SNF2H-MEDIATED CHROMATIN REMODELLING AND ITS REGULATION OF THE PLURIPOTENT STATE

By David Cook

This thesis is submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the M.Sc. program in Cellular and Molecular Medicine

Department of Cellular and Molecular Medicine,
Faculty of Medicine,
University of Ottawa

Submitted August, 2016

Table of Contents

Acknowledgements

I extend nothing but the greatest appreciation and acknowledgement to my supervisor, Dr. Barbara Vanderhyden, who is a constant reminder of what it is to be a great scientist, role model, and human being. I am forever grateful for the scientific freedom and encouragement she has provided during my studies. She is always willing to make herself available for support and advice, and it is clear that she cares deeply about her trainees and their success.

I would also like to acknowledge my wonderful colleagues in the Vanderhyden lab. Our staff, Olga Collins, Elizabeth Macdonald, and Ken Garson, have provided an outstanding foundation of expertise that keep the lab running. I would also like to thank my fellow Vanderhyden trainees, who are always available to lend an ear, a reagent, or advice. I would particularly like to thank Lauren Carter for taking me on as an Honour's student with no lab experience, and having the patience to work with me for a year.

Certainly not least, I am indebted to my loving fiancé, Andrea, for her support and never-ending patience through this journey—encouraging when I decided to attend adult high school seven years ago, and understanding when I leave to go the lab for a 10PM timepoint. She really is my better half and I could not have done this without her.

Abstract

In embryonic stem cells (ESCs), the SWI/SNF, CHD, and INO80 families of ATP-dependent chromatin remodellers have been implicated in maintaining pluripotency-associated gene expression, however the involvement of ISWI family remodellers has yet to be defined. Here, we explore the importance of the mammalian ISWI homologue SNF2H (*Smarca5*) by deriving a conditional knockout mouse ESC line and observing the consequences of SNF2H depletion on the pluripotent state. Cre-mediated deletion of *Snf2h* disrupts hallmark characteristics of pluripotency, resulting in distinct morphological changes; reduced expression of the master transcription factors *Oct4*, *Sox2*, and *Nanog*; and reduced alkaline phosphatase activity.

To understand the mechanisms of SNF2H-mediated regulation, we mapped SNF2H-bound nucleosomes genome-wide. SNF2H is broadly distributed across the genome, but is preferentially enriched at active regulatory regions and transcription factor binding sites. These findings demonstrate the importance of SNF2H in ESCs and shed light on genome-wide mechanisms of transcriptional regulation.

List of Figures

List of Abbreviations

^{o}C	Degrees Celsius
µg	Microgram(s)
µL	Microliter(s)
ACF1	ATP-utilizing chromatin assembly and remodelling factor 1
ANOVA	Analysis of variance
ATP	Adenosine triphosphate
BAF	Barrier-to-autointegration factor
bp	Base pair(s)
BRG1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily A, member 4
CBP	CREB-binding protein
cDNA	Complimentary DNA
Cdx2	Caudal type homeobox 2
CHD	Chromodomain helicase DNA binding protein
CHD1	Chromodomain helicase DNA binding protein 1
ChIP	Chromatin immunoprecipitation
ChIP-Seq	ChIP followed by next-generation sequencing
CHRAC	Chromatin accessibility complex
cm	Centimeter
Cre	Cre recombinase
CTCF	CCCTC-binding factor
d.p.c.	Days post coitum
DAPI	4',6-diamidino-2-phenylindole
ddH_2O	Double-distilled water
DEXDc domain	DEAD-like helicase
DNA	Deoxyribonucleic acid
DR4	Drug resistant 4
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
EBs	Embryoid bodies
EDTA	Ethylenediaminetetracetic acid
EP400	E1A binding protein 400
ESCs	Embryonic stem cells
ESRRB	Estrogen related receptor beta
EZH2	Enhancer of zeste 2 polycomb repressive complex 2 subunit

FACS	Fluorescence-activated cell sorting
FBS	Fetal bovine serum
Fgf5	Fibroblast growth factor 5
FITC	Fluorescein
FPKM	Fragments per kilobase of exon per million mapped reads
g	Grams
gDNA	Genomic DNA
GFF	General feature format
GFP	Green fluorescent protein
H3K27ac	Acetylation of lysine 27 on the tail of histone H3
H3K27me3	Trimethylation of lysine 27 on the tail of histone H3
H3K36me3	Trimethylated lysine 36 on the tail of histone H3
H3K4me1	Monomethylated lysine 4 on the tail of histone H3
H3K4me3	Trimethylation of lysine 4 on the tail of histone H3
H3K9me3	Trimethylation of lysine 9 on the tail of histone H3
HELICc	Helicase superfamily c-terminal domain
hiF	Human induced fibroblast-like cells
HOX	Homeobox
HRP	Horseradish peroxidase
HSA domain	Helicase/SANT-associated domain
IgG	Immunoglobulin G
INO80	INO80 Complex Subunit
iPSCs	Induced pluripotent stem cells
ISWI	Imitation Switch
kb	Kilobase(s)
KLF4	Kruppel-like factor 4
L	Liter
LIF	Leukemia inhibitory factor
lincRNA	Long intergenic non-coding RNA
M	Molar
MEFs	Mouse embryonic fibroblasts
mESCs	Mouse embryonic stem cells
mL	Milliliter
MNase	Micrococcal nuclease
MyoD	Myogenic differentiation 1
NANOG	Nanog homeobox
nL	Nanoliter
NoRC	Nucleolar remodelling complex

NURF	Nucleosome remodelling factor complex
OCT4	Octamer-binding protein 4
ORF	Open reading frame
P300	E1A Binding Protein P300
Pax3	Paired box 3
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
Pgk	Phosphoglycerate kinase-1
PHD domain	Plant homeodomain
PI	Propidium iodide
PIC	Protease inhibitor cocktail
pmol	Picomole
PRC2	Polycomb repressive complex 2
qPCR	Quantitative PCR
RNA	Ribonucleic acid
RNA PolII	RNA Polymerase II
RNA-Seq	Next-generation sequencing of RNA
RNAi	RNA interference
rpm	Reads per million mapped reads
rRNA	Ribosomal RNA
RSF	Remodelling and spacing factor
SchLAP1	Second chromosome locus associated with prostate-1
SDS	Sodium dodecyl sulfate
shRNA	Short hairpin RNA
siRNA	Small interfering RNA
SLIDE domain	SANT-life ISWI domain
SMC1	Structural maintenance of chromosome 1A
SNF2H	Sucrose nonfermenting protein 2 homolog
SNF2L	Sucrose nonfermenting 2-like protein 1
SOX2	(Sex Determining Region Y)-Box 2
STAT3	Signal transducer and activator of transcription 3
SUZ12	Supressor of zeste 12 protein homolog
SWI/SNF	Switch/Sucrose Non-Fermentable
TADs	Topologically associated domains
TBST	Tris-buffered saline with 0.1% Tween-20
TSS	Transcription start site
U	Units
V	Volts

WDR5	WD Repeat Domain 5
WICH	WSTF-ISWI chromatin remodelling complex
Xist	X inactive specific transcript

Introduction

Embryonic stem cells and the pluripotent state

Embryonic stem cells (ESCs) are a self-renewing population of cells derived from the inner cell mass of blastocyst-stage embryos[1]. During embryonic development, the inner cell mass continues to develop into the embryo proper, the amnion, yolk sac, and allantois, while the surrounding trophoblast cells form the fetal components of the placenta. Cultured ESCs maintain the differentiation potential of the inner cell mass, capable of differentiating into the vast array of cell types that comprise the adult body[2]. This capacity is referred to as pluripotency. Several hallmarks of pluripotent cells include the expression of the transcription factors OCT4, SOX2, and NANOG; high alkaline phosphatase activity; a tightly clustered morphology *in vitro*; and the ability to differentiate into embryonic tissues of the three germ layers[3].

Pluripotent cells are a powerful model for studying developmental processes and the regulation of mammalian cell state, and they have also been an integral tool for our understanding of disease. Their intrinsic proliferative and self-renewing characteristics have made them a favourable model system for studying fundamental cell biology. Also, the ability to perform genetic manipulations, re-incorporate these cells back into embryos, and have them contribute to adult tissues[4] has allowed for studying the genetic basis of various diseases, and the roles of thousands of genes *in vivo*[5,6].

Understanding the pluripotent state also holds promise for the field of regenerative medicine. Because of their self-renewal and pluripotency, ESCs could theoretically be used to generate an unlimited source of cells or tissues for transplantation, however, ethical and technical boundaries exist. Previous work to

understand the molecular foundation of ESCs enabled the discovery that the exogenous expression of four genes alone was sufficient to reprogram adult fibroblasts to an induced pluripotent state[7]. This production of induced pluripotent stem cells (iPSCs) breaks down many of the barriers towards the use of pluripotent stem cells in regenerative medicine: the use of human embryos is no longer required, and autologous iPSCs can be derived from patients with no more than a skin biopsy, reducing the risk of graft rejection.

Pluripotency is maintained by a coordinated gene expression program that maintains cells in a proliferative, self-renewing state, while preventing the expression of developmental genes. While full length transcripts for approximately half of all protein-coding genes are detectable by RNA sequencing (RNA-Seq) of ESCs[8], RNA interference (RNAi) screens have repeatedly shown that a subset of several hundred genes are indispensable for the maintenance of a pluripotent state[9–11]. At the core of this network are the transcription factors OCT4, SOX2, NANOG, KLF4, and ESRRB, which form an autoregulatory loop, driving their own expression along with many other actively transcribed genes[12,13]. Chromatin-level regulation also plays a vital role in maintaining ESC identity, involving various histone modifiers, chromatin remodellers, and structural protein complexes. Several of these factors have also been shown to be essential components of the ESC regulatory network, including the Trithorax-group protein WDR5[14]; the histone acetyltransferases P300/CBP[15]; the structural protein CTCF and the Cohesin complex[16,17]; and the nucleosome remodellers BRG1[18], CHD1[19], and INO80[20]. Interestingly, when the components of the polycomb repressive complex 2 (PRC2) are deleted in mouse ESCs (mESCs), the expression of pluripotency factors is not affected, however lineage-associated genes become activated and the cells fail to differentiate

properly[21]. This is likely due to a failure to maintain repression of developmental genes, as well as failure to repress the mESC gene expression program.

It is evident that the orchestration of cell identity gene expression programs is a complex feat involving the coordination of many regulatory factors. Understanding the complexity of global gene regulation in ESCs may provide important insights into mechanisms of mammalian development and cell reprogramming. Importantly, it may also provide broader insight into the general control of cell identity and how this may be misregulated in disease.

Chromatin organization and transcriptional regulation

The epigenome comprises the chemical and structural features that organize the genome within the nucleus[22]. These features include the positioning of histone proteins along DNA[23], covalent modifications to histones[24], covalent modifications to DNA[25], structural organization of chromatin in three-dimensional space[26], and likely more. Together, these coordinate how cues from a cell's intra- and extracellular milieu are translated into transcription output. This output largely defines a cell's state, and its dynamics are capable of driving transitions between states, for example to push a stem cell through differentiation or to respond to an extracellular stressor. Understanding the epigenome, the mechanisms that regulate it, and its impact on transcription are crucial to understanding contextual cell behaviour and the general notion of cell state.

Many associations between epigenetic characteristics, and functional chromatin organization have been established. The linear genome is folded and organized into two distinct spatial compartments (A/B compartments), with each compartment spanning several megabases of DNA[27]. Regions within a compartment exist within close spatial

proximity to each other, and are spatially separated from the other compartment. This organization places transcription-rich spans of DNA towards the nuclear center and condensed heterochromatin at the nuclear periphery, often associating with the nuclear lamina[27]. Within these compartments exists further organization into megabase-scale topologically associated domains (TADs) whose boundaries are enriched with housekeeping genes, and are relatively invariant between cell types[28]. There is even further compartmentalization into CTCF-anchored loop domains of several hundred kilobases[29], and even further looping mediated by CTCF, cohesin, and RNA PolII[30]. This allows for coordinated expression of many genes from concentrated regions of transcriptional machinery[31]. Disrupting the architecture of loops associated with high levels of transcription can also cause the upregulation of genes adjacent to the loops, suggesting that the looping serves to isolate and coordinate transcriptional regulation[32].

Along this organized chromatin, various histone modifications demarcate functional chromatin states (**Fig 1**). For example, H3K4me3 marks active promoters[33], H3K27ac is present at active enhancers[34], H3K27me3 at polycomb-repressed regions[35], and H3K9me3 at heterochromatin associated with the nuclear periphery[36]. The precise function of these modifications isn't clear, as most studies have focused on the correlation between these marks and chromatin states. It is clear, however, that these modifications are important, as deleting the factors that establish and maintain them often results in dramatic phenotypes. It is thought that either through structural changes to nucleosomes, or through acting as a molecular beacon for the recruitment of other regulatory factors, these modifications ultimately lead to changes in chromatin accessibility, affecting the ability of regulatory factors to interact with underlying DNA[37].

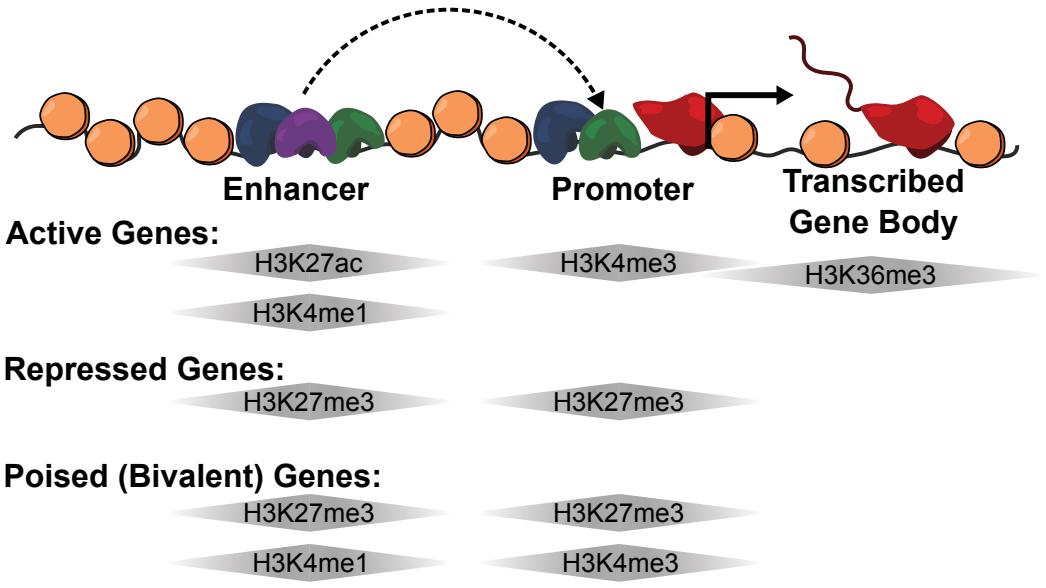


Figure 1. A simplistic epigenetic model.
This model demonstrates a simplified representation of the prevalence of various covalent histone modifications at specific genomic elements associated with active, repressed, and poised genes.

ATP-dependent chromatin remodelling

ATP-dependent chromatin remodelling enzymes are responsible for sliding, evicting, or exchanging histone proteins along DNA. This ultimately affects the accessibility of underlying DNA to regulatory factors[38], which is essential for the activation and inhibition of regulatory regions. This helps facilitate controlled transcriptional output by governing where transcription factors can converge along the genome, preventing promiscuous transcription of non-coding regions[39], while allowing recruitment of these factors to specific promoters and enhancers[23,40] across the genome.

There are four families of chromatin remodelling ATPases in mammals: SWI/SNF, ISWI, CHD, and INO80[38] (**Fig 2**). All families contain a highly conserved

ATPase domain, but vary in the flanking domains: SWI/SNF ATPases are characterized by a C-terminal bromodomain that allows interaction with acetylated histone tails, and an N-terminal helicase-SANT-associated (HSA) domain that interacts with nuclear actin-related proteins[41]; ISWI ATPases have C-terminal SANT and SLIDE domains, which allow them to interact with components of histone proteins and extranucleosomal linker DNA, respectively[42,43]; CHD remodellers have two N-terminal chromodomains, which interact with methylated tails of histone H3[44], and a C-terminal SLIDE domain; and the INO80 remodellers have a single HSA domain, as well a unique 300 amino acid insertion between the two components of the ATPase domain[45]. While these ATPases alone can affect nucleosome spacing *in vitro*, they exist in a variety of nucleosome remodelling complexes *in vivo*, ranging from small ISWI complexes with two to four subunits, to large SWI/SNF and INO80 complexes with over a dozen protein subunits[38]. The precise function of many of these subunits remains unknown, however it thought that many provide complexes with specificity across the genome, such as the BAF subunits of SWI/SNF remodelling complexes, which can directly bind DNA, and can interact with a variety of transcription factors to increase its specificity[46].

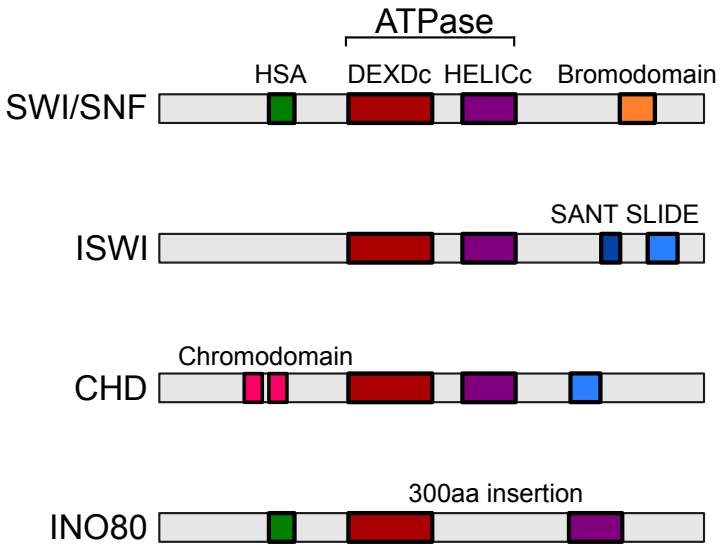


Figure 2. ATP-dependent chromatin remodelling subfamilies.
Schematic of the generalized protein structures of remodellers of the four major subfamilies: SWI/SNF, ISWI, CHD, and INO80. Remodellers have a highly conserved ATPase domain comprising DEAD-like helicase (DEXDc) and helicase superfamily c-terminal (HELICc) protein domains. Subfamilies vary in the composition of the regions flanking the ATPase domain. SWI/SNF remodellers are characterized by an N-terminal helicase-SANT (HSA) domain, and a C-terminal bromodomain. ISWI remodellers lack these domains, but contain C-terminal SANT and SLIDE domains. CHD remodellers contain two N-terminal chromodomains, and a C-terminal SLIDE domain. Like SWI/SNF remodellers, INO80 remodellers have an N-terminal HSA domain, but lack C-terminal domains. Rather, they are characterized by a 300 amino acid insertion between the two components of the ATPase domain.

Remodellers have proven critical for much of mammalian development. Deletion of many individual ATPases, representing each of the four families, precludes the production of a viable embryo, often due to early developmental deficits[46]. Disrupting the expression of many remodeller components in adult cells has highlighted their importance in adult tissues as well. For example, CHD4 is required to maintain the self-renewal of hematopoietic stem cells[47], and the SWI/SNF ATPase BRG1 is essential for the differentiation into neurons[48], lymphocytes[49], adipose[50], and cardiac tissue[51].

Components of the SWI/SNF, CHD, and INO80 families have each been implicated in maintaining the pluripotent state in mESCs[18–20], however the ISWI family has yet to be investigated in this context.

The mammalian ISWI homologue SNF2H

The ISWI family of chromatin remodellers in mammals comprises two ATPases, SNF2H (sucrose nonfermenting 2 homologue; *Smarca5*) and SNF2L (sucrose nonfermenting 2-like protein 1; *Smarca1*), which are homologues of the *Drosophila melanogaster* ATPase ISWI. SNF2L expression is limited to reproductive and neural tissues[52], and has been shown to be important for follicular development within the ovary[53], and in regulation of progenitor cell populations in the developing brain[54].

SNF2H is a catalytic subunit of six characterized remodelling complexes (ACF1, CHRAC, RSF, NoRC, WICH, and WCRF; **Fig 3**) that are broadly involved in transcriptional regulation, DNA replication and repair, and maintaining chromatin structure[55]. These tasks are performed primarily through the repositioning of nucleosomes along chromatin[55]. Unlike other remodellers, the ATPase assays suggest that the ATPase function of ISWI is not stimulated by free DNA or histones, but rather recognizes some structural feature of assembled nucleosomes[56]. Further mutagenesis experiments have demonstrated that the ATPase function depends on a critical hydrophilic epitope of the histone-H4 tail that interacts with nucleosomal DNA[57].

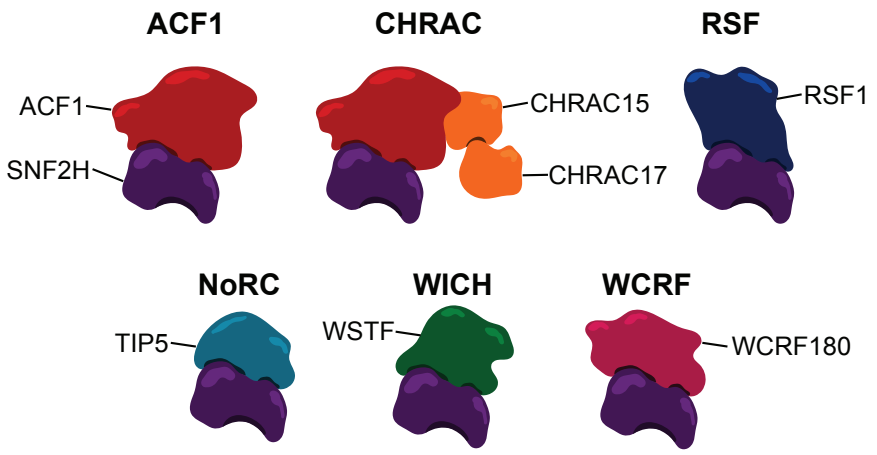


Figure 3. SNF2H chromatin remodelling complexes.
Schematics of the SNF2H-containing chromatin remodelling complexes ACF1, CHRAC, RSF, NoRC, WICH, and WCRF. These complexes have been implicated in various roles, such as transcriptional regulation, DNA replication and repair, and maintaining higher-order chromatin structure.

While the molecular mechanisms of SNF2H-mediated chromatin remodelling are fairly well understood, it is unclear how it is targeted to specific loci across the genome. A "continuous sampling" model has been proposed, in which mammalian ISWI proteins form transient interactions with chromatin at a high frequency (continuous sampling) until a molecular cue increases its binding affinity at a specific locus[55,58]. This model is supported by measurements dynamics of GFP-tagged SNF2H from fluorescence recovery after photobleaching, continuous photobleaching, and fluorescence (cross-)correlation spectroscopy-based experiments[55,58]. It is not known what can trigger this increased affinity, however various protein or RNA structures may play a role: the SNF2L-containing NURF complex has been shown to depend on H3K4me3 to maintain HOX gene expression[59], and the ATPase activity of the rRNA-regulating NoRC complex is reduced by non-coding RNA transcribed from the promoters of rRNA genes[60]. It seems

likely that diverse molecular cues could modulate SNF2H activity depending on the cellular context.

SNF2H also has important roles in mammalian development. It is most studied in the developing cerebellum, where conditional knockout in neural progenitors results in decreased expansion of the progenitor cell population, resulting in neurological deficits and death several weeks after birth[61]. While these cells have reduced proliferation, *Snf2h* deletion in post-mitotic Purkinje cells also produces cognitive alterations, suggesting that observed phenotypes are not solely due to the regulation of proliferation[61]. Its expression is also not limited to neural tissues, but rather, it's expressed across a variety of tissues[52]. Dynamic *Snf2h* expression during ovarian follicle development also suggests a possible role in regulating proliferating cells in growing follicles[62]. Stopka and Skoultchi[63] demonstrated that mouse embryos lacking SNF2H are not viable. Using a targeted deletion model that excises a portion of the ATPase domain, they found that homozygous-null embryos are not detectable after 3.5-d.p.c., and when cultured, the inner cell mass of null embryos fails to form a typical colony outgrowth and the cells undergo growth arrest[63]. Heterozygous mice are viable, however. This phenotype is consistent with other null models of remodellers in mice (reviewed by Ho and Crabtree[46]). These data suggest that SNF2H, like other remodellers, may be important in the maintenance of pluripotency and proper development of the early embryo.

Hypothesis

We hypothesize that through continuous sampling and targeted stabilization at regulatory regions, SNF2H serves as a broad regulator of the global gene expression program in mESCs, and is essential for the maintenance of the pluripotent state.

Objectives

1. Disrupt *Snf2h* expression in mESCs and assess the consequences on the pluripotent state through monitoring of cell viability, morphology, and gene expression
2. Use ChIP-Seq to map SNF2H localization across the mESC genome and determine if SNF2H preferentially localizes to specific genomic features

Materials and Methods

Cell culture and mESC maintenance

mESCs were cultured in standard serum-based culture conditions. Cells were maintained in DMEM with 4.5g/L glucose, L-glutamine, and sodium pyruvate (Corning), supplemented with 15% fetal calf serum, 1X PenStrep (ThermoFisher Scientific), 1X MEM Non-Essential Amino Acids (ThermoFisher Scientific), 1000U/mL ESGRO Leukemia Inhibitory Factor (LIF; EMD Millipore), and 8nL/mL 2-Mercaptoethanol (Sigma). Initial passages of the derived *Snf2h*$^{fl/fl}$-mESCs were maintained on an irradiated feeder layer of DR4 MEFs (ATCC, SCRC1045). After validating characteristics of pluripotency, the line was weaned off feeder cells and cultured on 0.1% gelatinized plates. Culture medium was changed every second day and passaged when confluent (usually every 2-3 days).

Derivation of a *Snf2h*$^{fl/fl}$-mESC line

Derivation of the *Snf2h*$^{fl/fl}$-mESC line was performed as described in Czechanski *et al.*[64]. Briefly, timed matings were performed with C57Bl/6 *Snf2h*$^{fl/fl}$ mice[61] (mice kindly provided by Arthur Skoultchi, Albert Einstein College of Medicine) and 3.5-d.p.c. pregnant females were collected. The uterus was then removed and fertilized embryos were flushed into petri dishes using M2 embryo media (Sigma-Aldrich). Embryos were transferred and washed through several M2 media droplets before being pooled in a petri dish with M2 media. At this stage, most embryos were at the expanded blastocyst stage. Under-developed embryos were transferred into M16 media (Sigma-Aldrich) for further development for up to 24 hours (any embryos that remained under-developed embryos at this point were discarded). No chemical removal of the zona pellucida was performed. Individual hatched blastocysts were then transferred to a well of a 24-well plate containing irradiated DR4 MEF feeders in normal FBS-containing mESC media, described above. Embryos were allowed to attach to the MEF layer for 3 days without disruption. Following attachment, culture medium changed daily. After 7 days in culture, primary outgrowths were disaggregated using trypsin and vigorous pipetting with a P2 pipette tip. While this tended not to completely break up the primary outgrowth, cells sloughed off in the process did seed the growth of new colonies. After several days (a judgement based on the size of the residual primary outgrowth and the development of secondary colonies), enzymatic disaggregation was repeated. As the number of mESC colonies increased, the cells were eventually transferred to 6-well plates, and finally 10cm plates. The number of disaggregation steps to reach this point was variable, but was approximately 4-6 times. Also, several embryos did not reach this final point, ultimately

resulting in two derived lines. One of the two lines was characterized to confirm characteristics of pluripotency, and was used for all experiments.

Western blot

Cells were cultured in 10cm tissue culture plates until confluent. Culture medium was then aspirated, and the cells were washed with PBS. Cells were then lysed with 500µL M-PER Mammalian Protein Extraction Reagent (ThermoFisher Scientific). Lysates were collected into a microcentrifuge tubes and sonicated for 30 seconds (Sonic Dismembrator Model 300) to increase yield, followed by centrifugation at 15,000xg at 4^{o}C to clarify the lysate. Protein was quantified using the Bradford Protein Assay (Bio-Rad) colorimetric assay, and an 30µg of protein was prepared for each lane by boiling in SDS loading buffer for 10 minutes. The NuPAGE Electrophoresis System (Invitrogen) was used with precast 4-12% graduated tris-glycine gels (Invitrogen) and the tris-glycine gel protocol described in the NuPAGE Technical Guide. Gels were ran for 2 hours at 120V with 1X NuPAGE MOPS running buffer (20X: 1M MOPS, 1M Tris, 69.3mM SDS, 54.9mM EDTA). Gels were transferred to a Hybond-C Extra nitrocellulose membrane (Amersham Biosciences) using the Mini Trans-Blot Cell (Bio-Rad) with Towbin buffer (25mM Tris, 192mM glycine, 20% methanol in ddH_2O) at 100V for 1 hour.

Protein membranes were blocked for 1 hour in 5% milk in Tris-buffered saline with 0.1% Tween-20 (TBST), and probed with primary antibodies overnight at 4^{o}C. Primary antibodies used include anti-SNF2H (1/2000; Abcam, ab72499) and anti-β-Actin (1/80,000; Sigma-Aldrich, A2228). Membranes were then washed three times for 5 minutes with TBST, and were then probed with HRP-conjugated secondary antibody for

1 hour at room temperature. Secondary antibodies used were rabbit anti-mouse IgG HRP (1/10,000; Abcam, Ab6728) and donkey anti-rabbit IgG HRP (1/10,000; Sigma, 711-035-152). The membrane was then washed three times for 5 minutes with TBST, and developed with ECL Advance Western Blotting Detection Kit (Amersham Biosciences).

PCR-based genotyping

Genomic DNA (gDNA) was isolated from cells using the GeneJET Genomic DNA Purification Kit (ThermoFisher Scientific) and 2μL of purified gDNA was added to 5μL of the 2X REDExtract-N-Amp PCR ReadyMix (Sigma-Aldrich), 0.2μL 10μM forward/reverse primer mix, and 2.8μL ddH_2O. The reaction was performed with the following parameters: initial denature, 94^{o}C, 3 minutes; denature, 94^{o}C, 30 seconds; anneal, 55^{o}C, 30 seconds; extension, 72^{o}C, 2 minutes; repeat anneal and denature for a total of 34 cycles; one final extension, 72^{o}C, 10 minutes. PCR products were then run down a 1% agarose gel containing RedSafe Nucleic Acid Staining Solution (Intron Biotechnology) and visualized with an EpiChem II Darkroom Transilluminator (UVP Laboratory).

Immunofluorescence of pluripotency factors

Glass coverslips were flame-sterilized, placed in the wells of 6-well plates, and coated with 0.1% gelatin for 30 minutes. The gelatin was then aspirated and *Snf2h*$^{fl/fl}$-mESCs were plated on the gelatinized coverslips. Once the cells had reached confluence, they were washed twice with cold PBS, and then fixed in 4% paraformaldehyde for 30 minutes at room temperature. After fixation, the paraformaldehyde was rinsed off with PBS, and washed three additional times with PBS, bathing the cells for 5 minutes each time. The cells were then permeabilized with 0.2% Triton X-100 in PBS for 10 minutes

at room temperature, which was then rinsed off, and the cells were washed for an additional three washes in PBS for 5 minutes each wash. The cells were then blocked in 5% normal goat serum in PBS for 1 hour at room temperature. After blocking, the primary antibodies (OCT4, Ab19857, 1/800; SOX2, Ab97959, 1/1000; NANOG, Ab80892, 1/100) were added and the cells were incubated at 4^{o}C overnight. The following morning, the cells were washed three times in PBS for 5 minutes. The secondary antibody (goat anti-rabbit IgG Alexa Fluor 488, ThermoFisher Scientific, 1/1000) was diluted in 5% normal goat serum in PBS and was added to the cells for 1 hour in the dark, at room temperature. Afterwards, the cells were washed three times for 10 minutes in PBS. Finally, coverslips were removed from the wells and mounted onto slides using ProLong Gold Antifade Mountant with DAPI (ThermoFisher Scientific). Images were taken on a Zeiss Axioskop 2 MOT upright microscope.

Embryoid body differentiation of *Snf2h*$^{fl/fl}$-mESCs

Snf2h$^{fl/fl}$-mESCs (5x10^{5}) were diluted into 10mL of mESC media lacking LIF supplementation, and were transferred into a 10cm bacterial petri dish to prevent attachment. Cell aggregates were left alone for 4 days, after which the culture medium was changed every second day. This was performed by transferring the aggregates into a 15mL conical tube, allowing the aggregates to settle at the bottom of the tube, aspirating the cleared media being careful not to aspirate any aggregates, and replacing with 10mL of LIF-free mESC media. This was repeated until 14 days after the initial plating. To prepare the embryoid bodies (EBs) for paraffin embedding and cross-sectioning, the EBs were collected in a 15mL conical tube, the medium was aspirated, and the EBs were washed with PBS. The PBS was then aspirated and the EBs were resuspended in 10%

formalin to fix for 30 minutes. Afterwards, the formalin was aspirated, the EBs were washed once in PBS, resuspended in 1mL of 1% agarose, and placed into a square-shaped mold to solidify. After the mix had solidified, the agarose blocks were paraffin-embedded, cut into 5μm sections, and stained with hematoxylin and eosin.

Co-transfection approach for Cre-mediated deletion of *Snf2h*

Snf2h$^{fl/fl}$-mESCs were trypsinized and 3x10^5 cells were plated per well in a gelatinized 6-well plate. The following day, the culture medium was replaced and the cells were transfected with either pKJ1-GFP alone or a 9:1 ratio of pKJ1-Cre and pKJ1-GFP, respectively (total mass of DNA transfected was the same for both conditions), respectively, using Lipofectamine 3000 (ThermoFisher Scientific) according to the manufacturer's protocol (7.5μL lipofectamine reagent per well). The pKJ1 vectors were a kind gift from Michael McBurney (pKJ1 backbone: Addgene plasmid #11333) and were specifically chosen due to high expression levels in mESCs from the *Phosphoglycerate kinase-1* (*Pgk1*) promoter. We herein refer to these expression vectors as pgk-GFP and pgk-Cre. One day following transfection, GFP expression was visually confirmed and all cells were trypsinized and transferred into a 10cm plate for further expansion. The following day, GFP-expressing cells were purified using fluorescence-activated cell sorting (FACS). Given that in our experimental group, 9X more pgk-Cre was transfected than pgk-GFP, all GFP-expressing cells should presumably express Cre as well. Sorted cells were then cultured and maintained in 6-well plates for downstream analysis.

siRNA transfection of *Snf2h*$^{fl/fl}$-mESCs

Snf2h$^{fl/fl}$-mESCs were trypsinized and 1x10^5 cells were plated in a gelatinized 6-well plate. The following day, the culture medium was replaced and the cells were

transfected with 75pmol of either siSmarca5 (SMARTpool of four siRNA targeting mouse *Smarca5*; Dharmacon, D-041484-03) or siScrambled (siGenome Non-Targeting siRNA #1; Dharmacon, D-001210-01-05) using Lipofectamine 3000 according to the manufacturer's protocol for siRNA transfection. Successful knockdown was confirmed by qPCR.

Quantitative real-time RT-PCR (qPCR)

Cells were trypsinized, centrifuged at 400xg, and washed with PBS before a final centrifugation to isolate a cell pellet. Pellets were frozen at -80^{o}C or immediately processed for RNA extraction on ice. RNA was extracted using the GenElute Mammalian Total RNA Miniprep Kit (Sigma-Aldrich) and RNA quality and quantity were assessed with the NanoDrop ND-1000 spectrophotometer (ThermoFisher Scientific). cDNA was then synthesized from 500ng of RNA using the iScript cDNA Synthesis Kit (Bio-Rad), along with "no reverse transcriptase" and "no RNA" negative controls. Samples were run on an ABI7500 Fast Real-Time PCR System (Applied Biosystems) with the SsoFast EvaGreen Supermix with Low ROX (Bio-Rad). The following parameters were used: initial denature, 95^{o}C, 30 seconds; denature, 95^{o}C, 5 seconds; anneal, 60^{o}C, 30 seconds; repeat denature and anneal steps for a total of 40 cycles. Primers were all designed with Primer3Plus[65] (Primer sequences are listed in Appendix I).

Alkaline phosphatase activity stain

Alkaline phosphatase activity was determined using the Alkaline Phosphatase Detection Kit (Millipore) according to the manufacturer's protocol, with a minor modification. *Snf2h*$^{fl/fl}$-mESCs were co-transfected with pgk-GFP and pgk-Cre, or pgk-GFP alone as a control, as described above. After samples were sorted for GFP-

expressing cells, cells were cultured for 4 days. The manufacturer's protocol recommends 5 days, however, given the rebound in *Snf2h*-expressing cells over time, we wanted to limit the effect of this on alkaline phosphatase activity. After 4 days, cells were fixed in 4% paraformaldehyde in PBS for 90 seconds. Cells were washed twice in 1X TBST and were then incubated in the Naphthol/Fast Red Violet Solution for 15 minutes in the dark at room temperature. After the incubation, cells were rinsed once with 1X TBST and then covered in 1X TBST to prevent drying. Stained colonies were then visualized using microscopy.

Annexin V apoptosis detection

Annexin V apoptosis and dead cell detection assay was performed with the Alexa Fluor 488 Annexin V/Dead Cell Apoptosis Kit (Invitrogen, cat# V13241) according to the manufacturer's protocol. Briefly, *Snf2h*$^{fl/fl}$-mESCs were transfected with either siRNA targeting *Snf2h*, or a non-targeting control for mouse cells, as described above, and cultured in normal conditions for four days. Cells were removed from the plate using trypsin, centrifuged at 300xg for 5 minutes, washed with cold PBS, re-centrifuged, and resuspended in 1X annexin-binding buffer at a density of $1x10^6$ cells/mL (100µL per sample). 5µL Alexa Fluor 488 annexin V and 1µL 100µg/mL propidium iodide (PI) solution were added to each 100µL sample. Cells were incubated for 15 minutes in the dark at room temperature. After the incubation, 400µL of 1X annexin-binding buffer was added, mixed gently, and immediately analyzed by flow cytometry. Cells were analyzed by measuring FITC and PI fluorescence. Gates were established to separate three distinct populations of cells: alive cells were FITC^{low} and PI^{low}, apoptotic cells were FITC^{high} and

PI^{low}, and dead cells were $FITC^{high}$ and PI^{high}. Three replicates from three independent transfections were performed and identical gates were used across all samples.

Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) was performed with the SimpleChIP Enzymatic Chromatin IP Kit (Cell Signaling Technology) according to the manufacturer's protocol. *Snf2h*$^{fl/fl}$-mESCs ($20x10^6$) were crosslinked in 1% methanol-free formaldehyde for 10 minutes. Cross linking was quenched in 1X glycine solution for 5 minutes and cells were washed twice with ice-cold 1X PBS. Cells were scraped from the plate into 1X PBS with 1X protease inhibitor cocktail (PIC) and pelleted by centrifugation at 400xg at 4°C. Nuclei were extracted by resuspending the pellet into 4mL of 1X Buffer A + DTT + PIC for 10 minutes on ice with gentle mixing every 3 minutes. Nuclei were pelleted by centrifugation at 500xg at 4°C, washed in 4mL 1X Buffer B + DTT, re-centrifuged, and resuspended in 400µL 1X Buffer B + DTT. Chromatin was digested using 4µL of micrococcal nuclease (MNase; Cell Signaling Technology, cat #10011) for 25 minutes at 37°C, mixing gently every 5 minutes, producing primarily mono-, and di-, and tri-nucleosomes. The reaction was stopped by adding 40µL of 0.5M EDTA and placing the sample on ice. Nuclei were then pelleted by centrifugation at 15,000xg for 1 minute at 4°C, resuspended in 400µL 1X ChIP Buffer + PIC, and incubated on ice for 10 minutes. In a microcentrifuge tube, nuclei were sonicated for 90 seconds (Misonix Sonicator 3000; 20s ON, 30s OFF) to break the nuclear membrane. Lysates were clarified by centrifugation at 10,000xg for 10 minutes at 4°C.

Clarified lysate (100µL) was transferred to four tubes containing 4µL of 1X ChIP Buffer + PIC and 10µL was removed from one of the tubes and frozen at -20°C for a 2%

Input Control. Antibody was added to each tube (10µL anti-Histone H3, Cell Signaling Technology, cat #4620; 4µg Normal Rabbit IgG, Cell Signaling Technology, cat #2729; 4µg anti-SNF2H, Abcam, ab72499). Two SNF2H IP replicates were performed. Samples were mixed with antibodies overnight at 4^{o}C, and 30µL Protein G Magnetic Beads (Cell Signaling Technology, cat #9006) were added for 2 hours at 4^{o}C. Beads were washed three times in low-salt buffer, and once in high-salt buffer. The IP samples were eluted from the beads in 150µL (per IP) 1X ChIP Elution Buffer at 65^{o}C, vortexing every 3 minutes. Crosslinks were then reversed by adding 6µL 5M NaCl and 2µL Proteinase K (Cell Signaling Technology, cat #10012) to each IP sample and incubating at 65^{o}C for 2 hours. DNA samples were column purified (provided by manufacturer), eluted in 50µL DNA Elution Buffer, and quantified using the Quant-iT PicoGreen dsDNA Assay (Invitrogen, cat# P11496). DNA quantifications are listed in Appendix II, along with a western blot of the SNF2H antibody used. These data demonstrate that the ChIP protocol efficiently removes non-specific genomic fragments (as indicated by the DNA-free IgG sample), and that the antibody used does not promiscuously interact with non-specific proteins.

qPCR was performed by adding 1µL sample with 3µL nuclease-free water, 1µL 5µM primers (See Appendix I for primer sequences), and 5µL SYBR-Green Reaction Mix. The reaction was run with the following parameters: Initial denature, 95^{o}C, 3 minutes; denature, 95^{o}C, 15 seconds; anneal and extension, 60^{o}C, 60 seconds; repeat denature and anneal steps for a total of 40 cycles.

Library preparation and sequencing

Immunoprecipitated and MNase-digested input DNA were submitted to The Centre for Applied Genomics (The Hospital for Sick Children, Toronto, ON) for ChIP-Seq library preparation using the New England Biolabs NEBNext ChIP-Seq Library Prep Reagent Set for Illumina (New England BioLabs) following the manufacturer's recommended protocol. In brief, DNA was end repaired, followed by dA-tailing to create an overhang A that can ligate to the corresponding overhang T of the double stranded Illumina-compatible adapters, and was PCR enriched with an initial denaturation step at 98^{o}C for 30 seconds, followed by 15 cycles of 98^{o}C for 10 seconds, 65^{o}C for 30 seconds and 72^{o}C for 30 seconds, and a final extension at 72^{o}C for 5 minutes. During adapter ligation, each sample received a different barcoded adapter included in the NEBNext Multiplex Oligos kit (New England BioLabs) to allow for multiplexed sequencing. 1μL of the ChIP-seq libraries was loaded on a Bioanalyzer 2100 DNA High Sensitivity chip (Agilent Technologies) to check for fragment size and absence of primer dimers. Libraries were quantified by qPCR using the Kapa Library Quantification Illumina/ABI Prism Kit protocol (KAPA Biosystems). Libraries were pooled in equimolar quantities and sequenced on an Illumina HiSeq 2500 platform using a Rapid Run Mode flowcell following Illumina's recommended protocol to generate single end reads of 100 bases in length.

Computational analysis

All scripts used to process data and generate figures are publicly available and maintained at http://github.com/dpcook/mscthesis.

Public data accession

All public data sets used are listed in Appendix III, along with GEO accession numbers or URLs where applicable.

ChIP-Seq alignment and peak calling

All ChIP-Seq data sets were aligned to the mm9 mouse genome assembly using Bowtie2[66] with the --sensitive-local parameter. With the exception of the SNF2H ChIP-Seq data sets, peak calling was performed with MACS2[67] using the callpeak --SPMR -q 0.05 options, along with the -c parameter specifying an appropriate input control when available (See Appendix III for ChIP-Seq data sets and matched controls). These options set a q-value threshold of 0.05 for peak calling and normalize the output signal track by sequencing depth, reporting values as reads per million mapped reads (rpm). Peak calling was omitted for the SNF2H ChIP-Seq data sets because we found it performed poorly at generating high-confidence peak calls. This could be due to its broad distribution across the genome with only modest enrichment at specific loci. Given the high quantification of immunoprecipitated DNA relative to a non-targeting IgG control, we believe this broad signal is true binding rather than technical noise—which is often the interpretation in peak calling—making peak calling inapplicable. Working under the null hypothesis that SNF2H has no regional specificity, binding across the genome with equal probability, evaluating the enrichment of normalized SNF2H signal over MNase-digested input signal at specific target loci proved more useful. For downstream analysis, the alignment files of the two SNF2H ChIP replicates were combined to increase power.

Calculating SNF2H enrichment regulatory factor binding sites

Factor binding sites were defined as the summit locations of all peaks identified by MACS (described above). All binding sites for a given factor were pooled into a GFF file and the regions were expanded by 1kb in both directions, producing a 2kb window, centred on the binding site. We used bamToGFF (https://github.com/BranderLab/pipeline) to break each 2kb window into 20bp bins and calculated the number of reads from the SNF2H ChIP and MNase-digested input data sets that map to each bin. This produced a *m* x *n* matrix for each the SNF2H ChIP and MNase input samples, where *m* represents each binding site, and *n* represents each of the 20bp bins (ordered from 5' to 3' across the 2kb window). This was accomplished using bamToGFF with the -t -e 0 -m 100 options.

Because the primary interest was to understand the prevalence of SNF2H at binding sites relative to flanking regions, we normalized the read counts so that the reads for each 2kb window sums to 100 (representing the percent of total reads mapping to the 2kb window that fall within each bin). This accounts for differences in coverage between each binding site that may mask patterns when producing an average profile of the 2kb windows. We found that this approach tends to produce more-consistent nucleosome patterns than when averaging either total counts, or rpm-normalized counts. The signal in all plots represents the average normalized read counts for each bin.

Calculating SNF2H enrichment at transcription start sites

All known transcription start sites (TSSs) across the mouse genome were collected from the Ensembl NCBIM37.67 annotation using the gene start coordinate for positive-strand genes, and the gene end coordinate for negative-strand genes (because the

start and end coordinates are annotated by the most 5' and 3' positions, respectively, on the positive strand, regardless of which strand the gene is located on). All TSSs were pooled into a GFF file and expanded 2kb on both sides, producing a 4kb window, centred on the TSS. Each window was broken into 20bp bins and the number of reads that mapped to each bin was calculated using bamToGFF using the -t -e 0 -m 200 options. The same row normalization approach was performed, as described above.

Because negative-strand genes are transcribed in the opposite direction to positive-strand genes, we reversed the bin order for all negative strand genes so that all open reading frames project to the right of the TSS. The list of TSSs was also filtered to only include protein-coding or lincRNA gene annotations. This removed certain classes of genes, such as pseudogenes and small-nuclear RNAs, that may not have typical chromatin structures or transcriptional features, which may mask binding patterns when profiles are averaged. It also limits the data set to genes whose expression is quantified during standard RNA-Seq experiments, allowing us to use the ENCODE RNA-Seq data for mESCs[68] to filter for genes that are expressed or non-expressed in mESCs and compare binding profiles.

Calculating SNF2H enrichment at DNase I hypersensitivity sites

DNase I hypersensitivty sites across the mESC genome were collected from the ENCODE ES-E14 DNase-Seq processed peak files[69]. We combined the BED files from the two replicates and merged overlapping peaks by running the BEDtools[70] merge function with default parameters. All 233,253 hypersensitive regions were pooled into a GFF file, and we also created two separate GFF files representing 1kb upstream and 1kb downstream of the hypersensitivity site. These sites are variable in size (mean: 166bp), so

to create an average hypersensitivity profile, each site was broken into 20 bins, regardless of the site's size. The number of SNF2H and MNase input reads mapping to each bin was calculated, and differences in bin size was accounted for by dividing the counts by the size of the bin, producing a density value (reads/bp). Read density was also determined for the 1kb flanking regions after breaking each flank into 50 bins. This was all performed used bamToGFF with the -r -e 0 -d -m 20 (for the hypersensitivity sites; -m 50 for flanking regions). The three resulting matrices were combined horizontally so that the resultant matrix spans from 1kb upstream, across the hypersensitivity sites, to 1kb downstream. Differences in coverage between hypersensitivity sites was accounted for by normalizing each row so that the reads across bins summed to 1. The mean normalized value for each bin was calculated and plotted. To make values easier to visually assess when plotted, we transformed these values proportionally so that the left-most input value was set to 1.

Calculating SNF2H enrichment across chromatin states

A chromatin state map of the mESC genome was produced by Bogu *et al.*[71] (https://github.com/gireeshkbogu/chromatin_states_chromHMM_mm9) using ChromHMM[72] to train a 15-state model based on H3K36me3, H3K4me1, H3K27ac, H3K4me3, H3K27me3, RNA PolII, and CTCF ChIP-Seq profiles from ENCODE[68,72]. A GFF file was produced for each of the 15 states, pooling all regions of the same state. The SNF2H ChIP and MNase input data sets were rpm-normalized and a single rpm/bp density value was calculated for each region using bamToGFF with the -r -e 0 -m 1 - d options. The log SNF2H enrichment over input was calculated for each region, and the average enrichment score for each state was plotted.

Results

Derivation of a *Snf2h*$^{fl/fl}$ mESC line

To begin to understand if SNF2H-mediated chromatin remodelling is an essential component of the pluripotency regulatory system, we generated a mESC line from mice with floxed *Snf2h* alleles (*Snf2h*$^{fl/fl}$-mESC) allowing for a conditional excision of exon 5 following the addition of Cre recombinase **(Fig 4A;** mice were obtained from Arthur Skoultchi, Albert Einstein College of Medicine). This excises a portion of the DNA encoding the ATPase domain and generates a premature stop codon, preventing expression by nonsense-mediated decay. This knockout system was preferred to knockdown systems because mice harboring only one wild-type *Snf2h* allele are still viable[63] and it is unclear what degree of knockdown would be required to disrupt SNF2H function. Timed matings were performed and blastocysts were recovered at 3.5-d.p.c. and cultured on a feeder layer of MEFs (**Fig 4B**). Cell outgrowths were periodically dissociated, resulting in a cell line with a colony morphology typical of mESCs **(Fig 4C)**. PCR-based genotyping was used to confirm homozygous floxed alleles **(Fig 4D)**.

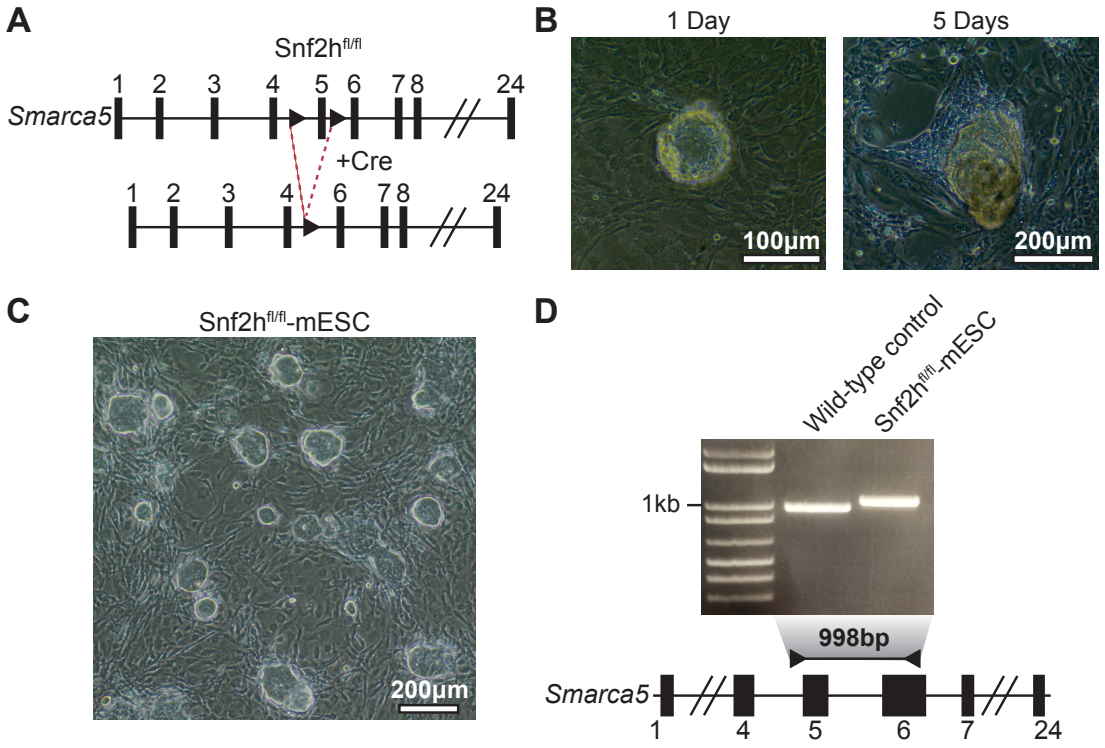


Figure 4. Derivation of a *Snf2h*$^{fl/fl}$-mESC line.
(A) Schematic of the *Snf2h* (*Smarca5*) locus in the *Snf2h*$^{fl/fl}$ mice. Arrowheads represent loxP sites. Cre-mediated recombination excises exon 5, inactivating the ATPase domain and generates a premature stop codon. **(B)** Phase contrast images of a *Snf2h*$^{fl/fl}$ embryo after 1 and 5 days (left and right panels, respectively) of culturing on a monolayer of irradiated MEFs. **(C)** Phase contrast image of the mESC line derived from the embryo in Figure 2B after several rounds of colony disaggregation and continued culture on MEFs. **(D)** PCR-based genotyping of the derived *Snf2h*$^{fl/fl}$-mESC line. The schematic at the bottom displays the location of the primers used in the reaction. A wild-type allele will produce a 998bp amplicon, however the floxed alleles contain a loxP site between exons 5 and 6, resulting in a 1032bp amplicon. This slight shift in size can be seen in the *Snf2h*$^{fl/fl}$-mESC sample. R1 mESCs were used as a wild-type control.

To determine if the derived *Snf2h*$^{fl/fl}$-mESC line displayed hallmarks of pluripotency, we first confirmed the expression of the core transcription factors OCT4, SOX2, and NANOG by immunofluorescence **(Fig 5)**. The cells were also able to differentiate into embryoid bodies comprising diverse histologies **(Fig 6)**, some of which

began to spontaneously contract after 12 days in culture (data not shown), consistent with differentiation into functional cardiac muscle tissue. Together, these data support that the derived *Snf2h*$^{fl/fl}$-mESCs are a valid cell line for studying the role of SNF2H in the regulation of pluripotency.

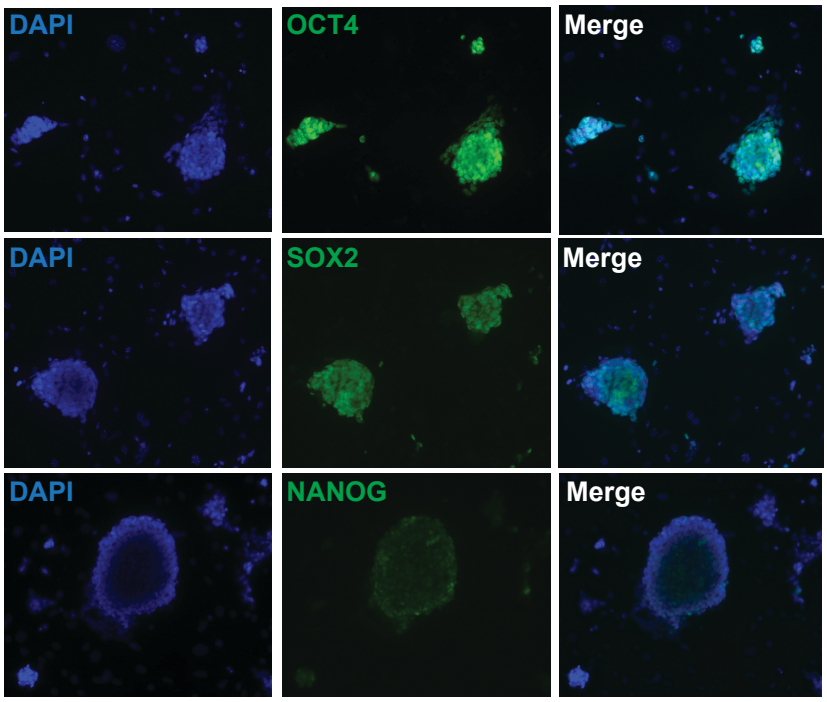


Figure 5. *Snf2h*$^{fl/fl}$-mESCs express markers of pluripotency.
Immunofluorescence of OCT4, SOX2, and NANOG in the *Snf2h*$^{fl/fl}$-mESCs. Expression of the pluripotency factors is restricted to the mESC colonies, and is absent from the surrounding MEFs.

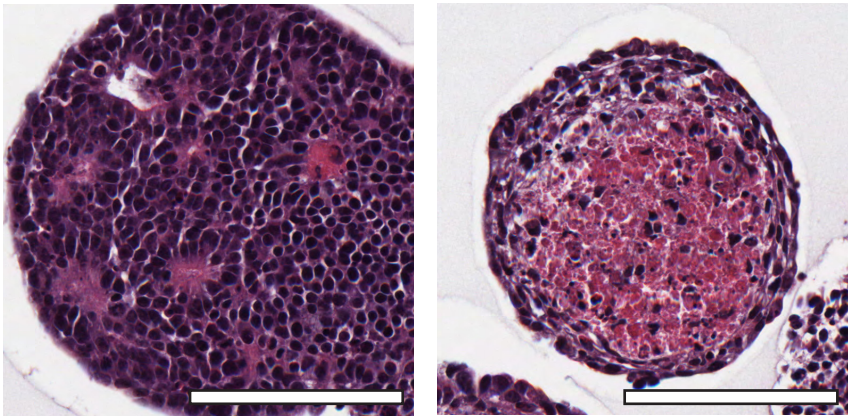


Figure 6. Embryoid body differentiation of *Snf2h*$^{fl/fl}$-mESCs.
Hematoxylin and eosin staining of cross-sections of agarose-embedded embryoid bodies. *Snf2h*$^{fl/fl}$-mESCs are able to differentiate to form distinct histologies, including rosette (left) and proteinaceous (right) structures, consistent with a pluripotent phenotype. Scale bar = 100μm

---

Cre-mediated deletion of *Snf2h* disrupts the pluripotent state

To begin to explore the importance of SNF2H in the regulation of the pluripotent state, we developed an approach to produce high-efficiency deletions of *Snf2h* within a narrow temporal window by co-transfecting *Snf2h*$^{fl/fl}$-mESCs with expression plasmids containing GFP and Cre recombinase driven by the *Pgk1* promoter (pgk-GFP and pgk-Cre, respectively), and then purifying GFP-expressing cells 48 hours later using fluorescence-activated cell sorting (FACS) (**Fig 7A;** See methods for further detail). This approach yields a highly enriched population of GFP expressing cells (**Fig 7B**), and because pgk-Cre and pgk-GFP plasmids were co-transfected at a 9:1 ratio, respectively, it can be assumed that all cells expressing GFP had also received Cre vector.

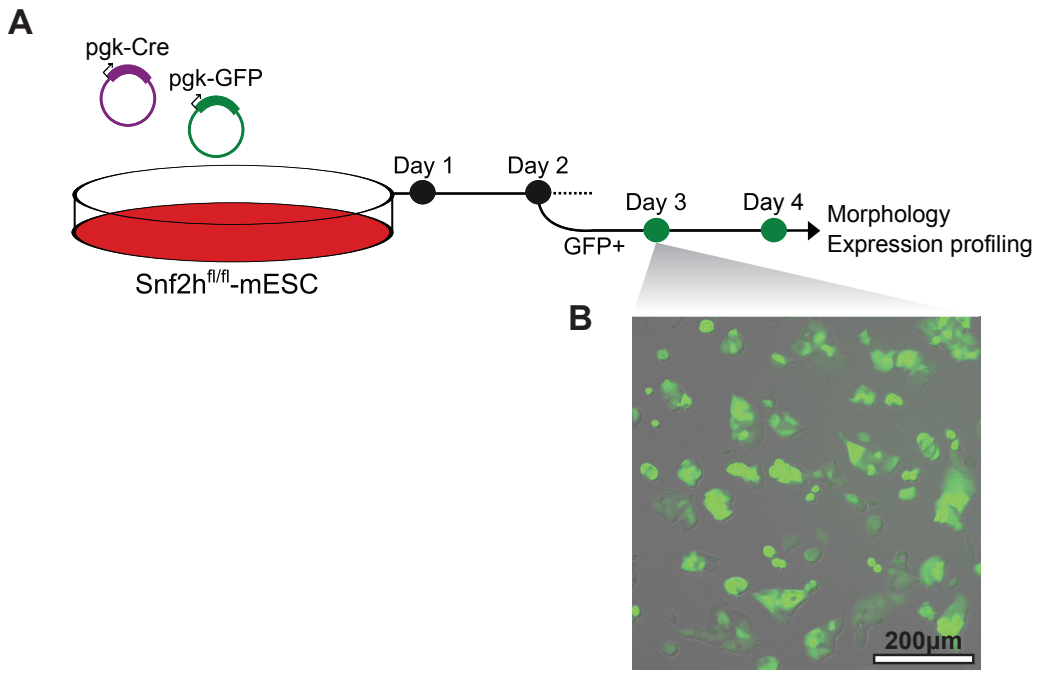


Figure 7. A co-transfection approach for deleting *Snf2h* in *Snf2h*$^{fl/fl}$-mESCs. **(A)** Cells are co-transfected with pgk-GFP and pgk-Cre expression plasmids, sorted for GFP expression 2 days following transfection to enrich for Cre-expressing cells. Cell morphology and gene expression was monitored following GFP enrichment. **(B)** GFP expression of control cells transfected with pgk-GFP alone one day following fluorescence-activated cell sorting (FACS). The cell population is highly pure for GFP-expressing cells.

Four days following Cre transfection, the cells begin to undergo a marked morphological change, losing their typical colony morphology, flattening out, and spreading across the culture dish (**Fig 8**). This morphological change is often observed during mESC differentiation, as well as following the disruption of other chromatin remodelling enzymes essential to mESCs[18–20]. RT-qPCR confirmed an 89% reduction in *Snf2h* expression across the population of cells, confirming that our co-transfection approach yields efficient, albeit not complete deletion throughout the population (**Fig 9**). Unfortunately, this analysis also showed that *Snf2h* expression levels recover over time,

and by 8 days following transfection, are 54% wild-type levels on average. The two most likely explanations for this rebound in the proportion of wild-type cells among the population would be *Snf2h* deletion producing a phenotype associated with slowed proliferation, or cell death.

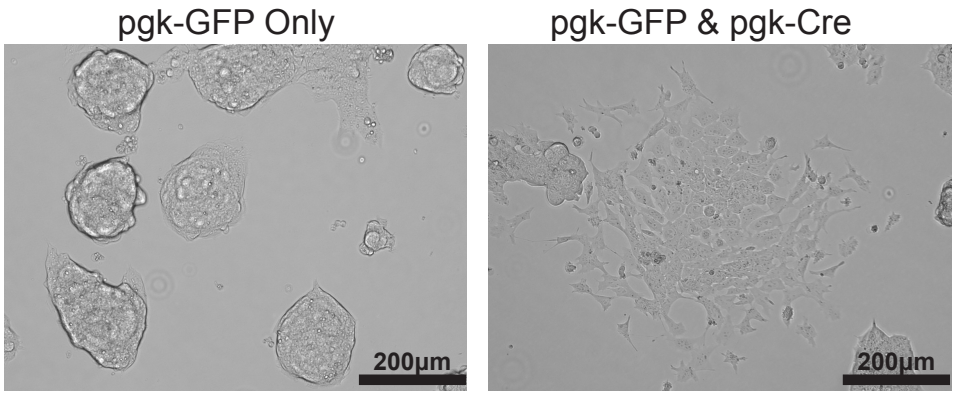


Figure 8. Cre-mediated deletion of *Snf2h* disrupts mESC morphology.
Images of *Snf2h*$^{fl/fl}$-mESCs 6 days after transfection with pgk-GFP and pgk-Cre expression vectors, or pgk-GFP alone as a control. Cells were processed as described in Figure 7. Cells transfected with Cre have an increased proportion of cells that fail to form typical mESC colonies. Colonies are still present, however, they are less prevalent.

We assessed the expression of the core transcription factors *Oct4*, *Sox2,* and *Nanog* 4, 6, and 8 days following Cre transfection. The expression of all three factors was reduced at 6 days, lagging behind the suppression of *Snf2h*, suggesting that SNF2H may be required to maintain the expression of key pluripotency genes (**Fig 9**). The cells with disrupted morphology also have reduced alkaline phosphatase activity 6 days following transfection, consistent with the loss of pluripotency (**Fig 10**).

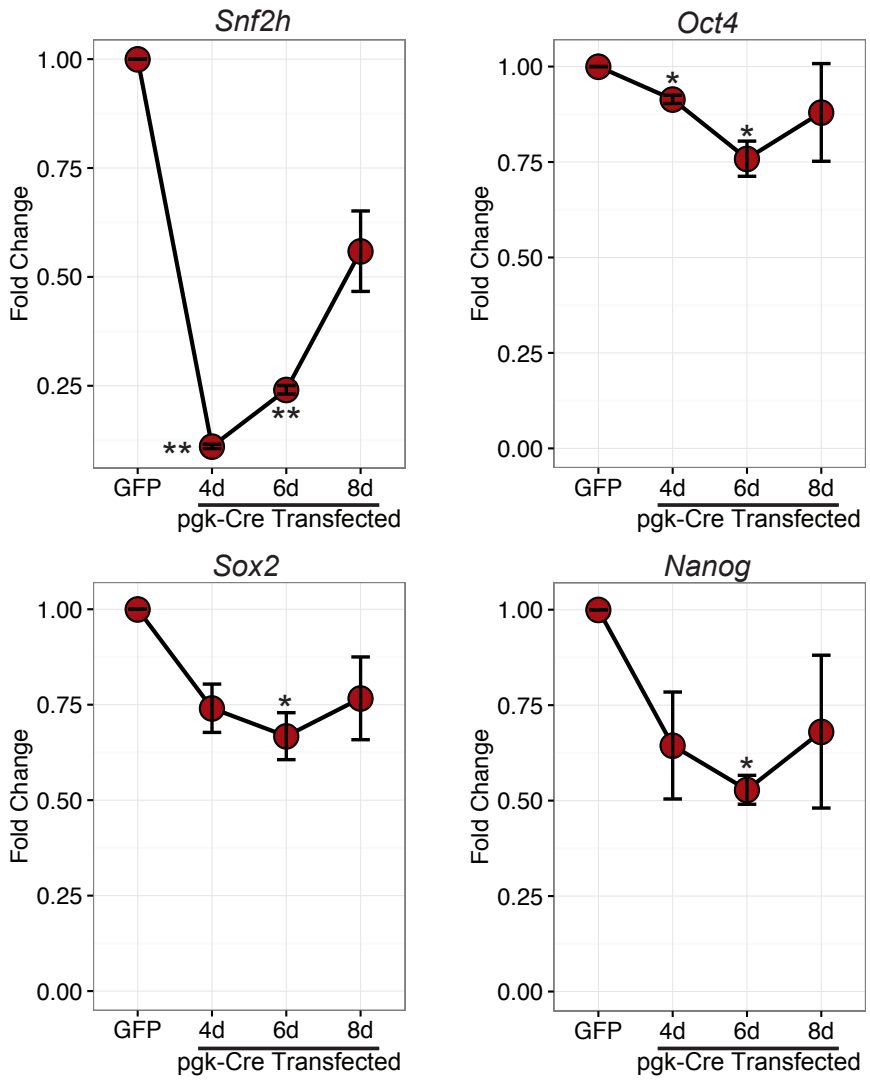


Figure 9. Cre-mediated deletion of *Snf2h* results in impaired expression of key pluripotency transcription factors.
qPCR expression of *Snf2h*, *Oct4*, *Sox2*, and *Nanog* at 4, 6, and 8 days(d) following Cre transfection. Values are represented as fold changes relative to a matched control sample. Error bars represent standard error of the mean. N=3. * P < 0.1, ** P < 0.05 ratio t-test (relative to null fold change value of 1). All P values were corrected for multiple hypothesis testing using a Benjamini-Hochberg false discovery rate correction, accounting for all tests performed.

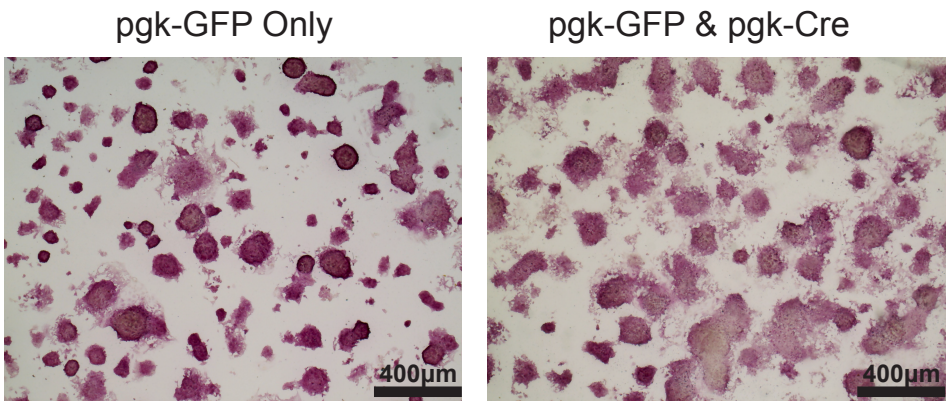


Figure 10. Cre-mediated deletion of *Snf2h* disrupts the pluripotent phenotype. Alkaline phosphatase activity staining of *Snf2h*$^{fl/fl}$-mESCs 6 days after transfection with either pgk-GFP as a control, or pgk-GFP and pgk-Cre. Cells transfected with Cre have a reduced staining intensity, particularly in cells with a flattened morphology, such as those seen in Figure 8.

Together, these data support the conclusion that the loss of SNF2H in mESCs disrupts the pluripotent state. The limited reduction of *Oct4, Sox2,* and *Nanog* expression is likely explained by the heterogeneity of *Snf2h* status within the population of cells at each time point. Assuming that cells with intact *Snf2h* alleles display a normal, pluripotent phenotype, their gene expression would dampen the observed effect size when collecting whole-population RNA samples. Additionally, full suppression of these factors may require more time than what can be allotted due to the relatively rapid resurgence of *Snf2h*-expressing cells.

Early morphological changes following SNF2H depletion are not associated with the activation of apoptosis

The role of SNF2H in DNA replication and repair, additional to its role in transcriptional regulation[55], suggests that long-term reduction of SNF2H in highly

proliferative cell types, such as mESCs, may not be supported. To determine if the early morphological changes are associated with transcriptional changes, rather than an apoptotic phenotype from the accumulation of DNA damage, or impaired proliferation due to impaired DNA replication, we assessed cell number and used Annexin V staining to assess early activation of apoptosis and cell death at the time of the morphological changes. We chose to use siRNA to disrupt *Snf2h* expression rather than our co-transfection approach. The rationale for this was to minimize confounding factors, such as differences in plasmid composition during transfection and the selection process during FACS (eg. higher GFP levels in control cells may produce differences in the cell populations recovered during FACS or during the Annexin V stain).

This siRNA approach yielded a 54% knockdown (**Fig 11A**) and morphological changes four days following transfection, consistent with our Cre-mediated deletion approach (**Fig 11B**). At this time point, there was no difference in cell number between cells transfected with siRNA targeting *Snf2h* and a non-targeting control (**Fig 12**). The Annexin V apoptosis assay also confirmed that *Snf2h* knockdown had no impact on the proportions of cells that were alive, undergoing apoptosis, or dead (**Fig 12**). This does not exclude the possibility that proliferation defects and apoptosis occur at later time points following SNF2H depletion. In fact, cell death or impaired proliferation are required to explain the resurgence of *Snf2h*-expressing cells over time. However, these data do suggest that the early phenotypes observed are likely the result of transcriptional disruption upon the loss of SNF2H.

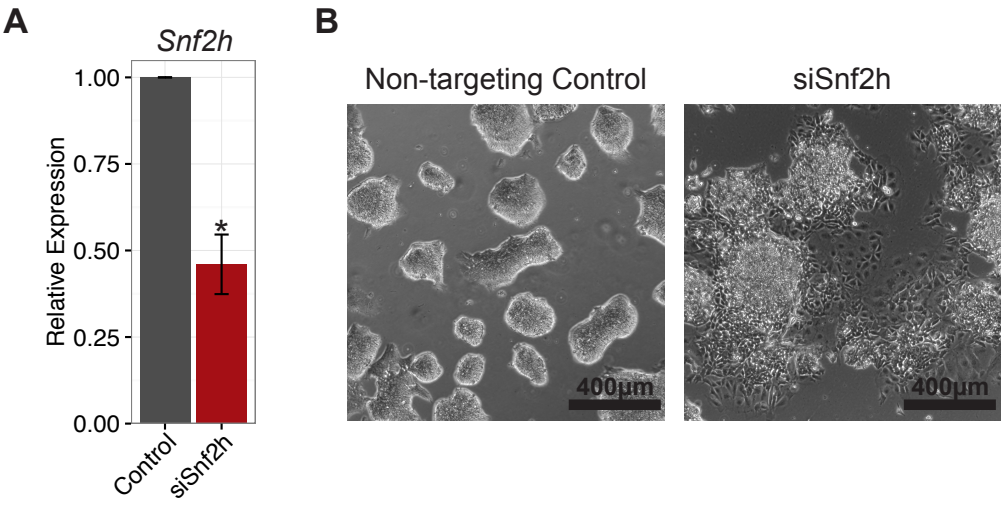


Figure 11. siRNA knockdown of *Snf2h* disrupts mESC morphology.
(A) Relative expression of *Snf2h* 4 days following siRNA transfection. Control cells were transfected with a non-targeting control siRNA. Gene expression was measured using RT-qPCR. N=3. * $P < 0.05$ (Student's t-test). **(B)** Phase contrast images *Snf2h*$^{fl/fl}$-mESCs 4 days following transfection with either a non-targeting control siRNA (left), or a siRNA targeting *Snf2h*. The morphological changes are consistent with those observed with Cre transfection.

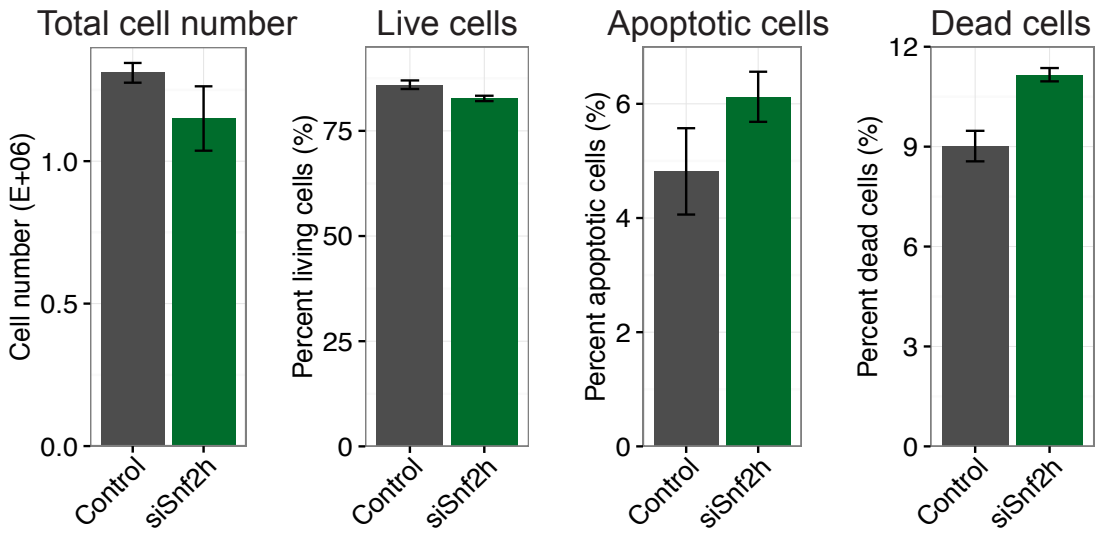


Figure 12. Snf2h knockdown does not affect apoptosis or cell viability at the time of morphological changes.
Annexin V and propidium iodide (PI) analysis of the proportion of living, apoptotic, and dead cells 4 days following the transfection of an siRNA targeting *Snf2h*, or a non-targeting control. Living cells were defined as Annexin V^{low}, PI^{low}; apoptotic cells were Annexin V^{high}, PI^{low}; and dead cells were Annexin V^{high}, PI^{high}. *Snf2h* knockdown does not significantly affect the apoptosis/viability status of cells (N=3, P=0.28, paired two-way ANOVA).

Snf2h expression is correlated with the pluripotent state

Given the evidence that ESC chromatin is highly dynamic and contains a large proportion of euchromatic regions[73–75], we predicted that if SNF2H were important for the maintenance of these dynamics, pluripotent ESCs would be associated with elevated expression of *Snf2h* relative to differentiated cell types. To explore this, we explored the expression levels of *Snf2h* throughout neural differentiation of human ESCs[76] and activin A-induced differentiation of mESCs[77] and found that *Snf2h* expression decreased over time in both cases **(Fig. 13A).** This suggests that its suppression may not be due to the specific approach used for differentiation, but rather, may be a general feature of differentiation.

mESCs can also be maintained in a more homogeneous and naive pluripotent state when cultured in the absence of serum, but with two small molecule kinase inhibitors (2i)[78,79]**.** This pluripotency "ground state" is characterized by a decreased deposition of the repressive H3K27me3 histone modification at the promoters of developmental genes and an overall reduction in the number of poised developmental genes[79]. By comparing single-cell RNA-Seq profiles from cells cultured in serum or 2i conditions, we found that cells cultured in 2i express approximately two-fold higher *Snf2h* expression **(Fig 13B).** The distribution of expression values across cells also suggests that this is not simply due to the decreased proportion of partially-differentiated cells seen in the 2i-cultured population. If this were the case, it would be expected that the distribution of *Snf2h* expression in 2i-cultured cells would be similar to serum-cultured cells, but with a reduced skew of low expression values. However, there is very little overlap in the distribution of expression values, suggesting this is not a result of decreased heterogeneity, but rather a feature of the ground state pluripotency. While the kinase inhibitors reduce differentiation cues, little is currently known about unique characteristics of chromatin state in 2i-cultured cells, and how SNF2H may be involved in their regulation.

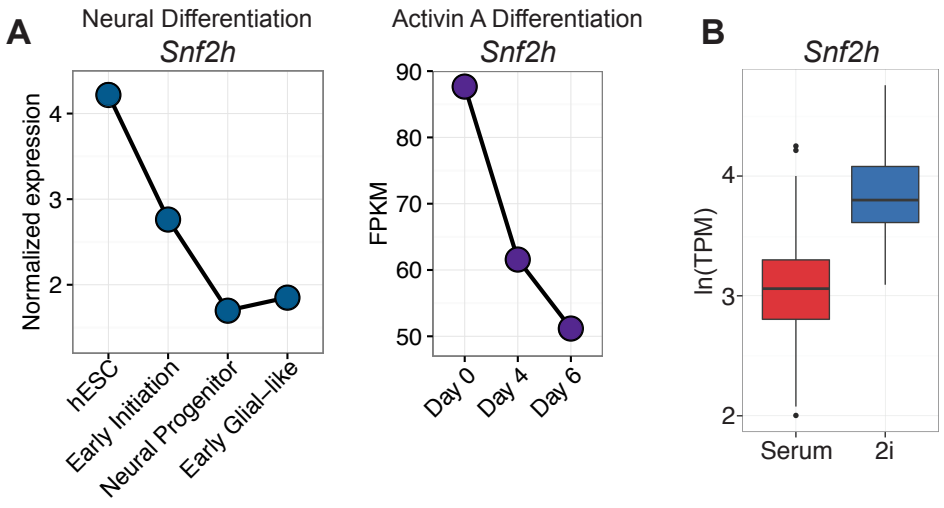


Figure 13. *Snf2h* expression is correlated with pluripotent state.
(A) RNA-Seq expression values of *Snf2h* throughout neural differentiation of human ESCs (left panel) and activin A-induced differentiation of mESCs. Data acquired from Wu *et al.*[76] (neural differentiation) and Xiao *et al.*[77] (activin A) **(B)** Boxplot of single-cell RNA-Seq expression values of mESCs cultured in either serum-containing media, or 2i media. Data acquired from Kumar *et al.*[80].

SNF2H is required for the establishment of pluripotency during reprogramming

While these data suggest that SNF2H is important for maintaining the pluripotent state, it is unclear if SNF2H is required for the acquisition of pluripotency during cellular reprogramming. Using RNA-Seq data from cells throughout the reprogramming process[81], we found that SNF2H is progressively upregulated throughout this process (**Fig 14A**), consistent with the observed correlation between its expression and pluripotency. We also explored data from a pooled lentiviral shRNA library screen used to identify critical genes for the reprogramming of human fibroblast-like cells to iPSCs[81] and found that cells expressing each of the 7 shRNAs within the library that target *Snf2h* were dramatically under-represented in the population of reprogrammed cells, suggesting a failure to reprogram (**Fig 14B**). These findings support that in addition to being

essential in maintaining pluripotency, SNF2H is also essential for the establishment of the pluripotent state.

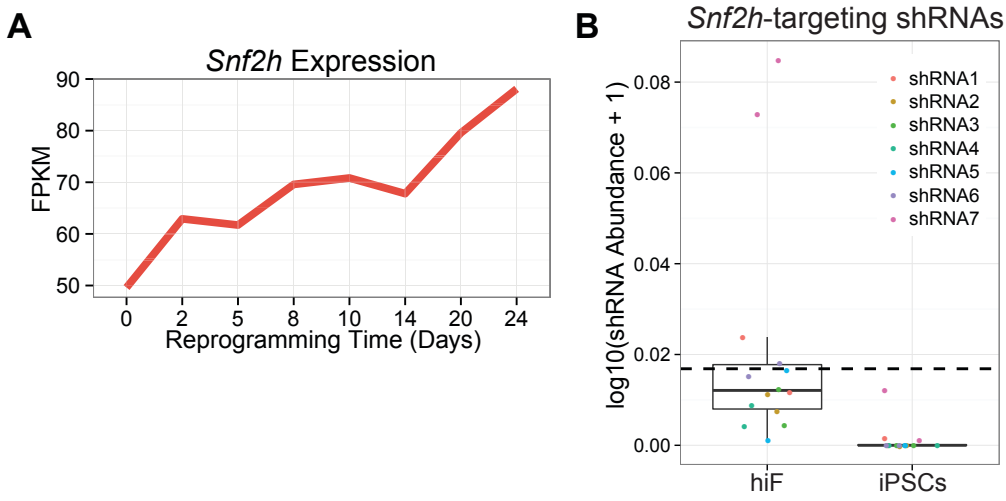


Figure 14. SNF2H and cellular reprogramming.
(A) RNA-Seq expression of *Snf2h* throughout 24 days of reprogramming human fibroblast-like cells (hiF) to induced pluripotent stem cells (iPSCs). Data acquired from Cacchiarelli *et al.*[81] **(B)** Proportion of integrated *Snf2h*-targeting shRNA vectors across a population of cells infected with a pooled shRNA library screen before (hiF) and after undergoing reprogramming to iPSCs. Dots represent seven unique shRNAs targeting *Snf2h*, and all are underrepresented following reprogramming. Data from Cacchiarelli *et al.*[81].

Generating a genome-wide profile of SNF2H localization

Despite its ubiquitous expression (**Fig 15**), there is an expanding body of evidence implicating SNF2H as an important regulator in distinct biological contexts, such as the regulation of neural stem cells[61], early embryonic development[63], and possibly developing follicles of the ovary[62]. We hypothesize that cell type-specific factors coordinate SNF2H localization across the genome, producing context-specific binding and regulation. These factors may serve to stabilize SNF2H at important regulatory loci,

facilitating efficient remodelling. In the context of pluripotency, we predicted that SNF2H would be enriched at promoters and enhancers of core pluripotency factors. To test this, we performed a preliminary ChIP-qPCR experiment to determine if it was bound to the promoters and super-enhancers[13] of the *Oct4, Sox2,* and *Nanog* genes, and found SNF2H enrichment relative to an IgG isotype control for all regulatory regions (**Fig 16**).

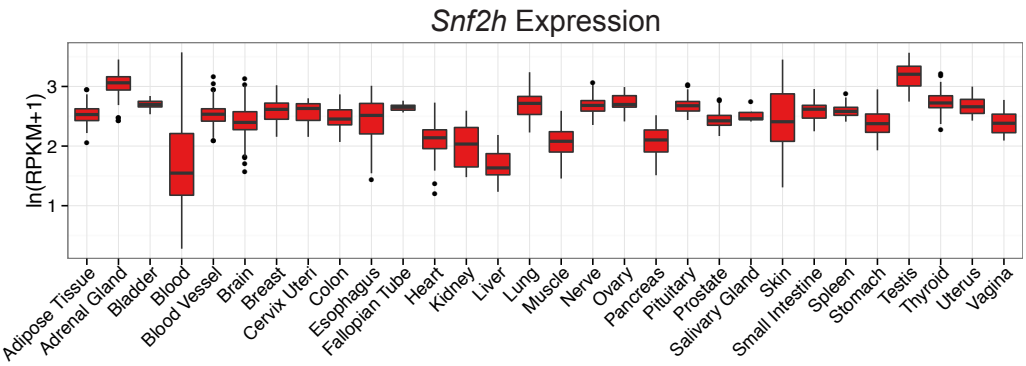


Figure 15. *Snf2h* is expressed ubiquitously across 30 tissues.
Boxplot representing RNA-seq expression values (RPKM) for *Snf2h* from 2923 samples, spanning 30 tissues. While there is a small degree of variability in expression between tissues, *Snf2h* expression is ubiquitous. RNA-Seq expression values were acquired from the GTEx Project[82] (Version V4; http://www.gtexportal.org/home/)

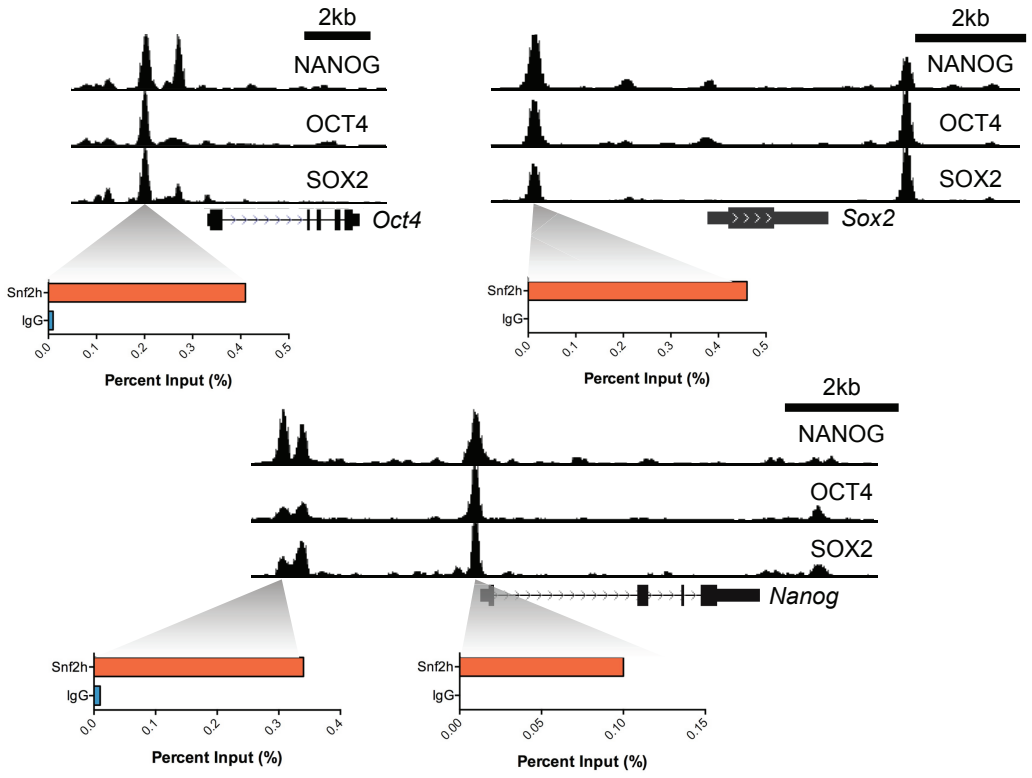


Figure 16. ChIP-qPCR analysis of SNF2H enrichment at key pluripotency regulatory regions.
Bar plots represent quantification (percent input) of immunoprecipitated DNA relative to an IgG isotype control. Promoters and enhancers were chosen by using ChIP-Seq data for OCT4, SOX2, and NANOG (signal tracks above bar plots) to identify regions where all three factors were co-localized upstream of the transcription start sites of the factors themselves. Primers were designed to amplify these regions and were used for qPCR analysis of the immunoprecipitated DNA.

While ChIP-qPCR allows for the interrogation of localization at specific loci, its resolution is limited. To capture genome-wide patterns of SNF2H binding, we performed MNase-based ChIP, followed by next-generation sequencing (ChIP-Seq). Strikingly, SNF2H is widely distributed across the genome (**Fig 17**). This introduces the interesting null hypothesis that SNF2H has no regional specificity across the genome at all and simply occupies all nucleosomes with equal probability. The use of MNase for DNA

fragmentation is beneficial because it enriches for nucleosomal DNA[23] from which chromatin is immunoprecipitated during the procedure. Thus under the null hypothesis that SNF2H has no apparent nucleosomal specificity, there would be no difference in the proportions of genomic regions represented in the input DNA prior to immunoprecipitation, and the DNA resulting from the ChIP procedure. All downstream analyses were performed with this consideration.

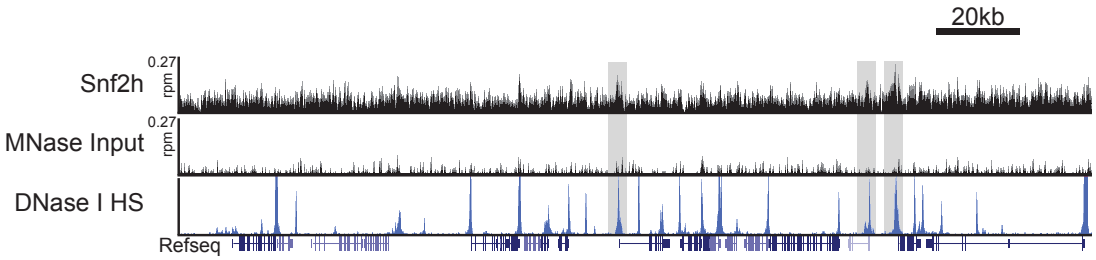


Figure 17. SNF2H is broadly distributed across the genome.
Genome browser track of SNF2H ChIP, MNase Input, and DNase I Hypersensitivity sequencing data across a region of chromosome 17. The x-axis represents a linear representation of the chromosome, while the y-axis represents the number of reads mapping along the region. Signal values for the SNF2H and MNase Input data sets have been normalized for sequencing depth and are represented as reads per million mapped reads (rpm). Shaded boxes represent examples of DNase I hypersensitivity sites that may have a higher proportion in the SNF2H data set than in the input data set. DNase I-Seq data was acquired from ENCODE (see Appendix III for source and reference).

SNF2H is enriched at regulatory elements and transcription factor binding sites

To explore the genomic specificity of SNF2H, we first compared the average profile of normalized read counts from the SNF2H ChIP and MNase-digested input samples around known transcription start sites (TSSs) and found that there was a genome-wide enrichment of SNF2H within an approximately 1kb window around the TSS, peaking at the TSS itself (**Fig 18A**). Segregating the data into two profiles of TSSs of genes that are expressed (FPKM > 0.1), or are not (FPKM < 0.1), based on mESC

RNA-Seq data[68], revealed that SNF2H preferentially localizes to the TSSs of transcriptionally active genes (**Fig 18B**). Given this difference, we wondered if there was a relationship between SNF2H enrichment and the transcription factors that drive gene expression. We generated profiles of average normalized read counts of SNF2H around the binding sites of a variety of transcription factors and cofactors (See Appendix III for a list of factors and sources). Genome-wide, there was an enrichment of SNF2H at the binding sites of the core transcription factors OCT4, SOX2, and NANOG (**Fig 19**), which tend to co-localize at the promoters and enhancers of actively transcribed genes[13,83]. Consistent with enhancer binding, SNF2H was also enriched at the binding sites of the enhancer-associated acetyltransferases P300 and CBP (**Fig 20**).

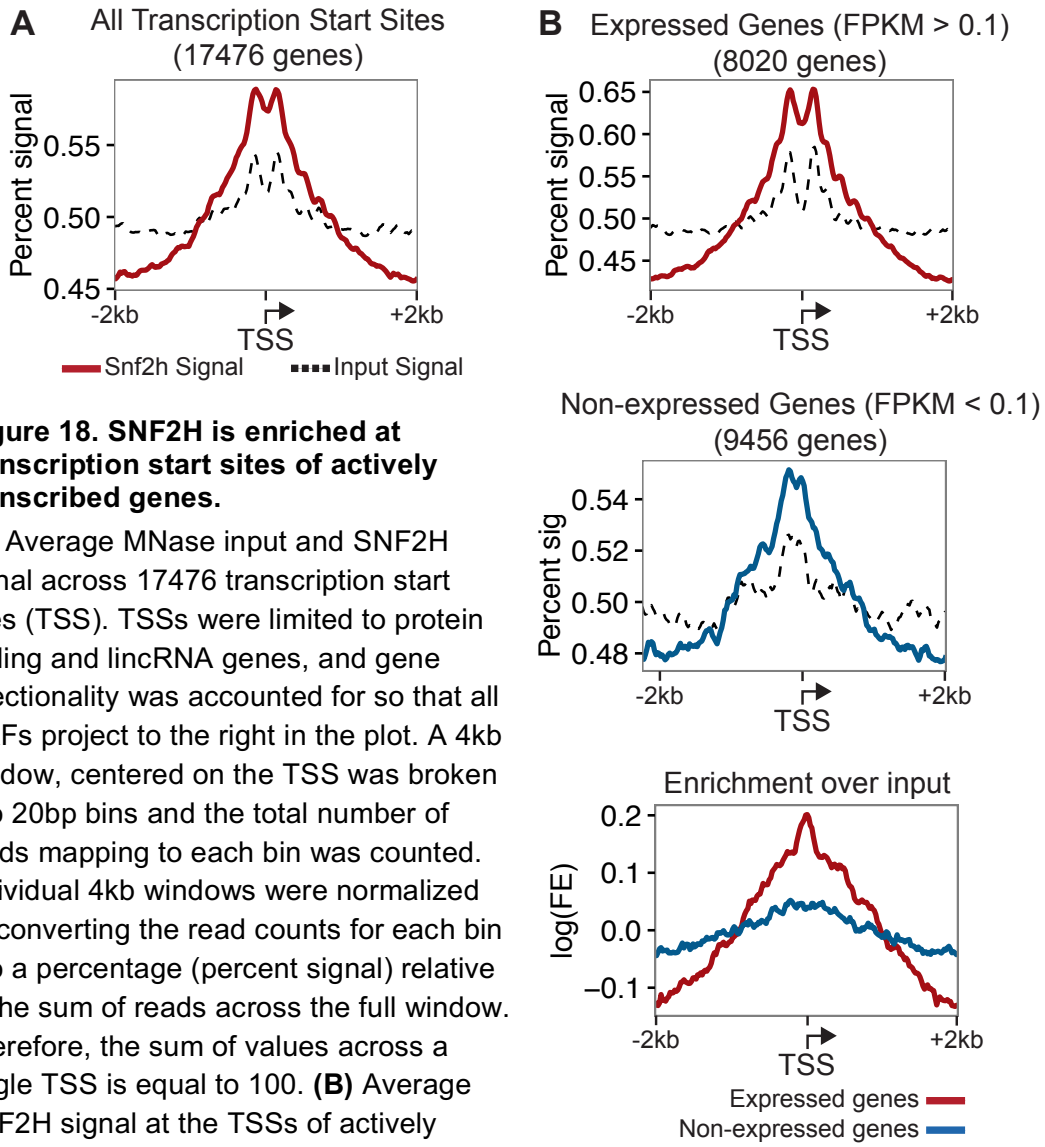


Figure 18. SNF2H is enriched at transcription start sites of actively transcribed genes.

(A) Average MNase input and SNF2H signal across 17476 transcription start sites (TSS). TSSs were limited to protein coding and lincRNA genes, and gene directionality was accounted for so that all ORFs project to the right in the plot. A 4kb window, centered on the TSS was broken into 20bp bins and the total number of reads mapping to each bin was counted. Individual 4kb windows were normalized by converting the read counts for each bin into a percentage (percent signal) relative to the sum of reads across the full window. Therefore, the sum of values across a single TSS is equal to 100. **(B)** Average SNF2H signal at the TSSs of actively transcribed (FPKM > 0.01; top panel, red) and non-expressed (FPKM < 0.01; middle panel, blue). The bottom panel demonstrates the preferential enrichment of SNF2H at expressed genes. Signal is represented as the log enrichment over input. At non-expressed genes, enrichment is maintained close to zero, consistent with little preferential binding towards the TSS. RNA-Seq data was acquired from ENCODE (see Appendix III for source and reference).

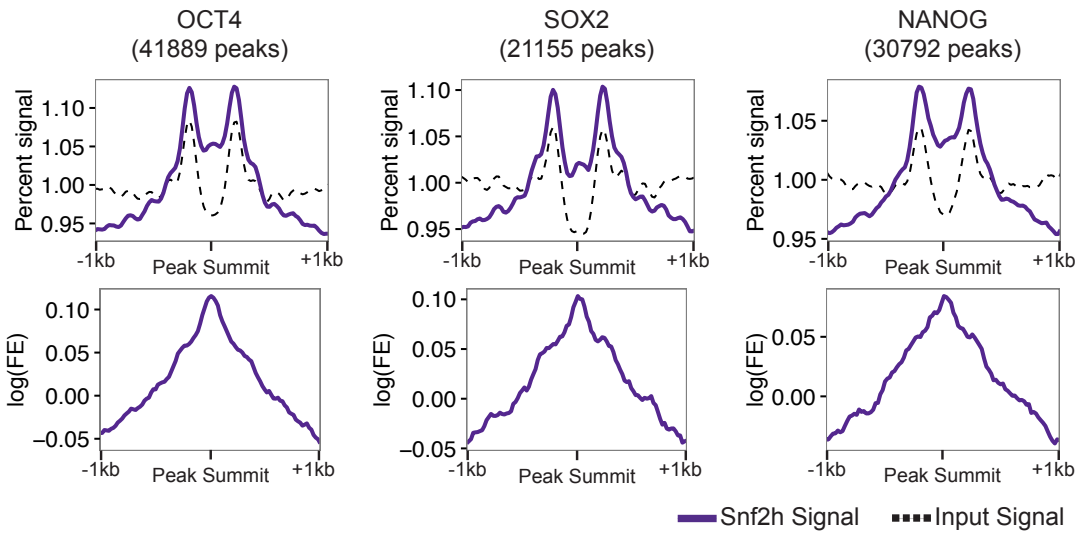


Figure 19. SNF2H enrichment at the binding sites of core transcription factors. Average MNase input and SNF2H signal across a 2kb window centered on the summit of OCT4, SOX2, and NANOG ChIP-Seq peaks. Top panels show average normalized SNF2H and Input signals across these windows, and the bottom panels represent the average log(SNF2H enrichment over input). SNF2H is most enriched at nucleosomes adjacent to the transcription factors' footprint, as well as in the nucleosome-poor region immediately surrounding the peak summit. Source of OCT4, SOX2, and NANOG ChIP-Seq data sets can be found in Appendix III.

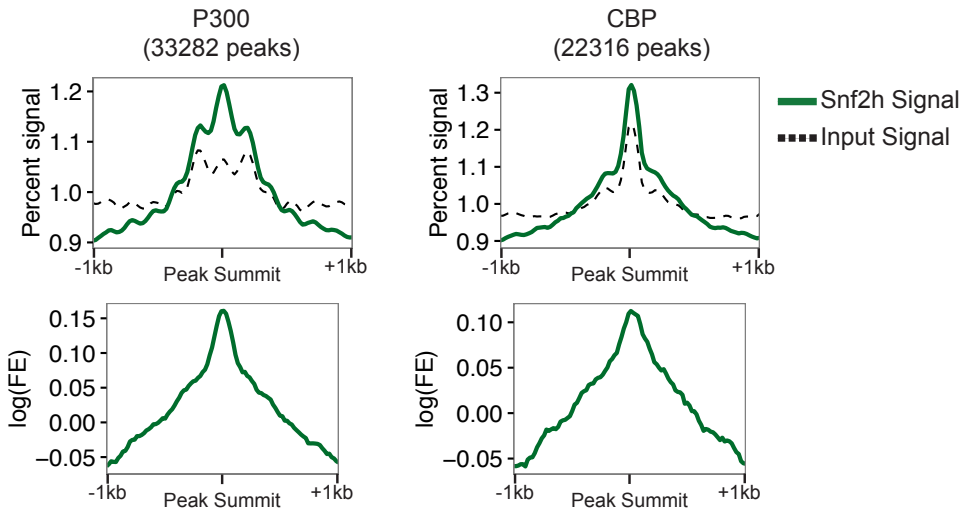


Figure 20. SNF2H enrichment at the binding sites of enhancer-associated factors. Average MNase input and SNF2H signal across a 2kb window centered on the summit of P300 and CBP ChIP-Seq peaks. Top panels show average normalized SNF2H and Input signals across these windows, and the bottom panels represent the average log(SNF2H enrichment over input). Source of P300 and CBP ChIP-Seq data sets can be found in Appendix III.

Polycomb group proteins are a group of highly-conserved factors that form repressive complexes to silence gene transcription through epigenetic modification[84]. To explore if SNF2H enrichment is also associated with polycomb-repressed regions, we calculated the enrichment of SNF2H at the binding sites of the polycomb repressive complex 2 (PRC2) components SUZ12 and EZH2 and found that SNF2H was enriched over the MNase-digested input (**Fig 21**). This suggests that SNF2H may be involved with the regulation of repressive states in addition to its relationship with active regulatory regions.

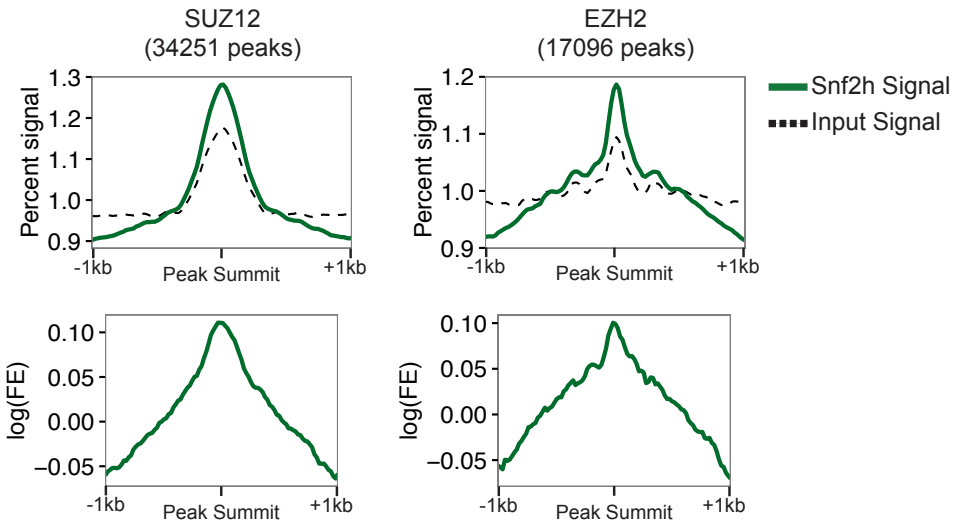


Figure 21. SNF2H enrichment at the binding sites of polycomb-associated factors. Average MNase input and SNF2H signal across a 2kb window centered on the summit of SUZ12 and EZH2 ChIP-Seq peaks. Top panels show average normalized SNF2H and Input signals across these windows, and the bottom panels represent the average log(SNF2H enrichment over input). Source of SUZ12 and EZH2 ChIP-Seq data sets can be found in Appendix III.

To further understand SNF2H localization more broadly across distinct chromatin states, we used a chromatin state map that segregates the mESC genome into 15 discrete states based on H3K36me3, H3K4m1, H3K27ac, H3K4me3, H3K27me3, RNA PolII, and CTCF ChIP-Seq profiles[85](**Fig 22A)**. This allows for distinction between different classes of regulatory regions that may be bound by some of the same factors, such as transcription factor binding at enhancers and promoters, or active promoters and repressed bivalent promoters both containing active chromatin marks[86]. The model successfully distinguished states with known characteristics of specific regulatory regions, such as active promoters marked with H3K4me3 but no H3K27me3[33], bivalent promoters with both modifications[86], and strong enhancers with the H3K27ac

modification[34]. To determine SNF2H enrichment across states, we pooled all regions of identical states, calculated SNF2H enrichment relative to the MNase-digested input for each region, and reported an average enrichment score for each state (**Fig 22B**). We found that SNF2H was most enriched at regions associated with regulating active transcription, such as strong enhancers and promoters. Interestingly, it is not enriched at repressed regions bearing the polycomb-associated H3K27me3 mark, but is enriched at bivalent regions containing both the H3K27me3 and H3K4me3 marks. SNF2H was not overly enriched across gene bodies of transcribed genes, or within heterochromatin, suggesting that it may not be involved with maintaining nucleosome positioning through stably-compacted chromatin. However, SNF2H was slightly enriched at CTCF-rich regions, consistent with insulators, suggesting that it may help maintaining the boundaries of heterochromatin and euchromatin, or other structural elements whose boundaries are enriched with CTCF, such as TADs[28], and CTCF- and RNA PolII-driven loop domains[30]. Indeed, we found that SNF2H was enriched at the binding sites of CTCF and SMC1 (**Fig 23**), a core component of the cohesin complex, which is involved in higher order chromatin structure, such as enhancer-promoter looping and sister chromatid cohesion[17,87].

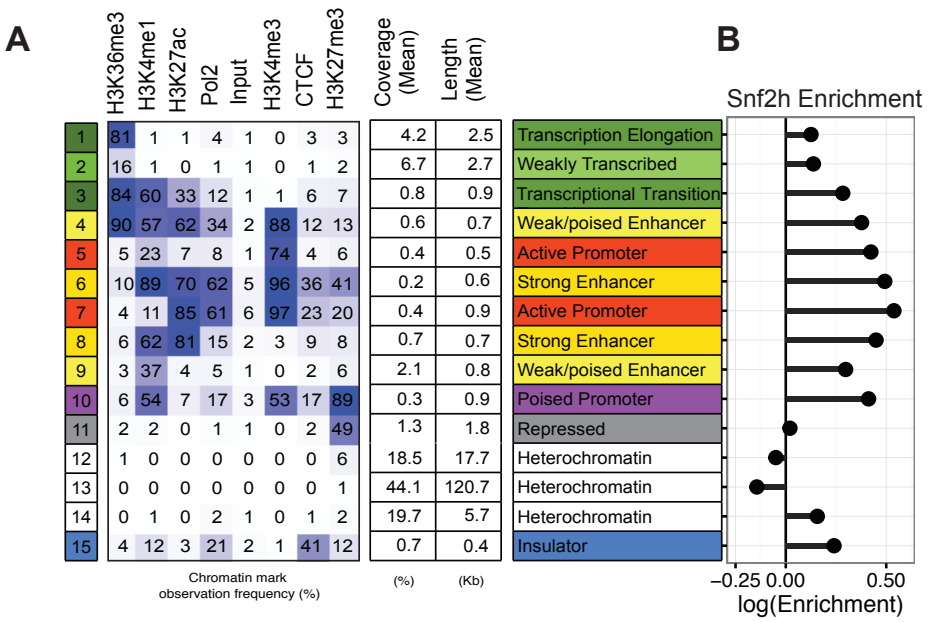


Figure 22. SNF2H enrichment across various chromatin states.
(A) Summary of a chromatin state map involving the segmentation of the genome into 15 discrete states. The heatmap displays the prevalence of various histone modifications and transcription factor binding within each state. Chromatin state annotation was produced with ChromHMM[88] and acquired from Bogu *et al.,* 2015[89]**.** Figure adapted from Bogu *et al.,* 2015[90]**. (B)** Plot of SNF2H enrichment over input for each chromatin state. All genomic regions of the same state were pooled, SNF2H and MNase input reads mapping to each region were counted, normalized for sequencing depth (rpm), and reported as a density (rpm/bp). Log(Enrichment) was calculated by dividing the density of SNF2H reads by the density of MNase input reads for each region and log transforming the resulting value. The mean of all log-transformed enrichment values for each state was reported.

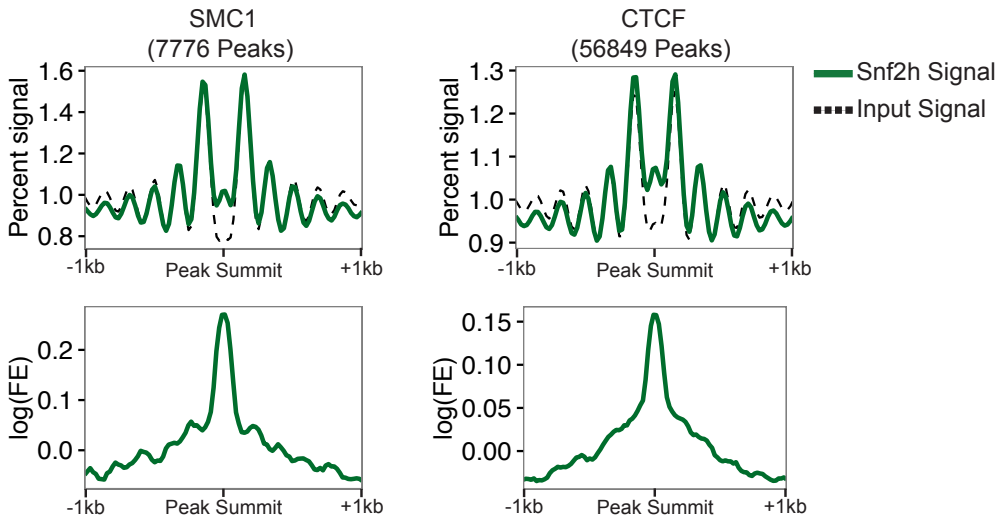


Figure 23. SNF2H enrichment at the binding sites of structural factors.
Average MNase input and SNF2H signal across a 2kb window centered on the summit of SMC1 (left panels) and CTCF (right panels) ChIP-Seq peaks. Top panels show average normalized SNF2H and Input signals across these windows, and the bottom panels represent the average log(SNF2H enrichment over input). Source of SMC1 and CTCF ChIP-Seq data sets can be found in Appendix III.

Along these lines, we also assessed SNF2H localization at the nucleosomes adjacent to broad nucleosome-free regions and throughout the regions themselves. We used a map of 233,253 DNase I hypersensitivity sites across the mESC genome[69], which define broad regions of nucleosome-free DNA, often broadly associated with regulatory regions and factor binding sites. We mapped SNF2H enrichment across the hypersensitivity sites and 1kb flanking regions and found that SNF2H was enriched at nucleosomes immediately adjacent to the site, as well as throughout the site itself, but not at more-distant nucleosomes (**Fig 24**). This supports that SNF2H may be involved with maintaining boundaries of genomic states, but further work will be required to understand the consequences of SNF2H deletion on nucleosome organization at these sites.

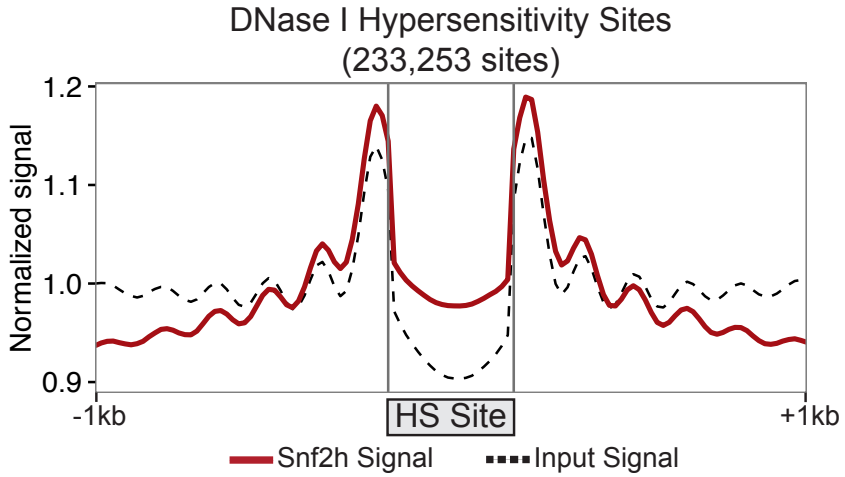


Figure 24. SNF2H enrichment at DNase I hypersensitivity sites.
Average MNase input and SNF2H signal at DNase I hypersenstivity sites and 1kb flanking regions. All DNase hypersensitivity sites, regardless of size (mean: 166bp), were broken into 20 bins, the number of reads mapping to each bin was calculated, normalized for sequencing depth by converting to reads per million mapped reads (rpm), and differences in bin size were accounted for by converting values to a density value (rpm/bp). This was repeated for 1kb flanking regions of each hypersensitivity site. To make visual interpretation of these values more accessible, the average rpm/bp for each window was calculated and the left-most input value was set to 1. The plot demonstrates that, like with transcription factor binding sites, SNF2H is most enriched at hypersensitivity sites and their flanking nucleosomes, suggesting a possible role in maintaining chromatin structure at these regions. Source of DNase-Seq data set can be found in Appendix III.

Discussion

Here, we have explored the importance of the ATP-dependent chromatin remodelling enzyme SNF2H in maintaining the pluripotent state in mESCs. Using both knockout and knockdown models, we found that SNF2H depletion disrupts characteristics of pluripotency, including cell morphology; expression of the transcription

factors OCT4, SOX2, and NANOG; and alkaline phosphatase activity. To better understand the implication of SNF2H in mESC regulation, we produced a genome-wide map of SNF2H localization and found that it is broadly distributed across the genome, but is enriched at active regulatory regions and transcription factor binding sites.

Snf2h deletion disrupts the pluripotent state

The pluripotent state is maintained by a set of several hundred indispensable genes[9–11], many of which are transcriptional regulators that support the pluripotency gene expression program. Disrupting the expression of these factors consistently results in morphological changes, leading to more flattened and dispersed cells, as well as reduced expression of the pluripotency transcription factors OCT4, SOX2, and NANOG. It should be noted, however, that these features are often involved in evaluating candidate "hits" in these screens (eg. the use of an *Oct4-GFP* reporter to look for shRNAs that reduced GFP expression), selecting for these phenotypic effects. It is likely the number of genes essential to maintain functional ESCs is much higher, but not all will result in positive hits with current screening strategies. For example, when the PRC2 component *Jarid2* is deleted, mESCs still express high levels of OCT4, SOX2, and NANOG, but are unable to undergo neural differentiation[21], which would fail to be picked up in most RNAi screens in mESCs.

Discovering factors that are indispensable to cell states is crucial to understand how the thousands of genes encoded in the genome work together to give rise to biological phenotypes. Here, we have identified the mammalian ISWI homologue SNF2H as an essential component of the pluripotent regulatory machinery, making the ISWI family of chromatin remodellers the last to be implicated in this role. Here, we have

demonstrated that disruption of SNF2H in mESCs results in morphological changes and a modest reduction in *Oct4*, *Sox2*, and *Nanog* expression. This has also been observed following the disruption of the BRG1[18], CHD1[19], and INO80[20] chromatin remodellers. Given the clear morphological effects, we suspect that the limited repression of pluripotency transcription factor is likely the combined result of early sample collection and heterogeneity in SNF2H depletion. If a pure population of *Snf2h*-null cells were cultured for an extended period of time, we predict that transcription factor expression would continue to fall over time, assuming the cells remain viable. This has been seen following the knockdown of *Brg1*, which only results in a complete reduction of OCT4, SOX2, and NANOG ten days after *Brg1* knockdown[18]. This prediction was unable to be tested in this study due to the resurgence of *Snf2h*-expressing cells in culture, which would increasingly dilute the measurable consequences of *Snf2h* deletion over time. A more homogeneous system, such as a clonal doxycycline-inducible Cre cell line, will be required to assess the long-term consequences of *Snf2h* deletion.

BRG1 depletion in mESCs also results in the outgrowth of wild-type cells in a co-culture system due to a defect in transitioning into S-phase[18]. The Annexin V assay four days following *Snf2h* knockdown confirmed that the observed morphological phenotype at this time did not coincide with changes in cell number, viability, or apoptosis. However, given the known roles of *Snf2h* in DNA replication and repair, it is highly possible that proliferative changes occur soon after, likely driving the decreased proportion of *Snf2h*-null cells observed over time. Given this, we suspect that the morphological change observed is a consequence of transcriptional changes from disrupted nucleosome positioning genome-wide, which could impair the ability to

maintain the pluripotency-associated gene expression program. Similar morphological changes are often associated with differentiation, however, this conclusion would be premature. Further work is required to determine if coordinated activation of lineage expression programs occurs in the absence of SNF2H, or if the observed phenotype is simply associated with sporadic transcriptional changes.

Given its ubiquitous expression and roles in distinct biological functions, we predict that cells lacking SNF2H would be unable to coordinate differentiation due to a general impairment in transcriptional regulation. This is in contrast to the disruption of specific transcription factors, which keeps the general regulatory machinery intact, allowing for coordinated transcriptional changes. Examples include the activation of endoderm and trophoblast expression patterns following the knockdown of *Oct4*[91], or trophectoderm differentiation following deletion of *Sox2* in mESCs[92]. Less is known about transcriptional consequences of disrupting chromatin regulators, which often regulate transcription more broadly, making it unclear if canonical expression changes can be expected. Knockdown of the chromatin remodeller INO80 in mESCs does result in the modest upregulation of the lineage markers *Cdx2*, *Fgf5*, *Nestin*, and *Pax3*[20], but it was not clear if coordinated gene sets were activated in individual cells, or simply if genes that happened to be repressed in mESCs became activated. Related to this, it is also not clear if these markers of distinct lineages are mutually exclusive in individual cells following Ino80 knockdown. This raises the question of whether individual cells are induced to differentiate down specific lineages, with some stochastic process determining which lineage cells commit to, or if individual cells undergo sporadic transcription changes that include the activation of repressed lineage markers. These two hypotheses

could lead to identical measurements of gene expression when using bulk RNA measurements across a population of cells, and therefore approaches with single-cell resolution may be required to capture the true transcriptional consequences following the deletion of *Snf2h*, or any other chromatin regulator.

SNF2H is broadly distributed across the mESC genome

SNF2H-mediated regulation of transcription is presumably through its activity as a nucleosome remodeller across the genome. ChIP-Seq analysis of its localization revealed a broad distribution across the genome. While the SWI/SNF, CHD, and INO80 families of remodellers are quite broadly distributed in mESCs themselves, they do tend to have a more sizable enrichment at target loci[20,93,94]. They all appear enriched at promoters of actively transcribed genes, but have been attributed several specific functions: INO80 localizes to the promoters of pluripotency-associated genes in an OCT4- and WDR5-dependent manner to increase transcription factor accessibility[20]; CHD remodellers localize to promoters of actively transcribed genes, with some having biases for the -1 or +1 nucleosome around the TSS[94]; and Brg1 has been primarily attributed with maintaining active transcription at LIF/STAT3 targets and antagonizing PRC2 from transcriptional repression[95]. The only directly repressive role that has been demonstrated has been the suppression of divergent noncoding transcription by Brg1 through maintaining high nucleosome occupancy adjacent to nucleosome-free regions[39]. It is unclear how many of these features are truly unique to the remodeller studied, or if they are simply biased due to specific experimental designs.

The broad distribution of *Snf2h* across the genome may be attributed to several factors. First, ISWI family remodellers lack protein domains that could provide them with

increased specificity, such as the bromodomains and chromodomains present in SWI/SNF and CHD family remodellers, respectively[38]. These domains allow the remodellers to interact with acetylated and methylated histones, which are prominent modifications at active promoters and enhancers. Second, SNF2H-containing remodelling complexes are relatively small, comprising 2-4 protein subunits, in contrast to complexes of other families, such as SWI/SNF BAF complexes, which are 12 proteins in size[38]. Increased numbers of protein subunits may provide additional specificity to the remodelling complex. For example, several cell-type specific BAF complexes have been characterized, in which different protein subunits are present and contain their own bromodomains, chromodomains, or PHD finger domains[46]. In several cases, these cell type-specific subunits have been shown to directly interact with key transcription factors of the given cell type[46]. And third, given the continuous sampling model of mammalian ISWI proteins, where they are constantly forming transient interactions with nucleosomes across the genome, but require additional stabilization for efficient remodelling[58], it could be expected that these transient interactions would be captured during the crosslinking step of the ChIP protocol, contributing to this genome-wide signal.

SNF2H preferentially localizes to active regulatory regions

Despite the broad distribution across the genome, we have shown that SNF2H exhibits preferential localization, albeit with modest enrichment, at the binding sites of several transcription factors and cofactors, and around the TSSs of actively transcribed genes. Localization at TSSs is consistent with observed localization of other remodellers, however it is currently unclear what role SNF2H plays at these promoters. BRG1 has been shown to maintain nucleosome-free regions at promoters in mESCs, competing with

the CHD4 remodeller, whose depletion causes an increase in chromatin accessibility, suggestive of role in chromatin compaction[94]. Depletion of INO80 in mESCs results in decreased DNase I sensitivity at regulatory regions and reduced localization of pluripotency-associated transcription factors and RNA PolII, suggesting a role in maintaining open chromatin and allowing transcriptional machinery access to regulatory DNA[20]. In contrast, depletion of the EP400 remodeller does not affect broad regions of accessible chromatin, but rather affects the precise location of the first nucleosome before the TSS (-1 nucleosome)[94]. There is evidence that SNF2H is involved in stabilizing the first nucleosome after the TSS (+1 nucleosome), increasing RNA PolII pausing and decreasing the rate of transcription[96]. We did find SNF2H enriched at the +1 nucleosome, but not any more than the -1 nucleosome. This does not exclude the possibility that it is involved with positioning the +1 nucleosome in mESCs, but it likely has additional roles given its enrichment elsewhere.

It should also be noted that despite these defined functions, gene expression analysis of cells lacking individual remodellers results in both the significant upregulation and downregulation of remodeller-bound genes[20,94,95], suggesting that it may not be appropriate to cast “activating” or “repressive” labels on remodellers. Genomic context, and in particular, recruitment and stabilization cues likely determine the regulatory effect of a remodeller. In fact, CHD4 and EP400 have been shown to facilitate active transcription of H3K4me3-marked promoters, but upon their depletion, repressed bivalent genes become activated[94], suggesting that some feature of bivalent promoters that is absent in active promoters may drive these remodellers to prevent transcription. We found that SNF2H was also roughly equally enriched at active and bivalent

promoters, suggesting it may be playing dual roles in promoter activation and repression as well. Gene expression analysis following SNF2H depletion will be important in future studies to provide insight into different regulatory roles that SNF2H may be involved with across active, bivalent, and inactive promoters.

Chromatin state analysis allowed for exploring SNF2H enrichment at non-promoter regions. This confirmed that, in addition to promoters, SNF2H also preferentially localizes to enhancer regions. Defined functions or localization patterns of remodellers at enhancer regions are less understood, likely due to the difficulty in identifying enhancer regions in a particular cell type. Depletion of BRG1 and CHD4 in mESCs affects the number of broad nucleosome-free regions genome-wide[94], and while enhancer regions weren't specifically addressed, they are included among these hypersensitivity regions. Ino80 has recently been shown to bind 90% of super-enhancers[13,97] in melanoma cells, where it reduces nucleosome occupancy and allows for the recruitment of the mediator complex[98]. Interestingly, its localization is dependent on the binding of melanoma-associated transcription factors[98]. In *Drosophila melanogaster*, the ISWI-containing NURF complex has been shown to interact with transcription factors at enhancers to establish nucleosome-depleted regions[99]. In contrast, SNF2H has been shown to bind to the enhancer upstream of the SPI1 proto-oncogene in human leukemia cells, where it represses its transcription[100]. Given that SNF2H was most enriched at strong enhancers bearing high levels of H3K27ac[34], and that *Oct4*, *Sox2*, and *Nanog* are repressed following SNF2H depletion, we suspect that it is unlikely that SNF2H is playing a repressive role at these regions.

Models of SNF2H recruitment genome-wide

Given its ubiquitous expression and roles in diverse cellular contexts, we predict that cell type-specific features or factors are crucial for targeting SNF2H to regulatory regions across the genome. One model of SNF2H recruitment across the genome involves protein-protein interactions increasing the affinity of SNF2H at key regulatory regions (**Fig 25**). By comparing genome-wide localization profiles of SNF2H and a variety of transcription factors and cofactors, we found that SNF2H was particularly enriched at the binding sites of OCT4, SOX2, and NANOG; the enhancer-associated acetyltransferases CBP and P300; and the PRC2 components SUZ12 and EZH2. While only correlative, these factors would be ideal candidates to explore the recruitment of SNF2H to active and bivalent promoters, and enhancer regions. Both BRG1 and INO80 have been shown to interact with OCT4 in co-immunoprecipitation studies[18,20], and INO80 localization to promoters is dependent on OCT4 and WDR5 binding[20], suggesting a role in remodeller recruitment. BRG1 has also been shown to interact with over 20 different transcription factors and cofactors across a variety of cell types[101].

During myogenesis, temporal studies have shown that, at the myogenin promoter, MYOD binding and histone acetylation precede BRG1-mediated remodelling[102], and that the histone acetylation may support BRG1 localization through interacting with its bromodomain[103]. This supports that transcription factor binding and histone modifications occur, at least in some contexts, prior to remodeller binding, and may serve as a recruitment signal. Paradoxically, functional BRG1 is also required for stable MYOD binding to the myogenin promoter[102]. Together, these findings suggest that there is dynamic interplay between remodellers, transcription factors, and histone modifications in establishing and maintaining regulatory regions. However, SNF2H lacks

the bromodomains and chromodomains that are thought to be the primary means of recruiting SWI/SNF[103] and CHD[44,104] family remodellers, respectively, to target loci. Therefore, SNF2H may not rely on histone modifications for recruitment, unless it forms interactions through undefined protein domains, or interacts with specific proteins that serve as adapters to anchor onto these modified histones. Co-immunoprecipitation studies and targeted loss-of-function assays will greatly improve our understanding of the SNF2H protein interactome and how it may impact its localization and/or function across the genome.

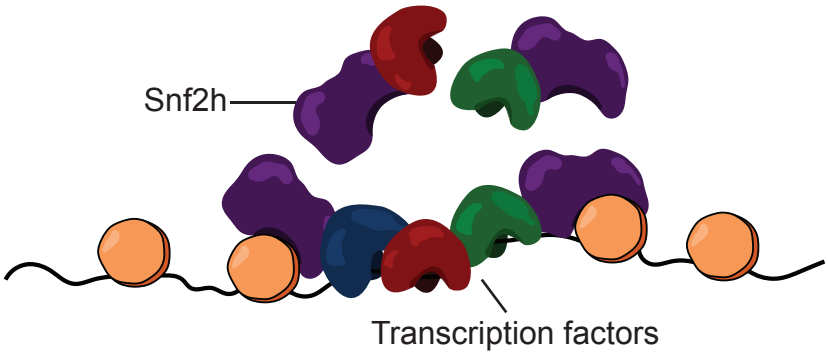


Figure 25. SNF2H recruitment model: protein-protein interactions.
In this model, SNF2H acquires affinity for regulatory regions across the genome through direct, or indirect interactions with DNA-binding factors. This may include transcription factors, cofactors, or structural proteins within the nucleus.

A second possible model of SNF2H recruitment across the genome may involve an increased affinity at regulatory regions through interactions with RNA (**Fig 26**). Non-coding RNAs have increasingly become a topic of interest for studies of epigenetic regulation and the recruitment of chromatin regulators. Perhaps the most studied example would be the non-coding RNA *Xist*, which is expressed off the X chromosome, coats the

entire chromosome, and recruits PRC2 to establish H3K27me3 modifications, facilitating X chromosome inactivation[105,106]. Isolation of RNAs from both the soluble, and chromatin pellet extract components of a nuclear fractionation has demonstrated that thousands of non-coding RNAs associate with chromatin at active regulatory regions, and that their presence positively correlates with the transcriptional activity of adjacent genes[107]. Additionally, many chromatin regulators interact with a large number of both coding and non-coding RNAs[108], and they do not all interact with the same sets of RNA, suggesting that it isn't simply an artifact of certain RNAs being more amenable to general protein interactions[108]. In several well-studied contexts, it has been demonstrated that interactions with non-coding RNAs facilitate the localization of chromatin regulators to specific loci across the genome[109–112]. For example, the SWI/SNF remodeller SNF5 has been shown to interact with the long intergenic non-coding RNA (lincRNA) *SchLAP1* in prostate cancer, which prevents SNF5 binding across the genome[113].

Guttman *et al.*[114] performed a shRNA knockdown screen against 226 lincRNAs expressed in mESCs and found that, of the 147 lincRNA successfully disrupted, 137 produced global gene expression changes (affecting 175 protein-coding mRNAs on average). RNA immunoprecipitation of 12 histone modifiers showed that many of these lincRNA interacted with histone modifiers, and all 12 modifiers interacted with at least 9 lincRNA[114]. It is possible these lincRNA exert their transcriptional regulation through anchoring these chromatin regulators, and perhaps other regulators, such as SNF2H, to specific target loci across the mESC genome. Assessing the interactions between SNF2H and various RNAs, along with approaches to disrupt SNF2H-RNA interactions would

provide valuable insight into the role of RNAs in stabilizing SNF2H at regulatory regions across the genome.

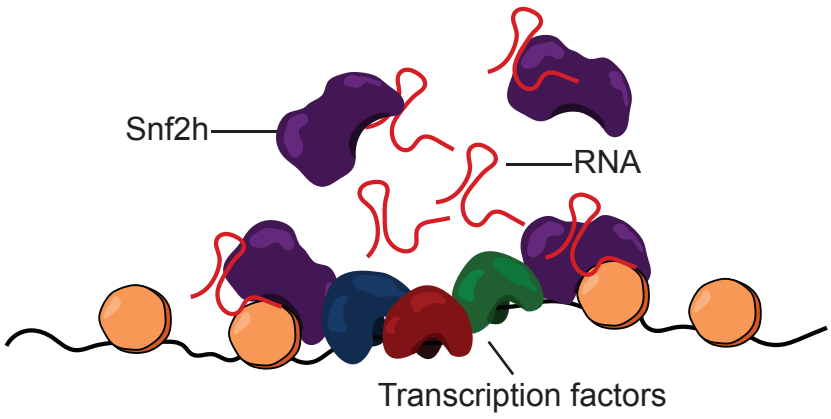


Figure 26. SNF2H recruitment model: protein-RNA interactions.
In this model, SNF2H acquires affinity for regulatory regions across the genome through direct, or indirect interactions with RNA molecules. This may include intergenic non-coding RNAs that recruit SNF2H to adjacent genes, or perhaps non-coding RNA transcribed directly from the regulatory regions bound by SNF2H. These interactions may directly provide specificity through RNA-DNA interactions, or by associating with DNA and other proteins through RNA secondary structure.

An intriguing third model of SNF2H recruitment is one of negative selection, where chromatin state directly impacts localization: nucleosomes at regulatory regions are preferentially bound because they are simply more accessible, while chromatin conformation, or even particular molecular cues serve to deter SNF2H localization from repressed chromatin (**Fig 27**). Unfortunately, there is no direct evidence for this model, likely because it is a difficult hypothesis to test, requiring thorough analysis of chromatin state after all perturbations. However, it does fit as an alternative hypothesis to many observations in the literature. For example, INO80 does not localize to the promoters of mESC genes following the depletion of OCT4 or WDR5[20], but in the absence of these

factors, mESCs undergo widespread transcriptional changes that have been associated with a differentiated phenotype[14,91]. It is highly likely that the promoters of mESC genes undergo changes to their histone modifications and three dimensional conformation, perhaps simply affecting the accessibility of these regions to INO80. It would be informative to see if INO80 simply localizes to newly accessible regions in cells lacking OCT4 and WDR5, but the problem with interpretation would persist: do newly expressed transcription factors change recruitment patterns, or is INO80 simply localizing to the most accessible regions? The ideal scenario for distinguishing between these two interpretations would be where it could be confirmed that the depletion of a remodeller and a candidate recruitment factor has no impact on chromatin state, and thus reduced localization of the remodeller was a consequence of factor depletion and not of epigenetic changes. Unfortunately, this is not likely to happen given that the depletion of regulatory factors almost always results in transcriptional changes to hundreds of genes, and are presumably accompanied with corresponding epigenetic changes. It will be important to always consider this model when studying the recruitment of chromatin modifiers, as it may help guide experimental design or may explain peculiar findings.

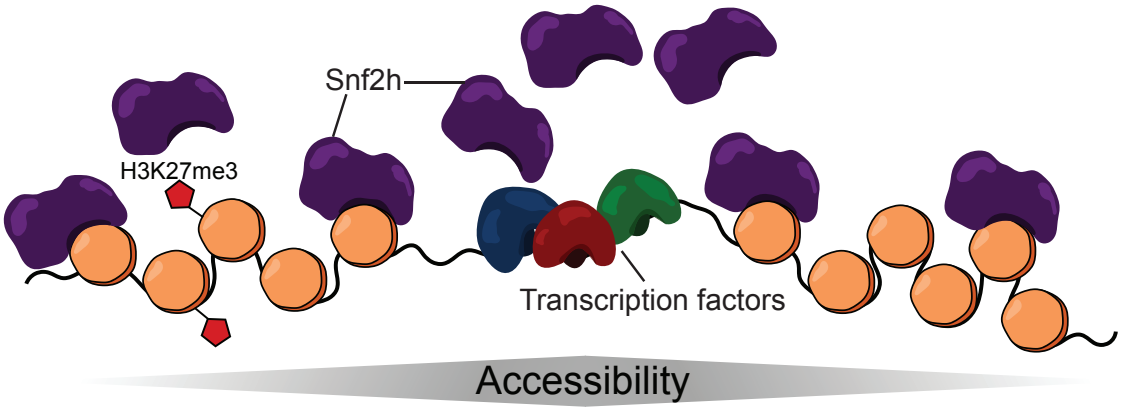


Figure 27. SNF2H recruitment model: negative selection.
In this model, SNF2H does not receive cues that increase its affinity at regulatory regions, but rather, SNF2H has a naturally high affinity for nucleosomes and preferentially binds regulatory regions due to their high accessibility. SNF2H localization to non-regulatory regions is hindered due to decreased accessibility from chromatin compaction, or directly through histone modifications that prevent its binding.

SNF2H and higher-order chromatin structure

We found that SNF2H was enriched at nucleosomes adjacent to DNase I hypersensitivity sites. This may be important for maintaining stable boundaries between distinct chromatin states, preventing nucleosomes from distributing within nucleosome-free regions. Similarly, it has been shown that BRG1 is important in maintaining nucleosome free regions[39,94], but is involved in ensuring high nucleosome occupancy immediately adjacent[39], establishing a border between nucleosome-free regions and adjacent compact chromatin. Beyond regulating transcription factor accessibility at regulatory regions, SNF2H may also be involved in maintaining chromatin structure through nucleosome positioning at the borders of chromatin states. Chromatin state analysis showed that SNF2H was modestly enriched at a state enriched with CTCF, a DNA-binding factor classically attributed to insulator regions[115]. Recent advances have

demonstrated that CTCF plays a crucial role in establishing functional boundaries of chromatin through facilitating chromatin looping[28–30,115]. RNA PolII, which is not typically associated with insulators, has also been shown to facilitate DNA looping[30] and is also enriched within the identified chromatin state. Wiechens *et al.*[116] recently demonstrated that SNF2H is important for positioning nucleosomes at CTCF binding sites, and cells lacking SNF2H have decreased CTCF binding. We confirmed that SNF2H is also localized to CTCF binding sites genome-wide in mESCs. Also, transmission electron microscopy of neurons lacking SNF2H has shown broad disruptions to chromatin organization following its deletion[61]. Together, these data suggest that in addition to regulating nucleosome dynamics at regulatory regions, SNF2H may also facilitate broad chromatin organization through regulating CTCF binding genome-wide. Integration of SNF2H and nucleosome positioning data with chromatin topology data would allow for further exploration of SNF2H enrichment and nucleosome positioning at loop boundaries. Chromatin conformation capture experiments following *Snf2h* deletion may also provide valuable insight into the consequences of *Snf2h* deletion on chromatin looping.

Conclusion

We have demonstrated that the ISWI chromatin remodeller SNF2H is an essential regulator of the pluripotent state—a member of the hundreds of factors that are crucial for maintaining pluripotency. This also makes the ISWI family of ATP-dependent chromatin remodelling enzymes the last to be implicated in this role. The work described here has provided a foundation of information that can be used to motivate further studies on the specific functions of SNF2H within ESCs. To this end, we believe there are three

important questions that deserve further focus: First, what are the consequences of *Snf2h* deletion on nucleosome positioning and chromatin states across the mESC genome? Second, and related, what are the transcriptional consequences of these changes? It would be particularly informative to study this at a single-cell resolution to determine if the consequences occur in a deterministic manner across cells, or if a stochastic process drives heterogeneous changes that are masked by bulk-cell measurements. And third, what targets SNF2H to active regulatory regions across the genome? We have proposed three models of how SNF2H may be recruited across the genome, none of which are necessarily exclusive. Studying the SNF2H interactome and using careful experimental design to disrupt interactions, while not confounding the results with unrelated chromatin state changes, will be highly valuable to address this question.

More broadly, we wonder how generalizable this model of SNF2H-mediated regulation is. There's no apparent reason to believe that the pluripotent state is special with regards to requiring chromatin remodelling to regulate nucleosome positioning. Given the ubiquitous expression of *Snf2h* across cell types, it would not be surprising to discover a role in all cell types. If this model of SNF2H as a general regulatory tool is true however, it becomes increasingly important to understand how SNF2H, and other ubiquitous regulators, are recruited across the genome: what determines its use?

Ultimately, this work has added to the increasingly complex model of the regulatory system in embryonic stem cells. Understanding this system is valuable for both theoretical and practical purposes. Learning new information about key regulators and regulatory principles allows us to revise our model cell state and how transitions between states occur. This allows for making informed predictions about how normal

cells can become pathological, or how we can drive specific cell state changes for therapeutics—to make brain cells from a patient's skin, or to turn cancer cells off.

Appendices

Appendix I. Primer sequences

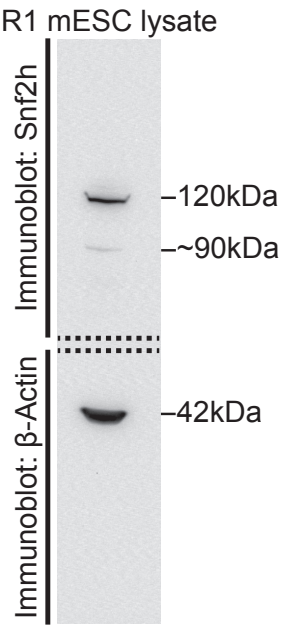
Application	Target	Sequence (5'-3')
Genotyping	*Smarca5*-F	GGCCAGCCCTGTCTTATATTT
	Smarca5-R	CAAGCAAACAGACCTAAGTGTTG
Gene expression (qPCR)	*Smarca5*-F	GTTGGAGATTACCGACACCGT
	Smarca5-R	GCCAATTCAATCCTCGCACC
	Oct4-F	ATTTTGTCCTCAGTGGGGCG
	Oct4-R	ACACCTTTCCAAAGAGAACGC
	Sox2-F	CAAAAACCGTGATGCCGACT
	Sox2-R	CGCCCTCAGGTTTTCTCTGT
	Nanog-F	TACCTCAGCCTCCAGCAGAT
	Nanog-R	CTGGTGCTGAGCCCTGAATC
ChIP-qPCR	*Oct4*-Enhancer-F	CGGCAGATGCATAACAAAGG
	Oct4-Enhancer-R	GCCAAGTTCACAAAGCTTCC
	Sox2-Enhancer-F	CTCAAATGCAGATGCAGGAG
	Sox2-Enhancer-R	GCGAGAACTAGCCAAGCATC
	Nanog-Enhancer-F	CAGAGGACCCACTTAACATTCC
	Nanog-Enhancer-R	CAGGGAAGCGGTTTGAATAG
	Nanog-Promoter-F	TGCAGGTGGGATTAACTGTG
	Nanog-Promoter-R	CTACCCACCCCCTATTCTCC

Appendix II. ChIP DNA quantification and antibody specificity

A

	DNA Quantification
Input (total)	36.85µg
Histone H3	834.5ng
IgG	0ng
Snf2h Rep #1	163.5ng
Snf2h Rep #2	190ng

B



Appendix III. Accession numbers of all public data sets used

Raw data used				
Factor	**Sample Accession Number**	**Series Accession Number**	**Read Count ($x10^6$)**	**Control used for peak calls**
Input (GSM1082343)	GSM1082343	GSE44286	59.8	
OCT4	GSM1082340	GSE44286	50	Input (GSM1082343)
SOX2	GSM1082341	GSE44286	68.1	Input (GSM1082343)
NANOG	GSM1082342	GSE44286	60.9	Input (GSM1082343)

Input (GSM594599)	GSM594599	GSE24164	13.7	
P300	GSM594600, GSM594601	GSE24164	15.2, 17.3	Input (GSM594599)
CBP	GSM1246866	GSE51522	22.2	N/A
Input(GSM480164)	GSM480164	GSE18776	6.4, 8.9, 9.3, 9.4	
EZH2	GSM480161	GSE18776	1.6, 5.5, 8.9	Input (GSM480164)
SUZ12	GSM480162	GSE18776	33	Input (GSM480164)
Input (GSM560357)	GSM560357	GSE22562	7.3	
SMC1	GSM560342	GSE22562	30.4	Input (GSM560357)
Input (GSM918754)	GSM918754	GSE36027	14.1, 19.0	
CTCF	GSM918748	GSE36027	19.4, 18.5	Input (GSM918754)
Processed data used				
Data Type	**Cell Type**	**Reference**	**Source (when applicable)**	
RNA-Seq	mESC	Mouse ENCODE Consortium *et al.*[69]	ENCODE (http://chromosome.sdsc.edu/mouse/download.html)	
RNA-Seq	30 Human Tissues	GTEx Consortium [82]	GTEx V4 (http://www.gtexportal.org/)	
RNA-Seq	hESC neural differentiation	Wu *et al.*[76]		
RNA-Seq	mESC activin-A differentiation	Xiao *et al.*[77]		
RNA-Seq	Reprogramming fibroblasts to	Cacchiarelli *et al.*[81]		

	iPSCs		
shRNA Library Screen (gDNA)	Reprogramming fibroblasts to iPSCs	Cacchiarelli *et al.*[81]	
DNase-Seq	mESC	Shen *et al.*[68]	ENCODE (http://chromosome.sdsc.edu/mouse/download.html)
ChromHMM State Map	mESC	Bogu *et al.*[85]	(https://github.com/gireeshkbogu/chromatin_states_chromHMM_mm9)

References

1. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292,** 154–156 (1981).
2. Chambers, I. & Smith, A. Self-renewal of teratocarcinoma and embryonic stem cells. *Oncogene* **23,** 7150–7160 (2004).
3. De Los Angeles, A. *et al.* Hallmarks of pluripotency. *Nature* **525,** 469–478 (2015).
4. Beddington, R. S. & Robertson, E. J. An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development* **105,** 733–737 (1989).
5. Avior, Y., Yishai, A., Ido, S. & Nissim, B. Pluripotent stem cells in disease modelling and drug discovery. *Nat. Rev. Mol. Cell Biol.* **17,** 170–182 (2016).
6. Hall, B., Bradford, H., Advait, L. & Kulkarni, A. B. in *Current Protocols in Cell Biology* (2009).
7. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).
8. Yan, L. *et al.* Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* **20,** 1131–1139 (2013).
9. Ding, L. *et al.* A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity. *Cell Stem Cell* **4,** 403–415 (2009).
10. Hu, G. *et al.* A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes Dev.* **23,** 837–848 (2009).

11. Chia, N.-Y. *et al.* A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature* **468,** 316–320 (2010).
12. Young, R. A. Control of the embryonic stem cell state. *Cell* **144,** 940–954 (2011).
13. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153,** 307–319 (2013).
14. Ang, Y.-S. *et al.* Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* **145,** 183–197 (2011).
15. Fang, F. *et al.* Coactivators p300 and CBP maintain the identity of mouse embryonic stem cells by mediating long-range chromatin structure. *Stem Cells* **32,** 1805–1816 (2014).
16. Balakrishnan, S. K., Witcher, M., Berggren, T. W. & Emerson, B. M. Functional and molecular characterization of the role of CTCF in human embryonic stem cell biology. *PLoS One* **7,** e42424 (2012).
17. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **472,** 247–247 (2011).
18. Ho, L. *et al.* An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 5181–5186 (2009).
19. Gaspar-Maia, A. *et al.* Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature* **460,** 863–868 (2009).
20. Wang, L. *et al.* INO80 facilitates pluripotency gene activation in embryonic stem cell self-renewal, reprogramming, and blastocyst development. *Cell Stem Cell* **14,** 575–591 (2014).

21. Landeira, D. *et al.* Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. *Nat. Cell Biol.* **12,** 618–624 (2010).

22. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* (2016). doi:10.1038/nrg.2016.59

23. Jiang, C. & Pugh, B. F. Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.* **10,** 161–172 (2009).

24. Bannister, A. J. & Tony, K. Regulation of chromatin by histone modifications. *Cell Res.* **21,** 381–395 (2011).

25. Smith, Z. D. & Alexander, M. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14,** 204–220 (2013).

26. Dekker, J. & Mirny, L. The 3D Genome as Moderator of Chromosomal Communication. *Cell* **164,** 1110–1121 (2016).

27. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326,** 289–293 (2009).

28. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).

29. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159,** 1665–1680 (2014).

30. Tang, Z. *et al.* CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell* **163,** 1611–1627 (2015).

31. Sutherland, H. & Bickmore, W. A. Transcription factories: gene expression in unions? *Nat. Rev. Genet.* **10,** 457–466 (2009).

32. Dowen, J. M. *et al.* Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159,** 374–387 (2014).

33. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129,** 823–837 (2007).

34. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 21931–21936 (2010).

35. Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441,** 349–353 (2006).

36. Harr, J. C. *et al.* Directed targeting of chromatin to the nuclear lamina is mediated by chromatin state and A-type lamins. *J. Cell Biol.* **208,** 33–52 (2015).

37. Henikoff, S., Steven, H. & Ali, S. Histone modification: cause or cog? *Trends Genet.* **27,** 389–396 (2011).

38. Becker, P. B. & Workman, J. L. Nucleosome Remodeling and Epigenetics. *Cold Spring Harb. Perspect. Biol.* **5,** a017905–a017905 (2013).

39. Hainer, S. J. *et al.* Suppression of pervasive noncoding transcription in embryonic stem cells by esBAF. *Genes Dev.* **29,** 362–378 (2015).

40. Bell, O., Tiwari, V. K., Thomä, N. H. & Schübeler, D. Determinants and dynamics of genome accessibility. *Nat. Rev. Genet.* **12,** 554–564 (2011).

41. Szerlong, H. *et al.* The HSA domain binds nuclear actin-related proteins to regulate chromatin-remodeling ATPases. *Nat. Struct. Mol. Biol.* **15,** 469–476 (2008).

42. Boyer, L. A., Latek, R. R. & Peterson, C. L. The SANT domain: a unique histone-tail-binding module? *Nat. Rev. Mol. Cell Biol.* **5,** 158–163 (2004).

43. Hota, S. K. *et al.* Nucleosome mobilization by ISW2 requires the concerted action of the ATPase and SLIDE domains. *Nat. Struct. Mol. Biol.* **20,** 222–229 (2013).

44. Flanagan, J. F. *et al.* Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature* **438,** 1181–1185 (2005).

45. Bakshi, R., Prakash, T., Dash, D. & Brahmachari, V. In silico characterization of the INO80 subfamily of SWI2/SNF2 chromatin remodeling proteins. *Biochem. Biophys. Res. Commun.* **320,** 197–204 (2004).

46. Ho, L. & Crabtree, G. R. Chromatin remodelling during development. *Nature* **463,** 474–484 (2010).

47. Yoshida, T. *et al.* The role of the chromatin remodeler Mi-2beta in hematopoietic stem cell self-renewal and multilineage differentiation. *Genes Dev.* **22,** 1174–1189 (2008).

48. Lessard, J. *et al.* An essential switch in subunit composition of a chromatin remodeling complex during neural development. *Neuron* **55,** 201–215 (2007).

49. Chi, T. H. *et al.* Sequential roles of Brg, the ATPase subunit of BAF chromatin remodeling complexes, in thymocyte development. *Immunity* **19,** 169–182 (2003).

50. Pedersen, T. A., Kowenz-Leutz, E., Leutz, A. & Nerlov, C. Cooperation between C/EBPalpha TBP/TFIIB and SWI/SNF recruiting domains is required for adipocyte differentiation. *Genes Dev.* **15,** 3208–3216 (2001).

51. Lickert, H. *et al.* Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. *Nature* **432,** 107–112 (2004).

52. Lazzaro, M. A. & Picketts, D. J. Cloning and characterization of the murine Imitation Switch (ISWI) genes: differential expression patterns suggest distinct

developmental roles for Snf2h and Snf2l. *J. Neurochem.* **77,** 1145–1156 (2001).

53. Pépin, D., Paradis, F., Perez-Iratxeta, C., Picketts, D. J. & Vanderhyden, B. C. The imitation switch ATPase Snf2l is required for superovulation and regulates Fgl2 in differentiating mouse granulosa cells. *Biol. Reprod.* **88,** 142 (2013).
54. Yip, D. J. *et al.* Snf2l regulates Foxg1-dependent progenitor cell expansion in the developing brain. *Dev. Cell* **22,** 871–878 (2012).
55. Erdel, F., Fabian, E. & Karsten, R. Chromatin remodelling in mammalian cells by ISWI-type complexes - where, when and why? *FEBS J.* **278,** 3608–3618 (2011).
56. Corona, D. F. V. *et al.* ISWI Is an ATP-Dependent Nucleosome Remodeling Factor. *Mol. Cell* **3,** 239–245 (1999).
57. Clapier, C. R., Nightingale, K. P. & Becker, P. B. A critical epitope for substrate recognition by the nucleosome remodeling ATPase ISWI. *Nucleic Acids Res.* **30,** 649–655 (2002).
58. Erdel, F., Schubert, T., Marth, C., Längst, G. & Rippe, K. Human ISWI chromatin-remodeling complexes sample nucleosomes via transient binding reactions and become immobilized at active sites. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 19873–19878 (2010).
59. Wysocka, J. *et al.* A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442,** 86–90 (2006).
60. Manelyte, L., Strohner, R., Gross, T. & Längst, G. Chromatin targeting signals, nucleosome positioning mechanism and non-coding RNA-mediated regulation of the chromatin remodeling complex NoRC. *PLoS Genet.* **10,** e1004157 (2014).
61. Alvarez-Saavedra, M. *et al.* Snf2h-mediated chromatin organization and histone H1

dynamics govern cerebellar morphogenesis and neural maturation. *Nat. Commun.* **5,** 4181 (2014).

62. Lazzaro, M. A. *et al.* The imitation switch protein SNF2L regulates steroidogenic acute regulatory protein expression during terminal differentiation of ovarian granulosa cells. *Mol. Endocrinol.* **20,** 2406–2417 (2006).
63. Stopka, T. & Skoultchi, A. I. The ISWI ATPase Snf2h is required for early mouse development. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 14097–14102 (2003).
64. Czechanski, A. *et al.* Derivation and characterization of mouse embryonic stem cells from permissive and nonpermissive strains. *Nat. Protoc.* **9,** 559–574 (2014).
65. Untergasser, A. *et al.* Primer3--new capabilities and interfaces. *Nucleic Acids Res.* **40,** e115–e115 (2012).
66. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).
67. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).
68. Shen, Y. *et al.* A map of the cis-regulatory sequences in the mouse genome. *Nature* **488,** 116–120 (2012).
69. Mouse ENCODE Consortium *et al.* An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol.* **13,** 418 (2012).
70. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
71. Bogu, G. K. *et al.* Chromatin and RNA Maps Reveal Regulatory Long Noncoding RNAs in Mouse. *Mol. Cell. Biol.* **36,** 809–819 (2015).

72. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9,** 215–216 (2012).

73. Meshorer, E. *et al.* Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev. Cell* **10,** 105–116 (2006).

74. Meshorer, E. & Misteli, T. Chromatin in pluripotent embryonic stem cells and differentiation. *Nat. Rev. Mol. Cell Biol.* **7,** 540–546 (2006).

75. Gaspar-Maia, A., Alajem, A., Meshorer, E. & Ramalho-Santos, M. Open chromatin in pluripotency and reprogramming. *Nat. Rev. Mol. Cell Biol.* **12,** 36–47 (2011).

76. Wu, J. Q. *et al.* Dynamic transcriptomes during neural differentiation of human embryonic stem cells revealed by short, long, and paired-end sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 5254–5259 (2010).

77. Xiao, S. *et al.* Comparative epigenomic annotation of regulatory DNA. *Cell* **149,** 1381–1392 (2012).

78. Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453,** 519–523 (2008).

79. Marks, H. *et al.* The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149,** 590–604 (2012).

80. Kumar, R. M. *et al.* Deconstructing transcriptional heterogeneity in pluripotent stem cells. *Nature* **516,** 56–61 (2014).

81. Cacchiarelli, D. *et al.* Integrative Analyses of Human Reprogramming Reveal Dynamic Nature of Induced Pluripotency. *Cell* **162,** 412–424 (2015).

82. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348,** 648–660 (2015).

83. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122,** 947–956 (2005).

84. Blackledge, N. P., Rose, N. R. & Klose, R. J. Targeting Polycomb systems to regulate gene expression: modifications to a complex story. *Nat. Rev. Mol. Cell Biol.* **16,** 643–649 (2015).

85. Bogu, G. K. *et al.* Chromatin and RNA Maps Reveal Regulatory Long Noncoding RNAs in Mouse. *Mol. Cell. Biol.* **36,** 809–819 (2015).

86. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125,** 315–326 (2006).

87. Peters, J.-M., J.-M., P., Tedeschi, A. & Schmitz, J. The cohesin complex and its roles in chromosome biology. *Genes Dev.* **22,** 3089–3114 (2008).

88. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9,** 215–216 (2012).

89. Bogu, G. K. *et al.* Chromatin and RNA Maps Reveal Regulatory Long Noncoding RNAs in Mouse. *Mol. Cell. Biol.* **36,** 809–819 (2015).

90. Bogu, G. K. *et al.* Chromatin and RNA Maps Reveal Regulatory Long Noncoding RNAs in Mouse. *Mol. Cell. Biol.* **36,** 809–819 (2015).

91. Hay, D. C., Sutherland, L., Clark, J. & Burdon, T. Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells. *Stem Cells* **22,** 225–235 (2004).

92. Masui, S. *et al.* Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat. Cell Biol.* **9,** 625–635 (2007).

93. Ho, L. *et al.* An embryonic stem cell chromatin remodeling complex, esBAF, is an

essential component of the core pluripotency transcriptional network. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 5187–5191 (2009).

94. de Dieuleveult, M. *et al.* Genome-wide nucleosome specificity and function of chromatin remodellers in ES cells. *Nature* **530,** 113–116 (2016).
95. Ho, L. *et al.* esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. *Nat. Cell Biol.* **13,** 903–913 (2011).
96. Jimeno-González, S., Ceballos-Chávez, M. & Reyes, J. C. A positioned +1 nucleosome enhances promoter-proximal pausing. *Nucleic Acids Res.* **43,** 3068–3078 (2015).
97. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155,** 934–947 (2013).
98. Zhou, B. *et al.* INO80 governs superenhancer-mediated oncogenic transcription and tumor growth in melanoma. *Genes Dev.* **30,** 1440–1453 (2016).
99. Kwon, S. Y., Grisan, V., Jang, B., Herbert, J. & Badenhorst, P. Genome-Wide Mapping Targets of the Metazoan Chromatin Remodeling Factor NURF Reveals Nucleosome Remodeling at Enhancers, Core Promoters and Gene Insulators. *PLoS Genet.* **12,** e1005969 (2016).
100. Dluhosova, M. *et al.* Epigenetic Control of SPI1 Gene by CTCF and ISWI ATPase SMARCA5. *PLoS One* **9,** e87448 (2014).
101. Trotter & Trotter. The BRG1 transcriptional coregulator. *Nucl. Recept. Signal.* **4,** (2007).
102. de la Serna, I. L. *et al.* MyoD targets chromatin remodeling complexes to the

myogenin locus prior to forming a stable DNA-bound complex. *Mol. Cell. Biol.* **25,** 3997–4009 (2005).

103. Hassan, A. H. *et al.* Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. *Cell* **111,** 369–379 (2002).

104. Sims, R. J. *et al.* Human but Not Yeast CHD1 Binds Directly and Selectively to Histone H3 Methylated at Lysine 4 via Its Tandem Chromodomains. *J. Biol. Chem.* **280,** 41789–41792 (2005).

105. Plath, K. Role of Histone H3 Lysine 27 Methylation in X Inactivation. *Science* **300,** 131–135 (2003).

106. Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341,** 1237973 (2013).

107. Werner, M. S. & Ruthenburg, A. J. Nuclear Fractionation Reveals Thousands of Chromatin-Tethered Noncoding RNAs Adjacent to Active Genes. *Cell Rep.* **12,** 1089–1098 (2015).

108. G Hendrickson, D., Kelley, D. R., Tenen, D., Bernstein, B. & Rinn, J. L. Widespread RNA binding by chromatin-associated proteins. *Genome Biol.* **17,** 28 (2016).

109. Nagano, T. *et al.* The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322,** 1717–1720 (2008).

110. Lai, F. *et al.* Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494,** 497–501 (2013).

111. Chalei, V. *et al.* The long non-coding RNA Dali is an epigenetic regulator of neural

differentiation. *Elife* **3,** (2014).

112. Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129,** 1311–1323 (2007).
113. Prensner, J. R. *et al.* The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nat. Genet.* **45,** 1392–1398 (2013).
114. Guttman, M. *et al.* lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477,** 295–300 (2011).
115. Ong, C.-T. & Corces, V. G. CTCF: an architectural protein bridging genome topology and function. *Nat. Rev. Genet.* **15,** 234–246 (2014).
116. Wiechens, N. *et al.* The Chromatin Remodelling Enzymes SNF2H and SNF2L Position Nucleosomes adjacent to CTCF and Other Transcription Factors. *PLoS Genet.* **12,** e1005940 (2016).