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

Maude de Dieuleveult1,*, Kuangyu Yen2,3,§, Isabelle Hmitou1,*, Arnaud Depaux1,*, Fayçal Boussouar4,†, Daria Bou Darghami1, Sylvie Jouinier1, Hélène Humbert-claude1, Florence Ribierrière4, Céline Baulard4, Nina P. Farrell3, Bongsoo Park3, Céline Keime5, Lucie Carrière1, Soizick Berlivet1, Marta Gut6, Ivo Gut6, Michel Werner1, Jean-François Deleuze4, Robert Olaso4, Jean-Christophe Aude1, Sophie Chantalat4, B. Franklin Pugh3§ & Matthieu Gérard1§

ATP-dependent chromatin remodellers allow access to DNA for transcription factors and the general transcription machinery, but whether mammalian chromatin remodellers1–3 target specific nucleosomes to regulate transcription is unclear. Here we present genome-wide remodeller–nucleosome interaction profiles for the chromatin remodellers Chd1, Chd2, Chd4, Chd6, Chd8, Brg1 and Ep400 in mouse embryonic stem (ES) cells. These remodellers bind one or both full nucleosomes that flank micrococcal nuclease (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 are nevertheless bound by non-nucleosomal or subnucleosomal histone variants (H3.3 and H2A.Z) and marked by H3K4me3 and H3K27ac modifications. RNA polymerase II therefore navigates hundreds of base pairs of altered chromatin in the sense direction before encountering an MNase-resistant nucleosome at the 3′ end of the NFR. Transcriptome analysis after remodeller depletion reveals reciprocal mechanisms of transcriptional regulation by remodellers. Whereas at active genes individual remodellers have 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 assay3 (MNase digestion to define nucleosomes, followed by remodeller chromatin immunoprecipitation and sequencing (ChIP-seq)) to ES cells, focusing on the 5′ end of genes (Extended Data Fig. 1 and Supplementary Table 1). We first examined remodeller co-enrichment with other factors such as RNA polymerase (Pol) II, selected histone marks and transcription factors, over broad (500-base-pair (bp)) windows centred on DNase-I hypersensitive sites (DHSs) (that is, promoters and enhancers; n = 138,582) (Fig. 1a). High Pearson correlation scores were observed among the remodellers Brg1 (also known as Smarca4), 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 focusing on annotated transcriptional start sites (TSSs) (Fig. 1c and Extended Data Fig. 3). Remarkably, some remodellers such as Brg1, Chd4 and Chd6 bound similar nucleosome positions at all active genes, regardless of their H3K4me3 enrichment (which is a mark of transcriptional activity), whereas others, such as Chd1, Chd2, Chd9 and Ep400, were tightly linked to histone 3 Lys4 trimethylation (H3K4me3)/transcription levels. Chd8 had an intermediate pattern. Chd1 and Chd2, which are both related to Saccharomyces cerevisiae (yeast) Chd1, showed markedly 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 works5,6, and thus mammalian Chd2 and yeast Chd1 may be functionally equivalent.

Figure 1 | Correlated occupancies across remodeller-bound nucleosomal regions. a, Heat map representing Pearson correlations between remodellers and other factors within 500 bp of 138,582 DHS midpoints. b, Same as in a but for 16,300 promoter-like, H3K4me3−, TBP- and Pol II S5ph-positive DHSs. c, Distribution of remodeller–nucleosome interactions (MNase ChIP-seq tags for the indicated remodellers in blue) aligned at 14,623 individual RefSeq TSSs (rows), sorted by H3K4me3 levels. Corresponding RNA expression levels (red) are shown.

1 Institute for Integrative Biology of the Cell (IBISC), IBITECS, CEA, CNRS, Université Paris-Sud, Université Paris-Saclay, 91198 Gif-sur-Yvette, France. 2Department of Veterinary Science, University of Illinois, Urbana-Champaign, Illinois 61801, USA. 3CNRS, UMR7104, Université de Strasbourg, 67400 Illkirch, France. 4Centre National de Généturgie, Institut de Génétique et de Biologie Moléculaire et Cellulaire, INSERM, U964, CNRS, UMR7104, Université de Strasbourg, 67400 Illkirch, France. 510515, China. 6Centre for Eukaryotic Gene Regulation, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. *These authors jointly supervised this work. §These authors contributed equally to this work.
We next investigated more closely the relationship between individual remodeller-bound nucleosomes and all nucleosomes defined by MNase-resistant mononucleosome-sized DNA fragments. Plots of individual genes were aligned by their 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. a, Same as in a for other genomic features. b, Same as in a for other genomic features. c, Averaged distribution of remodeller–nucleosome interactions from a and 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.

We examined the distribution of histone variants and marks in NFRs, measured by standard ChIP-seq. 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.

At narrow NFRs, the Ep400-bound and Chd4-bound nucleosomes 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. Chd4-enriched, and contain remarkably noncanonical chromatin (being DNase I resistant but MNase sensitive).

A subset of less active genes is marked by a combination of H3K4me3 and H3K27me3, defining them as bivalent17 (Fig. 1c). Notably, bivalency does not predominate on the same nucleosomes (that is, H3K4me3 and H3K27me3 are enriched in NFRs, whereas H3K27me3 resides downstream, over full nucleosomes) (Fig. 2b). Thus, wide 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). This −1 nucleosome also represented the peak of DNase I hypersensitivity (bottom left). These remodelers therefore might be involved in structural reorganization, ejection or repositioning of −1 to regulate sense and divergent transcription...
To investigate how remodelers, bound at these distinct classes of genes, are involved in transcription regulation in ES cells, we depleted each remodeler using short hairpin RNA (shRNA) vectors (Extended Data Fig. 6) and profiled messenger RNA expression or used a published deletion data set \(^{18}\) for Brg1 (Fig. 3a). We observed that Chd4, Ep400 and Brg1, among the tested remodelers, 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 \(^{18}\). Loss-of-function of the other remodelers resulted in more limited changes in gene expression. These results were validated by reverse transcriptase quantitative PCR (RT-qPCR) using two different shRNA vectors for each remodeler (Extended Data Fig. 7).

Since CpG islands help determine remodeler requirements \(^{19}\), 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 with 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, remodelers and chromatin architecture commonly have reciprocal relationships between H3K4me3 and bivalent genes.

To understand how specific remodelers might regulate genes, we used ATAC-seq (assay for transposase-accessible chromatin-sequencing) \(^{20}\) to examine regional chromatin accessibility after remodeler 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 after 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 remodeler 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 remodelers contribute to Pol II recruitment at subsets of active promoters.

Three contrasting stereotypes of remodeler control of gene expression in mouse ES cells arise from the data (Fig. 4), although not all

**Figure 3** | Remodellers differentially regulate active versus bivalent genes. a, The number of genes in which RNA was either downregulated (green) or upregulated (red) after remodeler depletion by >1.5-fold is shown for H3K4me3-only or bivalent genes. \( ^{*} P < 0.05, ^{*}{^{*}} P < 0.01, \) \( ^{****} P < 0.001 \) (two-sample test for equality of proportions with continuity correction). b, The percentages of H3K4me3 (left) or bivalent (right) genes upregulated (red) or downregulated (green) by remodeler 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 a) are colour-matched and applied to the first and last group. c, Averaged nucleosome distribution (MNase-seq) after control (black) or Ep400 (colour) knockdown at H3K4me3-only (narrow versus wide NFR) and bivalent gene groups. Top and bottom panels represent genes that are upregulated (red) or downregulated (green) (>1.5-fold) after Ep400 knockdown (KD).

**Figure 4** | Model of how remodelers might regulate distinct classes of genes in ES cells. Three gene classes are indicated, having remodeler-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 remodelers.
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 Ep-400. 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 using the H3K36me3 mark to organize nucleosomes analogously to Chd1 or Isw1b in budding yeast.

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 non-canonical chromatin or partial nucleosomes (for example, 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: an activating remodeller in one class of genes is an inhibitor remodeller in the other class; and 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.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 January; accepted 1 December 2015.
Published online 27 January 2016.

Author Contributions We thank J. C. Andraud, 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. Guerios and F. Ochsenbein for discussions, and 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.).

Author Information Data have been deposited in Gene Expression Omnibus (GEO) under accession numbers GSE64825 (ChIP-seq, ATAC-seq, MNase-seq and ChIP-exo data) and GSE64819 (microarray). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.G.e (matthieu.gerard@cea.fr), K.Y. (kuangyuyen@smu.edu.cn) or B.F.P. (bfp2@psu.edu).
METHODS

No statistical methods were used to predetermine sample size. Experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

**Knock-in of a TAP-tag in the genes encoding the remodelers through homologous recombination in ES cells.** The recombineering technique\(^*\) was adapted to construct all targeting vectors for homologous recombination in ES cells. Retriever vectors were obtained by using 3’ miniarm (NotI/SpeI), 3’ miniarm (SpeI/ BamHI) or the plasmid PL253 (NotI/BamHI). SV40 cells containing a BAC encompassing the carboxy-terminal part of the gene encoding the remodeler, were electroporated with the SpeI-linearized retrieval vector. This allowed the subcloning of genomic fragments of approximately 10 kilobases (kb) comprising the last exon of the gene encoding each remodeler. The next step was the insertion of a TAP-tag into the subcloned DNA, immediately 3’ to the coding sequence. The TAP-tag was (Flag)\(^*\)-TEV-HA for Chd1, Chd2, Chd4, Chd6, Chd8, Ep400, Brg1 and 6His-Flag-HA for Chd9. We first inserted the TAP-tag and an Ascl site into the PL452 vector, to clone 5’ homology arms as SalI/Ascl 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 on 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 haemagglutinin (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 5 days after plating, using the Millipore alkaline detection kit, following manufacturer’s instructions. In addition, we verified by immunofluorescence using an antibody against Oct4 (also known as Pou5f1) (Abcam ab18957, 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\(^*\). 46C ES cells and their tagged derivatives were cultured at 37°C, 5% CO\(_2\), on mitomycin C-inactivated mouse embryonic fibroblasts, in DMEM (Sigma) with 15% fetal bovine serum (Invitrogen), l-glutamine (Invitrogen), MEM non-essential amino acids (Invitrogen), penicillin/streptomycin (Invitrogen), l-mercaptoethanol (Sigma), and a saturating amount of leukaemia inhibitory factor (LIF), as described previously\(^*\).

**Reference ES cell nucleosome map and NFR categories.** Mouse ES cells were acquired from a published MNase-seq data set\(^*\) 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. Because 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 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 139,454 DHS regions. Each of these DHS regions consists of 10 kilobases (kb) and was assigned CpG island information was retrieved from UCSC (mm9 build) and assigned to the coding sequence.

**H3K4me3-only and bivalent gene lists.** The list of 14,623 genes used in Figs 1 and 2 was obtained by filtering all mm9 RefSeq genes\(^*\). We removed redundancies (that is, genes having the same start and end sites), unmappable genes, blacklisted genomic regions (those with artefact signal regardless of 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 defined 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: GMS590111). This was accomplished by operating with the seqMINER platform. Tag densities from this data set were collected in a −500/+1,000 bp window around the TSS, and subjected to three successive rounds of k-means clustering, 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 different settings (Supplementary Table 2). We used the top 3% of the ranked H3K4me3 scores as the starting point for this analysis, 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 data set were summed up

Tandem affinity purification of MNase-digested remodeler–nucleosome complexes. A detailed version of this protocol is available on the protocol exchange website: http://dx.doi.org/10.1038/protex.2014.04. In brief, about 400 million ES cells were fixed either with formaldehyde, or with a combination of disuccinimidyl glutarate (DSS) and formaldehyde (Supplementary Table 1), then permeabilized with IGEPAL, and incubated with 2,800 units of micrococcal nuclease (MNase, New England Biolabs) in order to fragment the genome into mononucleosomes (Extended Data Fig. 1). This nuclease preparation was next incubated with antibodies coupled with an antibody anti-HA or anti-Flag. Anti-HA-agarose (ref. A2295) and anti-Flag-agarose (ref. A2220) beads were purchased from Sigma. After a series of washes, tagged remodeler–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. Two biological replicates were used for each tagged and control ES cell line, using independent cell cultures and chromatin preparations.

**High-throughput sequencing of MNase remodeler ChIP samples.** After crosslink reversal, phenol–chloroform extraction and ethanol precipitation, the DNA from remodeler–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). Approximately 5–10 ng of ChIP DNA was used for library preparation according to the Illumina ChIP-seq protocol (ChIp-seq sample preparation kit). Following end-repair and adaptor 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. Sequence data analysis was performed using the seqMINER platform (Fig. 1c). To examine the distribution of remodelers at individual genes, we used WigMaker3 (default settings) to convert BED files into wig files, which were uploaded onto the IGV genome browser (Extended Data Fig. 2).

Nucleosome cells were made from MNase remodeler ChIP-seq tags using GeneTrack\(^*\) with the following parameters: sigma = 20, exclusion = 146. We then globally shifted tags to the median value of half distances of all nucleosome cells. GRO-seq tags\(^*\) sharing the same or opposite orientation with the ‘TSS’ were assigned as ‘sense’ and ‘divergent’ tags, respectively. The orientation of each ‘TSS’ was arranged so that sense transcription proceeds to the right. ES nucleosomal tags, globally shifted tags from MNase remodeler ChIP-seq (this current study), tags from DHS regions (Mouse ENCODE), GRO-seq oriented (from transcription start) and transcriptionally engaged Pol II and Cpg islands (UCSC, mm9 build) were then aligned to the midpoints of each ‘TSS’. Promoter regions were then sorted by ‘TSS’ 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\(^*\) (Fig. 2b).

Tags from reference nucleosome\(^*\) remodeler-interacting nucleosomes (this study) and transcriptionally engaged Pol II (GRO-seq)\(^*\) were also 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’, the orientation of which was arranged so that the transcription proceeds to the right. After normalization to the gene count in the two different NFR subcategories, 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 DNASel-Seq data from the mouse ENCODE consortium (GSM1004653) for the identification of DHS regions in the mouse ES cells (GSM881725, GSM881726 and GSM881727, 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 data set 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 heat map (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 SSF. We obtained 16,300 promoter-like DHS regions bfitting the criterria. Pair-wise 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 pgyHER shRNA vector for remodeller depletion in ES cells, as previously described. shRNA design was performed using DSIR software (http://biodivet.extra.cea.fr/DSIR/DSIR.html). Below are the shRNAs selected for each remodeller. The sense strand sequence is given; the rest of the shRNA sequence is as described previously. Chd1 shRNA 1: 5′-GGCAAGACCGGCGACTAAGAGA-3′; Chd1 shRNA 2: 5′-GCACTGGTCTAAATCAGTGC-3′; Chd4 shRNA 1: 5′-GGACGGCATTTAGTATGAGA-3′; Chd4 shRNA 2: 5′-GCTGCACTTCTCCCAAAGATG-3′; Chd6 shRNA 1: 5′-GATCACTGCTGCTCTA3′; Chd6 shRNA 2: 5′-GCTGCTCTGTA CCGCCTAAGA-3′; Chd8 shRNA 1: 5′-GATTACCCACATTTCCCAGAAGGC-3′; Chd9 shRNA 2: 5′-AGTTAATGCTACTCAAGATT-3′; Ep400 shRNA 1: 5′-GGTAAAGAGTCCAGATTAAAG-3′; Ep400 shRNA 2: 5′-GGTCCACACCTCAAAACAGGCC-3′; Smarcal shRNA 1: 5′-ATCTCTTGTAGATAGATTCACCT-3′; Smarcal shRNA 2: 5′-CCTCCGAATGGTCGTAAGAT-3′.

Each shRNA was transfected in its corresponding tagged ES cell line, 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 ab8245).

The pHYPER shRNA vectors were transfected in ES cell by electroporation, using an Anmaxa nucleofector (Lonza). Twenty-four hours after transfection, puromycin (2 μg/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 30th 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 adapters were ligated. The ligation product was then purified and enriched with 15 cycles of PCR to create a final library for single-read sequencing of 75 bp carried out on an Illumina GAIIx.

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/). Mapping of these sequences onto the mm9 assembly of mouse genome and RPKM computation were then performed using ERANGE v3.1.0 (ref. 29) and done using Genespring software (version 13.0-GX). The remaining 18,459 (41%) are for other transcripts. BeadChips were scanned using 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). Following analyses were done using Genespring software (version 13.0-GX).

For Brg1, we used a previously published transcriptome data set, in which loss of Brg1 function was obtained by genetic ablation. All array analyses were undertaken using the RMA method as implemented in the R affy package (R-Development-Core-Team, 2007). Microarray spot intensities were normalized using the RNA-seq method and as equivalent to shrinkage (or CpGisition) of the estimated sample variance towards a pooled estimate, resulting in a more stable inference. Finally, adjusted P values were calculated using the false discovery rate (FDR)-controlling procedure of Benjamini and 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 (ref. 18), 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 two-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 figure 3 of ref. 15, which we crossed with the 14,623 promoter list, to obtain a list of 6,317 promoters rank ordered according to GC content.

In Fig. 3b, we compared the percentages of genes either down- or upregulated 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 < 15bp), narrow NFR subclass 2 (15–115 bp NFR), wide NFR subclass 1 (116–504 bp) and wide NFR subclass 2 (505–1,500 bp). Genes in these groups were further subdivided into H3K4me3 and bivalent subgroups. (2) NFR subclass 1: genes were further 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 (1,829), wide NFR subclass 1 (2,613), subclass 2 (1,253), 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 (2,266), subclass 2 (1,184), GC content quartile 1 (220), quartile 2 (485), quartile 3 (750) and quartile 4 (1149).

FAIRE-seq. FAIRE was performed as described 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 and 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 s on/30 s 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. Before 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). Approximately 1 μg of FAIRE DNA was then purified and transferred with shRNA vectors as described above. Biological replicates were obtained by performing two independent transfection experiments for each shRNA vector. ATAC-seq libraries were constructed by adapting a published protocol. In brief, 50,000 cells were collected, washed with cold PBS and resuspended in 50 μl of ES buffer (10 mM Tris, pH 7.4, 10 mM NaCl, 3 mM MgCl2). Permeabilized cells were resuspended in 50 μl transposase reaction (1 x transposase buffer, 1000–1500 μl Tn5 transposase enzyme (Illumina)) and incubated for 30 min at 37 °C. Subsequent steps of the protocol were performed as previously described. 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 pair-ended sequenced (2 x 100 bp) on a HiSeq instrument (Illumina).

Analysis of ATAC-seq data. As ATAC-seq libraries are composed in large part of short genomic DNA fragments, reads were cropped to 50 bp using trimmomatic-0.32 to optimize mapping and alignment. Reads were aligned to the mouse genome (mm9) using Bowtie with the parameters -m1-best-strata -X2000, with two 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 data sets were next converted to BED. We used the seqMINER platform with the lists of 6,481 H3K4me3-only and 3,411 bivalent genes described above, to collect tag densities from ATAC-seq data sets, 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 following remodeller depletion in ES cells. ES cells were grown and transfected with shRNA vectors as described above. Biological replicates were obtained by performing two independent transfection experiments for each shRNA vector. For each experiment, 1 million cells were fixed 10 min in ES cell culture medium containing 1% formaldehyde, quenched with glycerine (125 mM), washed with PBS buffer, collected in 175 μl of solution I (15 mM Tris-HCl, pH 7.5, 0.3 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM MgCl2 and 0.1 mM EGTA), and stored on ice. Cells were permeabilized by adding 175 μl of solution I (solution I with 0.8% Igepal CA-630 (Sigma)) and incubating for 15 min on ice. We next added 700 μl of MNase digestion buffer (50 mM Tris-HCl, pH 7.5, 0.3 M sucrose, 15 mM KCl, 60 mM NaCl, 4 mM MgCl2, and 2 mM CaCl2), 4 U of MNase, and incubated for 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 25 G needle, followed by a 10 min centrifugation at 18,000g. The supernatant was collected and incubated for 1 h at 65 °C with 15 μg of RNase A. We next added 10 μg of proteinase K, adjusted each sample to 0.1% SDS (final concentration) and incubated for 2 h at 37 °C. After incubation, 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.

Purified DNA (20 ng) was used for library preparation according to manufacturer’s instructions, using Ultralow ovation library system (Nugen). Following end-repair and adaptor ligation, fragments were size-selected onto an agarose gel in order to purify genomic DNA fragments between 50 and 220 bp. LIBRARYs 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-ended 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 (5 bp) 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-seq. ES cells were grown and transfected with shRNA vectors as described above. Biological replicates were obtained by performing two independent transfection experiments for each shRNA vector. Following a 10 min fixation with 1% formaldehyde in ES cell culture medium, chromatin was prepared from 5–10 million cells and sonicated as described. ChIP-exo experiments were carried out essentially as described. 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 (ESAA2), and lambda and rec exonuclease digestion on the beads. After elution, a primer was annealed to ESAA2 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). 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 GenTrack, 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 carried out remodeller-affected genes. We assumed that having remodeller deletion bore no bulk change on Pol II occupancy, and hence total tags among wild type and all remodeller mutants were normalized to be the same. 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 (5 bp binned before 10-bin moving average) before plotting (Extended Data Fig. 9a).

To examine Pol II 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) 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 transformation and barcharm plotting (Extended Data Fig. 9b).

Average binding profiles (Extended Data Fig. 3). Genes were rank-ordered according to reads per kb of transcript per million mapped reads (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 data sets were collected in a window of −2 kb/+2 kb around the TSS, except for Chd2, for which densities were collected from the TSS until +4 kb. Tag output 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 and 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 = 0.137722; Q3 – Q2 = 0.002259; Q2 – Q1 = 7.985217 × 10−10; Chd2, Q4 – Q3 = 7.543473 × 10−33; Q3 – Q2 = 1.155223 × 10−25; Q2 – Q1 = 3.283427 × 10−38; Chd4, Q4 – Q3 = 0.2094255; Q3 – Q2 = 0.0184552; Q2 – Q1 = 0.0702886. Chd6, Q4 – Q3 = 0.4168748; Q3 – Q2 = 0.1534144; Q2 – Q1 = 0.0113803. Chd8, Q4 – Q3 = 0.4031995 × 10−15; Q3 – Q2 = 0.1231527 × 10−6; Q2 – Q1 = 1.34455 × 10−9. Chd9, Q4 – Q3 = 9.485857 × 10−4; Q3 – Q2 = 1.059783 × 10−16; Q2 – Q1 = 4.666352 × 10−10. Ep400, Q4 – Q3 = 3.046796 × 10−20; Q3 – Q2 = 1.215304 × 10−14; Q2 – Q1 = 6.466267 × 10−11. Brg1, Q4 – Q3 = 3.512021 × 10−6; Q3 – Q2 = 2.515217 × 10−12; Q2 – Q1 = 0.977422.

We concluded from this analysis that Chd1, Chd2, Chd8 and Ep400 binding at promoters is highly linked to gene expression level. By contrast, Brg1, Chd4 and Chd6 deposition showed little correlation with gene expression level (statistical test failed for at least one comparison for these remodellers). While 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 data sets used in Figs 1 and 2 and Extended Data Figs 1, 4 and 7. 

Brg1 (ref. 35): GSM359413; DNase-seq: GSM1014154; Ezh2 (ref. 15): GSM590132; GRO-seq10: GSM665994; H2A.Z:12,13; GSM958501, DRP001103; H3.3 (ref. 12): GSM1386359; H3K27ac (ref. 16): GSM594578; H3K27me3 (ref. 15): GSM590115; H3K36me3 (ref. 15): GSM590119; H3K4me3 (ref. 15): GSM590111; Med1 (ref. 36): GSM560347; Mi2b (Chd4): GSM687284; MNase-seq7: GSM1004653; Oct4, Sox2, Nanog: GSM1082340; Pol II S5ph (ref. 37): GSM515662; TBP: GSM958503.

21. Liu, P., Jenkins, N. A. & Copeland, N. G. A highly efficient recombineering-based method for generating conditional knockout mutations. Genome Res. 13, 476–484 (2003).
22. Ying, Q. L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nature Biotechnol. 21, 183–186 (2003).
23. Tessarollo, L. Manipulating mouse embryonic stem cells. Methods Mol. Biol. 158, 47–63 (2001).
24. Pruitt, K. D. et al. RefSeq: an update on mammalian reference sequences. Nucleic Acids Res. 42, D756–D763 (2014).
25. Ye, T. et al. seqMINER: an integrated ChIP-seq data interpretation platform. Nucleic Acids Res. 39, e35 (2011).
26. Albert, I., Wachi, S., Jiang, C. & Pugh, B. F. GeneTrack—a genomic data processing and visualization framework. Bioinformatics 24, 1305–1306 (2008).
27. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP-seq enrichment using MACS. Nature Protocols 7, 1728–1740 (2012).
28. Berlivet, S., Houard, M. & Gérard, M. Loss-of-function studies in mouse embryonic stem cells using the pHYPER shRNA plasmid vector. Methods Mol. Biol. 650, 85–100 (2010).
29. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods 5, 621–628 (2008).
30. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25 (2009).
31. Giresi, P. G. & Lieb, J. D. Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements). Methods 48, 233–239 (2009).
32. Carière, L. et al. Genomic binding of Pol III transcription machinery and relationship with TFIIIS transcription factor distribution in mouse embryonic stem cells. Nucleic Acids Res. 40, 270–283 (2012).
33. Rhee, H. S. & Pugh, B. F. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell 147, 1408–1419 (2011).
34. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009).
35. 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. USA 106, 5187–5191 (2009).
36. Kagey, M. H. et al. Mediator and cohesin connect gene expression and chromatin architecture. Nature 467, 430–435 (2010).
37. Rahl, P. B. et al. c-Myc regulates transcriptional pause release. Cell 141, 432–445 (2010).
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 haemagglutinin (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. Immunopurification 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 | Relationship between remodeler enrichment at promoters and RNA expression level. Average binding profile of remodelers 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 remodelers to bivalent promoters, which are transcribed at lower levels, showed a significant reduction compared to H3K4me3 promoters. By 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 panels 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 (data set 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 Fig. 1.
Extended Data Figure 7 | Validation of remodeller 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. Shown on the right of each panel are the results of RT–qPCR analysis that quantify RNA expression levels of the corresponding genes upon remodeller depletion in ES cells. Two distinct shRNA vectors (shRNA 1 and shRNA 2, see Methods) were used for each remodeller. Scores on the y axis indicate the relative expression of the indicated genes compared to reference genes. Values are mean and s.d. of three independent transfection experiments.
Extended Data Figure 8 | Effect of remodeller depletion on chromatin accessibility at promoters. Consequence of remodeller depletion by shRNA vectors on ATAC-seq average profiles at all H3K4me3-only (top) and bivalent (bottom) promoters. Two replicate experiments are shown on each graph for both remodeller 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 after indicated remodeller knockdown (colour) at H3K4me3-only (left) and bivalent (right) genes. Left and right panels within a set represent the set of genes that are most upregulated (red) or downregulated (green) after remodeller knockdown. Pol II occupancy is indicated within a window spanning 500 and 2,000 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 after remodeller knockdown relative to control, measured by ChIP-exo, at genes either downregulated (green) or upregulated (red) after the depletion of the indicated remodeller. c. Pol II distribution of remodeller 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 remodelers contribute to Pol II recruitment.