Supplemental Materials

The Pol II Pre-Initiation Complex (PIC) Influences Mediator Binding but Not Promoter-Enhancer Looping

Fei Sun, Terrence Sun, Michael Kronenberg, Xianglong Tan, Chengyang Huang, and Michael F. Carey

Supplemental Tables S1-S6 (see attached Excel files)

Supplemental Table S1. List of dysregulated genes in each KD.

Supplemental Table S2. Antibodies used in this study.

Supplemental Table S3. shRNA sequences.

Supplemental Table S4. qPCR primers.

Supplemental Table S5. Oligos used in degron cell line constructions.

Supplemental Table S6. Pearson’s correlation coefficients between biological replicates.

Supplemental Methods (see below)

Supplemental Figures S1-S5 (see below)
Supplemental Methods

Cell culture
Wild type V6.5 and degron cell lines were cultured on a MEF feeder layer (Invitrogen) in media containing DMEM-KO supplemented with 15% fetal bovine serum, 1000 U/ml LIF, 100 μM nonessential amino acids, 2 mM L-gultamine and 8 nL/ml of 2-mercaptoethanol at 37°C in a humidified 5% CO₂/95% air incubator.

Degron cell line construction
To construct the parental cell line constitutively expressing V5-tagged OsTir1, V6.5 cells grown in a 24-well plate were transfected with 1 µg each of pEN396 (Addgene #92142) and pX330-EN1201 (Addgene #92144) plasmid using Lipofectamine 2000. 6 hours post transfection, cells were passed into a 6-cm dish with DR4 MEF feeders (Thermo Fisher Scientific, cat#A34966). Selection started 2 days later by adding 1 µg/ml puromycin. Survival colonies were picked, expanded and verified by genotyping. Expression of OsTir1 was validated by Western blotting. This parental cell line was used for the construction of both TAF12 and MED4 degron cell lines.

Construction of TAF12 or MED4 degron cell lines required a CRISPR plasmid and a donor plasmid. To construct the CRISPR plasmid, a guide RNA sequence was designed targeting a locus adjacent to the stop codon of the target gene and the corresponding DNA sequence was cloned into pX330 (Addgene #42230). To generate the donor plasmid, 1) two homology arms of ~500 bp corresponding to upstream and downstream of the stop codon were amplified from the mouse genome by PCR; 2) a DNA fragment of the sequence encoding mini-AID and hygromycin-resistant gene (AID&Hyg) was amplified from plasmid pMK287 (Addgene cat#72825); 3) the homology arms and the AID&Hyg fragment were fused into a single fragment by overlapping PCR; 4) the fragment was cloned into pMD20 (Clontech cat#3270) with TA clone. The parental cells were transfected with 1 µg each of the CRISPR plasmid and the donor plasmid. Protocols for cell transfection, selection and colonies picking were the same way as above except 150 µg/ml hygromycin was used for selection. Homozygotes verified by genotyping were expanded and further confirmed by Western blotting.
**Nuclear extraction**

Cells were harvested and washed by cold DPBS. The cell pellet volume was determined. Cells were dounced 10 times in 5 volumes of Buffer A (10 mM HEPES-KOH pH 7.9, 10 mM KCl, 1.5 mM MgCl₂, 0.5 mM DTT, 0.5 mM PMSF). Nuclei were harvested, resuspended and dounced in Buffer C (0.2 mM EDTA, 25% glycerol, 20 mM HEPES-KOH pH 7.9, 420 mM NaCl, 1.5 mM MgCl₂, 0.5 mM DTT, 0.5 mM PMSF) followed by 1 h gentle stirring. After centrifugation, the supernatant was collected and saved without dialysis.

**Co-immunoprecipitation**

100 µl of nuclear extract was thawed and diluted with 450 µl of 1×PBS. After centrifugation, 450 µl of the supernatant was collected and incubated overnight at 4°C with 2-5 µg of antibody or IgG. The remaining supernatant was saved as input. Protein-antibody conjugates were precipitated using Dynabeads Protein A/G. Beads were washed 3 times with 1×PBS before proteins were eluted with 1×SDS loading buffer.

**mRNA-seq**

mRNA-seq libraries were generated from total RNA samples extracted with TRIzol Reagent using KAPA Stranded mRNA-Seq Kits following manufacturer’s instructions and sequenced on Illumina HiSeq 4000 using single-end 50 bp mode. All experiments were performed with 2 biological replicates. Sequence reads were mapped to mm9 using TopHat2 with option –g 1 –N 2. Gene transcription levels were normalized to RPKM using SAMMate with the mode of Read Assignment Expectation Maximization (RAEM). Pearson’s correlation coefficients were calculated between biological replicates to confirm the reproducibility. Data from one replicate were plotted in the figures. Fold change of a gene upon a knockdown was calculated by dividing RPKM of the control by that of the knockdown. To calculate of the percentages in Fig. 1C, for TFIID KD, the numbers of downregulated genes from all TFIID subunit KDs in Fig. 1B were summed and the sum was defined as N\text{down}. The numbers of upregulated genes from all TFIID subunit KDs were summed and the sum was defined as N\text{up}. The percentage of downregulated genes in all dysregulated ones was calculated using this formula: \%\text{down} = N\text{down}/(N\text{down}+N\text{up}). The percentage of upregulated genes was calculated using the formula \%\text{up} = N\text{up}/(N\text{down}+N\text{up}). For Mediator KD, the percentages were calculated identically.
**Nascent RNA-seq**

Mouse ESCs were harvested at ~80% confluency and washed with DPBS. Cells were counted and 10 million cells were collected for subsequent steps. Half a million human BEAS-2B cells were mixed into the 10 million mouse ESCs as the spike-in. Cells were lysed in 200 μL of ice-cold lysis buffer (10 mM Tris-HCl pH 7.5, 0.1% NP40, 150 mM NaCl). The cell lysate was gently layered over 500 μL of a chilled sucrose cushion (24% RNase-free sucrose in lysis buffer) in a new Eppendorf tube and centrifuged for 10 min at 4°C at 10,000xg. The supernatant (cytoplasmic fraction) was removed and the pellet (nuclei) was washed once with 200 μL of ice-cold 1×PBS/1 mM EDTA. The nuclear pellet was resuspended in 100 μL of pre-chilled glycerol buffer (20 mM Tris-HCl pH 7.9, 75 mM NaCl, 0.5 mM EDTA, 0.85 mM DTT, 0.125 mM PMSF, 50% glycerol) by gentle flicking of the tube. An equal volume (100 μl) of cold nuclear lysis buffer (10 mM HEPES pH 7.6, 1 mM DTT, 7.5 mM MgCl₂, 0.2 mM EDTA, 0.3 M NaCl, 1 M urea, 1% NP-40) was added. The mix was vortexed vigorously for 2 s. The sample was incubated for 2 min on ice, and then centrifuged for 2 min at 4°C at 10,000xg. The supernatant (nuclear fraction/nucleoplasm) was removed and the pellet (chromatin) was gently rinsed with ice-cold 1×PBS/1 mM EDTA. 1 mL of Trizol reagent was added to the chromatin and incubated for 30 min at 50°C to dissolve it. Nascent RNA was extracted following the manufacturer’s instructions. Libraries were constructed using KAPA Stranded RNA-Seq Kit with RiboErase (HMR). Single-end 50-bp sequencing by Illumina HiSeq 4000 was used. All experiments were performed with 2 biological replicates. Sequence reads were first mapped to mm9 as above. Reads within the entire gene body (including all exons and introns) were counted with a custom script. An RPK (reads per kilobase) value was calculated for each gene by dividing the read counts by the gene body size. Reads derived from the human cell spike-ins were counted by mapping sequence reads to hg19. RPK values were normalized to the spike-ins when comparing between different samples. Pearson’s correlation coefficients were calculated between biological replicates to confirm the reproducibility. Data of one replicate were plotted in the figures.

**ChIP-seq**

Cells were harvested at 70%–80% confluency by trypsinization. After washing with DPBS, cells were formaldehyde-crosslinked to a final concentration of 1% for 10 min at room temperature,
followed by 5 min quenching with 125 mM glycine. Cells were washed twice with cold DPBS. If not used immediately, the cell pellet was flash frozen in liquid nitrogen and stored at −80°C. Crosslinked cells were resuspended in swelling buffer (25 mM HEPES pH 7.9, 1.5 mM MgCl2, 10 mM KCl, 0.1% NP-40, and 1×Complete protease inhibitor cocktail) and incubated for 10 min on ice. Cells were centrifuged and the cell pellet was resuspended in Buffer A (50 mM HEPES-KOH pH 7.9, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS) plus 1×Complete protease inhibitor cocktail. Cell sonication was performed on a Qsonica Q800R2 sonicator with 20% amplitude for 20 cycles at 10 s each with 30 s between cycles at 4°C. Sonicated lysate was pre-cleared by incubating with Dynabeads Protein A/G. Part of the pre-cleared lysate was used as input and the remainder was incubated overnight at 4°C with 2-10 μg of antibody. DNA/Protein-antibody conjugates were precipitated using Dynabeads Protein A/G. Beads were washed twice each with Buffer A, Buffer B (50 mM HEPES-KOH pH 7.9, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS), LiCl buffer (20 mM Tris-HCl pH 8.0, 250 mM LiCl, 1 mM EDTA, 0.5% Na-deoxycholate, 0.5% NP-40) and TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). DNA was eluted in elution buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1% SDS). Cross-links were reversed overnight at 65°C. RNA and protein were digested using RNase A and Proteinase K, respectively, and DNA was purified by phenol-chloroform extraction and ethanol precipitation. Libraries were prepared using KAPA HyperPrep Kit and sequenced using the Illumina HiSeq 4000 platform for 50 bp single end reads. All experiments were performed with 2 biological replicates. Sequence reads were mapped to mm9 using Bowtie2 with the --very-sensitive option. Clonal reads were removed with a custom script. Pearson’s correlation coefficients were calculated between biological replicates to confirm the reproducibility. Data of the 2 replicates were pooled before downstream analyses. To call peaks, the mouse genome was segmented into 50-bp windows. A Poisson p value was calculated for each window based on the IP and normalized input counts in this window. A significant peak was retained only when p values of its own window and the 2 neighboring windows were all below 0.001. Sitepro from the CEAS package was used to determine the genome-wide binding profile of different proteins at TSSs and enhancers. The TSS bed file was derived from the mm9 refGene data downloaded from UCSC Genome Browser but only the promoters of active genes (norm. RPK > 10 from the nascent RNA-seq data) were plotted in the heatmaps. For enhancers, the bed
file was made from data of Whyte et al., 2013. Heatmaps were made with JavaTreeview. Genome browser tracks were made using IGB.

**Promoter capture Hi-C**

Five million cells were harvested and formaldehyde-crosslinked as for ChIP-seq. Crosslinked cells were lysed for 15 min on ice in 10 mM Tris-HCl pH 8.0, 10 mM NaCl, and 0.2% NP-40 supplemented with cOmplete protease inhibitor. Nuclei were isolated by centrifugation and by removing the supernatant. Nuclei were gently resuspended in 50 μL of 0.5% SDS and incubated at 62°C for 5-10 min. SDS was quenched by adding 170 μl of 1.5% Triton X-100 and incubating at 37°C for 15 min. 25 μl of 10×NEBuffer2 and 100 U of MboI restriction enzyme were added and the sample was digested overnight on a thermo-mixer. MboI was inactivated by heating to 62°C for 20 min. After the sample was cooled down to room temperature, a fill-in master mix (37.5 μl of 0.4 mM biotin-14-dATP, 1.5 μl of 10 mM dCTP, 1.5 μl of 10 mM dGTP, 1.5 μl of 10 mM dCTP, 8 μl of Klenow Fragment) was added to the sample and incubated at 37°C for 1 hour. Proximity ligation was carried out by adding a ligation master mix (663 μl of H₂O, 120 μl of 10X NEB T4 DNA ligase buffer, 100 μl of 10% Triton X-100, 12 μl of 10 mg/ml BSA, 5 μl of T4 DNA ligase) and incubating for 4 hours at room temperature. The sample was treated with Proteinase K before DNA purification with ethanol precipitation. Purified DNA was sheared on a Qsonica Q800R2 sonicator with 20% amplitude for 9 cycles at 10 s each with 30 s between cycles at 4°C. Size-selection of the sheared DNA was performed with KAPA pure beads to enrich fragments in the range of 300-500 bp. Size-selected biotinylated DNA was pulled down with Dynabeads MyOne Streptavidin T1 beads. After on-bead end repair and A-tailing, SureSelect Adaptor was added to the ends of DNA following Agilent’s protocol. The adaptor-ligated DNA was amplified directly off the beads with 6-8 PCR cycles. PCR products were purified and concentrated to 221 ng/μl using a vacuum concentrator. For capture, two 120-nt biotinylated RNA probes were designed to anneal specificity to the promoter region of one gene. In total, ~40,000 RNA probes targeting ~20,000 mouse genes were synthesized at Agilent. 3.4 μl (750 ng) of the concentrated PCR products were hybridized to the RNA probes following Agilent’s protocol. The hybridized library was captured by Dynabeads MyOne Streptavidin T1 beads and amplified off the beads with Agilent’s indexing primers. The library was purified and sequenced on NovaSeq 6000 S4 with 2×150bp mode. After analyzing the data, we found different RNA probes had very different
performances in capturing their target promoters. To increase the capture efficiency of the RNA probe pool, the probes with low capture ability were discarded from the pool. A second batch of PCHi-C assays were carried out using the new pool of ~4,000 probes targeting ~2,000 genes. All experiments were performed with 2 biological replicates. Pearson’s correlation coefficients were calculated between biological replicates to confirm the reproducibility. Data of the replicates in both batches of the same sample were pooled before downstream analyses.
Supplemental Fig. S1. Verification of shRNA KD efficiency and specificity. (A) qPCR test of different shRNAs for efficiency of KD of each TFIID/Mediator subunit. Bar graphs show the average of 3 technical replicates. The shRNA generating the most efficient KD is highlighted. (B) Heatmap showing log2 ratio of RPKM(KD/control) in subunit mRNAs for each KD. X axis, gene of each subunit. Y axis, KD experiment. mRNA-seq data from the shRNA screen are plotted here. (C) Heatmap showing the significance of all pairwise overlaps based on the upregulated genes. Related to Fig. 1.
Supplemental Fig S2

Comparison between WT V6.5 and degron experiments and between KD and degron experiments. (A) Heatmaps of TAF12 ChIP-seq signal in WT V6.5 and TAF12 degron cells (no Auxin treatment). Color scales indicate significance (-log_{10}P). (B) Scatter plot of nascent RNA levels of each gene in WT V6.5 and TAF12 degron cells (no Auxin treatment). (C) Browser track examples showing TAF12 ChIP-seq and nascent RNA-seq data in WT V6.5 versus TAF12 degron. (D) Heatmaps of MED4 ChIP-seq signal in WT V6.5 and MED4 degron cells (no Auxin treatment). (E) Scatter plot of nascent RNA levels of each gene in WT V6.5 and MED4 degron cells (no Auxin treatment). (F) Browser track examples showing MED4 ChIP-seq and nascent RNA-seq data in WT V6.5 versus MED4 degron. (G) Overlap of downregulated genes in KD and degron. (H) Percentage of down- and up-regulated genes in dysregulated genes in degron or KD. (I) GO analysis on upregulated genes in TAF12 KD and MED4 KD.
degron cells (no Auxin treatment). (D-F) Same as (A-C) but comparing WT V6.5 with MED4 degron cells. (G) Venn diagrams showing overlap of downregulated genes upon KD and degradation. P-values were calculated based on a hypergeometric distribution. (H) Pie charts showing percentages of down-regulated and up-regulated genes in KD and degron experiments. (I) Gene Ontology analyses on upregulated genes in TAF12 KD and MED4 KD. Related to Fig. 2.
Supplemental Fig. S3. KD of GCN5 or USP22 has limited effects on transcription. (A) Western blotting showing protein levels of TBP, TAF3, TAF4, TAF4B, TAF12, TAF5, TAF1 and Pol II in the input and TAF1 IP samples of the co-IP experiments in TAF12 degron cell extracts. IgG was used as a negative control for the IP. Three independent replicates are shown. (B) Heatmaps showing binding patterns of TAF12 and MED4 at promoters and enhancers in WT V6.5 cells. H3K4me3, H3K27ac and P300 (Chronis et al 2017) are shown to indicate promoters and enhancers.
Promoters were sorted by the mRNA expression level of the genes. (C) qPCR test of different shRNAs in KD of GCN5 and USP22. Bar graphs show the average of 3 technical replicates. The shRNA giving the most efficient KD is highlighted. The sample treated with the most effective shRNA was used for mRNA-seq. (D) Heatmap showing specificity of shRNAs for GCN5 KD and USP22 KD. (E) Bar plots summarizing numbers of genes down-regulated and up-regulated by each KD. (F) Heatmap showing significance of pairwise overlap between each KD based on the down-regulated genes. (G) Heatmaps showing binding of MED1 and ESRRB at super-enhancers (SE) and typical enhancers (TE) in TAF12 degron cells. (H) Average profiles of MED1 and ESRRB binding at SE and TE. Related to Fig. 3.
Supplemental Fig. S4. Western blotting levels of MED6, MED18, MED1, MED4, MED9, MED10, MED23, MED24, MED12 and Pol II in the input and MED6 IP samples of the co-IP experiments. IgG was used as a negative control for the IP. Three independent replicates are shown. Related to Fig. 4.
**Supplemental Fig. S5.** Enrichment of promoter-interacting regions (PIRs) with different chromatin states and additional browser tracks of examples showing impact of TAF12 or MED4 degradation on promoter-interacting chromatin looping. (A) Pie chart showing percentages of PIRs related to each chromatin state. (B) Pie chart showing percentages of each chromatin state in the genome. (C) Fold enrichment of PIRs at different chromatin states. (D,E) Browser tracks of examples showing impact of TAF12 (D) or MED4 (E) degradation on promoter-interacting looping. Data of ChromHMM, TAF12/MED4 ChIP-seq, virtual 4C and nascent RNA-seq are shown. Related to Fig. 5.