



Published in final edited form as:

Nature. 2016 February 4; 530(7588): 113–116. doi:10.1038/nature16505.

## Genome-wide nucleosome specificity and function of chromatin remodellers in ES cells

Maud de Dieuleveult<sup>1,\*</sup>, Kuangyu Yen<sup>2,3,\*,#</sup>, Isabelle Hmitou<sup>1,\*</sup>, Arnaud Depaux<sup>1,\*</sup>, Fayçal Boussouar<sup>1,4</sup>, Daria Bou Dargham<sup>1</sup>, Sylvie Jounier<sup>1</sup>, H el ene Humbertclaude<sup>1</sup>, Florence Ribierre<sup>5</sup>, C eline Baulard<sup>5</sup>, Nina P. Farrell<sup>3</sup>, Bongsoo Park<sup>3</sup>, C eline Keime<sup>6</sup>, Lucie Carri ere<sup>1</sup>, Soizick Berlivet<sup>1</sup>, Marta Gut<sup>7</sup>, Ivo Gut<sup>7</sup>, Michel Werner<sup>1</sup>, Jean-Fran ois Deleuze<sup>5</sup>, Robert Olaso<sup>5</sup>, Jean-Christophe Aude<sup>1</sup>, Sophie Chantalat<sup>5</sup>, B. Franklin Pugh<sup>3,#</sup>, and Matthieu G erard<sup>1,#</sup>

<sup>1</sup>Institute for Integrative Biology of the Cell (I2BC), IBITECS, CEA, CNRS, Universit  Paris-Sud, Universit  Paris-Saclay, 91198 Gif-sur-Yvette, France

<sup>2</sup>Key Laboratory of Zebrafish Modeling and Drug Screening for Human Diseases of Guangdong Higher Education Institutes, Department of Developmental Biology, School of Basic Medical Sciences, Southern Medical University, Guangzhou, China

<sup>3</sup>Center for Eukaryotic Gene Regulation, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

<sup>5</sup>Centre National de G enotypage, Institut de G enomique, CEA, Evry, France

<sup>6</sup>IGBMC (Institut de G en tique et de Biologie Mol culaire et Cellulaire), INSERM, U964, CNRS, UMR7104, Universit  de Strasbourg, 67404 Illkirch, France

<sup>7</sup>Centre Nacional D'An lisi Gen mica, Barcelona, Spain

### Summary

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: [http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

#Correspondence and requests for materials should be addressed to M.G. (; Email: [matthieu.gerard@cea.fr](mailto:matthieu.gerard@cea.fr)), K.Y. (; Email: [kuangyuyen@smu.edu.cn](mailto:kuangyuyen@smu.edu.cn)), or B.F.P. (; Email: [bf2@psu.edu](mailto:bf2@psu.edu))

<sup>4</sup>Present address: INSERM, U823, Universit  Joseph Fourier - Grenoble 1, Institut Albert Bonniot, Facult  de M decine, La Tronche, France

\*These authors contributed equally to this work

#### Author contribution

M.D.D., I.H., A.D., F.B., D.B.D., S.J. and H.H. built cell lines and performed ChIP-seq and shRNA transfections. I.H., C.B. and R.O. conducted transcriptome experiments and analysis. F.R. and S.C. performed FAIRE-seq, ATAC-seq and RT-qPCR experiments. M.Ge and S.C. carried out MNase-seq experiments. S.B. provided the pHyper shRNA vector. K.Y., M.Ge., L.C., C.K., B.P., and J.C.A. conducted bioinformatics analysis. N.P.F. performed ChIP-exo and DNA sequencing at PSU. I.G., M.Gu. and J.F.D supervised DNA sequencing at CNAG and CNG. M.Ge., K.Y., S.C., M.W., and B.F.P. supervised the project. M.Ge., K.Y. and B.F.P. co-wrote the manuscript.

Data have been deposited in Gene Expression Omnibus under accession numbers GSE64825 (ChIP-seq, ATAC-seq, MNase-seq and ChIP-exo data) and GSE64819 (microarray).

The authors declare no competing financial interests.

Supplementary Information

Supplementary Tables (Excel file)

This file contains Supplementary Tables 1 and 2. Table 1 describes sequencing experiments characteristics. Table 2 reports the primers used in RT-qPCR experiments

ATP-dependent chromatin remodellers allow access to DNA for transcription factors and the general transcription machinery, but whether mammalian chromatin remodellers 1–3 target specific nucleosomes to regulate transcription is unclear. Here, we present genome-wide remodeller-nucleosome interaction profiles for Chd1, Chd2, Chd4, Chd6, Chd8, Chd9, Brg1 and Ep400 in mouse embryonic stem (ES) cells. These remodellers bind one or both full nucleosomes that flank MNase-defined nucleosome-free promoter regions (NFRs), where they separate divergent transcription. Surprisingly, large CpG-rich NFRs that extend downstream of annotated transcriptional start sites (TSSs) are nevertheless chromatinized with non-nucleosomal or subnucleosomal histone variants (H3.3 and H2A.Z) and modifications (H3K4me3 and H3K27ac). RNA polymerase (pol) II therefore navigates hundreds of bp of altered chromatin in the sense direction before encountering an MNase-resistant nucleosome at the 3' end of the NFR. Transcriptome analysis upon remodeller depletion reveals reciprocal mechanisms of transcriptional regulation by remodellers. Whereas at active genes individual remodellers play either positive or negative roles via altering nucleosome stability, at polycomb-enriched bivalent genes the same remodellers act in an opposite manner. These findings indicate that remodellers target specific nucleosomes at the edge of NFRs, where they regulate ES cell transcriptional programs.

---

We applied a genome-wide remodeller-nucleosome interaction assay<sup>4</sup> (MNase digestion to define nucleosomes followed by remodeller ChIP-seq) to ES cells, focusing on the 5' ends of genes (Extended Data Fig. 1 and Supplementary Table 1). We first examined remodeller co-enrichment with other factors such as pol II, selected histone marks, and transcription factors, over broad (500-bp) windows centred on DNase-I hypersensitive sites (DHS) (i.e., promoters and enhancers; N = 138,582) (Fig. 1a). High Pearson correlation scores were observed among the remodellers Brg1, Ep400, Chd1, Chd4, Chd6 and Chd8, suggesting that these factors tend to occupy the same genomic regions in ES cells. When we focused on active promoter regions within DHSs, most remodellers were correlated with components of the general transcription machinery, including pol II S5ph and TBP (Fig. 1b and Extended Data Fig. 2).

We next examined remodeller distribution in more detail by focussing on annotated TSSs (Fig. 1c and Extended Data Fig. 3). Remarkably, some remodellers like Brg1, Chd4, Chd6 bound similar nucleosome positions at all active genes, regardless of their H3K4me3 enrichment (which is a mark of transcriptional activity), while others, such as Chd1, Chd2, Chd9 and Ep400, were tightly linked to H3K4me3/transcription levels. Chd8 had an intermediate pattern. Chd1 and Chd2, which are both related to *S. cerevisiae* (yeast) Chd1, showed strikingly different distributions. Whereas Chd1 is present near the 5' ends of genes, the Chd2-nucleosome enrichment pattern encompassed the entire transcription unit and shared high correlation with H3K36me3. (Fig. 1a, c and Extended Data Fig. 2). This is consistent with how yeast Chd1 works<sup>5,6</sup>, and thus mammalian Chd2 and yeast Chd1 may be functionally equivalent.

We next investigated more closely the relationship between individual remodeller-bound nucleosomes and all nucleosomes defined by MNase-resistant mononucleosome-sized DNA fragments<sup>7</sup>. Plots of individual genes were aligned by their NFR midpoint and sorted by

NFR width into narrow and wide groups (Fig. 2a). We validated the experimental approach and its improved resolution by comparison to an existing sonication-based (rather than MNase) ChIP-seq approach<sup>8</sup> (Extended Data Fig. 4). Importantly, this sonication-based method, which reports on both nucleosomal and non-nucleosomal interactions, demonstrated that Chd4 was not bound within NFRs in a non-nucleosomal manner.

At narrow NFRs, Ep400 and Chd4 crosslinked predominantly nucleosomes -1 and +1 that flank the NFR (Fig. 2a). Chd6, Chd8 and Brg1 interacted predominantly with +1 nucleosomes, and at lower levels with -1 and -2. Chd1 was also enriched at +1, and had a diffuse distribution on several additional nucleosomes on both sides of the TSS (Fig. 2a,c). Thus, at short NFRs, the first nucleosome (+1) encountered by pol II after release from the pause state is one that is highly enriched with remodellers. These remodellers might play a role in the passage of pol II through these nucleosome barriers.

At wide NFRs, Ep400 and Chd4 were preferentially bound to -1 nucleosomes (Fig. 2a), and relatively less to the first detectable full nucleosome downstream of the NFR. More strikingly, Chd6, Chd8 and Brg1 had shifted from their preferential binding to +1 at narrow NFRs toward a predominant enrichment at the -1 position of wide NFRs. These NFRs also define the boundaries of CpG islands, as reported previously<sup>9</sup>. Thus, mammalian remodellers interact with nucleosomes in a position-specific manner, with a distribution pattern adapting to the local chromatin architecture and DNA composition.

The pattern of remodeller-bound nucleosomes contrasted with TSSs, including divergent pol II GRO-seq transcripts<sup>10</sup>, which generally stayed towards the upstream side of NFRs (i.e., distal to annotated gene bodies) (Fig. 2b). DHS and FAIRE<sup>11</sup> patterns, which demarcate chromatin accessibility, matched the narrower regions of annotated TSSs, rather than reflecting the dimensions of MNase-defined NFRs. Thus, DNase I (or FAIRE) and MNase define regions of differing dimensions. At the enzyme concentrations used and DNA fragment sizes analysed, DNase I (and FAIRE) released histone-free regions (termed HFRs) generating a positive signal, whereas MNase destroyed HFRs and noncanonical chromatin thereby generating a lower signal. We find that promoter HFRs have a roughly fixed width (<115 bp). Where NFRs are narrow (Fig. 2a), HFRs and NFRs are essentially the same. At wide NFRs, HFRs are embedded in the upstream portion of NFRs that are variably wider, CpG-enriched, and contain remarkably noncanonical chromatin (being DNase I resistant but MNase sensitive).

We examined the distribution of histone variants and marks in NFRs, measured by standard ChIP-seq<sup>12-16</sup>. Narrow NFRs had H3.3, H2A.Z, H3K4me3 and H3K27ac enriched primarily at the bordering +1 nucleosome, with H3K4me3 extending to nucleosome +3. Some enrichment occurred upstream, largely commensurate with the level of divergent transcription (Fig. 2b). Remarkably, at wide NFRs, these variants and marks were largely restricted to the noncanonical chromatin region downstream of promoter HFRs, but still within NFRs.

A subset of less active genes is marked by a combination of H3K4me3 and H3K27me3, defining them as bivalent<sup>17</sup> (Fig. 1c). Strikingly, bivalency does not predominate on the

same nucleosomes (i.e., H3K4me3 and H3K27ac are enriched in NFRs, whereas H3K27me3 resides downstream, over full nucleosomes) (Fig. 2b). Thus, mammalian NFRs are largely chromatinized with non-nucleosomal (MNase-defined) transcription-associated histone modifications and variants that may be spatially adjacent to repressive chromatin.

At narrow NFRs, the Ep400-bound and Chd4-bound  $-1$  nucleosomes separated sense-directed pol II from upstream divergently transcribed pol II (Fig. 2c, top left panel). This  $-1$  nucleosome also represented the peak of DNase I hypersensitivity (bottom left panel). These remodellers therefore might be involved in structural reorganization, ejection or repositioning of  $-1$  to regulate sense and divergent transcription (examined below). At bivalent genes, Brg1, Chd4, Chd6, Chd8 (weakly), and Ep400 (weakly) bound specific nucleosomes at a level commensurate with their transcription levels, akin to what was observed at non-bivalent “H3K4me3-only” gene class (Extended Data Fig. 5). In contrast, Chd1 was enriched only at the H3K4me3 class.

To investigate how remodellers, bound at these distinct classes of genes, are involved in transcription regulation in ES cells, we depleted each remodeller using shRNA vectors (Extended Data Fig. 6) and profiled mRNA expression or used a published deletion dataset<sup>18</sup> for Brg1 (Fig. 3a). We observed that Chd4, Ep400 and Brg1, among the tested remodellers, were the most required for transcriptional expression, in both the H3K4me3 and bivalent classes. Ep400 and Chd4 were predominantly involved in transcriptional activation of H3K4me3 promoters, whereas Brg1 showed a preference for repression. At bivalent promoters, Chd4, Chd6, Chd8 and Ep400 were mostly involved in transcriptional repression, whereas Brg1 counteracted transcriptional repression, as previously described<sup>18</sup>. Loss-of-function of the other remodellers resulted in more limited changes in gene expression. These results were validated by RT-qPCR using two different shRNA vectors for each remodeller (Extended Data Fig. 7).

Since CpG islands help determine remodeller requirements<sup>19</sup>, we counted the percentage of genes most regulated by Chd4, Ep400 and Brg1 as a function of C+G content or NFR width, which is related to C+G (and CpG) content (Fig. 3b). Chd4 and Ep400 preferentially activated H3K4me3 genes having narrow NFRs (low CpG content), whereas at bivalent promoters the trend was doubly reversed: genes were preferentially repressed at wide NFRs (high CpG content). Brg1 preferentially activated genes with long NFRs, and repressed genes with narrow NFRs (low CpG content), at both H3K4me3 and bivalent classes. Thus, remodellers and chromatin architecture commonly have reciprocal relationships between H3K4me3 and bivalent genes.

To start understanding how specific remodellers might regulate genes, we used ATAC-seq<sup>20</sup> to examine regional chromatin accessibility upon remodeller depletion. Brg1 depletion resulted in decreased ATAC-seq signal at bivalent promoters (Extended Data Fig. 8), whereas Chd4 depletion produced the opposite effect. Thus, Brg1 maintains accessible chromatin and Chd4 restricts it. Despite Ep400 positively regulating H3K4me3 promoters and negatively regulating bivalent promoters, we detected no effect on regional accessibility by Ep400 using ATAC-seq, indicating that it acts by another mechanism.

To further examine how Ep400 might work, we turned to MNase-seq upon Ep400 depletion. Remarkably, its depletion resulted in increased MNase-resistance particularly at the  $-1$  nucleosome, where Ep400 was enriched at both positively and negatively regulated genes (Fig. 3c). This effect was most evident at H3K4me3 promoters, which are bound by high Ep400 levels, compared to bivalent promoters (Fig. 3c and Extended Data Fig. 3). Ep400 therefore may act to alter the structure of the  $-1$  nucleosome.

We also examined the consequences of remodeller depletion on pol II occupancy at promoters, using ChIP-exo. We found that depletion of either Ep400 or Chd4 resulted in a reduction of pol II levels at the H3K4me3 promoters they activate (Extended Data Fig. 9), showing that these remodellers contribute to pol II recruitment at subsets of active promoters.

Three contrasting stereotypes of remodeller control of gene expression in mouse ES cells arise from the data (Fig. 4), although not all genes fall into these stereotypes. First, active (H3K4me3-only) genes with narrow NFRs are flanked by nucleosomes bound and destabilized by positive-acting Chd4 and Ep400. The  $+1$  nucleosome is further engaged with negatively acting Brg1. Other remodellers bind there as well, but their function is less clear. Further downstream, Chd2-nucleosome interactions may be utilizing the H3K36me3 mark to organize nucleosomes analogously to Chd1 or Isw1b in budding yeast<sup>6</sup>.

The second stereotype has similar but not identical remodeller-nucleosome interactions as the first, such that Brg1 acts positively through  $-1$  instead of negatively through  $+1$ . This stereotype also has wide CpG-rich NFRs that are chromatinized with remodelled noncanonical chromatin or partial nucleosomes (e.g. hexasomes, tetrasomes, or half-nucleosomes), and includes short upstream HFRs where bidirectional transcription originates. The third stereotype is similar to the second but is enriched with bivalent genes having H3K4me3 within the NFR, and H3K27me3 and polycomb downstream within genic nucleosomal arrays. Thus bivalency is spatially separated on the same gene. Two trends emerge: 1) An activating remodeller in one class of genes is an inhibitor remodeller in the other class; 2) Within the same class, an activating remodeller can be counteracted by an inhibitor remodeller. Taken together, remodellers work together at specific nucleosome positions adjacent to promoter region NFRs to elicit proper gene control.

## METHODS

### Knock-in of a TAP-tag in the genes encoding the remodellers through homologous recombination in ES cells

The recombineering technique<sup>21</sup> was adapted to construct all targeting vectors for homologous recombination in ES cells. Retrieval vectors were obtained by combining 5' miniarm (NotI/SpeI), 3' miniarm (SpeI/BamHI) and the plasmid PL253 (NotI/BamHI). SW102 cells<sup>21</sup> containing a BAC encompassing the C terminal part of the gene encoding the remodeller, were electroporated with the SpeI-linearized retrieval vector. This allowed the subcloning of genomic fragments of approximately 10 kb comprising the last exon of the gene encoding each remodeller. The next step was the insertion of a TAP-tag into the subcloned DNA, immediately 3' to coding sequence. The TAP-tag was (FLAG)<sub>3</sub>-TEV-HA

for Chd1, Chd2, Chd4, Chd6, Chd8, Ep400, Brg1, and 6His-FLAG-HA for Chd9. We first inserted the TAP-tag and an *AscI* site into the PL452 vector, in order to clone 5' homology arms as *Sall/AscI* fragments into the PL452TAP-tag vector. 46C ES cells were electroporated with *NotI*-linearized targeting constructs and selected with G418. In all cases, G418-positive clones were screened by Southern blot. Details on the Southern genotyping strategy, as well as sequences of primers and plasmids used in this study are available upon request. Correctly targeted ES cell clones were karyotyped, and the expression of each tagged remodeler was controlled by western blot analysis, using antibodies against FLAG and HA epitopes (see Extended Data Fig. 6). We also verified by immunofluorescence, using monoclonal antibodies anti-FLAG (M2, Sigma F1804) and anti-HA (HA.11, Covance MMS-101P) epitopes, that each tagged remodeler was properly localized in the nucleus of ES cells.

### Verification of pluripotency in tagged ES cell line

ES cell lines expressing a tagged remodeler were all indistinguishable in culture from their mother cell line (46C). Pluripotency of tagged ES cell lines was verified by detecting alkaline phosphatase activity on ES cell colonies five days after plating, using the Millipore alkaline detection kit, following manufacturer's instructions. In addition, we verified by immunofluorescence using an antibody against Oct4/Pou5f1 (Abcam ab19857, lot 943333) that expression of this pluripotency-associated transcription factor was uniform in each tagged ES cell line.

### Cell lines and ES cell culture condition

Mouse 46C ES cells have been described previously<sup>22</sup>. 46C ES cells and their tagged derivatives were cultured at 37°C, 5% CO<sub>2</sub>, on mitomycin C-inactivated mouse embryonic fibroblasts, in DMEM (Sigma) with 15% foetal bovine serum (Invitrogen), L-Glutamine (Invitrogen), MEM non-essential amino acids (Invitrogen), pen/strep (Invitrogen),  $\beta$ -mercaptoethanol (Sigma), and a saturating amount of leukemia inhibitory factor (LIF), as described in reference<sup>23</sup>.

### Reference ES cell nucleosome map and NFR categories

Mouse ES nucleosomal tags were acquired from a published MNase-seq dataset<sup>7</sup> to make the reference map shown in Fig. 2. Reference nucleosomes were called using MACS 2.0 before assigning the first MNase-resistant nucleosome upstream and downstream of TSSs as -1 and +1, respectively. Since long NFRs, may actually contain MNase-sensitive nucleosome-like structures or histone-containing complexes, defining the first downstream MNase-resistant nucleosome as "+1" is problematic, and so we refer to it as the "first stable nucleosome". Regions between the associated -1 and +1 (or first stable) nucleosomes were defined as nucleosome free regions (NFR). We define histone-free regions (HFRs) as lacking histones as defined by ChIP-seq.

According to the distance between -1 and the first stable nucleosome dyad locations, we further defined narrow and wide NFR categories, which have the median width of 175 bp and 954 bp respectively.

### H3K4me3-only and bivalent gene lists

The list of 14,623 genes used in Fig. 1 and 2 was obtained by filtering all mm9 RefSeq genes<sup>24</sup>. We removed redundancies (that is, genes having the same start and end sites), unmappable genes, blacklisted genomic regions (those with artefact signal regardless which NGS techniques were used), and genes shorter than 2 kb. The purpose of this last filtering step was to unambiguously distinguish the promoter region from the end of the genes in heat-maps.

Lists of genes define as having H3K4me3 and bivalent promoters: We first defined, among the 14,623 RefSeq genes, those with a promoter that was positive for H3K4me3 (accession number: GSM590111). This was accomplished by operating with the seqMINER platform. Tag densities from this dataset were collected in a -500/+1,000 bp window around the TSS, and subjected to three successive rounds of k-means clustering, in order to remove all genes with a promoter that was clustered with low H3K4me3. We next conducted on this series of H3K4me3-positive promoters three successive rounds of k-means clustering, using several published datasets for H3K27me3. The genes with a promoter positive for H3K27me3 in four distinct H3K27me3 datasets (accession numbers: GSM590115, GSM590116, GSM307619 and GSM392046/GSM392047) were considered as bivalent. We eventually obtained a list of 6,481 genes with H3K4me3-only promoters, and a list of 3,411 bivalent genes.

### Tandem affinity purification of MNase-digested remodeller-nucleosome complexes

A detailed version of this protocol is available on the protocol exchange website: <http://dx.doi.org/10.1038/protex.2014.040>. In brief, about 400 million ES cells were fixed either with formaldehyde, or with a combination of disuccinimidyl glutarate (DSG) and formaldehyde (Supplementary Table 1), then permeabilized with IGEPAL, and incubated with 70 units of micrococcal nuclease (MNase) in order to fragment the genome into mononucleosomes (Extended Data Fig. 1). This nucleosome preparation was next incubated with agarose beads coupled with an antibody anti-HA or anti-FLAG. Anti-HA-agarose (ref. A2095) and anti-FLAG-agarose (ref. A2220) beads were purchased from Sigma. After a series of washes, tagged-remodeller-nucleosome complexes were eluted, either by TEV protease cleavage or by peptide competition (Supplementary Table 1). The eluted complexes were then subjected to a second immunopurification step, using beads coupled to the antibody specific of the second HA or FLAG epitope. After elution, DNA was extracted from the highly purified mononucleosome fraction, and processed for high-throughput sequencing (see below). As a negative control, chromatin from untagged ES cells was subjected to the same protocol to define background signal.

### High-throughput sequencing of MNase remodeller ChIP samples

After crosslink reversion, phenol-chloroform extraction and ethanol precipitation, the DNA from remodeller-nucleosome complexes was quantified using the picogreen method (Invitrogen) or by running 1/20 of the ChIP material on a High sensitivity DNA chip on a 2100 Bioanalyzer (Agilent, USA). 5 to 10 ng of ChIP DNA were used for library preparation according to the Illumina ChIP-seq protocol (ChIP-seq sample preparation kit). Following end-repair and adapter ligation, fragments were size-selected on an agarose gel in

order to purify nucleosome-sized genomic DNA fragments between 140 and 180 bp. Purified fragments were next amplified (18 cycles) and verified on a 2100 Bioanalyzer before clustering and single-read sequencing on an Illumina Genome Analyzer (GA) or GA II, according to manufacturer's instructions. Sequencing characteristics are shown in Supplementary Table 1.

### MNase Remodeller ChIP-seq data analysis

Chd1, Chd2, Chd4, Chd6, Chd8, Chd9, Ep400 and Brg1 MNase remodeller ChIP-seq short reads were mapped to mouse mm9 genome using Bowtie 0.12.7 with the following settings: -a -m1 --best --strata -v2 -p3. Datasets were next converted to BED format files, and data analysis was performed using the seqMINER platform<sup>25</sup> (Fig 1c). In order to examine the distribution of remodellers at individual genes, we used WigMaker3 (default settings) to convert BED files into wig files, which were uploaded onto the IGV genome browser (ED Fig. 2).

Nucleosome calls were made from MNase remodeller ChIP-seq tags using GeneTrack<sup>26</sup> with the following parameters: sigma = 20, exclusion = 146. We then globally shifted tags to the median value of half distances of all nucleosome calls. GRO-seq tags<sup>10</sup> sharing the same or opposite orientation with the TSS were assigned as "Sense" and "Divergent" tags, respectively. The orientation of each NFR was arranged so that sense transcription proceeds to the right. ES nucleosomal tags, globally shifted tags from MNase remodeller ChIP-seq (this current study), tags from DHS regions (Mouse ENCODE), GRO-seq oriented tags from transcriptionally engaged pol II and CpG islands (UCSC, mm9 build) were then aligned to the midpoint of each NFR. Promoter regions were then sorted by NFR length and visualized by Java TreeView (Fig 2a, b).

CpG island information was retrieved from UCSC (mm9 build) and assigned to the closest TSS by using bedtools. We noticed that promoters with wide NFRs were mostly CpG island (CpGI)-rich, while those with narrow NFRs were globally CpGI-poor, in agreement with a previous report showing that CpGIs induce nucleosome exclusion<sup>9</sup> (Fig 2b).

Tags from reference nucleosomes<sup>7</sup>, remodeller-interacting nucleosomes (this study) and transcriptionally engaged pol II (GRO-seq)<sup>10</sup> were aligned to nucleosome -1 and +1 (or the first stable nucleosome) dyad positions. The direction of each dyad was assigned according to the orientation of its associated TSS, whose orientation was arranged so that the transcription proceeds to the right. After normalization to the gene count in the two different NFR subclasses, tags were plotted from the NFR midpoint to 500 bp distal to the reference nucleosome. An x-axis gap in the NFR was introduced to normalize variations in NFR length inside each class.

### Pearson correlation analysis

We used DNaseI-Seq data from the Mouse ENCODE Consortium (GSM1004653) for the identification of DNase hypersensitive (DHS) regions in the mouse ES cell genome. DHS regions were defined using MACS 2.0<sup>27</sup> (default setting), which resulted in the identification of 139,454 DHS regions. Each of these DHS regions was represented as a 500-bp window (-250bp / +250bp) centred on the midpoint of the DHS peak. DHS regions overlapping with



the blacklisted (high background signal) genomic areas (mm9) were removed, resulting in a final list of 138,582 DHS regions. Tags from each tested ChIP-seq dataset were summed up for each DHS region before pair-wise Pearson correlation comparison. The  $R^2$  value from each pair-wise Pearson correlation was then visualized by heatmap (Fig. 1a).

Pearson correlation analysis at promoter-like DHS regions. Operating with the seqMINER platform, we retrieved, from the 138,582 DHS regions list, those positive for H3K4me3, TBP and Pol II S5ph. We obtained 16,300 promoter-like DHS regions befitting the criteria. Pairwise Pearson correlation was performed and plotted (Fig. 1b) as described for Fig 1a.

### RNA preparation from ES cells depleted of each remodeller by shRNA

We used the pHYPER shRNA vector for remodeller depletion in ES cells, as previously described<sup>28</sup>. shRNA design was performed using DESIR software (<http://biodev.extra.cea.fr/DSIR/DSIR.html>). Below are listed the shRNA selected for each remodeller. The sense strand sequence is given; the rest of the shRNA sequence is as described<sup>28</sup>.

Chd1 shRNA 1: GCAAAGACGGCGACTAGAAGA

Chd1 shRNA 2: GACAGTGCTTAATCAAGATCG

Chd4 shRNA 1: GGACGACGATTTAGATGTAGA

Chd4 shRNA 2: GCTGACGTCTTCAAGAATATG

Chd6 shRNA 1: GTACTATCGTGCTATCCTAGA

Chd6 shRNA 2: CAGTCAGAACCCACAATAACT

Chd8 shRNA 1: GCAGTTACTACTGACGTCTACA

Chd8 shRNA 2: GACTTTCTGTACCGCTCAAGA

Chd9 shRNA 1: TATACCAATTGAACAAGAGCC

Chd9 shRNA 2: AGTTAAAGTCTACAGATTAGT

Ep400 shRNA 1: GGTAAGAGTCCAGATTAAAG

Ep400 shRNA 2: GGTCCACACTCAACAACGAGC

Smarca4 shRNA 1: ACTTCTTGATAGAATTCTACC

Smarca4 shRNA 2: CCTTCGAACAGTGGTTCAATG

Each shRNA was transfected in its corresponding tagged ES cell line, in order to follow remodeller depletion by Western blotting using monoclonal antibodies anti-FLAG (M2, Sigma F1804), or anti-HA (H7, Sigma H3663) epitopes (Extended Data Fig. 6), in comparison with the signal obtained with a control antibody anti-GAPDH (Abcam ab9485). The pHYPER shRNA vectors were transfected in ES cell by electroporation, using an Amaxa nucleofector (Lonza). 24 h after transfection, puromycin (2  $\mu$ g/ml) selection was applied for an additional 48 h period, before cell collection and RNA preparation, except for Chd4, for which cells were collected after 30 h of selection. Total RNA was extracted using an RNeasy Kit (Qiagen). Total RNA yield was determined using a NanoDrop ND-100 (Labtech). Total RNA profiles were recorded using a Bioanalyzer 2100 (Agilent). For each

remodeller, RNA was prepared from three independent transfection experiments, and processed for transcriptome analysis.

### Analysis of gene expression in 46C ES cells by RNA-seq

46C ES cells were amplified on feeder cells except for the last passage, at which point cells were plated onto 60-mm dishes coated with gelatine, and grown to 70% confluence in D15 medium with LIF. Total RNA was extracted using an RNeasy Kit (Qiagen). The RNA quality was verified on a 2100 Bioanalyzer. Library preparation was performed using the Illumina mRNAseq sample preparation kit according to manufacturer's instructions. Briefly, the total RNA was depleted of ribosomal RNA using the Sera-mag Magnetic Oligo (dT) Beads (Illumina) and after mRNA fragmentation, reverse transcription and second strand cDNA synthesis the Illumina specific adaptors were ligated. The ligation product was then purified and enriched with 15 cycles of PCR to create the final library for single-read sequencing of 75 bp carried out on an Illumina GAIIx.

In order to keep only sequences of good quality we retained the first 40 bp of each read and discarded all sequences with more than 10% of bases having a quality score below 20, using FASTX-Toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)). Mapping of these sequences onto the mm9 assembly of mouse genome and RPKM computation were then performed using ERANGE v3.1.0<sup>29</sup> and bowtie v0.12.0<sup>30</sup>. Briefly, a splice file was created with UCSC known genes and maxBorder=36. We created an expanded genome containing genomic and splice-spanning sequences using bowtie-build and bowtie was used to map the reads onto this expanded genome. Then the ERANGE runStandardAnalysis.sh script was used to compute RPKM values following steps previously described<sup>29</sup>, using a consolidation radius of 20 kb.

### Quantitative RT-PCR analysis of gene expression

Random-primed reverse transcription was performed at 52°C in 20 µl using Maxima First strand cDNA synthesis kit (Thermo Scientific) with 1 µg of total RNA isolated from ES cells (Qiagen), quantified with NanoDrop instrument (Thermo Scientific). Reverse transcription products were diluted 40-fold before use. Composition of quantitative PCR assay included 2.5 µl of the diluted RT reaction, 0.2 to 0.5 mM forward and reverse primers, and 1X Maxima SYBR Green qPCR Master Mix (Thermo Scientific). Reactions were performed in a 10µl total volume. Amplification was performed as follows: 2 min at 95°C, 40 cycles at 95°C for 15 sec and 60°C for 60 sec in the ABI/Prism 7900HT real-time PCR machine (Applied Biosystems). The real-time fluorescent data from quantitative PCR were analyzed with the Sequence Detection System 2.3 (Applied Biosystems). Each quantitative real-time PCR was performed using the set of primer pairs listed in Supplementary Table 2, validated for their specificity and efficiency of amplification. All reactions were performed in triplicates, using RNA prepared from three independent cell transfection experiments. Control reactions without enzyme were verified to be negative. Relative expression was calculated after normalization with three reference genes (*Actb*, *Nmt1* and *Ddb1*), validated for this study.

## Transcriptome analysis in remodeller-depleted ES cells

cRNA was synthesized, amplified, and purified using the Illumina TotalPrep RNA Amplification Kit (Life Technologies) following Manufacturer's instructions. Briefly, 200 ng of RNA were used to prepare double-stranded cDNA using a T7 oligo (dT) primer. Second strand synthesis was followed by in vitro transcription in the presence of biotinylated nucleotides. cRNA samples were hybridized to the Illumina BeadChips Mouse WG-6v2.0 arrays. These BeadChips contain 45,281 unique 50-mer oligonucleotides in total, with hybridization to each probe assessed at 30 different beads on average. 26,822 probes (59%) are targeted at RefSeq transcripts and the remaining 18,459 (41 %) are for other transcripts. BeadChips were scanned on the Illumina iScan scanner using Illumina BeadScan image data acquisition software (version 2.3). Data were then normalized using the 'normalize quantiles' function in the GenomeStudio Software (version 1.9.0). Following analyses were done using Genespring software (version 13.0-GX)

For Brg1, we used a previously published transcriptome dataset, in which loss of Brg1 function was obtained by genetic ablation<sup>18</sup>. All array analyses were undertaken using the Limma package from the R/Bioconductor software (R-Development-Core-Team, 2007). Microarray spot intensities were normalized using the RMA method as implemented in the R affy package. Normalized measures served to compute the log2-ratios for each gene between the wild-type strain and the Brg1 KO mutant. Then, to identify genes with a log2-ratios significantly different between the mutant and wild-type strain, p-values were calculated for each gene using a moderated t-test. The moderated t-test applied here was based on an empirical Bayes analysis and was equivalent to shrinkage (or expansion) of the estimated sample variances towards a pooled estimate, resulting in a more stable inference. Finally, adjusted p-values were calculated using the FDR-controlling procedure of Benjamini & Hochberg.

### Analysis of gene deregulation

We identified deregulated genes using the thresholds of 0.05 for the p-value and 1.5 for the fold-change (FC 1.5). This FC 1.5 threshold was chosen based on a previous study on Brg1<sup>18</sup> and also because it was compatible with the analysis of the remodellers more modestly involved in transcriptional control in ES cells such as Chd1, Chd6 and Chd8. Note that seemingly modest fold changes might arise from many sources including a response lag, residual remodelling activity, and relatively high experimental background. Using a FC 2 threshold, we could however confirm that Ep400, Chd4 and Brg1 are important transcriptional regulators in ES cells, with 535, 293 and 570 genes deregulated, respectively. This level of deregulation is indicative of a context-specific function of remodellers in transcriptional activation or repression, which is distinct from the function of general transcription factors, whose depletion is expected to affect most genes.

Statistical analysis of the differences in transcriptional activation and repression by remodellers was performed using a 2-sample test for equality of proportions with continuity correction.

For the generation of GC-content-based lists of promoters, we used the list of promoters defined in Fig. 3 of reference 15, that we crossed with the 14,623 promoter list, to obtain a list of 6,317 promoters rank ordered according to GC content.

In Figure 3b, we compared the percentages of genes either down- or up-regulated by loss of function of each remodeller in the following two groups: (1) NFR length classes: Genes from the narrow and wide NFR classes shown in Fig. 2a were each further divided into two subclasses, which resulted in the following four categories: narrow NFR subclass 1 (NFR < 15 bp), narrow NFR subclass 2 (15–115 bp NFR), wide NFR subclass 1 (116–504 bp) and wide NFR subclass 2 (505–1500 bp). Genes in these groups were further subdivided into H3K4me3 and bivalent subgroups. (2) GC content classes: Genes were divided into four quartiles based on GC content at promoters and further subdivided into H3K4me3 and bivalent subclasses. The number of genes analysed in Fig. 3b is indicated in brackets for the following subgroups. H3K4me3 genes: narrow NFR subclass 1 (739), subclass 2 (1829), wide NFR subclass 1 (2613), subclass 2 (1253), GC content quartile 1 (low GC content) (450), quartile 2 (719), quartile 3 (644), quartile 4 (high GC content) (430). Bivalent genes: narrow NFR subclass 1 (271), subclass 2 (866), wide NFR subclass 1 (2266), subclass 2 (1184), GC content quartile 1 (220), quartile 2 (485), quartile 3 (750), quartile 4 (1149).

### FAIRE-seq

FAIRE was performed as described<sup>31</sup> with modifications. 46C ES cells were amplified as described above for RNA preparation. Formaldehyde was added directly to the growth media (final concentration 1%), and cells were fixed for 5 min at room temperature. After quenching with glycine (125 mM) and several washes, cells were collected, resuspended in 500 µl of cold lysis buffer (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA) and disrupted using glass beads for five 1-min sessions with 2 min incubations on ice between disruption sessions. Samples were then sonicated for 16 sessions of 1 min (30 sec on/30 sec off) using a bioruptor (Diagenode) at max intensity, at 4°C. After centrifugation, the supernatant was extracted twice with phenol-chloroform. The aqueous fractions were collected and pooled, and a final phenol-chloroform extraction was performed before DNA precipitation. FAIRE experiments were realized in triplicate, using independent ES cell cultures. Prior to sequencing, FAIRE DNA was analysed and quantified by running 1/25 of the FAIRE material on a High sensitivity DNA chip on a 2100 Bioanalyzer (Agilent, USA). 20 ng of FAIRE DNA was used for library preparation according to manufacturer's instructions using the ChIP-seq sample preparation kit (Illumina). Single-read sequencing (36bp) was performed on a Genome Analyzer II (Illumina).

### ATAC-seq

ES cells were grown and transfected with shRNA vectors as described for RNA analysis. Two independent transfection experiments were performed for each shRNA vector. ATAC-seq libraries were constructed by adapting a published protocol<sup>20</sup>. Briefly, 50,000 cells were collected, washed with cold PBS and resuspended in 50 µl of ES buffer (10µmM Tris pH 7.4, 10µmM NaCl, 3µmM MgCl<sub>2</sub>). Permeabilized cells were resuspended in 50 µl Transposase reaction (1X Tagmentation buffer, 1.0–1.5 µl Tn5 transposase enzyme (Illumina)) and incubated for 30min at 37°C. Subsequent steps of the protocol were

performed as previously described<sup>20</sup>. Libraries were purified using a Qiagen MinElute kit and Ampure XP magnetic beads (1:1.6 ratio) to remove remaining adapters. Libraries were controlled using a 2100 Bioanalyzer, and an aliquot of each library was sequenced at low depth onto a MiSeq platform to control duplicate level and estimate DNA concentration. Each library was then paired-end sequenced ( $2 \times 100$  bp) on a HiSeq instrument (Illumina).

### Analysis of ATAC-seq data

As ATACseq libraries are composed in large part of short genomic DNA fragments, reads were cropped to 50pb using trimmomatic-0.32 to optimize paired-end alignment. Reads were aligned to the mouse genome (mm9) using Bowtie with the parameters -m1 --best --strata -X2000, with 2-mismatches permitted in the seed (default value). The -X2000 option allows the fragments  $< 2$ kb to align and -m1 parameter keeps only unique aligning reads. Duplicated reads were removed with picard-tools-1.85. To perform differential analysis, libraries were adjusted to 33 million aligned reads using samtools-1.2 and by making a random permutation of initial input libraries (shuf linux command line). Adjusted BAM datasets were next converted to BED. We used the seqMINER platform with the lists of 6481 H3K4me3-only and 3411 bivalent genes described above, to collect tag densities from ATAC-seq datasets, in a window of  $-2$ kb/ $+2$ kb around the TSS. Output tag density files were analysed using R software to establish average ATAC-seq signal profiles shown in Extended Data Fig. 8.

### MNase-seq

ES cells were grown and transfected with shRNA vectors as described above. Two independent transfection experiments were performed for each shRNA vector. For each experiment, 1 million cells were fixed 10 min in ES cell culture medium containing 1% formaldehyde, quenched with glycine (125mM), washed with PBS buffer, collected in 175  $\mu$ l of solution I (15 mM Tris-HCl PH7.5, 0.3 M sucrose, 60 mM KCl, 15 mM NaCl, 5mM MgCl<sub>2</sub>, 0.1 mM EGTA), and stored on ice. Cells were permeabilized by adding 175  $\mu$ l of solution II (solution I with 0.8% Igepal CA-630 (Sigma)) and incubating 15 min on ice. We next added 700  $\mu$ l of MNase digestion buffer (50 mM Tris-HCl PH7.5, 0.3 M sucrose, 15 mM KCl, 60 mM NaCl, 4 mM MgCl<sub>2</sub>, 2 mM CaCl<sub>2</sub>), 4 units of MNase, and incubated 10 min at 37°C. MNase digestion was stopped by adding 10 mM EDTA (final concentration), and storing on ice. Cells were then disrupted by 15 passages through a 25G needle, followed by a 10 min centrifugation at 18,000 g. The supernatant was collected and incubated 1h at 65°C with 15  $\mu$ g of RNase A. We next added 10  $\mu$ g of proteinase K, adjusted each sample to 0.1% SDS (final concentration) and incubated 2 h at 55°C. NaCl concentration was then adjusted to 200 mM and the samples were incubated overnight at 65°C for crosslink reversal. DNA was purified from each sample by phenol-chloroform extraction followed by ethanol precipitation. 20 ng of purified DNA was used for library preparation according to manufacturer's instructions, using Ultralow ovation library system, Nugen. Following end-repair and adapter ligation, fragments were size-selected onto an agarose gel in order to purify genomic DNA fragments between  $\sim 60$  and 220 bp. Libraries were verified using a 2100 Bioanalyzer before clustering and paired-read sequencing. Sequencing of each sample was performed in a single lane of a HiSeq instrument (Illumina).

### Analysis of MNase-seq data

The midpoint of each pair-end sequencing read was used to represent dyad location of each nucleosomal tag. We assumed that remodeller depletion has no bulk effect on nucleosome occupancy, hence the total reads of control and remodeller-depleted cells were adjusted to be the same. The adjusted tags were aligned to  $-1$  nucleosome dyads (determined by the first MNase-defined peak upstream of annotated RefSeq TSS), or the first stable (MNase-defined) nucleosome dyad position downstream of the TSS for different NFR categories. These tags were further normalized to the amount of genes involved in each NFR class. The normalized tags were then binned (5bp) and smoothed (10 bin moving average) before plotting (Fig 3c). Distances (bp) are indicated relative to these reference points. An x-axis gap in the NFR was introduced to normalize variations in NFR length inside each class.

### Pol II ChIP-exo

ES cells were grown and transfected with shRNA vectors as described above. Two independent transfection experiments were performed for each shRNA vector. Following a 10 min fixation with 1% formaldehyde in ES cell culture medium, chromatin was prepared from 5 to 10 million cells and sonicated as described<sup>32</sup>. ChIP-exo experiments were carried out essentially as described<sup>33</sup>. This included an immunoprecipitation step using antibodies against Pol II (sc-899, Santa Cruz Biotechnology) attached to magnetic beads, followed by DNA polishing, A-tailing, Illumina adaptor ligation (ExA2), and lambda and recJ exonuclease digestion on the beads. After elution, a primer was annealed to EXA2 and extended with phi29 DNA polymerase, then A-tailed. A second Illumina adaptor was then ligated, and the products PCR-amplified and gel-purified. Sequencing was performed using NextSeq500. Uniquely aligned sequence tags were mapped to the mouse genome (mm9) using BWA-MEM (version 0.7.9a-r786)<sup>34</sup>. The uniquely aligned sequence tags were used for the downstream analysis.

### Analysis of Pol II ChIP-exo data

The 5' end of mapped tags, representing exonuclease stop sites, were consolidated into peak calls ( $\sigma = 5$ , exclusion = 20) using GeneTrack<sup>26</sup>, and peak pairs were matched when found on opposite strands and 0–100 bp apart in the 3' direction. Tags were globally shifted to the median value of half distance between all peak pairs.

These global shifted tags were then aligned relative to the annotated RefSeq TSSs for H3K4me3-only and bivalent promoters separately before further carved out remodeller-affected genes. We assumed that having remodeller deletion bore no bulk change on PolII occupancy, and hence total tags among wild type and all remodeller mutants were normalized to be the same. In order to make direct comparison between different gene groups, we further normalized tags to the amount of genes within the group. These normalized tags were then smoothed (5bp binned before 10 bin moving average) before plotting (Extended Data Fig 9a).

To examine PolII occupancy change in remodeller mutants among different promoter groups, we first calculated total Pol II occupancy by summing up tags from transcript start to end sites (annotated RefSeq TSS and TES, respectively<sup>24</sup>) for the tested genes. Change in

Pol II occupancy was calculated by dividing the total Pol II occupancy of mutant by that of wild type before log<sub>2</sub> transformation and bargraph plotting (Extended Data Fig. 9b).

### Average binding profiles (Extended Data Fig. 3)

Genes were rank ordered according to rpkm and divided in four quartiles (highest:Q4, second:Q3, third:Q2 and lowest:Q1). Operating with the k-means clustering function of seqMINER, genes in each quartile were further subdivided in H3K4me3-only and bivalent genes, as described above.

Using these lists of genes, tag densities from remodeller ChIP-seq datasets were collected in a window of -2kb/+2kb around the TSS, except for Chd2, for which densities were collected from the TSS until +4kb. Output tag density files were first analysed using R software to establish average binding profiles. Statistical comparisons were performed between remodeller distributions at H3K4me3 promoters, to assess a significant increasing trend among distributions. Differences between successive pairs of quartiles (Q4 - Q3, Q3 - Q2, Q2 - Q1) were compared against a null distribution using a one side t-test.

The respective p values are reported for each remodeller: Chd1, Q4 - Q3 p = 1.371138e-27 ; Q3 - Q2 p = 1.728126e-16 ; Q2 - Q1 p = 7.985217e-23.

Chd2, Q4 - Q3 p = 7.543473e-33 ; Q3 - Q2 p = 1.115223e-25 ; Q2 - Q1 p = 3.283427e-38.

Chd4, Q4 - Q3 p = 0.2094255 ; Q3 - Q2 p = 0.1081455 ; Q2 - Q1 p = 0.07202865.

Chd6, Q4 - Q3 p = 0.4168748 ; Q3 - Q2 p = 0.1534144 ; Q2 - Q1 p = 0.01138035.

Chd8, Q4 - Q3 p = 4.031959e-15 ; Q3 - Q2 p = 1.231527e-06 ; Q2 - Q1 p = 1.34455e-09.

Chd9, Q4 - Q3 p = 9.484578e-44 ; Q3 - Q2 p = 1.059783e-14 ; Q2 - Q1 p = 4.646352e-28.

Ep400, Q4 - Q3 p = 3.046796e-20 ; Q3 - Q2 p = 1.215304e-14 ; Q2 - Q1 p = 6.462667e-11.

Brg1, Q4 - Q3 p = 3.512021e-24 ; Q3 - Q2 p = 2.515217e-07 ; Q2 - Q1 p = 0.977422.

We concluded from this analysis that Chd1, Chd2, Chd9 and Ep400 binding at promoters is tightly linked to gene expression level. In contrast, Brg1, Chd4 and Chd6 deposition showed little correlation with gene expression level (statistical test failed for at least one comparison for these remodelers). Whilst statistical analysis of Chd8 distributions concluded to significant differences between quartiles, inspection of distributions in Extended Data Fig. 3 showed that Chd8 binding profile was intermediate between these two categories.

### Accession numbers and references of the publicly available Datasets used in Fig. 1 and 2 and Extended Data Fig. 1, 4 and 7

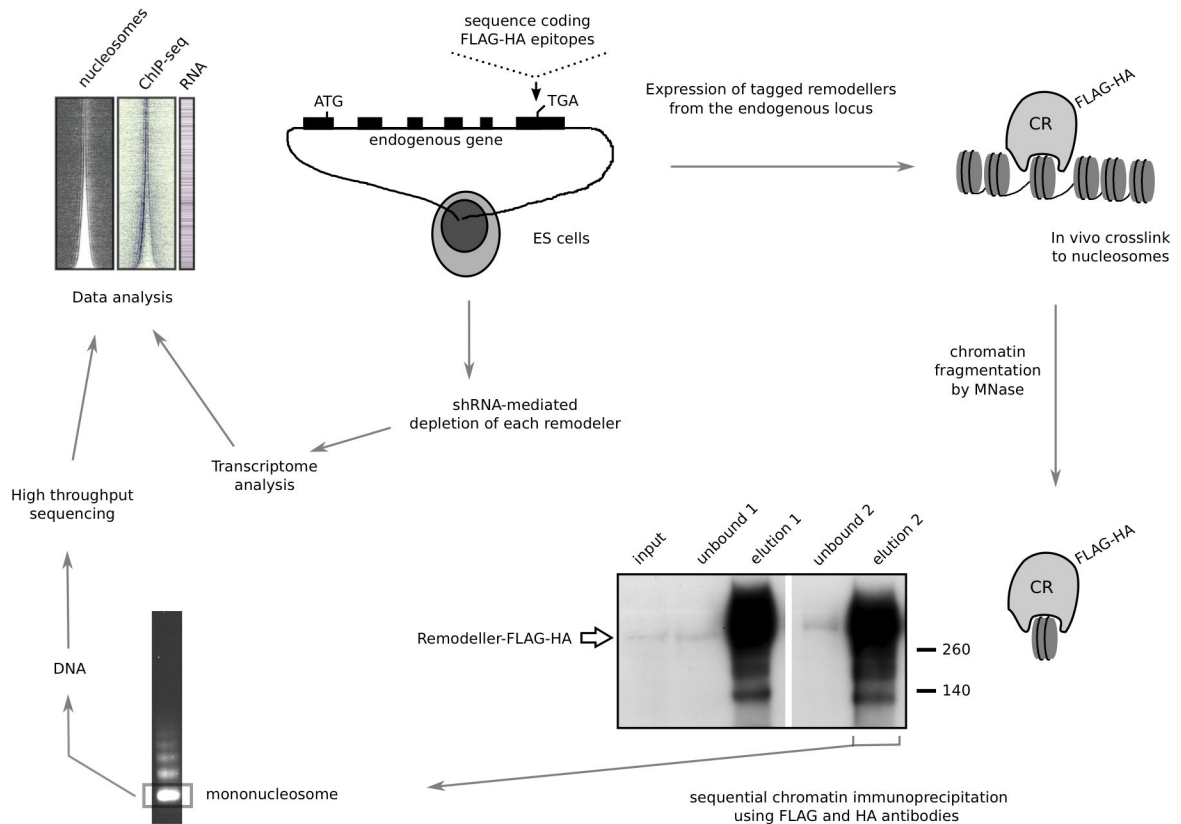
Brg1<sup>35</sup> : GSM359413

DNase-seq : GSM1014154

Ezh2<sup>15</sup>: GSM590132  
GRO-seq<sup>10</sup>: GSM665994  
H2A.Z<sup>13,14</sup>: GSM958501, DRP001103  
H3.3<sup>12</sup>: GSM1386359  
H3K27ac<sup>16</sup>: GSM594578  
H3K27me3<sup>15</sup>: GSM590115  
H3K36me3<sup>15</sup>: GSM590119  
H3K4me3<sup>15</sup>: GSM590111  
Med1<sup>36</sup>: GSM560347  
Mi2b (Chd4)<sup>8</sup>: GSM687284  
MNase-seq<sup>7</sup>: GSM1004653  
Oct4/Pou5f1, Sox2, Nanog : GSM1082340  
Pol II S5ph<sup>37</sup>: GSM515662  
TBP : GSM958503



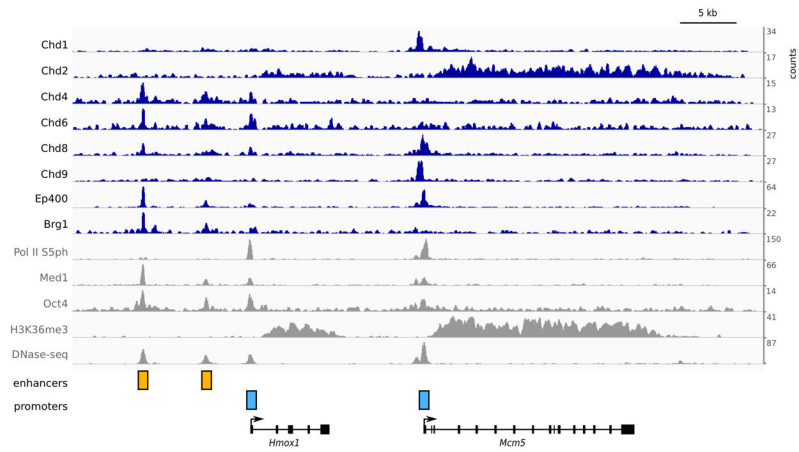
## Extended Data



### Extended Data Figure 1. Experimental strategy for genome-wide remodeller-nucleosome interactions and transcriptome analysis in ES cells

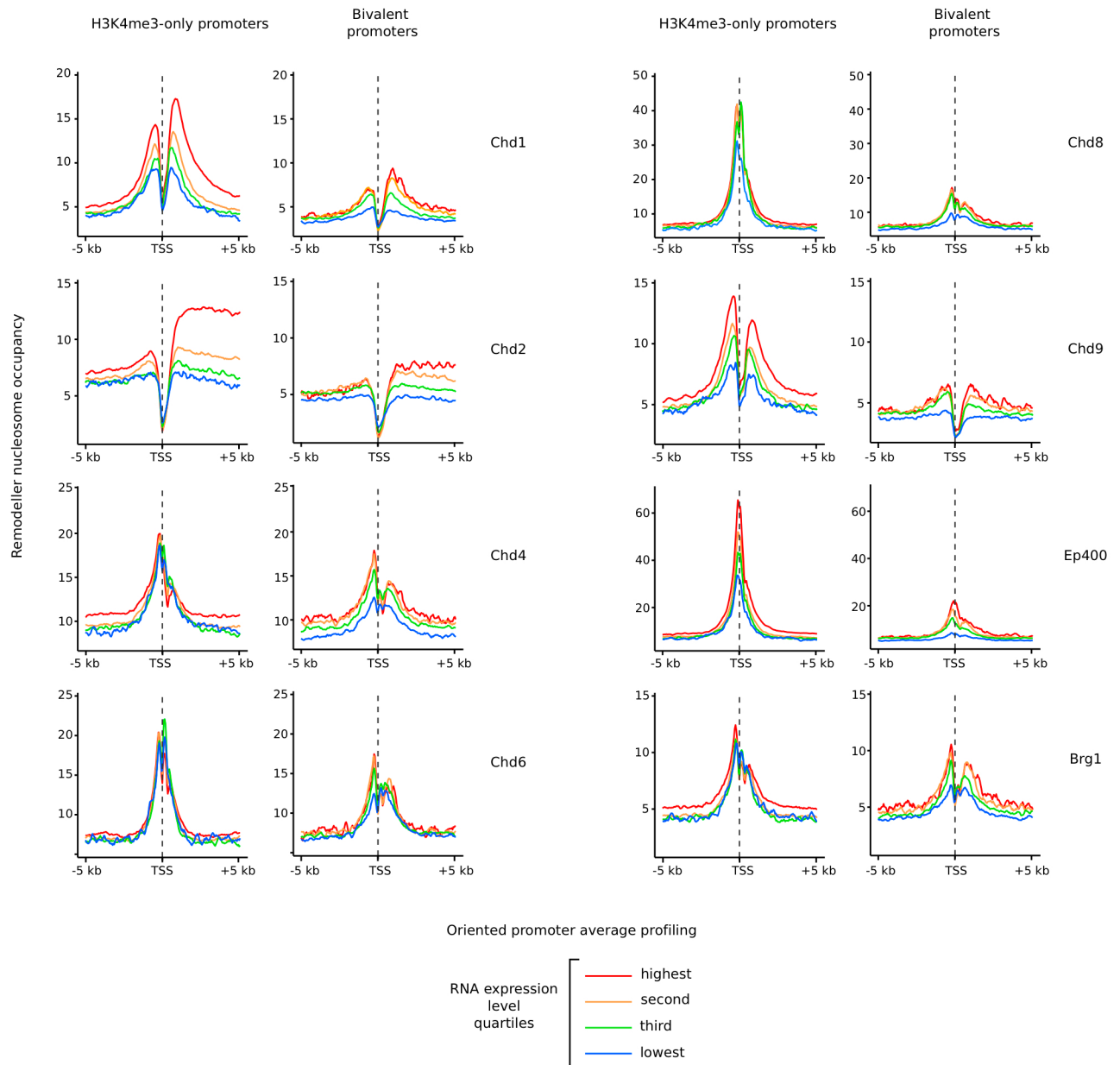
Using homologous recombination in ES cells, a sequence encoding a combination of FLAG and hemagglutinin (HA) epitopes was introduced at the 3' end of the coding sequence of the genes encoding the catalytic subunit of each remodeller. After in vivo crosslinking, chromatin was prepared and fragmented to mononucleosomes by MNase. Remodeller-bound mononucleosomes were isolated using a double immunoaffinity procedure.

Immunoprecipitation efficiency was assessed by Western blotting. Deep sequencing of the DNA from purified nucleosomes allowed the mapping of remodeller-bound nucleosomes across the mouse genome. The same tagged ES cell lines were used for shRNA-mediated depletion of remodellers and transcriptome analysis.



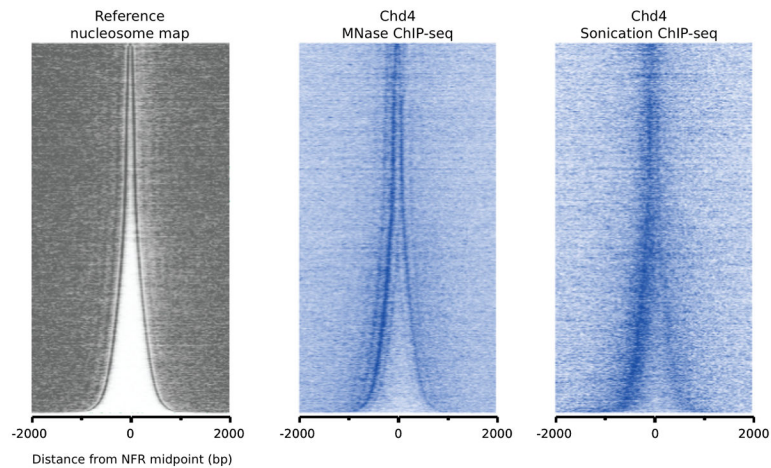
**Extended Data Figure 2. Remodeller binding profile at a representative locus**

Counts indicate reads per 10 million. Promoters and enhancers are highlighted by blue and orange squares, respectively



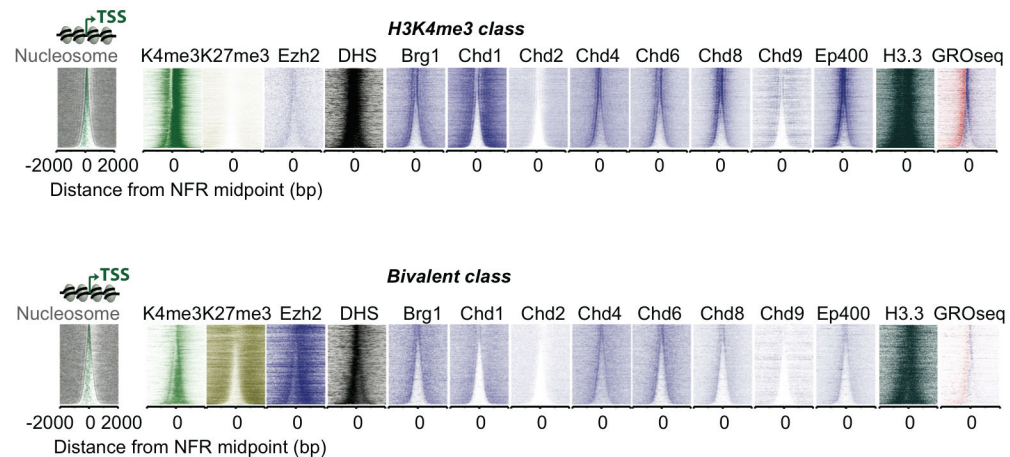
### Extended Data Figure 3. Relation between remodeller enrichment at promoters and RNA expression level

Average binding profile of remodellers at promoters, divided in four quartiles based on RNA expression level of the corresponding genes. All promoters are transcribed from left to right. Promoter binding intensity of Chd1, Chd2, Chd9 and Ep400 at H3K4me3 promoters was correlated with RNA expression (see Methods). Consequently, binding of these remodellers to bivalent promoters, which are transcribed at lower levels, showed a significant reduction compared to H3K4me3 promoters. In contrast, Chd4, Chd6, and Brg1 enrichment at promoters showed little correlation with the transcription level of the corresponding genes, and was only slightly lower at bivalent, compared to H3K4me3 promoters.



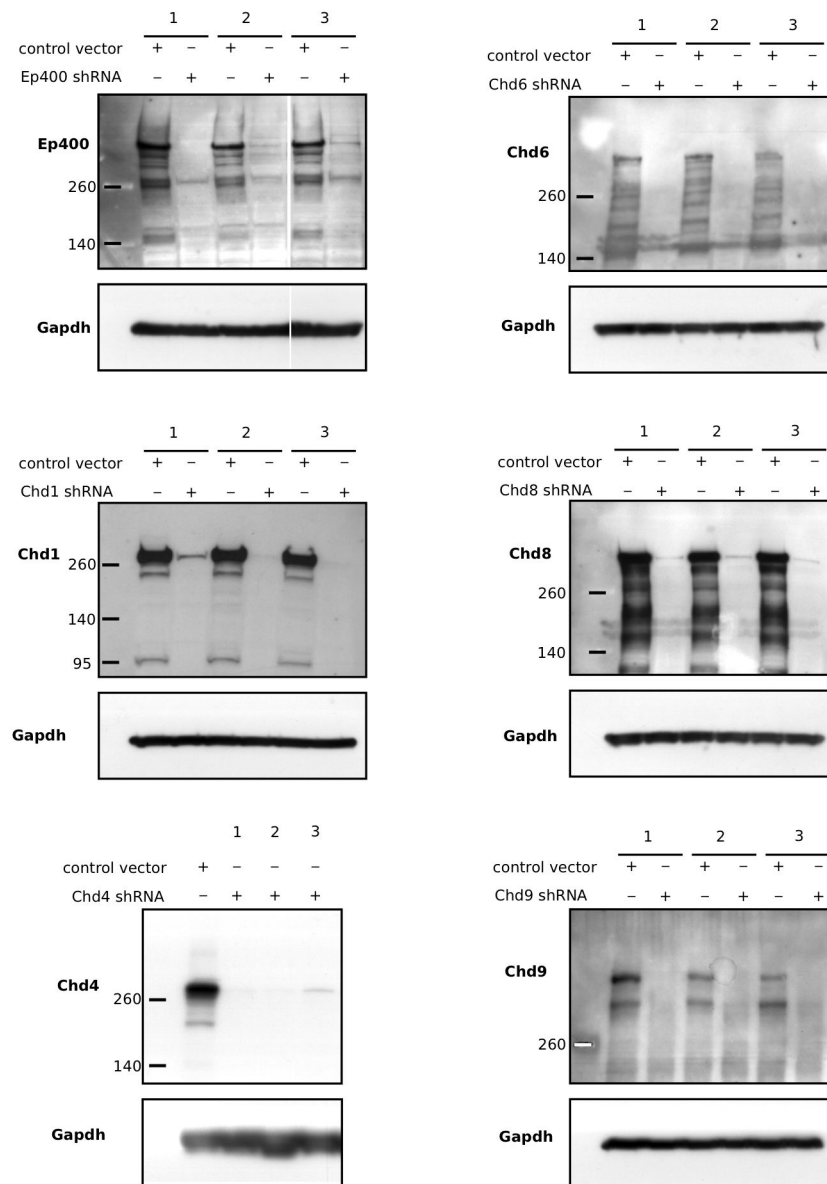
**Extended Data Figure 4. Comparison of MNase ChIP-seq and sonication ChIP-seq for Chd4**

The left panel shows the reference nucleosome map of 14,623 RefSeq genes, rank-ordered from smallest to largest NFR length, as in Fig. 2. The two on the right compare the distribution patterns obtained for Chd4 either by MNase ChIP-seq, with chromatin prepared from Chd4-tagged ES cells, or by ChIP-seq with sonicated chromatin (dataset accession number: GSM687284).



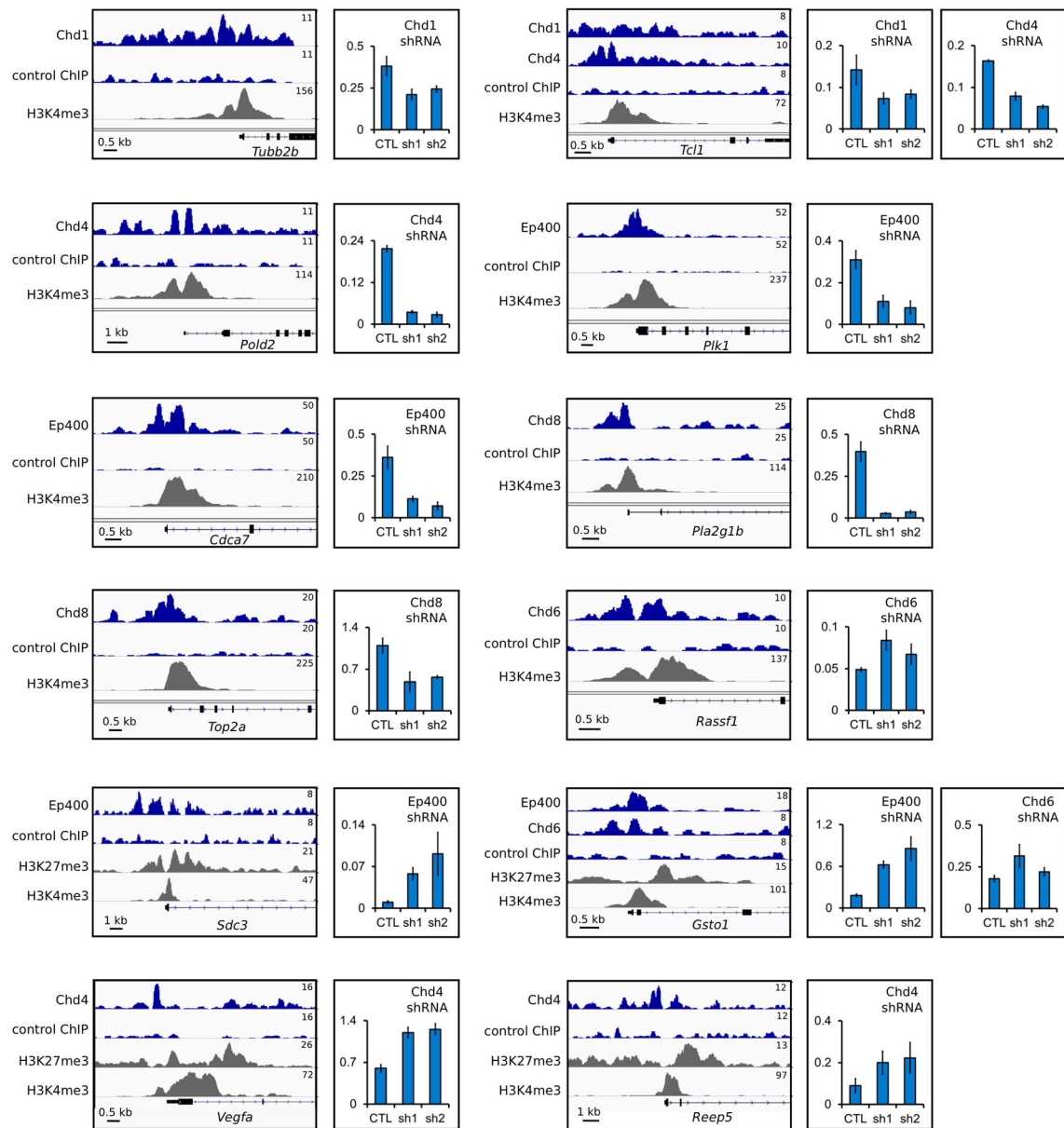
**Extended Data Figure 5. Nucleosome targeting by remodellers at H3K4me3-only and bivalent promoters**

Remodeller-bound nucleosomal tags were aligned to the promoters of 6,481 active (H3K4me3 promoters) or 3,411 bivalent genes, rank-ordered from narrow to wide NFR. Corresponding reference nucleosomes, remodeller occupancy and the other indicated features are shown as in Fig. 2.

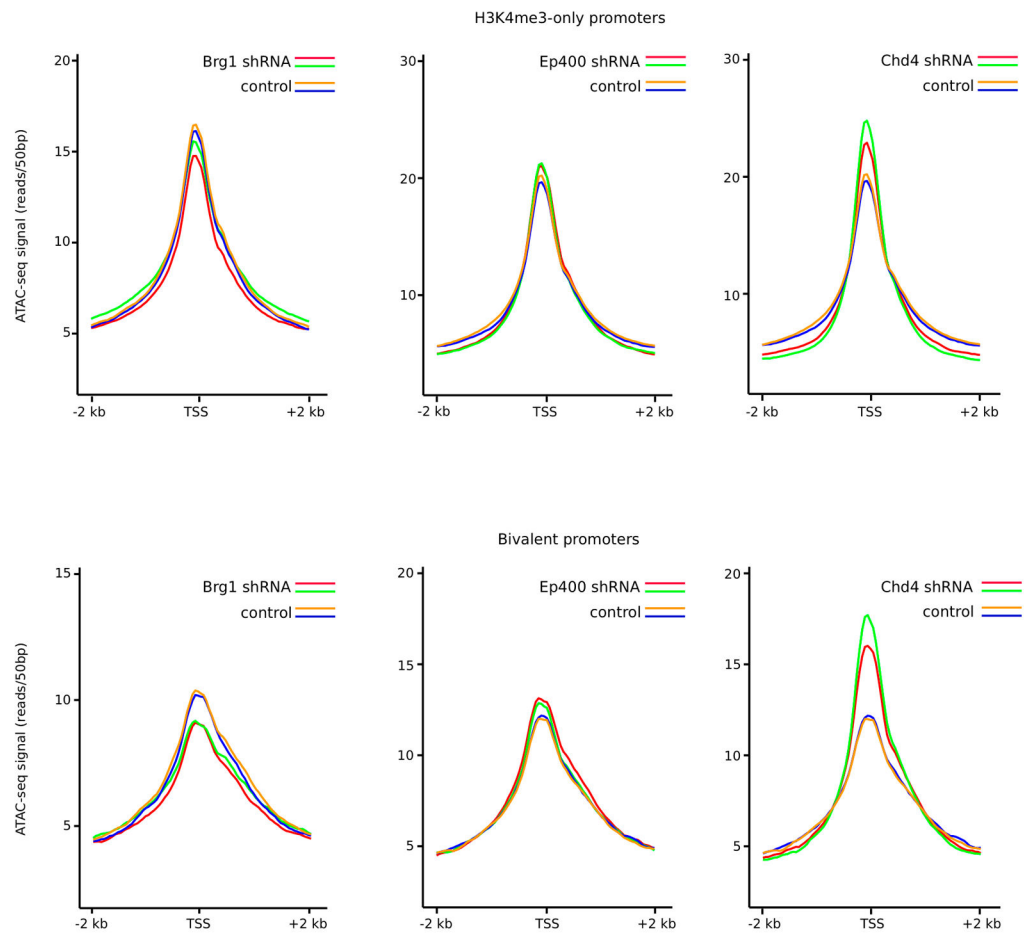


**Extended Data Figure 6. Western blot analysis of remodeller depletion by shRNA for transcriptome analysis**

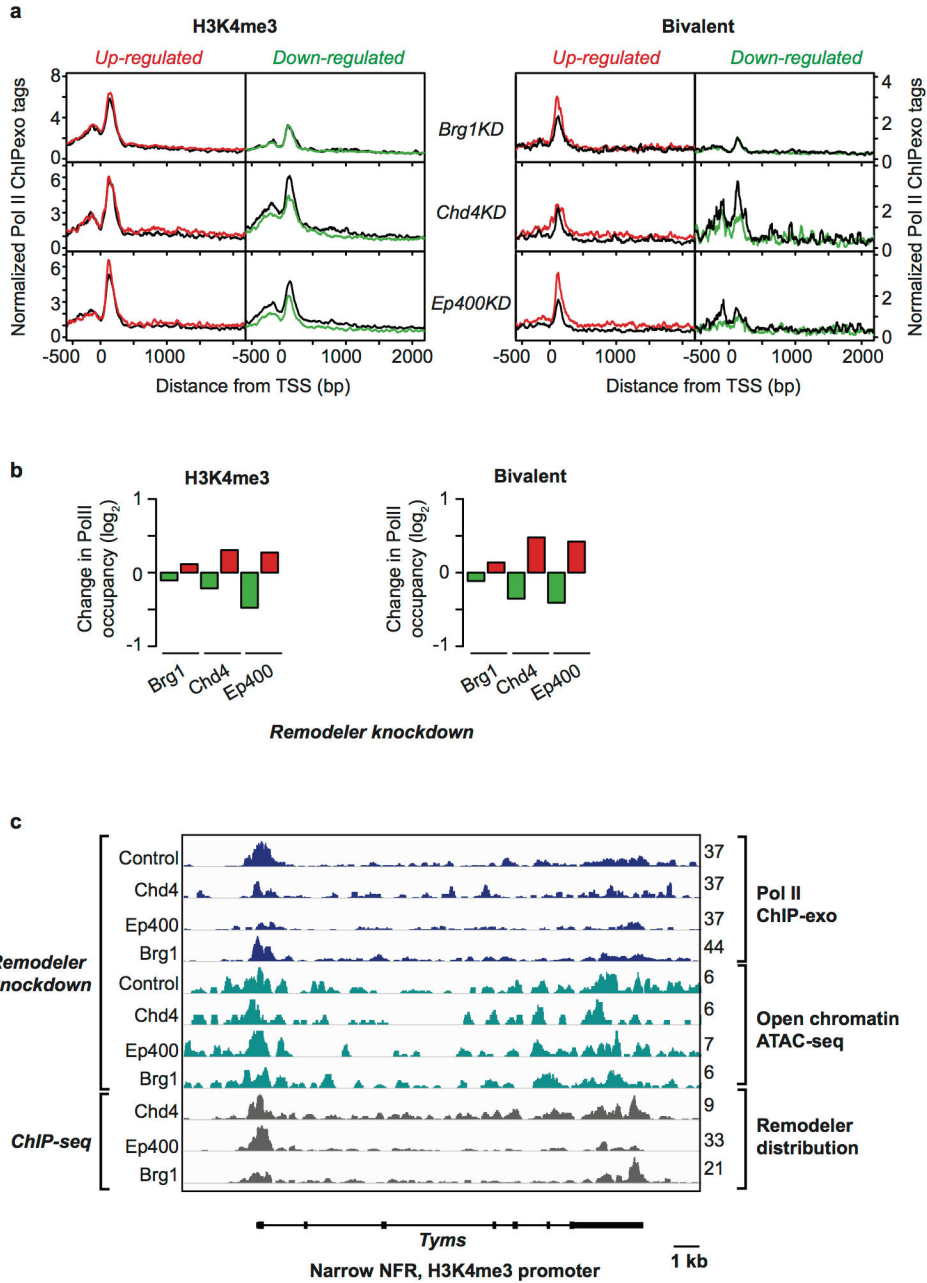
ES cells tagged for each remodeller were transfected with the corresponding shRNA vector, or a control plasmid. After puromycin selection, ES cells were collected for RNA preparation and Western blot analysis. Three independent experiments (indicated as 1, 2, and 3) were performed for each remodeller. Remodeller depletion was assessed using antibodies against FLAG or HA epitopes. Loading control: Gapdh. For gel source data, see Supplementary Figure 1.



**Extended Data Figure 7. Validation of remodeler depletion effects on transcription by RT-qPCR**  
Remodellers and histone marks enrichment profiles are shown as indicated on the left of each panel. A control ChIP profile, obtained with untagged ES cells, is shown for comparison. Scores indicate reads per 10 million. On the right of each panel are shown the results of RT-qPCR analysis that quantify RNA expression levels of the corresponding genes upon remodeler depletion in ES cells. Two distinct shRNA vectors (shRNA1 and shRNA2, see Methods) were used for each remodeler. Scores on the y-axis indicate the relative expression of the indicated genes compared to reference genes. Values are means and standard deviations of three independent transfection experiments.



**Extended Data Figure 8. Impact of remodeler depletion on chromatin accessibility at promoters**  
 Consequence of remodeler depletion by shRNA vectors on ATAC-seq average profiles at all H3K4me3-only (top panels) and bivalent (bottom panels) promoters. Two replicate experiments are shown on each graph for both remodeler knockdown and controls.



**Extended Data Figure 9. Analysis of pol II distribution at promoters in remodeller depleted ES cells**

**a**, Average pol II distribution (ChIP-exo) profile in control ES cells (black) or upon indicated remodeller knockdown (colour) at H3K4me3-only (left set) and bivalent genes (right set). Left and right panels within a set represent the set of genes that are most up- (red) or down-regulated (green) upon remodeller knockdown. Pol II occupancy is indicated within a window spanning 500 and 2000 bp on the upstream and downstream side of the TSS, respectively. All promoters are transcribed from left to right. **b**, Bargraphs showing Pol II occupancy change upon remodeller knockdown relative to control, measured by ChIP-exo, at genes either down-regulated (green) or up-regulated (red) following the depletion of the



indicated remodeler. **c**, Pol II distribution of remodeler knockdown at a representative locus. Counts indicate reads per 10 million. Pol II loading is markedly reduced at the *Tyms* narrow NFR, H3K4me3 promoter by either Ep400 and Chd4 depletion, suggesting that these two remodellers contribute to Pol II recruitment.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank J.C. Andrau, S. Ravens, L. Tora and M. de Chaldée for critical reading, J.B. Charbonnier's team for sharing material, A. Krebs, T. Ye, I. Davidson, R. Guerois and F. Ochsenbein for discussions, A. Martel for computational assistance. This work was supported by the CEA, and by grants from the Association pour la Recherche sur le Cancer (grant 3164), the Agence Nationale de la Recherche (ANR-05-BLAN-0396), the Fondation pour la Recherche Médicale (FRM), the National Institutes of Health (grant HG004160, to B.F.P.) and from Southern Medical University (grant B1000465, to K.Y.).

## References

- Narlikar GJ, Sundaramoorthy R, Owen-Hughes T. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell*. 2013; 154:490–503. DOI: 10.1016/j.cell.2013.07.011 [PubMed: 23911317]
- Becker PB, Workman JL. Nucleosome remodeling and epigenetics. *Cold Spring Harbor perspectives in biology*. 2013; 5
- Cairns BR. The logic of chromatin architecture and remodelling at promoters. *Nature*. 2009; 461:193–198. DOI: 10.1038/nature08450 [PubMed: 19741699]
- Yen K, Vinayachandran V, Batta K, Koerber RT, Pugh BF. Genome-wide nucleosome specificity and directionality of chromatin remodelers. *Cell*. 2012; 149:1461–1473. DOI: 10.1016/j.cell.2012.04.036 [PubMed: 22726434]
- Simic R, et al. Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. *The EMBO journal*. 2003; 22:1846–1856. DOI: 10.1093/emboj/cdg179 [PubMed: 12682017]
- Smolle M, et al. Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. *Nature structural & molecular biology*. 2012; 19:884–892. DOI: 10.1038/nsmb.2312
- Teif VB, et al. Genome-wide nucleosome positioning during embryonic stem cell development. *Nature structural & molecular biology*. 2012; 19:1185–1192. DOI: 10.1038/nsmb.2419
- Whyte WA, et al. Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature*. 2012; 482:221–225. DOI: 10.1038/nature10805 [PubMed: 22297846]
- Fenouil R, et al. CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome research*. 2012; 22:2399–2408. DOI: 10.1101/gr.138776.112 [PubMed: 23100115]
- Min IM, et al. Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes & development*. 2011; 25:742–754. DOI: 10.1101/gad.2005511 [PubMed: 21460038]
- Giresi PG, Kim J, McDaniel RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome research*. 2007; 17:877–885. DOI: 10.1101/gr.5533506 [PubMed: 17179217]
- Yildirim O, et al. A system for genome-wide histone variant dynamics in ES cells reveals dynamic MacroH2A2 replacement at promoters. *PLoS genetics*. 2014; 10:e1004515. [PubMed: 25102063]
- Yukawa M, et al. Genome-wide analysis of the chromatin composition of histone H2A and H3 variants in mouse embryonic stem cells. *PloS one*. 2014; 9:e92689. [PubMed: 24658136]

14. Hu G, et al. H2A.Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation. *Cell stem cell*. 2013; 12:180–192. DOI: 10.1016/j.stem.2012.11.003 [PubMed: 23260488]
15. Marks H, et al. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell*. 2012; 149:590–604. DOI: 10.1016/j.cell.2012.03.026 [PubMed: 22541430]
16. Creighton MP, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107:21931–21936. DOI: 10.1073/pnas.1016071107 [PubMed: 21106759]
17. Bernstein BE, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*. 2006; 125:315–326. DOI: 10.1016/j.cell.2006.02.041 [PubMed: 16630819]
18. Ho L, et al. esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. *Nature cell biology*. 2011; 13:903–913. DOI: 10.1038/ncb2285 [PubMed: 21785422]
19. Ramirez-Carrozzi VR, et al. A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell*. 2009; 138:114–128. DOI: 10.1016/j.cell.2009.04.020 [PubMed: 19596239]
20. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open.
21. Liu P, Jenkins NA, Copeland NG. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome research*. 2003; 13:476–484. DOI: 10.1101/gr.749203 [PubMed: 12618378]
22. Ying QL, Stavridis M, Griffiths D, Li M, Smith A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nature biotechnology*. 2003; 21:183–186. DOI: 10.1038/nbt780
23. Tessarollo L. Manipulating mouse embryonic stem cells. *Methods Mol Biol*. 2001; 158:47–63. DOI: 10.1385/1-59259-220-1:47 [PubMed: 11236671]
24. Pruitt KD, et al. RefSeq: an update on mammalian reference sequences. *Nucleic acids research*. 2014; 42:D756–763. DOI: 10.1093/nar/gkt1114 [PubMed: 24259432]
25. Ye T, et al. seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic acids research*. 2011; 39:e35. [PubMed: 21177645]
26. Albert I, Wachi S, Jiang C, Pugh BF. GeneTrack--a genomic data processing and visualization framework. *Bioinformatics*. 2008; 24:1305–1306. DOI: 10.1093/bioinformatics/btn119 [PubMed: 18388141]
27. Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. *Nature protocols*. 2012; 7:1728–1740. DOI: 10.1038/nprot.2012.101 [PubMed: 22936215]
28. Berlivet S, Houliard M, Gerard M. Loss-of-function studies in mouse embryonic stem cells using the pHYPER shRNA plasmid vector. *Methods Mol Biol*. 2010; 650:85–100. DOI: 10.1007/978-1-60761-769-3\_7 [PubMed: 20686945]
29. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods*. 2008; 5:621–628. DOI: 10.1038/nmeth.1226 [PubMed: 18516045]
30. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*. 2009; 10:R25. [PubMed: 19261174]
31. Giresi PG, Lieb JD. Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements). *Methods*. 2009; 48:233–239. DOI: 10.1016/j.ymeth.2009.03.003 [PubMed: 19303047]
32. Carriere L, et al. Genomic binding of Pol III transcription machinery and relationship with TFIIS transcription factor distribution in mouse embryonic stem cells. *Nucleic acids research*. 2012; 40:270–283. DOI: 10.1093/nar/gkr737 [PubMed: 21911356]
33. Rhee HS, Pugh BF. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell*. 2011; 147:1408–1419. DOI: 10.1016/j.cell.2011.11.013 [PubMed: 22153082]
34. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25:1754–1760. DOI: 10.1093/bioinformatics/btp324 [PubMed: 19451168]

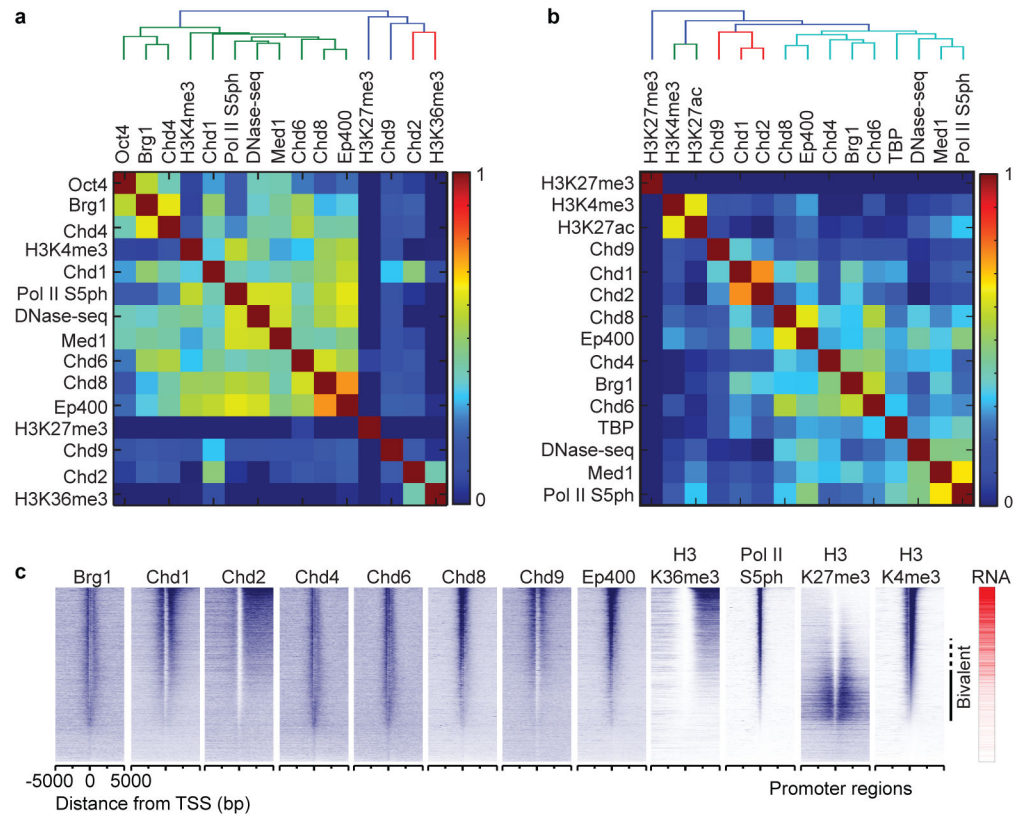
35. Ho L, et al. An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106:5187–5191. DOI: 10.1073/pnas.0812888106 [PubMed: 19279218]
36. Kagey MH, et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*. 2010; 467:430–435. DOI: 10.1038/nature09380 [PubMed: 20720539]
37. Rahl PB, et al. c-Myc regulates transcriptional pause release. *Cell*. 2010; 141:432–445. DOI: 10.1016/j.cell.2010.03.030 [PubMed: 20434984]

Author Manuscript

Author Manuscript

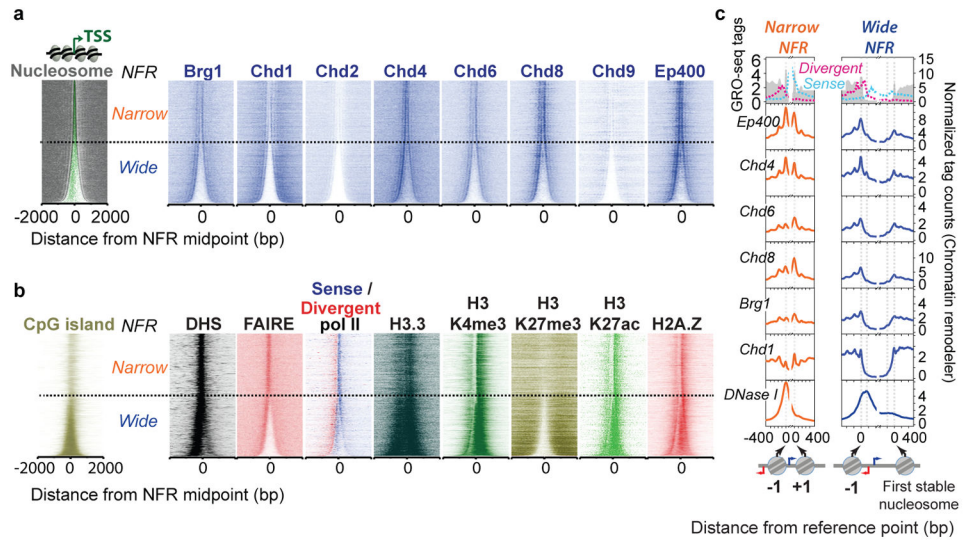
Author Manuscript

Author Manuscript



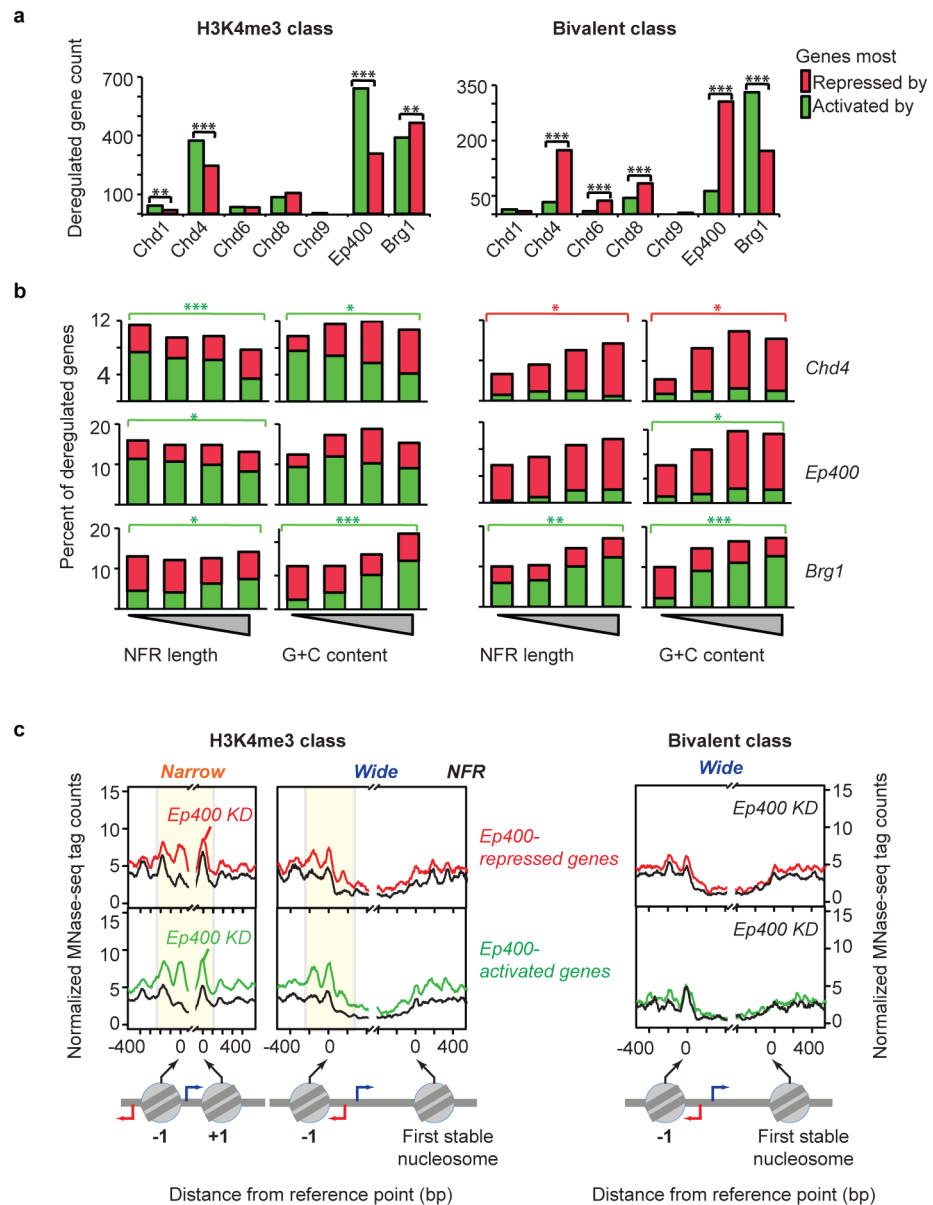
**Figure 1. Correlated occupancies across remodeler-bound nucleosomal regions**

**a**, Heat map representing Pearson correlations between remodelers and other factors within 500 bp of 138,582 DHS midpoints. **b**, Same as (a) but for 16,300 promoter-like, H3K4me3-, TBP- and Pol II S5ph-positive DHSs. **c**, Distribution of remodeler-nucleosome interactions (MNase ChIP-seq tags for the indicated remodelers in blue) aligned at 14,623 individual RefSeq TSSs (rows), sorted by H3K4me3 levels. Corresponding RNA expression levels (red) are shown.



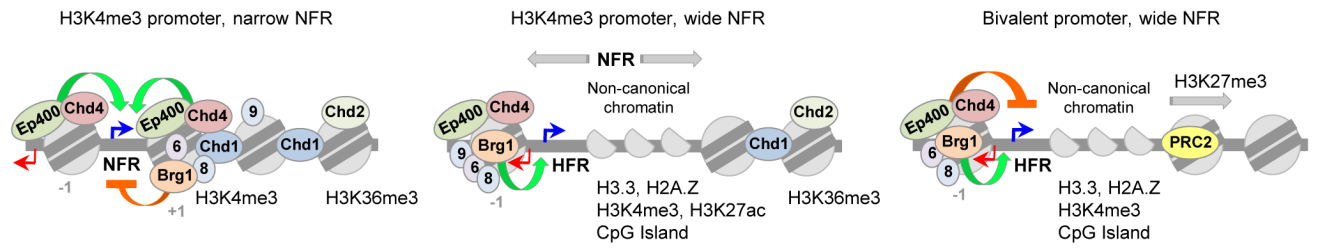
**Figure 2. Patterns of remodeller-nucleosome interactions and chromatin features around promoter NFRs**

**a.** Distribution of remodeller-nucleosome interactions, as in Fig. 1c, except aligned by NFR midpoint and sorted by NFR length. Standard MNase-defined nucleosomes (grey) and TSS (green) are shown. Narrow and wide NFRs are delineated by the dashed line. **b.** Same as in (a) for other genomic features. **c.** Averaged distribution of remodeller-nucleosome interactions from (a, b) at narrow and wide NFRs, aligned to the dyad of  $-1$  (left portion of each graph) or the first MNase-resistant nucleosome downstream of the noncanonical chromatin (right portion). Standard nucleosomes (grey fill) and GRO-Seq RNA (blue and red dashed lines) are shown. A gap in the NFR midpoint was introduced to account for variations in NFR length inside each class.



**Figure 3. Remodellers differentially regulate active versus bivalent genes**

**a.** The number of genes whose RNA was either down- (green) or up-regulated (red) following remodeller depletion by >1.5 fold is shown for H3K4me3-only or bivalent genes. Statistical analysis involved a 2-sample test for equality of proportions with continuity correction: \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ . **b.** The percentages of H3K4me3 (left) or bivalent (right) genes up- (red) or down-regulated (green) by remodeller depletion are shown in four subgroups based on NFR length, as defined in Fig. 2, and in quartiles for C+G content. Statistical significance metrics (described in panel **a**) are colour-matched and applied to the first and last group. **c.** Averaged nucleosome distribution (MNase-seq) upon control (black) or Ep400 (colour) knockdown at H3K4me3-only (narrow versus wide NFR) and bivalent gene groups. Upper and lower panels represent genes that are up- (red) or down-regulated (green) (>1.5 fold) upon Ep400 knockdown.



**Figure 4. Model of how remodellers might regulate distinct classes of genes in ES cells**

Three gene classes are indicated, having remodeller-bound nucleosomes (coloured circles on top of grey circles) at specific positions relative to the TSS (horizontal blue arrow). MNase-sensitive noncanonical chromatin, having histone variants and active marks, is shown as half circles. Curved green and orange ribbons indicate transcriptional activation and repression, respectively. Single digit numbers denote corresponding Chd remodellers.