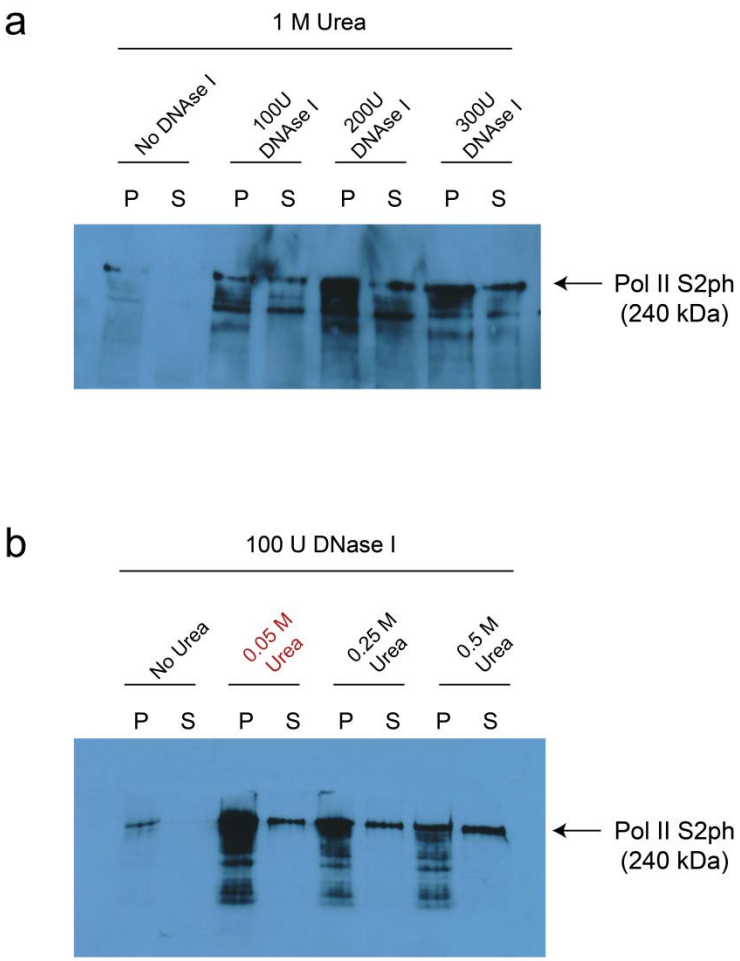
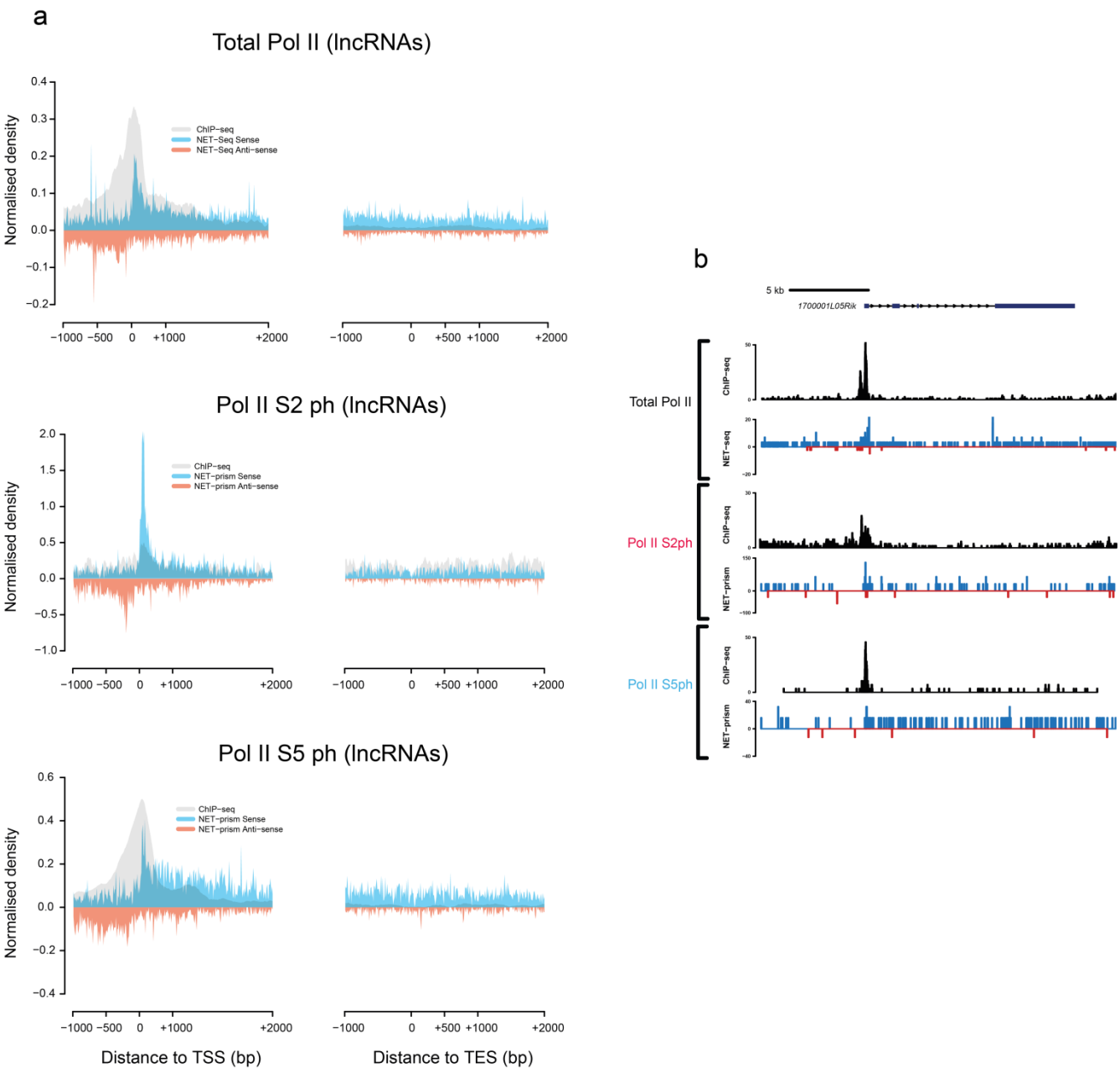
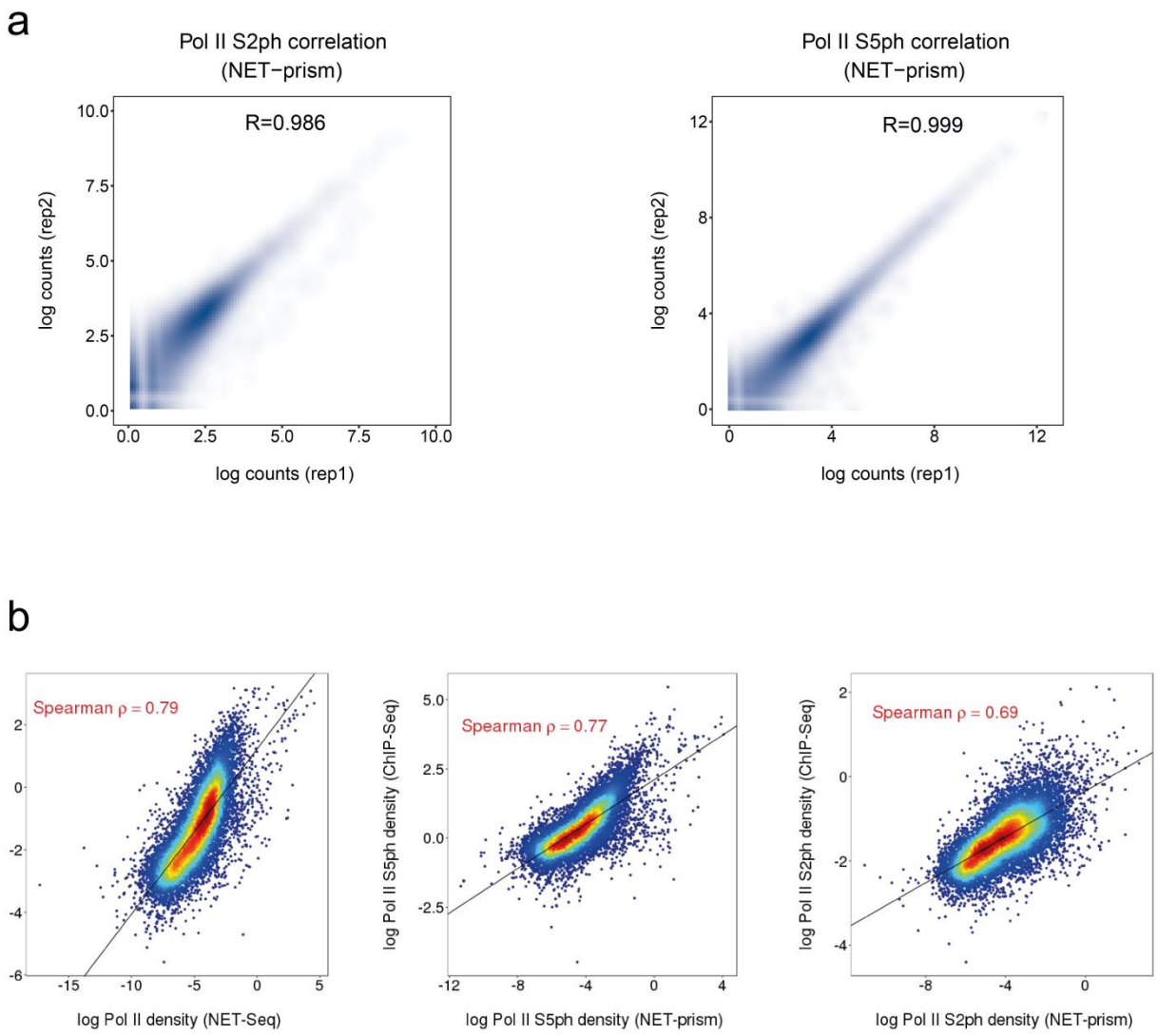


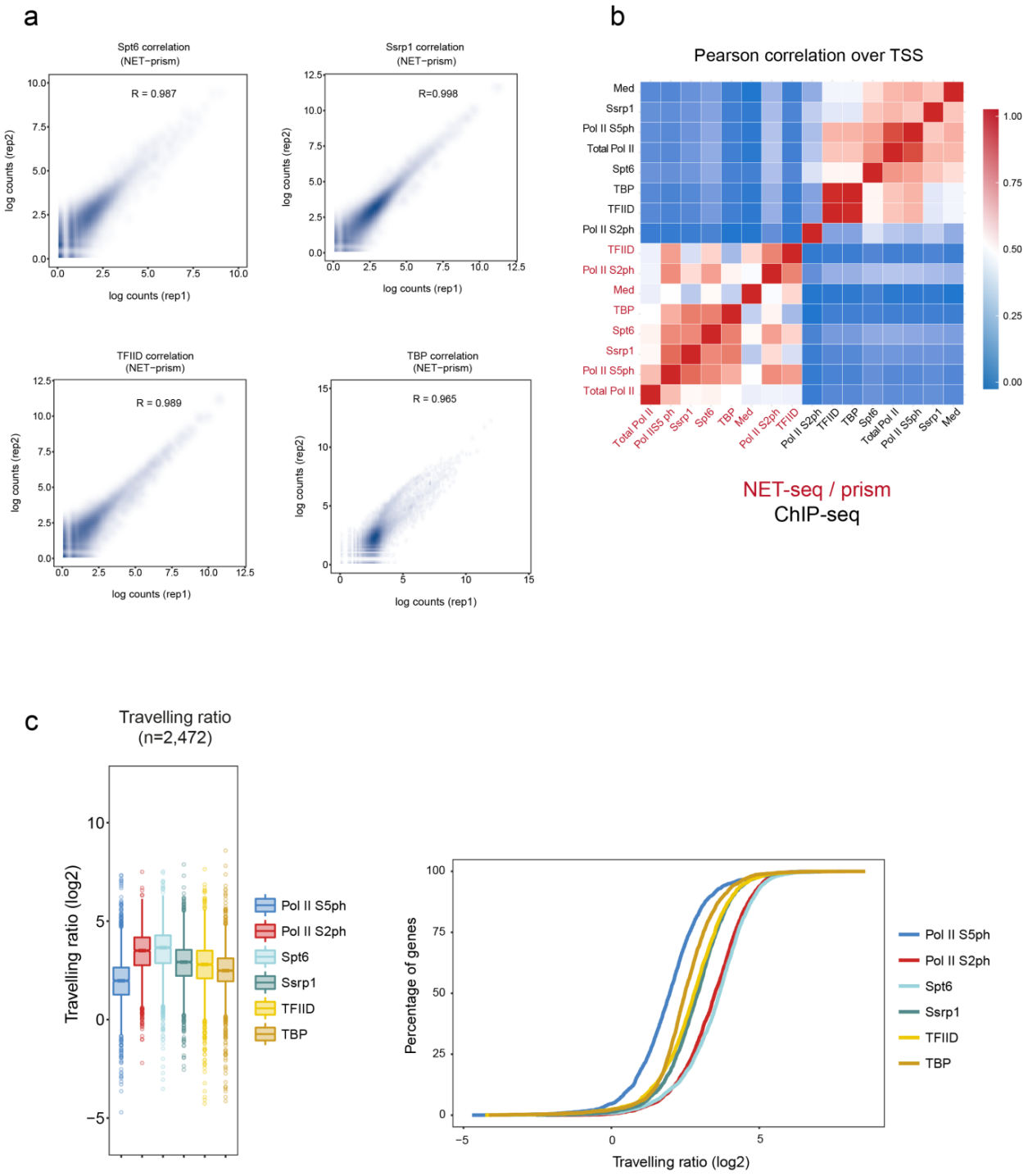
Supplementary Figure 1: Release of Pol II from chromatin by DNase I treatment. **(a)** Western blot assessing Pol II release after treatment of $2x10^7$ mouse ES cells with different DNase I concentrations (0, 100U, 200U, 300U) in the presence of 1M Urea. All different DNase I treatments cause the same level of Pol II release from chromatin (P; Pellet, S; Supernatant). **(b)** Western blot assessing Pol II release after treatment of $2x10^7$ mouse ES cells with different Urea concentrations (0, 0.05 M, 0.25 M, 0.5 M) in the presence of 100U of DNase I. The conditions highlighted in red (0.05 M Urea & 100U DNAse I) were used for the generation of all NET-prism libraries (P; Pellet, S; Supernatant).



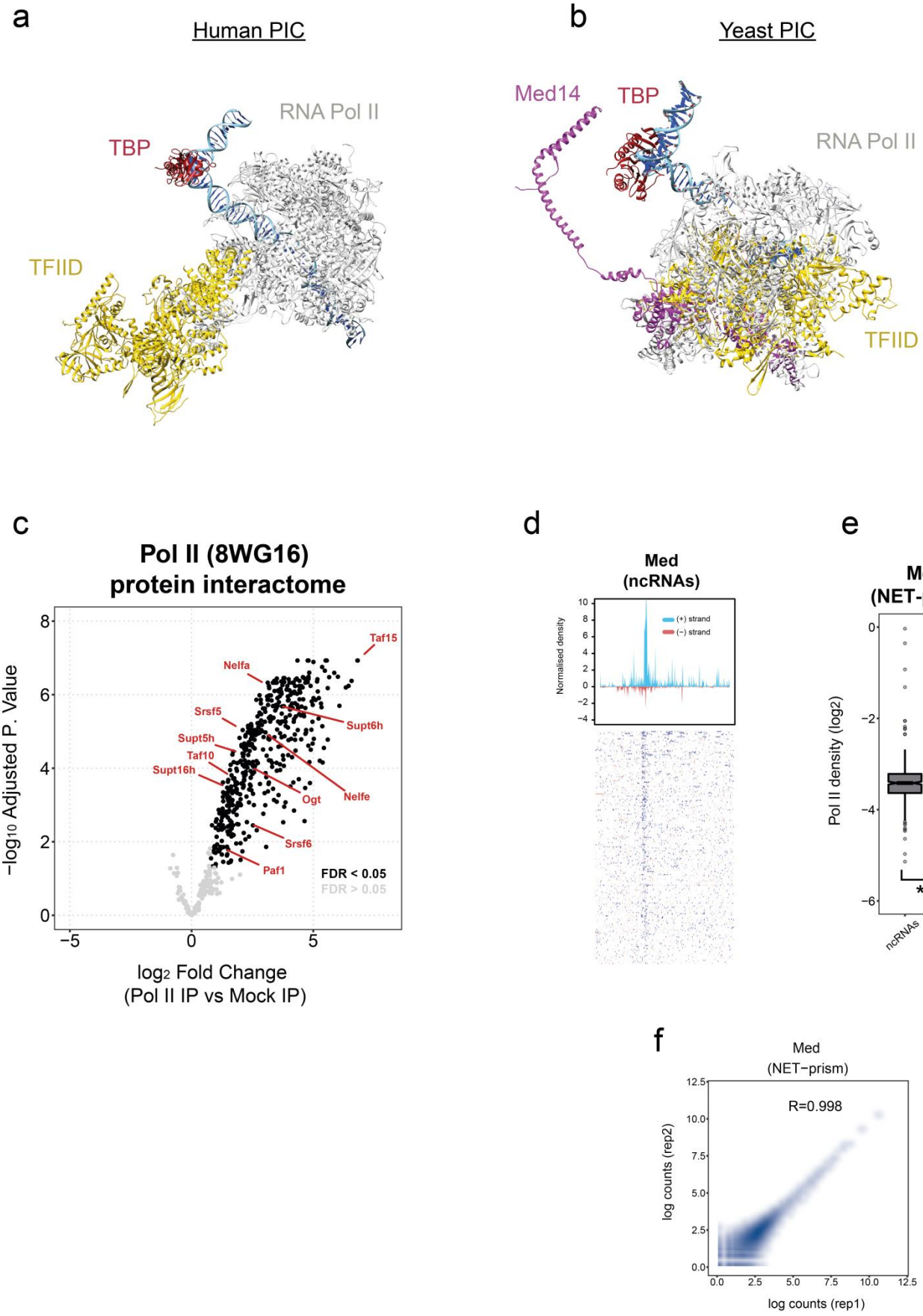
Supplementary Figure 2: NET-prism enables polymerase-specific transcriptional interrogation at a high resolution over long non-coding RNAs (lncRNAs). **a)** Metaplot profiles for total Pol II (ChIP-seq & NET-seq) and Pol II S2ph / Pol II S5ph (ChIP-seq & NET-prism) over the TSS and TES of lncRNAs (n = 127). A 10-bp smoothing window has been applied. **(b)** Pol II density (Total Pol II, Pol II S2ph, Pol II S5ph), assessed either by ChIP-seq or NET-seq/prism, over a single gene (*1700001L05Rik*). Black = ChIP-seq density, Blue = Sense transcription (NET-seq/prism), Red = Anti-sense transcription (NET-seq/prism)



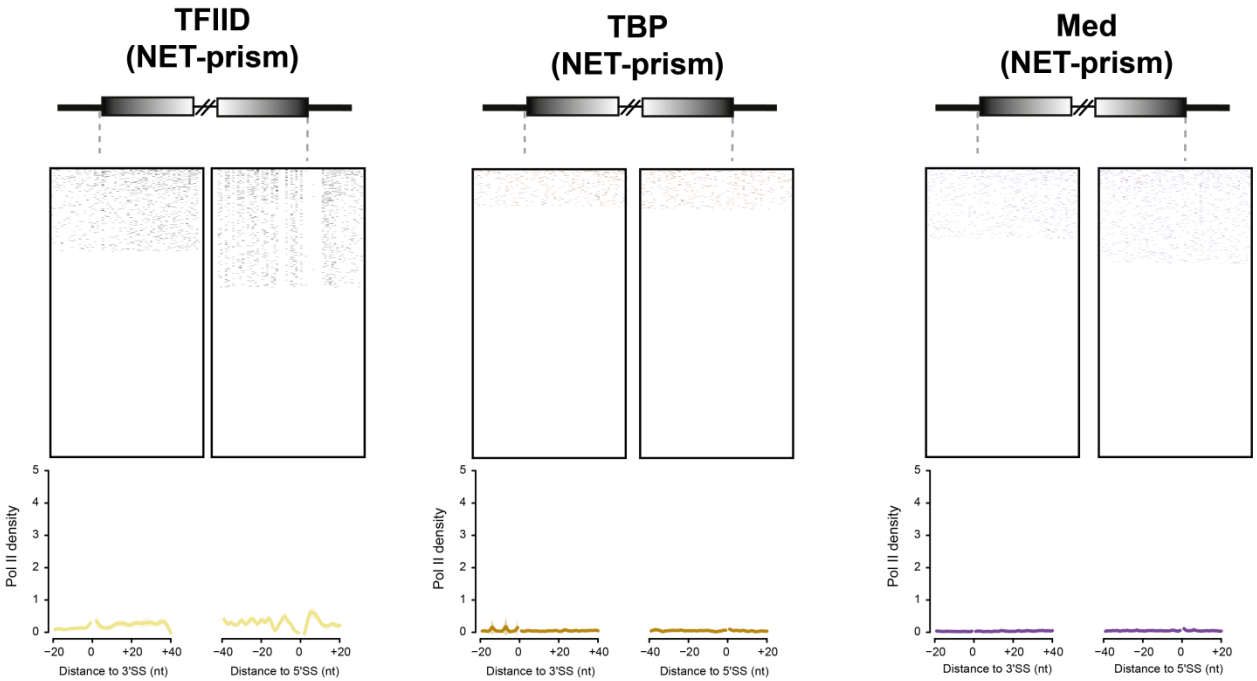
Supplementary Figure 3: High reproducibility and resolution of NET-prism as compared to ChIP-seq. **(a)** Correlation plots assessing reproducibility between replicates of Pol II S2ph and Pol II S5ph libraries (NET-prism). R corresponds to Pearson correlation. **(b)** Scatterplots between NET-seq/prism and ChIP-seq densities. Spearman ρ is indicated on each plot.



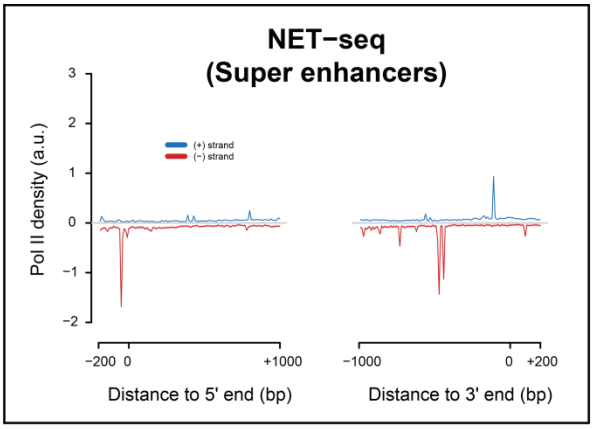
Supplementary Figure 4: Transcriptional activity assessed by NET-prism **(a)** Correlation plots assessing reproducibility between replicates of Spt6, Ssrp1, TFIID, and TBP libraries (NET-prism). R corresponds to Pearson correlation. **(b)** Correlation heatmap of all NET-seq/prism and ChIP-seq libraries on promoter regions (± 500 bp around TSS) of all uniquely annotated genes (n = 12,737). **(c)** Boxplot and cumulative distribution of Pol II travelling ratio as assessed by NET-prism.

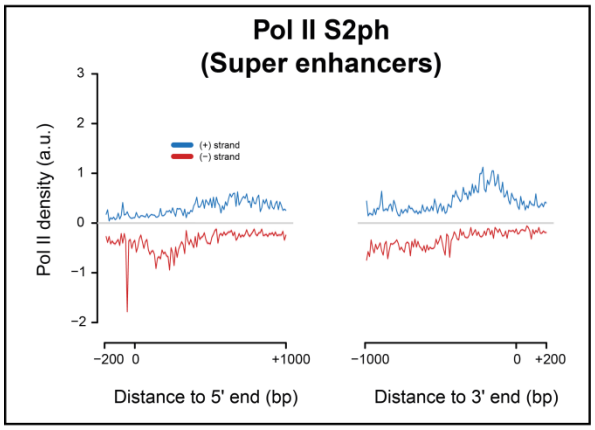

a
Human PIC
RNA Pol II
TBP
TFIID
b
Yeast PIC
Med14
TBP
RNA Pol II
TFIID
c
Pol II (8WG16)
protein interactome
−log10 Adjusted P. Value
log2 Fold Change
(Pol II IP vs Mock IP)
Taf15
Nelfa
Srsf5
Supt6h
Supt5h
Taf10
Supt16h
Ogt
Nelfe
Srsf6
Paf1
FDR < 0.05
FDR > 0.05
d
Med
(ncRNAs)
Normalised density
(+) strand
(−) strand
e
Med
(NET-prism)
Pol II density (log2)
**
ncRNAs
protein-coding
f
Med
(NET−prism)
R=0.998
log counts (rep2)
log counts (rep1)

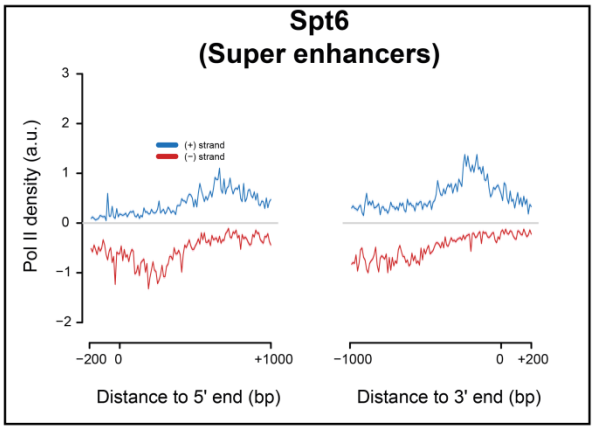
Supplementary Figure 5: NET-prism is affected by the binding affinity of each immunoprecipitated protein to the RNA Pol II. **(a)** Crystal structure of the human pre-initiation complex (PIC). Proteins have been super-imposed and visualised via UCSF chimera. **(b)** Same as **(a)** but for yeast PIC. The crystal structure of Med14 has also been implemented and is indicated in purple. **(c)** Volcano plot of Pol II IP vs Mock (IgG) IP depicting a whole RNA Pol II – protein interactome as assessed by Mass spectrometry. Significant values (FDR < 0.05) are coloured in black. **(d)** Metaplot profile and heatmap over annotated non-coding RNAs (ncRNAs) (n =434) for Pol II bound by Mediator (Med). A 10-bp smoothing window has been applied. Blue = Sense transcription, Red = Anti-sense transcription. **(e)** Boxplot comparing Pol II density assessed by Mediator immunoprecipitation over ncRNAs and protein-coding genes. Significance was tested via the Wilcoxon rank test (** $p < 1.0e^{-10}$). **(f)** Correlation plot assessing reproducibility between replicates of Mediator libraries (NET-prism). R corresponds to Pearson correlation.

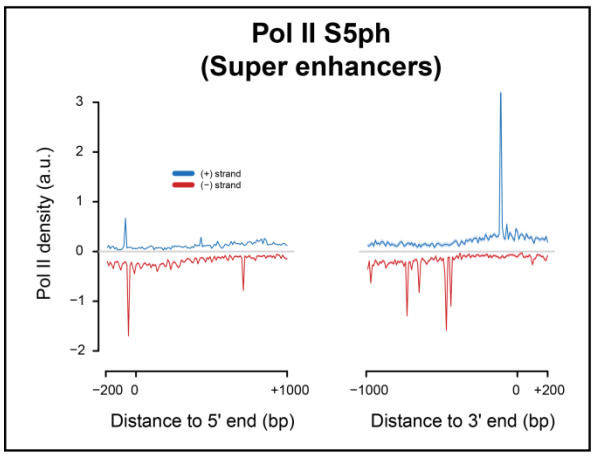


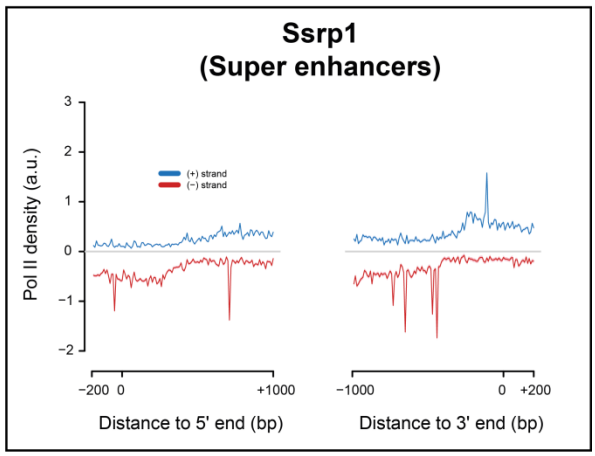
Supplementary Figure 6: No association of the PIC complex with co-transcriptional splicing. Heatmaps and metaplots assessing polymerase pausing for TFIID, TBP, and Mediator over exon boundaries (n = 2,586). Solid lines indicate the mean values, whereas the shading represents the 95% confidence interval.

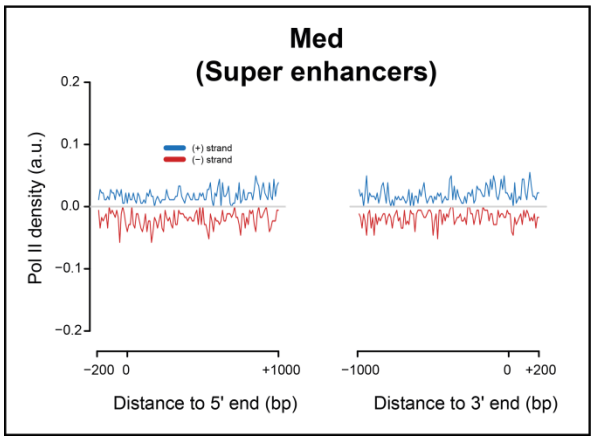

NET−seq
(Super enhancers)
Pol II density (a.u.)
(+) strand
(−) strand
−200 0 +1000
Distance to 5' end (bp)
−1000 0 +200
Distance to 3' end (bp)

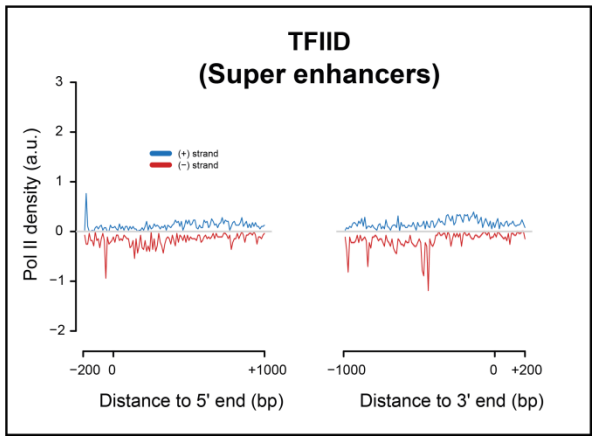

Pol II S2ph
(Super enhancers)
Pol II density (a.u.)
(+) strand
(−) strand
−200 0 +1000
Distance to 5' end (bp)
−1000 0 +200
Distance to 3' end (bp)


Spt6
(Super enhancers)
Pol II density (a.u.)
(+) strand
(−) strand
−200 0 +1000
Distance to 5' end (bp)
−1000 0 +200
Distance to 3' end (bp)

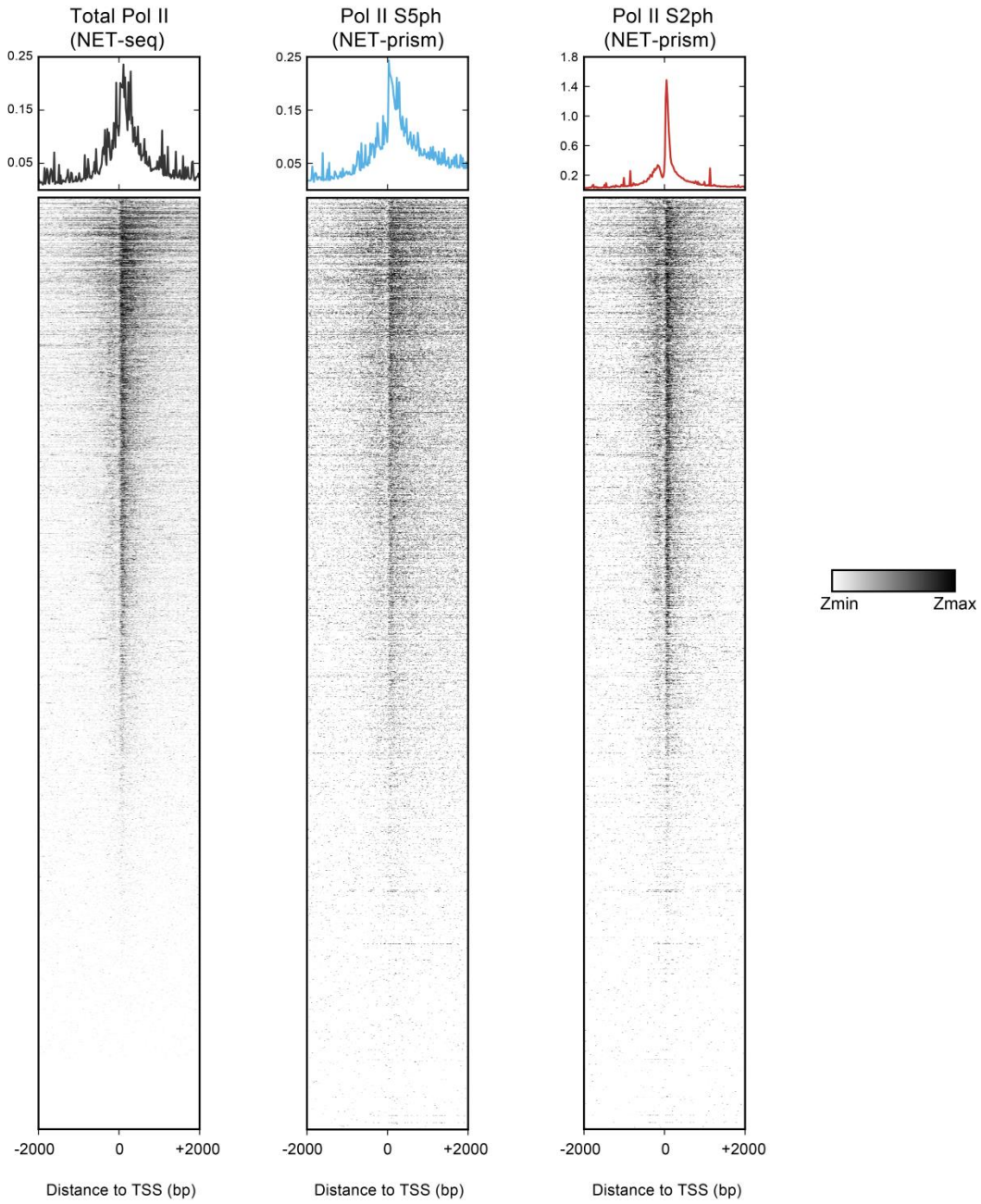

Pol II S5ph
(Super enhancers)
Pol II density (a.u.)
(+) strand
(−) strand
−200 0 +1000
Distance to 5' end (bp)
−1000 0 +200
Distance to 3' end (bp)


Ssrp1
(Super enhancers)
Pol II density (a.u.)
(+) strand
(−) strand
−200 0 +1000
Distance to 5' end (bp)
−1000 0 +200
Distance to 3' end (bp)


Med
(Super enhancers)
Pol II density (a.u.)
(+) strand
(−) strand
−200 0 +1000
Distance to 5' end (bp)
−1000 0 +200
Distance to 3' end (bp)


TFIID
(Super enhancers)
Pol II density (a.u.)
(+) strand
(−) strand
−200 0 +1000
Distance to 5' end (bp)
−1000 0 +200
Distance to 3' end (bp)

Supplementary Figure 7: Unique patterns of transcriptional pausing over super-enhancers. Metaplots assessing polymerase pausing for total Pol II, Pol II S2ph, Pol II S5ph, Ssrp1, Spt6, TFIID, and Mediator over super-enhancers (n = 226). Solid lines indicate the mean values, whereas the shading represents the 95% confidence interval.



Supplementary Figure 8: Heatmaps and density metaplots of total Pol II (NET-seq), Pol II S5ph (NET-prism), and Pol II S2ph (NET-prism) over 11,177 uniquely annotated protein-coding genes. Metaplot profiles are smoothed with a 20-bp sliding window. Genes are sorted by total Pol II (NET-seq) levels.

Materials and Methods

Cell culture. The E14 cell line (mESCs) was cultured at 37 °C, 7.5% CO_2, on 0.1% gelatin coated plates, in DMEM + GlutaMax™ (Gibco) with 15% fetal bovine serum (Gibco), MEM non- essential amino acids (Gibco), penicillin/streptomycin (Gibco), 550 µM 2-mercaptoethanol (Gibco), and 10 ng/ml of leukaemia inhibitory factor (LIF) (eBioscience).

Antibody-bead coating. 50 µl of Dynabeads G were washed twice in 200 µl of IP buffer (50 mM Tris-HCl (pH 7.0), 50 mM NaCl, 1% NP-40) and 5-10 µg of antibody was added. Antibodies used in this study: Pol II S2ph (Cat No: 61083 – Active Motif); Pol II S5ph (ab5408 – Abcam); Spt6 (Cell Signalling); Ssrp1 (Biolegends); TFIID (Santa Cruz); TBP (ab818 – Abcam); Med14 (Invitrogen), total Pol II (ab817 – Abcam).

Nuclear extraction and DNase treament. 10^8 ES cells were used for each IP. It is important to split cells down to five batches of $2x10^7$ each when performing nuclei extraction. All extraction steps are performed on ice to avoid degradation of the nascent RNA. $2x10^7$ were treated with 200 µl of cytoplasmic lysis buffer (0.15% (vol/vol) NP-40, 10 mM Tris-HCl (pH 7.0), 150 mM NaCl, 25 µM α-amanitin (Epichem), 10 U RNasin Ribonuclease inhibitor (Promega) and 1× protease inhibitor mix (Thermo)) for 5 min on ice. Lysate was layered on 500 µl of sucrose buffer (10 mM Tris-HCl (pH 7.0), 150 mM NaCl, 25% (wt/vol) sucrose, 25 µM α-amanitin, 20 U RNasin Ribonuclease inhibitor and 1× protease inhibitor mix) and spun down for 5 min at 16,000g (4ºC). Supernatant was carefully removed and nuclei were resuspended in 100 µl of DNase digestion buffer (1x DNase buffer (NEB), 25 µM α-amanitin, 20 U RNasin Ribonuclease inhibitor and 1× protease inhibitor mix) and further treated with 100 U of DNase I (NEB) for 20 min on ice. It is important for nuclei to be fully resuspended in the DNase digestion buffer. Non resuspended nuclei are an indication of harsh cytoplasmic lysis conditions – In this case reduce the volume of cytoplasmic lysis buffer.

Chromatin-solubilised nuclei were spun down at 6,000g (4ºC) for 2 min and supernatant was carefully removed. Nuclei were further treated with 200 µl of nuclei lysis buffer (1% (vol/vol) NP-40, 20 mM HEPES (pH 7.5), 125 mM NaCl, 50 mM urea, 0.2 mM EDTA, 0.625 mM DTT, 25 µM α-amanitin, 20 U RNasin Ribonuclease inhibitor and 1× protease inhibitor mix) for 5 min on ice. Nuclei lysate was spun down at 18,500g (4ºC) for 2 min and supernatant from five different batches was combined. Phophatase inhibitor mix (x1) (Thermo) was implemented on all the above extraction steps for batches intended for Pol II S2ph and Pol II S5ph immunoprecipitations.

Chromatin Immunoprecipitation (IP) and nascent RNA extraction. Combined supernatants from the previous step were incubated in a final 1/10 dilution in IP buffer for 2 hours at 4ºC. Beads were washed 4 times with 1 ml of IP buffer and 700 µl of Qiazol (Qiagen) was directly added to the beads, followed by 140 µl of Chloroform. Samples were spun down and supernatant was ethanol precipitated (0.3M NaOAc, 2 µl Glycoblue). Concentration and size of nascent RNA was assessed by Nanodrop and TapeStation 2200, respectively. An IP from 10^8 ES cells usually yields ~200-1000 ng of nascent RNA. Assessment of RNA size is important in order to evaluate the fragmentation time during the library preparation.

NET-prism library preparation. Two biological replicates were processed for each IP and library preparation. NET-prism libraries were prepared similarly to the human NET-seq protocol [1] with a few modifications. The random barcode was ligated overnight at 16 ºC to maximise ligation efficiency. Fragmentation of the ligated nascent RNA varies depending on the size of the RNA fragments obtained from each IP. IPs for Pol II S5ph, Pol II S2ph, Ssrp1, and Spt6 yielded large RNA fragments and therefore the ligated nascent RNA was fragmented until all RNA transcripts were within the range of ~35-200 nucleotides. IPs for TBP, TFIID, and Mediator yielded fragments < 200 nt and therefore no fragmentation was performed. Maximum recovery of ligated RNA and cDNA was achieved from 15 % TBE-Urea (Invitrogen) and 10% TBE-Urea (Invitrogen),

respectively, by adding RNA recovery buffer (Zymo Research, R1070-1-10) to the excised gel slices and further incubating at 70°C (1500 rpm) for 15 min. Gel slurry was transferred through a Zymo-Spin IV Column (Zymo Research, C1007-50) and further precipitated for subsequent library preparation steps. cDNA containing the 3' end sequences of a subset of mature and heavily sequenced snRNAs, snoRNAs, and rRNAs, were specifically depleted using biotinylated DNA oligos[2]. Oligo-depleted circularised cDNA was amplified via PCR (9-12 cycles) and double stranded DNA was run on a 4% low melt agarose gel. The final NET-seq library running at ~150 bp was extracted and further purified using the ZymoClean Gel DNA recovery kit (Zymo Research). Sample purity and concentration was assessed in a 2200 TapeStation and further sequenced in a HiSeq 2500 Illumina Platform.

NET-prism analysis. All the NET-prism fastq files were processed using custom Python scripts to remove PCR duplicates and reads arising from RT bias (mm10 genome). Reads mapping exactly to the last nucleotide of each intron and exon (Splicing intermediates) were further removed from the analysis. The final NET-prism BAM files were converted to bigwig (1 bp bin), separated by strand, and normalized to x1 sequencing depth using Deeptools (v 2.4) with an "–Offset 1" in order to record the position of the 5' end of the sequencing read which corresponds to the 3' end of the nascent RNA. NET-seq/prism tags sharing the same or opposite orientation with the TSS were assigned as 'sense' and 'anti-sense' tags, respectively. Promoter-proximal regions were carefully selected for analysis to ensure that there is minimal contamination from transcription arising from other transcription units. Genes overlapping within a region of 2.5 kb upstream of the TSS were removed from the analysis. For the NET-seq/prism metaplots, genes underwent several rounds of k-means clustering in order to filter regions; in a 2kb window around the TSS, rows displaying very high Pol II occupancy within a <100 bp region were removed from the analysis as they represent non-annotated short non-coding RNAs. Average Pol II occupancy profiles were visualised using R (v 3.3.0).

Travelling ratio & Termination index. The travelling ratio is calculated via:

$$Travelling\ ratio = \frac{\text{Proximal Promoter}}{\text{Gene body}}$$

with Proximal Promoter defined as the Pol II coverage -30 bp and +250 bp around the TSS whereas Gene body region as the Pol II coverage +300 bp downstream of TSS and -200 bp upstream of TES.

The termination index is calculated via:

$$Termination\ index = \frac{\text{Termination signal}}{\text{Gene body}}$$

with the Termination signal defined as the Pol II coverage +2000 bp downstream of the TES whereas Gene body region as the Pol II coverage +300 bp downstream of TSS and -200 bp upstream of TES.

The reverse travelling ratio is calculated via:

$$Reverse\ Travelling\ ratio = \frac{\text{Divergent Elongation}}{\text{Divergent Initiation}}$$

with Divergent Elongation is defined as the Pol II coverage between -1500 bp and -330 bp upstream of the TSS whereas Divergent Initiation as the Pol II coverage between -300 bp and -30 bp upstream of the TSS.

ChIP-seq data processing. All ChIP-seq fastq files were aligned to the mm10 genome using Bowtie2 (v 2.2.6) with default parameters[3] . All BAM files were converted to bigwig (10 bp bin) and normalised to x1 sequencing depth using Deeptools (v 2.4)[4]. Duplicated reads were removed. Blacklisted mm9 co-ordinates were converted to mm10 using the LiftOver tool from UCSC and were further removed from the analysis. Average binding profiles were visualised using R (v 3.3.0).

Mass spectrometry sample preparation. Nuclei were treated with DNase I as described above and the supernatant was incubated for 2 hours with a total Pol II antibody (ab817 – Abcam) or IgG (Cell Signalling) at 4ºC. After thorough washing of beads with IP buffer, samples were incubated overnight at 37°C with Tris pH 8.8 and 300 ng Trypsin Gold (Promega). In total, four samples were prepared for each IP (Total Pol II, IgG). Peptides were desalted using StageTips[5] and dried. The peptides were resuspended in 0.1% formic acid and analysed using liquid chromatography - mass spectrometry (LC-MS/MS).

LC-MS/MS analysis. Peptides were separated on a 25 cm, 75 µm internal diameter PicoFrit analytical column (New Objective) packed with 1.9 µm ReproSil-Pur 120 C18-AQ media (Dr. Maisch,) using an EASY-nLC 1200 (Thermo Fisher Scientific). The column was maintained at 50°C. Buffer A and B were 0.1% formic acid in water and 0.1% formic acid in 80% acetonitrile. Peptides were separated on a segmented gradient from 6% to 31% buffer B for 45 min and from 31% to 50% buffer B for 5 min at 200 nl / min. Eluting peptides were analyzed on a QExactive HF mass spectrometer (Thermo Fisher Scientific). Peptide precursor m/z measurements were carried out at 60000 resolution in the 300 to 1800 m/z range. The top ten most intense precursors with charge state from 2 to 7 only were selected for HCD fragmentation using 25% normalized collision energy. The m/z values of the peptide fragments were measured at a resolution of 30000 using a minimum AGC target of 8e3 and 55 ms maximum

injection time. Upon fragmentation, precursors were put on a dynamic exclusion list for 45 sec.

Protein identification and quantification. The raw data were analyzed with MaxQuant version 1.6.0.13 [6] using the integrated Andromeda search engine [7]. Peptide fragmentation spectra were searched against the canonical and isoform sequences of the mouse reference proteome (proteome ID UP000000589, downloaded December 2017 from UniProt). Methionine oxidation and protein N-terminal acetylation were set as variable modifications; cysteine carbamidomethylation was set as fixed modification. The digestion parameters were set to “specific” and “Trypsin/P,” The minimum number of peptides and razor peptides for protein identification was 1; the minimum number of unique peptides was 0. Protein identification was performed at a peptide spectrum matches and protein false discovery rate of 0.01. The “second peptide” option was on. Successful identifications were transferred between the different raw files using the “Match between runs” option. Label-free quantification (LFQ) [8] was performed using an LFQ minimum ratio count of 2. LFQ intensities were filtered for at least three valid values in at least one group and imputed from a normal distribution with a width of 0.3 and down shift of 1.8. The median value of the log2 LFQ intensities for the RNA Pol II IPs was used for the imputation of the missing values in the IgG IPs. Differential abundance analysis was performed using limma [9] .

Enhancers and Super-enhancers. BED files containing typical enhancer and super-enhancer coordinates in mESCs were downloaded from Whyte et.al.[10]. Distal enhancers were defined as regions that are not overlapping with any annotated gene within a 2000 bp window. Only the distal enhancers that displayed an RPKM > 1 for Pol II were kept for subsequent analyses.

Publicly available datasets

The human Pol II - TBP (5IY6)[11] , and TFIID (5FUR)[12] structures, as well as yeast Pol II – TBP (5FYW)[13], Med14 (5N9J)[14], and TFIID (4OY2)[15] structures were downloaded from the RCSB Protein Data Bank and superimposed using UCSF CHIMERA.

NET-seq dataset (GSE90906[2])

ChIP-seq (Pol II; GSE28247[16], Pol II S2ph; GSE20530[17], Ssrp1; GSE90906, Spt6; GSE103180[18], Med14; GSE44288[10], TFIID; GSE39237[19], TBP; GSE22557[20], Pol II S5ph, H3K27Ac (Encode Consortium – E14 cell line))

References

1. Mayer, A. & Churchman, L. S. Genome-wide profiling of RNA polymerase transcription at nucleotide resolution in human cells with native elongating transcript sequencing. *Nat Protoc.* **11,** 813–833 (2016).

2. Mylonas, C. &Tessarz, P. The repressive and alleviating nature of FACT shapes the transcriptional landscape in ES cells. *In Revision* (2018).

3. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9,** 357–359 (2012).

4. Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44,** 160–165 (2016).

5. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and Go Extraction Tips for Matrix-Assisted Laser Desorption/Ionization, Nanoelectrospray, and LC/MS Sample Pretreatment in Proteomics. *Anal. Chem.* **75,** 663–670 (2003).

6. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotech* **26,** 1367–1372 (2008).

7. Cox, J. *et al.* Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J. Proteome Res.* **10,** 1794–1805 (2011).

8. Cox, J., Hein, M. Y., Luber, C. a & Paron, I. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. …* **13,** 2513–2526 (2014).

9. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43,** e47 (2015).

10. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153,** 307–319 (2013).

11. He, Y. *et al.* Near-atomic resolution visualization of human transcription promoter opening. *Nature* **533,** 359 (2016).

12. Louder, R. K. *et al.* Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature* **531,** 604–609 (2016).

13. Plaschka, C. *et al.* Transcription initiation complex structures elucidate DNA opening. *Nature* **533,** 353 (2016).

14. Nozawa, K., Schneider, T. R. & Cramer, P. Core Mediator structure at 3.4 Å extends model of transcription initiation complex. *Nature* **545,** 248 (2017).

15. Bhattacharya, S. *et al.* Structural and functional insight into TAF1-TAF7, a subcomplex of transcription factor II D. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 9103–8 (2014).

16. Handoko, L. *et al.* CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat. Genet.* **43,** 630–638 (2011).

17. Rahl, P. B. *et al.* C-Myc regulates transcriptional pause release. *Cell* **141,** 432–445 (2010).

18. Wang, A. H. *et al.* The Elongation Factor Spt6 Maintains ESC Pluripotency by Controlling Super-Enhancers and Counteracting Polycomb Proteins. *Mol. Cell* **68,** 398–413.e6 (2017).

19. Ku, M. *et al.* H2A.Z landscapes and dual modifications in pluripotent and multipotent stem cells underlie complex genome regulatory functions. *Genome Biol.* **13,** R85 (2012).

20. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467,** 430–435 (2010).