NET-prism enables RNA polymerase-dedicated transcriptional interrogation at nucleotide resolution

Constantine Mylonas[1] and Peter Tessarz[1,2,3]

1 - Max Planck Institute for Biology of Ageing, Joseph-Stelzmann-Str. 9b, 50931 Cologne, Germany

2 - Cologne Excellence Cluster on Cellular Stress Responses in Ageing Associated Diseases (CECAD), University of Cologne, Joseph-Stelzmann-Strasse 26, 50931 Cologne, Germany

3 - corresponding author: ptessarz@age.mpg.de

The advent of quantitative approaches that enable interrogation of transcription at single nucleotide resolution has allowed a novel understanding of transcriptional regulation previously undefined. To better map transcription genome-wide at base pair resolution and with transcription/elongation factor dependency we developed an adapted NET-seq protocol called NET-prism (Native Elongating Transcription by Polymerase-Regulated Immunoprecipitants in the Mammalian genome). NET-prism introduces an immunoprecipitation to capture RNA Pol II – associated proteins, which reveals the interaction of these proteins with active RNA Pol II. Application of NET-prism on different Pol II subunits (Pol II S2ph, Pol II S5ph), elongation factors (Spt6, Ssrp1), and components of the pre-initiation complex (PIC) (TFIID, TBP, and Mediator) reveals diverse Pol II signals, at a single nucleotide resolution, with regards to directionality and intensity over promoters, splice sites, and enhancers/super-enhancers. NET-prism will be broadly applicable as it exposes transcription factor/Pol II dependent topographic specificity and thus, a new degree of regulatory complexity.

Approaches that precisely map the position of RNA Pol II at a high resolution are considered the cradle of transcriptional cartography as they provide a deeper insight into transcriptional regulatory mechanisms[1–6]. The human NET-seq protocol quantitatively purifies Pol II in the presence of a strong Pol II inhibitor hence omitting the utilisation of an antibody [2]. Although, it successfully maps the 3'end of nascent RNA to reveal the strand-specific position of Pol II with single nucleotide resolution, it cannot distinguish between different Pol II variants or specific protein-dependent transcriptional processes. On the other hand, the mammalian NET-seq protocol (mNET-seq) uses an IP to capture the nascent RNA produced by different C-terminal domain (CTD) phosphorylated forms of Pol II [7]. However, mNET-seq relies on the release of Pol II complexes from chromatin via digestion with micrococcal nuclease (MNase); a potent

nuclease that digests both DNA and nascent RNA. Consequently, short nascent RNA fragments may not be incorporated into the library.

Here we describe an approach (NET-prism) to capture nascent RNA transcripts generated by different Pol II variants and transcription/elongation factors associated with active Pol II. Nuclei are extracted in the presence of a strong inhibitor for Pol II (α-amanitin) to prevent run-on of the polymerase. Subsequently, nuclei are treated with DNase I to solubilise chromatin and promote release of the RNA Pol II complex while keeping the nascent RNA intact (**Supplementary Fig. 1a, b**). An antibody is then used to immunoprecipitate either differentially phosphorylated RNA Pol II variants or proteins bound to RNA Pol II. The successfully purified nascent RNA is later parsed into the NET-seq library preparation as previously described [2]. The detailed protocol is outlined in **Figure 1a** and is also available in the Methods section.

Results

Nascent RNA transcripts by differently phosphorylated Pol II variants

Initially, we applied NET-prism to map the strand-specific location of two differently phosphorylated Pol II variants (Pol II S2ph, and Pol II S5ph) at single nucleotide resolution. Over protein-coding genes, the Pol II S2ph variant was found to be highly enriched close to the TSS (Transcription Start Site), and after the TES (Transcription Termination Site), whereas the Pol II S5ph variant exhibited similar distribution to total Pol II as assessed by NET-seq (**Fig. 1b,c**). Similar Pol II occupancies were also

detected over long non-coding RNAs (**Supplementary Fig. 2)**. Very high correlations were observed between the different replicates (R = 0.99; Pol II S5ph, R = 0.98; Pol S2ph – **Supplementary Fig. 3a**) confirming the robustness and reproducibility of NET-prism. Moreover, the validity of the data and its robustness was confirmed by obtaining relatively high correlations between ChIP-seq and NET-prism (R = 0.75; Pol II S5ph, R = 0.69; Pol S2ph – **Supplementary Fig. 3b**). To further assess the density distribution of both Pol II variants, we calculated the travelling ratios (Pol II density over proximal promoter versus gene body), and termination indices (Pol II density over termination versus gene body). No significant difference was observed between the travelling ratio of total Pol II (NET-seq) and Pol II S5ph (NET-prism), as opposed to Pol II S2ph ($p < 10^{-16}$). Conversely, Pol II S5ph displayed a significantly lower termination index whereas the Pol II S2ph exhibited the highest, confirming enrichment of Pol II S2ph over promoters and termination sites and that of Pol II S5ph over promoters and gene body regions (**Fig. 1d).**

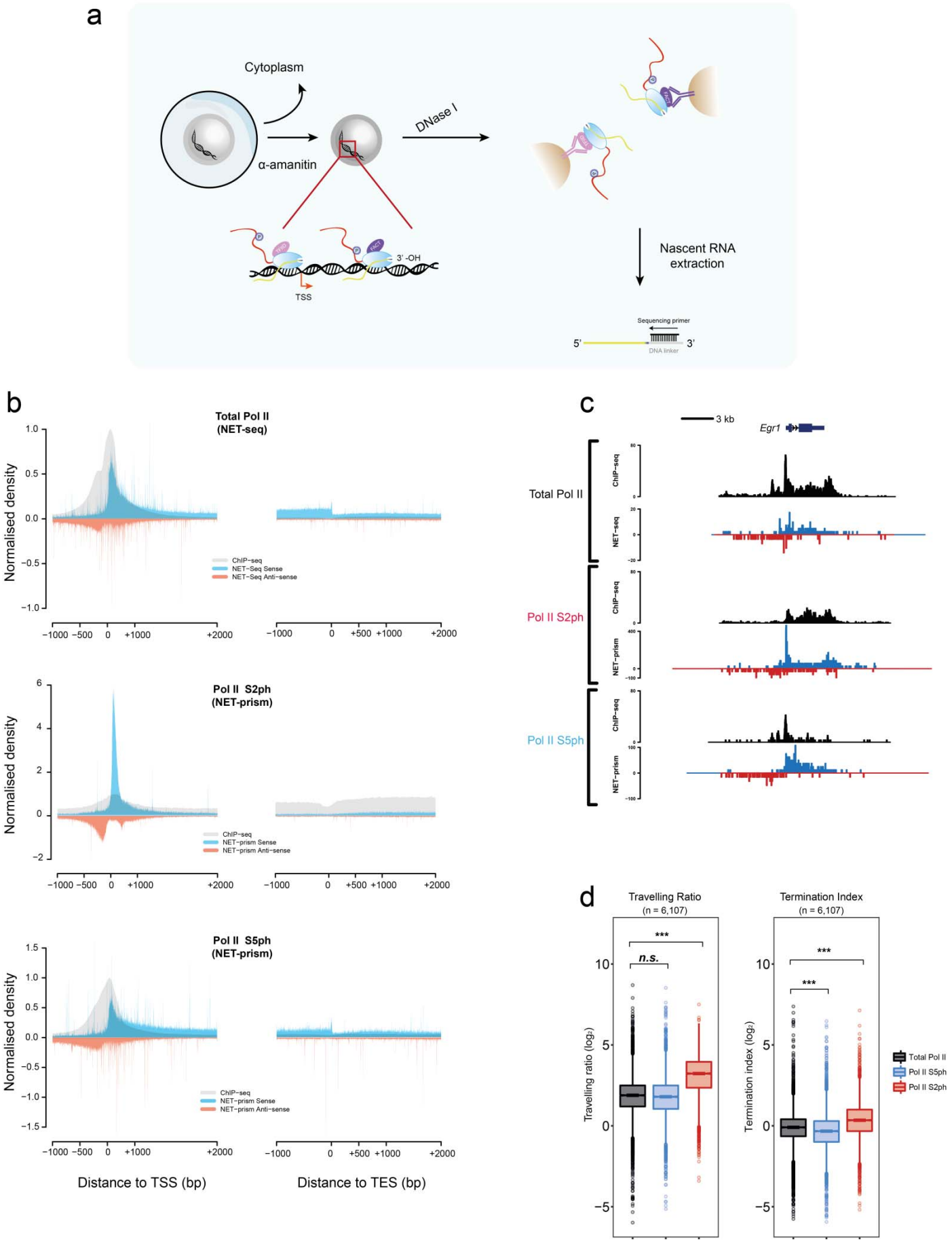


Figure 1: NET-prism enables polymerase-specific transcriptional interrogation at a high resolution. **(a)** NET-prism protocol overview. **(b)** Metaplot profiles for total Pol II (ChIP-seq & NET-seq) and Pol II S2ph / Pol II S5ph (ChIP-seq & NET-prism) over the TSS and TES of protein-coding genes (n = 6,107). No smoothing has been applied. **(c)** 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 (*Egr1*). Black = ChIP-seq density, Blue = Sense transcription (NET-seq/prism), Red = Anti-sense transcription (NET-seq/prism). **(d)** Boxplot comparison of the travelling ratios and termination indices for total Pol II, Pol II S2ph, and Pol II S5ph assessed via NET-seq/prism. Significance was tested via the Wilcoxon rank test (*** $p < 2.2e^{-16}$, n.s. = non-significant).

Transcription factor – Pol II dependent nascent transcription

Given the high concordance between ChIP-seq and NET-prism for both RNA Pol II variants, we next applied NET-prism on the elongation factors Spt6, and Ssrp1 (subunit of the FACT heterodimer), as well as on Transcription factor IID (TFIID), TATA-binding box protein (TBP), and Mediator (Med) with the latter serving as fundamental components of the pre-initiation complex (PIC). The data were highly reproducible among replicates (**Supplementary Fig. 4a**) and exhibited diverse levels of correlations over promoter regions (**Supplementary Fig. 4b**) indicating that different TFs establish unique Pol II footprints. Indeed, aligned and averaged NET-prism profiles over the TSS expose diversity in transcriptional initiation and elongation, suggesting that TF binding specificity directly affects RNA Pol II travelling. IPs for elongation factors Spt6 and Ssrp1 show strong and broad enrichment of the Pol II complex reminiscent of Pol II S2ph and Pol II S5ph distribution, respectively. Conversely, TFIID, and TBP IPs display sharper Pol II signals centred around the TSS (**Fig. 2a**). Similar Pol II patterns were also confirmed at a single gene level (**Fig. 2b**).

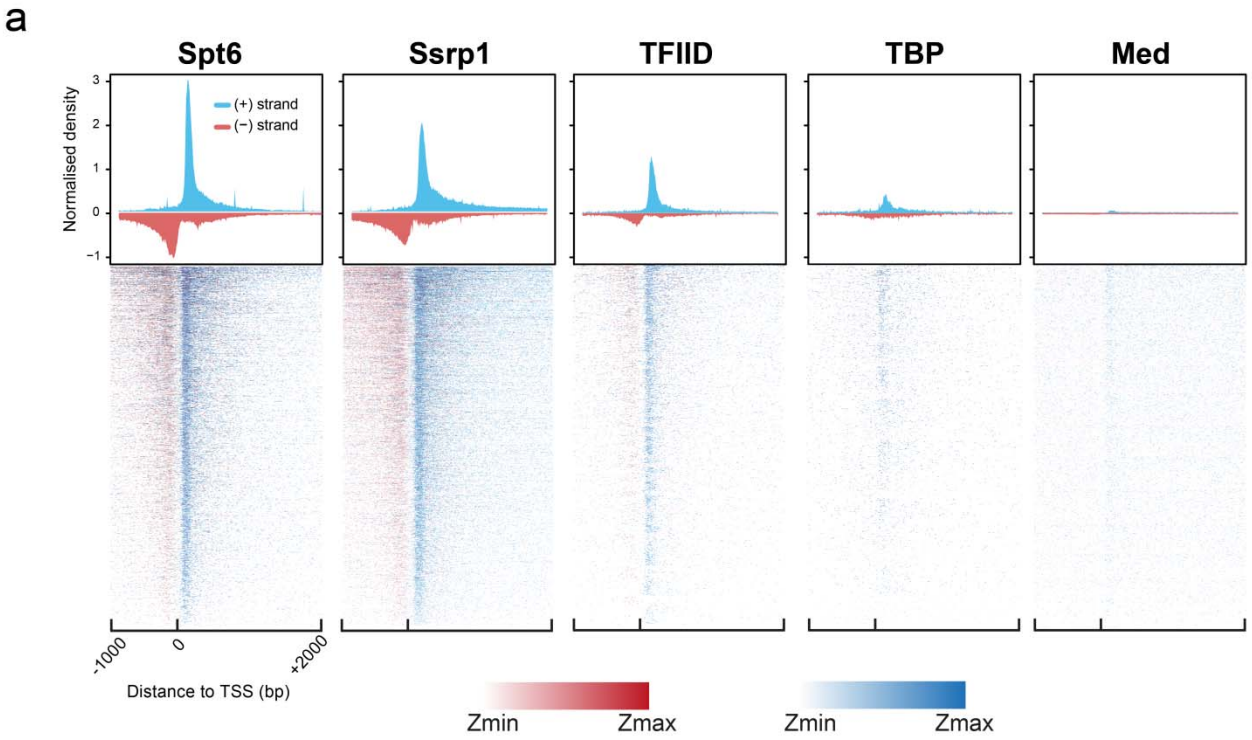
Nascent RNA transcripts upstream of the TSS are too short to produce mappable sequencing reads as the minimum read length is ~18 nt for unique alignment to the mammalian genome. Moreover, formation of the pre-initiation complex (PIC) at promoter regions occurs before nascent RNA synthesis Therefore, in order to characterise Pol II distribution over the PIC we coined the “Reverse travelling ratio” which is defined as the density of elongating divergent Pol II versus the density of divergent Pol II at initiation (**Fig. 2c** – schematic diagram). Interrogation of the regions with the highest Pol II density (n = 823) revealed a confinement of anti-sense Pol II

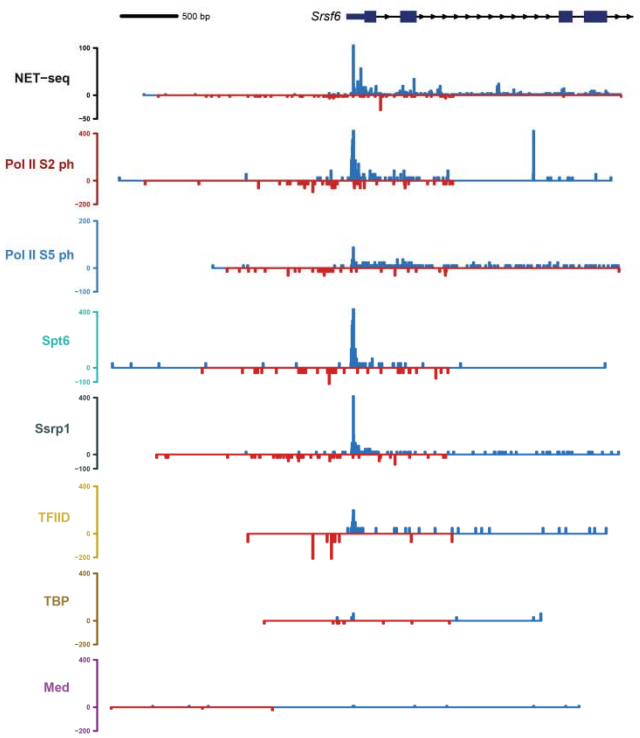
(lower reverse travelling ratios), bound by either TFIID or TBP, adjacent to TSS, thus confirming their role in pre-initiation. In addition, Pol II bound by either Spt6 or Ssrp1 exhibit higher reverse travelling ratios indicative of a broader anti-sense Pol II distribution (**Fig. 2c**). This is in agreement with ChIP-seq densities for both elongation factors[8,9].

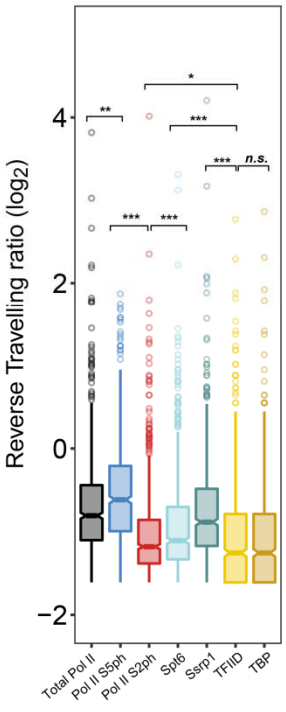
To test more systematically whether different NET-prism profiles generate exclusive Pol II distributions with regard to broadness and directionality, we calculated the travelling ratio in the sense direction for all the above NET-prism libraries (**Supplementary Fig. 4c**). Similarly, to the reverse travelling ratio, we confirmed the notion of a restrained Pol II at the TSS that is exclusively bound by TFIID and TBP, as opposed to Spt6 and Ssrp1 that support an involvement in transcriptional elongation. Surprisingly, both the reverse and normal travelling ratios expose a closer association of Spt6 to Pol II S2ph whereas that of Ssrp1 to Pol II S5ph. Corroborative evidence supporting this association arises from structural studies where the SH2 domain of Spt6 displays high affinity to Pol II S2ph[10,11].

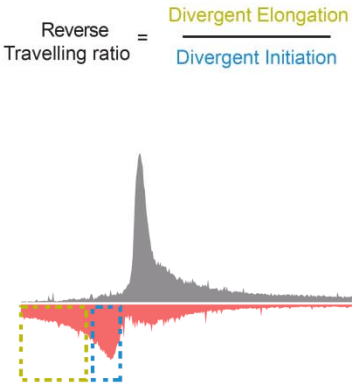
Out of all the NET-prism libraries that we generated, Med IP displayed the lowest Pol II density over protein-coding promoters despite its sequencing depth degree (137 million total reads, 49 million uniquely aligned). One likely explanation might be that the nascent RNA obtained by IP is strongly dependent on the binding affinity of each TF to RNA Pol II, explaining the lower Pol II read count over promoter regions for these specific IPs. Indeed, the crystal structures of human and yeast PIC reveal that TBP does not directly contact RNA Pol II, whereas the binding surface between Med14 and RNA Pol II is limited (**Supplementary Fig. 5a,b**). To confirm this, we interrogated the total

Pol II protein interactome via Mass spectrometry using the same extraction conditions as NET-prism. Positive (Supt5, Supt6, FACT, Paf1) and negative (NELF) elongation factors as well as splicing (Srsf5, Srsf6) and TFIID (Taf10, Taf15) components displayed a significant association with Pol II (**Supplementary Fig. 5c**). Neither TBP nor Mediator components were observed in the dataset suggesting either absent or weak interactions. Nevertheless, Pol II occupancy facilitated by Med is abundant over lncRNAs, snRNAs, and snoRNAs (**Supplementary Fig. 5d,e**) whereas the high replicate reproducibility (**Supplementary Fig. 5f**) confirms the robustness of the NET-prism protocol.









151

152

153 **Figure 2:** NET-prism application on polymerase-bound transcription factors. **(a)**
154 Metaplot profiles and heatmaps over protein-coding genes (n =4,314) for polymerase
155 associated elongation (Spt6, Ssrp1) and initiation (TFIID, TBP, Med) factors. A 10-bp
156 smoothing window has been applied. Blue = Sense transcription, Red = Anti-sense
157 transcription. **(b)** Polymerase density of all NET-seq/prism libraries over a single gene
158 (*Srsf6*). **(c)** Boxplot comparison and schematic diagram of the reverse travelling ratio
159 among different NET-prism libraries of protein-coding genes (n = 823). Significance was
160 tested via the Wilcoxon rank test (* $p< 0.01$, *** $p< 2.2e^{-16}$, n.s. = non-significant).

161

162

163

Assessment of kinetic splicing by NET-prism

Transcriptional elongation rates can affect splicing outcomes suggesting the proposal of the kinetic model of transcription and splicing coupling [12,13]. Data generated by human NET-seq, mNET-seq, and PRO-seq are consistent with this kinetic model[2,4,7]. Therefore, we sought to determine, via NET-prism, how transcriptional pausing is facilitated by different Pol II variants and elongation factors over exon boundaries. As splicing intermediates are known NET-seq contaminants due to the presence of 3'-OH groups in these RNAs [2], we removed them from the analysis to avoid bias. Total RNA Pol II, as assessed by NET-seq, in mouse ES cells showed increased pausing at exon boundaries similarly to human cells[2] (**Fig. 3a** – Total Pol II). Application of NET-prism confirmed that only the Pol II S5ph exhibited similar pausing, although less defined, at the 3' Splice Site (3'SS). When we focused on the 5' Splice Site (5'SS) we identified that Pol II pausing at the last nucleotides of the exon boundary was prominently absent for all IPs (**Fig. 3a**). To methodically compare transcriptional pausing for the different IPs, we introduced the 'Splicing Score', which derives from the Pol II density within 10 nucleotides around the 3'SS versus the density within 10 nucleotides around the 5'SS (**Fig. 3b**). The Pol II S5ph seemed to exhibit the strongest association with transcriptional splicing (present in 20.7% of total exons examined) compared to the other libraries (Pol II S2ph; 15%, Ssrp1; 13.3%, Spt6: 5.8%) (**Fig. 3c**). In addition, components of the PIC did not associate with Pol II pausing over spliced sites (**Supplementary Fig. 6**). This is in agreement with previous reports that support the involvement of Pol II S5ph [7] and Ssrp1[8,14] in the regulation of the transcriptional machinery during splicing. Our data therefore suggest that transcriptional splicing

mechanics is facilitated by Pol II variants and elongation factors differently. At present it is not clear why the phosphorylated Pol II isoforms and elongation factor IPs show different occupancies in comparison with total Pol II. It is however tempting to speculate that this represents some form of regulation for splicing catalysis. We also envision that NET-prism might be particularly useful to address the interplay of splicing factors with RNA Pol II at splice sites.

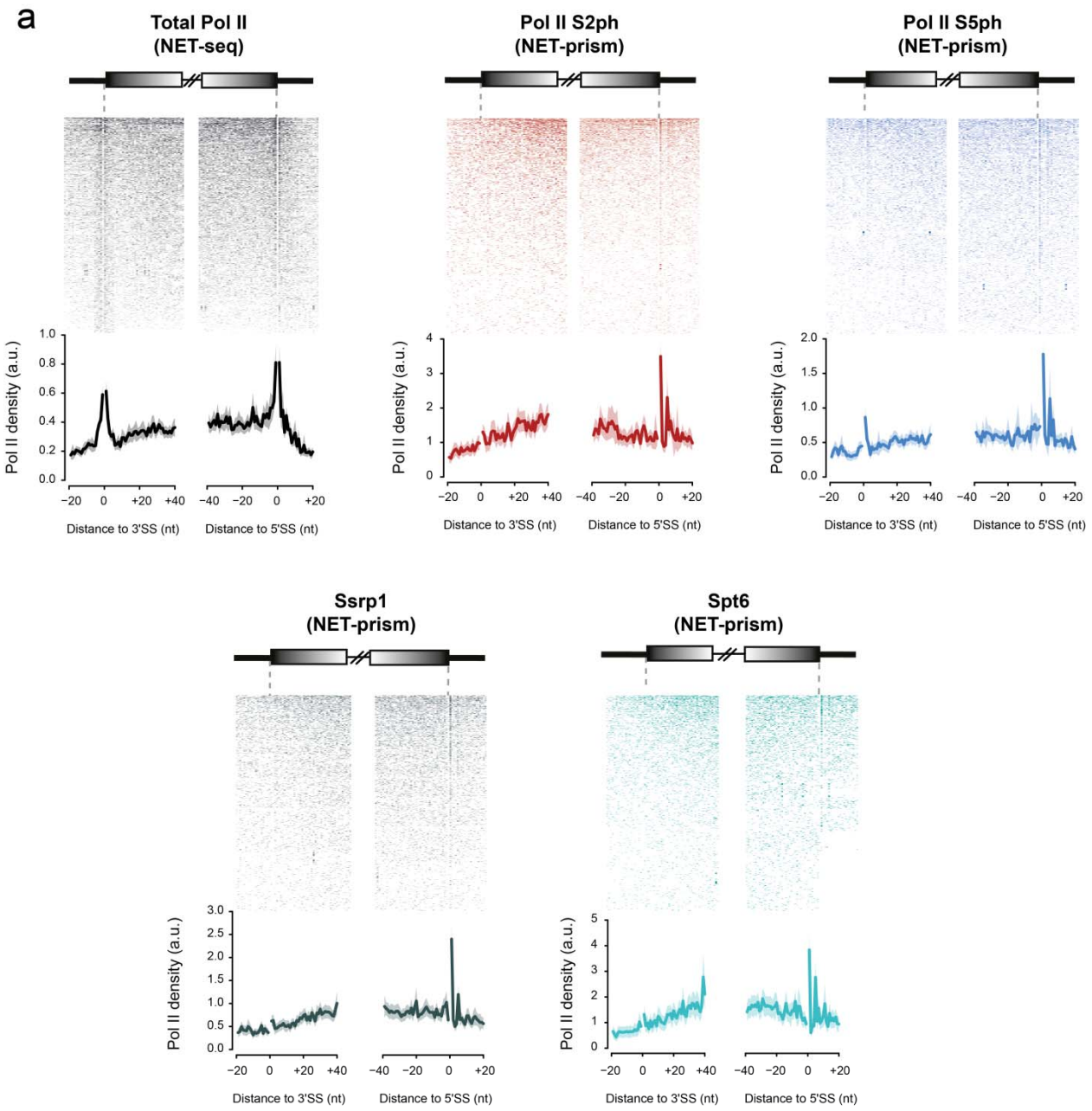


Figure 3: Association of different proteins with transcriptional splicing as assessed by NET-prism. **(a)** Heatmaps and metaplots assessing polymerase pausing for total Pol II, Pol II S2ph, Pol II S5ph, Ssrp1, and Spt6 over exon boundaries (n = 2,586). Solid lines indicate the mean values, whereas the shading represents the 95% confidence interval. **(b)** Schematic diagram depicting the calculation of the splicing score. Boxes denote the 10 nucleotide window around the 3'SS (blue) or 5'SS (red). **(c)** Cumulative distribution of the splicing score for each NET-prism library over all the assessed exons in **(a)**.

Diverse enrichment of RNA Pol II over enhancer regions

Enhancers and super-enhancers have been shown to play a prominent role in the control of gene expression programs essential for cell identity across many mammalian cell types [15]. Production of enhancer RNAs (eRNAs) is bidirectional and is governed by distinctive patterns of chromatin accessibility [16] but it is not well characterised whether the same transcriptional rules apply over enhancers as in promoters, in terms of initiation and elongation. We therefore extended our analysis over distal and super-enhancers and interrogated NET-prism density. Highest correlations were identified among Pol II S5ph – Ssrp1 and Pol II S2ph – Spt6 both for distal and super-enhancers (**Fig. 4a**). All Pol II variants and TFs exhibited significantly higher ChIP-seq density over super-enhancers as opposed to distal enhancers. Significantly increased transcriptional activity was confirmed over super-enhancers via NET-prism suggesting TF density being proportional to the degree of Pol II recruitment (**Fig. 4b**). Strikingly, both metaplot profiling (**Supplementary Fig. 7**) and single enhancer (**Fig. 4c**) interrogation of NET-prism transcriptional activity exposed distinctive topographic footprints; Pol II S5ph and Ssrp1 displayed patterns similar to transcriptional initiation whereas Pol II S2ph and Spt6 imitated a trail reminiscent of transcriptional elongation. Moreover, transcriptional activity prompted by TFIID also supports, to some degree, a notion of transcriptional initiation over enhancers (**Fig. 4c**).

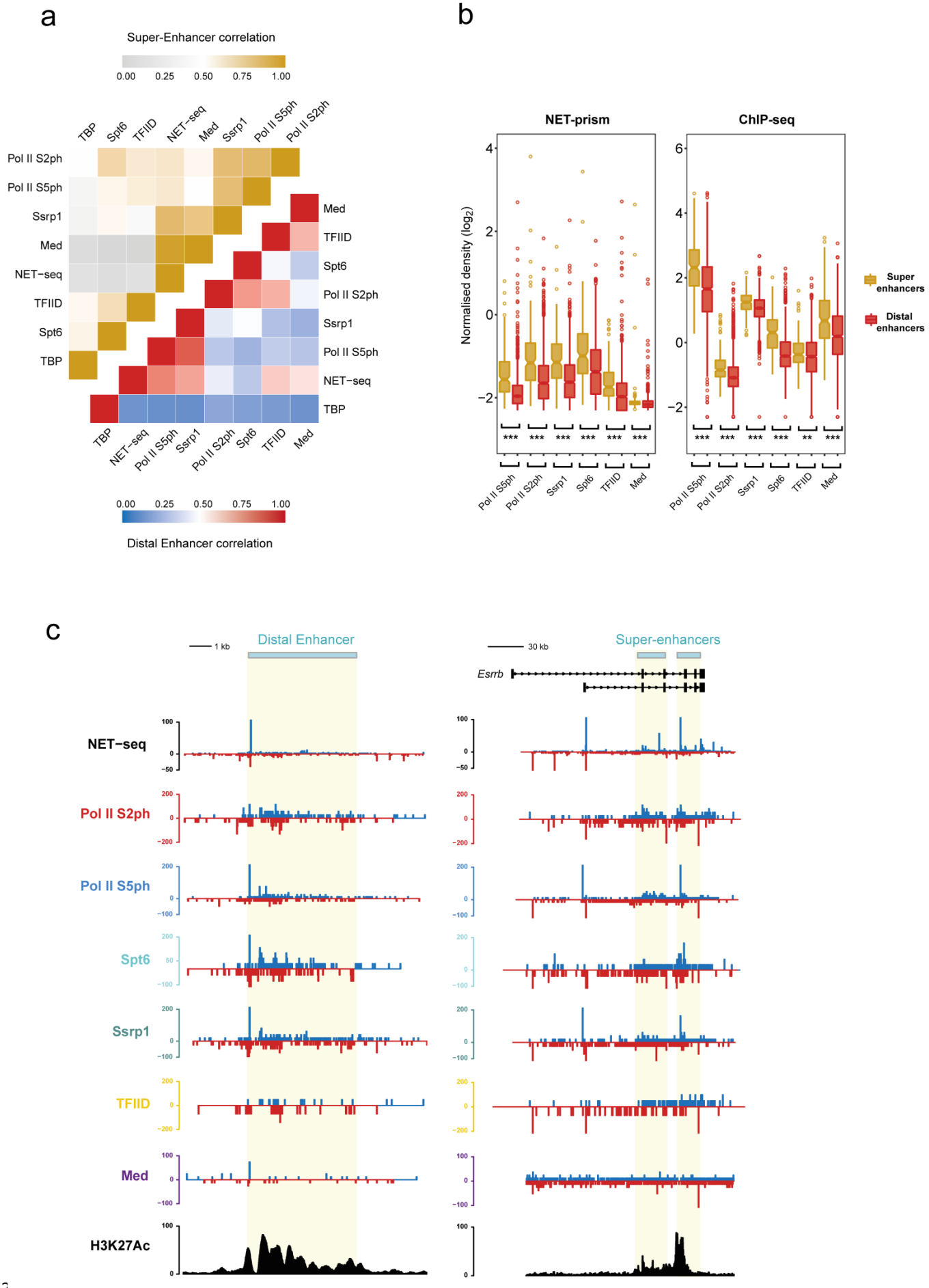


Figure 4: Distinctive patterns of transcriptional regulation over enhancers and super-enhancers. **(a)** Pearson's correlation heatmap among NET-seq/prism libraries over distal enhancers (blue – red) and super-enhancers (grey – gold). **(b)** Boxplots assessing either transcription factor (ChIP-seq) and Pol II (NET-prism) density over distal enhancers (red) and super-enhancers (gold). Significance was tested via the Wilcoxon rank test (** $p< 1.0e^{-10}$, *** $p< 2.2e^{-16}$). **(c)** Pol II distribution over a distal (chr1: 86,484,171 – 86,495,700) or super-enhancer as assessed by NET-seq/prism. H3K27Ac density is depicted in black colour. Blue and red depict RNA Pol II pausing in the positive and negative strand, respectively.

Discussion

Here, we have developed a new approach to accurately assess transcriptional topography at a high resolution. In summary, NET-prism allows the direct strand-specific interrogation of the transcriptional landscape at single nucleotide resolution of any protein of interest in complex with RNA Pol II. Its robustness enables a deeper insight into the interplay of transcriptional mechanisms conferred by different Pol II variants and proteins that are bound to Pol II. The comprehensive Pol II - protein interactome that we provide here facilitates the choice of the protein of interest when applying NET-prism. In addition, given the right RNA polymerase inhibitors and antibodies, NET-prism can be extended to specifically interrogate nascent transcription governed by either RNA Pol I or Pol III.

Although our approach relies on the release of Pol II from chromatin, NET-prism yields very similar results to NET-seq as the potency of the DNase is capable of liberating Pol II from all active genes (**Supplementary Fig. 8**).

Similarly to the human NET-seq [2], we expect the adaptation of NET-prism to be equally straightforward in any higher eukaryotic cell type. The combination of NET-prism with a high resolution ChIP-seq technique, such as ChIP-nexus [17], can illuminate how exactly *in vivo* binding of transcription factors correlates with transcriptional activity over different cell states and conditions. Therefore, NET-prism could become a valuable tool for unravelling unspecified transcriptional and regulatory complexity.

Author Contributions

C.M. and P.T. designed the study, C.M. performed all experiments and analysed data, C.M. and P.T. interpreted results and wrote the manuscript.

Acknowledgments

We would like to thank Stirling Churchman for critical reading and comments. We are also particularly grateful to Ilian Attanassov of the Max Planck Institute for Biology of Ageing Proteomics Core Facility for Mass Spectrometry Analysis. Sequencing was performed at the Max Planck Genome core centre in Cologne and data analysis was done on servers of the GWDG, Göttingen and the MPI-AGE cluster. We thank members of the Tessarz laboratory for discussion and comments on the manuscript. This work was funded by the Max Planck Society.

References

1. Churchman, L. S. & Weissman, J. S. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469,** 368–373 (2011).

2. Mayer, A. *et al.* Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* **161,** 541–544 (2015).

3. Min, I. M. *et al.* Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes Dev.* **25,** 742–754 (2011).

4. Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise Maps of RNA Polymerase Reveal How Promoters Direct Initiation and Pausing. *Science (80-.).* **339,** 950 LP-953 (2013).

5. Scruggs, B. S. *et al.* Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. *Mol. Cell* **58,** 1101–1112 (2015).

6. Fischl, H., Howe, F. S., Furger, A. & Mellor, J. Paf1 Has Distinct Roles in Transcription Elongation

and Differential Transcript Fate. *Mol. Cell* **65,** 685–698.e8 (2017).

7. Nojima, T. *et al.* Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell* **161,** 526–540 (2015).

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

9. 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).

10. Sun, M., Lariviere, L., Dengl, S., Mayer, A. & Cramer, P. A tandem SH2 domain in transcription elongation factor Spt6 binds the phosphorylated RNA polymerase II C-terminal repeat domain (CTD). *J. Biol. Chem.* **285,** 41597–41603 (2010).

11. Yoh, S. M., Cho, H., Pickle, L., Evans, R. M. & Jones, K. A. The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. *Genes Dev.* **21,** 160–174 (2007).

12. Dujardin, G. *et al.* How Slow RNA Polymerase II Elongation Favors Alternative Exon Skipping. *Mol. Cell* **54,** 683–690 (2014).

13. Fong, N. *et al.* Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes & Development* **28,** 2663–2676 (2014).

14. Sims, R. J. *et al.* Recognition of Trimethylated Histone H3 Lysine 4 Facilitates the Recruitment of Transcription Postinitiation Factors and Pre-mRNA Splicing. *Mol. Cell* **28,** 665–676 (2007).

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

16. Core, L. J. *et al.* Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat. Genet.* **46,** 1311–1320 (2014).

17. He, Q., Johnston, J. & Zeitlinger, J. ChIP-nexus enables improved detection of in vivo transcription factor binding footprints. *Nat. Biotechnol.* **33,** 395–401 (2015).