ARID1B is a Dosage-sensitive Regulator of Polycomb Repressive Complex Distribution and HOX Gene Regulation in Patient-derived Neural Progenitors

Gerald Crabtree (✉ crabtree@stanford.edu)
Stanford University   https://orcid.org/0000-0001-9685-7911

Esther Son
Stanford University

Andrey Krokhotin
Howard Hughes Medical Institute, Stanford School of Medicine

Sai Gourisankar
Department of Chemical Engineering, Stanford School of Engineering

Chiung-Ying Chang
Howard Hughes Medical Institute, Stanford School of Medicine

1 ***ARID1B* is a Dosage-sensitive Regulator of Polycomb Repressive Complex**
2 **Distribution and *HOX* Gene Regulation in Patient-derived Neural Progenitors**
3
4 Esther Y. Son[1,2,3]*, Andrey Krokhotin[1,2,]*, Sai Gourisankar[1,4], Chiung-Ying Chang[1,2], Gerald R.
5 Crabtree[1,2]
6
7 **Affiliations:**
8 [1]Howard Hughes Medical Institute, Stanford School of Medicine
9 [2]Department of Pathology, Stanford School of Medicine
10 [3]Department of Ophthalmology, University of California San Francisco
11 [4]Department of Chemical Engineering, Stanford School of Engineering
12 *Equal contribution
13
14 **Keywords**: Altism, intellectual disability, hploinsufficiency, chromatin, SWI/SNF, BAF
15 complexes
16
17

Abstract

18
19 Recent unbiased exome and whole-genome sequencing studies have identified *ARID1B* (originally
20 *BAF250b*) as the most frequently mutated gene in human de novo neurodevelopmental disorders
21 and a high confidence autism gene. ARID1B is a subunit of the multimeric SWI/SNF or
22 Brg/Brahma-Associated Factor (BAF) ATP-dependent chromatin remodeling complex. Studies of
23 *Arid1b*[+/-] mice as well as other BAF subunit mutants have found defects in neural progenitor
24 proliferation and activity-dependent neuronal dendritogenesis; however, to date, the molecular
25 impact of *ARID1B* mutations on the human neural lineage has not been investigated. Remarkably,
26 *ARID1B* is required for expression of *HOX* genes, including anterior *HOX* genes necessary for
27 brain development. Despite the high homology with ARID1A and the fact that ARID1A is
28 expressed at about 3-fold higher levels, it is unable to compensate for heterozygous loss of
29 ARID1B. These changes in gene expression were paralleled by dosage-sensitive altered deposition
30 of histone H3 lysine-27 trimethylation (H3K27me3) and histone H2A lysine-119 ubiquitination
31 (H2AK119ub) indicating that an evolutionarily conserved pathway of *HOX* gene regulation
32 underlies the neurodevelopmental defects accompanying ARID1B haploinsufficiency. Using
33 FIRE-Cas9, we show that the unmutated *ARID1B* allele can be activated to near normal and
34 potentially therapeutic levels.

35
36 **INTRODUCTION**
37

38 The human genome encodes 31 ATP-dependent chromatin regulatory enzymes homologous to
39 yeast SWI2/SNF2. The members of this class investigated to date exert their functions on
40 overlapping aspects of nucleosome dynamics, yet play highly specific biologic roles. The first to
41 be discovered was the yeast SWI/SNF complex [1,2], named after its roles in yeast mating type
42 switching and sucrose fermentation. In flies, related complexes containing the Brahma (Brm)
43 ATPase were discovered to suppress mutations in Polycomb repressive complexes (PRC) and to
44 influence development [3]. In mammals, related Brahma-associated factor (BAF), or mammalian
45 SWI/SNF (mSWI/SNF), complexes consist of 15 to 17 subunits encoded by 29 to 31 genes that
46 are assembled in combinatorial fashion. In the normal development of the mammalian nervous
47 system, BAF complexes exchange subunits to generate neuronal BAF (nBAF) complexes found
48 only in neurons [4–7]. This switch in subunit composition as cells progress from neural progenitors
49 to neurons is critical for cell cycle exit and maturation [8,9]. ARID1B, the largest core subunit of
50 BAF complexes, is implicated in several neurodevelopmental disorders. Mutations in the *ARID1B*
51 gene were found to be the most common cause of Coffin-Siris syndrome [10–13]. *ARID1B* is
52 frequently mutated in patients with nonsydnromic intellectual disability [14,15], autism spectrum
53 disorder (ASD) [16–18], and unclassified neurological developmental disorders [19]. Importantly,
54 *ARID1B* is haploinsufficient for normal human neurodevelopment [15], while loss of both alleles
55 leads to early postnatal death in mice [20]. Genetic surveys of mutations in the normal human
56 population performed on a cohort of 141,456 unrelated individuals revealed that *ARID1B* is
57 intolerant to loss of function mutations in one allele, raising the question of the nature of the dosage
58 sensitive biologic mechanism [21]. Dosage sensitivity of a gene often reflects a rate-limiting
59 biochemical step in a developmental, metabolic, or other pathway and hence can be
60 mechanistically informative.
61   At the molecular level, the BAF complex slides and evicts nucleosomes in vitro in an ATP
62 dependent manner [22], thereby creating and maintaining genomic accessibility at its target sites [23–
63 29]. BAF is also implicated in regulation of PRC [3,23,30–32] both directly, through PRC eviction [33], and

64   indirectly, by promoting genome-wide PRC redistribution possibly as a passive result of
65   widespread direct eviction [29]. PRC eviction is rapid, occurring within minutes of BAF recruitment
66   to endogenous PRC-repressed loci in somatic cells [32–34]. Oncogenic mutations in the ATPase
67   domain of Brg (SMARCA4) prevent rapid PRC eviction [32], and PRC eviction does not appear to
68   be a property of other chromatin regulatory complexes [35].

69   In flies, BAF opposes Polycomb-mediated repression at the *HOX* loci, thereby allowing
70   normal topological and temporal development of the body plan [3]. Intriguingly, some of the clinical
71   characteristics of *ARID1B* patients resemble phenotypes seen upon *HOX* gene inactivation. For
72   example, Coffin-Siris patients are characterized by hypoplasia of the distal phalanx or nail of the
73   fifth and other digits. *HOX* gene groups 11, 12, and 13 control size and number of digits in dose
74   dependent manner. In *HOX* mutant mice, the most commonly observed digit alteration involves
75   reduction in digit size and loss of phalanges [36]. Similar symptoms were observed in humans
76   carrying mutations in *HOXA13* and *HOXC13* [37]. Most Coffin-Siris patients have hypotonia, and
77   delayed development of motor skills [10]. Reduced muscle strength and deficits in motor skill
78   coordination are also observed in the *Arid1b*[+/-] mouse model [38], and *HOX* genes are important
79   players in the differentiation of motor neurons, responsible for proper muscle innervation [39] and
80   motor skill development [40].

81   Several *Arid1b* deficient mouse models consistently exhibit neurodevelopmental
82   phenotypes seen in patients with *ARID1B* mutations, including reduced cortical volume and
83   thickness, as well as deficits in learning, memory, and socialization [20,38,41–43]. Neuron subtype-
84   specific analysis revealed that Arid1b mutant mice have decreased numbers of cortical GABAergic
85   interneurons, which shifts the balance between excitatory and inhibitory synapses in the cerebral
86   cortex [41]. Remarkably, mice with selective *Arid1b* deletion either in parvalbumin or somatostatin
87   interneurons, two subtypes of inhibitory interneurons, exhibit distinct phenotypes, characterized
88   either by impairment of social interactions or by stereotypic behavior as well as learning and
89   memory dysfunction [42]. Finally, cortical and ventral neural progenitors with homozygous *Arid1b*
90   deletion displayed reduced proliferation rate, altered cell cycle regulation, and increased cell death
91   [44]. Recently, ARID1b has been shown to execute its social functions in neurons of the dorsal raphe
92   of adult mice, raising the possibility of effective therapy in adults[45].

93   Despite these recent advances in recapitulating the human phenotypes in murine models,
94   the molecular function of ARID1B remains unclear and the nature of the presumably rate-limiting
95   biochemical step performed by ARID1B is unknown. To understand the roles of BAF complexes
96   in the development of the human nervous system, we used human induced pluripotent stem cell
97   (iPSC)-derived neural progenitor cells (NPCs) containing inactivating mutations within the genetic
98   context of the *ARID1B* patient. The mutant NPCs exhibited reduced proliferation and increased
99   differentiation, had impaired WNT signaling and reduced binding of SOX family transcription
100 factors and several nuclear receptors. Importantly, the *ARID1B*-mutant human NPCs have striking
101 haploinsufficient deficits in transcription, chromatin accessibility, and PRC placement over the
102 genome. Indeed, the loss of one or two alleles appeared to have very similar effects on gene
103 expression, chromatin accessibility, *HOX* gene activation, and PRC placement over the genome.
104 Our studies suggest that BAF's ability to evict PRC from multiple sites over the genome and
105 redistribute PRC to the four *HOX* loci is a rate-limiting step in the development of the human
106 nervous system. As was previously shown in yeast [46], the expression level of haploinsufficient
107 genes is tightly regulated, potentially making these genes resistant against attempts to
108 programmatically alter their expression. Nevertheless, we demonstrate that expression of the wild
109 type endogenous *ARID1B* allele can be boosted to compensate for the lack of the expression from

110 the second endogenous allele lost due to deleterious mutations, paving the way for potential
111 therapeutical applications.
112
113 **RESULTS**
114 **Proliferation of *ARID1B* deficient NPCs is impaired.** Because the phenotypes of *ARID1B*
115 mutations are highly genetically context-dependent, we used an iPSC line derived from a patient
116 with Coffin-Siris syndrome, a neurodevelopmental disorder characterized by intellectual
117 disability, language delay, and social deficits, as well as distinct facial features and hypoplastic
118 fifth fingernails. The patient had a frameshift mutation in exon 12 of the *ARID1B* gene; using
119 CRISPR/Cas9 on the original patient iPSC line ('HET1'), we then derived isogenic iPSC lines
120 with two wild-type copies of *ARID1B* ('WT1' and 'WT2') and isogenic lines with two non-
121 functional copies of the gene (called 'KO1' and 'KO2'). An additional isogenic line heterozygous
122 for *ARID1B* ('HET2') bearing one wild-type allele and a new truncated allele was identified from
123 the pool of CRISPR/Cas9-edited clones (Supplementary Fig. 1a). These efforts gave us a full
124 isogenic series to allow normalization for genetic context. Analysis of *ARID1B* mRNA levels by
125 qRT-PCR showed an approximately 50% and <20% expression in the heterozygous and double-
126 mutant iPSC lines, respectively, relative to their wild-type counterparts (Supplementary Fig. 1b).
127 Western blot revealed that the ARID1B protein was expressed at about 50% of normal in the
128 heterozygous iPSCs (Supplementary Fig. 1c). The mutant iPSCs proliferated at about the same
129 rate as the wild type cells (data is not shown). Upon neuronal induction, the mutant iPSCs
130 generated NESTIN+ SOX1+ neural progenitors with similar efficiencies as the wild-type
131 (Supplementary Fig. 1d) but showed a mild reduction in proliferative rate (Supplementary Fig.
132 1e).
133
134 ***ARID1B* deficient NPCs exhibit strong haploinsufficient phenotypes and increased**
135 **differentiation to neurons.** To define the genes dependent upon ARID1B, we carried out RNA-
136 seq studies of the NPCs at day 40 after the start of neuronal induction in iPSCs. The heterozygous
137 and double-mutant samples were seen clustered together, away from wild-type samples (Fig. 1a).
138 Similarly, principal component analysis indicated that the major determinant of variation was the
139 genotype among the independently derived NPCs (Fig 1b). Hundreds of genes were misregulated
140 in *ARID1B* mutant cells as compared to the isogenic wild type (Fig. 1c), while only a few genes
141 were differentially expressed between *ARID1B* heterozygous and double-mutant NPCs
142 (Supplementary Fig. 1f). Gene Ontology (GO) term analysis revealed that the most significant
143 terms associated with genes increased in *ARID1B* mutant NPCs are related to neurogenesis, while
144 significant terms associated with decreased genes include cell proliferation and positive regulation
145 of cell proliferation (Fig. 1d). The previous studies showed that neurogenesis in neural progenitors
146 is accompanied by switching of neural progenitor-specific BAF (npBAF) subunits to neuronal-
147 specific BAF subunits (nBAF) [6]. Indeed, we observed a statistically significant increase in
148 expression of *ACTL6B* (log2FC=1.3, FDR=5.9 10^{-4}), an nBAF specific subunit. Other nBAF
149 specific subunits (*DPF1*, *DPF3*, *SS18L1*) showed minor but consistent increase in expression,
150 while all npBAF specific subunits (*PHF10*, *DPF2*, *SS18*, *ACTL6A*) exhibited decreased
151 expression. Despite increased neurogenesis in *ARID1B* mutant cells, neuronal progenitors
152 predominated in the cell population, such that the *ACTL6B* neuronal paralog was only expressed
153 at 5% of *ACTL6A*. Specific GO terms associated with upregulated genes include synapse formation
154 and function, axonal guidance, and dendritic development (Fig. 1d). Subunits of the BAF complex
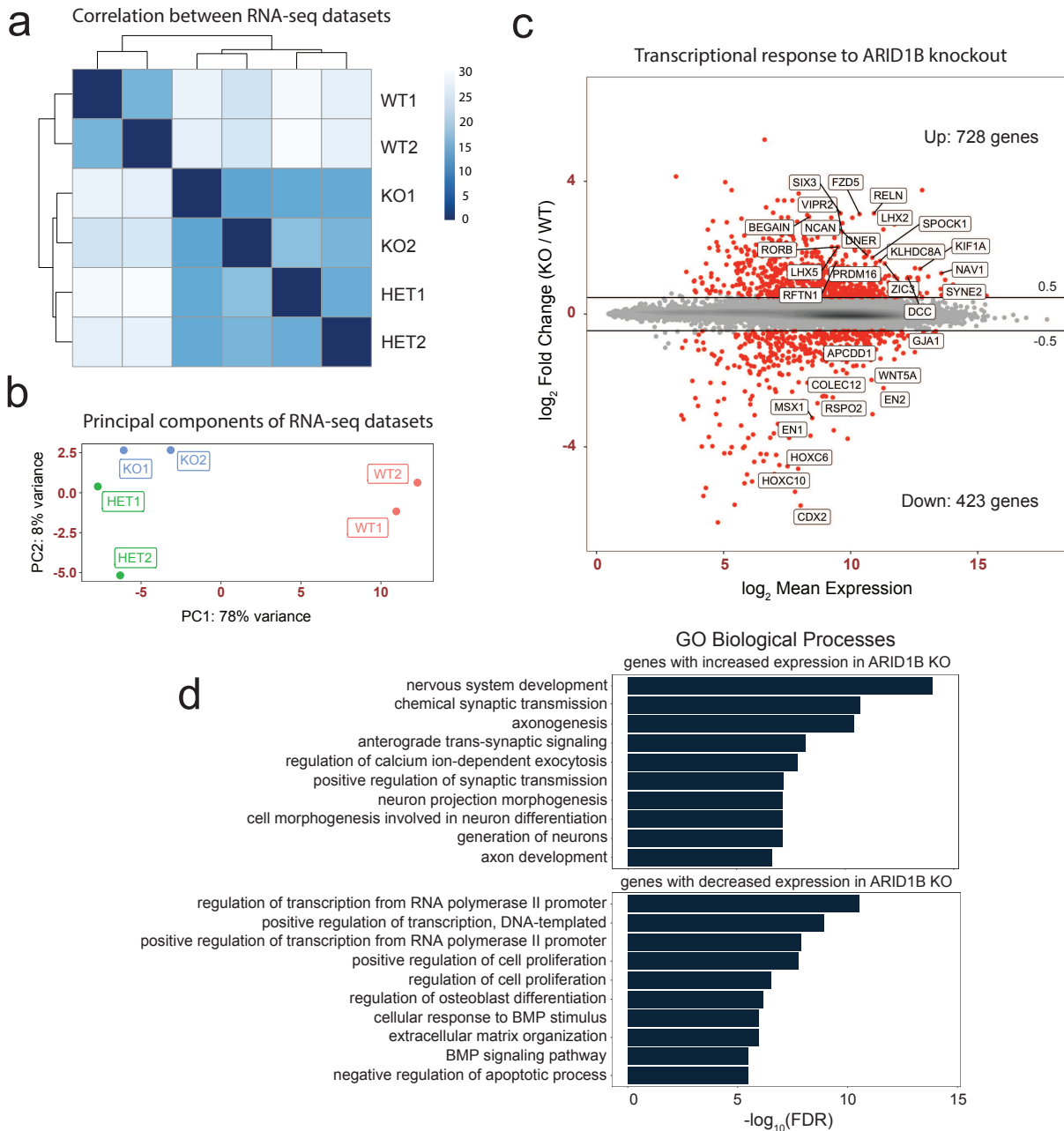
Figure 1: ARID1B dosage-sensitive effect on transcription of NPCs. (a) Correlation between RNA-seq datasets. (b) Principal component analysis of RNA-seq datasets. (c) MA plot showing gene expression changes between ARID1B WT and KO datasets. Genes with $|log2FC|>0.5$ and $FDR < 0.05$ are colored in red. The top 30 genes with the lowest FDRs are labeled. (d) Gene Ontology Biological Processes enrichments analysis for genes with altered expression in KO NPCs. All results are shown at Day 40 after start of hESCs differentiation to NPCs.

155 have been previously implicated in activity-dependent dendritic outgrowth and synaptogenesis in
156 mice [5,47,48], which speaks to the evolutionary conservation of function in humans. Together, these
157 results indicate that ARID1B maintains NPCs stemness. Impairment of ARID1B functions causes

158 early differentiation of NPCs to neurons, while wild-type NPCs exhibit neuronal differentiation at
159 later time points.
160
161 **WNT and TGF-β signaling is impaired in *ARID1B* deficient NPCs.** Previous reports implicated
162 ARID1B in regulation of the Wnt signaling pathway, which is mediated through direct interaction
163 between SMRACA4 ATPase subunit of the BAF complex and β-catenin [49]. The effect of ARID1B
164 on Wnt signaling is context-dependent. As an example, the analysis of peripheral lymphocytes
165 from individuals with intellectual disability harboring *ARID1B* mutations as well as in vitro studies
166 in several cancer cell lines demonstrated a repressive role of ARID1B on Wnt/β-catenin signaling
167 [49]. The same conclusion was reached in another study on HEK293T and ATDC5 cells, which
168 observed an increase in *AXIN2* and *LEF1* expression upon *ARID1B* knock out [50]. Contrary to these
169 findings, the analysis of gene expression in cerebral cortex of *Arid1b* deficient mice showed that
170 components of the Wnt pathway are downregulated [41]. It was also shown that ARID1B is required
171 for induction of *MYC* expression, a target of Wnt/β-catenin signaling, in MC3T3-E1 pre-osteoblast
172 cell line [51]. Finally, decreased nuclear localization of β-catenin was observed in *Arid1b* deficient
173 neurons [44]. Surprisingly, deletion of *ARID1B* in human neural progenitors had a programmatic,
174 but complex role in WNT regulation resulting in increased expression of some components of the
175 WNT pathway and decreased expression of others (Supplementary Fig. 1g). Remarkably, we
176 found that many of the genes whose expression was reduced are known downstream WNT targets,
177 including *AXIN2*, *LEF1*, *MYC*, *NKD1* and *WNT1*. This is consistent with the finding that BAF is
178 required in murine neural progenitors for effective Wnt signaling [8]. In contrast, genes with
179 increased expression are enriched for negative regulators of the Wnt pathway, including APC2, a
180 component of the β-catenin destruction complex, SFRP1, multiple cadherins, including E-cadherin
181 and multiple protocadherins [52]. Thus, ARID1B exerts a programmatic and positive role in WNT
182 signaling in neural progenitors.
183 TGF-β is another signaling pathway whose components are downregulated in *ARID1B*
184 deficient NPCs (Supplementary Fig. 1g). This is consistent with previous observations that BAF
185 complexes are required for activation of TGF-β downstream target genes [53], and that Smad2/3, the
186 mediators of TGF-β signal transduction, directly interact with BAF complex subunits [53,54].
187
188 **Loss of accessibility is observed at BAF binding sites in *ARID1B* deficient NPCs.** We used
189 ATAC-seq to investigate the role of ARID1B on producing genomic accessibility in neural
190 progenitors. Again, the major component of variability among data samples was the genotype,
191 with the heterozygous and double-mutant samples clustered together (Fig. 2a,b). Hundreds of
192 ATAC-seq peaks were significantly altered between wild-type and mutant NPCs (Fig. 2c). GO
193 term analysis revealed that peaks with decreased accessibility in the *ARID1B* mutants significantly
194 associate with genes involved in negative regulation of neurogenesis and neuron differentiation
195 (Fig. 2d). This is consistent with our RNA-seq data on increased differentiation of mutant *ARID1B*
196 NPCs to neurons. We note that peaks with altered accessibility are depleted from promoter regions
197 as compared to the total fraction of accessibility peaks found in promoters (Fig. 2e). To explore
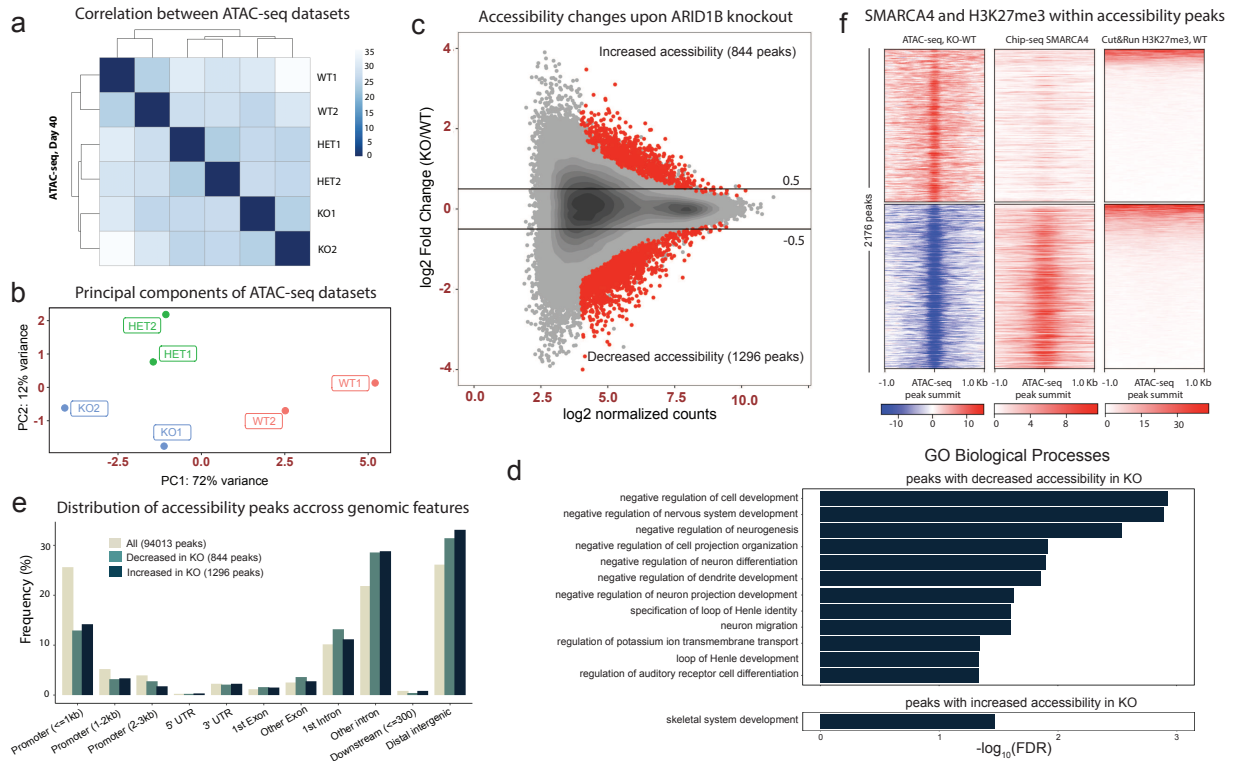
Figure 2: ARID1B promotes accessibility at BAF binding sites. (a) Correlation between ATAC-seq datasets. (b) Principal component analysis of ATAC-seq datasets. (c) MA plot showing changes of accessibility between ARID1B WT and KO datasets. Peaks with $|log2FC|>0.5$ and $FDR < 0.1$ are colored in red. (d) Gene Ontology Biological Processes associated with ATAC-seq peaks altered upon ARID1B knock out. (e) Distribution of ATAC-seq peaks over genomic features. (f) A heatmap showing association between accessibility changes ($FDR < 0.1$), SMARCA4 binding and H3K27me3 marks around summits of ATAC-seq peaks. All results are shown at Day 40 after start of hESCs differentiation to NPCs.

198   the relationship between accessibility changes and BAF occupancy we utilized SMARCA4
199   chromatin immunoprecipitation (ChIP)-seq dataset previously collected in the wild-type human
200   NPCs [55]. Strikingly, we found that peaks with decreased accessibility in *ARID1B* mutants are
201   strongly associated with BAF binding in the wild-type NPCs, which is not the case for peaks with
202   increased accessibility (Fig. 2f). This result is consistent with a role of BAF complexes in creating
203   and maintaining genomic accessibility [27,28]. Another important role of BAF is to counteract
204   repression mediated by PRC, which is achieved by direct and ATP-dependent PRC eviction [32,33].
205   However, accessibility changes induced by BAF loss only partially overlap with H3K27me3
206   marked regions (Fig. 2f), which demonstrates that most of ARID1B action is not directly related
207   to PRC redistribution.
208
209   **Transcription factor networks are heavily mis-regulated upon ARID1B loss: SOX family**
210   **transcription factors and family of nuclear receptors are among the most affected.** To explore
211   the association between accessibility and transcription factor (TF) binding, we used a modified
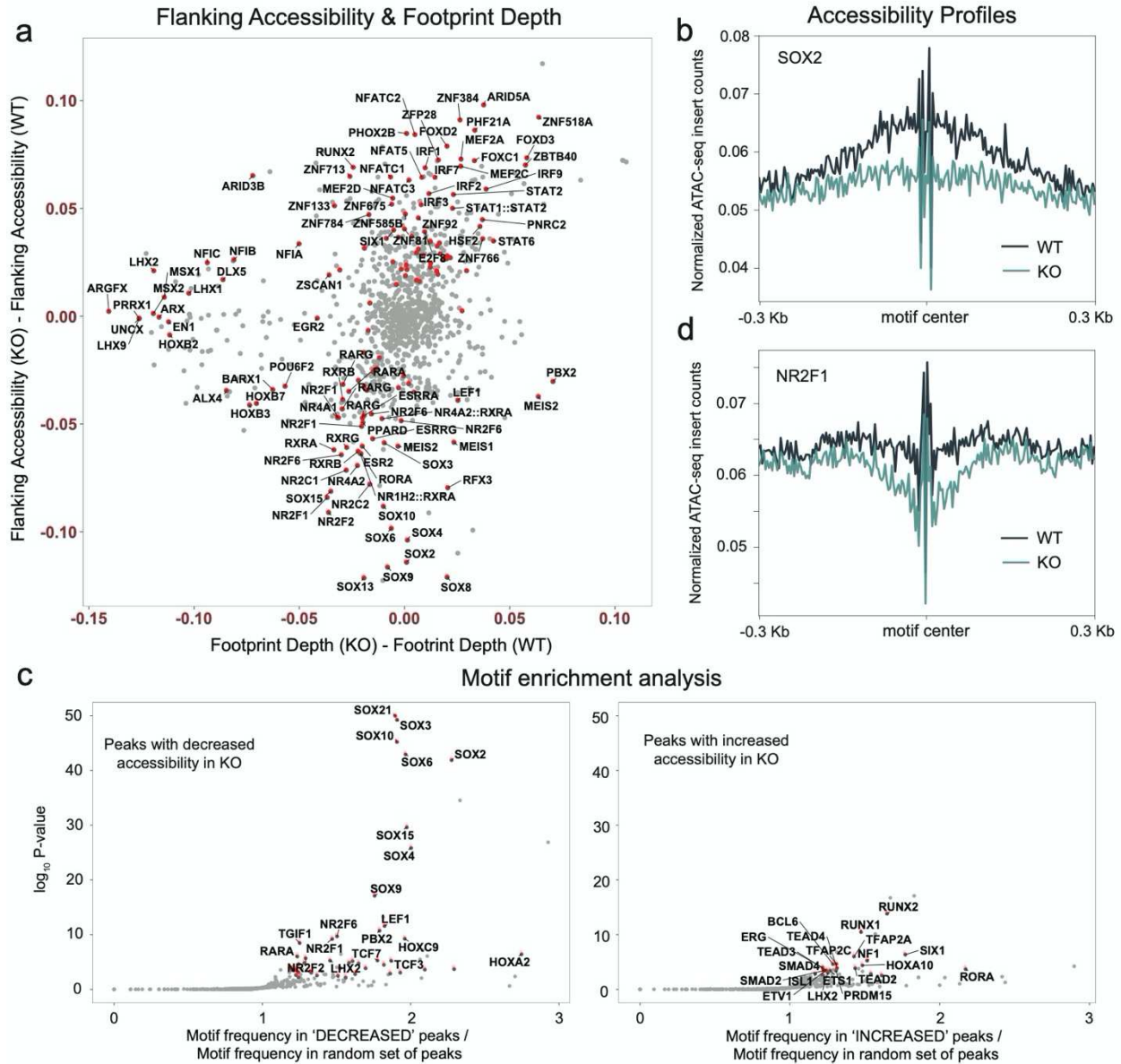
Figure 3: ATAC-seq accessibility profiling reflects extensive perturbation of transcription factor regulatory networks. (a) Changes of Flanking Accessibility and Footprint Depth between ARID1B WT and KO. Motifs with *FDR < 0.05* corresponding to genes with at least minimal expression at NPCs are colored in red. The minimal expression is defined from RNA-seq data as expression exceeding a random threshold in 50 normalized counts calculated by DESeq2 [83]. (b) Accessibility profile of SOX2 and (c) NR2F1 (COUP-TF I). (d) Motif enrichment analysis found within peaks with decreased accessibility *(left)* and increased accessibility *(right)*. All results are shown at Day 40 after start of hESCs differentiation to NPCs.

212 version of an approach quantifying TF binding based on flanking accessibility, a measure of
213 accessibility in the region adjacent to the TF binding site, and footprint depth, a measure of
214 protection from Tn5 access by the TF [56,57]. A change in flanking accessibility and footprint depth
215 reflects the effect of *ARID1B* genomic deletion on TF binding (Fig. 3a and Supplementary Fig.

216   2b). A group of SOX TFs exhibited the largest loss of flanking accessibility. In particular, SOX2
217   and SOX3, which are involved in the generation and maintenance of neural stem and progenitor
218   cells, exhibited loss of flanking accessibility, which is confirmed by comparing their accessibility
219   profiles between wild-type and *ARID1B* mutant NPCs. (Fig. 3b, Supplementary Fig. 2c). This
220   result is supported by enrichment analysis of TF binding motifs present within peaks with
221   decreased accessibility. SOX TFs are the most enriched motifs compared to all other motifs
222   enriched in peaks with either decreased or increased accessibility (Fig. 3c).
223   Multiple nuclear receptors expressed in NPCs represent another group of TFs affected by
224   ARID1B deletion. This group includes NR2F1 (COUP-TF I), NR2F2 (COUP-TF II), RARA,
225   RARG, RORA, NR2F6, NR4A2, NR1H2, NR2C1, NR2C2, RXRA, and RXRB. Depending on
226   context and ligand binding status, nuclear receptors can either be repressors or activators. To
227   elucidate potential repressor or activator status of these nuclear receptors, we looked at their
228   accessibility profiles (Fig. 3d and Supplementary Fig. 2d-g). We found that all accessibility
229   profiles have a deep valley in the vicinity of the nuclear receptor binding sites, which suggests
230   these nuclear receptors reduce accessibility and promote chromatin compaction, thus playing a
231   repressive role. Remarkably, we found that upon *ARID1B* deletion, these nuclear receptors became
232   even stronger repressors, which is reflected in reduced accessibility around their binding sites.
233   Thus, in the wild-type NPCs, ARID1B counteracts repression induced by nuclear receptors.
234   Accessibility profiles of several other TFs are affected both by heterozygous and
235   homozygous *ARID1B* loss of function mutations (Supplementary Data 1), which includes MEIS2
236   (Supplementary Fig. 2h), LEF1, SIX1, ARX, ZEB1, LHX1, and LHX2. LEF1, a known
237   downstream WNT target, has reduced flanking accessibility and increased footprint depth,
238   suggesting reduced LEF1 binding. This is consistent with decrease in *LEF1* transcription observed
239   in *ARID1B* heterozygous and double-mutant NPCs, and further supports our observation that
240   ARID1B positively regulates WNT signaling in neural progenitors.
241   While the number of peaks with significant changes in accessibility is much smaller in
242   heterozygous as compared to double-mutant *ARID1B* datasets (605 vs. 2140 peaks) (Fig. 2c and
243   Supplementary Fig. 2a), the footprint depth and flanking accessibility analyses produced similar
244   results for both conditions. This somewhat surprising result demonstrates the ability of the
245   footprint depth and flanking accessibility analyses to detect subtle, but highly consistent and likely
246   functional variations in accessibility to TFs that bind in these regions, which otherwise do not
247   exhibit bulk accessibility changes [56].
248
249   **ARID1B loss affects distribution of PRC.** Previous studies in both flies and mammals have
250   revealed that BAF opposes PRC to balance genomic accessibility during development [31,58]. The
251   opposition is produced by direct and dynamic ATP-dependent eviction of PRCs from
252   developmentally repressed loci [32,33]. Thus, to more clearly understand the mechanism underlying
253   the gene expression changes found in NPCs, we analyzed the localization pattern of H3K27me3
254   and H2AK119ub, the histone modifications produced by PRC2 and PRC1, respectively [59].
255   Correlational analysis and principal component analysis revealed that genotype was the most
256   significant variable in the samples and that independent biologic replicates were highly correlated
257   (Fig. 4a,b and Supplementary Fig. 3a,b). *ARID1B* heterozygous and double-mutant datasets
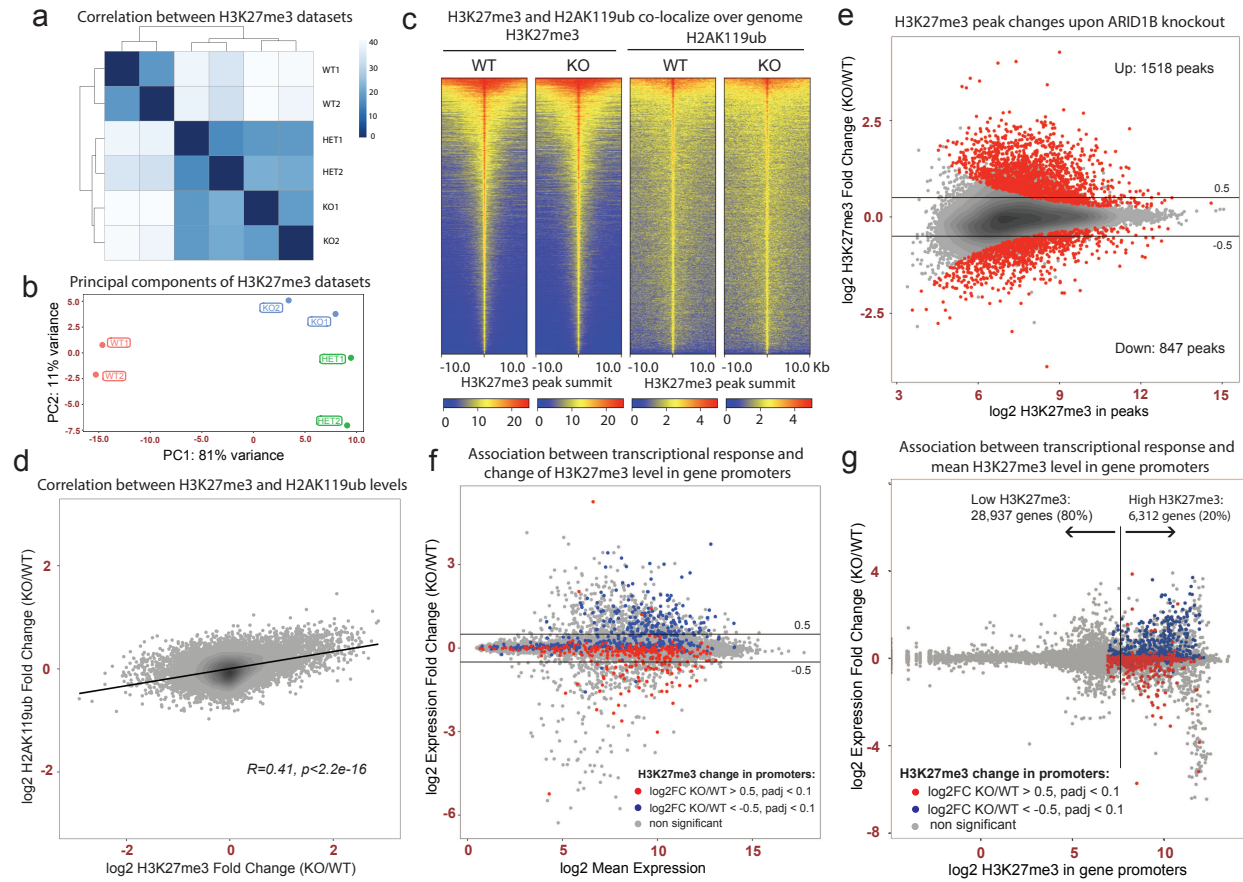
Figure 4: ARID1B dosage sensitive effect on PRC distribution. (a) Correlation between H3K27me3 Cut&Run datasets. (b) Principal component analysis of H3K27me3 Cut&Run datasets. (c) Heatmap displaying H3K27me3 and H2AK119ub distribution around summits of H3K27me3 peaks. (d) Scatter plot displaying log2 fold changes of H3K27me3 versus log2 fold changes of H2AK119ub between KO and WT ARID1B conditions. H3K27me3 and H2AK119ub levels are calculated in ±3kb window around H3K27me3 summits. (e) MA plot showing H3K27me3 peak changes between ARID1B WT and KO datasets. Peaks with $|log2FC|>0.5$ and $FDR < 0.05$ are colored in red. (f) MA plot showing gene expression changes between ARID1B WT and KO datasets. Genes with significant changes of H3K27me3 level within ±3kb window around their TSS ($|log2FC|>0.5$ and $FDR < 0.1$) are colored in red and blue for increased and decreased H3K27me3 levels respectively. (g) Log2 fold changes of gene expression versus H3K27me3 level within ±3kb window around their TSS. Genes with significant changes of H3K27me3 are colored red and blue for increased and decreased H3K27me3 levels respectively. All results are shown for Day 40 after start of hESCs differentiation to NPCs.

258   clustered together away from wild-type, as was seen for gene expression and accessibility. We
259   found that H3K27me3 and H2AK119ub marks co-localize over the genome as expected from the
260   coordinated action of PRC1 and PRC2 (Fig. 4c) and exhibit correlated changes across wild-type
261   and mutant conditions (Fig. 4d). Most of the peaks with significant changes were associated with
262   increased H3K27me3 level in *ARID1B* double-mutant NPCs (Fig. 4e) consistent with direct ATP-
263   dependent PRC eviction [32,33]. GO Biological Process enrichment analysis revealed that decreased
264   H3K27me3 peaks associate with neural differentiation, confirming our data on enhanced

265 differentiation of *ARID1B* mutant cells (Supplementary Fig. 3c). Specific examples of
266 neurodevelopmental genes that have increased H3K27me3 level in *ARID1B* mutant NPCs include
267 *RBFOX1*, *FOXP2*, *ANK2*, *NR2F1*, and an intron peak within *ARID1B* itself, suggesting that
268 ARID1A might act within a positive feedback loop. (Supplementary Fig. 3d-h). Among those,
269 peaks at *RBFOX1*, *FOXP2* and *NR2F1* also appear in human embryonic stem cells (ESCs), which
270 indicates failure of *ARID1B* mutant NPCs to remove the PRC mark, while peaks at *ANK2* and
271 *ARID1B* appeared *de novo*. Specific examples of neurodevelopmental genes with reduced levels
272 of H3K27me3 in *ARID1B* mutant NPCs include *FOXG1*, *LHX2*, *FZD5*, *OTX1* and *EMX2*
273 (Supplementary Fig. 3i-m).
274
275 **Promoters of differentially expressed genes are associated with high H3K27me3 levels, which**
276 **is not affected by ARID1B loss.** Next, we explored the relationship between changes in
277 H3K27me3 levels within promoter regions (±3 kb from transcription start sites (TSS)) and gene
278 expression. While we found that genes with decreased (increased) expression are more likely to
279 be associated with increased (decreased) level of H3K27me3 mark (Fig. 4f), the amplitude of
280 H3K27me3 variation was not predictive of gene expression changes. We observed that significant
281 changes in H3K27me3 level in promoters have only minor effects on gene expression levels in
282 most cases. In addition, we found that the majority of the most differentially expressed genes are
283 located in the regions heavily decorated with H3K27me3, which did not exhibit significant
284 changes in *ARID1B* mutant NPCs (Fig. 4g). Indeed, while only 20% of all genes have high
285 H3K27me3 levels within their promoters, they represent 64% of all differentially expressed genes.
286 Of note, the genes with the highest expression changes are associated with the highest level of
287 promoter H3K27me3 that does not change in *ARID1B* mutants. The list of examples of such genes
288 include *EN1*, *EN2*, *CDX2*, *WNT1*, *LMX1A*, *LMX1B*, *RSPO2*, *GDF7*, *DMRT3* (Supplementary Fig.
289 4).
290
291 ***HOX* genes are the most affected by *ARID1B* loss.** In ESCs, *HOX* genes are bivalent and contain
292 histone marks associated with active and repressed chromatin [60]. The bivalent genes are trapped in
293 a state with very low expression until the underlying chromatin is resolved to either fully active or
294 repressed states upon ESC differentiation. Consistent with this paradigm, we observe activation of
295 *HOX* genes upon differentiation of the wild-type ESCs to NPCs. However, *HOX* genes remained
296 silent in the NPCs derived from *ARID1B* heterozygous or double-mutant ESCs (Fig 5a and
297 Supplementary Fig. 5). Activation of the *HOX* genes in the mouse NPCs is accompanied by
298 saltatory and complete clearance of the H3K27me3 histone mark from the underlying genomic
299 loci [61]. After 40 days of differentiation, we observed only marginal and statistically insignificant
300 decreases in H3K27me3 marks in the wild-type NPCs as compared to *ARID1B* heterozygous or
301 double-mutant NPCs (Fig. 5b and Supplementary Fig. 5) despite robust differences in Hox gene
302 expression (Fig 5a). To investigate the dynamics of *HOX* genes expression and H3K27me3 marks
303 within the *HOX* loci, we extended our NPC culture for another 20 days and generated RNA-seq
304 and H3K27me3 datasets at day 60 after start of neuronal induction of iPSCs. Analogous to day 40,
305 the expression of *HOX* genes was reduced in *ARID1B* double-mutant NPCs as compared to the
306 wild-type NPCs at day 60 (Fig. 5c). We also observed clearance of H3K27me3 from the underlying
307 genomic loci in the wild-type NPCs but not in the *ARID1B* double-mutant NPCs (Fig. 5d and
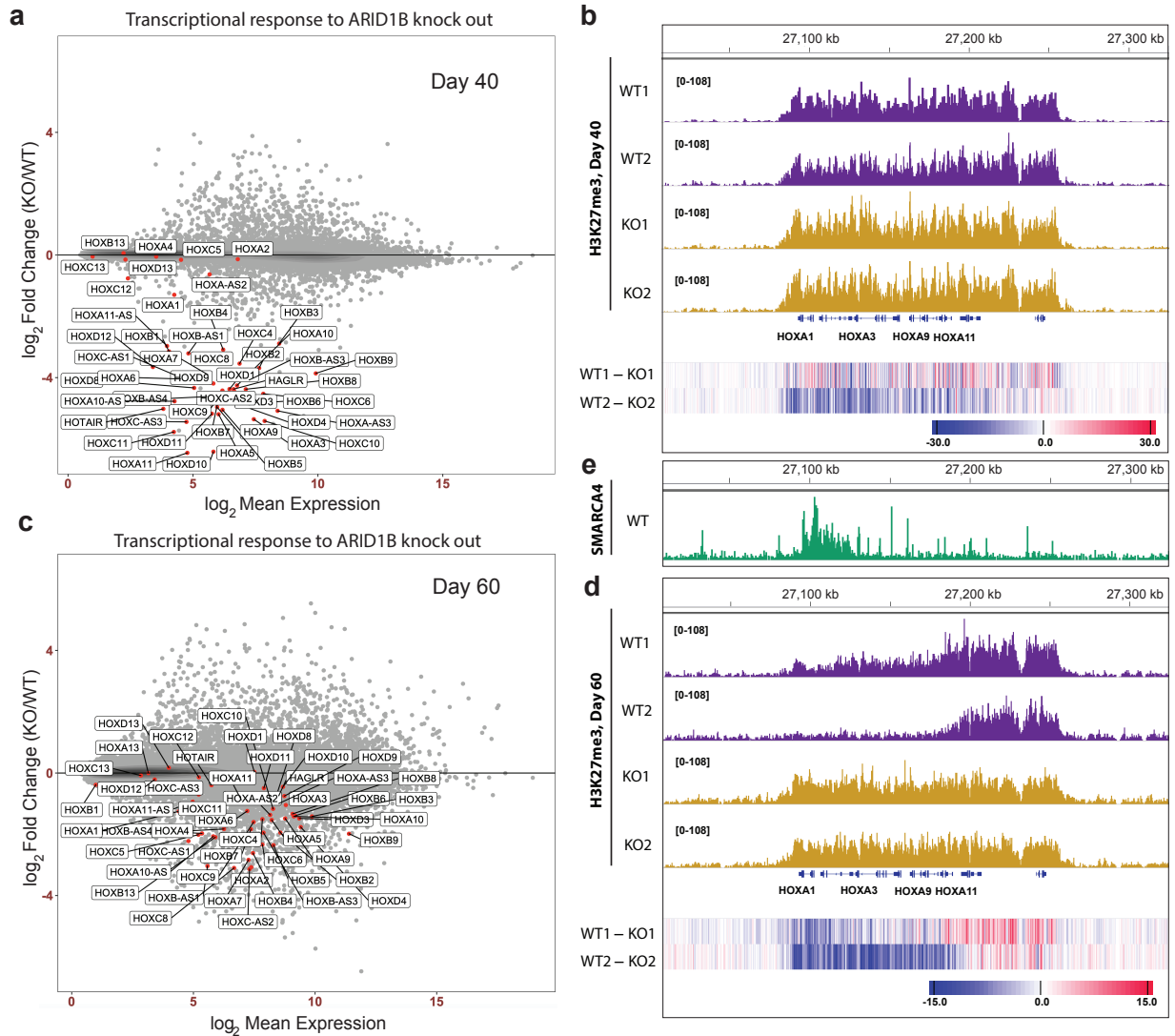
Figure 5: Expression of HOX genes is suppressed in ARID1B mutant NPCs. (a) MA plot showing gene expression changes between ARID1B WT and KO datasets at Day 40 and (c) at Day 60 after start of hESCs differentiation to NPCs. HOX genes and other coding and non-coding genes located within HOX clusters are labeled. (b) H3K27me3 coverage of HOXA locus shown separately for two KO and WT biological replicas at Day 40 and (d) at Day 60. The result of WT and KO subtraction is shown at the bottom of the coverage tracks. (e) SMARCA4 binding within HOXA locus.

308 Supplementary Fig. 6). Surprisingly, these observations suggest that the initial activation of *HOX*
309 genes occurs prior to H3K27me3 clearance.
310   We observed substantial SMARCA4 presence at all four *HOX* loci in the wild-type NPCs,
311 with a broad coverage at *HOXA* and *HOXB* domains (Fig. 5e and Supplementary Fig. 5). Together,
312 our findings suggest that ARID1B is necessary for *HOX* gene activation and for subsequent
313 heterochromatin resolution to euchromatin.
314

315 ***ARID1B* mutation affects genes associated with autism spectrum disorders (ASD).** *ARID1B*
316 and other subunits of BAF complexes are one of the most frequently mutated groups of genes in
317 ASD [47]. To elucidate functional associations between ARID1B and ASD related genes, we
318 examined our data against the SFARI database, a diverse collection of genes whose *de novo*
319 mutations are linked to ASD diagnosis. Since most of these mutations are presumably loss-of-
320 function mutations, we looked for overlap between the SFARI autism genes and genes with
321 decreased expression in *ARID1B* mutant NPCs. We identified 15 high-confidence ASD genes that
322 were downregulated (log2FC < -0.5 and FDR < 0.05) both in heterozygous and double-mutant
323 *ARID1B* NPCs: *CASZ1*, *NR3C2*, *PAX5* (SFARI score 1), *CGNL1*, *ICA1*, *LMX1B* (SFARI score 2),
324 and *CDH11*, *EN2*, *ERG*, *MSX2*, *MUC12*, *NXPH1*, *SATB2*, *TNS2*, *WNT1* (SFARI score 3). We also
325 found that some TFs whose expression is not affected by *ARID1B* mutation nevertheless have
326 altered flanking accessibility and footprint depth, which suggests defective binding. This group of
327 TFs include *RFX3*, *SOX5*, *NR4A2*, *MEIS2* (SFARI score 1), *NR2F1*, *NR2F2* (SFARI score 3). In

328 addition, some high-confidence ASD genes
329 with very low expression in NPCs (e.g.,
330 *RBFOX1* (SFARI score 2)) have increased
331 levels of H3K27me3, potentially preventing
332 their activation later in the course of
333 development.

335 **Enhancing the expression of the**
336 **endogenous wild-type *ARID1B* allele.**
337 Most neurodevelopmental disorders
338 associated with *ARID1B* mutations are due to
339 loss-of-function mutations in one allele. Our
340 results above indicate that haploinsufficency
341 is paralleled by dosage-sensitive changes in
342 gene expression, accessibility, and PRC
343 distribution in human NPCs.
344 Haploinsufficiency is also seen for a large
345 number of autism genes, indicating that one
346 might treat these diseases by inducing the
347 expression of the remaining functional wild-
348 type allele. For example, some of the autism-
349 like phenotypes of *Arid1b* mutant mice can be
350 attributed to deficits in the dorsal raphe of
351 adult mice and can be rescued with serotonin
352 analogues [45]. However, genes are subject to
353 complex positive and negative feedback
354 pathways that might fix the level of

Figure 6: Correction of ARID1B expression to the WT level in HET ARID1B cell line. (a) Schematic representation of rapamycin inducible FIRE-Cas9 system used to recruit VPR to the ARID1B promoter. (b) Time-course of ARID1B expression after addition of rapamycin; n=4 independent repeats Two primer sets located at exon 12-13 and exon 18-19 were used for quantifying ARID1B expression with qPCR.

355 expression at a stable level. Indeed, *ARID1B* has an intronic regulatory PRC target site sensitive
356 to the level of the ARID1B protein (Supplementary Fig 3e). In addition, studies in yeast have
357 shown that haploinsufficient genes are often resistant to or intolerant to overexpression [46]. Thus,
358 we were interested to see if the level of expression from the endogenous gene could be increased
359 to potentially therapeutic levels, i.e. 2-fold. We employed a rapamycin-inducible FIRE-Cas9
360 system [34] targeted to the *ARID1B* promoter region in iPSCs by custom-designed sgRNAs (Fig.
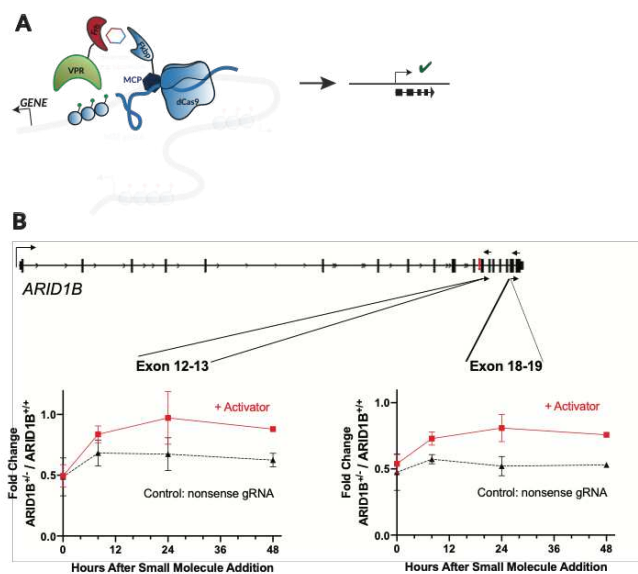
361 6A). After 24 hours of rapamycin treatment (3 nM), we observed an increase in *ARID1B* mRNA
362 expression to the level comparable to the wild-type (Fig. 6B) as assessed by qPCR using two
363 independent primer sets showing that the full transcript is produced at about 1.8-fold over
364 background (Fig. 6B). Thus, even though *ARID1B* appears to be subject to autoregulation, its
365 mRNA expression can be increased to levels that could be therapeutic using FIRE-Cas9. While
366 the immunogenicity of Cas9 poses a challenge in using the current system as a therapeutic in
367 patients, our proof-of-concept demonstrates the feasibility of activating the endogenous wild-type
368 allele which may be achieved by alternative means such as small molecules.
369
370 **DISCUSSION**
371
372   Our studies show that the haploinsufficient roles that ARID1B plays in human
373 neurodevelopment are paralleled by dosage-sensitive functions in gene expression, genomic
374 accessibility, and distribution of PRC over the genome of human NPCs. Remarkably, ARID1B is
375 required for expression of all four *HOX* clusters. The induction of neural differentiation of human
376 wild type iPSCs is accompanied by increased expression of *HOX* genes and by reduction of the
377 H3K27me3 histone mark from the expressed *HOX* loci. However, *ARID1B*-deficient iPSCs
378 maintain uniform H3K27me3 coverage, characteristic of embryonic stem cells. We find that BAF
379 complexes form broad domains within the *HOX* loci of human NPCs, indicating that clearance of
380 H3K27me3 is likely a result of direct eviction of PRCs by BAF as demonstrated previously [32,33].
381 These results parallel classic studies in *Drosophila* which showed that the fly BAP or dSWI/SNF
382 complex maintains expression of homeotic genes during development. Remarkably, the drop in
383 HOX gene expression is not only limited to neural progenitors. The analysis of the previously
384 published RNA-seq dataset from the whole brain extract of Arid1b heterozygous male mice
385 revealed that Hoxb2 is the most downregulated gene (log2FC=-8.78) in the whole dataset with
386 Hoxd3 exhibiting significant downregulation (log2FC=-3.29) [38]. Thus, our studies call attention to
387 an evolutionarily conserved pattern played out in the development of the nervous system that
388 underlies the high frequency of human mutations in *ARID1B* and other BAF subunits in human
389 neurodevelopment.
390   Our findings indicate that the mechanism underlying the clearance of Polycomb marks by
391 ARID1B during differentiation of human iPSCs to NPCs appears to result from a complex and
392 temporally specific interplay with critical TFs. Previous work has shown that activation of anterior
393 *Hox* genes (*Hox1-Hox5*) in NPCs is mediated by retinoid acid receptors, which have multiple
394 binding sites within *Hox* loci [61]. Remarkably, we showed that in *ARID1B* mutant NPCs, chromatin
395 becomes more inaccessible by ATAC-seq in the vicinity of the binding sites of two retinoic acid
396 receptors RARA and RARG, which suggests that interaction between ARID1B-containing BAF
397 complex and retinoic acid receptors is required for the RA-dependent activation of *HOX* genes.
398 This conclusion is also supported by the evidence of the direct interaction between RARA and
399 ARID1 [62]. The activation of posterior *Hox* genes (*Hox6-Hox9*) in the NPCs requires the Cdx2
400 transcription factor binding within *Hox* loci [61]. Remarkably, *CDX2* is the most significantly
401 downregulated gene in the *ARID1B*-deficient NPCs in our studies, which could be a consequence
402 of impaired WNT signaling, and downregulation of *WNT3A* as well as several members of the
403 WNT signaling pathway that are required for *CDX2* activation. This observation indicates that
404 CDX2 may be interacting with BAF, which has the ability to rapidly clear PRC and its marks
405 within minutes [32,33]. Consistent with this mechanism, CDX2 was previously shown to directly
406 interact with the BAF complex [63,64]. Thus, our studies and those previously reported indicate that

407 ARID1B-containing BAF complexes act early in iPSC-to-NPC differentiation by first cooperating
408 with retinoid acid receptors localized to the *HOX* loci. A second mechanism used by ARID1B is
409 to contribute substantially to the activation of *CDX2*, whose gene product appears to guide BAF
410 to the *HOX* loci, which then rapidly and directly evict PRC in an ATP-dependent fashion [32,33].

411 The BAF complex counteracts repression mediated by PRC1 and PRC2. Indeed, we
412 observe robust changes in H3K27me3 distribution in *ARID1B* deficient NPCs. We also found that
413 increase (decrease) in gene expression was associated with decreased (increased) H3K27me3 level
414 in gene promoters. However, the amplitude of H3K27me3 changes within gene promoters was not
415 predictive of the amplitude of transcriptional changes. Remarkably, we found that the genes
416 exhibiting the largest changes in transcription have high H3K27me3 level that does not change in
417 *ARID1B* mutants, which indicates that regulation of transcription by PRC is a more complicated
418 process than simple changes in H3K27me3 levels [65]. These findings indicate that in mammals,
419 BAF complexes use additional mechanisms to control critical developmental genes, such as
420 nucleosome remodeling, regulation of histone acetylation by interaction with CBP [66–68], and
421 regulation of chromosomal topology with Topoisomerase 2 [25].

422 We also find that ARID1B has programmatic roles in regulating both the expression and
423 function of WNT and TGF-β signaling pathways. During the course of iPSC-to-NPC
424 differentiation, ARID1B directly regulates many genes required for these signaling pathways
425 leading to well characterized phenotypes that might be related to the neurodevelopmental
426 abnormalities found in individuals having mutations in one allele of *ARID1B*. In addition to the
427 regulation of *HOX* genes, we also found programmatic roles in controlling the accessibility of
428 binding sites of SOX family of transcription factors and multiple nuclear receptors important in
429 human neurodevelopment. *ARID1B* is a high confidence autism gene and we observe that several
430 genes previously associated with ASD have decreased expression in *ARID1B* heterozygous and
431 double-mutant NPCs. In addition, several TFs associated with ASD have altered accessibility
432 profiles in *ARID1B* mutant NPCs indicating their deficient binding and function.

433 BAF complexes are combinatorially assembled with 15 to17 subunits encoded by 29 to 31
434 genes [69–73]. Based on immunofluorescence studies and single cell RNA sequencing it appears that
435 each cell contains perhaps 100 or more distinct complexes [74], raising the issue of how functionally
436 diverse or redundant these complexes might be. Remarkably, while the *ARID1B* expression level
437 is only about 30% of *ARID1A* expression in our samples, ARID1A cannot compensate for
438 ARID1B deletion, which indicates unique roles for ARID1B-containing complexes. ARID1A-
439 containing complexes have a dramatically different function in the nervous system [75]. The unique
440 and non-redundant functionality of different combinatorially assembled BAF complexes is further
441 supported by the fact that both *ARID1A* and *ARID1B* are haploinsufficient and dosage-sensitive in
442 the nervous system. The molecular mechanism by which these homologous proteins play such
443 different functions is not yet clear, but localization, differential TF interactions and perhaps
444 different roles in nucleosome dynamics or PRC opposition are all possibilities.

445 Based on recent sequencing studies, it appears that *de novo* loss of function mutations in
446 haploinsufficient genes are common causes of human disease. These diseases might be treated by
447 boosting transcription of the functional endogenous allele to overcome the reduction of the normal
448 gene product due to the defective allele. Concomitant increase in the expression of the defective
449 allele should not be an obstacle for this strategy as long as the defective allele is not a dominant
450 negative. However, Amon and colleagues have found that in yeast, haploinsufficiency is often
451 accompanied by intolerance or resistance to overexpression [46], which would foil this simple
452 therapeutic strategy. In addition, many genes including *ARID1B* appear to operate within powerful

453   feedback pathways that could make a cell resistant to altering the level of expression of an
454   endogenous gene. To test this possibility for the remarkably dosage-sensitive ARID1B gene, we
455   used FIRE-Cas9 to conditionally activate the expression of *ARID1B*. Remarkably, in heterozygous
456   *ARID1B* iPSCs, 3 nM rapamycin activated transcription along the entire 400 kb gene to a stable
457   level of expression about 1.8-fold above that found in the mutant cells. While the immunogenicity
458   of Cas9 makes the system unsuitable for clinical studies, these preliminary experiments suggest
459   that no intrinsic barrier exists to limit expression of *ARID1B* mRNA and that perhaps a small
460   molecule could specifically induce expression of this gene, much in the same way that topotecan
461   unsilences the paternally encoded Ube3a in Angelman syndrome [76]. Furthermore, we have recently
462   found that ARID1B executes its social functions in neurons of the dorsal raphe of the adult mouse[45]
463   indicating that this strategy could evolve into an effective treatment.
464
465   **METHODS**
466
467   **Generation of *ARID1B* isogenic iPSC lines**
468   Human iPSCs derived from an *ARID1B*-heterozygous Coffin-Siris patient (*ARID1B* c.2598del)
469   were a generous gift from the Santen laboratory (Leiden University). The iPSCs were cultured in
470   mTeSR1 medium (Stemcell Tech.) on plates coated with Geltrex (ThermoFisher) and passaged
471   with Accutase (Stemcell Tech.) every 4-5 days. Genome editing was performed using the protocol
472   described in Ran FA, HsuPD, Wright J, Agarwala V, Scott DA, Zhang F. Nat Protoc 2013;
473   8(11):2281-308 (PMID 24157548). A brief protocol is as follows. **Molecular cloning.** Single-
474   guide RNAs (sgRNAs) sequences targeting exon 9 of the human *ARID1B* gene were designed
475   using the tool developed in the Zhang lab (crispr.mit.edu) and cloned into the pSpCas9-2A-puro
476   (Addgene, catalog no. 48139). The following sgRNA was chosen for the experiments following
477   optimization: 5'-CTGGGCACCCCACTATACGC. HDR (homology directed repair) templates
478   were generated by inserting into pUC19 either the wild-type (5'-
479   agctctcctgctgccgtgtgccgtgcttgtgctgggacagttttccgagatgtggcagccagcacaaaccagctgctcgatctggtactcaag
480   caaaaaaagaaagaaagaaagaaaaatattattcccaacttagaggaaaactggggctcatgtcaggttcccgcgtgacaccttcctgtgg
481   ggttttaaaggatgattttgaccatatgtgcatgtcgtcttacacgttgtctttttgagcccctttcacatgactgttgaaggtagcagtgatctagtg
482   gtgtgatctcaagagtggtgctcaccattttggttaaaatataagaatatgtggtgttgaatggattgaggcagcactcggtgtcccttgcctatt
483   catagggtttgtttttttgttttttttgtttgtttcttttttatgaccagcctgaatttctcttcatcagtgcagtggtcctctggggctgtctgtatttgagtca
484   ctctctgagggcctgggagtgtgcagagagagcatgtttgagactggcagctgcaccagcagcaacaggaagggcctataacggtcatg
485   actaatactccgtgctgatcgcattgttggacaaaaagtatttccagtgaatgttgtcacaagtttaaataaaaggctttactttgtgtttgctttttgt
486   ttaggtaactactccag**g**cc**t**ccagcgtatagtggggtgcccagtgcaag**C**tacagcggcccagggcccggtatgggtatcagtgc*caa*
487   *caaccagatgcatggacaag*ggccaagccagccatgtggtgctgtgcccctgggacgaatgccatcagctgggatgcagaacagacc
488   atttcctggaaatatgagcagcatgacccccagttctcctggcatgtctcagcagggagggccaggaatggggccgccaatgccaactgtg
489   aaccgtaaggcacaggaggcagccgcagcagtgatgcaggctgctgcgaactcagcacaaagcaggtacgccacccaggagcacgcc
490   ccgggcaggtacgctgtgtgtctacccgtgaccacgtgactgcgcacatagctgcattgttccctggggtcacacagagcagtagaatgtc
491   actgtgcttggccgttctttttgtgttaaatatattctgtagcatcgaggtcagagaactaagtgacaaaccatgtttatgatacatgaggaagg
492   aatgtgagcggtgagcacatcagtcgtgggaggtgatttgaaatgcctgtgaatagtcattgcagcagcaaaccagatctggttgcacagg
493   gaagtgtctgcgtggtaccaattgaagcagtgacttttaagcaggaaatacgcttaggaacgatttaaaatctaaaacattcaaaaggaacac
494   actccccacctttttttcaccctaccacaagaatcttcctaaaatacaaacttcgtatgtatacatacgtgtattgtgtgtttgtctaaatgtatatgtg
495   tgagatttatatacttgacattcttatgttatattttatttttccaggcatatctatctgtccaaacca)   or   mutant   (5'-
496   agctctcctgctgccgtgtgccgtgcttgtgctgggacagttttccgagatgtggcagccagcacaaaccagctgctcgatctggtactcaag
497   caaaaaaagaaagaaagaaagaaaaatattattcccaacttagaggaaaactggggctcatgtcaggttcccgcgtgacaccttcctgtgg
498   ggttttaaaggatgattttgaccatatgtgcatgtcgtcttacacgttgtctttttgagcccctttcacatgactgttgaaggtagcagtgatctagtg

499   gtgtgatctcaagagtggtgctcaccattttggttaaaatataagaatatgtggtgttgaatggattgaggcagcactcggtgtcccttgcctatt

500   cataggggtttgtttttttgtttttttgtttgtttctttttatgaccagcctgaatttctcttcatcagtgcagtggtcctctggggctgtctgtatttgagtca

501   ctctctgagggcctgggagtgtgcagagagagcatgtttgagactggcagctgcaccagcagcaacaggaagggcctataacggtcatg

502   actaatactccgtgctgatcgcattgttggacaaaaagtatttccagtgaatgttgtcacaagtttaaataaaaggctttactttgtgtttgctttttgt

503   ttaggtaactactccag**g**cc**t**ccagcgtatagtgggggtgcccagtgcaagtacagcggcccagggcccggtatgggtatcagtgc*caaca*

504   *accagatgcatggacaag*ggccaagccagccatgtggtgctgtgcccctgggacgaatgccatcagctgggatgcagaacagaccattt

505   cctggaaatatgagcagcatgacccccagttctcctggcatgtctcagcagggagggccaggaatggggccgccaatgccaactgtgaac

506   cgtaaggcacaggaggcagccgcagcagtgatgcaggctgctgcgaactcagcacaaagcaggtacgccacccaggagcacgccc

507   gggcaggtacgctgtgtgtctacccgtgaccacgtgactgcgcacatagctgcattgttccctggggtcacacagagcagtagaatgtcact

508   gtgcttggccgttcttttgtgttaaatatattctgtagcatcgaggtcagagaactaagtgacaaaccatgttttatgatacatgaggaaggaat

509   gtgagcggtgagcacatcagtcgtgggaggtgatttgaaatgcctgtgaatagtcattgcagcagcaaaccagatctggttgcacagggaa

510   gtgtctgcgtggtaccaattgaagcagtgacttttaagcaggaaatacgcttaggaacgatttaaaatctaaaacattcaaaaggaacacactc

511   cccacctttttttcaccctaccacaagaatcttcctaaaatacaaacttcgtatgtatacatacgtgtattgtgtgtttgtctaaatgtatatgtgtga

512   gatttatatacttgacattcttatgttatatttttatttttccaggcatatctatctgtccaaacca) *ARID1B* gene sequence via the

513   BamHI site; note that a new silent restriction site was incorporated into the HDR templates (5'-

514   ag**gcc**t from the native sequence 5'-agaccc). **Transfection.** Human pluripotent stem cells were

515   grown in mTeSR1 medium (Stem Cell Technologies) on plates coated with Geltrex (Invitrogen).

516   At ~70% confluency, cells were dissociated to near single cells using Accutase (Life

517   Technologies). 10^6 cells were transfected with 5 ug total DNA (1:1 w/w mixture of pSpCas9-

518   sgRNA with or without HDR repair template) using Amexa Human Stem Cell Nucleofector™ Kit

519   1 (Lonza) according to the manufacturer's protocol. **Drug selection, genotyping, and qRT-PCR.**

520   At 24h post-transfection, puromycin (0.5-0.75 ug/ml) was added to culture medium for 48h.

521   Individual colonies were manually picked after 7-10 days and expanded. Genomic DNA for each

522   clone was prepared by incubating cells overnight at 55'C in 0.1 mg/ml proteinase K in digestion

523   buffer (100 mM NaCl, 10 mM TrisCl pH 8, 25 mM EDTA pH 8, 0.5% SDS) followed by heat

524   inactivation at 85'C for 10 minutes. The target region in *ARID1B* was amplified by PCR (Fw

525   primer: 5'- ctggggctgtctgtatttgagtcactctc; Rv primer: 5'-aaatgagtaagtaattgtagctaaagccttt) and

526   sequenced using the following primer: 5'- attcctggccctccctgctgagacatg.

527

528   **Quantitative real-time polymerase chain reaction (qRT-PCR)**

529   The *ARID1B* isogenic iPSC lines with the desired genotypes (wild-type, heterozygous, or double

530   mutant) were verified by analyzing *ARID1B* mRNA transcription by qRT-PCR. Briefly, total RNA

531   was isolated using TRIzol™ Reagent (Thermo Fisher) and cDNAs were synthesized using

532   SuperScriptTM Double-Stranded cDNA Synthesis Kit (Thermo Fisher) following the

533   manufacturer's instructions. qRT-PCR was performed in QuantStuio 6 Flex Real-Time PCR

534   System (Thermo Fisher) with SensiFAST™ SYBR Lo-ROX Kit (Bioline) using primer sets

535   targeting   exons   12-13   *(Fw: 5'-*CTACGTCTGCGTCAAAGAGATCG;   Rv: 5'-

536   GTTGCCAGCTCACGCCACTTCTTG)   or   exons   18-19   (Fw: 5'-

537   GCATCACAACCAGTCTTGAAACAA; Rv: 5'- ATCACACGCCACGCCTCAGGAGTA) of

538   the *ARID1B* gene. *GAPDH* was used as a control gene (Fw: 5'- GGTGGTCTCCTCTGACTTCAA

539   C; Rv: 5'- TTCGTTGTCATACCAGGAAATG).

540

541   **Human Pluripotent Stem Cell Differentiation into the Neural Lineage**

542   Human iPSC cultures were maintained in mTeSR1 medium (StemCell Technologies). To start

543   NPC differentiation, cells were dissociated to single cells using Accutase (Life Technologies) and

544   re-aggregated for 48h in Ultra-Low Cluster 96-well round bottom plates (CoStar) at a density of

545   7500 cells per well in mTeST1 supplemented with 1 uM Thiazovivin. On day 1 of differentiation,
546   an equal volume of Neural Induction Medium (NIM) was added: 10% Knock-Out Serum
547   Replacement, Penicillin/Streptomycin, and GlutaMax in Dulbecco's Modified Eagle Medium
548   (DMEM) (Life Technologies) supplemented with 10 uM RepSox (Tocris), 1 uM LDN-193189
549   (StemCell Technologies), and 1 uM Thiazovivin. From day 2 to day 5, half of the spent medium
550   was removed daily and replaced with fresh NIM without Thiazovivin. Starting on day 6, Neural
551   Progenitor Medium (NPM) was used, containing 2% B27 Supplement without vitamin A,
552   Penicillin/Steptomycin, and GlutaMax in DMEM/F12 (Life Technologies) supplemented with 20
553   ng/ml EGF and 20 ng/ml FGF (PeproTech); 10 uM RepSox was only added until day 10. On day
554   7, cell aggregates were transferred to Geltrex (Life-Technologies)-coated 6-well plates. Neural
555   rosettes were carefully dissected from the adherent culture on days 12-14 and maintained as
556   neurospheres in suspension, with medium change every 1-2 days and gentle dissociation by
557   trituration every 4-5 days until the cells were used in experiments. EGF and FGF were omitted
558   from the NPM medium starting on day 40.
559
560

561   **FIRE-Cas9 Rapid and Inducible Activation of *ARID1B***
562   Activation of *ARID1B* in human iPSC *ARID1B^{+/-}* cultures were carried out in four independent
563   experiments using the FIRE-Cas9 activation (VPR) system described in [34]. Briefly, iPSC cultures
564   were sequentially infected with lentiviral constructs and selected using an appropriate antibiotic
565   (blasticidin: 5ug/mL; hygromycin: 25ug/mL, puromycin: 0.75ug/mL, zeocin: 25ug/mL).
566   Throughout, infection and selection concentrations were carefully titred and cultures were
567   monitored for maintenance of pluripotency with microscopy. 3nM rapamycin was added for 8-48
568   hrs to induce activation. Cultures were washed 1X with ice-cold PBS and total RNA was extracted
569   with Trisure (Bioline), purified, and DNAse-treated using the Direct-zol RNA kit (Zymo). cDNA
570   was synthesized from 1 ug RNA using the SensiFAST kit (Bioline). For RT-qPCR, samples were
571   prepared using the SYBR LO-ROX kit (Bioline) and analyzed on a QuantStudio 6 Flex (Life
572   Technologies). Two sets of *ARID1B* primers targeting two different exons were chosen based on
573   a standard curve analysis and Ct values were normalized to Ct values for the housekeeping gene
574   *GAPDH* (dCt = Ct$_{ARID1B}$ – Ct$_{GAPDH}$) for each sample before further analysis. Primer sequences are:
575   *ARID1B* Exon12-13 fwd: CTACGTCTGCGTCAAAGAGATCG
576   *ARID1B* Exon12-13 rev: GTTGCCAGCTCACGCCACTTCTTG
577   *ARID1B* Exon18-19 fwd: GCATCACAACCAGTCTTGAAACAA
578   *ARID1B* Exon18-19 rev: ATCACACGCCACGCCTCAGGAGTA
579   *GAPDH* fwd: GGTGGTCTCCTCTGACTTCAAC
580   *GAPDH* rev: TTCGTTGTCATACCAGGAAATG
581   Guide Design: spCas9 guide RNAs (gRNAs) targeting the *ARID1B* promoter were designed using
582   the GPP sgRNA Design Tool/CRISPick [77,78] and cloned into lentiviral constructs (Lv U6 sgRNA-
583   2xMS2-sgRNA EF1 Zeo) from [34]. sgRNA sequence: CCGCGCGGCCATGATCGCCG
584   Lentivirus Production: HEK293T cells were transfected with 18ug of the respective lentiviral
585   construct (Lv U6 sgRNA-2xMS2-sgRNA EF1 Zeo, Lv EF1a dCas9-2A-Blast, Lv EF1a MCP-
586   2xFkbp-2A-Hygro, Lv EF1a VPR-2xFrb Pgk Puro; described in [34] and packaging plasmids
587   (psPAX2 and pMD2.G) using PEI (Polysciences). 48hrs after transfection, the media was
588   collected, filtered with a 0.45 um filter (Millipore), and ultra-centrifuged at 50,000 x *g* for 2hrs at
589   4°C. The viral pellet was resuspended in PBS and used for infection.
590

RNA-seq Library Preparation
Cultures were washed 1X with ice-cold PBS and total RNA was extracted with Trisure (Bioline), purified and DNAse-treated using the Direct-zol RNA kit (Zymo). Libraries were prepared from 500ng RNA with dual-index unique barcodes using the SMARTer Stranded Total RNA Sample Prep HI Kit (Clontech), quantified by qPCR using a PhiX (Illumina) standard curve, and checked for appropriate fragment size distribution by Tapestation (Agilent). Libraries were pooled and multiplexed 1:1 and run on a HiSeq3000 (Illumina) for 2x150 paired-end sequencing (Novogene) with a 20% PhiX spike-in.

ATAC-seq library preparation

Cells were washed and collected in ice-cold PBS using a cell lifter (Corning #3008), and then counted using a Countess II FL cell counter (Applied Biosystems). Accessible chromatin from 75,000 cells per biological sample were transposed and tagged using Tn5 (Nextera DNA Sample Prep Kit from Illumina, cat# FC-121-1030), then barcoded and amplified following a published protocol (PMID: 24097267). DNA fragment sizes were determined using a TapeStation (Agilent) with High Sensitivity D1000 Screentapes (Agilent cat# 5067-5584) and showed the expected pattern for non-nucleosomal and nucleosomal sizes. Libraries were quantified by PCR using a PhiX library (Illumina cat# TG-110-3001) standard curve and then multiplexed at a 1:1 ratio. Pooled libraries were run on a HiSeq3000 (Illumina) for 2x75 paired-end sequencing at the Stanford Functional Genomics Facility.

CUT&RUN Sample Preparation
Cultures were washed 1X with ice-cold PBS and dissociated to single cells using Accutase (Life Technologies). CUT&RUN was performed with an anti-H3K27me3 antibody (ActiveMotif rabbit polyclonal #39155) as described in [79]. Briefly, exactly 250,000 cells were counted with an automated cell counter (Countess, Life Technologies), bound to Concanavalin A beads (BioMagPlus, Bangs Laboratories #BP531). The anti-H3K27me3 antibody was added (1:50 dilution) and incubated with cells under 800rpm shaking at 4^OC for 2 hrs. After wash steps, a 1:200 dilution of proteinA-MNase was added and incubated for 1hr at 4^OC. After washing, 100mM $CaCl_2$ was added at 0^OC and incubated for 15 mins to activate the MNase. The reaction was stopped by addition of a 10mM EDTA/EGTA Stop Buffer (see Meers et al. 2019 for details), incubated at 37^OC for 10mins, and centrifuged at 4^OC, 5000xg for 15 mins. The supernatant containing DNA was taken and purified with the DNA Clean and Concentrator Kit (Zymo). Samples were quantified on a HS DNA Qubit Fluorometer (Life Technologies) and fragment size distribution checked with a Tapestation (Agilent).

CUT&RUN Library Preparation
Dual-index libraries were prepared from CUT&RUN eluted DNA using the NEBNext Ultra II DNA Library Prep (E7103S) with NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 1, E7600S) following manufacturer instructions. Libraries were quantified by qPCR using a PhiX (Illumina) standard curve, and checked for appropriate fragment size distribution by Tapestation (Agilent). Libraries were pooled and multiplexed 1:1 and run on a HiSeq3000 (Illumina) for 2x150 paired-end sequencing (Novogene).

635 **RNA-seq data analysis**
636 Raw reads were pseudoaligned to GENCODE v33 human transcriptome assembly [80] and transcript
637 abundance was quantified using kallisto (version 0.44.0) [81]. Transcript abundance files from
638 kallisto output were imported to R using tximport [82]. DESeq2 [83] was used for analysis of
639 differential gene expression. For visualization and gene ranking we used effect size shrinkage for
640 Log2FC estimates as encoded by *ashr()* function [84]. Enrichr [85] was used for GO-term analysis of
641 differentially expressed genes, defined as genes with |log2FC| > 0.5 and FDR < 0.05. PANTHER
642 Pathway analysis was performed using PANTHER [86].
643
644 **ATAC-seq data analysis**
645 The adapters were trimmed from the raw reads using *trim_galore* and *cutadapt* [87]. The trimmed
646 reads were aligned to GRCh38 human genome assembly using bowtie2 [88] with *--very-sensitive -*
647 *X2000* parameters. Not aligned reads and paired reads with at least one read in a pair falling below
648 quality score threshold 10 were removed using *samtools* [89] and *awk* text editor. Additionally,
649 paired reads were removed if each read in a pair have multiple alignment sites. PCR and optical
650 duplicates were removed using *Picard*. Tn5 insertion sites were defined as the start read position
651 offset by +4 bp for the reads aligned to + strand and as the end read position offset by -5 bp for the
652 reads aligned to - strand [90]. Macs2 (v2.1.1) [91] was used to reconstruct accessibility peaks around
653 Tn5 insertion sites with the following parameters *--shift -75 --extsize 150 --nomodel --call-summits*
654 *--nolambda --keep-dup all -p 0.01*. To derive consensus peak set we (1) merged individual
655 biological replica data sets across each condition (WT, HET, KO), (2) applied *macs2* on the
656 merged data sets and filtered out all peaks falling below the random quality threshold 100, (3)
657 merged resulting peaks using *bedtools* [92]. The raw counts in the peak regions were calculated using
658 *bedtools*. Peak differential analysis was performed using DESeq2 [83]. GO-term enrichment analysis
659 was performed using GREAT web server [93] on a set of peaks with differential accessibility, with
660 a set of all accessibility peaks used as a background. HOMER [94] was used to perform TF
661 enrichment analysis in the differentially accessible peaks, defined as peaks with |log2FC| > 0.5 and
662 FDR < 0.05. A heatmap was produced using deeptools (version 3.3.0) [95].
663
664 **Footprint Depth and Flanking accessibility analysis**
665 We used non-redundant position frequency matrices from JASPAR 2020 [96] CORE (vertebrates)
666 and UNVALIDATED collections comprising 746 and 337 transcription factors respectively.
667 FIMO tool from the MEME Suite [97] was used to locate individual transcription factor binding sites.
668 The sites with identical genomic coordinates located on the opposite strands were merged. For
669 each TF within each individual ATAC-seq dataset we built an accessibility profile, defined as the
670 number of Tn5 insertion sites counted as function of distance from the TF motif borders. The
671 counting was done over all TF motifs present within accessibility peaks. Based on TF accessibility
672 profiles we defined Flanking Accessibility (FA) and Footprint Depth (FPD).

$$FA = \sum_i \log_2 \frac{\overline{N}_{FLANK}}{\overline{N}_{BG}}$$

674 where $\overline{N}_{FLANK}$ is a number of Tn5 insertion sites found in the region adjacent to TF motif, [-55 bp,
675 -6 bp] U [6 bp, 55 bp], and divided by the region length. – and + refer to nucleotides located
676 upstream and downstream the lower and upper motif boundaries respectively. $\overline{N}_{BG}$ is a number of
677 Tn5 insertion sites in the background region, [-250 bp, -200 bp] U [200 bp, 250 bp], divided by
678 the region length. The index *i* runs over all biological replicas in a single condition (KO, HET,
679 WT).

$$FPD = \sum_i \log_2 \frac{\overline{N}_{MOTIF}}{\overline{N}_{FLANK}}$$

680

681 where $\overline{N}_{MOTIF}$ is a number of Tn5 insertion sites in the region centered on the motif, [-5 bp, 5 bp],
682 divided by the region length. To evaluate changes in TF binding across conditions, we calculated
683 $\Delta FA_{TF} = FA^{KO\ (HET)} - FA^{WT}$ and $\Delta FPD_{TF} = FPD^{KO\ (HET)} - FPD^{WT}$. A statistical significance
684 (p-value) of $(\Delta FA, \Delta FPD)_{TF}$ was evaluated using bootstrapping technique as follows. For each TF
685 and each condition (KO, HET, WT), we built an artificial accessibility profile from a set of motifs
686 randomly selected from a set we used for $(\Delta FA, \Delta FPD)_{TF}$ calculations. A single motif could be
687 selected multiple times. The total number of selected motifs was equal to the size of the motif set
688 used for $(\Delta FA, \Delta FPD)_{TF}$ calculations. For each selected motif Tn5 insertion sites from a randomly
689 chosen biological replica within a single condition were merged into the artificial accessibility
690 profile. These artificial accessibility profiles were used to calculate new ΔFA and ΔFPD values.
691 The procedure was repeated 10,000 times resulting in a bootstrapped $(\Delta FA, \Delta FPD)$ distribution
692 built separately for each TF, which reflects the statistical uncertainty in evaluation of ΔFA and
693 ΔFPD for that TF. To evaluate p-value, we created a 1-dimensional histogram, filled with the
694 number of $(\Delta FA, \Delta FPD)$ points found within bins along the axis connecting a center of
695 bootstrapped $(\Delta FA, \Delta FPD)$ distribution and $(\Delta FA, \Delta FPD)_{TF}$ observed for specific TF. FDRs
696 were calculated using Benjamini & Hochberg correction for multiple comparisons using R p.adjust
697 function. For visualization purpose we placed the center of $(\Delta FA, \Delta FPD)$ distribution into the
698 origin of the coordinate system and labeled TFs, with FDR less than 0.05 and random threshold
699 on minimal expression at 50 normalized counts.

700

701 **CUT&RUN H3K27me3 data analysis**
702 The adapters were trimmed from the raw reads using *trim_galore* and *cutadapt* [87]. The trimmed
703 reads were aligned to GRCh38 human genome assembly using bowtie2 [88] with *--very-sensitive -*
704 *X2000* parameters. Not aligned reads and paired reads with at least one read in a pair falling below
705 quality score threshold 10 were removed using *samtools* [89] and *awk* text editor. Additionally,
706 paired reads were removed if each read in a pair have multiple alignment sites. PCR and optical
707 duplicates were removed using *Picard*. Macs2 (v2.1.1) [91] was used to reconstruct peaks from
708 fragments with the following parameters --format BEDPE -B --call-summits --nolambda -p 0.01.
709 To derive consensus peak set we (1) merged individual biological replica data sets across each
710 condition (WT, HET, KO), (2) applied *macs2* on the merged data sets and filtered out all peaks
711 falling below the random quality threshold 200, (3) merged resulting peaks using *bedtools* [92]. The
712 raw counts in the peak regions were calculated using *bedtools*. Peak differential analysis was
713 performed using DESeq2 [83]. GO-term enrichment analysis was performed using GREAT web
714 server [93] on a set of peaks with significant change in H3K27me3 level across conditions. A
715 significant change was defined as |log2FC| > 0.5 and FDR < 0.05. A set of all peaks was used as a
716 background.

717

718 **Cut&Run H2AK119ub data analysis**
719 The adapters were trimmed from the raw reads using *trim_galore* and *cutadapt* [87]. The trimmed
720 reads were aligned to GRCh38 human genome assembly using bowtie2 [88] with *--very-sensitive -*
721 *X2000* parameters. Not aligned reads and paired reads with at least one read in a pair falling below
722 quality score threshold 10 were removed using *samtools* [89] and *awk* text editor. Additionally,
723 paired reads were removed if each read in a pair have multiple alignment sites. PCR and optical

duplicates were removed using *Picard*. Macs2 (v2.1.1) [91] was used to reconstruct peaks from fragments with the following parameters --format BEDPE -B --call-summits --nolambda -p 0.01.

Analysis of autism related genes
We used SFARI gene database of autism related genes (release 1/13/2021).

Author Contributions
E.Y.S. and A.K. conceived the project. E.Y.S. derived the cell lines and performed RNA-seq and ATAC-seq experiments. E.Y.S. and S.G. designed and performed the *ARID1B* activation experiment. E.Y.S., A.K. and C.Y.C. performed the CUT&RUN experiments. A.K. performed data analysis. A.K., E.Y.S., and G.R.C. wrote the manuscript with contribution from all authors. G.R.C. designed experiments and supervised the project.

Competing interests
G.R.C. is founder and stockholder of Foghorn Therapeutics. The other authors declare no competing interests.

REFERENCES

1. Abrams, E., Neigeborn, L. & Carlson, M. Molecular analysis of SNF2 and SNF5, genes required for expression of glucose-repressible genes in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **6**, 3643–3651 (1986).
2. Stern, M., Jensen, R. & Herskowitz, I. Five SWI genes are required for expression of the HO gene in yeast. *J. Mol. Biol.* **178**, 853–868 (1984).
3. Tamkun, J. W. *et al.* brahma: A regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2 SWI2. *Cell* **68**, 561–572 (1992).
4. Lessard, J. *et al.* An Essential Switch in Subunit Composition of a Chromatin Remodeling Complex during Neural Development. *Neuron* **55**, 201–215 (2007).
5. Wu, J. I. *et al.* Regulation of Dendritic Development by Neuron-Specific Chromatin Remodeling Complexes. *Neuron* **56**, 94–108 (2007).
6. Yoo, A. S., Staahl, B. T., Chen, L. & Crabtree, G. R. MicroRNA-mediated switching of chromatin-remodelling complexes in neural development. *Nature* **460**, 642–646 (2009).
7. Yoo, A. S. *et al.* MicroRNA-mediated conversion of human fibroblasts to neurons. 3–7 (2011). doi:10.1038/nature10323
8. 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).
9. Son, E. Y. & Crabtree, G. R. The role of BAF (mSWI/SNF) complexes in mammalian neural development. *Am. J. Med. Genet. Part C Semin. Med. Genet.* **166**, 333–349 (2014).
10. 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).

770 11. Kosho, T., Miyake, N. & Carey, J. C. Coffin-Siris syndrome and related disorders
771   involving components of the BAF (mSWI/SNF) complex: Historical review and recent
772   advances using next generation sequencing. *Am. J. Med. Genet. Part C Semin. Med.*
773   *Genet.* **166**, 241–251 (2014).
774 12. Cheng, S. S. W., Luk, H. M., Mok, M. T. S., Leung, S. S. & Lo, I. F. M. Genotype and
775   phenotype in 18 Chinese patients with Coffin-Siris syndrome. *Am. J. Med. Genet. Part A*
776   (2021). doi:10.1002/ajmg.a.62187
777 13. Santen, G. W. E. *et al.* The ARID1B phenotype: What we have learned so far. *Am. J. Med.*
778   *Genet. Part C Semin. Med. Genet.* **166**, 276–289 (2014).
779 14. Wright, C. F. *et al.* Genetic diagnosis of developmental disorders in the DDD study: A
780   scalable analysis of genome-wide research data. *Lancet* **385**, 1305–1314 (2015).
781 15. Hoyer, J. *et al.* Haploinsufficiency of ARID1B, a member of the SWI/SNF-A chromatin-
782   remodeling complex, is a frequent cause of intellectual disability. *Am. J. Hum. Genet.* **90**,
783   565–572 (2012).
784 16. Halgren, C. *et al.* Corpus callosum abnormalities, intellectual disability, speech
785   impairment, and autism in patients with haploinsufficiency of ARID1B. *Clin. Genet.* **82**,
786   248–255 (2012).
787 17. Roak, B. J. O. *et al.* Multiplex Targeted Sequencing Identifies Recurrently Mutated Genes
788   in Autism Spectrum Disorders. *Science (80-.).* **23**, 1619–1623 (2012).
789 18. Satterstrom, F. K. *et al.* Large-Scale Exome Sequencing Study Implicates Both
790   Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568-
791   584.e23 (2020).
792 19. Fitzgerald, T. W. *et al.* Large-scale discovery of novel genetic causes of developmental
793   disorders. *Nature* **519**, 223–228 (2015).
794 20. Celen, C. *et al.* \textit{Arid1b} haploinsufficient mice reveal neuropsychiatric phenotypes
795   and reversible causes of growth impairment. *Elife* **6**, e25730 (2017).
796 21. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in
797   141,456 humans. *Nature* **581**, 434–443 (2020).
798 22. Clapier, C. R., Iwasa, J., Cairns, B. R. & Peterson, C. L. Mechanisms of action and
799   regulation of ATP-dependent chromatin-remodelling complexes. *Nat. Rev. Mol. Cell Biol.*
800   **18**, 407–422 (2017).
801 23. Ho, L. *et al.* EsBAF facilitates pluripotency by conditioning the genome for LIF/STAT3
802   signalling and by regulating polycomb function. *Nat. Cell Biol.* **13**, 903–913 (2011).
803 24. King, H. W. & Klose, R. J. The pioneer factor OCT4 requires the chromatin remodeller
804   BRG1 to support gene regulatory element function in mouse embryonic stem cells. *Elife*
805   **6**, 1–24 (2017).
806 25. Miller, E. L. *et al.* TOP2 synergizes with BAF chromatin remodeling for both resolution
807   and formation of facultative heterochromatin. *Nat. Struct. Mol. Biol.* **24**, 1–11 (2017).
808 26. Hodges, H. C. *et al.* Dominant-negative SMARCA4 mutants alter the accessibility
809   landscape of tissue-unrestricted enhancers. *Nat. Struct. Mol. Biol.* **25**, 61–72 (2018).
810 27. Iurlaro, M. *et al.* Mammalian SWI / SNF continuously restores local accessibility to
811   chromatin. *Nat. Genet.* (2021). doi:10.1038/s41588-020-00768-w
812 28. Schick, S. *et al.* Acute BAF perturbation causes immediate changes in chromatin
813   accessibility. *Nat. Genet.* **53**, (2021).
814 29. Weber, C. M. *et al. mSWI/SNF promotes Polycomb repression both directly and through*
815   *genome-wide redistribution. Nature Structural and Molecular Biology* **28**, (Springer US,

816   2021).

817   30.   Kia, S. K., Gorski, M. M., Giannakopoulos, S. & Verrijzer, C. P. SWI/SNF Mediates
818   Polycomb Eviction and Epigenetic Reprogramming of the INK4b-ARF-INK4a Locus.
819   *Mol. Cell. Biol.* **28**, 3457–3464 (2008).

820   31.   Wilson, B. G. *et al.* Epigenetic antagonism between polycomb and SWI/SNF complexes
821   during oncogenic transformation. *Cancer Cell* **18**, 316–328 (2010).

822   32.   Stanton, B. Z. *et al.* Smarca4 ATPase mutations disrupt direct eviction of PRC1 from
823   chromatin. *Nat. Publ. Gr.* **49**, 282–288 (2016).

824   33.   Kadoch, C. *et al.* Dynamics of BAF – Polycomb complex opposition on heterochromatin
825   in normal and oncogenic states. *Nat. Genet.* **49**, 213–222 (2016).

826   34.   Braun, S. M. G. *et al.* Rapid and reversible epigenome editing by endogenous chromatin
827   regulators. *Nat. Commun.* **8**, (2017).

828   35.   Ni, K. *et al.* LSH mediates gene repression through macroH2A deposition. *Nat. Commun.*
829   **11**, (2020).

830   36.   Zákány, J., Fromental-Ramain, C., Warot, X. & Duboule, D. Regulation of number and
831   size of digits by posterior Hox genes: A dose-dependent mechanism with potential
832   evolutionary implications. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 13695–13700 (1997).

833   37.   Quinonez, S. C. & Innis, J. W. Human HOX gene disorders. *Mol. Genet. Metab.* **111**, 4–
834   15 (2014).

835   38.   Shibutani, M. *et al.* Arid1b haploinsufficiency causes abnormal brain gene expression and
836   autism-related behaviors in mice. *Int. J. Mol. Sci.* **18**, (2017).

837   39.   Catela, C., Shin, M. M., Lee, D. H., Liu, J. P. & Dasen, J. S. Hox Proteins Coordinate
838   Motor Neuron Differentiation and Connectivity Programs through Ret/Gfrα Genes. *Cell*
839   *Rep.* **14**, 1901–1915 (2016).

840   40.   Mendelsohn, A. I., Dasen, J. S. & Jessell, T. M. Divergent Hox Coding and Evasion of
841   Retinoid Signaling Specifies Motor Neurons Innervating Digit Muscles. *Neuron* **93**, 792-
842   805.e4 (2017).

843   41.   Jung, E. M. *et al.* Arid1b haploinsufficiency disrupts cortical interneuron development and
844   mouse behavior. *Nat. Neurosci.* **20**, 1694–1707 (2017).

845   42.   Smith, A. L., Jung, E. M., Jeon, B. T. & Kim, W. Y. Arid1b haploinsufficiency in
846   parvalbumin- or somatostatin-expressing interneurons leads to distinct ASD-like and ID-
847   like behavior. *Sci. Rep.* **10**, 1–13 (2020).

848   43.   Ellegood, J. *et al.* Neuroanatomy and behavior in mice with a haploinsufficiency of AT-
849   rich interactive domain 1B (ARID1B) throughout development. *Mol. Autism* **12**, 1–24
850   (2021).

851   44.   Moffat, J. J. *et al.* Differential roles of ARID1B in excitatory and inhibitory neural
852   progenitors in the developing cortex. *Sci. Rep.* **11**, 1–17 (2021).

853   45.   Walsh, J. J. *et al.* Systemic enhancement of serotonin signaling reverses social deficits in
854   multiple mouse models for ASD. *Neuropsychopharmacology* (2021). doi:10.1038/s41386-
855   021-01091-6

856   46.   Morrill, S. A. & Amon, A. Why haploinsufficiency persists. *Proc. Natl. Acad. Sci. U. S. A.*
857   **116**, 11866–11871 (2019).

858   47.   Wenderski, W. *et al.* Loss of the neural-specific BAF subunit ACTL6B relieves repression
859   of early response genes and causes recessive autism. *Proc. Natl. Acad. Sci. U. S. A.* **117**,
860   10055–10066 (2020).

861   48.   Aizawa, H. *et al.* Dendrite Development Regulated by CREST, a Calcium-Regulated

862   Transcriptional Activator. *Science (80-.).* **303**, 197–203 (2004).

863   49.   Vasileiou, G. *et al.* Chromatin-Remodeling-Factor ARID1B Represses Wnt/β-Catenin
864   Signaling. *Am. J. Hum. Genet.* **97**, 445–456 (2015).

865   50.   Liu, X. *et al.* De Novo ARID1B mutations cause growth delay associated with aberrant
866   Wnt/β–catenin signaling. *Hum. Mutat.* **41**, 1012–1024 (2020).

867   51.   Nagl, N. G., Wang, X., Patsialou, A., Van Scoy, M. & Moran, E. Distinct mammalian
868   SWI/SNF chromatin remodeling complexes with opposing roles in cell-cycle control.
869   *EMBO J.* **26**, 752–763 (2007).

870   52.   Mah, K. M. & Weiner, J. A. Regulation of Wnt signaling by protocadherins. *Semin. Cell*
871   *Dev. Biol.* **69**, 158–171 (2017).

872   53.   Xi, Q., He, W., Zhang, X. H. F., Le, H. Van & Massague, J. Genome-wide impact of the
873   BRG1 SWI/SNF chromatin remodeler on the transforming growth factor β transcriptional
874   program. *J. Biol. Chem.* **283**, 1146–1155 (2008).

875   54.   He, W. *et al.* Hematopoiesis Controlled by Distinct TIF1γ and Smad4 Branches of the
876   TGFβ Pathway. *Cell* **125**, 929–941 (2006).

877   55.   Gao, F. *et al.* Heterozygous Mutations in SMARCA2 Reprogram the Enhancer Landscape
878   by Global Retargeting of Article Heterozygous Mutations in SMARCA2 Reprogram the
879   Enhancer Landscape by Global Retargeting of SMARCA4. *Mol. Cell* 1–14 (2019).
880   doi:10.1016/j.molcel.2019.06.024

881   56.   Baek, S. *et al.* Bivariate Genomic Footprinting Detects Changes in Resource Bivariate
882   Genomic Footprinting Detects Changes in Transcription Factor Activity. *CellReports* **19**,
883   1710–1722 (2017).

884   57.   Corces, M. R. *et al.* The chromatin accessibility landscape of primary human cancers.
885   *Science (80-.).* **362**, eaav1898 (2018).

886   58.   Kennison, J. A. & Tamkun, J. W. Dosage-dependent modifiers of Polycomb and
887   Antennapedia mutations in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 8136–8140
888   (1988).

889   59.   Schuettengruber, B., Bourbon, H., Croce, L. Di & Cavalli, G. Review Genome Regulation
890   by Polycomb and Trithorax : 70 Years and Counting. *Cell* **171**, 34–57 (2017).

891   60.   Bernstein, B. E. *et al.* A Bivalent Chromatin Structure Marks Key Developmental Genes
892   in Embryonic Stem Cells. *Cell* **125**, 315–326 (2006).

893   61.   Mazzoni, E. O. *et al.* Saltatory remodeling of Hox chromatin in response to rostrocaudal
894   patterning signals. *Nat. Neurosci.* **16**, 1191–1198 (2013).

895   62.   Flajollet, S., Lefebvre, B., Cudejko, C., Staels, B. & Lefebvre, P. The core component of
896   the mammalian SWI/SNF complex SMARCD3/BAF60c is a coactivator for the nuclear
897   retinoic acid receptor. *Mol. Cell. Endocrinol.* **270**, 23–32 (2007).

898   63.   Yamamichi, N. *et al.* Cdx2 and the Brm-type SWI/SNF complex cooperatively regulate
899   villin expression in gastrointestinal cells. *Exp. Cell Res.* **315**, 1779–1789 (2009).

900   64.   Wang, K. *et al.* Brg1 is required for Cdx2-mediated repression of Oct4 expression in
901   mouse blastocysts. *PLoS One* **5**, (2010).

902   65.   Blackledge, N. P. & Klose, R. J. The molecular principles of gene regulation by Polycomb
903   repressive complexes. *Nat. Rev. Mol. Cell Biol.* **0123456789**, (2021).

904   66.   Huang, Z. Q., Li, J., Sachs, L. M., Cole, P. A. & Wong, J. A role for cofactor-cofactor and
905   cofactor-histone interactions in targeting p300, SWI/SNF and Mediator for transcription.
906   *EMBO J.* **22**, 2146–2155 (2003).

907   67.   Ogiwara, H. *et al.* Histone acetylation by CBP and p300 at double-strand break sites

908   facilitates SWI/SNF chromatin remodeling and the recruitment of non-homologous end
909   joining factors. *Oncogene* **30**, 2135–2146 (2011).

910 68.   Alver, B. H. *et al.* The SWI/SNF chromatin remodelling complex is required for
911   maintenance of lineage specific enhancers. *Nat. Commun.* **8**, (2017).

912 69.   Wu, J. I., Lessard, J. & Crabtree, G. R. Understanding the Words of Chromatin
913   Regulation. *Cell* **136**, 200–206 (2009).

914 70.   Mashtalir, N. *et al.* A Structural Model of the Endogenous Human BAF Complex Informs
915   Disease Mechanisms. *Cell* **183**, 802-817.e24 (2020).

916 71.   Pan, J. *et al.* The ATPase module of mammalian SWI/SNF family complexes mediates
917   subcomplex identity and catalytic activity–independent genomic targeting. *Nat. Genet.* **51**,
918   618–626 (2019).

919 72.   Wang, W. *et al.* Diversity and specialization of mammalian SWI / SNF complexes. *Genes*
920   *Dev.* **10**, 2117–2130 (1996).

921 73.   Wang, W. *et al.* Purification and biochemical heterogeneity of the mammalian SWI-SNF
922   complex. *EMBO J.* **15**, 5370–5382 (1996).

923 74.   Staahl, B. T. *et al.* Kinetic analysis of npBAF to nBAF switching reveals exchange of
924   SS18 with CREST and integration with neural developmental pathways. *J. Neurosci.* **33**,
925   10348–10361 (2013).

926 75.   Shi, H. *et al.* ARID1A loss in neuroblastoma promotes the adrenergic-to-mesenchymal
927   transition by regulating enhancer-mediated gene expression. *Sci. Adv.* **6**, 1–13 (2020).

928 76.   Huang, H. S. *et al.* Topoisomerase inhibitors unsilence the dormant allele of Ube3a in
929   neurons. *Nature* **481**, 185–191 (2012).

930 77.   Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-
931   target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).

932 78.   Sanson, K. R. *et al.* Optimized libraries for CRISPR-Cas9 genetic screens with multiple
933   modalities. *Nat. Commun.* **9**, 1–15 (2018).

934 79.   Meers, M. P., Bryson, T. D., Henikoff, J. G. & Henikoff, S. Improved CUT&RUN
935   chromatin profiling tools. *Elife* **8**, 1–16 (2019).

936 80.   Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes.
937   *Nucleic Acids Res.* **47**, D766–D773 (2019).

938 81.   Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq
939   quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).

940 82.   Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq:
941   transcript-level estimates improve gene-level inferences. *F1000Research* **4**, 1521 (2015).

942 83.   Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion
943   for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

944 84.   Stephens, M. False discovery rates: A new deal. *Biostatistics* **18**, 275–294 (2017).

945 85.   Kuleshov, M. V *et al.* Enrichr : a comprehensive gene set enrichment analysis web server
946   2016 update. *Nucleic Acids Res.* **44**, 90–97 (2016).

947 86.   Mi, H., Muruganujan, A. & Thomas, P. D. PANTHER in 2013: Modeling the evolution of
948   gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic*
949   *Acids Res.* **41**, 377–386 (2013).

950 87.   Marcel, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
951   *EMBnet. J.* **17**, 10–12 (2011).

952 88.   Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*
953   **9**, 357–359 (2012).

954   89.   Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,
955   2078–2079 (2009).
956   90.   Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J.
957   Transposition of native chromatin for fast and sensitive epigenomic profiling of open
958   chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218
959   (2013).
960   91.   Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). (2008). doi:10.1186/gb-
961   2008-9-9-r137
962   92.   Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing
963   genomic features. *Bioinformatics* **26**, 841–842 (2010).
964   93.   McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions.
965   *Nat. Biotechnol.* **28**, 495–501 (2010).
966   94.   Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime
967   cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* **38**,
968   576–589 (2010).
969   95.   Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data
970   analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
971   96.   Fornes, O. *et al.* JASPAR 2020: Update of the open-Access database of transcription
972   factor binding profiles. *Nucleic Acids Res.* **48**, D87–D92 (2020).
973   97.   Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: Scanning for occurrences of a given
974   motif. *Bioinformatics* **27**, 1017–1018 (2011).
975
976

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryData1.pdf
- SupplementaryFigures.docx