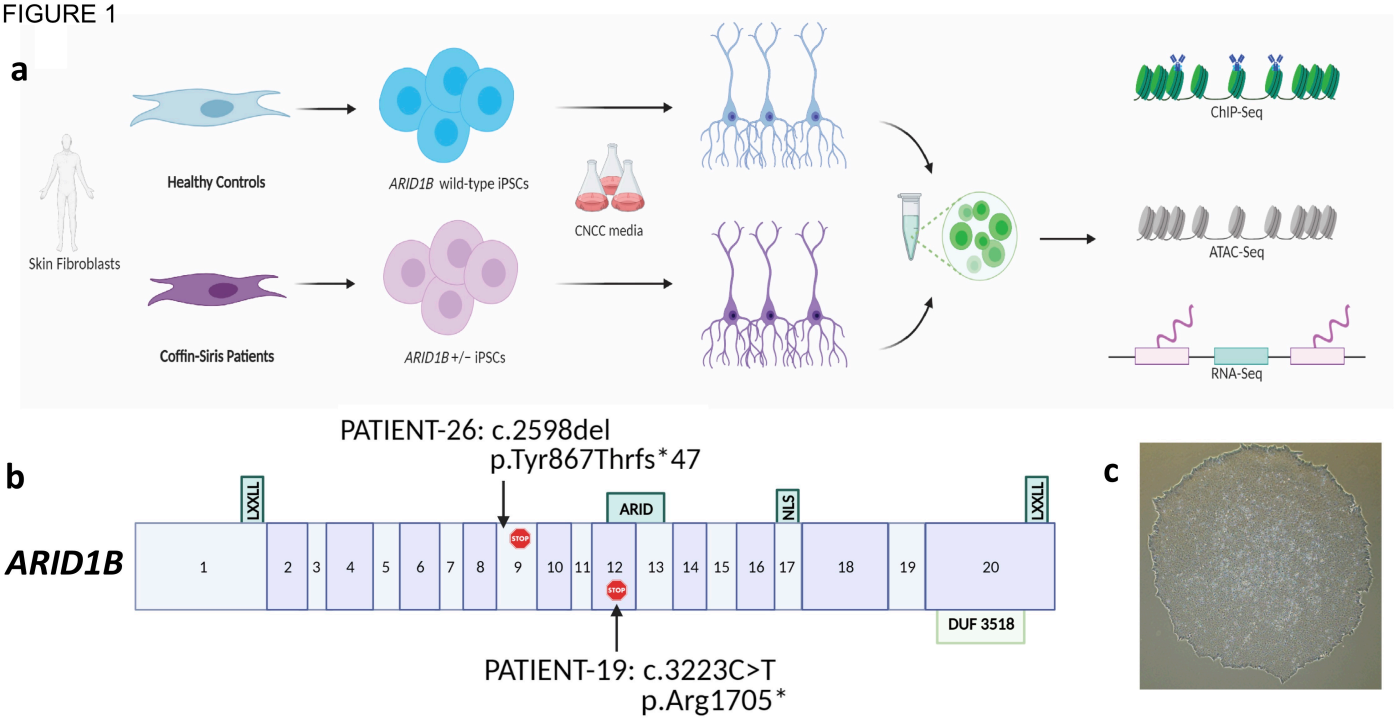
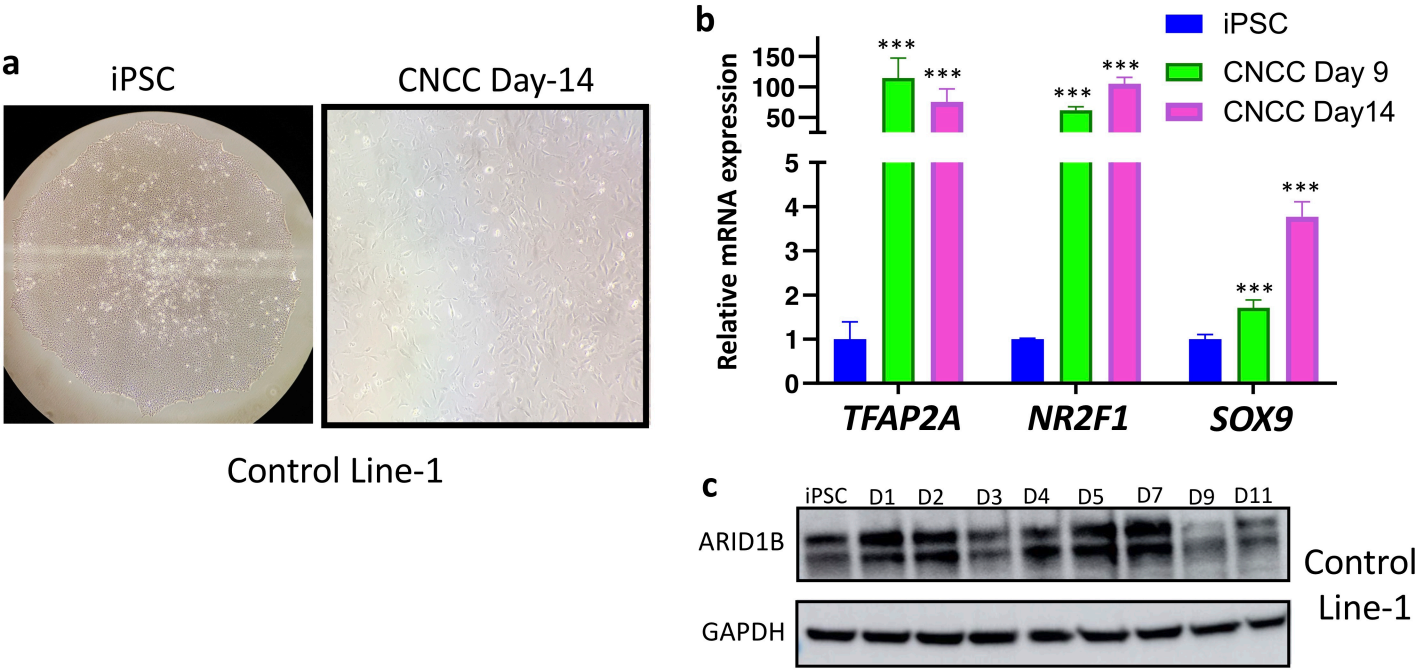
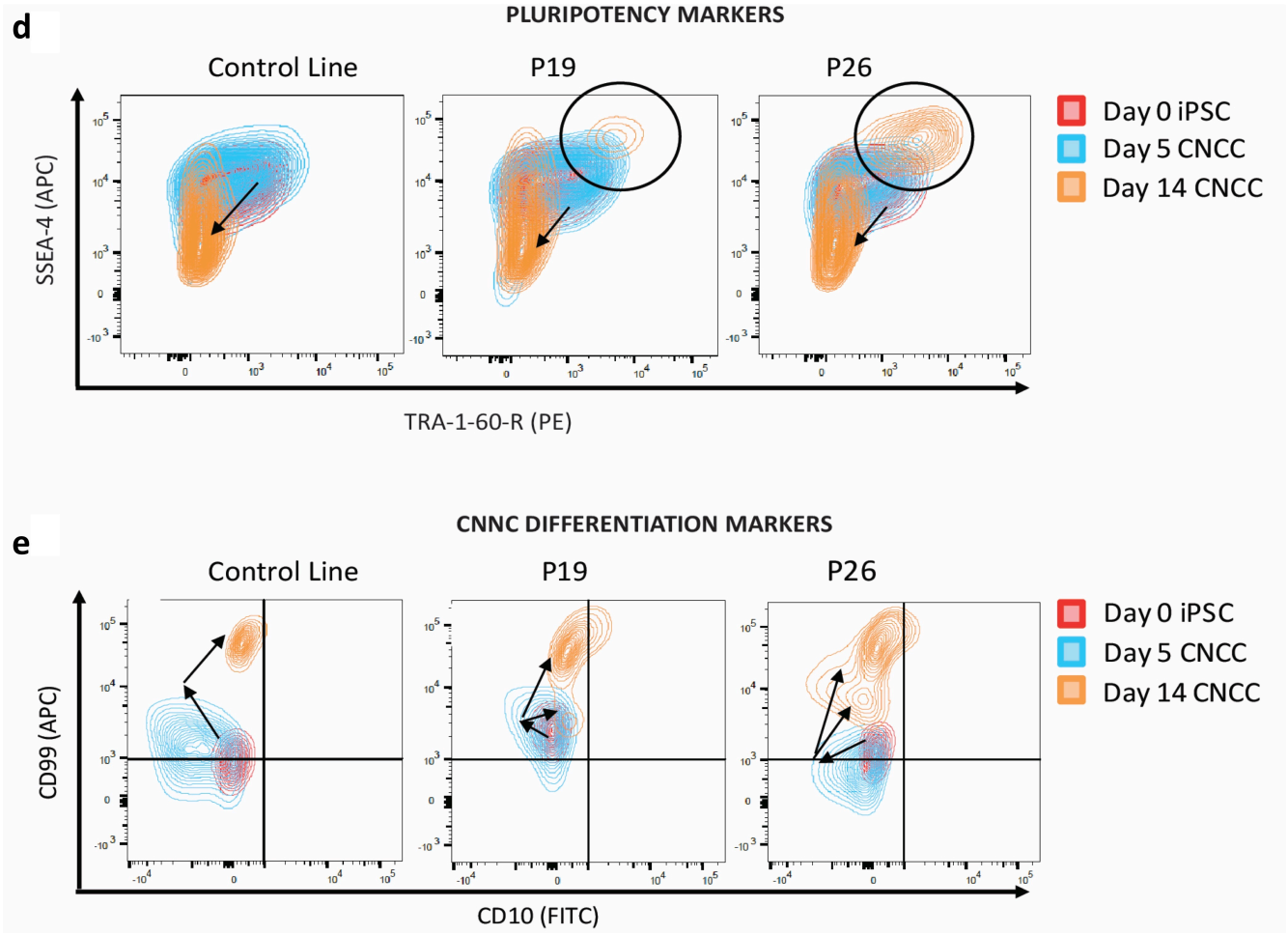


FIGURE 2





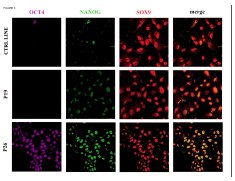
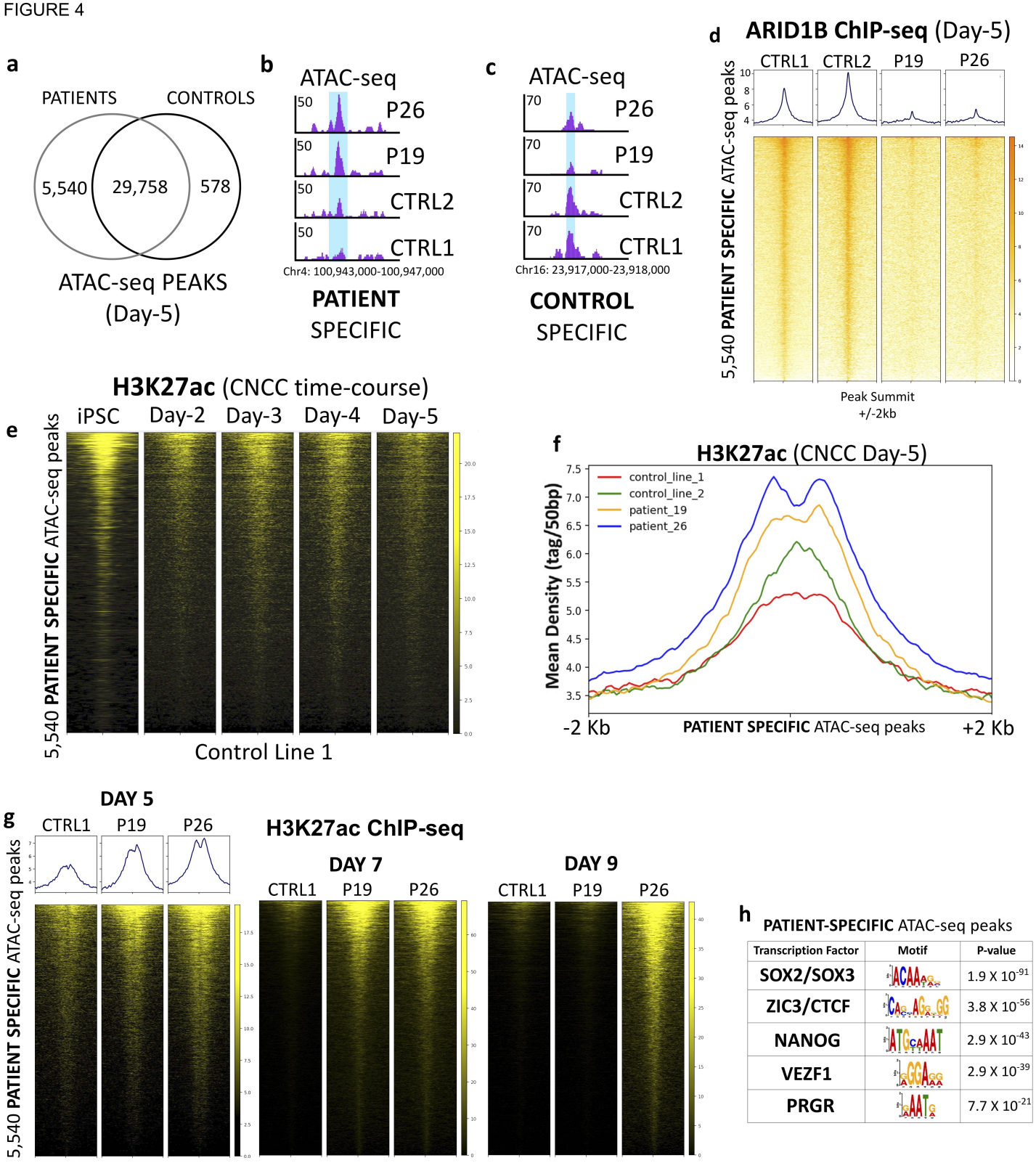


FIGURE 4

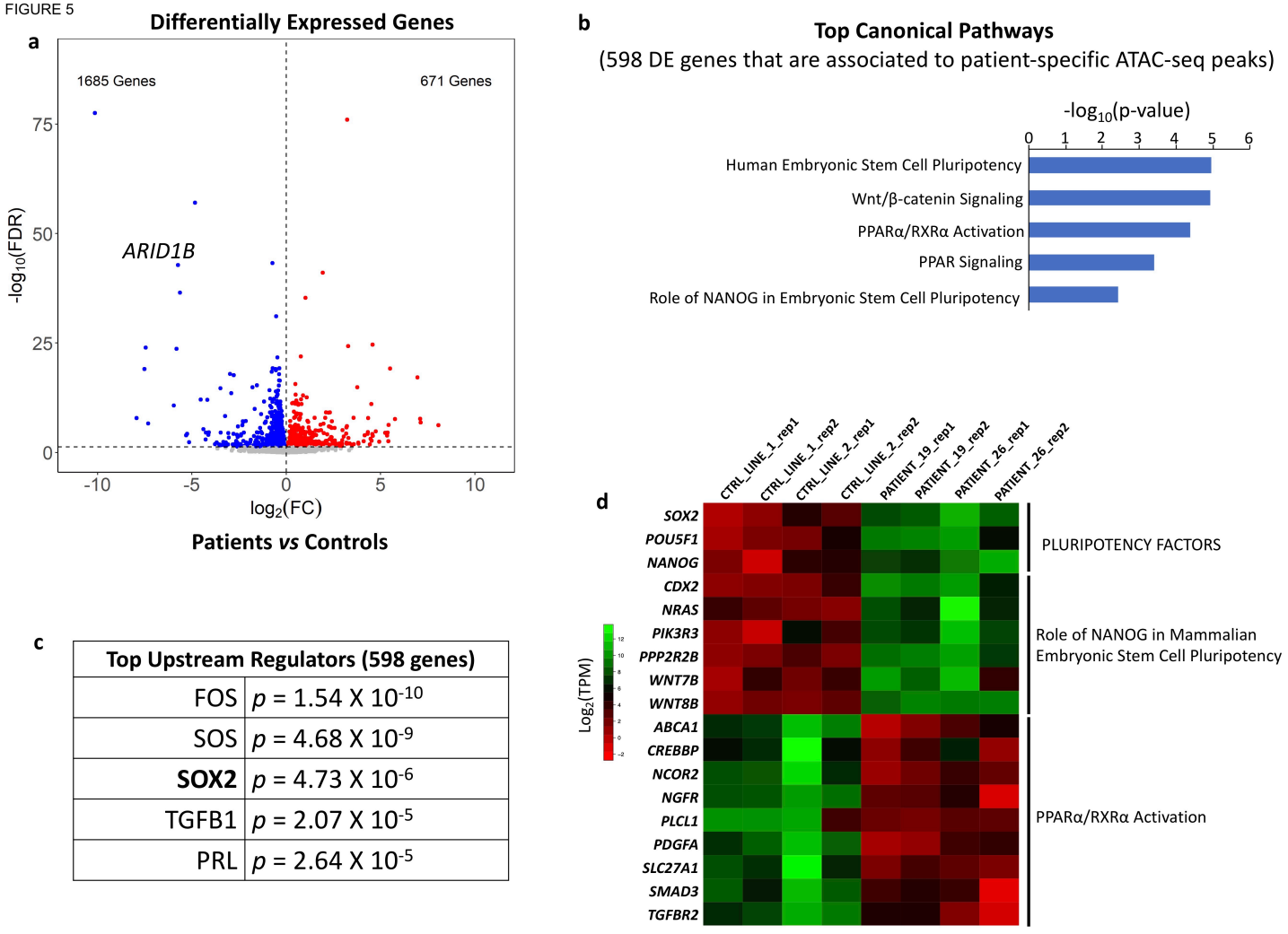


h

PATIENT-SPECIFIC ATAC-seq peaks

Transcription Factor	Motif	P-value
SOX2/SOX3		1.9×10^{-91}
ZIC3/CTCF		3.8×10^{-56}
NANOG		2.9×10^{-43}
VEZF1		2.9×10^{-39}
PRGR		7.7×10^{-21}

FIGURE 5



c

Top Upstream Regulators (598 genes)	
FOS	$p = 1.54 \times 10^{-10}$
SOS	$p = 4.68 \times 10^{-9}$
SOX2	$p = 4.73 \times 10^{-6}$
TGFB1	$p = 2.07 \times 10^{-5}$
PRL	$p = 2.64 \times 10^{-5}$

FIGURE 6

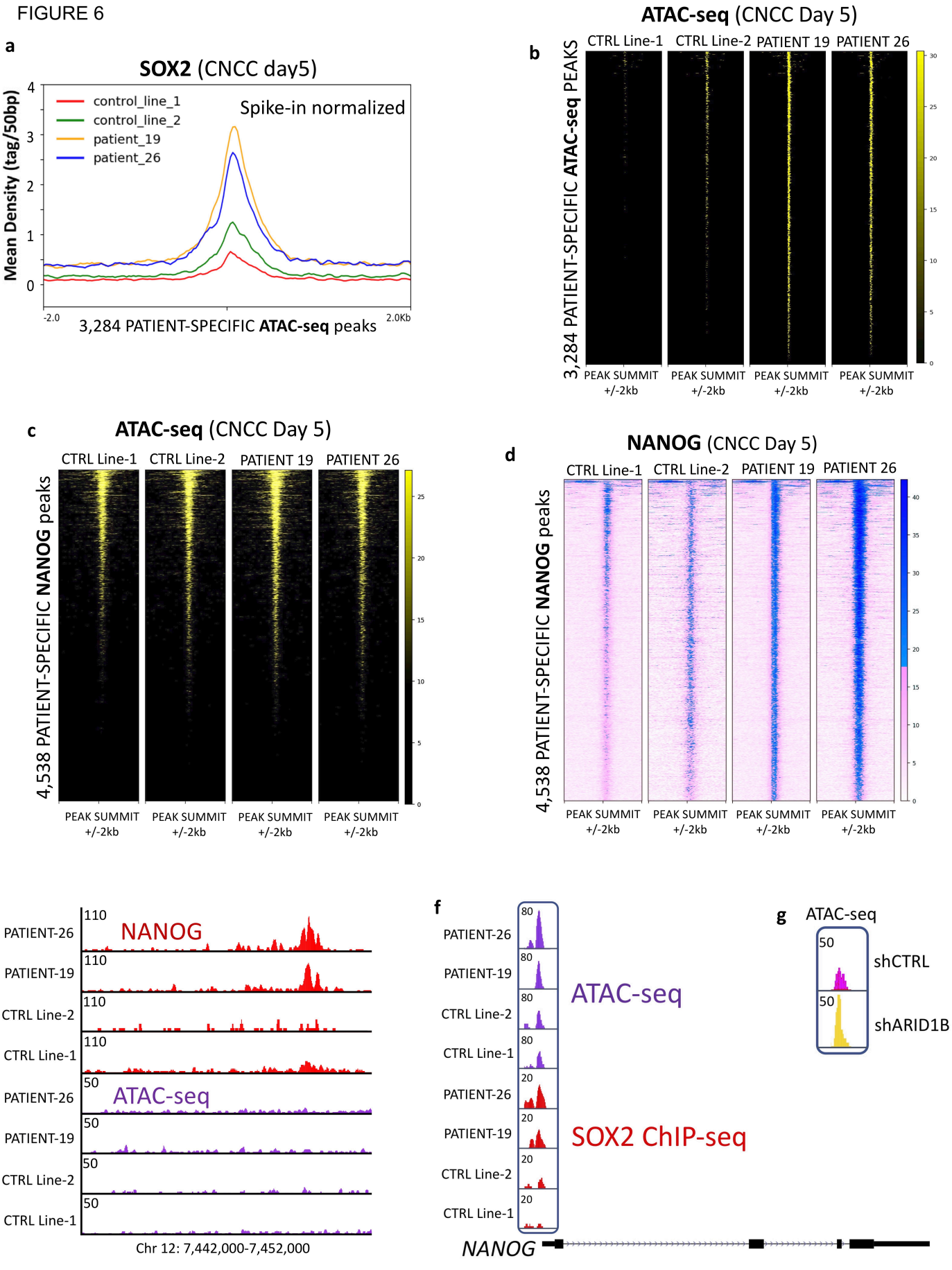


FIGURE 7