Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The AT-hook is an evolutionarily conserved auto-regulatory domain of SWI/SNF required for cell lineage priming

Dhurjhoti Saha
The University of Texas MD Anderson Cancer Center
Solomon Hailu
The University of Texas MD Anderson Cancer Center
Arjan Hada
The University of Texas MD Anderson Cancer Center
Junwoo Lee
The University of Texas MD Anderson Cancer Center
Jie Luo
Institute for Systems Biology
Jeff Ranish
Institute for Systems Biology https://orcid.org/0000-0001-7181-0287
Kyle Feola
The University of Texas MD Anderson Cancer Center
Yuan-chi Lin
The University of Texas MD Anderson Cancer Center
Abhinav Jain
MD Anderson Cancer Center https://orcid.org/0000-0003-3268-514X
Bin Liu
Department of Genetics, University of Texas MD Anderson Cancer Center, Houston, TX
Yue Lu
The University of Texas MD Anderson Cancer Center
Payel Sen
National Institute on Aging, NIH https://orcid.org/0000-0003-2809-0901
Blaine Bartholomew (bbartholomew@mdanderson.org)
The University of Texas MD Anderson Cancer Center https://orcid.org/0000-0003-4014-9750

Article

Keywords: SWI/SNF, Snf2/Brg1/Smarca4, naïve, primed/EpiSCs, MNase-seq, Enhancer, PRO-seq, ATAC-seq

Posted Date: January 11th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2429800/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Abstract

The SWI/SNF ATP-dependent chromatin remodeler is a master regulator of the epigenome; controlling pluripotency and differentiation. Towards the C-terminus of the catalytic subunit of SWI/SNF is a motif called the AT-hook that is evolutionary conserved. The AT-hook is present in many chromatin modifiers and generally thought to help anchor them to DNA. We observe the AT-hook however regulates the intrinsic DNA-stimulated ATPase activity without promoting SWI/SNF recruitment to DNA or nucleosomes by increasing the reaction velocity a factor of 13 with no accompanying change in substrate affinity (K_M). The changes in ATP hydrolysis causes an equivalent change in nucleosome movement, confirming they are tightly coupled. Attenuation of SWI/SNF remodeling activity by the AT-hook is important in vivo for SWI/SNF regulation of chromatin structure and gene expression in yeast and mouse embryonic stem cells. The AT-hook in SWI/SNF is required for transcription regulation and activation of state-specific enhancers critical in cell lineage priming. Similarly, the AT-hook is required in yeast SWI/SNF for activation of genes involved in amino acid biosynthesis and metabolizing ethanol. Our findings highlight the importance of studying SWI/SNF attenuation versus eliminating the catalytic subunit or completely shutting down its enzymatic activity.

Introduction

In all SWI/SNF complexes, the AT-hook is located between the SnAC and bromo domains at the C-terminus of the catalytic subunit. The AT-hook motif is found in a number of other chromatin modifiers and has been suggested to anchor these proteins to chromatin[1,2]. The AT-hook motif is a tripeptide of arginine-glycine-arginine flanked on either one or both sides by proline, first identified in HMGA1 (High mobility group AT-hook 1). The AT-hook binds to the minor groove of DNA, alters DNA architecture and facilitates binding of other proteins[3–5]. Sequence alignment of a large number of AT-hook containing proteins has revealed a 10–15 amino acid motif that has a G-R-P core surrounded by K and R residues [2,6,7]. A special class of AT-hook motifs called extended AT-hooks are approximately 3 times longer than the conventional AT-hook with basic residues extending symmetrically 12–15 amino acids from the core AT-hook motif and have an order of magnitude higher affinity for RNA than DNA[2,8]. The most characterized ATP-dependent chromatin remodeling complex in terms of the AT-hook is the nucleolar remodeling complex (NoRC) with Snf2h, a Snf2-like ATPase subunit, and Tip5 (TTF-I interacting protein 5), a regulatory subunit containing 4 AT-hooks and a TAM (Tip5/ARBP/MBD) domain[9,10]. The NoRC complex represses rRNA transcription by positioning nucleosomes at the promoters of rRNA genes and facilitating the association of rRNA genes with the nuclear matrix. The AT-hook within the Tip5 subunit of the NoRC complex is required for NoRC binding to rRNA gene promoters[11]. The AT-hooks in the auxiliary NURF 301 subunit[12] and Rsc1 and 2 subunits in yeast RSC complex[13] are important for efficient remodeling, but the mechanistic basis of these AT-hooks promoting remodeling is not understood. Even though the AT-hook in the catalytic subunit of the SWI/SNF complex is conserved throughout eukaryotes, it is not known if it is required for nucleosome remodeling. We find the AT-hook in the catalytic subunit of SWI/SNF positively auto-regulates remodeling by modulating the intrinsic DNA-dependent ATPase activity of the catalytic subunit without impacting the DNA affinity of SWI/SNF. The AT-hook is evolutionary important for SWI/SNF regulation as shown by being required for cell lineage priming in mouse embryonic stem cells and for regulating genes involved in amino acid biosynthesis in yeast. The AT-hook has a role in SWI/SNF regulation that is clearly distinct from the complete absence or inactivity of SWI/SNF.

Results

AT-hooks positively regulate the nucleosome mobilization activity of SWI/SNF and are not essential for complex integrity

Two AT-hooks are located at the C-terminus of yeast Snf2 between the Snf2 ATPase Coupling (SnAC) domain and bromodomain (Fig. 1a and b). First, we find the AT-hooks are not required for complex integrity by deleting both AT-hooks (ΔAT) and purifying the mutant complex by M2 agarose immunoaffinity chromatography and analyzing the purified complex by SDS-PAGE (Fig. 1c). We initially examine the nucleosome remodeling activity of WT and DAT SWI/SNF to determine if there are differences when SWI/SNF is limiting. We tested remodeling efficiencies by examining the extent of nucleosome movement while varying the concentration of SWI/SNF with a fixed reaction time of 10 minutes (Fig. 1d). The lower concentrations of SWI/SNF used in these experiments are limiting as seen by the extent of mobilized nucleosomes. The similar efficiencies in nucleosome mobilization observed when varying the concentration of wild type versus mutant SWI/SNF suggest the AT-hook does not significantly impact SWI/SNF binding. Next, we find the AT-hook is required for efficient remodeling when SWI/SNF is saturating and reaction times are varied (Fig. 1e). The DAT SWI/SNF mobilizes nucleosomes 13 times slower than WT SWI/SNF (0.26 nM/s for WT compared to 0.0019 nM/s for DAT) and is similar to the decrease in the rate of ATP hydrolyzed, thus suggesting they are tightly coupled (Fig. 1f-g). Given nucleosomes are fully bound by WT and DAT SWI/SNF, the differences in remodeling are not due to any reduction in binding of DAT SWI/SNF.

AT-hooks positively regulate the DNA- and nucleosome-stimulated ATPase activity of SWI/SNF

The coupling of the rate of ATP hydrolysis and nucleosome mobilization in DAT SWI/SNF prompted us to test the role of the AT-hook in regulating the DNA- and nucleosome-dependent ATPase activity of SWI/SNF by determining the velocity or reaction rate (V_{max}) and binding affinity for substrate (K_M) using the Michaelis-Menten approach for WT and DAT SWI/SNF (Fig. 2a-d). Loss of the AT-hook does not significantly impact the substrate affinity of SWI/SNF but does decrease the reaction rate by a factor of 13 with free DNA (Table 1). SWI/SNF binding to nucleosomes increases the binding affinity of WT SWI/SNF for substrate over 3-times that with DNA (311 versus 98 nM), which was not observed when the AT-hook was deleted from the complex (Table 1). The rate of ATP hydrolysis is decreased 14-fold with DAT versus WT SWI/SNF in the presence of nucleosomes. In summary, the reduction in ATP hydrolysis that occurs when the AT-hook is deleted is not histone dependent, although there does seem to be evidence for some interplay between AT-hooks and histones.

Table 1

	Km (nm)	95% CI of Km	Vmax (mM/min)	95% CI of Vmax
WT + DNA	311	188–541	0.795	0.664 to 1.04
dAT + DNA	340	206–593	0.0604	0.0489 to 0.0792
WT + NCP	98	55.3–178	4.19	3.36 to 5.47
dAT + NCP	364	218–681	0.301	0.246–0.405

The AT-hooks moderately contribute to the affinity of yeast SWI/SNF for nucleosomes

The affinity of ΔAT and WT SWI/SNF for DNA and nucleosomes was measured by electrophoretic mobility shift assays (EMSA) to directly assess the AT-hook contribution to SWI/SNF's affinity for free DNA and nucleosomes. The apparent K_D of ΔAT was slightly more than WT for free 601 DNA (12.8 versus 7.87 nM, 1.6-fold difference); whereas for nucleosomes there was a 2.7-fold decrease in affinity for ΔAT compared to WT SWI/SNF (20.5 versus 7.67 nM) (Figs. 3a-b and S1a-b). These data suggest the AT-hook has a modest role in promoting the affinity SWI/SNF for nucleosomes but not significantly for DNA consistent with that observed in the Michaelis-Menten data. We examined the ability of plasmid DNA to compete for SWI/SNF binding to either 601 DNA or nucleosomes. The Ki for competing SWI/SNF from DNA is ~ 2-fold lower for WT than ΔAT SWI/SNF (12 versus 25 nM), indicating the AT-hook promotes competition of plasmid DNA for binding SWI/SNF (Figure S2a-b). In contrast there is no difference in the estimated Ki observed for WT and ΔAT SWI/SNF when bound to nucleosomes which indicates the AT-hook does not promote the transfer of SWI/SNF from nucleosomes to free DNA and likely behaves differently when SWI/SNF is bound to nucleosomes.

Next, SWI/SNF recruitment to nucleosomes by the VP-16 activation domain fused to the Gal4 DNA binding domain (Gal4-VP16) did not require AT-hooks as seen by EMSA (Figure S1c). SWI/SNF binding to nucleosomes under these conditions is highly dependent on Gal4-VP16 binding to extranucleosomal DNA 29 bp from the edge of nucleosomes (Figure S1c compare lanes 3–5 to 6–8). The ΔAT complexes are recruited the same as wild type SWI/SNF under these conditions and AT-hooks are not required for efficient recruitment of SWI/SNF by an acidic transcription factor (Figure S1c lanes 9–20).

The AT-hooks of yeast Snf2 interact with the N-terminal tail of histone H3

To understand how SWI/SNF interacts with nucleosomes and the role of the AT-hooks in these interactions, we mapped protein-protein interactions (PPIs) in nucleosome-bound SWI/SNF by crosslinking-mass spectrometry (CX-MS) using the amine-reactive crosslinker bis[sulfosuccinimidyl] suberate (BS3). There is good concordance of the CX-MS data with the cryo-EM structure of nucleosome-bound SWI/SNF as seen by the majority of the lysine pairs observed in the intra- and inter-links having a Cα-Cα distance of ≤30 angstroms in the solved structure and are similar to that observed earlier for CX-MS of free SWI/SNF [14](Figure S3). The Snf2, Snf5, Arp7, Swi1 and Snf6 subunits crosslinked to histones, consistent with the recent cryo-EM structures of nucleosome-bound ySWI/SNF[14,15] (Figure S3a). The histones crosslink the C-lobe, the linker region between the two lobes of the ATPase domain and the HSA domain in full agreement with the cryo-EM structure. We observe the C-lobe of the ATPase domain interacts with the N-terminal tail of histone H4 by crosslinking to lysines 9, 13 and 21 of H4, consistent with other data, but not detected in the cryo-EM structure[16] (Fig. 3c). These data demonstrate the H4 N-terminal tail may influence SWI/SNF remodeling in a manner similar to that previously observed for the ISWI family of chromatin

remodelers[17–19] including interacting with the C-lobe of the ATPase domain[20]. The HSA domain interacts with the N-terminal tail of H3 which agrees with the HSA domain and the H3 tail both contacting extranucleosomal DNA [21,22] (Fig. 3c).

The CX-MS data reveals novel interactions not previously observed in the cryo-EM structure regarding the AT-hook and SnAC domain. We find the C-terminus of the SnAC domain crosslinks to the C-terminus of H2B at lysines 117 and 121 and the AT-hooks crosslinks to the N-terminal tail of histones H3 at lysines 15 and 28 of H3. AT-hook interactions with the H3 tail is consistent with acetylation of H3 histone tail and the AT-hook positively regulating SWI/SNF recruitment and remodeling activity, similarly ubiquitination of the C-terminus of H2B interfering with SWI/SNF remodeling is also congruent with SnAC binding at this site [16,23–26] (Fig. 3c). The SnAC domain also interacts with the C-lobe of the ATPase domain as seen by lysines 1320 and 1336 in SnAC crosslinking to lysines 1028, 1040 and 1041 in the N-terminus of the C-lobe when SWI/SNF is bound to nucleosomes (Fig. 3d). The interactions of the SnAC domain with the C-lobe of the ATPase domain is consistent with biochemical data showing the SnAC domain is a positive regulator of the ATPase activity of SWI/SNF[14,15,27]. The AT-hook and SnAC domain also associate with each other as observed by lysine 1314 in the SnAC domain crosslinking lysine 1441 in the first AT-hook and additional crosslinks between the regions flanking the AT-hooks and SnAC domain (Fig. 3e). The AT-hook does not appear to directly interact with the ATPase domain as no crosslinks are detected between them and the AT-hook may regulate the catalytic activity of SWI/SNF through its interactions with the SnAC domain.

Disruption of the AT hooks alters the in vivo activity of SWI/SNF

The impact of loss of the AT-hooks on the in vivo activity of SWI/SNF is assessed by examining cell growth under conditions where the SWI/SNF complex is required for viability such as amino acid deprivation or switching metabolism to alternative carbon sources. The Snf2 subunit of yeast SWI/SNF is required for cell viability when the amino acids isoleucine and valine are omitted, but the catalytic activity of Snf2 is not (K798A) nor the AT-hooks and SnAC domain are required (Fig. 4a, SC-Ile-Val). In this case the physical presence of SWI/SNF is required rather than its chromatin remodeling activity. In contrast, we observe the catalytic activity of Snf2 is important for Gcn4 mediated activation of genes involved in amino acid biosynthesis and switching to alternative carbon sources for glucose (compare snf2-CD versus Snf2D in Fig. 4a, SM and Fig. 4b). In these conditions, we observe varied effects when the catalytic activity of Snf2 is partially down regulated due to the loss of the AT-hooks or SnAC domain in contrast to the catalytically dead mutant (K798A). The ATPase activity of the catalytically dead Snf2 mutant has previously been shown to be at least an order of magnitude more inhibited than the AT hook or SnAC domain deletion mutant complexes[28]. High efficiency of SWI/SNF remodeling is particularly required for Gcn4 mediated gene activation, less for using ethanol as an energy source and is not required for using raffinose (Fig. 4a). These results show attenuation of SWI/SNF remodeling activity will have varied context dependent effects in vivo that are missed when merely deleting SWI/SNF or fully inhibiting the complex.

We find the chromatin structure at promoters is altered more when SWI/SNF activity is downregulated ~ 10-fold than > 100-fold or the catalytic subunit is absent. Nucleosome occupancy is reduced upstream of the transcription start site (TSS) at the − 1 and − 2 positions of 1,697 and 1,291 genes, respectively and in the nucleosome-free region (NFR) when the AT-hook of Snf2 is deleted as shown by MNase-seq, i.e. micrococcal digestion and next generation DNA sequencing (Fig. 4c; blue-reduced, red-gained, grey-unchanged). These changes in nucleosome occupancy are negligible when Snf2 is deleted or less when the activity of SWI/SNF is more severely blocked (compare Fig. 4c to Figs. 4d-f). SWI/SNF therefore has a more nuanced role in promoter structure that is best revealed when attenuated rather than completely inhibited and the AT-hook is required for the in vitro and in vivo activities of SWI/SNF.

The AT-hook is also required in mammalian SWI/SNF for state-specific activation of enhancers

The AT-hook is a motif evolutionarily conserved in all eukaryotic versions of SWI/SNF which prompted us to examine the role of the AT-hook in the mammalian SWI/SNF complex by deleting exon 33 containing the AT-hook by CRISPR/Cas9 in both copies of BRG1 in mouse embryonic stem cells (mESCs). The genome-wide binding patterns of WT and two independent clones of the BRG1 AT-hook deletion mutant (DAT1 and DAT2) are mapped in two distinct pluripotent states referred to as naïve and primed, representative of the pre- and post-implantation stages using Brg1 ChIP-seq. We observe in the two different states most Brg1 sites are cell state-specific in naïve and primed pluripotent cells with significantly more Brg1 binding detected in the primed state (Figs. 5a and S4a). Brg1 in mESCs is assembled into esBAF (embryonic BAF) and GBAF or ncBAF (non-canonical BAF) complexes, and the esBAF complex preferentially binds to cis-regulatory regions and the GBAF complex to promoter regions[29]. Approximately 70–80% of the naïve- and primed-specific Brg1 peaks are at intronic or intergenic regions and correspond to the esBAF complex and 13–18% of the peaks are associated with promoter regions (Figure S4b). Similar to ySWI/SNF, deletion of the AT-hook does not affect SWI/SNF complex integrity in either naïve or primed cells as detected by Western blot and mass spectrometry respectively, although a modest reduction in the ARID1a and 1b subunits in the DAT complex in

primed cells is observed (Figures S4c,e). The expression and transport of Brg1 into the nucleus is also not affected by loss of the AT-hook (Figure S4d).

Next, we focused on the intronic and intragenic regions to identify those that are active enhancers by mapping TF binding/DNA accessibility (ATAC-seq) [30] and the localization of acetylated lysine 27 (H3K27ac) and monomethylation of lysine 4 (H3K4me1) of histone H3 by ChIP-seq. A total of 38,416 and 42,303 ATAC peaks are detected in naïve and primed cells of which 18,814 peaks are in common between naïve and primed (Figure S5a). The majority of state-specific ATAC peaks reside in intronic and intergenic regions, approximately 14,000 peaks in naïve or primed cells, and based on their co-localization with histone marks H3K27ac and H3K4me1 are in active enhancers (Figs. 5b and S5a-b). These accessible sites reflect binding of stage-specific TFs and motif analysis reveals the pluripotency TFs Oct4, Nanog, Klf family and Sox2 are most enriched at these sites that are naïve-specific and the epiblast-specific TFs Zic2, Zic3, Otx2 and Glis3 are most enriched in the primed state[31–34] (Figure S5c-d). Pluripotency TFs also bind at other sites in both the naïve and primed states that are not stage specific which shows there are two classes of pluripotency TF binding sites (Fig. 5e).

The stage-specific binding of many TFs depends on Brg1 and its AT-hook with 5,061 and 5,350 accessible sites in respectively the naïve and primed states depending on the AT-hook of Brg1 and more than 7,000 sites that remain unaltered upon deletion of the AT-hook (Figures S5f-g). Naïve-specific binding of the core pluripotency TFs (Oct4, Sox2, Nanog and Tcf) is AT-hook dependent as observed by DNA footprint analysis, while binding of these TFs at sites common to both states does not depend on the AT-hook of Brg1 (Figs. 5c-d). The epiblast TF Zic3 is dependent on the AT-hook but not for Otx2, another epiblast-specific TF (Figs. 5e-f). A more thorough breakdown of the AT-hook dependency of naïve- and primed-specific TF binding is shown in Fig. 5g. While Brg1 had been shown previously to be required for recruitment of Oct4, Sox2 and Nanog, this is the first time Brg1 and its AT-hook is shown to promote binding of these TFs specifically in the naïve state and to promote binding of a select set of epiblast-specific TFs in the primed state[35–38].

Beyond mediating TF binding, the connection of Brg1 and its AT-hook to enhancer activation is also examined for the AT-hook dependency of H3K27ac and H3K4me1. Monomethylation of lysine 4 of histone H3 at the naïve- and primed-specific enhancers requires the AT-hook of Brg1; whereas acetylation of lysine 27 of histone H3 is not reduced by loss of the AT-hook (compare Figs. 6a-b to 6c-d). There is no tight connection between TF binding and H3K4me1 as H3K4me1 is lost at sites where TF factor binding is not affected (Figs. 6a-b compare upper and lower panels). Brg1 recruitment at these sites is not changed by loss of the AT-hook and these data demonstrate the catalytic activity of Brg1 is required for monomethylation of H3K4 and not for acetylation of H3K27 (Figure S6).

Brg1 and its AT-hook are required to activate transcription in the naïve and primed

To determine if these altered enhancers are indeed involved in transcription regulation, we map actively transcribing RNAP II with base-pair resolution using Precision Run-On Sequencing (PRO-seq)[39]. In the DAT1 clone there are a set of 3,037 genes (1,454 genes in DAT2) normally expressed in naïve and repressed in the primed state that fail to be actively transcribed in the naïve state when the AT-hook of BRG1 is deleted (Figs. 7a and S7a). This group of genes are involved in such processes like signal transduction, metabolism and cell communication (Fig. 7b). Similarly, there is a group of 1,003 genes in DAT1 (720 genes in DAT2) that are up regulated in the primed versus the naïve state that fails to be activated in the AT mutant (Fig. 7c and S7b). The genes that fail to be up regulated in the AT-hook mutant in the primed stated are those important in cell development and differentiation and are most likely involved in the early step of cell fate determination or cell lineage priming. These changes in nascent transcription reflects well the changes in enhancer activation that are observed earlier when the AT-hook mutant is deleted and points to Brg1 and its AT-hook for regulating gene transcription through enhancer activation.

At the promoters of these genes, enrichment of trimethylated lysine 4 of histone H3 and DNA accessibility is not altered upon deletion of the AT-hook, consistent with the gene dysregulation being due to altered enhancer activity rather than restructuring of promoters (Figures S8a-f). The AT-hook dependent enhancers are in the correct proximity (250 kb-1 Mb) to regulate these genes in naïve and primed cells (Figure S8g). In summary, we find the AT-hook of Brg1 is vital for Brg1's role in activating stage-specific enhancers which regulates expression of genes important in differentiation and cell development.

There are other changes in gene transcription when the AT-hook is deleted that instead of impacting activation are more consistent with Brg1 having a dual role in repression. A group of 1,468 genes expressed in the naïve state that are repressed in the primed state fail to be repressed in DAT1 (2,166 genes in DAT2) in the primed state (Figs. 7e and S7d). Some of these genes encode factors involved in oxidative phosphorylation (OX-PHOS) and pentose phosphate pathways (PPP) whose expression is repressed in the transition from naïve to primed accompanying the shift to glycolysis [40,41] (Fig. 7f). There are other genes involved in lipid metabolism [42–44], cell cycle control [45] and G protein-coupled receptor signaling [46,47] we observe to be down regulated in the transition from naïve to primed, as part of exiting

pluripotency, that fail to be repressed in the primed state when the AT-hook of Brg1 is absent. These observations further illustrate the importance of Brg1 and its AT-hook in cell lineage priming because not only is it required to activate cell fate related genes but also to start the down regulation of pluripotency. There is a second set of genes active in the primed state that are repressed in the naïve state which when the AT-hook is deleted are up regulated in the naïve state, consistent with these genes being repressed by Brg1 in the naïve state (Figs. 7g and S7d).

Loss of the AT-hook from Brg1 disrupts cell lineage priming

Although many transcriptional changes occur in the naïve state when the AT-hook is deleted, these do not affect stem cell proliferation, clonal expansion, alkaline phosphatase staining or colony morphology (Figs. 8a-c and S9a-c). The expression and nuclear localization of the core pluripotency TFs Oct4, Nanog and Sox2 are also not altered by loss of the AT-hook (Figure S9d). These results highlight the differences between the complete loss of Brg1 versus loss of the AT-hook since the absence of Brg1 blocks expression of the core pluripotency TFs and stem cell maintenance and proliferation[48].

Brg1 and it AT-hook are however important for cell lineage priming as shown by measuring the expression of three distinct classes of lineage specific markers by qRT-PCR after culturing cells in the absence of LIF and two inhibitors for 7 days. Ectodermal markers Sox1 and Nestin, mesoderm marker Tbxt and the endoderm marker Sox17 are all down regulated in both DAT mutant clones compared to WT cells, consistent with defects in cell lineage priming (Fig. 8d). The endoderm specific gene Gata4 is aberrantly highly expressed in both clones of the DAT mutants compared to wild type (Fig. 8d). The intracellular levels of Sox1 and Gata4 in the two DAT mutant clones are respectively lower and higher in wild type cells as shown by immunofluorescence, thus suggesting the DAT mutant has a bias toward differentiating into endoderm compared to wild type (Fig. 8e). These results further confirm the importance of the AT-hook in cell lineage priming consistent with our profiling of enhancer activation and gene regulation in the primed state.

Discussion

Although the AT-hook binds to A/T-rich DNA as originally found in the HMGB family of proteins, we observe when the AT-hook regulates the nucleosome remodeling activity of SWI/SNF it does so outside of promoting binding to DNA or chromatin. Most eukaryotic AT-hooks are found in chromatin modifiers and for more than 2 decades have been considered to anchor these proteins to A/T-rich sequences[1,2]. The intrinsic DNA-dependent ATPase activity of SWI/SNF depends on the AT-hook as seen using free DNA and yeast SWI/SNF. The reduction in ATPase activity is directly coupled to a reduction in nucleosome mobilization in contrast to the uncoupling previously observed for the SnAC domain[49]. The AT-hook has a modest secondary role in enhancing SWI/SNF affinity for nucleosomes through its interactions with the N-terminal tail of histone H3 that doesn't significantly impact the nucleosome mobilization activity of SWI/SNF. Given the proximity of the bromodomain to the AT-hook the bromodomain because of its ability to bind H3K14ac could be a bridge between the AT-hook and histone H3 tail, but protein crosslinking indicates these interactions are most likely direct between the H3 tail and AT-hook [50].

The AT-hook is also shown to be important for the in vivo activity of SWI/SNF in yeast and mouse embryonic stem cells and has provided novel insights into SWI/SNF not observed previously by either deleting its catalytic subunit or completely inhibiting its ATPase/remodeling activity. Our studies show attenuation of SWI/SNF activity can have outcomes that are distinct from complete loss or shutdown of SWI/SNF and has helped uncover for the first time Brg1 being required for early cell lineage priming. Whereas deletion of Brg1 blocks stem cell proliferation/self-renewal and expression and recruitment of the core pluripotency transcription factors; deletion of the AT-hook does not[35,37,48]. The pluripotency circuitry shared in naïve and primed states where the pluripotency TFs are bound at the same sites in both states are not perturbed by loss of the AT-hook. The AT-hook is instead required for stage-specific binding of pluripotency TFs to naïve-specific sites and a subset of epiblast TFs to primed-specific sites. We don't not know yet the difference between the two sets of pluripotency TFs binding sites that cause them to have different SWI/SNF requirements. Similarly, we find the co-dependency between Brg1 and H3K27ac is not dependent on the remodeling efficiency of Brg1 and therefore on the AT-hook[51–53]. Surprisingly, we observe H3K4me1 at stage-specific enhancers is dependent on efficient SWI/SNF remodeling and is therefore independent of H3K27ac. They are due to catalytic differences since there is no significant change in Brg1 localization at these sites when the AT-hook is deleted and the yeast SWI/SNF biochemical data clearly shows the AT-hook regulates the catalytic activity.

We also observe in yeast that attenuation of SWI/SNF activity by the AT-hook reveals in vivo aspects of SWI/SNF not observed by deletion of the catalytic subunit or complete shutdown of its enzymatic activity. Highly efficient SWI/SNF remodeling is shown to be required for positioning nucleosomes upstream of the TSS and for positioning of fragile nucleosomes inside the NFR, which is not seen when the catalytic subunit is deleted or is less obvious with the catalytically dead subunit. We also observed that phenotypes well known to be SWI/SNF dependent have varied dependencies on the AT-hook thus making it possible to subdivide these based on how strictly they

depend on the remodeling efficiency of SWI/SNF. For example, switching from glucose to raffinose does not depend on the efficiency of SWI/SNF to mobilize nucleosomes, whereas Gcn4 activation of amino acid biosynthetic genes is dependent.

We show for the first time by extensive biochemical, genetic and genomic analyses the importance of attenuating SWI/SNF remodeling and the role of the AT-hook in this process. These studies highlight the need to not only study SWI/SNF in vivo by subunit deletion or chemical inhibitors to completely shut down its enzymatic activity, but to also keep in mind the power of attenuation of SWI/SNF for the in vivo regulation of gene expression and chromatin structure.

Materials And Methods

Yeast strains and SWI/SNF preparation

Mutant SWI/SNF strains were generated using a ylpLac128 plasmid containing Snf2 along with C terminal double Flag epitope and LEU2 marker. Domain deletions were generated by gap repair where 2 PCR products flanking the domain of interest (either both AT hook domains or only the N terminal AT hook domain) were made using ylpLac128-SNF2-2FLAG-LEU2 plasmid as template and were co-transformed into a Δsnf2 strain. Primer pairs used for generating the PCR products to delete both N and C terminal AT hooks (ΔAT) are summarized in Table 2. After co transformation of both primer pairs, yeast was grown in a Synthetic complete (-Leu) plate and replica plated onto a YPD-Kanamycin plate. Colonies that grew on the Sc-Leu plate but not on the YPD-Kan plate were then streaked onto a fresh Sc-Leu plate and sequenced for confirmation of the domain deletion and correct integration of the mutant snf2 sequence into the genome at the N and C terminus of the point of insertion. Clones which were sequenced, and integration positive were then grown separately in 100 ml YPD media and western blotted with Anti-Flag antibody to confirm expression of the mutant Snf2. SWI/SNF was purified as described previously in [54].

Yeast Snf2 Strains

Strain	Genotype
scSG001	MATa, his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 SNF2-2xFLAG::LEU2
scSG002	MATa, his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 snf2-N147S-R993H-K1279R-R1385K-S1689T-A1701T-ΔAT-2xFLAG::LEU2
scSG003	MATa, his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 SNF2::LEU2
scSG004	MATa, his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 snf2-ΔAT::LEU2
scSG005	MATa, his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 snf2-K798A::LEU2
scSG006	MATa, his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 snf2-ΔSnAC::LEU2
scSG007	MATa, his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 snf2Δ::LEU2

Nucleosome reconstitution

Homogenous mononucleosomes were reconstituted with 10 ug of PCR generated DNA using p159 plasmid containing a 601 nucleosome positioning sequence (NPS) as template and 29 bp and 59 bp flanking either side of the 601 sequence (29N59), 10 µg of wild type *Xenopus laevis* (Xenopus nucleosomes) or wild type *Saccharomyces cerevisiae* (Yeast nucleosomes) octamer, and 100 fmol of ^{32}P labeled 29N59 DNA at 37 C by a rapid salt dilution method [55]. Heterogenous nucleosomes were reconstituted with 100 fmol of ^{32}P labeled 29N59 DNA, 10 µg of wild type Xenopus laevis octamer and 9 µg of salmon sperm DNA by the same method. The radiolabeled DNA was generated by PCR using an oligonucleotide labeled using Optikinase (USB) and γ ^{32}P ATP (6000 Ci/mol). Homogenous yeast di-nucleosomes were reconstituted with 5 µg of PCR generated DNA using p199 plasmid containing 601b and 603 NPS separated by 31 bp and with 40 bp and 6 bp flanking the dinucleosome NPS (40N31N6) as template, 12 µg of Yeast octamer, 100 fmol of ^{32}P labeled 40N31N6 DNA, and 4 µg of BSA (NEB) by rapid salt dilution method.

Electrophoretic mobility shift assays (EMSA)

A titration was done with purified SWI/SNF (WT or mutant) pre-bound to homogenous / heterogenous nucleosomes (20 nM) or DNA (10 nM) at 30 C for 15 mins in the presence or absence of 3.2 nM Gal4-VP16 to determine the amount of SWI/SNF at which > 90% of the

nucleosome (or DNA) was fully bound as seen by gel-shift in a 4% native polyacrylamide gel (acrylamide:bis-acrylamide ratio 37.35:1) run at 200 V at 0.5x Tris-Borate-EDTA. Using these conditions, remodeling assays were performed where after prebinding for 15 mins at 30 C, ATP was added to a final desired concentration and then the reaction stopped at defined time points by addition of salmon sperm DNA and EDTA (final concentration 0.5 µg/ µl and 40 mM respectively in reaction). The remodeled products were analyzed on 5% native polyacrylamide gels (acrylamide:bis-acrylamide ratio 60:1) at 100 V in 0.5x Tris-Borate-EDTA.

ATPase assays

ATPase kinetic assays were performed similar to remodeling assays with γ ^{32}P ATP (0.02 µCi) mixed with cold ATP with homogenous Xenopus mono / di- nucleosomes (nucleosome stimulated) or with 25, 50, 100 bp or 3kbp of DNA (DNA stimulated). Reactions were stopped using a mixture of SDS and EDTA (final concentration 2.5% and 50 mM in reaction respectively). 1 µl of Stopped reactions were then spotted on JD Baker PEI-Cellulose TLC plates and run in TLC run solution (0.8M glacial Acetic acid, 0.8M Lithium chloride). Plates were dried and imaged by autoradiography.

Growth assays

WT and mutant SWI/SNF were grown overnight in 5 ml YPD cultures and the OD600 was adjusted to 1 by addition of fresh YPD. 100 ul of each culture was taken out and 5–10 fold serial dilutions were made. 10 ul of the original culture (at OD600 = 1) and each of the serial dilutions were then aliquoted onto YPD Dextrose, Sucrose, Galactose, Raffinose, Glycerol and Ethanol plates and dried. Plates were then turned over and left in 30C for 2–3 days until appearance of yeast colonies.

Chemical crosslinking and mass spectrometry analysis

SWI/SNF with Snf2 tagged at its C-terminus with two copies of the FLAG peptide was purified in a two-step process. SWI/SNF was purified by M2 immunoaffinity chromatography and the FLAG peptide removed by cation exchange chromatography using SP-Sepharose [56]. Crosslinking reactions contained 35 nM FLAG tagged SWI/SNF, 35 nM 29-N-59 nucleosomes and 200 nM Gal4-VP16 in a final volume of 520 mls and BS^3 (bis(sulfosuccinimidyl)suberate, an amine homobifunctional crosslinker from Thermo Scientific) was added to a final concentration of 115 mM. After crosslinking at room temperature for 2 hr, the reaction was quenched by adding 20 µL of 1 M ammonium bicarbonate, precipitated by addition of trichloroacetic acid (TCA) to 16.7% and further processed as described[56]. The integrity of the SWI/SNF-nucleosome complex was tested after crosslinking by SDS-PAGE to ensure BS^3 did not disrupt SWI/SNF binding.

RAW files were converted to mzXML by RawConverter software. We used the Comet search engine and the Trans-Proteomic Pipline (TPP, http://tools.proteomecenter.org/wiki/index.php?title=Software:TPP) were used for identification of unmodified and BS3 mono-modified (monolinks) peptides. Two crosslink database search algorithms were used for crosslinked peptide searches: pLink20and in-house designed Nexus with default settings. The database used for crosslink identification contained the wild type and modified protein sequences of the yeast SWI/SNF subunits and their reverse decoy sequences. A 5% FDR cutoff was used for both pLink and Nexus searches. The results were combined, and each spectrum was evaluated for the quality of the match to each peptide using the COMET/Lorikeet Spectrum Viewer (TPP). Crosslinked peptides were considered confidently if the majority of the observed ions are accounted for and at least 4 consecutive b or y ions were observed for each peptide. Finally, interaction plots are made using Circos online tool [57].

Growth assays

WT and mutant SWI/SNF were grown overnight in 5 ml YPD cultures and the OD600 was adjusted to 1 by addition of fresh YPD. 100 ul of each culture was taken out and 5 10 fold serial dilutions were made. 10 ul of the original culture (at OD600 = 1) and each of the serial dilutions were then aliquoted onto YPD Dextrose, Sucrose, Galactose, Raffinose, Glycerol and Ethanol plates and kept drying. Plates were then turned over and left in 30C for several days annotated until appearance of yeast colonies.

MNase-seq

Cells from WT Snf2 and AT hook domain deletion strain in BY4741 background were grown at 30^0C in YPD to an OD_{600} of 0.8. After digestion using MNase, mono-nucleosomal DNA was sequenced using high throughput paired-end sequencing as in [58] using Illumina Genome Analyzer IIx.

Mouse embryonic stem cell culture

Mouse embryonic stem cells (E14Tg2a, ATCC) were maintained on 1% gelatin-coated plates in the ESGRO complete plus clonal grade medium (Millipore), as previously described[59,60]. Embryonic stem cells (ESCs) were cultured on gelatin-coated plates in DMEM (Invitrogen)

supplemented with 15% FBS, 1X-Gultamax (Gibco), Na-Pyruvate (Gibco), 10mM 2-mercaptoethanol, 0.1 mM nonessential amino acids (Gibco), 1U/ml of ESGRO mLIF (Millipore), and 2i inhibitors (MEK inhibitor PD0325901, Gsk3b inhibitor (CHIR99021 – Stem Cell technology) in naïve (or + 2i) condition. In epiblast stem cells (EpiSCs) or primed condition, cells were cultured in chemically defined medium (IMDM and F12-Invitrogen) supplemented with 2%-BSA (Sigma), Insulin (Roche), Transferrin (Roche), CD-lipid concentrate (Gibco), FGF2 (R&D), and Activin-A (R&D) (Ref). For spontaneous differentiation (SD), cells were cultured in gelatin coated plate in previously described DMEM-FBS media without LIF and 2i-inhibitors. Cell lines were continuously monitored under microscope and confirmed to be free of mycoplasma contamination by using a MycoAlert mycoplasma detection kit (Lonza) and DAPI staining.

CRISPR mediated deletion of the AT-hook of SMARCA4/BRG1 and insertion of the HA-tag

Guide RNA used for CRISPR-Cas9 editing were designed usin the CRISPR Design Tool (http://crispr.mit.edu) to minimize off-target effects. G-blocks containing the guide RNA directed to the AT-hook domain of SMARCA4/BRG1 (exon 30) came from IDT and were PCR amplified, cloned into two Cas9 containing plasmids (pX330; Addgene#158973) using Zero Blunt TOPO Cloning Kit (Invitrogen) and sequence verified. After 72 hrs of transfection in mESCs using electroporation, positive colonies were selected based on puromycin (1mg/ml) selection. Plasmids were separately test in trial transfection in E14 cells to determine the efficiency of guide RNA cleavage by isolating genomic DNA (QuickExtract DNA Extraction from Epicentre), the region of interest being amplified by PCR and editing identified using EnGen Mutation Detection Kit (NEB). All guide RNA used in HA-tag insertion or AT-hook deletion had a cleavage efficiency of > 60%. ESCs were seeded at low density to allow for selection of individual colonies. Colonies were individually expanded and split for future culture or genomic DNA isolation. Genomic DNA (100 ng) from these colonies was used to confirm the desired targeted deletion by PCR. The donor DNA for HA-tag insertion contained a BamHI cut site that was used for screening for positive clones. We obtained several clones with homozygous and heterozygous insertion of the HA-tag. We used a single BfuAI cut site in the AT-hook region to screen for positive AT-hook deletion clones and from 129 clones obtained 10 homozygous knockouts of the AT-hook in E14 cells.

Cell Proliferation, Self-renewal assay and Alkaline phosphatase staining

Growth assay was done with $1x10^4$ cells seeded in gelatin coated six-wells plate and monitored for the next 5 days in the respective culture condition (i.e. naïve and primed). Cell counting was done every other day using hemocytometer and viability was checked using trypan blue at the time of counting. Self-renewal and alkaline-phosphatase staining assays were done with 200 cells seeded in gelatin coated 12-wells plate and maintained in naïve media condition. Colonies were counted and stained with alkaline phosphatase (ALP) after 5 days. ALP staining was done following manufacturer's protocol (SBI-AP100B-1). Media was changed every other day.

Immunofluorescence Staining

Cells were seeded in poly-ornithine (EMD-Millipore) and laminin (Sigma) coated 8-chambered slide (ibidi; cat#80806) and maintained in naïve and differentiation media independently. Next, cells were fixed with 4%-paraformaldehyde for 15 min at room temperature (RT) and washed with 1XPBS and permeabilized with 0.5% Triton for 5 min at RT. Blocking was done for 30 min at RT with 5% bovine serum albumin (Sigma). Primary antibodies were used with recommended dilutions and incubated for overnight at 4°C. After primary antibody incubation, cells were washed with 1XPBS (Phosphate buffer solution) and 1XPBST (PBS + 0.1%Tween20) and then follow secondary antibody incubation. For immunostaining, antibodies including anti-Oct4 (abcam; ab107156), anti-Sox2 (abcam; ab107156), anti-Nanog (abcam; ab107156), anti-Brg1/Smarca4 (abcam; ab110641), anti-Sox1 (CST; 4194S), anti-Gata4 (scbt; sc-25310) were used. Images were captured using a LSM-880 confocal microscope (Zeiss) and processed with Zen-blue software.

RNA isolation and quantitative real time PCR (qRT-PCR)

Total RNA was isolated using Trizol (Invitrogen), following the manufacturer's protocol. One mg of RNA was used to prepare cDNA and were synthesized using the iScript kit (Bio-rad) according to the manufacturer's protocol. For each biological replicate, quantitative PCR reactions were performed in technical triplicates using the iTaq Universal Sybr Green Supermix (Bio-Rad) on the Bio-Rad CFX-96 Real-Time PCR System, and the data normalized to *Gapdh*. Primers used in this study are summarized in Table 2.

Immunopurification of wild type and DAT mutant Smarca4/Brg1 complexes

Purification of HA-tag protein was performed following a previously described protocol ([61]). Briefly, cells (~ 2–3 x 10^8) were cultured in gelatin coated plates and maintained in naïve and primed media independently. The packed cell volume (PCV) was estimated after cell harvesting and gently resuspended in buffer with 10mM HEPES-KOH pH 7.9, 10mM KCl, 1.5mM $MgCl_2$ 1mM DTT, 1mM PMSF, 1uM pepstatin, 10uM chymostatin. Next, the cell suspension was transferred to a Dounce homogenizer fitted with a B-type pestle and the cells lysed with 20 strokes followed by centrifugation for 5 min at 900 x g. After removing the supernatant, the nuclei pellet was resuspended in

buffer containing 0.2mM EDTA, 20% Glycerol, 20mM HEPES-KOH pH 7.9, 420mM NaCl, 1.5mM $MgCl_2$, 1mM DTT, 1mM PMSF, 1uM pepstatin, 10uM chymostatin, protease inhibitor cocktail and the pellet lysed by gentle Dounce homogenization (i.e., 10–20 strokes with type-B pestle). The tube was mounted on a vortex mixer and agitated very gently for 30 minutes to 1hr at 4°C degree and centrifuged for 15 minutes at 20,000 x g and the supernatant collected. Nuclear lysates were diluted with two-thirds of original volume of 20mM HEPES, pH 7.9, and 0.3% NP-40 to adjust to an appropriate NaCl concentration. Anti-HA agarose beads were used following the manufacturer's protocol [62] and the nuclear extract was incubated with HA-beads overnight at 4ºC with gentle end-over-end mixing or on a rocking platform. Next, beads were washed three times with wash buffer (50mM Tris pH 8.0, 150mM NaCl, 1mM EDTA, 10% Glycerol, and 0.5% Triton X-100) and one bed volume of HA-peptide was added to the beads and incubated at 4ºC for 4 hr prior for elution of the complex. Untagged cells were lysed using 1X lysis buffer (CST) following manufacturer's protocol to prepare the whole cell extract. The efficiency of the immunoprecipitation was checked by Western blotting following standard protocols with the following primary antibodies anti-Brg1/Smarca4 (abcam; ab110641), anti-HA (Invitrogen; Cat #26183), anti-Tubulin (ThermoFisher; A11126), and anti-Gapdh (CST; 2118). The immunoblot were visualized using Super Signal Pico chemiluminescent reagent.

PRO-seq

PRO-seq was performed as previously described with minor modifications[39]. Briefly, nuclei were isolated using Dounce homogenizer (1 million cells per mL) and nuclear run-on was performed with all four biotin-NTPs. RNAs were extracted by Trizol LS (Ambion) and fragmentated by base hydrolysis. From the fragmentated RNAs, biotin RNAs were enriched by streptavidin beads. The biotin RNAs were enriched twice more in each 3' and 5' RNA adaptor ligation processes. Reverse transcription, PCR amplification, and library size selection were performed to obtain 140 ~ 350bp, 5ng/uL libraries. These libraries were submitted for sequencing (75bp, single-endreads) on an Illumina NextSeq 500.

ChIP-seq

Chromatin immunoprecipitation (ChIP) was performed following previously described high-throughput ChIP protocol with some modifications[63]. Cells were crosslinked in 1% formaldehyde for 10 min at room temperature, before quenching with 125 mM glycine for 5 min. After 2 washes with ice cold PBS, cells were incubated in lysis buffer (5 mM PIPES pH 8.0, 85 mM KCl, 0.5% NP-40 supplemented with protease inhibitor) for 10 min, and nuclei were collected after centrifugation. The nuclear pellet was re-suspended in shearing buffer (12 mM Tris-HCl pH 7.5, 6 mM EDTA pH 8.0, 0.5% SDS supplemented with protease inhibitor) and chromatin was fragmented using ME220 focused ultra-sonicator (Covaris) to obtain DNA fragments ranging 200–600 bp. The chromatin lysate was collected after centrifugation and incubated overnight at 4°C with Brg1 and respective histone antibodies conjugated with Dynabeads Protein G (Invitrogen). Next day, antibody-bound DNA were collected using Dynamag, washed extensively as described in the protocol, treated with RNase and Proteinase K, and reverse crosslinked overnight followed by DNA extraction using Ampure X beads (Beckman Coulter). Purified ChIP DNA was used for library construction using NEB Ultra II DNA library prep kit (New England Biolabs) and submitted for sequencing (75bp paired end and 50 bp single reads) on an Illumina HiSeq3000.

ATAC-seq

ATAC-seq was performed as previously described[64]. Briefly, 50,000 cells were washed with cold PBS, collected by centrifugation then resuspended in resuspension buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2). After collection, cells were lysed in lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% NP-40) and collected before incubating in transposition mix containing Tn5 transposase (Illumina). Purified DNA was then ligated with adapters, amplified and size selected for sequencing. Library DNA was sequenced with paired end 50 bp reads.

FASP Methods – Orbitrap Exploris DIA

Protein samples were reduced, alkylated, and digested using filter-aided sample preparation with sequencing grade modified porcine trypsin (Promega). Tryptic peptides were then separated by reverse phase XSelect CSH C18 2.5 um resin (Waters) on an in-line 150 x 0.075 mm column using an UltiMate 3000 RSLCnano system (Thermo). Peptides were eluted using a 60 min gradient from 98:2 to 65:35 buffer A:B ratio (Buffer A = 0.1% formic acid, 0.5% acetonitrile; Buffer B = 0.1% formic acid, 99.9% acetonitrile). Eluted peptides were ionized by electrospray (2.2 kV) followed by mass spectrometric analysis on an Orbitrap Exploris 480 mass spectrometer (Thermo). To assemble a chromatogram library, six gas-phase fractions were acquired on the Orbitrap Exploris with 4 m/z DIA spectra (4 m/z precursor isolation windows at 30,000 resolution, normalized AGC target 100%, maximum inject time 66 ms) using a staggered window pattern from narrow mass ranges using optimized window placements. Precursor spectra were acquired after each DIA duty cycle, spanning the m/z range of the gas-phase fraction (i.e., 496–602 m/z, 60,000 resolution, normalized AGC target 100%, maximum injection time 50 ms). For wide-window acquisitions, the Orbitrap Exploris was configured to acquire a precursor scan (385–1015 m/z, 60,000 resolution, normalized AGC

target 100%, maximum injection time 50 ms) followed by 50x 12 m/z DIA spectra (12 m/z precursor isolation windows at 15,000 resolution, normalized AGC target 100%, maximum injection time 33 ms) using a staggered window pattern with optimized window placements. Precursor spectra were acquired after each DIA duty cycle.

Data-analysis

MNase-seq

Paired end sequence reads were mapped to sacCer3 reference genome using Bowtie for Illumina version 2.4.1. Reads that are mapped to the repetitive rRNA locus (chrXII: 451275–469084) were filtered out. Dyad density maps were obtained for the mapped reads by considering the center of the paired sequence reads after normalizing to the mean genome wide coverage. Stereotypic nucleosome positions are identified using a greedy algorithm in a similar fashion as described for Gene Track [65]. These nucleosome calls include standard deviation (fuzziness) and nucleosome occupancy values. The data was processed using web based genomic data analysis software "Galaxy" [66–69] and additional resources used can be found at http://palpant.us/java-genomics-toolkit/. Further processing of data was performed in Matlab Ver. 7.12.0.635 (R2011a). Called nucleosomes that showed two or more-fold reduction or increase in nucleosome occupancy between WT Snf2 and ΔAT hook, or ΔSnAC or catalytic dead were aligned back to the transcription start and end sites (TSS/TES). A total of 6,289 genes with well-defined transcription start and end sites were used [70,71]. Genes were aligned to each other based on the position of the + 1 nucleosomes and a plot generated to display the extent of change in nucleosome occupancy at each position.

PRO-seq analysis

Adaptor trimming and low-quality reads were removed by Cutadapt[72]. The filtered reads were aligned on mm10 genome by bowtie2[73] with "–very-sensitive" option to discard reads mapped to more than one region. The mapped reads were compressed as binary form using samtools[74] and rRNAs were removed from those reads by bedtools intersect[75]. The 3' end of the filtered reads were captured and in strand-specific manner using bedtools genomecov. These files were used to calculate raw readcount of annotated genes by bedtools map. Similarity between biological replicates and two AT hook deletion clones, and difference between samples were confirmed by pearson correlation with clustering and PCA analysis using R package DESeq2[76]. After the validation of replicates, all replicates were merged and normalized to meet final 30 million reads. The normalized reads were used to obtain differential genes through log2 fold-change calculation, create MA plots using R package ggplot2[77] and perform meta-plot analysis using deepTools2[78] with gaussian smoothing by R package Smoother[79]. Statistically significant gene ontology (GO) terms for each differential genes were annotated by GO enrichment analysis (Adaptor trimming and low-quality reads were removed by Cutadapt[72]. The filtered reads were aligned on mm10 genome by bowtie2[73] with "–very-sensitive" option to discard reads mapped to more than one region. The mapped reads were compressed as binary form using samtools[74] and rRNAs were removed from those reads by bedtools intersect[75]. The 3' end of the filtered reads were captured and in strand-specific manner using bedtools genomecov. These files were used to calculate raw readcount of annotated genes by bedtools map. Similarity between biological replicates and two AT hook deletion clones, and difference between samples were confirmed by pearson correlation with clustering and PCA analysis using R package DESeq2[76]. After the validation of replicates, all replicates were merged and normalized to meet final 30 million reads. The normalized reads were used to obtain differential genes through log2 fold-change calculation, create MA plots using R package ggplot2[77] and perform meta-plot analysis using deepTools2[78] with gaussian smoothing by R package Smoother[79]. Statistically significant gene ontology (GO) terms for each differential genes were annotated by GO enrichment analysis (http://geneontology.org/)[80–82].)[80–82].

Gene list identification

Annotation of TSSs and TESs for 55487 genes are downloaded from Gencode vM22 and active TSSs for mESC identified through START-seq are provided by Dr. Karen Adelman's lab[83]. If active TSS located within ± 1kb region from the TSS of the 55487 genes, the TSS of this gene is replaced by the position of this active TSS. Genes with length (TSS-TES) < 2kb and without PRO-seq signal at either promoter proximal pausing (TSS-100bp + TSS + 300bp) or early elongating regions (TSS + 300bp ~ TSS + 2kbs) were removed to avoid genes with unclear transcripts. Thus, 15265 genes with the length of transcription (TSS-TTS) > 2kb with valid expression are remained for further study.

ATAC-seq and ChIP-seq

Data from two biological replicates were first compared ($R^2 > 0.9$), and then merged into a single read file for each time point. ATAC-seq peaks (were then called using MACS2 [84] with the following parameters: -q 0.01–nomodel–shift 75 –extsize 150. To get a union set of peaks from all samples (WT and dAT mutants), MACS2 peaks from each condition were merged using mergePeaks module from HOMER[85] (default parameters). For identifying differentially accessible regions, the union set of peaks was annotated by Homer and then divided into promoter (-1 kb to + 1 kb), and intronic-intergenic regions. Read counts for all peaks in the union set were obtained using the featureCount module of Subread package [86] and were differential anlysis was done using edgeR [87]. Heatmaps were created using deeptools[78] plotHeatmap function.Data from two biological replicates were first compared ($R^2 > 0.9$), and then merged into a single read file for each time point. ATAC-seq peaks (were then called using MACS2 [84] with the following parameters: -q 0.01–nomodel–shift 75 –extsize 150. To get a union set of peaks from all samples (WT and dAT mutants), MACS2 peaks from each condition were merged using mergePeaks module from HOMER[85] (default parameters). For identifying differentially accessible regions, the union set of peaks was annotated by Homer and then divided into promoter (-1 kb to + 1 kb), and intronic-intergenic regions. Read counts for all peaks in the union set were obtained using the featureCount module of Subread package [86] and were differential anlysis was done using edgeR [87]. Heatmaps were created using deeptools[78] plotHeatmap function.

To determine the known motif enrichment findMotifsGenome module of the HOMER package was used. For motif heatmap analysis, differential ATAC-seq peaks between WT and dAT mutants were used to identify the known motifs using 'findMotifsGenome.pl' from Homer. Then, p-value of identified motifs were transformed into Z-score and plotted as a heatmaps using the R ggplot package. For ATAC footprint analysis, normalized ATAC files were corrected using 'ATACorrect' module from TOBIAS ([30]). Next, the average ATAC-signals were calculated (around +/- 100bp of the center of the motif enriched peaks) and plotted using plotProfile from deeptools.

For ChIP-seq analysis, Paired-end 75 bp reads were aligned to mm10 using Bowtie 2[73] alignment tool using *–very-sensitive-local* preset parameter. Data from two biological replicates were first compared to check for concordance ($R^2 > 0.9$), and then merged into a single read file for each cell type for further downstream analysis. The high confidence peak sets were selected from biological replicates using the intersectBed function from BEDTools[88] with default parameters. For histone and Brg1 ChIP, peaks were called using MACS2[84] with the following parameters: –broad -q 0.05 –nomodel -extsize 500. To compare the signals between IP and input, 'bdgcmp' from MACS2 option was used. Peaks were annotated using the annotatePeaks module of HOMER package[85]. All the heatmaps were drawn using 'plotHeatmap' from deeptools[78].For ChIP-seq analysis, Paired-end 75 bp reads were aligned to mm10 using Bowtie 2[73] alignment tool using *–very-sensitive-local* preset parameter. Data from two biological replicates were first compared to check for concordance ($R^2 > 0.9$), and then merged into a single read file for each cell type for further downstream analysis. The high confidence peak sets were selected from biological replicates using the intersectBed function from BEDTools[88] with default parameters. For histone and Brg1 ChIP, peaks were called using MACS2[84] with the following parameters: –broad -q 0.05 –nomodel -extsize 500. To compare the signals between IP and input, 'bdgcmp' from MACS2 option was used. Peaks were annotated using the annotatePeaks module of HOMER package[85]. All the heatmaps were drawn using 'plotHeatmap' from deeptools[78].

Mass spectrometry

Following data acquisition, the data was searched using an empirically corrected library and a quantitative analysis was performed to obtain a comprehensive proteomic profile. Proteins were identified and quantified using EncyclopeDIA and visualized with Scaffold DIA using 1% false discovery thresholds at both the protein and peptide level[89] .The UniProt database for Mus musculus was used for the database search. Protein exclusive intensity values were assessed for quality using ProteiNorm, a user-friendly tool for a systematic evaluation of normalization methods, imputation of missing values and comparisons of different differential abundance methods [90]. Popular normalization methods were evaluated including log2 normalization (Log2), median normalization (Median), mean normalization (Mean), variance stabilizing normalization[91, 92], quantile normalization (Quantile), Cyclic loess normalization (Cyclic Loess), global robust linear regression normalization (RLR), and global intensity normalization (Global Intensity). The individual performance of each method was evaluated by comparing of the following metrices: total intensity, Pooled intragroup Coefficient of Variation (PCV), Pooled intragroup Median Absolute Deviation (PMAD), Pooled intragroup estimate of variance (PEV), intragroup correlation, sample correlation heatmap (Pearson), and log2-ratio distributions. The data was normalized using Cyclic Loess and statistical analysis was performed using Linear Models for Microarray Data (limma) with empirical Bayes (eBayes) smoothing to the standard errors [91].[91]. Proteins with an FDR adjusted p-value < 0.05 and a fold change > 2 were considered to be significant.

Table 2

List of primers

Organism	Description	Name	primers (5'-3')
Yeast	Delete both -N and -C terminal AT-hooks	Snf2 ΔAT hook F	tctaatcgcgactttctgctattttcacgactttcgattaattatctgcc atgaacataccacagcgtcaatttagcaacgaa
		Snf2 ΔAT hook r1	taatatcaataggactcgttgccgcagcgggcgtgcggacataatcaag tgcctttctctttttagacaagaaatcatc
		Snf2 ΔAT hook f2	cggatttggcaatgaatgacgatgatttcttgtctaaaaagagaaaggca cttgattatgtccgcacgcccgctgcggca
		Snf2 ΔAT hook R	tgtttgtctacgtataaacgaataagtacttatattgctttaggaaggtatgacgtCTAagaaaccattattatcatgac
	Delete -N terminal AT-hook	Snf2 ΔAT hook F	tctaatcgcgactttctgctattttcacgactttcgattaattatctgcc ATGaacataccacagcgtcaatttagcaacgaa
		Snf2 ΔAT hook 1 r1	ccagtaactggacttgattcaagcgctggtggttcagaattttcagat tgcctttctctttttagacaagaaatcatc
		Snf2 ΔAT hook 1 f2	cggatttggcaatgaatgacgatgatttcttgtctaaaaagagaaaggca ggatctgaaaattctgaaccaccagcgctt
		Snf2 ΔAT hook R	tgtttgtctacgtataaacgaataagtacttatattgctttaggaaggtatgacgtCTAagaaaccattattatcatgac
Mouse	Ectoderm lineage	Cdh2_Fwd	CAGGGTGGACGTCATTGTAG
		Cdh2_Rev	AGGGTCTCCACCACTGATTC
		Otx2_Fwd	CTTCATGAGGGAAGAGGTGG
		Otx2_Rev	GGCCTCACTTTGTTCTGACC
		Sox1_Fwd	CCTCGGATCTCTGGTCAAGT
		Sox1_Rev	GCAGGTACATGCTGATCATCTC
		Pax6_Fwd	AGTGAATGGGCGGAGTTATG
		Pax6_Rev	ACTTGGACGGGAACTGACAC
		Nestin_Fwd	CCCTGAAGTCGAGGAGCTG
		Nestin_Rev	CTGCTGCACCTCTAAGCGA
		Fgf5_Fwd	GCGATCCACAGAACTGAAAA
		Fgf5_Rev	ACTGCTTGAACCTGGGTAGG
	Mesoderm lineage	Gsc_Fwd	GCCACCGTACCATCTTCAGC
		Gsc_Rev	TACGTCGGGATACTGGTTCTG
		Brach_Fwd	CTGGGAGCTCAGTTCTTTCG
		Brach_Rev	CCCCTTCATACATCGGAGAA
		FoxA2_Fwd	GAGCAGCAACATCACCACAG
		FoxA2_Rev	CGTAGGCCTTGAGGTCCAT
	Endoderm lineage	Gata6_Fwd	CAAAAGCTTGCTCCGGTAAC
		Gata6_Rev	TGAGGTGGTCGCTTGTGTAG
		Gata4_Fwd	TCTCACTATGGGCACAGCAG
		Gata4_Rev	GCGATGTCTGAGTGACAGGA
		Sox17_Fwd	GCTTCTCTGCCAAGGTCAAC
		Sox17_Rev	CTCGGGGATGTAAAGGTGAA

control	Gapdh_Fwd_set3	TGACCACAGTCCATGCCATC
	Gapdh_Rev_set3	GACGGACACATTGGGGGTAG

Declarations

Funding and additional information

Funding support was provided by the National Institutes of Health (grant no.: R01GM136974 to J. A. R and grant number R01GM108908 to B.B.). NGS sequencing was performed at the Science Park NGS Core, with support from CPRIT Core Facility Support Grant RP120348. Mass spectrometry was done at the University of Arkansas Medical Center Proteomics Core with support from NIH grant R24GM137786. Immunofluorescence was done at the Science Park imaging core, with support from FCCIC Core and FCCIC CPRIT grant RP170628. DS and SH received support from the Center of Cancer Epigenetics at MD Anderson Cancer Center.

Author Contributions

The yeast Snf2 mutant strains were prepared and SWI/SNF complexes purified by P.S. and S.H. and the remodeling and ATPase assays were done by A.H and P.S. The spot assays were done by J. Lee and the MNase-seq by S.H. and J. Lee. The CX-MS were conducted by S.H and J.L. with assistance from J. A. R. K.F. prepared the PRO-seq, ATAC-seq and Brg1 ChIP-seq samples. The H3K4me1, H3K4me3 and H3K27ac samples were prepared by K.F., D.S and A.J. Cell imaging, qRT-PCR and Brg1 complex purification experiments were done by D.S. J.L performed the bioinformatic analysis of PRO-seq and D.S. did the bioinformatic analysis for ChIP-seq, and ATAC-seq. Y.C.L made and characterized all the mESC clones with HA tag and knock-in of the AT-hook deletion using CRISPR-Cas9. Y.L. did all the initial processing of NGS data. B.L. assisted and supervised the bioinformatic analysis and B.B. supervised this work. B.B., D.S., J.L., J.A.R and P.S. assisted in the manuscript preparation.

Acknowledgements

NGS sequencing was performed at the Science Park NGS Core, with support from CPRIT Core Facility Support Grant RP120348. Mass spectrometry was done at the University of Arkansas Medical Center Proteomics Core with support from NIH grant R24GM137786. Immunofluorescence was done at the Science Park imaging core, with support from FCCIC Core and FCCIC CPRIT grant RP170628. This work was supported by NIH grant R01GM108908. DS received support from the Center of Cancer Epigenetics at MD Anderson Cancer Center and SA from the Cockrell Foundation.

References

1. Aravind, L. & Landsman, D. AT-hook motifs identified in a wide variety of DNA-binding proteins. Nucleic Acids Res **26**, 4413–21 (1998).
2. Filarsky, M. et al. The extended AT-hook is a novel RNA binding motif. RNA Biol **12**, 864–76 (2015).
3. Huth, J.R. et al. The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif. Nat Struct Biol **4**, 657–65 (1997).
4. Bewley, C.A., Gronenborn, A.M. & Clore, G.M. Minor groove-binding architectural proteins: structure, function, and DNA recognition. Annu Rev Biophys Biomol Struct **27**, 105–31 (1998).
5. Reeves, R. & Nissen, M.S. The A.T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. J Biol Chem **265**, 8573–82 (1990).
6. Dickinson, L.A. & Kohwi-Shigematsu, T. Nucleolin is a matrix attachment region DNA-binding protein that specifically recognizes a region with high base-unpairing potential. Mol Cell Biol **15**, 456–65 (1995).
7. Sears, J. et al. The amino terminus of Epstein-Barr Virus (EBV) nuclear antigen 1 contains AT hooks that facilitate the replication and partitioning of latent EBV genomes by tethering them to cellular chromosomes. J Virol **78**, 11487–505 (2004).
8. Benecke, A.G. & Eilebrecht, S. RNA-Mediated Regulation of HMGA1 Function. Biomolecules **5**, 943–57 (2015).
9. Strohner, R. et al. NoRC–a novel member of mammalian ISWI-containing chromatin remodeling machines. EMBO J **20**, 4892–900 (2001).
10. Zillner, K. et al. Large-scale organization of ribosomal DNA chromatin is regulated by Tip5. Nucleic Acids Res **41**, 5251–62 (2013).

11. Zhou, Y. et al. Reversible acetylation of the chromatin remodelling complex NoRC is required for non-coding RNA-dependent silencing. Nat Cell Biol **11**, 1010–6 (2009).
12. Xiao, H. et al. Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions. Mol Cell **8**, 531–43 (2001).
13. Cairns, B.R. et al. Two functionally distinct forms of the RSC nucleosome-remodeling complex, containing essential AT hook, BAH, and bromodomains. Mol Cell **4**, 715–23 (1999).
14. Han, Y., Reyes, A.A., Malik, S. & He, Y. Cryo-EM structure of SWI/SNF complex bound to a nucleosome. Nature **579**, 452–455 (2020).
15. He, Z., Chen, K., Ye, Y. & Chen, Z. Structure of the SWI/SNF complex bound to the nucleosome and insights into the functional modularity. Cell Discov **7**, 28 (2021).
16. Mashtalir, N. et al. Chromatin landscape signals differentially dictate the activities of mSWI/SNF family complexes. Science **373**, 306–315 (2021).
17. Clapier, C.R. & Cairns, B.R. Regulation of ISWI involves inhibitory modules antagonized by nucleosomal epitopes. Nature **492**, 280–4 (2012).
18. Mueller-Planitz, F., Klinker, H., Ludwigsen, J. & Becker, P.B. The ATPase domain of ISWI is an autonomous nucleosome remodeling machine. Nat Struct Mol Biol **20**, 82–9 (2013).
19. Racki, L.R. et al. The histone H4 tail regulates the conformation of the ATP-binding pocket in the SNF2h chromatin remodeling enzyme. J Mol Biol **426**, 2034–44 (2014).
20. Ludwigsen, J. et al. Concerted regulation of ISWI by an autoinhibitory domain and the H4 N-terminal tail. Elife **6**(2017).
21. Angelov, D., Vitolo, J.M., Mutskov, V., Dimitrov, S. & Hayes, J.J. Preferential interaction of the core histone tail domains with linker DNA. Proc Natl Acad Sci U S A **98**, 6599–604 (2001).
22. Lehmann, K. et al. Dynamics of the nucleosomal histone H3 N-terminal tail revealed by high precision single-molecule FRET. Nucleic Acids Res **48**, 1551–1571 (2020).
23. Chatterjee, N. et al. Histone H3 tail acetylation modulates ATP-dependent remodeling through multiple mechanisms. Nucleic Acids Res (2011).
24. Ferreira, H., Flaus, A. & Owen-Hughes, T. Histone modifications influence the action of Snf2 family remodelling enzymes by different mechanisms. J Mol Biol **374**, 563–79 (2007).
25. Agalioti, T., Chen, G. & Thanos, D. Deciphering the transcriptional histone acetylation code for a human gene. *Cell* **111**, 381 – 92 (2002).
26. Hassan, A.H. et al. Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. Cell **111**, 369–79 (2002).
27. Sen, P., Ghosh, S., Pugh, B.F. & Bartholomew, B. A new, highly conserved domain in Swi2/Snf2 is required for SWI/SNF remodeling. Nucleic Acids Res (2011).
28. Smith, C.L. & Peterson, C.L. A conserved Swi2/Snf2 ATPase motif couples ATP hydrolysis to chromatin remodeling. Mol Cell Biol **25**, 5880–92 (2005).
29. (!!! INVALID CITATION !!! (Alpsoy and Dykhuizen, 2018; Gatchalian et al., 2018)).
30. Bentsen, M. et al. ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. Nat Commun **11**, 4267 (2020).
31. Yang, S.H. et al. ZIC3 Controls the Transition from Naive to Primed Pluripotency. Cell Rep **27**, 3215–3227 e6 (2019).
32. Luo, Z. et al. Zic2 is an enhancer-binding factor required for embryonic stem cell specification. Mol Cell **57**, 685–694 (2015).
33. Matsuda, K. et al. ChIP-seq analysis of genomic binding regions of five major transcription factors highlights a central role for ZIC2 in the mouse epiblast stem cell gene regulatory network. Development **144**, 1948–1958 (2017).
34. Acampora, D. et al. Loss of the Otx2-Binding Site in the Nanog Promoter Affects the Integrity of Embryonic Stem Cell Subtypes and Specification of Inner Cell Mass-Derived Epiblast. Cell Rep **15**, 2651–64 (2016).
35. Barisic, D., Stadler, M.B., Iurlaro, M. & Schubeler, D. Mammalian ISWI and SWI/SNF selectively mediate binding of distinct transcription factors. Nature **569**, 136–140 (2019).
36. Ho, L. et al. esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. Nat Cell Biol **13**, 903–13 (2011).

37. King, H.W. & Klose, R.J. The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. Elife **6**(2017).
38. Miller, E.L. et al. TOP2 synergizes with BAF chromatin remodeling for both resolution and formation of facultative heterochromatin. Nat Struct Mol Biol **24**, 344–352 (2017).
39. Mahat, D.B. et al. Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). Nat Protoc **11**, 1455–76 (2016).
40. Tsogtbaatar, E., Landin, C., Minter-Dykhouse, K. & Folmes, C.D.L. Energy Metabolism Regulates Stem Cell Pluripotency. Front Cell Dev Biol **8**, 87 (2020).
41. Stincone, A. et al. The return of metabolism: biochemistry and physiology of the pentose phosphate pathway. Biol Rev Camb Philos Soc **90**, 927–63 (2015).
42. Wang, L. et al. Fatty acid synthesis is critical for stem cell pluripotency via promoting mitochondrial fission. EMBO J **36**, 1330–1347 (2017).
43. Cornacchia, D. et al. Lipid Deprivation Induces a Stable, Naive-to-Primed Intermediate State of Pluripotency in Human PSCs. *Cell Stem Cell* **25**, 120–136 e10 (2019).
44. Tanosaki, S. et al. Fatty Acid Synthesis Is Indispensable for Survival of Human Pluripotent Stem Cells. iScience **23**, 101535 (2020).
45. Ter Huurne, M., Chappell, J., Dalton, S. & Stunnenberg, H.G. Distinct Cell-Cycle Control in Two Different States of Mouse Pluripotency. *Cell Stem Cell* **21**, 449–455 e4 (2017).
46. Dolatshad, N.F., Hellen, N., Jabbour, R.J., Harding, S.E. & Foldes, G. G-protein Coupled Receptor Signaling in Pluripotent Stem Cell-derived Cardiovascular Cells: Implications for Disease Modeling. Front Cell Dev Biol **3**, 76 (2015).
47. Callihan, P., Mumaw, J., Machacek, D.W., Stice, S.L. & Hooks, S.B. Regulation of stem cell pluripotency and differentiation by G protein coupled receptors. Pharmacol Ther **129**, 290–306 (2011).
48. Kidder, B.L., Palmer, S. & Knott, J.G. SWI/SNF-Brg1 regulates self-renewal and occupies core pluripotency-related genes in embryonic stem cells. Stem Cells **27**, 317–28 (2009).
49. Sen, P. et al. The SnAC domain of SWI/SNF is a histone anchor required for remodeling. Mol Cell Biol **33**, 360–70 (2013).
50. Morrison, E.A. et al. DNA binding drives the association of BRG1/hBRM bromodomains with nucleosomes. Nat Commun **8**, 16080 (2017).
51. Alexander, J.M. et al. Brg1 modulates enhancer activation in mesoderm lineage commitment. Development **142**, 1418–30 (2015).
52. Alver, B.H. et al. The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. Nat Commun **8**, 14648 (2017).
53. Blumli, S. et al. Acute depletion of the ARID1A subunit of SWI/SNF complexes reveals distinct pathways for activation and repression of transcription. Cell Rep **37**, 109943 (2021).
54. Dechassa, M.L. et al. Architecture of the SWI/SNF-nucleosome complex. Molecular and cellular biology **28**, 6010–21 (2008).
55. Kassabov, S.R., Henry, N.M., Zofall, M., Tsukiyama, T. & Bartholomew, B. High-resolution mapping of changes in histone-DNA contacts of nucleosomes remodeled by ISW2. Molecular and cellular biology **22**, 7524–34 (2002).
56. Sen, P. et al. Loss of Snf5 Induces Formation of an Aberrant SWI/SNF Complex. Cell Rep **18**, 2135–2147 (2017).
57. Krzywinski, M. et al. Circos: an information aesthetic for comparative genomics. Genome Res **19**, 1639–45 (2009).
58. Gossett, A.J. & Lieb, J.D. In vivo effects of histone H3 depletion on nucleosome occupancy and position in Saccharomyces cerevisiae. PLoS Genet **8**, e1002771 (2012).
59. Cinghu, S. et al. Integrative framework for identification of key cell identity genes uncovers determinants of ES cell identity and homeostasis. Proc Natl Acad Sci U S A **111**, E1581-90 (2014).
60. Oldfield, A.J. et al. Histone-fold domain protein NF-Y promotes chromatin accessibility for cell type-specific master transcription factors. Mol Cell **55**, 708–22 (2014).
61. Gatchalian, J. et al. A non-canonical BRD9-containing BAF chromatin remodeling complex regulates naive pluripotency in mouse embryonic stem cells. Nat Commun **9**, 5139 (2018).
62. Mashtalir, N. et al. A Structural Model of the Endogenous Human BAF Complex Informs Disease Mechanisms. Cell **183**, 802–817 e24 (2020).
63. Blecher-Gonen, R. et al. High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. Nat Protoc **8**, 539–54 (2013).

64. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. & Greenleaf, W.J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods **10**, 1213–8 (2013).
65. Albert, I., Wachi, S., Jiang, C. & Pugh, B.F. GeneTrack–a genomic data processing and visualization framework. Bioinformatics **24**, 1305–6 (2008).
66. (!!! INVALID CITATION !!! 43–45).
67. Giardine, B. et al. Galaxy: a platform for interactive large-scale genome analysis. Genome Res **15**, 1451–5 (2005).
68. Blankenberg, D. et al. Galaxy: a web-based genome analysis tool for experimentalists. *Curr Protoc Mol Biol Chap. 1***9**, Unit 19 10 1–21 (2010).
69. Goecks, J., Nekrutenko, A., Taylor, J. & Galaxy, T. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol **11**, R86 (2010).
70. (!!! INVALID CITATION !!! 46).
71. Nagalakshmi, U. et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science **320**, 1344–9 (2008).
72. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011 **17**, 3 (2011).
73. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. Nat Methods **9**, 357–9 (2012).
74. Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**, 2078–9 (2009).
75. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26**, 841–842 (2010).
76. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol **15**, 550 (2014).
77. Wickham, H., Sievert, C. & SpringerLink. ggplot2: Elegant Graphics for Data Analysis. in *Use R!* 2nd 2016. edn (Springer Springer International Publishing: Imprint: Springer, Cham, 2016).
78. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res **44**, W160-5 (2016).
79. Spiess, A.N. et al. Impact of smoothing on parameter estimation in quantitative DNA amplification experiments. Clin Chem **61**, 379–88 (2015).
80. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet **25**, 25–9 (2000).
81. Consortium, T.G.O. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Research **49**, D325-D334 (2020).
82. Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P.D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Research **47**, D419-D426 (2018).
83. Henriques, T. et al. Widespread transcriptional pausing and elongation control at enhancers. Genes Dev **32**, 26–41 (2018).
84. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol **9**, R137 (2008).
85. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell **38**, 576–89 (2010).
86. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics **30**, 923–30 (2014).
87. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26**, 139–40 (2010).
88. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26**, 841–2 (2010).
89. Searle, B.C. et al. Chromatogram libraries improve peptide detection and quantification by data independent acquisition mass spectrometry. Nat Commun **9**, 5128 (2018).
90. Graw, S. et al. proteiNorm - A User-Friendly Tool for Normalization and Analysis of TMT and Label-Free Protein Quantification. ACS Omega **5**, 25625–25633 (2020).
91. Ritchie, M.E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res **43**, e47 (2015).
92. Chawade, A., Alexandersson, E. & Levander, F. Normalyzer: a tool for rapid evaluation of normalization methods for omics data sets. J Proteome Res **13**, 3114–20 (2014).

Figures

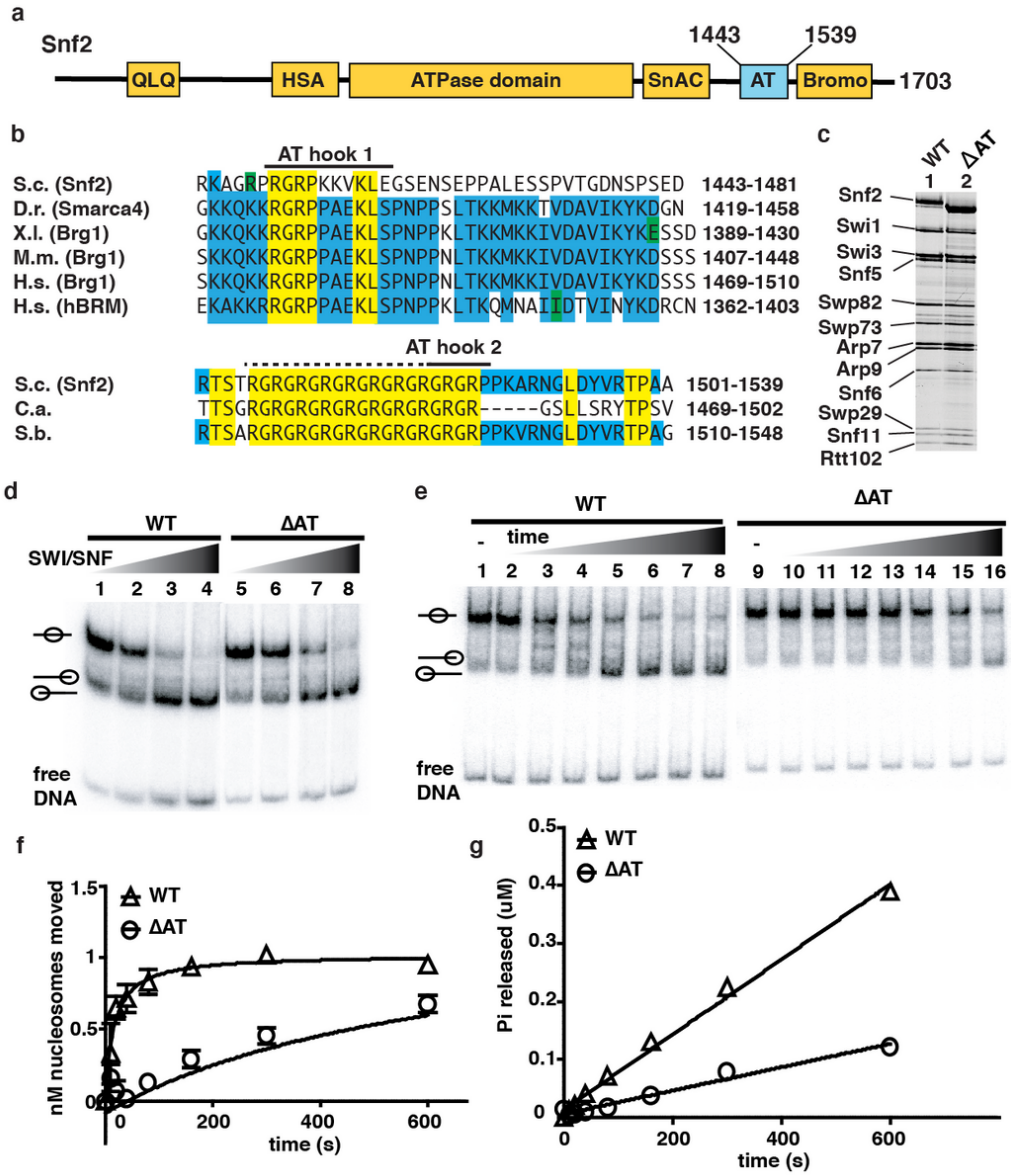

Figure 1

The AT-hook of Snf2 is required for efficient ATP hydrolysis and nucleosome mobilization by the yeast SWI/SNF complex. (a) The domain organization of catalytic subunit of yeast Snf2 contains AT hook motifs at C-terminus. **(b)**The amino acid sequence is shown for the two AT-hooks in Snf2 of *S. cerevisiae* and its homologs in zebra fish (D.r.), Xenopus laevis (X.l.), mouse (M.m), human (H.s.), and two fungal species *Candida albicans*(C.s.) and *Saccharomyces bayanus* (S.b). Conserved residues are highlighted in blue and yellow. **(c)** Wild (WT) and the AT-hooks deletion mutant (DAT) SWI/SNF complexes were immunoaffinity purified and analyzed on a 4-20% gradient SDS-PAGE. **(d)**The nucleosome mobilizing activity of WT and DAT were compared using an electrophoretic mobility shift assay on a 5% native polyacrylamide gel. The reactions contained 6.4 nM 29N59 nucleosomes, yeast SWI/SNF that varied from 1.6 to 13 nM, and 800 mM ATP and were incubated at 30°C for 10 min. **(e)** These remodeling reactions had saturating amounts of SWI/SNF (7.5 nM) relative to nucleosomes (2.5 nM) with limiting ATP (4.4 mM) and incubated for 0 to 10 min. **(f)** The extent of remodeling was quantitated from the gel shown in (d) and plotted relative to the reaction time. **(g)** The rate of ATP hydrolyzed under the same conditions as in (d) were measured by thin layer chromatography and plotted with the amount of inorganic phosphate (Pi) released versus reaction time.

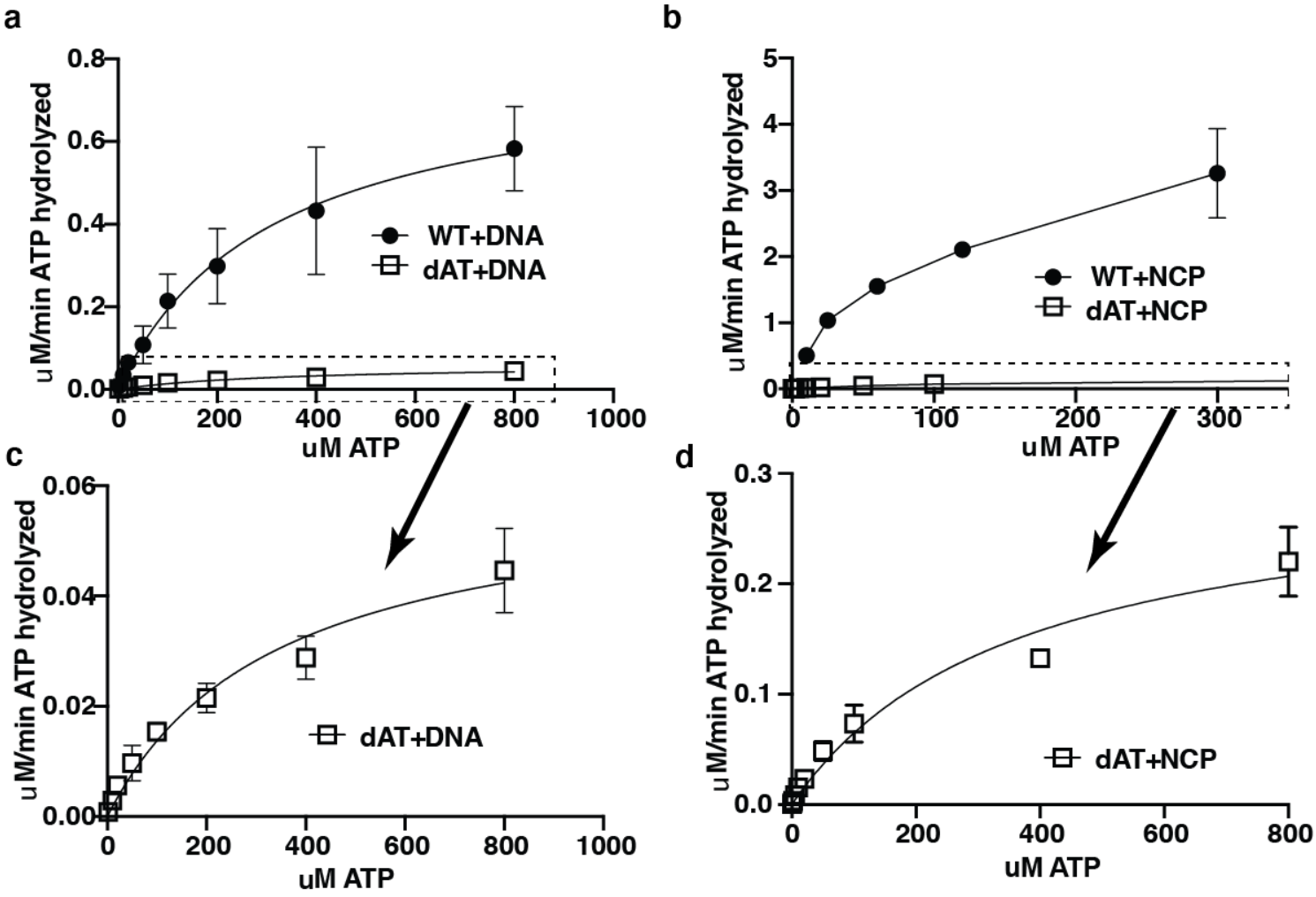

Figure 2

The AT-hooks enhance the reaction velocity of ATP hydrolysis by SWI/SNF in the presence of DNA or nucleosomes.

(a-b) The rate of ATP hydrolysis was measured for SWI/SNF (1.6 nM SWI/SNF) with either (a) free DNA (35 ng pUC18 plasmid DNA) or (b) 29N59 nucleosomes (8 nM) with ATP that varied from 0.2 to 800 mM to determine the maximum velocity (V_{max}) and substrate affinity (K_M) for both wild type and DAT mutant SWI/SNF. (c-d) The data for DAT mutant SWI/SNF is shown with a different y-scale to better visualize the trend not readily seen in (a-b).

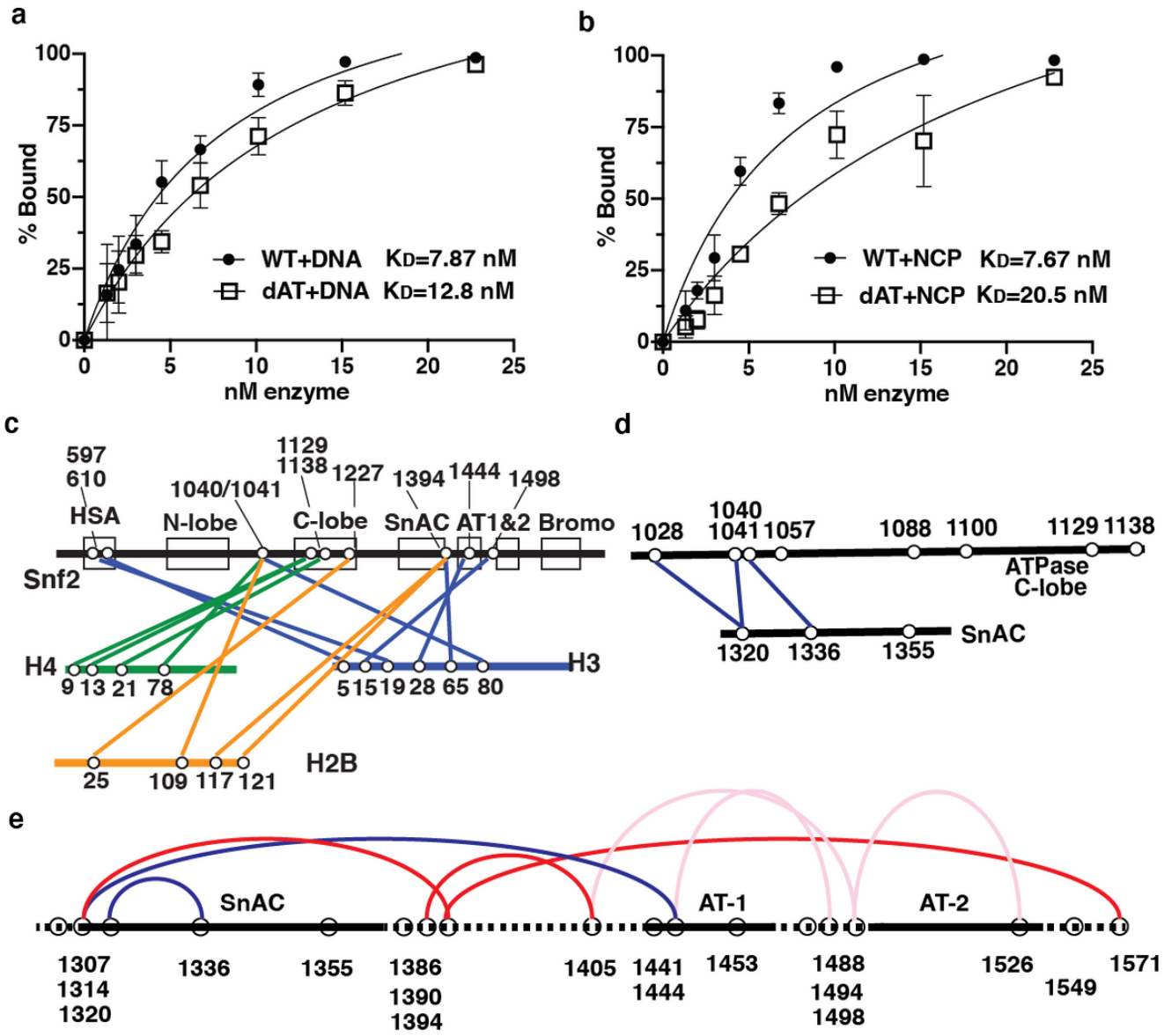

Figure 3

AT-hook involved in SWI/SNF docking onto nucleosomes.

(a-b) The relative affinity of WT and dAT SWI/SNF for (a) free DNA or (b) nucleosomes was measured by EMSA on a 4% native polyacrylamide gel (79:1 acrylamide:bisacrylamide). Reactions contained either 6.7 nM 235 bp 601 DNA or 29N59 nucleosomes and WT and dAT SWI/SNF was varied from 1.3 to 23 nM. (c-d) The Lys-Lys crosslinking pattern between the C-lobe of the ATPase and SnAC domains and the AT-hooks is shown for (d) nucleosome-bound SWI/SNF. The open circles indicate the positions of the crosslinked lysines and the colored lines show the Lys-ys crosslinked pairs. (e) The Lys-Lys crosslinking pattern between Snf2 and histones is shown for nucleosome-bound SWI/SNF, similar to that shown in (c-d).

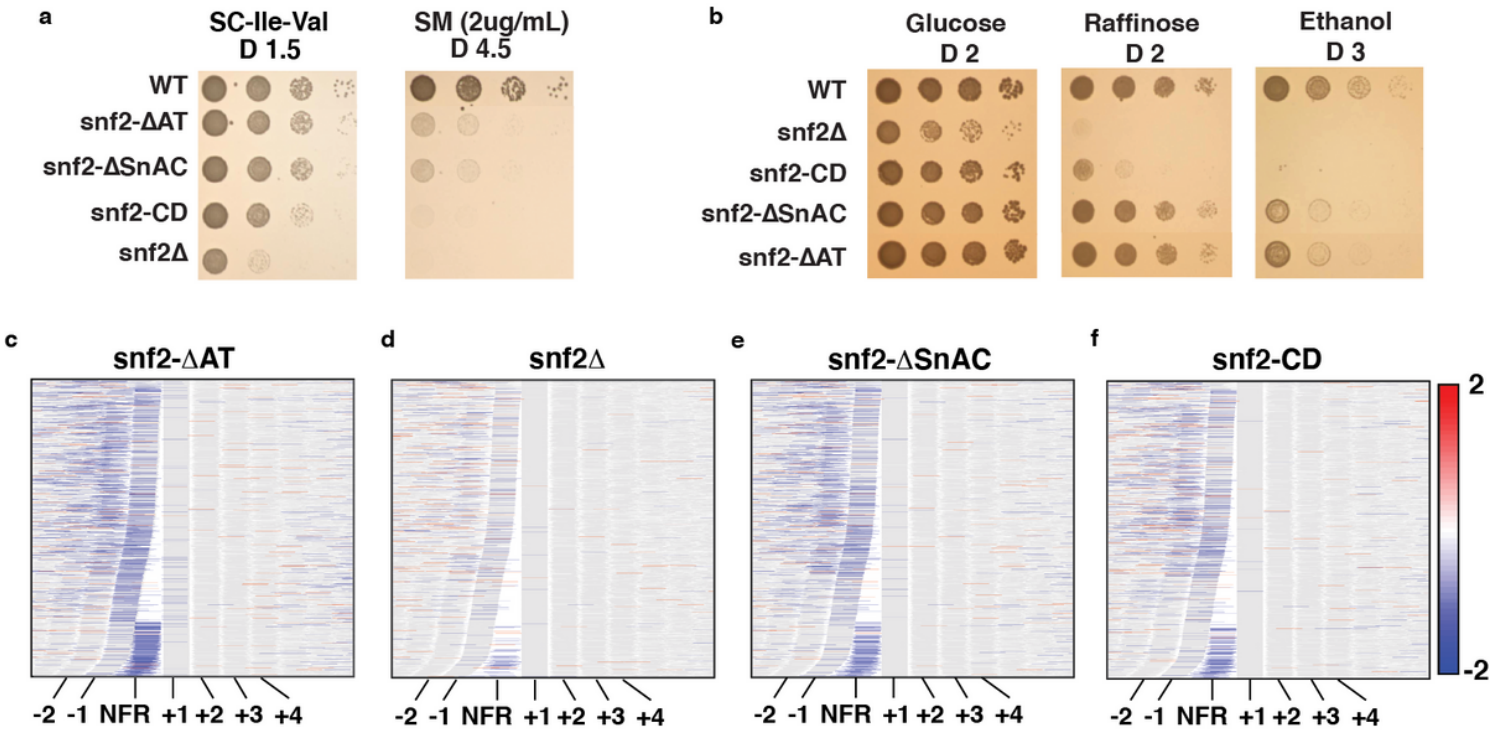

Figure 4

The AT-hooks of Snf2 are required for the in vivo activity of SWI/SNF in yeast.

(a-b) Spot growth assays are shown for determining how deletion of the AT-hooks compare to deletion or other mutations in the Snf2 subunit known to impact the in vivo activity of SWI/SNF. Cells are grown in (a) synthetic media lacking isoleucine and valine (SC-Ile-Val), complete synthetic media plus sulfometuron methyl (SM) or (b) other media with the following carbon sources: ethanol, glucose and raffinose. (c-f) The extent of changes in nucleosome occupancy mapped by MNase-seq are shown for the region surrounding the TSS and the grey highlights the mapped positions of nucleosomes upstream (-1, -2, …) and downstream (+1, +2, …) of the nucleosome-free region or NFR for wild type Snf2. The ratio of nucleosome signal of mutant versus wild type Snf2 is shown for each of these positions with blue representing a loss and red a gain of nucleosomes at that position.

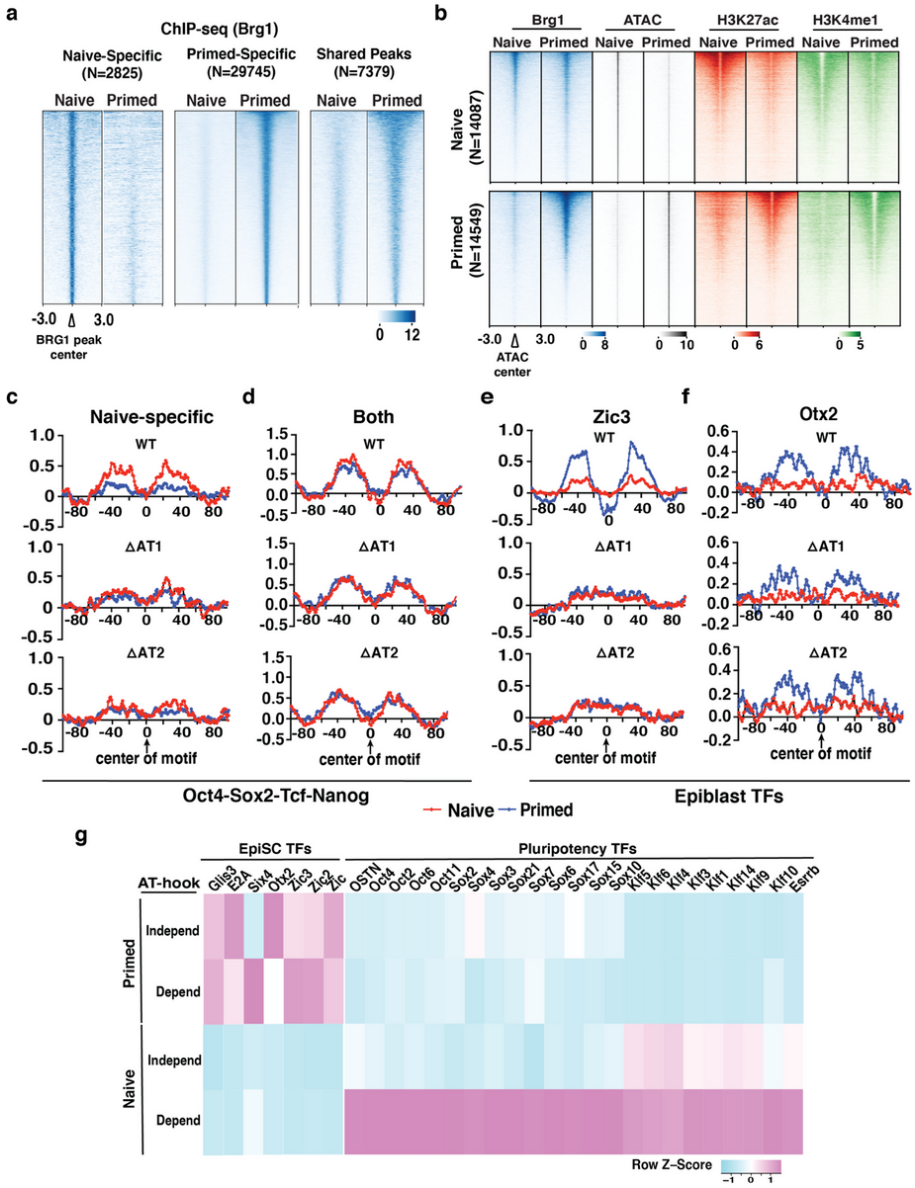

Figure 5

The AT-hook of Brg1 requires stage specific transcription factors binding at the enhancers.

(a) Heatmaps show Brg1 localization in naïve (left), primed (middle) and shared (right) ChIP-seq peaks. (b) Heatmap showing Brg1 localization (blue), ATAC-signals (grey), and active enhancer histone marks (H3K27ac [red], and H3K4me1 [green]) at naïve (top) and primed (bottom) specific intronic-intergenic ATAC-seq peaks. ChIP-seq signals are sorted based on WT Brg1 in each condition. (c-f) DNA foot-printing shows loss of naïve specific pluripotency TFs Oct4-Sox2-Tcf-Nanog (c) and primed/EpiSC specific TF Zic3 (e) binding in the AT-hook deletion mutants. DNA foot-printing of pluripotency Oct4-Sox2-Tcf-Nanog TF (d) and EpiSC specific Otx2 TF (f) binding sites that are unaltered in the dAT mutants. (g) Heatmap showing the pluripotency and epiblast specific (EpiSC) specific transcription factors (TFs) motif enrichment between AT-hook dependent -vs.- AT-hook independent ATAC-seq intronic-intergenic peaks in naïve and primed states.

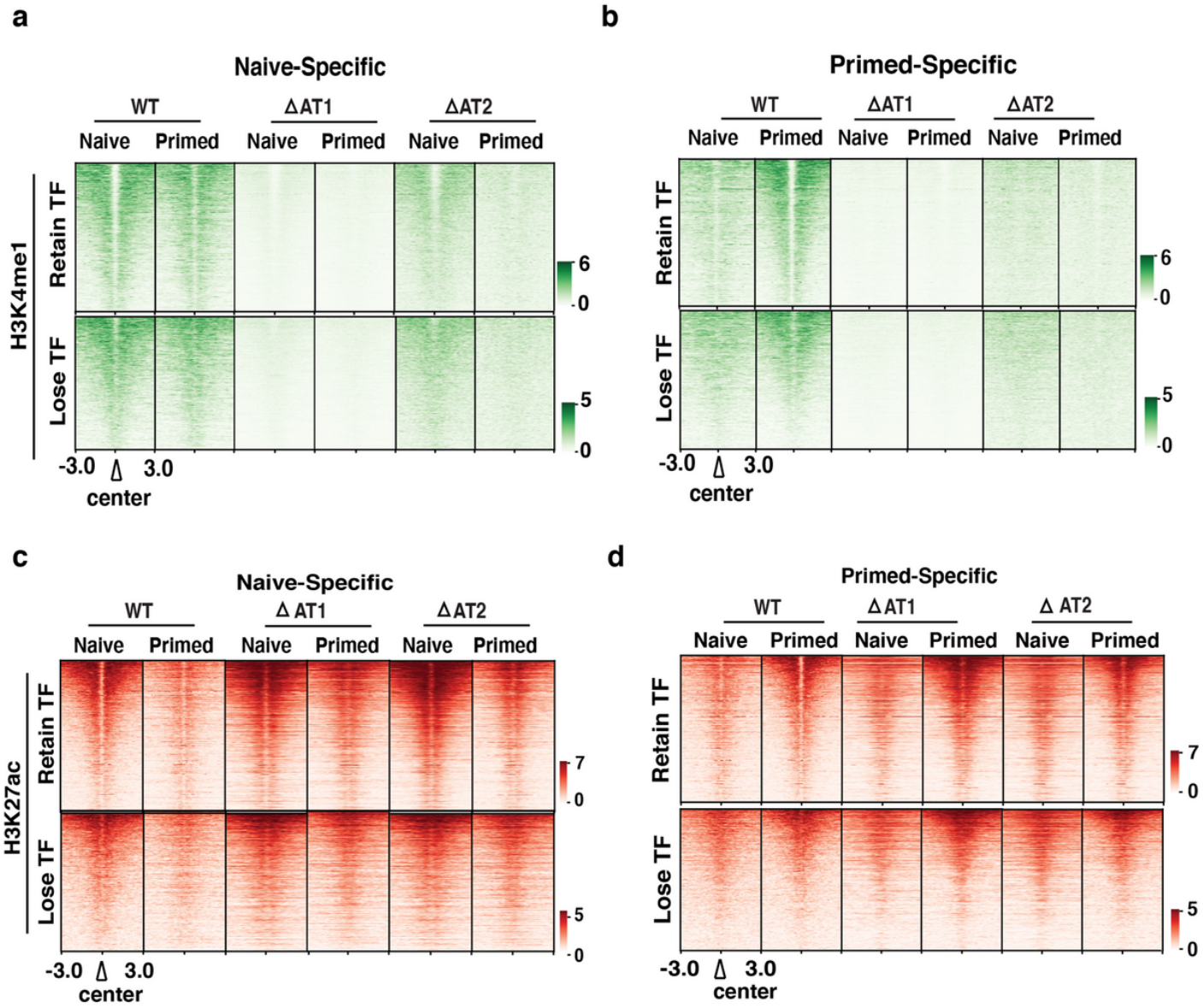

Figure 6

The AT-hook of Brg1 is required for H3K4me1 at stage specific enhancers.

(a-b) Heatmap shows ChIP-seq signals for H3K4me1 at naïve (a) and primed (b) specific enhancers. The regions are divided into those where the ATAC accessibility is either AT-hook dependent (top panel) or independent (bottom panel). ChIP signals are sorted based on WT in each group and on naïve and primed WT in each group. (c-d) Heatmap shows ChIP-seq signals for H3K27ac at naïve (f) and primed (g) specific enhancers. The regions are divided into those where the ATAC accessibility is either AT-hook dependent (top) or independent (bottom). ChIP signals are sorted based on naïve and primed WT in each group, respectively.

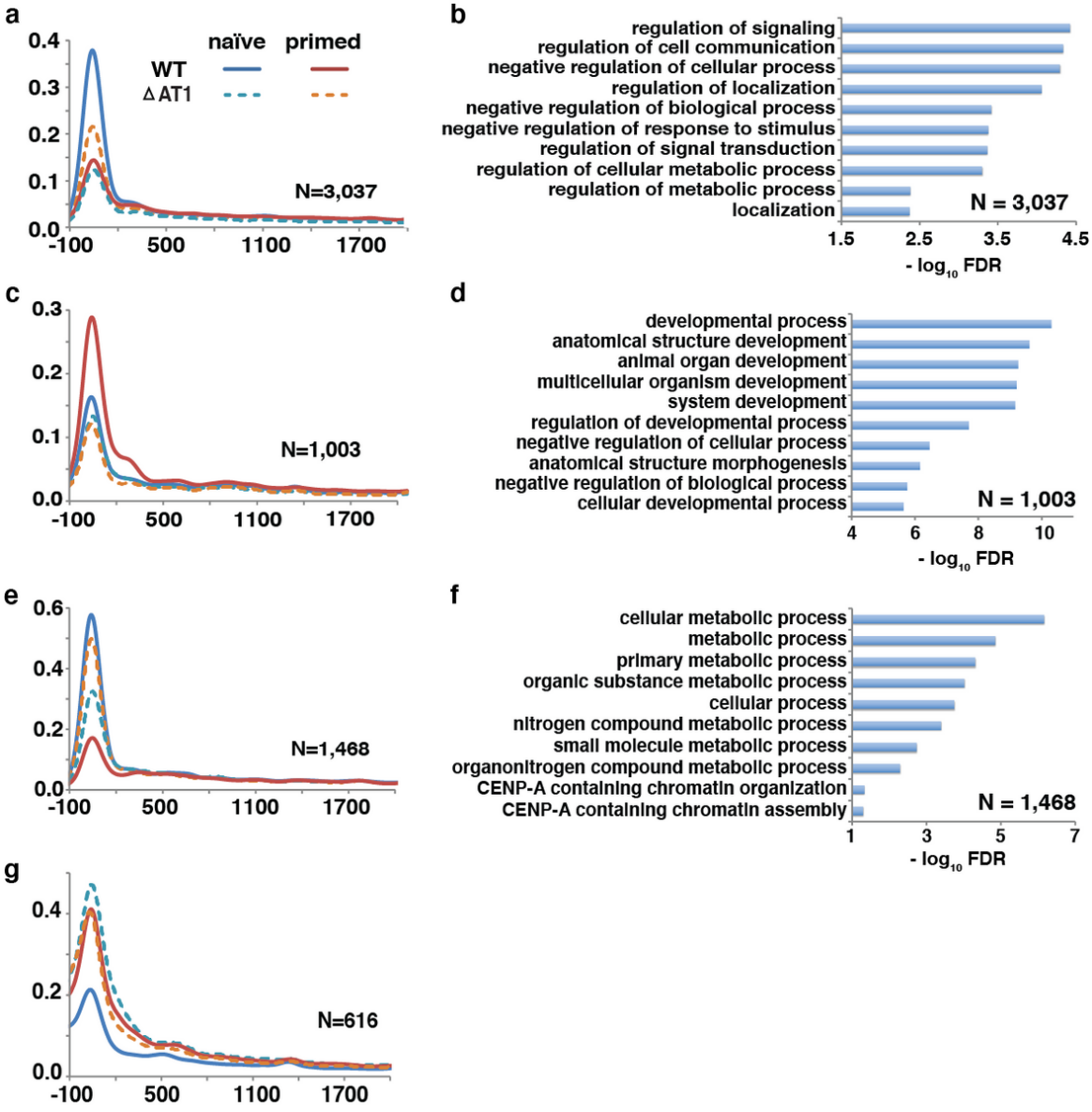

Figure 7

Loss of the AT-hook of Brg1 causes transcription dysregulation in both the naïve and primed states.

(a-d) Meta-analysis of PRO-seq signals (upstream TSS -100 bp to +300 bp downstream of TSS) shows differential pausing genes between WT and dAT1 mutant in naïve (a, b) and primed states (c, d). Paused Up-regulated genes in WT are shown in panel (a) and (c) and panel (b) and (d) show paused genes that are mis regulated in dAT1 mutant in naïve (b) and primed (d) respectively (e-g) Bar graphs show gene ontology (GO) term enriched in WT pausing up-regulated genes in naïve and primed (group-I and group - III) and mis regulated genes in dAT1 (group - II).

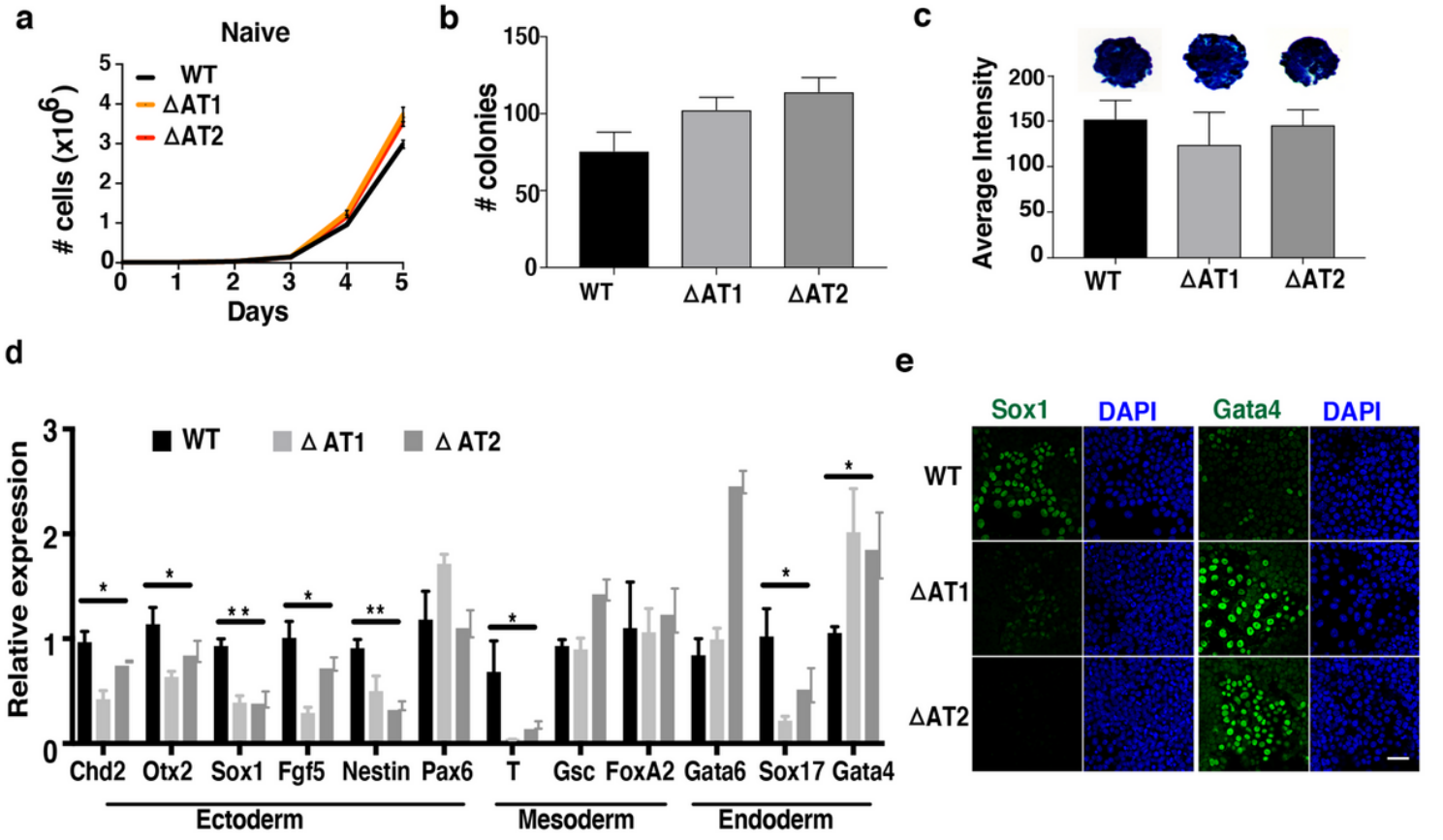

Figure 8

The AT-hook domain of Brg1 is required for cell lineage priming. (a) Growth curves of WT and AT-hook deletion mutant clones (dAT1, and dAT2) cultured in naïve condition are shown. (b-c) Bar graphs showing the number of colonies formed in a self-renewal assay (b), and average signal intensity of the colonies in alkaline phosphatase assay (c) of WT and dAT mESCs in naïve condition. (d) Bar graph shows the expression of lineage-specific markers in WT and dAT mutants after culturing for seven days in the absence of LIF/2i . Gene expression analysis was done by quantitative reverse transcription PCR (qRT-PCR) and the values were normalized with GAPDH. Results are presented as means ± sd (n=3); *p<0.05; **p<0.001; (unpaired student's t-test). (e) Immunofluorescence images showing expression of SOX1 (ectoderm lineage marker) and GATA4 (endoderm lineage marker) in WT and dAT mutants after seven days of LIF/2i withdrawal. Scale bar, 20 microns.

Supplementary Files

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

- SupplementalFigures.docx