Supplemental Material

Supplemental Figure legends

**Figure S1. Nup153 deficiency causes the differentiation of mESCs.** (A) Relative mRNA levels measured by qRT-PCR of Nup153 after 6, 9 and 13 days Nup153 knockdown. The relative expression levels are normalized to actin and expressed as fold change relative to the respective control (shCTRL). (B) Western blot analysis for Nup153 and Nup107 in control (empty construct) and FLAG.mCherry.Nup153 (rescue construct) expressing mESCs. Tubulin was used as the loading control. (C) Immunofluorescence for Cherry and FLAG was performed to confirm the localization of the FLAG.mcherry.Nup153 rescue construct at the nuclear envelope. Scale bars, 10µm. (D) Quantification of Nup153 protein levels accessed by western blot in control, Nup153-knockdown and FLAG.mCherry.Nup153 expressing cells using scanning and ImageJ software. Results are expressed as integrated optical density. Each sample was normalized to tubulin content. Each bar represents the mean ± SEM of duplicate values. p values were obtained from Student's t-test (* p<0.05) (E) Western blot analysis for Nup153 and Parp1 in whole cell extracts from mESCs after 3 and 6 days upon Nup153 depletion. A 75 KDa band corresponding to Parp1 cleavage (apoptotic marker) was observed after a 24h Puro selection due to the presence of uninfected cells. Tubulin was used as the loading control.

**Figure S2. Nup153 knockdown causes the upregulation of early ectodermal markers.** (A) RNA-seq scatter plot showing the upregulated and downregulated genes upon differentiation of mESCs into neural precursor (NeuP) cells. Two independent differentiated NeuP cell lines were analyzed, merged and compared to the control. Arrows indicate representative ectodermal markers upregulated in both Nup153-
knockdown mESCs and NeuP cells. (B) Venn diagram showing the overlap between the induced genes in Nup153-knockdown mESCs and differentiated NeuP cells. p value for the overlap computed using the Fisher’s exact test. (C) Western blot analysis for Tuj1 in whole cell extracts from mESCs after 6 days Nup153 knockdown. Gapdh was used as the loading control.

**Figure S3.** Nup153 deficiency impairs the proper differentiation of mESCs

(A) qRT-PCR analysis showing the relative expression levels of representative mesodermal markers in control and Nup153-knockdown embryoid bodies (EBS) at day 0, 8 and 12 of EB differentiation. The relative expression levels are normalized to actin and expressed as fold change to the respective control. The mean ± SD from three independent experiments are shown. p values were obtained from Student’s t-test (* p<0.05). (B) Immunofluorescence for the mesododermal marker α-Sma in control and Nup153-depleted EBs at day 12 of EB differentiation. Scale bars, 200µm. (C) Immunofluorescence for the endodermal marker Foxa2 in control and Nup153-knockdown EBs plated in matrigel at day 12 of EB differentiation. Scale bars, 50µm. (D) Representative immunohistochemical (ICH) data for Tuj1 protein expression on paraffin-embedded sections of teratomas originated from control and Nup153-depleted mESCs. The teratoma sections were immunostained, applying a Tuj1 antibody followed by detection using an HRP-based method with DAB colorimetric substrate (brown). Nuclei were counterstained with hematoxylin (blue). Original Magnification, 400X.

**Figure S4. Global nuclear transport is not affected in Nup153-knockdown mESCs**

(A) Immunofluorescence for the pluripotency marker Sox2 in control and Nup153-knockdown mESCs after 6 days Nup153 depletion. Scale bars, 10µm. (B) Relative mRNA levels measured by qRT-PCR of Nup153 and Nup50 in control, Nup153-
knockdown and Nup50-knockdown mESCs upon 5 days knockdown. The relative expression levels are normalized to actin and expressed as fold change relative to the respective control (shCTRL). (C) qRT-PCR analysis showing the relative expression of Nup153 and Nup107 in control, Nup153-knockdown and Nup107-knockdown mESCs at different time points upon shRNA infection. The relative expression levels are normalized to actin and expressed as fold change relative to the respective control. The mean ± SD from three independent experiments are shown. p values were obtained from Student's t-test (** p<0.01; *** p<0.001). (D) Quantification of the total number of individual pores per cell from control and Nup107 knockdown nuclei. Individual pores were stained using an antibody against Nup153. p value was obtained from Student's t-test (*** p<0.001). (E) AP staining and total number of positive AP stem cell colonies (right graph) in control, Nup153-knockdown and Nup50-knockdown mESC colonies. Data are represented as mean ± SD; n=3. p values were obtained from Student's t-test (*** p<0.001).

Figure S5. Validation of the DamID-seq approach in mESCs. (A) Immunofluorescence for the V5-tag in stable mESCs expressing V5-Dam-Nup153 or V5-Dam-Oct4 fusion proteins after heat-shock induced expression. Scale bars, 10µm. (B) Representative example of a pluripotency gene containing an Oct4 DamID-seq peak around its TSS. Two independent biological replicates are shown for both Dam-Oct4 and Dam-GFP. Oct4 peaks obtained by ChiP-seq are shown. Schematic illustration of the genomic structure of the gene loci, its genomic location and the associated GATC content are also shown. (C) Genome-wide occupancy of Oct4 DamID-seq peaks at all identified Oct4 ChiP-seq peaks. (E) Nup153 DamID-seq results at one representative differentiation gene (Nestin). Two independent biological replicates for both Dam-Nup153 and Dam-GFP are shown. (F) Fraction of genes containing a Nup153 binding
site within -2Kb from TSS to +1Kb from TES that are upregulated after Nup153 knockdown versus non-regulated. (G) Gene Ontology (GO) analysis of the genes containing a Nup153 binding site within -2Kb from TSS to +1Kb from TES. (H) Representative IF-FISH image showing the localization of a "non-peripheral" Nup153 target genes (Nes) in control and Nup153-depleted mESCs after 6 days knockdown. FISH signals are shown in red, Nup153 staining in green, Lmnb2 staining in gray and Hoechst in blue. (I) Percentage of FISH signals from several Nup153 target genes localized at 0-0.5µm (blue), 0.5-2µm (yellow) or 2-6µm (red) from the nuclear envelope (stained with Lmnb2) in control and Nup153-depleted mESCs after 6 days knockdown.

**Figure S6. Nup153 and PRC1 interact and co-occupy common chromatin sites.** (A) Normalized Ring1b read count +/- 1Kb from the transcription start site (TSS) in Nup153 bound TSSs versus unbound TSSs. p value was calculated using the Mann-Whitney test. (B) Distribution of Ring1b ChiP-seq peaks relative to RefSeq genes. (C) Gene Ontology (GO) analysis of the genes containing a Ring1b ChIP-seq peak within -2Kb from the transcription start site (TSS) to +1Kb from termination end site (TES). (D) Anti-FLAG immunoprecipitation (IP) of FLAG.GFP, FLAG.mCherry.Nup153 and FLAG.GFP.Nup50 ectopically expressed in mESCs, followed by western blot using antibodies against Nup50 (positive control) and Foxo1 (negative control) proteins. (E) Co-Immunoprecipitations (co-IP) of endogenous Nup153, Cbx7 and Nup50 proteins in mESCs, followed by western blot using antibodies against Nup153 and Cbx7. (F) Co-IP of endogenous Nup153, Ring1b, Mel-18 (an additional component of the PRC1) and Nup50 proteins in mESCs, followed by western blot using an antibody against Ring1b.

**Figure S7. Nup153 deficiency impairs the recruitment of PRC1 to the TSS of several developmental genes.** (A) Genome-wide occupancy of Ring1b at all annotated
TSSs in control, Nup153-knockdown mESCs and NeuP cells. (B) Genome-wide occupancy of Oct4 at all identified Nup153 DamID-seq peaks in control and Nup153-knockdown mESCs. (C) Nup153 DamID-seq, Ring1b ChiP-seq and RNA-seq profiles at the Gata6 genomic loci in control, Nup153-knockdown mESCs and NeuP cells are shown. For visualization, the biological replicates used to generate the Nup153 and GFP DamID profiles were merged. The genomic location of the gene loci and the associated GATC content are also shown. The y-axis in all Ring1b ChIP-seq and RNA-seq profiles shown represents the normalized number of reads. (D) Immunofluorescence for two different components (Cbx7 and Rybp) of the PRC1 in control and Nup153-knockdown mESCs after 6 days Nup153 depletion. Scale bars, 10µm. (E) Western blot analysis of protein lysates prepared from mESCs and NeuP cells using antibodies specific for the indicated proteins. Tubulin served as a loading control for total protein.

**Extended materials and methods**

**RNA-seq and data analysis**

RNA was extracted from Nup153-knockdown and control mESCs using Trizol (Ambion) and column purified using RNeasy kit (Qiagen). RNA-Seq strand-specific libraries were generated from total RNA with polyA+ selection of mRNA using the TruSeq Stranded RNA LT Kit (Illumina). Libraries were sequenced on the Illumina HiSeq 2500 according to the manufacturer’s instructions. Reads were aligned to the mouse genome (mm9, NCBI37) using STAR (version 2.2.0.c) (Dobin et al. 2013). Only reads that aligned uniquely to a single genomic location were used for downstream analysis. Gene expression values were calculated for annotated RefSeq genes using HOMER by counting reads found overlapping exons ((Heinz et al. 2010)). Differentially expressed genes were found using EdgeR (Robinson et al. 2010). Gene Ontology functional enrichment analysis was performed using DAVID (Dennis et al. 2003).
ChIP-seq and data analysis

ChIP was performed as previously described (Liang et al. 2013). Briefly, for ChIP-seq, approximately 10^7 cells were cross-linked with 1% formaldehyde at room temperature for 10 min and neutralized with 0.125 M glycine. After sonication, 20–30 µg of soluble chromatin was incubated with 5–10 µg of antibody at 4°C overnight. Immunoprecipitated complexes were collected using Dynabeads M280 sheep-anti-rabbit IgG or sheep-anti-mouse IgG (Invitrogen). Subsequently, immuno-complexes were washed, and DNA was extracted and purified by QIAquick Spin columns (Qiagen). DNA libraries were generated using the TruSeq ChIP sample Prep kit (Illumina) followed by deep sequencing with the Illumina’s HiSeq 2500 system according to the manufacturer’s instructions. Reads were aligned to the mouse genome (mm9, NCBI37) using Bowtie2 (version 2.2.3) (Langmead and Salzberg 2012). Only reads that aligned uniquely to a single genomic location were used for downstream analysis. ChIP-Seq peaks were identified using HOMER using an FDR or 0.1% and fold-enrichment over input of at least 4-fold (Heinz et al. 2010). Oct4 ChIP-Seq peaks were found in ‘factor’ mode, which looks for focal peaks, while Ring1B and Cbx7 ChIP-Seq peaks were found in “histone” mode since their enrichment patterns covered variable-sized domains. Normalized bedGraph files were created for each experiment using HOMER and visualized using the UCSC Genome Browser (Kent et al. 2002). Comparisons between ChIP-Seq peaks, DamID-Seq peaks, and other genomics features such as TSS were performed using the ‘mergePeaks’ program in HOMER. De novo motif discovery was carried out using HOMER.

DamID-seq
The heat shock promoter, ecodam tag and RFC.1 region of the pLgw EcoDam-V5-RFC1 vector (a kind gift of Dr. Bas Van Steensel) was cloned into the polylinker of the pMSCV-puro vector yielding pMSCV-ED-puro. All ORFs (GFP, Oct4 and Nup153) were PCR amplified with gateway compatible sites and cloned into the pDonr207 vector. The ORFs were transferred to the pMSCV-ED-puro vector by LR reaction. Around $5 \times 10^6$ mESCs per 10cm plate were infected with Dam-fusion proteins lentivirus at a multiplicity of infection (MOI) of 50. At 48 hours after transduction, puromycin (2µg/ml) was added to the ES media to select transduced mESCs. After a one-week selection, mECS were harvested and gDNA was isolated using the Qiagen DNeasy Blood & Tissue Kit. 3µg of gDNA was then digested with 0.5µl of DpnI (NEB, 20 U/ µl) and ligated with the adaptor AdR, which was made by mixing and slowly annealing AdR-top (5’ CTAATACGACTCACTATAGG GCAGCGTGGTCGCGGCCGAGGA 3’) and AdR-bottom (5’ TCCTCGGCCG 3’). This ligation was completed using 1uL T4 Ligase (Roche, 5U/µl) for 2 hours at 16°C. The regions flanked by adaptors were amplified by PCR using the Bio-Adr-primer (5’ Bio-GTTCGCGCGAGGATC 3’) and purified using the Qiagen MinElute PCR purification kit. 4µg of amplified products were sonicated using a Bioruptor (Diagenode) to obtain DNA fragments of sizes between 100-500bp. Dynabeads MyOne Streptavidin T1 (Invitrogen) beads were used to pull down the biotinylated ends of the sonicated PCR products. The bound DNA was removed off the beads by digestion with DpnII at 37°C for 1 hour and cleaned using the MinElute PCR Purification kit. DamID libraries were generated using the TruSeq ChIP sample Prep kit (Illumina) followed by deep sequencing with the Illumina’s HiSeq 2500 system according to the manufacturer’s instructions. When sequencing DamID-Seq libraries, 4 “dark cycles” were performed at the beginning of sequencing such that the actual sequencing data would start after the expected GATC at the beginning of each read. The greatly increased the performance...
of the HiSeq 2500 since it expects relatively even nucleotide distribution at the beginning of a sequencing run.

**Identification of DamID-seq peaks and data analysis**

Reads were aligned to the mouse genome (mm9, NCBI37) using Bowtie2 (version 2.2.3) (Langmead and Salzberg 2012). Only reads that aligned uniquely to a single genomic location were used for downstream analysis. In addition, only reads aligning just downstream of a GATC sequence in the genome were kept for further analysis. Reads aligning to non-GATC locations were much more likely to non-specific noise. Putative DamID-Seq peaks were identified using HOMER in a manner similar to variable-size ChIP-Seq peaks, with several modifications. First, duplicate reads, which are normally discarded to avoid artifacts in ChIP-Seq, were retained for DamID-Seq analysis since most of the reads align to a limited number of GATC sites. Second, the initial peak size used for enrichment calculations was set to 2500 bp based on the enrichment pattern of Oct4 DamID-Seq reads near Oct4 ChIP-Seq peaks. Third, DamID-Seq peaks regions were required to have 2-fold more normalized reads than GFP DamID-Seq controls. In addition, peaks were required to contain at least 50 normalized reads per peak (per 10 million reads sequenced) to remove low magnitude sites. Finally, duplicate target DamID-Seq experiments and GFP DamID-Seq experiments were analyzed at putative peaks with edgeR to identify target peaks that were statistically enriched relative to GFP controls (FDR < 0.1%, Fold-change > 2). Normalized bedGraph files were created for DamID by extending reads 1 kb both upstream and downstream to reflect the relative size of DamID enriched regions. Comparisons between ChIP-Seq peaks, DamID-Seq peaks, and other genomics features such as TSS were performed using the `mergePeaks` program in HOMER. De novo motif discovery, META-gene enrichment
profiles, and annotation to nearest or overlapping RefSeq genes were carried out using HOMER.

**Supplemental Reference**

Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 4: P3.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15–21.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell* 38: 576–589.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Research* 12: 996–1006.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Meth* 9: 357–359.

Liang Y, Franks TM, Marchetto MC, Gage FH, Hetzer MW. 2013. Dynamic association of NUP98 with the human genome. ed. J. Brickner. *PLoS Genetics* 9: e1003308.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–140.