LETTERS
https://doi.org/10.1038/s41588-019-0363-5
nature genetics

The ATPase module of mammalian SWI/SNF family complexes mediates subcomplex identity and catalytic activity–independent genomic targeting

Joshua Pan[1,2,3], Zachary M. McKenzie[1,4], Andrew R. D'Avino[1,2], Nazar Mashtalir[1,2], Caleb A. Lareau[2,5,6], Roodolph St. Pierre[1,2,7], Lu Wang[8], Ali Shilatifard[8] and Cigall Kadoch[1,2,4*]

Perturbations to mammalian switch/sucrose non-fermentable (mSWI/SNF) chromatin remodeling complexes have been widely implicated as driving events in cancer[1]. One such perturbation is the dual loss of the SMARCA4 and SMARCA2 ATPase subunits in small cell carcinoma of the ovary, hypercalcemic type (SCCOHT)[2–5], SMARCA4-deficient thoracic sarcomas[6] and dedifferentiated endometrial carcinomas[7]. However, the consequences of dual ATPase subunit loss on mSWI/SNF complex subunit composition, chromatin targeting, DNA accessibility and gene expression remain unknown. Here we identify an ATPase module of subunits that is required for functional specification of the Brahma-related gene–associated factor (BAF) and polybromo-associated BAF (PBAF) mSWI/SNF family subcomplexes. Using SMARCA4/2 ATPase mutant variants, we define the catalytic activity–dependent and catalytic activity–independent contributions of the ATPase module to the targeting of BAF and PBAF complexes on chromatin genome-wide. Finally, by linking distinct mSWI/SNF complex target sites to tumor-suppressive gene expression programs, we clarify the transcriptional consequences of SMARCA4/2 dual loss in SCCOHT.

To define mSWI/SNF complex size and composition following dual loss of the SMARCA4 and SMARCA2 ATPases, we leveraged two cell line models of SCCOHT: BIN-67 (refs. [8,9]) and SCCOHT-1 (ref. [10]), which we confirmed to be deficient in the proteins SMARCA4 and SMARCA2 (Supplementary Fig. 1a). Using density sedimentation of nuclear extracts coupled with immunoblotting, we observed that a subset of remaining mSWI/SNF subunits co-sedimented in the absence of SMARCA4 and SMARCA2 (Fig. 1a and Supplementary Fig. 1b), eluting in lower molecular-weight fractions than wild-type (WT) BAF and PBAF mSWI/SNF family subcomplexes (Supplementary Fig. 1c). We confirmed that this co-sedimentation profile was a result of direct interaction between remaining subunits, and that residual complexes preserved mutually exclusive occupancy of subcomplex-specific subunits such as ARID1A (BAF-specific) and ARID2 (PBAF-specific) (Supplementary Fig. 1d).

To comprehensively define the components of the residual complex, we performed immunoprecipitations from SCCOHT-1 nuclear extracts and subjected eluted proteins to mass-spectrometry-based proteomic analysis (Fig. 1b and Supplementary Table 1). These studies defined two groups of mSWI/SNF subunits: residual complex components whose interactions were preserved in the absence of SMARCA4/2 ATPase subunits (ARID1A, ARID2, SMARCB1, SMARCC1/2, SMARCD1/2/3, SMARCE1, DPF2, GLTSCR1/L and BRD7/9), and subunits that required SMARCA4/2 for complex binding (ACTL6A, BCL7A/B/C, PBRM1 and SS18), which we termed 'ATPase module' subunits. We replicated these results in the SMARCA4/2-deficient SW-13 cell line[11], in which purified mSWI/SNF complexes similarly eluted at a lower molecular weight (Fig. 1c) and selectively lacked ATPase module subunits (Supplementary Fig. 1e), consistent with parallel studies from our laboratory[12] as well as independent demonstrations that ACTL6A is destabilized following loss of SMARCA4/2 (refs. [13,14]).

Given that the subunits of the residual complex contain numerous DNA-binding domains (such as the HMG domain of SMARCE1, winged helix domain of SMARCB1), we tested for residual genomic targeting in the BIN-67 cell line by using chromatin immunoprecipitation followed by sequencing (ChIP-seq). Subunits in the residual complex (SMARCC1, DPF2, ARID2) bound accessible chromatin regions defined by an assay for transposase-accessible chromatin using sequencing (ATAC-seq) (Fig. 1d) such as active transcriptional start sites (TSSs) (Supplementary Fig. 1f,g). In contrast, ChIP-seq performed on ATPase module subunits (SS18, SMARCA4) returned almost no significant peaks in these regions (Fig. 1e), consistent with their absence from the complex and confirming the high specificity of the mSWI/SNF ChIP-seq studies performed.

To test whether loss of SMARCA4/2 would alter the essentiality of other mSWI/SNF subunits for cell proliferation, we performed genome-scale CRISPR–Cas9 fitness screening in the BIN-67 cell line[15]. Using a statistical framework to call genetic interactions from cancer cell line fitness screening data[16], we found that mSWI/SNF complex members were significantly less essential in SMARCA4/2-dual-deficient cell lines (BIN-67, as well as COV434, OVK18, H1581, H23 cell lines) compared to mSWI/SNF-intact cell lines (*SMARCB1*, $P = 1.3 \times 10^{-3}$, *SMARCE1*, $P = 1.0 \times 10^{-3}$, Wilcoxon rank-sum test) (Fig. 1f, Supplementary Fig. 1h and Supplementary

[1]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. [2]Broad Institute of Harvard and MIT, Cambridge, MA, USA. [3]Biological and Biomedical Science, Harvard Medical School, Boston, MA, USA. [4]Harvard Medical School, Boston, MA, USA. [5]Molecular Pathology Unit, Massachusetts General Hospital, Charlestown, MA, USA. [6]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [7]Chemical Biology Program, Harvard Medical School, Boston, MA, USA. [8]Department of Biochemistry and Molecular Genetics, Northwestern Feinberg School of Medicine, Chicago, IL, USA. *e-mail: Cigall_kadoch@dfci.harvard.edu

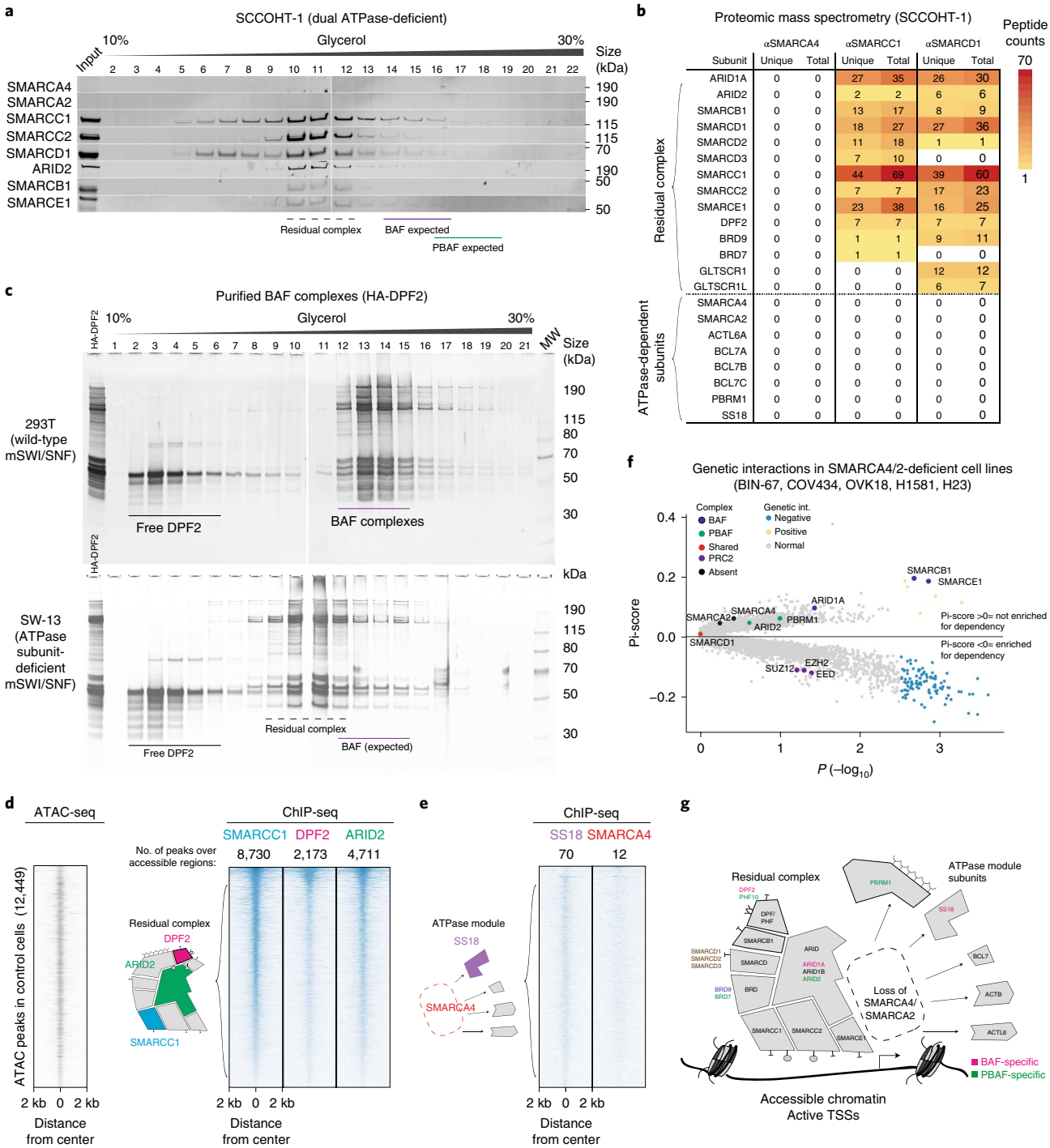

Fig. 1 | Biochemical and genomic characterization of a residual complex in SMARCA4/SMARCA2-dual-deficient cancer cell lines. a, Density sedimentation and immunoblot performed on SCCOHT-1 nuclear extracts. The residual mSWI/SNF complexes of both BAF and PBAF types exhibit similar elution profiles. See also Supplementary Fig. 7a. All molecular weights are reported in kilodaltons (kDa). **b**, Peptides corresponding to mSWI/SNF subunits identified in immunoprecipitation/mass spectrometry from SCCOHT-1 nuclear extract. SMARCC1 and SMARCD1 immunoprecipitations identified a cohort of subunits that remained stably bound to complexes in the absence of both ATPase subunits. **c**, Silver stain of density sedimented, endogenously purified BAF complexes from SMARCA4/2-intact (293T) and SMARCA4/2-deficient (SW-13) cell lines, using an HA-tagged DPF2 subunit as bait. **d**, Heatmaps of chromatin accessibility (ATAC-seq) and residual mSWI/SNF occupancy (ChIP-seq) in the BIN-67 cell line, treated with a control GFP vector. Each row of the heatmap corresponds to a peak from the union of ATAC-seq and SMARCC1 ChIP-seq peaks in this condition. Rows are rank ordered by SMARCC1 occupancy. Each column shows the normalized read density for different experimental conditions across a 4-kilobase (kb) window centered at each peak. **e**, Heatmaps reflecting SS18 and SMARCA4 ChIP-seq experiments performed in the BIN-67 cell line. **f**, Genetic interaction data derived from CRISPR–Cas9-based screens performed in SMARCA4/2- dual-deficient and SMARCA4/2-intact cell lines reported as a Pi-score, which measures the difference between the expected and observed fitness phenotypes upon knockout. Negative scores denote synthetic lethality, and positive scores denote alleviating interactions. Significance was calculated by a two-sided Wilcoxon rank-sum test on the SMARCA4/2-deficient lines ($n = 5$) against the WT lines ($n = 386$). The false discovery rate cutoff was 0.25 after Benjamini–Hochberg correction was used. **g**, Schematic of residual mSWI/SNF complex composition in the absence of the SMARCA4/2 ATPase subunits.

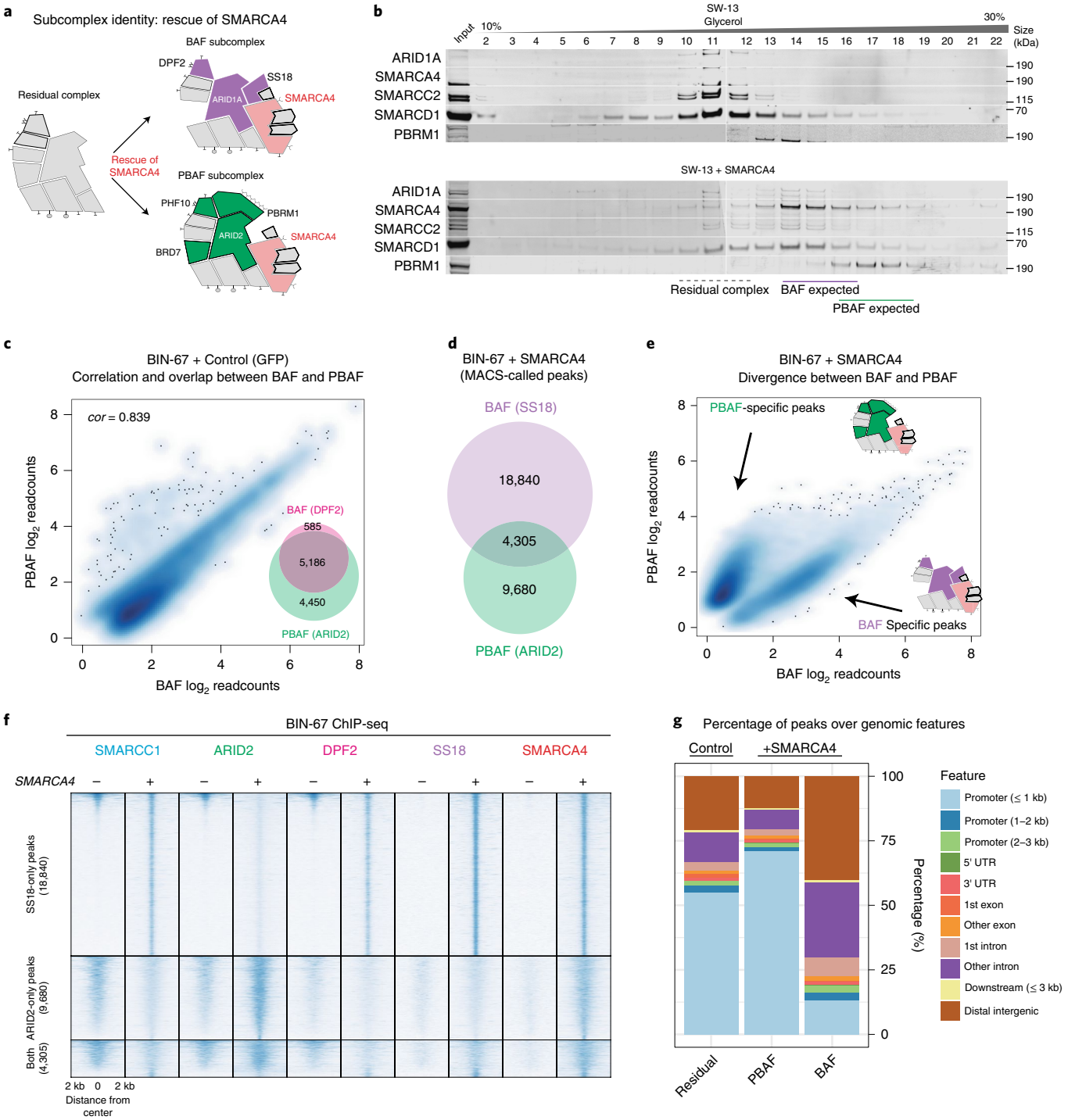

Fig. 2 | The ATPase module is required for BAF and PBAF subcomplex identity. a, Schematic of BAF and PBAF subcomplexes and their formation following rescue of SMARCA4 ATPase. **b**, Density sedimentation experiments performed on nuclear extracts of SW-13 cells with and without SMARCA4 rescue. Expected fraction numbers for BAF and PBAF subcomplexes are indicated. See also Supplementary Fig. 7b. **c**, Log-scale scatterplot showing the normalized read counts from DPF2 and ARID2 ChIP-seq across a union set of DPF2 and ARID2 peaks (inset). Spearman correlation (cor) is shown. **d**, Overlap between SS18 and ARID2 peaks in the SMARCA4 rescue condition in BIN-67 cells. **e**, Log-scale scatterplot showing the normalized read counts from SS18 and ARID2 ChIP-seq across a union set of SS18 and ARID2 peaks following SMARCA4 rescue. **f**. Heatmaps reflecting ChIP-seq occupancy for SMARCC1, ARID2, DPF2, SS18 and SMARCA4 in control and SMARCA4 rescue conditions, split between SS18-only, ARID2-only and shared occupancy. Rows are rank ordered by SMARCC1 occupancy in the control condition. **g**, Distribution of mSWI/SNF peaks over genomic features in control and SMARCA4 rescue conditions.

Table 2). We also recovered synthetic lethal relationships with PRC2 complex subunits, which have recently been reported[17,18] (Fig. 1f and Supplementary Fig. 1h). Taken together, these results suggest that mSWI/SNF complexes exhibit a modular biochemical organization, and that the absence of the ATPase module causes BAF- and PBAF-specific subunits to co-elute as similarly sized

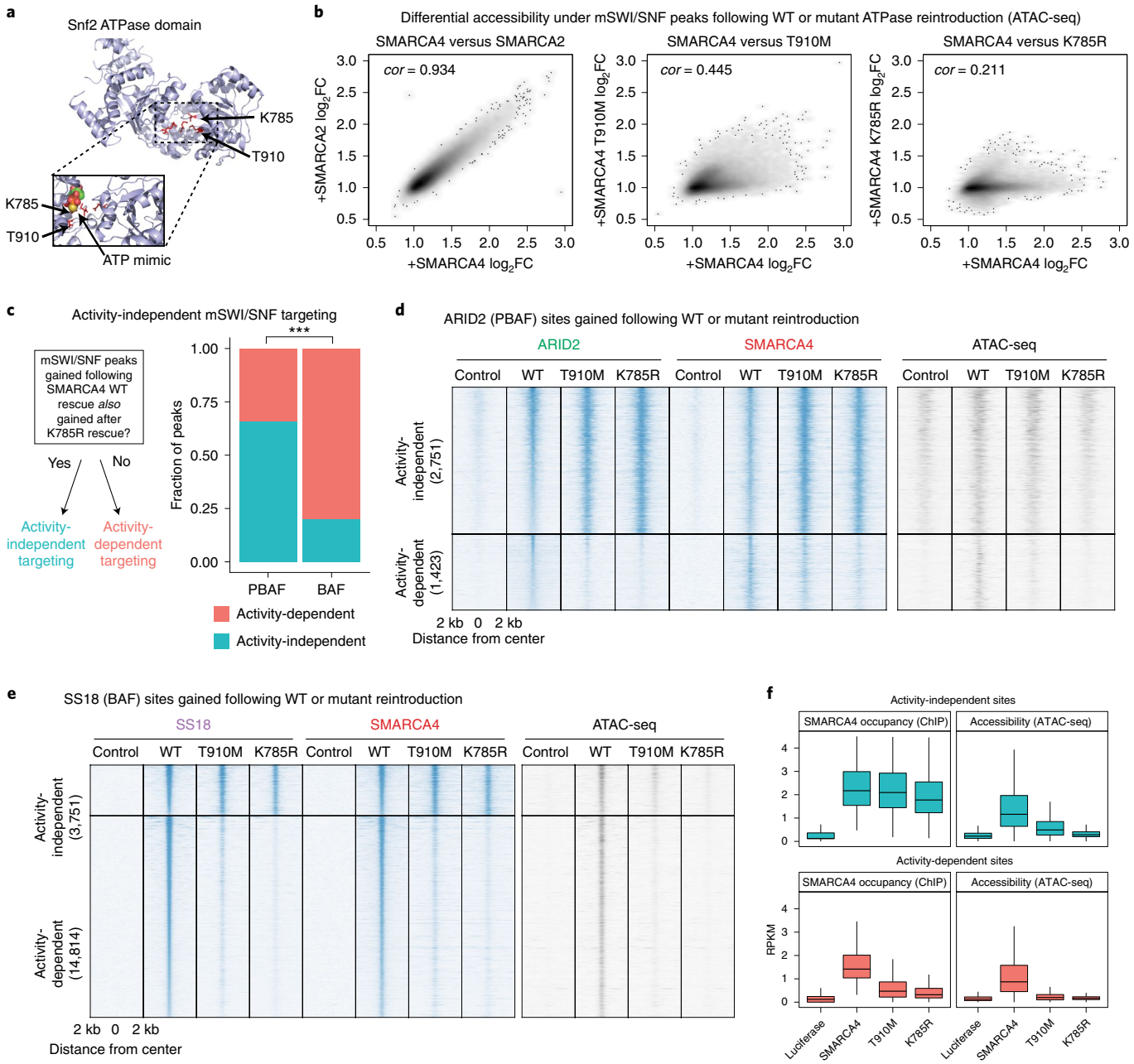

Fig. 3 | Rescue of a catalytic activity–deficient ATPase module is sufficient to localize PBAF and BAF complexes to a subset of target sites genome wide. a, Mapping of T910M and K785R mutant residues to the ATP-binding pocket of the yeast Swi2/Snf2 crystal structure (as performed in ref. [23]) shows proximity to a crystalized ATP mimic. Top: overall structure of the Swi2 domain, with mutated residues highlighted. Bottom: enhanced view of ATP-binding pocket in the presence of an ATP mimic. **b**, Scatterplots of differential accessibility under SMARCC1 peaks following rescue of different ATPase WT and mutant variants. Each point in the scatterplot corresponds to a SMARCC1 peak called in the BIN-67 SMARCA4 rescue condition. Each axis displays the fold change in accessibility under each peak (as measured by ATAC-seq reads per million mapped to that peak) as compared to the control condition. Spearman correlation is shown. **c**, Stacked bar plot of the proportion of BAF and PBAF peaks that are activity-dependent and activity-independent. Fisher's two-sided exact test performed on total ChIP-seq peaks ($n = 22{,}709$), $^{***}P < 1 \times 10^{-3}$. **d**, Heatmaps of chromatin accessibility and mSWI/SNF occupancy across ARID2 peaks gained following rescue of a SMARCA4 allelic series. The ARID2 peak set is split into two groups: activity-independent and activity-dependent peaks. Rows are rank ordered by ARID2 occupancy in the WT rescue condition. **e**, Heatmaps of chromatin accessibility and mSWI/SNF occupancy across SS18 peaks that were gained following rescue of a SMARCA4 allelic series. The SS18 peak set is split between activity-independent and activity-dependent peaks. Rows are rank ordered by SS18 occupancy in the WT rescue condition. **f**, Box plots showing normalized read counts across SS18 activity-independent ($n = 3{,}721$) and activity-dependent peaks ($n = 14{,}814$). First and third quartiles are denoted by lower and upper hinges. Whiskers extend to the smallest and largest values within 1.5× the interquartile range from the lower and upper hinges.

non-essential residual complexes that retain binding to accessible chromatin (Fig. 1g).

Given that the large majority (three-quarters) of the subunits are common to both BAF and PBAF subcomplexes, the mechanisms by which mSWI/SNF subunits form functionally differentiated assemblies remain unknown. The modularity of the mSWI/SNF complex we identified suggested the possibility that the residual complex, once loaded with subcomplex-specific subunits, could attach to an

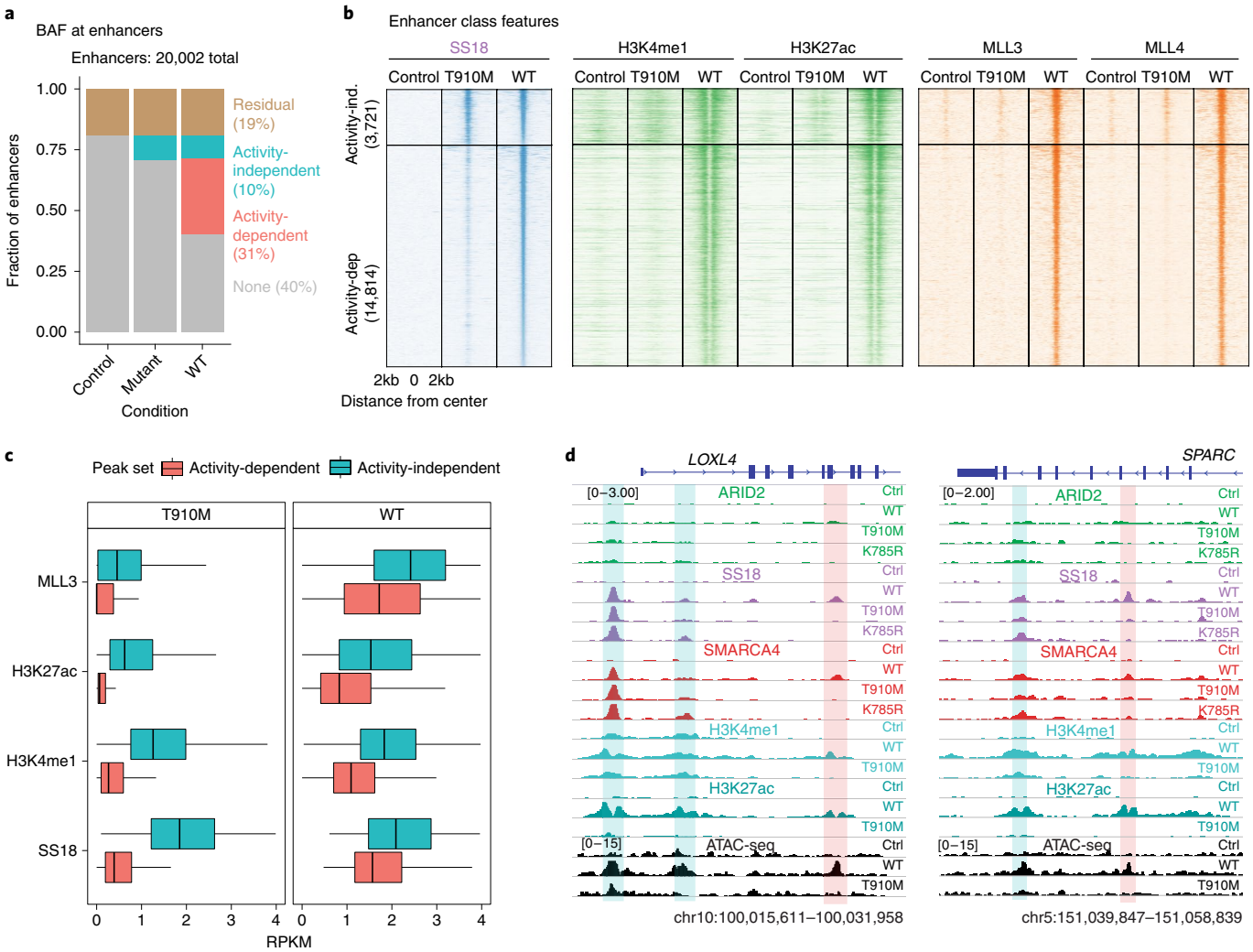

Fig. 4 | Defining activity-dependent and activity-independent targeting of BAF complexes at enhancers. a, Proportion of active enhancers occupied by BAF following control and ATPase rescues. Enhancers were defined as the overlap between H3K4me1 peaks and ATAC-seq peaks in the SMARCA4 WT condition. **b**, Heatmaps showing each gained SS18 site (ordered by SS18 occupancy) and the ChIP-seq profiles of H3K4me1, H3K27ac and MLL3/4. **c**, Box plot of ChIP-seq read counts for MLL3, H3K27ac, H3K4me1 and SS18 epitopes over activity-independent ($n=3,721$) and activity-dependent ($n=14,814$) peaks across T910M and WT rescues. First and third quartiles are denoted by lower and upper hinges. Whiskers extend to the smallest and largest values within 1.5× the interquartile range from the lower and upper hinges. **d**, Example browser track for the genes *LOXL4* and *SPARC*. Activity-independent sites are highlighted in blue while activity-dependent sites are shown in red.

ATPase module containing either SS18 or PBRM1 (known to be mutually exclusive in BAF and PBAF complexes, respectively[19,20]), enabling functional specificity (Fig. 2a). We confirmed that rescue of SMARCA4/2 in SW-13 cells sufficiently restored the protein level abundance of ATPase module subunits (Supplementary Fig. 2a) in a manner that was dose-dependent (Supplementary Fig. 2b) and correlated with the overall abundance of SMARCA4 (correlation in immunofluorescence signal, Spearman $P=0.860$, Supplementary Fig. 2c). The presence of SMARCA4 also restored interactions between ATPase module subunits and residual complex components (Supplementary Fig. 2d), including BAF-specific ARID1A-SS18 binding and PBAF-specific ARID2-PBRM1 binding (Supplementary Fig. 2e,f).

Further, we found that BAF and PBAF complexes required the ATPase module for their distinct biochemical size profiles as well as their genomic targeting patterns on chromatin. Restoration of the ATPase module in SW-13 cells resulted in clear size separation between BAF and PBAF complex subunits (Fig. 2b), implying complete assembly of both complexes. In BIN-67 control cells, genome-wide occupancies of BAF- and PBAF-specific subunits were strongly correlated (Spearman $P=0.839$, Fig. 2c). This suggested that the residual complexes lacked the distinguishing features of BAF or PBAF despite the presence of subcomplex-specific subunits (such as ARID1A/DPF2 for BAF or ARID2/BRD7 for PBAF). Restoration of the ATPase module increased chromatin affinity of both complexes (Supplementary Fig. 2g) and led to gains in the number of significant mSWI/SNF peaks (Supplementary Fig. 2h), which were predominantly non-overlapping between BAF and PBAF (Fig. 2d,e). This finding is consistent with those from previous mSWI/SNF subunit reintroduction studies in cancer cells[21] (although higher overlap between subcomplexes at promoters has been observed in mSWI/SNF-intact cell lines[22]). These results indicate that integration of the ATPase module is required for BAF and PBAF complexes to target divergent genomic loci (Fig. 2e,f), as demonstrated by the shift in PBAF localization toward TSSs and enrichment of BAF peaks distal to TSSs following SMARCA4 rescue (Fig. 2g).

We next investigated how the ATPase module mediates divergent BAF and PBAF complex localization. Rescue of the ATPase subunit

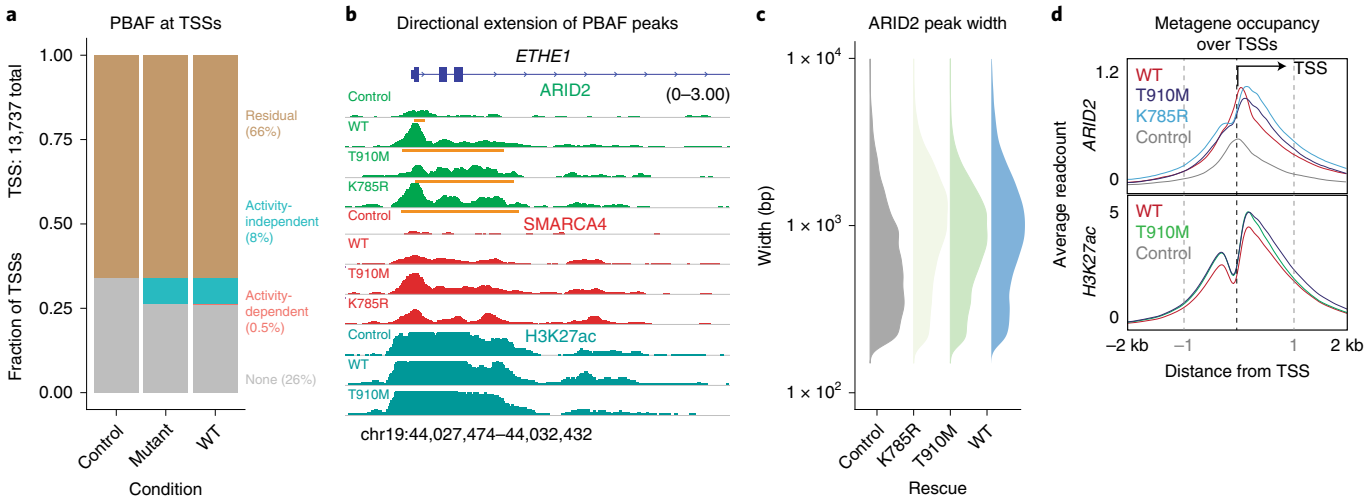

Fig. 5 | Directional PBAF complex occupancy along transcriptional initiation signatures requires the ATPase module, but not ATPase activity. **a**, Proportion of active TSSs occupied by PBAF (within 4 kb) following control and ATPase rescue. Active TSSs are defined as TSSs whose genes are expressed in the SMARCA4 WT rescue. **b**. Example browser track for ETHE1, whose TSS is occupied by a narrow ARID2 peak in the control condition and wide ARID2 peaks following ATPase subunit rescue. Broad peaks called by MACS 2.0 are shown in orange under ARID2 tracks. **c**, Peak width distributions of ARID2 peaks called following rescue of mutant and WT SMARCA4. bp, base pairs. **d**, Metagene plot of ARID2 and H3K27ac occupancy over TSSs.

introduces two concurrent changes to mSWI/SNF complexes: (1) restored biochemical composition of the complex and (2) restored ATPase catalytic activity. To distinguish the relative contributions of these two events on BAF and PBAF localization and function, we rescued BIN-67 cells with either WT or ATPase-mutant SMARCA4 (Fig. 3a)[23,24]. SMARCA4 T910M and SMARCA4 K785R mutants exhibited levels of protein expression similar to those of WT SMARCA4 (Supplementary Fig. 3a) and recapitulated WT complex biochemical composition (Supplementary Fig. 3b) as previously shown[24,25]. However, they exhibited either partial (T910M) or complete (K785R) losses of catalytic activity as assessed by in vitro ATPase assays (Supplementary Fig. 3c) and genome-wide chromatin accessibility profiles following expression BIN-67 cells (Fig. 3b). Expression of either mutant variant resulted in partially reduced affinity of mSWI/SNF complexes to chromatin as compared to expression of WT SMARCA4 (Supplementary Fig. 3d).

To assess whether gains in BAF and PBAF complex targeting can occur in the presence of the ATPase module but in the absence of ATPase activity, we mapped the localization of mSWI/SNF complexes by using ChIP-seq following rescue of ATPase-dead SMARCA4 K785R (Fig. 3c). Strong experimental agreement between SMARCA4 and SMARCC1 antibodies targeting mSWI/SNF complexes was detected in all reintroduction conditions (Supplementary Fig. 3e). Unexpectedly, rescue with SMARCA4 K785R in BIN-67 cells was sufficient to target mSWI/SNF complexes to a subset of sites that were gained following rescue of WT SMARCA4, and these sites were similar to the BAF- and PBAF-specific chromatin binding profiles observed with WT SMARCA4 rescue (Supplementary Fig. 3f).

From this observation, we defined catalytic activity–dependent sites as those that were gained following SMARCA4 WT rescue but not following SMARCA4 K785R rescue. We defined catalytic activity–independent sites as those that were gained in both rescue conditions (Fig. 3a). We further subdivided these peaks into BAF- and PBAF-specific sites by using the overlap between SMARCA4 and either SS18 (BAF) or ARID2 (PBAF) in the same condition, ensuring that each site represented high-confidence co-localization between at least two subunits of the same complex. We found that BAF and PBAF complexes exhibited significantly different proportions of activity-independent sites, with the majority of PBAF sites (~60%) but the minority of BAF sites (~20%) exhibiting activity-independent localization (Fisher's test $P < 2.2 \times 10^{-16}$, Fig. 3c). Over these BAF and PBAF activity-independent sites (Fig. 3d,e), both the SMARCA4 K785R and SMARCA4 T910M mutants exhibited genomic occupancy comparable to that of WT SMARCA4, although only the T910M mutant created partial chromatin accessibility under those sites, due to its partial ATPase activity (Fig. 3f). In contrast, at activity-dependent sites, only WT SMARCA4 was capable of generating mSWI/SNF de novo genomic occupancy and creating accessible sites (Fig. 3f).

We then examined the role of these sites at BAF complex-occupied enhancers. We found that 10% of all enhancers in BIN-67 cells were targeted by BAF in an activity-independent manner and an additional ~30% in an activity-dependent manner (Fig. 4a). We found that catalytically deficient BAF complexes containing SMARCA4 T910M could target sites that were enriched for H3K4me1 and, to a lesser extent, H3K27ac activating marks (Fig. 4b), suggesting that the ATPase module directly or indirectly recognizes pre-established enhancer elements. However, catalytically deficient BAF complexes could not target sites where these activating marks were absent (Fig. 4b). These sites are neither heterochromatic nor silenced, as they lacked the H3K9me3 or H3K27me3 mark, respectively (Supplementary Fig. 4a). At these sites, WT ATPase activity was essential for targeting, and substantial increases in H3K4me1 and H3K27ac marks as well as recruitment of MLL3/4 methyltransferases (Fig. 4c) occurred following remodeling (ATAC-seq data shown in Fig. 3e). In particular, because the SMARCA4 T910M mutant exhibited partial remodeling ability (Fig. 3b) yet did not localize to activity-dependent sites, we conclude that full remodeling activity is necessary for enhancer activation at these sites (Fig. 4c,d). These data suggest that previous models positing that H3K4me1 is required for BAF targeting are probably overly simplistic[26], because BAF complex ATPase activity following SMARCA4 rescue precedes the establishment of H3K4me1 by MLL3/4 in the majority of gained sites.

We also observed activity-independent effects for PBAF complexes following rescue of the ATPase module. At PBAF-occupied TSS regions (within 4 kb of the TSS, Fig. 5a), rescue of the ATPase module resulted in the directional extension of PBAF occupancy into gene bodies, for example at the *ETHE1* locus (Fig. 5b). This directional extension resulted in marked increases in genome-wide PBAF peak width and occurred irrespective of whether the

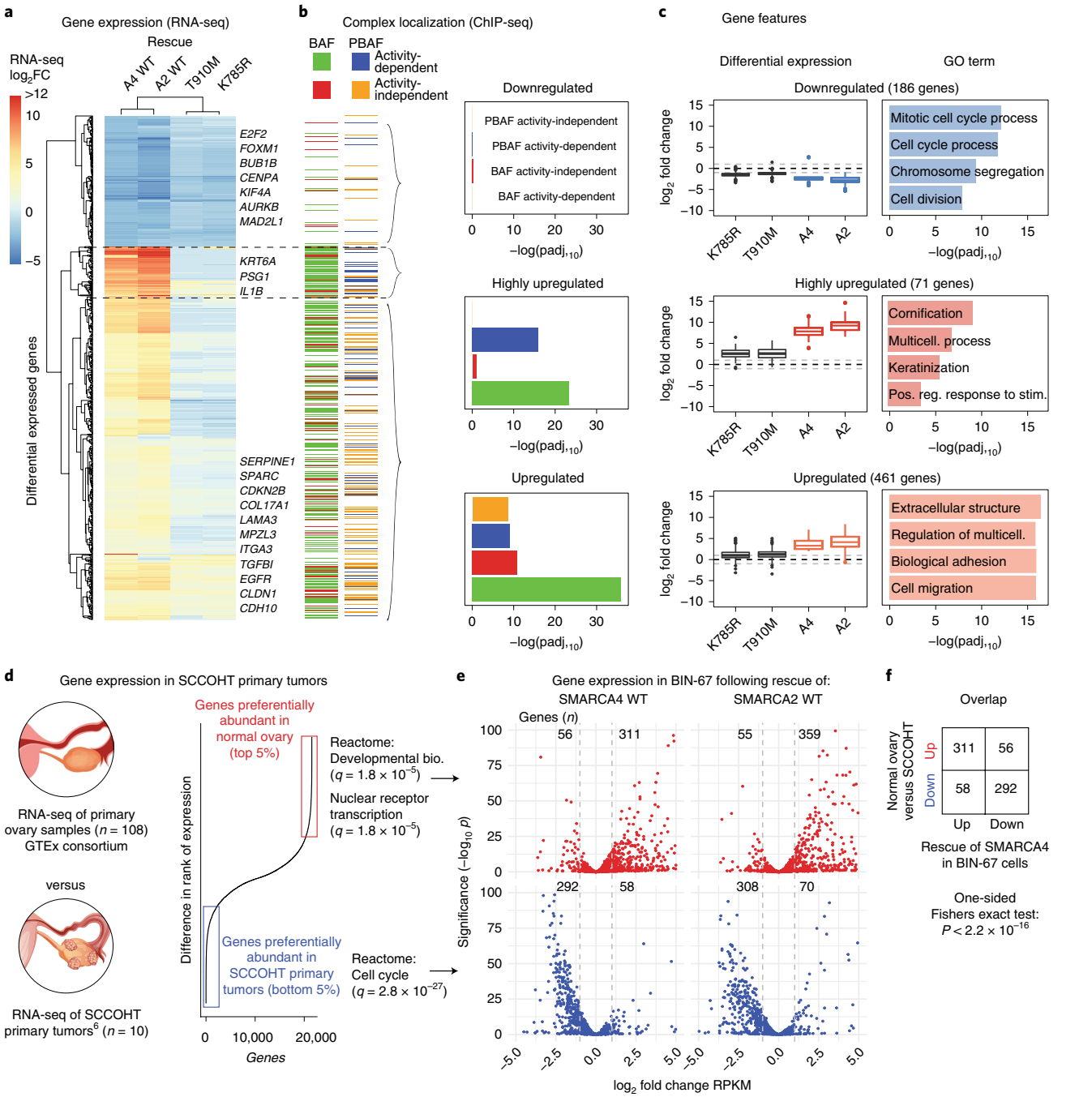

Fig. 6 | Catalytically active BAF and PBAF complexes collaborate to activate transcriptional programs that underlie differences between normal ovarian tissue and SCCOHT primary human tumors. a, Heatmap of differential expression of the 700+ significantly changed genes following SMARCA4 WT rescue, shown across all four rescue conditions in BIN-67 cells. Hierarchical clustering identified three gene clusters, with representative genes labeled. **b**, Left: heatmap showing the mSWI/SNF chromatin occupancy status of the genes in **a**. BAF and PBAF occupancies are shown separately, and activity-independent and activity-dependent peaks are indicated. Right: enrichment of each of the four peak groups in the three gene clusters using one-sided Fisher's exact test, corrected for multiple hypotheses with Benjamini–Hochberg correction (downregulated: $n = 186$; highly upregulated: $n = 71$; upregulated: $n = 461$). **c**, Left: box plots showing differential gene expression across each of the three clusters. First and third quartiles are denoted by lower and upper hinges. Whiskers extend to the smallest and largest values within 1.5× the interquartile range from the lower and upper hinges. Outliers beyond 1.5× the interquartile range are plotted as individual points. Gray dashed lines denote +1 or −1 logFC. Right: top four gene ontology (GO) enrichments for genes in these three clusters. Enrichment was calculated by hypergeometric testing after Benjamini–Hochberg false discovery rate correction by the GOrilla package. **d**, RNA-seq data from SCCOHT patient tumors were compared to primary ovarian samples from the GTEx consortium to generate two gene sets: genes that were preferentially abundant in normal ovarian tissue (red) and those that were preferentially abundant in SCCOHT tumor samples (blue). **e**, Top: scatterplot of BIN-67 differential expression data restricted to genes that were preferentially abundant in normal ovarian tissue. Bottom: scatterplot of BIN-67 differential expression data restricted to genes that were preferentially abundant in SCCOHT tumor samples. The Wald test followed by Benjamini–Hochberg false discovery rate correction was used to report significance (implemented by the DESeq2 package). The number of genes with $|\log_2 FC| > 1$ are annotated above each plot, where FC is the fold change. **f**, Restoration of SMARCA4 in BIN-67 cells causes gene expression changes that significantly overlap the changes seen between normal ovarian tissue and SCCOHT tumors (one-sided Fisher's exact test, $P < 2.2 \times 10^{-16}$).

ATPase was catalytically active, dead (K785R) or deficient (T910M) (Fig. 5c). Notably, the extent of PBAF spreading was concomitant with (and did not alter) the directional signatures of active transcription present at most promoters[27], high H3K27Ac (Fig. 5d) and depletion of H3K4me1 (Supplementary Fig. 4b). A comparatively small percentage (<1%, Fig. 5a) of TSS regions encompassed activity-dependent sites (Supplementary Fig. 4c). Taken together, these data suggest that the ATPase module on PBAF complexes enables recognition of transcriptional initiation signatures in a manner independent of catalytic activity.

Next, to assess the gene regulatory effects of mSWI/SNF catalytic activity at these TSS and enhancer elements, we performed RNA sequencing (RNA-seq) on BIN-67 cells in SMARCA4 WT and rescue conditions. In contrast to other enzymatic chromatin regulatory complexes such as MLL3/4 (ref. [28]), we found that the large majority of mSWI/SNF-driven gene expression changes in BIN-67 required mSWI/SNF enzymatic activity, with a sevenfold increase in the number of upregulated genes identified between WT SMARCA4/2 and SMARCA4 mutant rescue conditions (Fig. 6a, Supplementary Fig. 5a and Supplementary Table 3). The downregulated, upregulated and highly upregulated gene sets as defined by hierarchical clustering differed in their requirements for direct targeting by mSWI/SNF family complexes. Downregulated genes, which were predominantly involved in cell cycle and chromosome segregation (that is, *BUB1B*, *CENPA*, *E2F2*, *FOXM1*, et cetera), were not enriched for nearby BAF or PBAF complex localization relative to non-changing genes (Fig. 6b), suggesting that these gene changes were secondary downstream effects of SMARCA4 rescue. Highly upregulated genes involved in keratinization and other stimulus response genes (Fig. 6c and Supplementary Fig. 5b) were positively enriched for nearby BAF and PBAF complexes whose localization depended on their ATPase activity (Fig. 6b), implying that these genes are targets of de novo enhancers created by both complexes. Finally, we found that PBAF and BAF activity-dependent and activity-independent targeting collaborated in upregulation of the remaining upregulated genes (Fig. 6b), which were involved in senescence and extracellular structure, cell migration and epithelial–mesenchymal transition (Fig. 6c and Supplementary Fig. 5c,d).

Importantly, the gene expression changes following SMARCA4/2 rescue corroborate transcriptional signatures found in SCCOHT patient tumors. We analyzed tumor RNA-seq from SCCOHT primary tumors from a previous study[6] and compared these to transcriptional profiles from normal ovarian tissue[29] (Fig. 6d). We found that genes that were preferentially abundant in normal ovarian tissue highly overlapped with genes that were upregulated following the restoration of SMARCA4 or SMARCA2 to BIN-67 cells (Fig. 6e), and that genes that were preferentially abundant in SCCOHT tumors were downregulated following SMARCA4/2 rescue in BIN-67 cells (Fig. 6e). This overlap was statistically significant (Fig. 6f, one-sided Fisher's exact test, $P < 2.2 \times 10^{-16}$).

To summarize, by leveraging cell line models of the rare ovarian cancer SCCOHT, we have uncovered several principles underlying mSWI/SNF complex targeting and function. In the absence of the SMARCA4/2 subunits, mSWI/SNF complexes were reduced to a residual assembly with a baseline affinity to chromatin but lacking the distinguishing properties of BAF and PBAF subcomplexes. (Supplementary Fig. 6). Surprisingly, reintroduction of ATPase-deficient SMARCA4 restored mSWI/SNF localization to ~25% of its target sites and re-established differential localization of BAF and PBAF complexes to pre-marked enhancers and transcription initiation sites, respectively. ATPase-driven DNA accessibility is responsible for the remaining 75% of gained sites, particularly for BAF complexes, which also do not require pre-marked H3K4me1 for targeting to these sites. These ATPase activity-mediated events rewire the transcriptional state of SCCOHT cells toward tumor-suppressive gene expression programs found in normal ovarian tissue, providing a mechanistic context behind SMARCA4/2 dual loss in SCCOHT.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41588-019-0363-5.

Received: 8 August 2018; Accepted: 29 January 2019;
Published online: 11 March 2019

References

1. Kadoch, C. & Crabtree, G. R. Mammalian SWI/SNF chromatin remodeling complexes and cancer: mechanistic insights gained from human genomics. *Sci. Adv.* **1**, e1500447 (2015).
2. Karnezis, A. N. et al. Dual loss of the SWI/SNF complex ATPases SMARCA4/BRG1 and SMARCA2/BRM is highly sensitive and specific for small cell carcinoma of the ovary, hypercalcaemic type. *J. Pathol.* **238**, 389–400 (2016).
3. Jelinic, P. et al. Recurrent SMARCA4 mutations in small cell carcinoma of the ovary. *Nat. Genet.* **46**, 424–426 (2014).
4. Ramos, P. et al. Small cell carcinoma of the ovary, hypercalcemic type, displays frequent inactivating germline and somatic mutations in SMARCA4. *Nat. Genet.* **46**, 427–429 (2014).
5. Witkowski, L. et al. Germline and somatic SMARCA4 mutations characterize small cell carcinoma of the ovary, hypercalcemic type. *Nat. Genet.* **46**, 438–443 (2014).
6. Le Loarer, F. et al. SMARCA4 inactivation defines a group of undifferentiated thoracic malignancies transcriptionally related to BAF-deficient sarcomas. *Nat. Genet.* **47**, 1200–1205 (2015).
7. Karnezis, A. N. et al. Loss of switch/sucrose non-fermenting complex protein expression is associated with dedifferentiation in endometrial carcinomas. *Mod. Pathol.* **29**, 302–314 (2016).
8. Upchurch, K. S., Parker, L. M., Scully, R. E. & Krane, S. M. Differential cyclic AMP responses to calcitonin among human ovarian carcinoma cell lines: a calcitonin-responsive line derived from a rare tumor type. *J. Bone Miner. Res.* **1**, 299–304 (1986).
9. Gamwell, L. F. et al. Small cell ovarian carcinoma: genomic stability and responsiveness to therapeutics. *Orphanet J. Rare Dis.* **8**, 33 (2013).
10. Otte, A. et al. A tumor-derived population (SCCOHT-1) as cellular model for a small cell ovarian carcinoma of the hypercalcemic type. *Int. J. Oncol.* **41**, 765–775 (2012).
11. Leibovitz, A., McCombs, W. M., Johnston, D., McCoy, C. E. & Stinson, J. C. New human cancer cell culture lines. I. SW-13, small-cell carcinoma of the adrenal cortex. *J. Natl Cancer Inst.* **51**, 691–697 (1973).
12. Mashtalir, N. et al. Modular organization and assembly of SWI/SNF family chromatin remodeling complexes. *Cell* **175**, 1–17 (2018).
13. Wilson, B. G. et al. Residual complexes containing SMARCA2 (BRM) underlie the oncogenic drive of SMARCA4 (BRG1) mutation. *Mol. Cell Biol.* **34**, 1136–1144 (2014).
14. Hoffman, G. R. et al. Functional epigenetics approach identifies BRM/SMARCA2 as a critical synthetic lethal target in BRG1-deficient cancers. *Proc. Natl Acad. Sci. USA* **111**, 3128–3133 (2014).
15. Meyers, R. M. et al. Computational correction of copy-number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).
16. Rauscher, B. et al. Toward an integrated map of genetic interactions in cancer cells. *Mol. Syst. Biol.* **14**, e7656 (2018).
17. Chan-Penebre, E. et al. Selective killing of SMARCA2- and SMARCA4-deficient small cell carcinoma of the ovary, hypercalcemic type cells by inhibition of EZH2: in vitro and in vivo preclinical models. *Mol. Cancer Ther.* **16**, 850–860 (2017).
18. Wang, Y. et al. The histone methyltransferase EZH2 is a therapeutic target in small cell carcinoma of the ovary, hypercalcaemic type. *J. Pathol.* **242**, 371–383 (2017).
19. Middeljans, E. et al. SS18 together with animal-specific factors defines human BAF-Type SWI/SNF complexes. *PLoS ONE* **7**, e33834 (2012).
20. Kadoch, C. et al. Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat. Genet.* **45**, 592–601 (2013).
21. Nakayama, R. T. et al. SMARCB1 is required for widespread BAF complex-mediated activation of enhancers and bivalent promoters. *Nat. Genet.* **49**, 1613–1623 (2017).
22. Raab, J. R., Resnick, S. & Magnuson, T. Genome-wide transcriptional regulation mediated by biochemically distinct SWI/SNF complexes. *PLoS Genet.* **11**, e1005748 (2015).

23. Hodges, H. C. et al. Dominant-negative SMARCA4 mutants alter the accessibility landscape of tissue-unrestricted enhancers. *Nat. Struct. Mol. Biol.* **25**, 61–72 (2018).
24. Dykhuizen, E. C. et al. BAF complexes facilitate decatenation of DNA by topoisomerase IIα. *Nature* **497**, 624–627 (2013).
25. Bultman, S. J., Gebuhr, T. C. & Magnuson, T. A Brg1 mutation that uncouples ATPase activity from chromatin remodeling reveals an essential role for SWI/SNF-related complexes in beta-globin expression and erythroid development. *Genes Dev.* **19**, 2849–2861 (2005).
26. Local, A. et al. Identification of H3K4me1-associated proteins at mammalian enhancers. *Nat. Gen.* **50**, 73–82 (2018).
27. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77–88 (2007).
28. Dorighi, K. M. et al. Mll3 and Mll4 facilitate enhancer RNA synthesis and transcription from promoters independently of H3K4 monomethylation. *Mol. Cell* **66**, 568–5760000 (2017).
29. Consortium, G. Human genomics. the genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).

Acknowledgements

We thank members of the Kadoch Laboratory for helpful conceptual and experimental advice throughout the development of this study. We thank the DFCI Molecular Biology Core Facility, particularly Z. Herbert, for library preparation and sequencing, G. Boulay for advice regarding ChIP-seq optimization and the Taplin Mass Spectrometry Facility for mass-spectrometry analysis and data processing. We thank B. Vanderhyden (Ottawa Hospital Research Institute) and R. Hass (Hannover Medical School) for providing the BIN-67 and SCCOHT-1 cell lines, respectively. This work was supported in part by funding from the National Science Foundation Graduate Research Fellowship (No. 2015185722) and NIH T32 Training Grant in Genetics and Genomics to J.P.; the NIH DP2 New Innovator Award (No. 1DP2CA195762-01) to C.K.; the American Cancer Society Research Scholar Award (No. RSG-14-051-01-DMC) to C.K.; and the Pew–Stewart Scholars in Cancer Research Grant to C.K.

Author contributions

J.P and C.K conceived of and designed the study. J.P., Z.M.M, N.M. and R.S.P performed experiments. J.P., A.R.D. and C.A.L performed bioinformatic and statistical analyses. C.K. supervised the study. L.W. and A.S. provided novel in-house-generated MLL3/4 antibodies and experimental advice. J.P. and C.K. wrote the manuscript with editing by Z.M.M, A.R.D and N.M.

Competing interests

C.K. is a Scientific Founder, fiduciary Board of Directors member, Scientific Advisory Board member, consultant and shareholder of Foghorn Therapeutics, Inc.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-019-0363-5.

Correspondence and requests for materials should be addressed to C.K.

Methods

Cell lines. BIN-67 cell lines were grown in custom media (20% fetal bovine serum (FBS), 40% Dulbecco's modified Eagle's medium, 40% Dulbecco's modified Eagle's medium/Ham's F12). SCCOHT-1 cell lines were grown in Roswell Park Memorial Institute medium supplemented with 10% FBS. SW-13 and HEK-293T cell lines were grown in mouse embryonic fibroblast media (Dulbecco's modified Eagle's medium supplemented with GlutaMax, HEPES, sodium pyruvate and mouse embryonic fibroblast non-essential amino acids with 10% FBS). All media were supplemented with pen/strep.

Constructs. The pLEX307 constitutive overexpression backbone was used for all cloning and overexpression experiments. The WT SMARCA4 gene was subcloned from the Mammalian Gene Collection Human SMARCA4 Sequence-Verified complementary DNA purchased from GE Dharmacon (accession No. BC136644, clone No. 9020634). The WT SMARCA2 clone was subcloned from pBABE hBRM, a gift from R. Kingston (Addgene plasmid No. 1961). The K785R SMARCA4 mutant was subcloned from pBJ5 BRG1 DN, a gift from J. Crabtree (Addgene plasmid No. 17874). The T910M SMARCA4 mutant was created using overlap PCR performed on the WT SMARCA4 clone using a custom set of internal primers to introduce the T910M mutation, which was subcloned into a modified pTight (Clontech) and pLEX307 backbone. Oligonucleotides used for this protocol are included in Supplementary Table 4. A pLEX307 vector expressing green fluorescent protein was used as an overexpression control.

Lentiviral infection. The packaging vectors pspax2 and pMD2.G were co-transfected with pLEX307 vector containing the clone of interest into HEK-293T cells, using polyethylenimine as a transfection reagent. Cells were incubated for 72 h, and the media were filtered with a 0.4 µm filter before being either concentrated with an ultracentrifuge (20,000 r.p.m. for 2.5 h) or added directly to cells plated at 70% confluence with 1:1,000 of 10 mg ml^{-1} protamine sulfate.

Whole-cell lysates. Trypsinized cells (1×10^6) were washed twice in PBS and suspended in cold radioimmunoprecipitation assay buffer (10 mM Tris/Cl (pH 7.4), 1 mM EDTA (pH 7.4), 300 mM NaCl, 1% Triton X-100, 0.1% sodium dodecyl sulfate, 0.1% sodium deoxycholate, 1 mM dithiothreitol and 1 mM phenylmethanesulfonyl fluoride (PMSF) and rotated for 30 min at 4 °C.

Nuclear extracts. After trypsinization, cells were incubated in Buffer A (25 mM HEPES pH 7.6, 5 mM $MgCl_2$, 25 mM KCl, 0.05 mM EDTA, 10% glycerol and 0.1% NP-40 with protease inhibitor (Roche), 1 mM dithiothreitol (DTT) and 1 mM PMSF) for 10 min, and the pellets were resuspended in 600 µl of Buffer C (10 mM HEPES pH 7.6, 3 mM $MgCl_2$, 100 mM KCl, 0.5 mM EDTA and 10% glycerol with protease inhibitor, 1 mM DTT and 1 mM PMSF) with 67 µl of 3 M $(NH_4)_2SO_4$ for 20 min. The lysates were spun down using a tabletop ultracentrifuge at 100,000 r.p.m. at 4 °C for 10 min. Nuclear extracts were precipitated with 200 mg of $(NH_4)_2SO_4$ on ice for 20 min and finally purified as pellets by ultracentrifugation at 100,000 r.p.m. at 4 °C for 10 min. The pellets were resuspended in IP buffer (150 mM NaCl, 50 mM Tris-HCl pH 7.5, 1 mM EDTA and 1% Triton X-100 with protease inhibitor, 1 mM DTT and 1 mM PMSF) for subsequent experiments.

Density sedimentation gradient. Nuclear extract (800 µg, quantified by Bradford assay) was resuspended in 200 µl of 0% glycerol HEMG buffer (supplemented with protease inhibitors and DTT) and overlaid onto a 11 ml 10–30% glycerol (in HEMG buffer) gradient prepared in a 14 × 89 mm polyallomer centrifuge tube (Beckman Coulter, No. 331327). Tubes were centrifuged in an SW40 rotor at 4 °C for 16 h at 40,000 r.p.m. Fractions (0.550 ml) were collected and used in immunoblot analyses.

Immunoprecipitation. Immunoprecipitations were performed with 100 ug of nuclear extract resuspended in standard immunoprecipitation buffer (150 mM NaCl, 1 mM EDTA, 1% Triton X-100) and rotated overnight with 1.25 µg antibody. Antibodies to the following were used: SMARCA4 (Santa Cruz, No. G-7; Abcam, No. ab110641, used in Supplementary Fig. 2f); SMARCA2 (Bethyl, No. A301-015); SMARCC1 (Santa Cruz, No. H-76); SMARCD1 (Santa Cruz, No. 23); ARID2 (Santa Cruz, No. E-3); PBRM1 (EMD/Millipore, No. ABE70; Bethyl, No. A301-591A); ARID1A (Santa Cruz, No. C-7); SS18 (Cell Signaling Technologies, No. D6I4Z); and SMARCC2 (Santa Cruz, No. E-6), with mouse IgG (Santa Cruz, No. sc-2025) as a negative control.

Mass-spectrometry. For proteomic analysis, antibodies against SMARCA4, SMARCD1 and SMARCC1 were crosslinked with dimethyl pimelimidate to Protein G GammaBind Sepharose beads (GE) before immunoprecipitation from SCCOHT-1 nuclear extract. Bait protein was eluted with high salt and urea buffer, followed by trypsin digest and mass-spectrometry (Thermo Exactive Plus Orbitrap) as performed previously[12].

Purification of complexes. Complexes were purified as previously described, with few modifications[30]. SW-13 and 293 T cell lines stably expressing HA-DPF2 constructs were cultured in 150 mm dishes and expanded. Cells were scraped and washed with cold PBS. The suspension was centrifuged at 3,000 r.p.m. for 5 min at 4°C and pellets were resuspended in hypotonic buffer (HB) containing 10 mM Tris-HCl pH 7.5, 10 mM KCl, 1.5 mM $MgCl_2$, 1 mM DTT and 1 mM PMSF, and incubated on ice for 5 min. Nuclear pellets were resuspended in high salt buffer containing 50 mM Tris-HCl pH 7.5, 300 mM KCl, 1 mM $MgCl_2$, 1 mM EDTA, 1 mM, 1% NP-40, 1 mM DTT, 1 mM PMSF and protease inhibitor cocktail. The homogenate was incubated on a rotator for 1 h. Homogenates then were centrifuged at 20,000 r.p.m. (30,000 *g*) for 1 h at 4 °C using a SW32Ti rotor. The nuclear extract was filtered through a 0.45 µm filter and incubated overnight with anti-HA beads (Pierce, 88836) before elution with HA peptide. Eluted material was then subjected to density gradient centrifugation.

Immunoblot. Protein was loaded onto Bis-Tris 4–12% gradient Novex gels and run for 150 V for 90 min. A wet transfer was performed for 2.5 h at 165 mA at 4 °C onto polyvinylidene difluoride membrane. After transfer, membranes were blocked in milk for 1 h at room temperature before applying primary and fluorescent secondary antibodies (goat anti-mouse IgG antibody, IRDye 680RD conjugated, LICOR and goat anti-rabbit IgG antibody, IRDye 800CW conjugated, LICOR) for visualization on a LICOR Odyssey. Antibodies to the following were used: SMARCA4 (Santa Cruz, No. G-7); SMARCA2 (Bethyl, No. A301-015); SMARCC1 (Santa Cruz, No. H-76); SMARCC2 (Santa Cruz, No. E-6); SMARCD1 (Santa Cruz, No. 23); ARID2 (Santa Cruz, No. E-3); PBRM1 (EMD/Millipore, No. ABE70); ARID1A (Santa Cruz, No. C-7;) SS18 (Cell Signaling Technologies, No. D6I4Z); BRD7 (Santa Cruz, No. B-8); TBP (Abcam, No. 5184); SMARCB1 (Santa Cruz, No. A-5); GAPDH (Santa Cruz, No. G-9); ACTL6A (Santa Cruz, No. E-3); and CTCF (Santa Cruz, B-5).

Immunofluorescence. SW-13 cells expressing SMARCA4 after lentiviral treatment at low multiplicity of infection were split to 50% confluency on a 24-well plate. Cells were washed with PBS, incubated in 3% paraformaldehyde in phosphate buffered saline for 20 min at room temperature, then washed twice with IF wash buffer (0.1% NP-40, 1 mM sodium azide, 1× PBS) and blocked overnight in blocking buffer (IF wash buffer + 10% FBS). Antibodies against SMARCA4 (Santa Cruz, No. G-7) and SS18 (Cell Signaling Technologies, No. D6I4Z) were diluted into the blocking buffer and incubated for 3 h at room temperature. Cells were then washed three times with IF wash buffer and incubated with secondary antibody for 1 h (goat anti-rabbit IgG Alexa Fluor 546, Cell Signaling Technologies, No. A-11010; goat anti-mouse IgG Alexa Fluor Plus 555, Cell Signaling Technologies, No. A32727). Slides were mounted with ProLong Gold Antifade Mountant with DAPI (Thermo, No. P36935).

Senescence staining. BIN-67 cells treated with SMARCA4 and SMARCA2 lentivirus for two weeks post selection were fixed and stained as described in the Senescence β-Galactosidase Staining Kit (Cell Signaling Technologies) alongside a naïve control.

Generation of knockout HEK-293T cell lines. CRISPR–Cas9 KO constructs targeting SMARCA4 and SMARCA2 were purchased from Santa Cruz and co-transfected into HEK-293T cells by using the reagent Lipofectamine 3000. Cells were selected with puromycin at 2 µg ml^{-1} for 5 d. Single-cell clones were selected and screened for loss of subunit expression by using immunoblotting.

ATP consumption assay. Constructs expressing SMARCA4 WT, K785R, T910M and an empty vector control were transiently expressed in naïve 293 T cells for 72 h. Cells were trypsinized and pooled in batches of 50 million cells each, followed by PBS wash. Cells were treated with nuclear extract buffer A (25 mM HEPES pH 8.0, 5 mM $MgCl_2$, 25 mM KCl, 0.05 mM EDTA, 10% glycerol, 0.1% NP-40) and resuspended in 200–300 µl of lysis buffer (50 mM Tris pH 8.0, 150 mM NaCl, 0.1% NP-40) with protease inhibitor and incubated in a cold room with constant rotation for 30 min. Cell debris was discarded by ultracentrifuging at 100,000 r.p.m. for 5 min. The supernatant was collected, quantified and used for the subsequent overnight immunoprecipitation step in a cold room with 30 µl of V5 bead slurry and constant rotation.

On the following day, immunoprecipitations were washed three times with 1 ml of wash buffer (10 mM Tris, pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$) and resuspended in 25 µl of complete 1× ATPase Assay Reaction Buffer (10 mM Tris pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$, 20% glycerol) with reaction reagents (0.1 mg ml^{-1} BSA, 1 mM DTT, 4 mM ATP, 0.1–0.5 µg µl^{-1} DNA) and transferred to PCR tubes. The consumption reaction was incubated at 37 C for 1 h. Following ATP hydrolysis, beads were separated from the assay and 20 µl of ATPase reaction was transferred to an opaque, white 384-well plate.ADP-Glo reagent (20 µl) was added and incubated at room temp for 40 min. Then, 4 µl of ADP-Glo detection substrate was added to the reaction and incubated for 1 h. Luminescence was recorded on SpectraMax with an integration time of 0.25–1.0 s well^{-1}.

Differential salt extraction. Differential salt extraction was performed as previously described[21].

Genetic interactions. To call SMARCA4/2 dual loss genetic interactions from screening data, we utilized the MINGLE framework[16] with a few modifications. CRISPR screening data and cell line data were downloaded from the Depmap Public 18Q2 dataset (https://depmap.org/portal/download/all/). Cell lines were

designated as SMARCA4/2-dual-deficient from literature references (BIN-67, COV434, OVK18)[17] or if the cell line exhibited low gene expression of SMARCA2 (reads per kilobase of transcript, per million mapped reads (RPKM) < 2), exhibited biallelic damaging mutations on SMARCA4 and did not exhibit a strong dependency on SMARCA4 or SMARCA4 (H1581 or H23). Differential dependency was calculated directly from CERES scores, significance was assessed using a Wilcoxon rank-sum test and *P* values were adjusted for multiple comparisons with Benjamini–Hochberg correction. Differential dependencies with false discovery rate <0.25 were reported as significant. Pi scores and graphical output were performed as in the MINGLE pipeline.

ChIP-seq. We performed mSWI/SNF chromatin immunoprecipitation using a modified version of a protocol previously described[21,31,32]. In brief, cells were fixed in 1% formaldehyde at 37 °C for 10 min, quenched in 2.5 M glycine and snap frozen. Cells were thawed and lysed, and then the prepared nuclei were placed in a Covaris miliTube and sonicated for 20 min on a Covaris AFA Sonicator at standard settings (140 peak incident power(PIP), 5 W, 10% duty factor). The equivalent of 6–10 million cells were used in each immunoprecipitation. Three micrograms of antibodies to the following were used: DPF2 (Abcam, No. ab134942); SMARCC1 (homemade rabbit antibody raised against amino acids 998–1073 of human protein); SS18 (Cell Signaling Technologies, No. D6I4Z); ARID2 (Cell Signaling Technologies, No. D8D8U); H3K4me1 (Abcam, No. ab176877); H3K27ac (Abcam, No. ab4729); H3K27me3 (Cell Signaling Technologies, No. 9733 S); H3K9me3 (Abcam, No. ab8898); and MLL3/4 (homemade[33]). For SMARCA4 IP, 4 µl of the antibody was used (Abcam, No. ab110641). Immunoprecipitates were washed in 150 and 500 mM NaCl wash buffers as well as a lithium chloride wash. Protein-DNA fragments were eluted using an sodium dodecyl sulfate/DTT buffer and reverse crosslinked overnight. DNA was captured with SPRI beads (Agilent). Rubicon library prep was used to generate libraries, which were sequenced on an Illumina NextSeq 500. ChIP-seq immunoprecipitations for each antibody were performed in singlicate across each of the rescue conditions; overlaps of antibodies targeting two distinct epitopes or subunits were used for analyses.

ChIP analysis. For alignment of ChIP-Seq data, Bowtie2, version 2.1.0 was used to map reads to the hg19 human reference genome, using the parameter –k 1 to search for at most one distinct alignment. All downstream analysis was performed on bam files, with duplicates removed using the samtools rmdup command.

Peaks were called using MACS 2.0 at a *q* value cutoff of 0.001 using the narrow peak setting, with the exception of ARID2 peaks,which were called using the broad peak setting. Peaks that fell within ENCODE blacklisted regions or were mapped to unmappable chromosomes (not chr1-22, X or Y) were removed. ChIP-seq track densities were generated per million mapped reads with MACS 2.0 using the –B –SPMR options, sorted with the bedtools sort command, and converted to bigwig files using the bedGraphToBigWig tool from ENCODE. Tracks were visualized using the Integrative Genomics Viewer (IGV). Overlaps for ChIP Venn diagrams were created using the ChIPPeakAnno bioconductor package.

ChIP peak sets. All peak sets used in the paper were derived from at least two complementary ChIP-seq experiments (using antibodies designed to two distinct subunits on mSWI/SNF complexes) from the same experimental condition.

The residual complex peak set was defined as the union between peaks identified in BIN-67 + control rescue (any of peaks ARID2, DPF2 or SMARCC1). Both control DPF2 and ARID2 peaks analyzed in Fig. 2 were intersected with SMARCC1 peaks from the same experimental condition. Both SS18 and PBAF peaks following rescue of SMARCA4 in Fig. 2 were intersected with SMARCA4 peaks from the same experimental condition. Intersections were performed by using pybedtools with SS18 or ARID2 considered the 'A' peak in the comparison and the -wa option set to True.

BAF activity-independent peaks were defined as SS18 peaks (intersected with SMARCA4 peaks as noted above) that were also found in the SS18 ChIP performed in SMARCA4 K785R rescue. The remaining SS18 peaks were annotated as activity-dependent. PBAF activity-independent peaks were similarly defined, but with ARID2 rather than SS18. Enhancers (Fig. 4a) were defined as H3K4me1 peaks in the WT SMARCA4 rescue condition that overlapped ATAC-seq peaks in the same condition. TSSs were defined by using annotables 0.1.91 grch37, and active TSSs (used in Fig. 5a) were defined as those whose transcripts had RPKM > 2 in the SMARCA4 WT rescue in BIN-67 cells.

ChIP occupancy scatterplots. Read counts across peak sets of interest were calculated by calling Rsubread featureCounts() on duplicate removed bam files. These values were divided by the total number of mapped reads divided by one million, giving a normalized value of reads per million mapped reads for each interval in the input bed. Scatterplots were plotted with the R function smoothScatter.

ChIP feature distributions. Peak sets in Fig. 2g were analyzed with the ChIPSeeker package using default settings[34].

ChIP heatmaps and metagene plots. Heatmaps of ChIP occupancy were generated created using HTSeq. To account for the 200 base pairs (bp) average fragment length selected for in sonication, fragment length was extended by 200 bp from the edge of each genomic interval. Total read counts for each interval were normalized to reads per million mapped reads. For each antibody the resulting matrix has a width of 4 kb and the number of rows corresponding to the number of peaks in the indicated set. Strandedness of the interval was not considered except for the directional TSS metagene plots, in which the average readcount per million mapped reads was plotted for each sample to construct an average profile over TSSs (Fig. 5d and Supplementary Fig. 4b). Heatmaps were visualized using matplotlib in Python. Heatmaps were ordered by the maximum value in each matrix row of the indicated antibody.

ATAC-seq. ATAC-seq was performed as described previously[35]. In particular, 50,000 cells were harvested for each condition as suggested. Libraries were prepared using the suggested barcodes and sequenced on a paired-end Illumina NextSeq 500, with four samples on a single paired-end lane. All rescue and control experiments were performed in duplicate, and sequencing data were aligned as previously described[36]. In detail, ATAC-seq reads were paired and trimmed to 30 bp with Trimmomatic 0.36, then aligned to the hg19 genome with Bowtie2 v2.2.9 with the -X parameter set to 2000. Picard MarkDuplicates was used remove duplicate reads, and then bam files were further filtered by using the samtools view function with settings -F 256 -f 2 -q 30. Reads in ENCODE blacklisted regions were removed by using bedtools v2.26 intersect with the –v parameter True. Replicate bam files were merged by using Samtools v0.1.19 merge. These merged bam files were used for all downstream analysis. Bedgraph files were generated with the BedTools function genomeCoverageBed with the –pc option true. These bedgraphs were converted to BigWig format by using the bedGraphToBigWig script downloaded from UCSC. ATAC peaks were called by using MACS2 callpeak with a *q* value cutoff of 1×10^{-3} and the paramaters --nomodel --shift -150 --extsize 200. Peaks that fell into unmappable chromosomes were removed. All rescue and control experiments were performed in duplicate, and sequencing data merged into a single file for processing.

RNA-seq. One million cells of each condition were harvested using the Qiagen RNeasy kit, then 1 µg was DNase treated and used in the Illumina TruSeq Stranded mRNA library prep kit. All samples were validated as having an RNA integrity number (RIN) > 0.6 on a BioAnalyzer. Libraries were sequenced on an Illumina NextSeq 500. All rescue conditions were performed in replicate, with the SMARCA4 rescue performed in triplicate.

After sequencing, RNA-seq data reads were mapped using default parameters to hg19 using STAR version 2.5.2a. Differential genes were called using DESeq2. RPKM values were calculated using GFOLD version 1.1.4. Unless otherwise noted, $\log_2$ fold change and Bonferri-corrected *P* values were generated using DESEQ2, with reads mapped using Rsubread. Genes were considered to have changed significantly if they had an adjusted $P < 0.05$ and a $\log_2$ fold change of at least 2.

Clustering of differentially expressed genes. The $\log_2$ fold change of differentially expressed genes in the SMARCA4 rescue condition was plotted with the pheatmap R package across all four experimental conditions tested (SMARCA4 K785R, SMARCA4 T910M, SMARCA4 WT and SMARCA2 WT). Clustering was performed with complete linkage clustering and Euclidian distance, and a cluster number of 3 was chosen for further analysis.

Gene Ontology enrichment. Gene Ontology term enrichment was performed with the GOrilla web tool (http://cbl-gorilla.cs.technion.ac.il)[37], using each of the three differentially expressed gene clusters as target gene lists, and the set of expressed genes among BIN-67 RNA-seq experiments (RPKM > 2, resulting in 13,189 genes) as background.

Target gene assignment and enrichment. Genes were annotated as targets of BAF regulation if there was at least one BAF peak within 30 kb of the TSS of that gene. Genes were annotated as PBAF targets in the same fashion. Target genes were split between activity-dependent and activity-independent groups based on which peak set was more abundant within the 30-kb window around the TSS.

For each of the three differentially regulated gene clusters, over-representation of BAF and PBAF target genes was assessed with a one-sided Fisher's exact test against the background of all expressed genes (13,189 in total).

Gene set enrichment analysis of target genes. Gene set enrichment analysis was performed with GSEA software[38] on the MSigDB C2 gene sets using differential expression data following the reintroduction of SMARCA4 WT. Gene sets that were enriched for upregulated genes with a family-wise error rate $P < 0.01$ were used for downstream analysis. Over-representation of BAF activity-dependent or activity-independent target genes among upregulated genes within these gene sets was calculated with a one-sided Fisher's exact test against the background of all upregulated genes (533 in total).

Primary tumor and normal ovarian data. Raw RNA abundance counts for ten primary SCCOHT tumors (SRA SRP052896) and 108 primary ovary samples (GTEx) were computed using the recount2 pipeline[39] (PMID: 28398307). MicroRNA and small non-coding RNA transcripts were removed from RNA-seq for SCCOHT tumors and wild type ovary tissue. Due to considerable batch differences

between the two samples, we used a more conservative non-parametric rank-based method to identify differentially expressed genes. Transcripts were ranked by using a mean rank across all tumor or wild-type samples, which was calculated for each transcript. The difference in mean rank was used to determine differentially expressed transcripts between the samples. Specifically, the top and bottom 5% of differentially ranked genes between conditions were used for further analysis.

Statistics and reproducibility. The following panels are representative of experiments repeated the following number of times:

Figure 1a—duplicate
Figure 1c—DPF2, 10+ replicates, SW-13, triplicate
Figure 2b—duplicate of both conditions
Supplementary Fig. 1a—10+ replicates
Supplementary Fig. 1b—duplicate
Supplementary Fig. 1c—duplicate
Supplementary Fig. 1d—singlicate
Supplementary Fig. 1e—SW-13 triplicate, DPF2 10+ replicates, SMARCD1 10+ replicates
Supplementary Fig. 2a—singlicate
Supplementary Fig. 2b—singlicate
Supplementary Fig. 2d—duplicate (second shown in Supplementary Fig. 1f)
Supplementary Fig. 2e—duplicate
Supplementary Fig. 2f—duplicate with two different antibodies (both shown on same blot)
Supplementary Fig. 2g—duplicate
Supplementary Fig. 3a—5+ replicates
Supplementary Fig. 3b—singlicate
Supplementary Fig. 3d—singlicate
Supplementary Fig. 5c—representative of three images from same experiment

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Code availability

Code to generate figures is available at https://github.com/joshbiology/sccoht/.

Data availability

All sequencing data are available at GEO: GSE117735. All peaks called from raw data can be found at https://figshare.com/s/d74eccb73f20af21a6da. All custom-defined peak sets can be found at https://figshare.com/s/00ac067cf47edea9805d.

References

30. Mashtalir, N. et al. Autodeubiquitination protects the tumor suppressor bap1 from cytoplasmic sequestration mediated by the atypical ubiquitin ligase UBE2O. *Mol. Cell* **54**, 392–406 (2014).
31. Takaku, M. et al. GATA3-dependent cellular reprogramming requires activation-domain dependent recruitment of a chromatin remodeler. *Genome Biol.* **17**, 36 (2016).
32. McBride, M. J. et al. The SS18-SSX fusion oncoprotein hijacks BAF complex targeting and function to drive synovial sarcoma. *Cancer Cell* **33**, 1128–1141.e7 (2018).
33. Wang, L. et al. Resetting the epigenetic balance of Polycomb and COMPASS function at enhancers for cancer therapy. *Nat. Med.* **24**, 758–769 (2018).
34. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).
35. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 9 (2015).
36. Denny, S. K. et al. Nfib promotes metastasis through a widespread increase in chromatin accessibility. *Cell* **166**, 328–342 (2016).
37. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).
38. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
39. Collado-Torres, L. et al. Reproducible RNA-seq analysis using recount2. *Nat. Biotechnol.* **35**, 319–321 (2017).

nature research

Corresponding author(s): Cigall Kadoch

Last updated by author(s): 1-22-2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used.
Data analysis	R 3.3.3 MINGLE pipeline (https://github.com/boutroslab/Supplemental-Material/tree/master/Rauscher_2017), Bowtie2 version 2.1.0, MACS 2.0, STAR v. 2.5.2a, pybedtools, BedTools v2.26.0, SamTools v0.1.19 DESeq2, GOrilla web tool, GSEA software, htseq and matplotlib (Python) for heatmaps Custom code to generate figures are available upon request, and a public repository will be included at https://github.com/joshbiology/sccoht.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The sequencing data sets generated during the current study have been deposited in the Gene Expression Omnibus (GEO) repository under accession number GSE117735 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE117735).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	RNA-seq studies were performed at least in duplicate for use with the DESeq2 pipeline. ATAC-seq studies were performed in duplicate and merged. Due to cost, individual antibody epitopes (SMARCA4, SS18 etc) were performed in singlicate per condition, and overlapping peaks taken between these two epitopes to define a high-quality BAF binding dataset. The same approach was taken to PBAF (SMARCA4 and ARID2). Biochemical figures were generated with a variety of replicates as reported in the Statistics and Reproducibility section of our manuscript.
Data exclusions	No data exclusions were performed.
Replication	ATPase module composition was confirmed in four cell lines (SCCOHT-1, BIN-67, SW-13, 293T double SMARCA4/2 knockout). SMARCA4 and SMARCA2 paralogs were used for rescue experiments in ATAC-seq and RNA-seq experiments, showing similar results. All RNA-seq and ATAC-seq was performed in duplicate, and peaks for ChIP-seq (performed in singlicate because of cost) were discarded unless they appeared as significant in multiple distinct epitopes (eg. SS18 and SMARCA4).
Randomization	Randomization was not relevant to our study; our techniques were all molecular biological techniques where the experimenter designs and executes the experimental conditions so randomization is not possible.
Blinding	Blinding was not relevant to our study; our techniques were all molecular biological techniques where the experimenter designs and executes the experimental conditions so blinding is not possible.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☐ Flow cytometry
☐	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Immunoprecipitation: SMARCA4 (Santa Cruz, G-7); SMARCA2, (Bethyl A301-015); SMARCC1 (Santa Cruz, H-76); SMARCD1 (Santa Cruz, 23); ARID2 (Santa Cruz, E-3); PBRM1 (EMD/Millipore, ABE70, Bethyl, A301-591A); ARID1A (Santa Cruz, C1), SS18 (Cell Signalling Technologies, D6I4Z), SMARCC2 (Santa Cruz, E-6), and mouse IgG (Santa Cruz, sc-2025) Western blot: Primary: SMARCA4 (Santa Cruz, G-7); SMARCA2, (Bethyl A301-015); SMARCC1 (Santa Cruz, H-76); SMARCC2 (Santa Cruz, E-6); SMARCD1 (Santa Cruz, 23); ARID2 (Santa Cruz, E-3); PBRM1 (EMD/Millipore, ABE70); ARID1A (Santa Cruz, C7), SS18

(Cell Signaling Technologies, D6I4Z), BRD7 (Santa Cruz, B-8), TBP (Abcam, 5184), SMARCB1 (Santa Cruz, A-5), GAPDH (Santa Cruz, G-9), ACTL6A (Santa Cruz, E-3), and CTCF (EMD/Millipore, 07-729). Secondary: Goat Anti-Mouse IgG Antibody, IRDye® 680RD Conjugated, LICOR, Goat Anti-Rabbit IgG Antibody, IRDye® 800CW Conjugated, LICOR
ChIP: SMARCA4 (Abcam ab110641); DPF2 (Abcam, ab134942); SMARCC1 (homemade rabbit antibody raised against amino acids 998-1073 of human protein); SS18 (Cell Signaling Technologies D6I4Z); ARID2 (Cell Signaling Technologies D8D8U); H3K4me1 (Abcam, ab176877); H3K27ac (Abcam, ab4729); H3K27me3 (Cell Signaling Technologies 9733S); H3K9me3 (Abcam, ab8898), MLL3/4 (homemade)
IF: Goat anti-Rabbit IgG Alexa Fluor 546, Cell Signaling Technologies A-11010; Goat anti-Mouse IgG Alexa Fluor Plus 555, Cell Signaling Technologies A32727

Validation

All antibodies against mSWI/SNF subunits have been validated by our lab to recognize targets specifically (using CRISPR-Cas9-mediated knock-out or shRNA-mediated knock-down approaches in cell lines).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

293T and SW-13 were purchased from ATCC. BIN-67 was a gift from Barbara Vanderhyden and SCCOHT-1 was a gift from Ralf Hass.

Authentication

None

Mycoplasma contamination

All cell lines tested neg. for mycoplasma.

Commonly misidentified lines
(See ICLAC register)

None

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.

All sequencing data is available at GEO: GSE117735.
Raw peaks can be found at: https://figshare.com/s/d74eccb73f20af21a6da
Derived peaksets for each complex can be found here: https://figshare.com/s/00ac067cf47edea9805d

Files in database submission

BIN-67_BRG1_Luciferase_ChIP-Seq_Rep1
BIN-67_Input_Luciferase_ChIP-Seq
BIN-67_BRG1_SMARCA4_ChIP-Seq
BIN-67_BAF155_SMARCA4_ChIP-Seq
BIN-67_ARID2_SMARCA4_ChIP-Seq
BIN-67_SS18_SMARCA4_ChIP-Seq
BIN-67_Input_SMARCA4_ChIP-Seq
BIN-67_BAF155_Null_ChIP-Seq
BIN-67_45D_Null_ChIP-Seq
SCCOHT-1_BAF155_Null_ChIP-Seq
SCCOHT-1_45D_Null_ChIP-Seq
BIN-67_BRG1_Luciferase_ChIP-Seq_Rep2
BIN-67_SS18_Luciferase_ChIP-Seq
BIN-67_BAF155_Luciferase_ChIP-Seq
BIN-67_ARID2_Luciferase_ChIP-Seq
BIN-67_45D_Luciferase_ChIP-Seq
BIN-67_BRG1_SMARCA4.T910M_ChIP-Seq
BIN-67_ARID2_SMARCA4.T910M_ChIP-Seq
BIN-67_BAF155_SMARCA4.T910M_ChIP-Seq
BIN-67_SS18_SMARCA4.T910M_ChIP-Seq
BIN-67_45D_SMARCA4.T910M_ChIP-Seq
BIN-67_Input_SMARCA4.K785R_ChIP-Seq
BIN-67_BRG1_SMARCA4.K785R_ChIP-Seq
BIN-67_BAF155_SMARCA4.K785R_ChIP-Seq
BIN-67_ARID2_SMARCA4.K785R_ChIP-Seq
BIN-67_45D_SMARCA4.K785R_ChIP-Seq
BIN-67_SS18_SMARCA4.K785R_ChIP-Seq
BIN-67_Input_SMARCA4.T910M_ChIP-Seq
BIN-67_H3K4me1_Luciferase_ChIP-Seq
BIN-67_H3K27ac_Luciferase_ChIP-Seq
BIN-67_H3K27me3_Luciferase_ChIP-Seq
BIN-67_H3K9me3_Luciferase_ChIP-Seq
BIN-67_H3K4me1_SMARCA4.T910M_ChIP-Seq
BIN-67_H3K27ac_SMARCA4.T910M_ChIP-Seq

BIN-67_H3K4me1_SMARCA4_ChIP-Seq
BIN-67_H3K27ac_SMARCA4_ChIP-Seq
BIN-67_BAF155_Luciferase_ChIP-Seq_Rep2
BIN-67_DPF2_SMARCA4_ChIP-Seq
BIN-67_MLL3_Luciferase_ChIP-Seq
BIN-67_MLL3_SMARCA4_ChIP-Seq
BIN-67_MLL3_SMARCA4.T910M_ChIP-Seq
BIN-67_MLL4_Luciferase_ChIP-Seq
BIN-67_MLL4_SMARCA4_ChIP-Seq
BIN-67_MLL4_SMARCA4.T910M_ChIP-Seq

Genome browser session
(e.g. UCSC)

Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology

Replicates

All RNA and ATAC-seq was performed in duplicate. All ChIP-seq was performed in singlicate, so only peaks that were common in multiple overlapping epitopes (e.g. SS18 and SMARCA4) used for analysis.

Sequencing depth

All ChIP-seq and RNA-seq libraries were sequenced with ~50 million single end reads. All ATAC-seq libraries were sequenced with 150 million reads with paired end.

Antibodies

BIN-67_BRG1_Luciferase_ChIP-Seq_Rep1 Abcam ab110641 GR150844-25
BIN-67_Input_Luciferase_ChIP-Seq - - -
BIN-67_BRG1_SMARCA4_ChIP-Seq Abcam ab110641 GR150844-25
BIN-67_BAF155_SMARCA4_ChIP-Seq Homemade Fx3 -
BIN-67_ARID2_SMARCA4_ChIP-Seq CST D8D8U 1
BIN-67_SS18_SMARCA4_ChIP-Seq Cell Signaling Technologies D6I4Z 1
BIN-67_Input_SMARCA4_ChIP-Seq - - -
BIN-67_BAF155_Null_ChIP-Seq Homemade Fx3 -
BIN-67_45D_Null_ChIP-Seq Abcam ab134942 GR3185597-1
SCCOHT-1_BAF155_Null_ChIP-Seq Homemade Fx3 -
SCCOHT-1_45D_Null_ChIP-Seq Abcam ab134942 GR3185597-1
BIN-67_BRG1_Luciferase_ChIP-Seq_Rep2 Abcam ab110641 GR150844-25
BIN-67_SS18_Luciferase_ChIP-Seq Cell Signaling Technologies D6I4Z 1
BIN-67_BAF155_Luciferase_ChIP-Seq Homemade Fx3 -
BIN-67_ARID2_Luciferase_ChIP-Seq CST D8D8U 1
BIN-67_45D_Luciferase_ChIP-Seq Abcam ab134942 GR3185597-1
BIN-67_BRG1_SMARCA4.T910M_ChIP-Seq Abcam ab110641 GR150844-25
BIN-67_ARID2_SMARCA4.T910M_ChIP-Seq CST D8D8U 1
BIN-67_BAF155_SMARCA4.T910M_ChIP-Seq Homemade Fx3 -
BIN-67_SS18_SMARCA4.T910M_ChIP-Seq Cell Signaling Technologies D6I4Z 1
BIN-67_45D_SMARCA4.T910M_ChIP-Seq Abcam ab134942 GR3185597-1
BIN-67_Input_SMARCA4.K785R_ChIP-Seq - - -
BIN-67_BRG1_SMARCA4.K785R_ChIP-Seq Abcam ab110641 GR150844-25
BIN-67_BAF155_SMARCA4.K785R_ChIP-Seq Homemade Fx3 -
BIN-67_ARID2_SMARCA4.K785R_ChIP-Seq CST D8D8U 1
BIN-67_45D_SMARCA4.K785R_ChIP-Seq Abcam ab134942 GR3185597-1
BIN-67_SS18_SMARCA4.K785R_ChIP-Seq Cell Signaling Technologies D6I4Z 1
BIN-67_Input_SMARCA4.T910M_ChIP-Seq - - -
BIN-67_H3K4me1_Luciferase_ChIP-Seq Abcam ab176877 GR208955-5
BIN-67_H3K27ac_Luciferase_ChIP-Seq Abcam ab4729 GR288071-1
BIN-67_H3K27me3_Luciferase_ChIP-Seq Cell Signaling Technologies 9733S 8
BIN-67_H3K9me3_Luciferase_ChIP-Seq Abcam ab8898 GR3176466-6
BIN-67_H3K4me1_SMARCA4.T910M_ChIP-Seq Abcam ab176877 GR208955-5
BIN-67_H3K27ac_SMARCA4.T910M_ChIP-Seq Abcam ab4729 GR288071-1
BIN-67_H3K4me1_SMARCA4_ChIP-Seq Abcam ab176877 GR208955-5
BIN-67_H3K27ac_SMARCA4_ChIP-Seq Abcam ab4729 GR288071-1
BIN-67_BAF155_Luciferase_ChIP-Seq_Rep2 homemade (PMID: 28945250)
BIN-67_DPF2_SMARCA4_ChIP-Seq Abcam, ab134942
BIN-67_MLL3_Luciferase_ChIP-Seq homemade (PMID: 29404406)
BIN-67_MLL3_SMARCA4_ChIP-Seq homemade (PMID: 29404406)
BIN-67_MLL3_SMARCA4.T910M_ChIP-Seq homemade (PMID: 29404406)
BIN-67_MLL4_Luciferase_ChIP-Seq homemade (PMID: 29404406)
BIN-67_MLL4_SMARCA4_ChIP-Seq homemade (PMID: 29404406)
BIN-67_MLL4_SMARCA4.T910M_ChIP-Seq homemade (PMID: 29404406)

Peak calling parameters

Peaks were called using MACS 2.0 at a q value cutoff of 0.001 using the narrow peak setting, with the exception of ARID2 peaks, which were called using the broad peak setting. Peaks that fell in ENCODE blacklisted regions or were mapped to unmappable chromosomes (not chr1-22, X or Y) were removed. The input for each experimental condition are indicated above.

Data quality

All synthetic peaksets (residual complex, activity-independent, activity-dependent) were derived from multiple ChIP-seq

Data quality	epitopes targeting the same complex (SS18 and SMARCA4, ARID2 and SMARCA4).
Software	ChIP occupancy scatterplots Read count across peak sets of interest were calculated by calling the Rsubread bioconductor package function featureCounts() on duplicate removed bam files. These values were divided by the total number of mapped reads divided by one million, giving a normalized value of reads per million mapped reads for each interval in the input bed. Scatterplots were plotted with the R function smoothScatter. ChIP heatmaps Heatmaps of ChIP occupancy were generated created using HTSeq. To account for the 200bp average fragment length selected for in sonication, fragment length was extended 200 bp from the edge of each genomic interval. Total read counts for each interval were normalized to reads per million mapped reads (RPM). For each antibody the resulting matrix has a width of 4 kb and a height of the number of peaks in the indicated set. Strandedness of the interval was not considered except for the direction TSS plots (Figure 4D, Supplementary Figure 4A). Heatmaps were visualized using matplotlib in 5 nature research | reporting summary April 2018 python. Heatmaps were ordered by the maximum value in each matrix row of the indicated antibody. All other plots were generated with ggplot2 in R. Custom code to generate figures are available upon request, and a public repository will be included at https://github.com/joshbiology/sccoht.

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	*Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*
Instrument	*Identify the instrument used for data collection, specifying make and model number.*
Software	*Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*
Cell population abundance	*Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*
Gating strategy	*Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.*

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s)	*Specify: functional, structural, diffusion, perfusion.*
Field strength	*Specify in Tesla*
Sequence & imaging parameters	*Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*
Area of acquisition	*State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software — *Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*

Normalization — *If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*

Normalization template — *Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*

Noise and artifact removal — *Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*

Volume censoring — *Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.*

Statistical modeling & inference

Model type and settings — *Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).*

Effect(s) tested — *Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.*

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016) — *Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.*

Correction — *Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).*

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity — *Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).*

Graph analysis — *Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).*

Multivariate modeling and predictive analysis — *Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.*