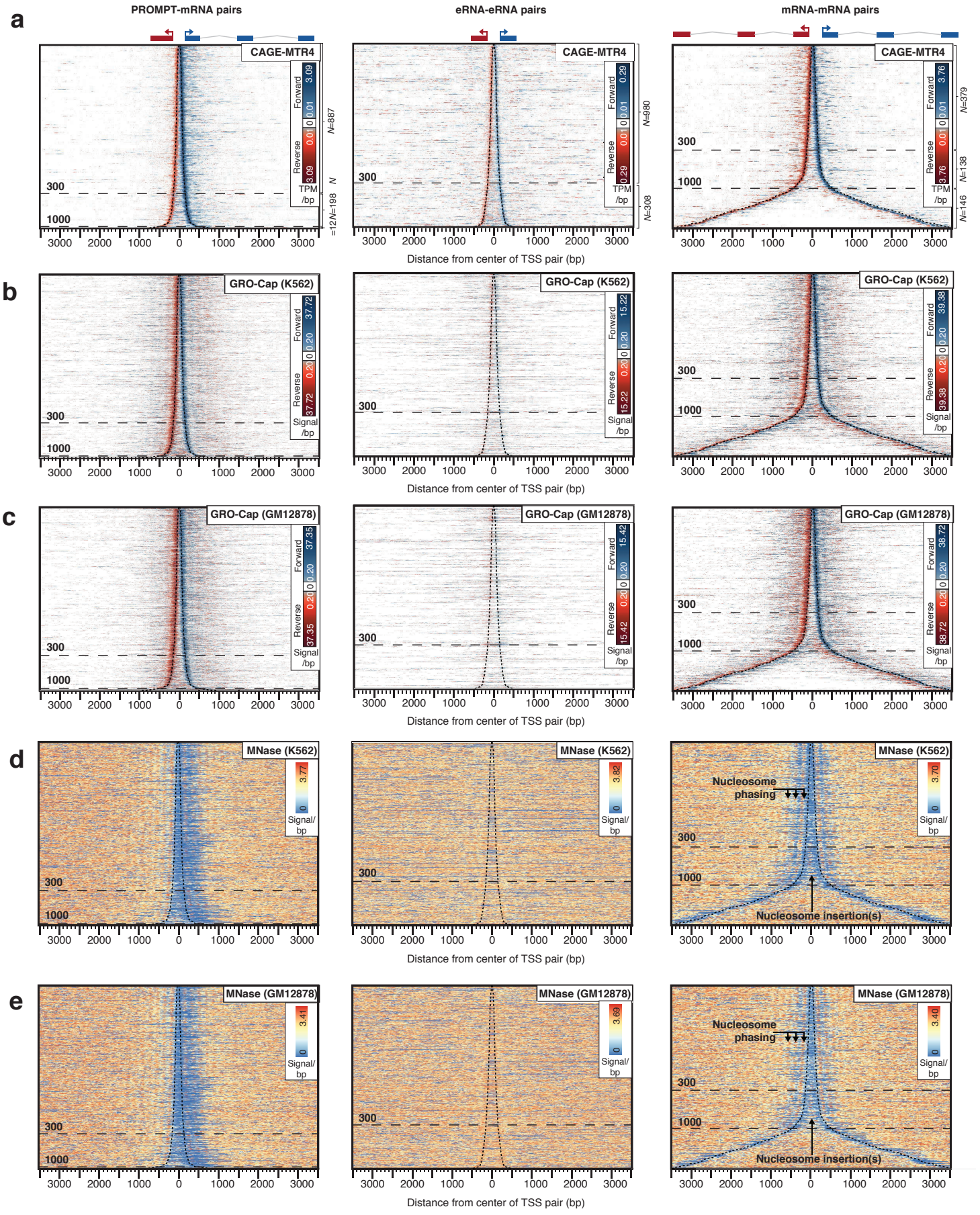


Supplementary Figure 1: Organization of divergent RNA-RNA TSS pairs

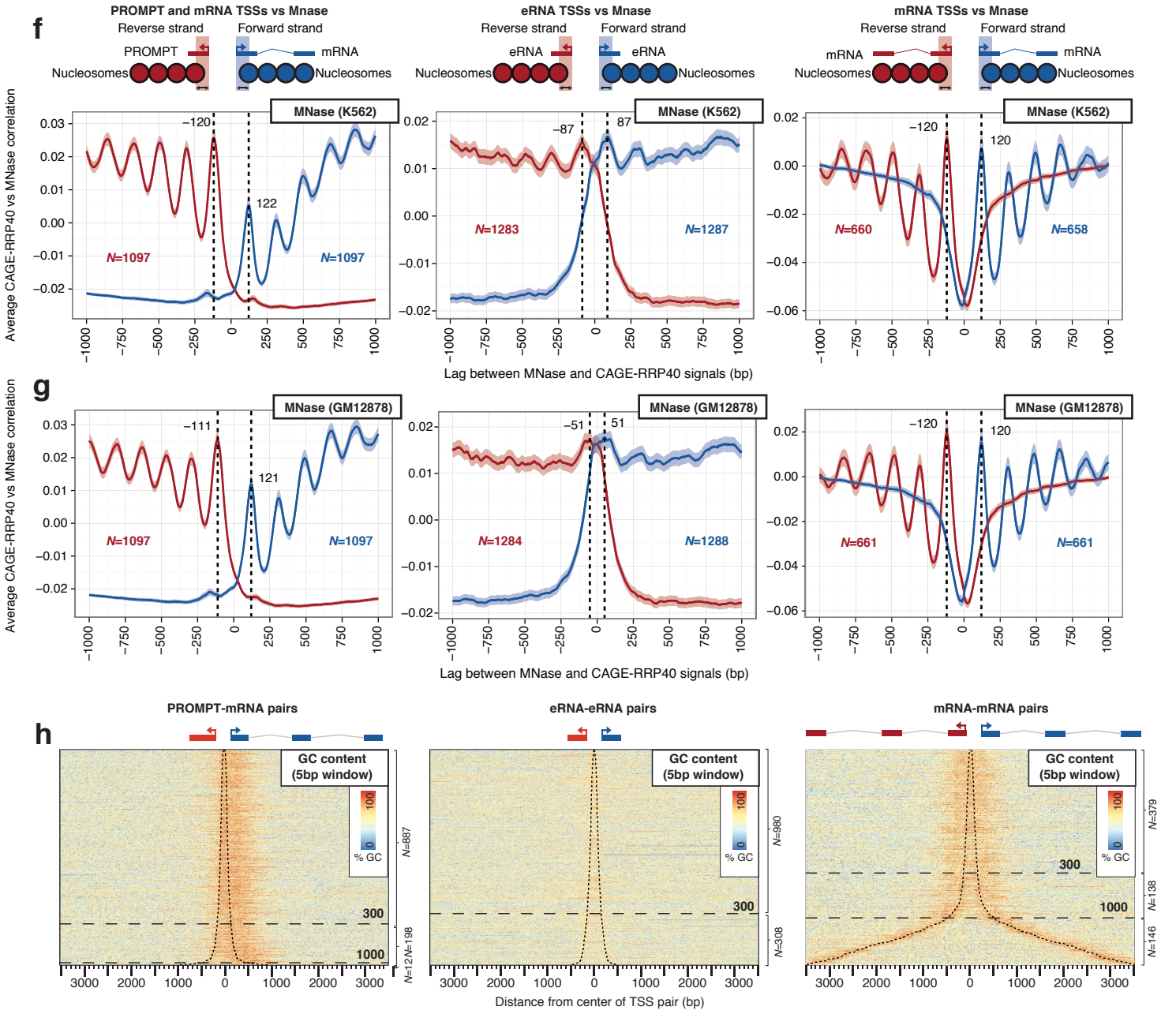
a: Heat maps showing forward (blue) and reverse (red) strand CAGE-MTR4 (CAGE following MTR4 depletion) signals at TSSs as in Fig. 2a. CAGE-defined TSS positions are marked with dashed black lines.

b: Heat maps organized as in (a), showing global nuclear run-on sequencing followed by cap-enrichment (GRO-Cap) data from K652 cells[1].

c: Heat maps organized as in (a), showing GRO-Cap data from GM12878 cells[1].

d: Heat maps organized as in (a), showing MNase cleavage data from K652 cells[2,3]. Note that ends of DNA regions protected by nucleosomes are sequenced, so in effect a high signal indicates nucleosome presence. Nucleosome phasing and insertions around mRNA-mRNA TSS pairs are indicated by black arrows.

e: Heat maps organized as in (a), showing MNase cleavage data from G12878 cells[2,3] as in (d).

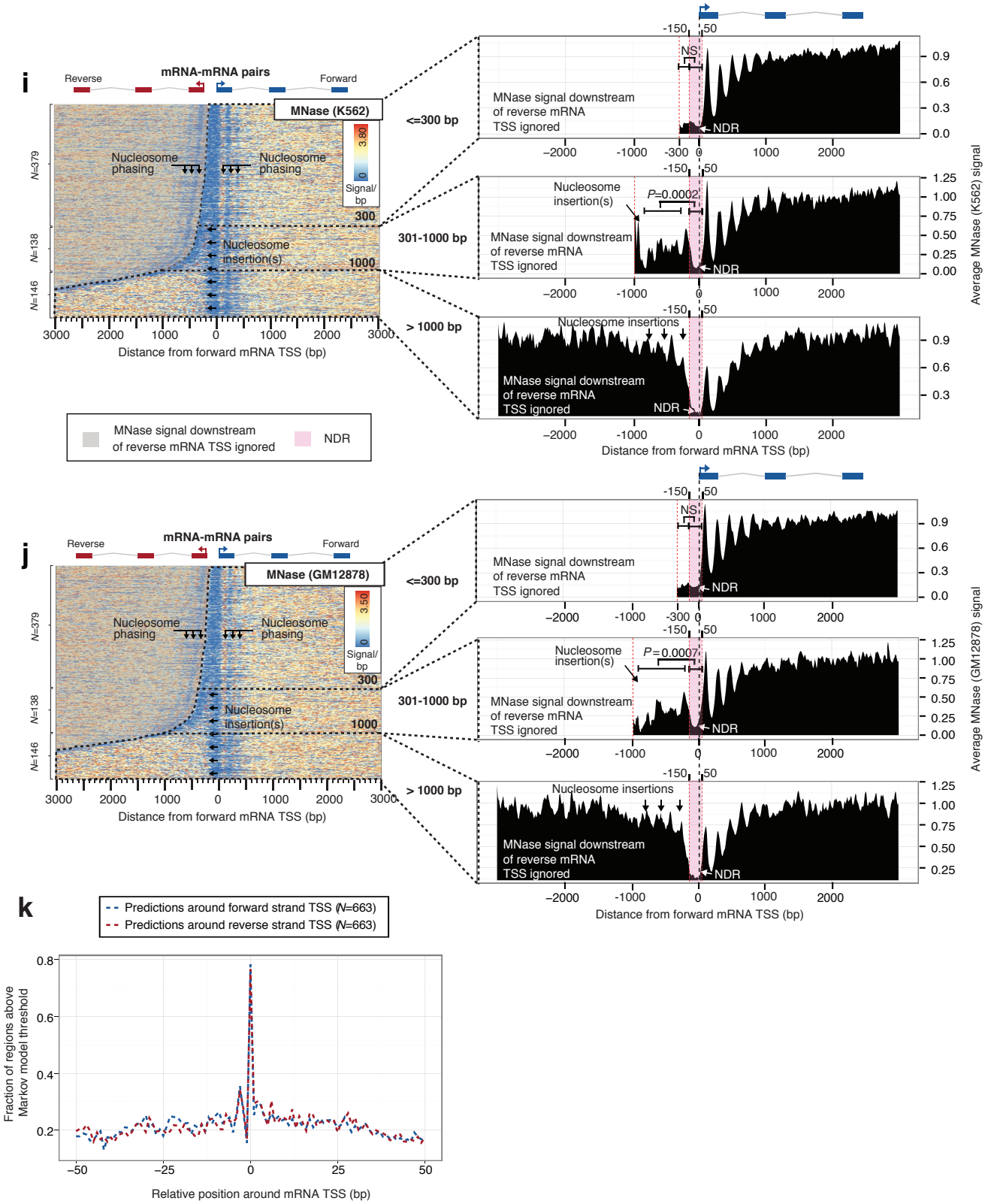


Supplementary Figure 1: Organization of divergent RNA-RNA TSS pairs

f: Cross-correlation analyses between CAGE-RRP40 TSS signals and MNase (K562) cleavage data for the RNA classes from (a). One dataset was slid across the other, recording the average Pearson correlation (Y axes) as a function of lag (X axes). The schematics on top show the analyzed TSSs, strands and regions. Red and blue boxes indicate the comparison between TSSs and their respective +1 nucleosomes, which in turn correspond to the correlation peaks indicated by dotted lines in the panels below. Numbers denote the CAGE-RRP40/MNase lags that produced these peaks.

g: Cross-correlation analyses between CAGE-RRP40 TSS signals and MNase (GM12878) cleavage data for the RNA classes from (a), shown as in (e).

h: Heat maps organized as in (a), showing GC content.

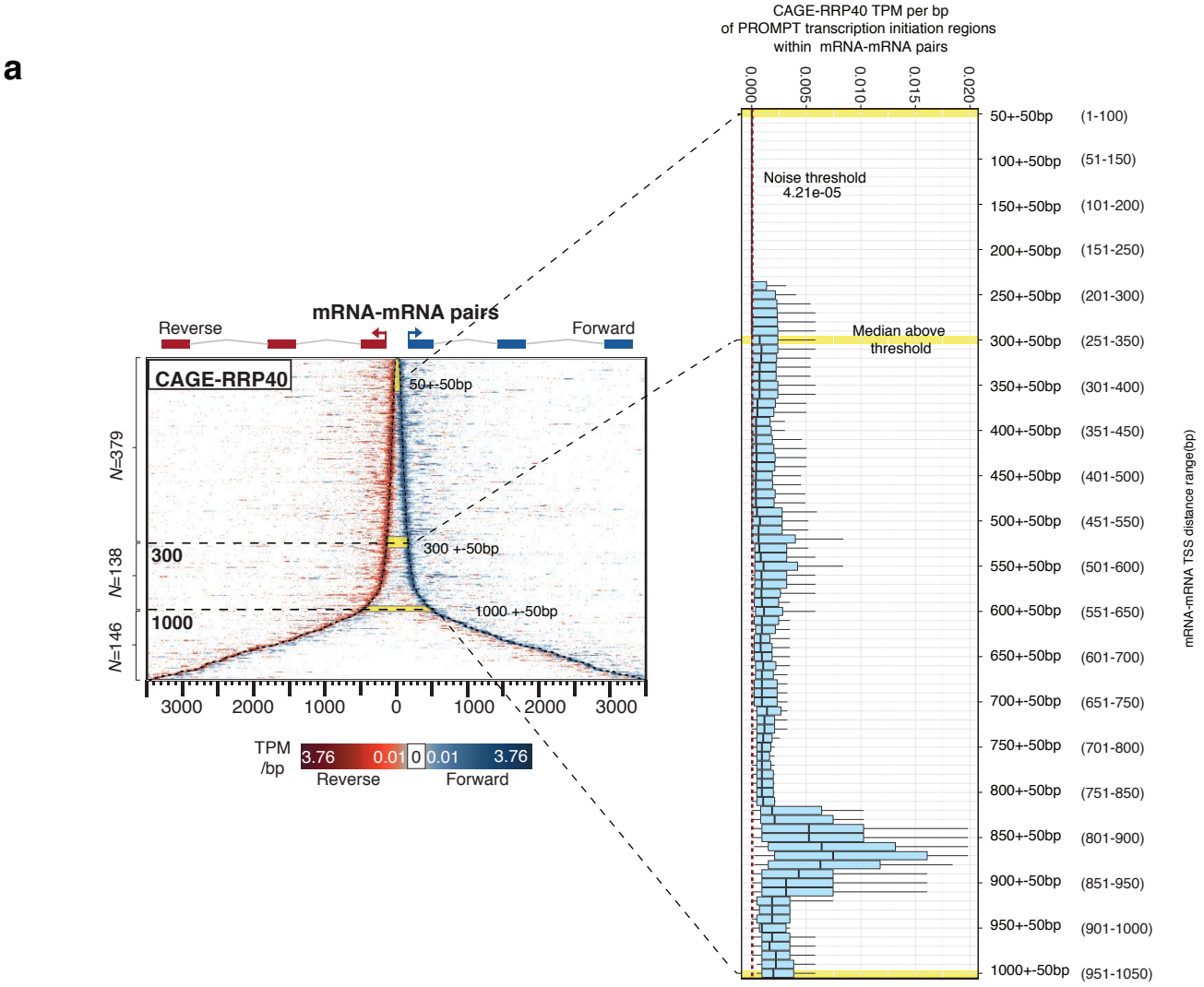


Supplementary Figure 1: Organization of divergent RNA-RNA TSS pairs

i: Nucleosome density between mRNA-mRNA TSSs. Left panel shows MNase (K562)[2] data across mRNA-mRNA TSS regions as in (d), but centered on the forward mRNA TSS. Nucleosome phasing and insertions are indicated by black arrows. The average MNase cleavage signals of the regions indicated by dotted lines are shown to the right, excluding signals downstream of the reverse mRNA TSS (excluded region shaded in grey in left panel). NDRs adjacent to the center point forward TSS and nucleosome insertions between the two mRNA TSSs are indicated. Statistical tests for the difference in MNase (K562) signal in the indicated NDR region (-150 to +50 of the center TSS) vs. the remaining region in between the mRNA TSSs are indicated (P values refer to two-sided Mann-Whitney tests; NS=non significant).

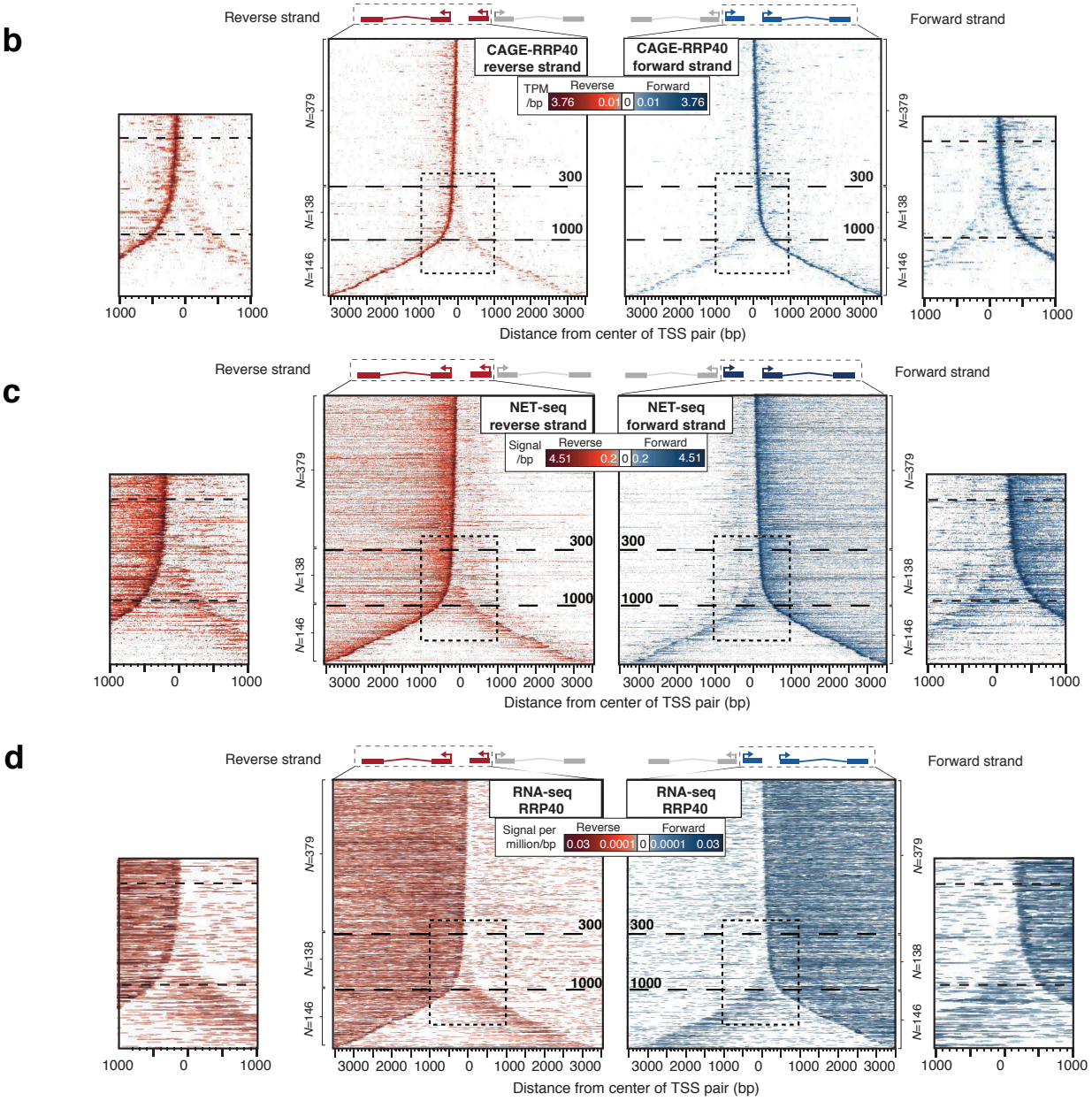
j: As in (i), but showing MNase data from GM12878 cells[2].

k: TSS position predictability based on DNA sequence around mRNA TSSs. A k-mer Markov Model trained on sharp CAGE TSSs as in[4] was slid over mRNA TSSs within divergent loci. Y-axis shows the fraction of regions that have a model score over 0 (corresponding to a higher 'TSS' than 'no TSS' likelihood). X-axis shows relative positions around mRNA TSSs. Forward (red) and reverse (blue) strand TSSs were analyzed separately but oriented so that transcription is left to right. Numbers of analyzed regions are indicated.



Supplementary Figure 2: PROMPT biogenesis, fate and properties within divergent mRNA TSSs constellations

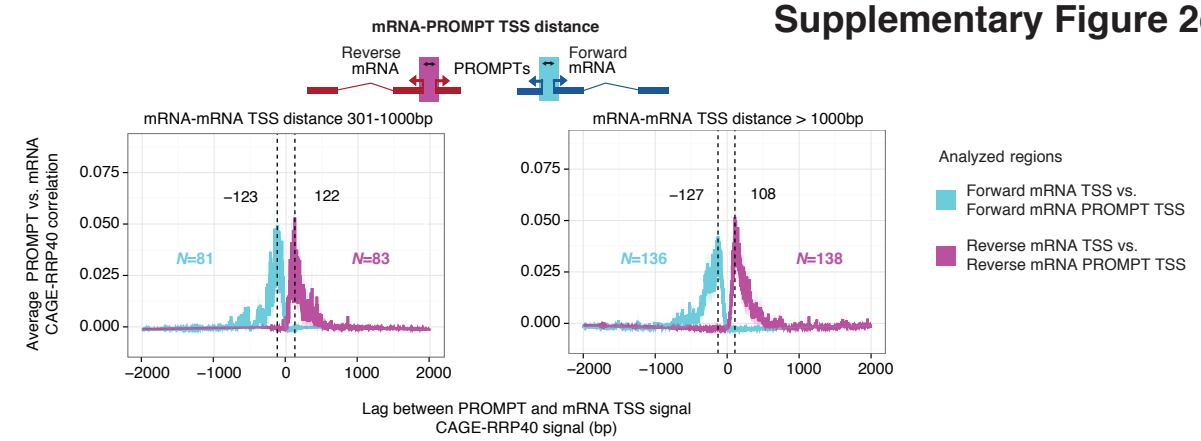
a: Rationale for the 300bp threshold selection. Starting from the CAGE-RRP40 mRNA-mRNA pairs heat map shown to the left (based on Fig. 2a), CAGE–RRP40 signals were analyzed within PROMPT transcription initiation regions split by mRNA-mRNA TSS distance. A windowing approach was used, exemplified by the yellow highlights in the left panel, corresponding to three of the analyzed windows. The first window includes PROMPT transcription initiation regions between mRNA TSSs separated by 1-100bp, while the highlighted middle and bottom windows include PROMPT transcription initiation regions between mRNA TSSs separated by 251-350bp and 951-1050bp, respectively. The right panel shows the distribution of CAGE-RRP40 signal (TPM/bp) of PROMPT transcription initiation regions for consecutive windows as blue boxplots. A CAGE-RRP40 noise threshold based on non-genic background sampling (see Methods) is indicated as a horizontal red dotted line. At the 251-350 bp separation window the median CAGE-RRP40 signal exceeds the noise threshold (see middle highlight in right panel). The midpoint of this distance range (300bp) was selected as the threshold for further analysis.



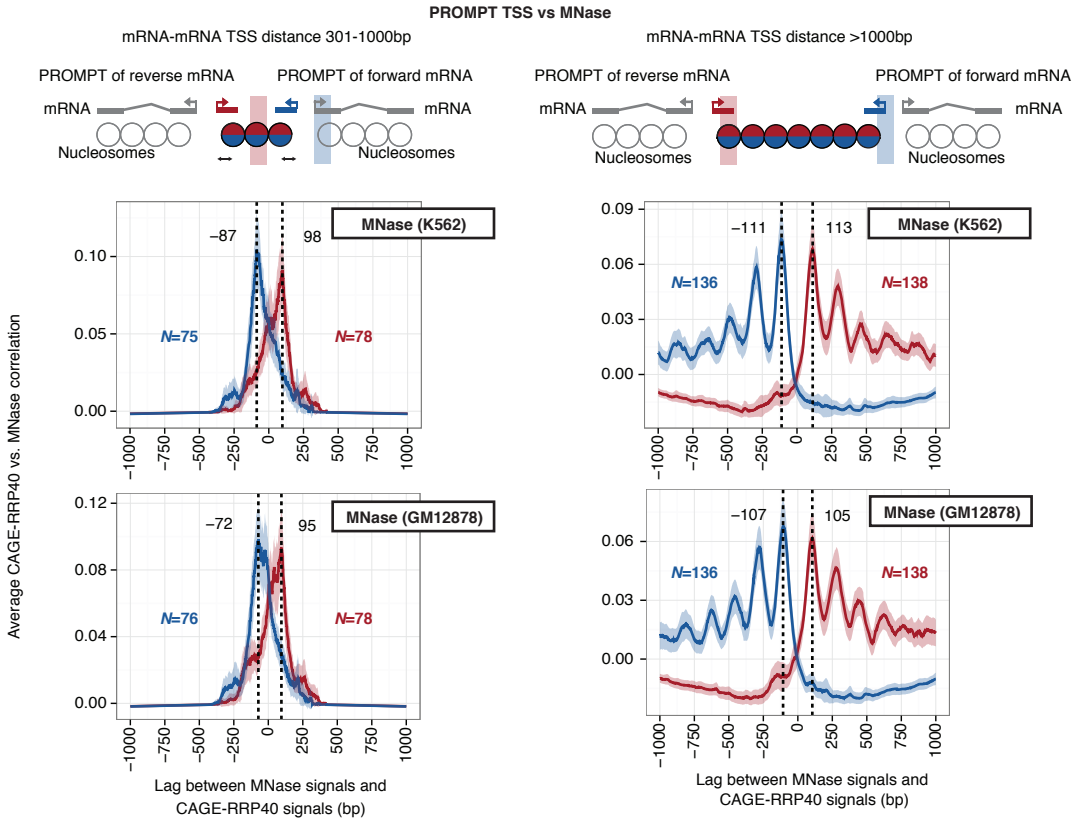
Supplementary Figure 2: PROMPT biogenesis, fate and properties within divergent mRNA TSSs constellations

b: Heat maps showing forward (blue) and reverse (red) strand CAGE-RRP40 data over mRNA-mRNA TSS regions organized as in Fig. 2a. Maps are split by reverse (left) and forward (right) strands. Zoom-ins of indicated areas are shown.
c: Heat maps as in (b) showing NET-seq data[5].
d: Heat maps as in (b) showing RNA-seq data from cells depleted for RRP40 (RNA-seq-RRP40)[6].

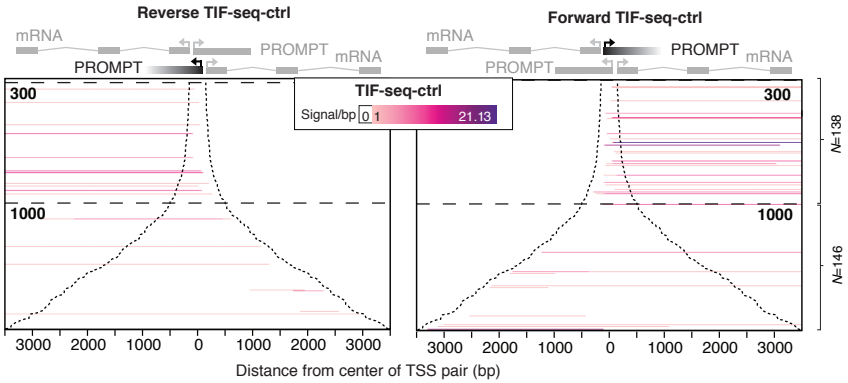
e



f



g



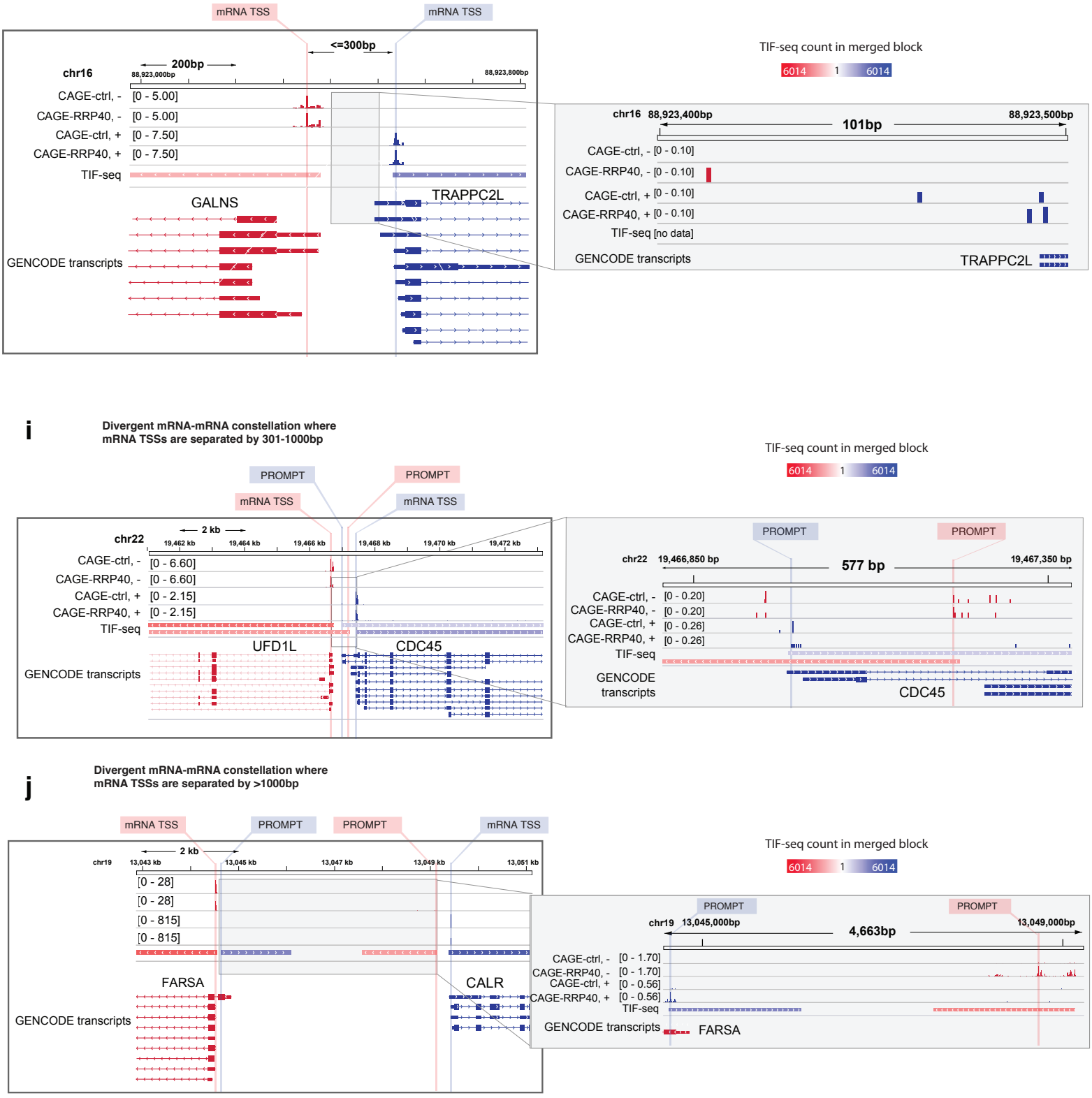
Supplementary Figure 2: PROMPT biogenesis, fate and properties within divergent mRNA TSSs constellations

e: Cross-correlation analyses between CAGE-RRP40 signals corresponding to TSSs of mRNAs and PROMPTs at opposite strands within divergent mRNA-mRNA TSS regions separated by 301-1000bp (left panel) or >1000bp (right panel). Schematic on top shows the analyzed TSSs, strands and regions. Red and blue boxes indicate the comparison between respective TSSs, which in turn corresponds to the correlation peaks indicated by dotted lines in the panels below. Numbers denote the lags that produced these peaks. Y-axes show average correlation between mRNA and PROMPT TSS signals as a function of lag on the X axes. Dashed lines and numbers indicate lags with the highest correlation.

f: Cross-correlation analysis as in Supplementary Fig. 1f-g, but correlating PROMPT CAGE-RRP40 with MNase (K652)[2] (upper row) and MNase (GM12878)[2] (lower row) signals downstream of the PROMPT region within divergent mRNA-mRNA TSS constellations, as indicated by the schematics on top. Analyses were split by mRNA-mRNA TSS distance (left panel: 301-1000bp, right panel: >1000bp). Schematics on top show the analyzed TSSs, strands and regions.

g: Heat maps of reads derived from transcript isoform sequencing of RNA from control cells (TIF-seq-ctrl) initiating within PROMPT transcription initiation regions of forward (left panel) and reverse (right panel) mRNA strands. Heat maps were organized as in Fig. 3b.

Supplementary Figure 2h-j



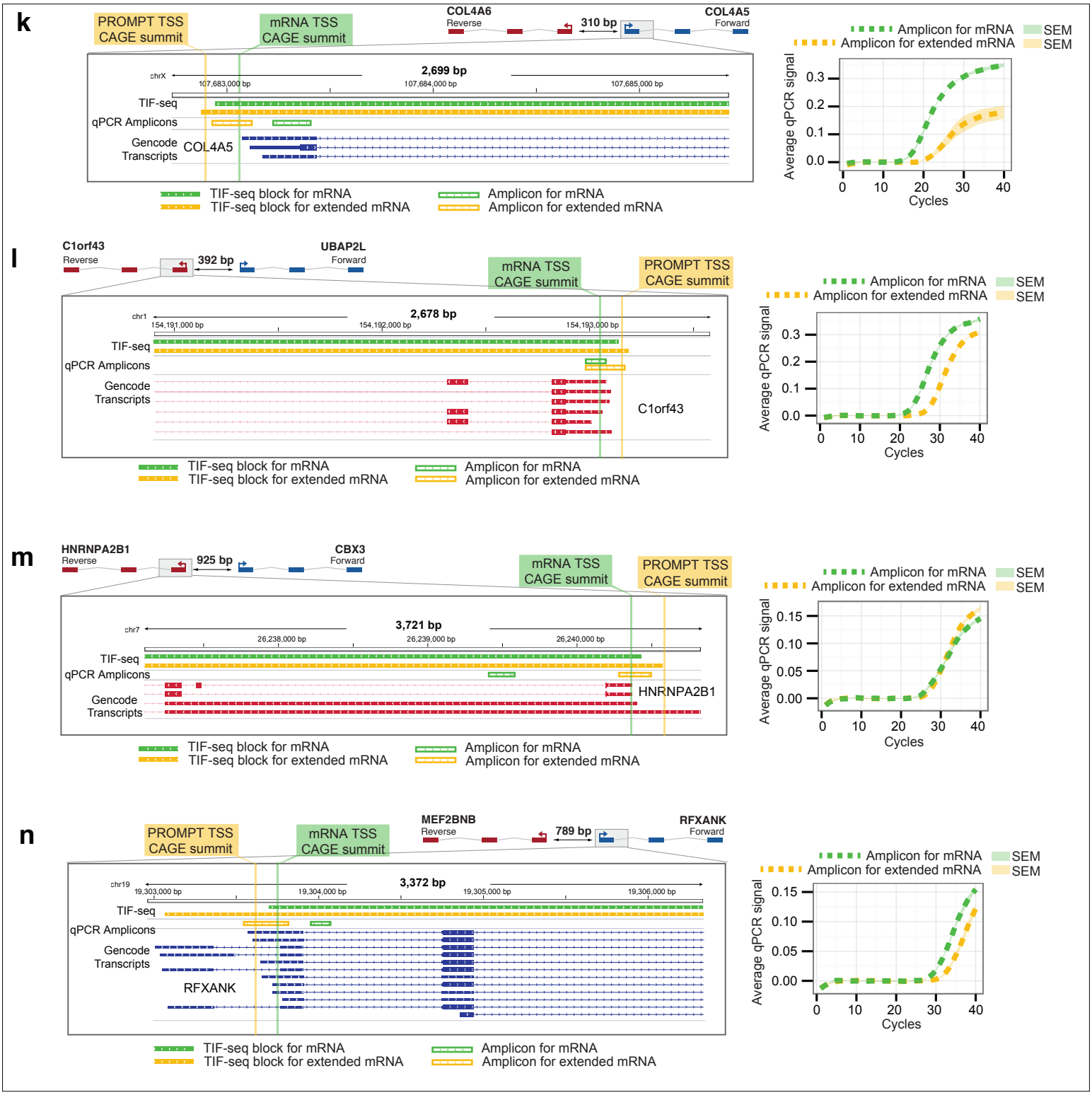
Supplementary Figure 2: PROMPT biogenesis, fate and properties within divergent mRNA TSSs constellations

h: Genome-browser example of a divergent mRNA-mRNA promoter (GALNS and TRAPPC2L genes) with a TSS separation ≤300bp, visualized by the IGV browser[7]. CAGE TPM intensities per bp are shown as bar plots with ranges in brackets. TIF-seq-RRP40+ZCCHC8 data are shown as TSS-associated blocks (see Methods) linking their most upstream 5'ends and most downstream 3'ends, and with color intensity denoting the relative number of reads. Reverse (hg19 minus) and forward (hg19 plus) strand data are shown in red and blue, respectively. mRNA TSSs, as defined by CAGE data, are indicated on top with red and blue callouts. GENCODE v17 transcripts are shown below. A zoom-in is shown in grey. Note that no substantial PROMPT signal is present.

i: Genome-browser example of a divergent mRNA-mRNA constellation (UFD1L and CDC45 genes) with TSSs separated by 301-1000bp, organized as in (h). PROMPT TSSs are visible in the zoom-in and indicated on top by red and blue callouts.

j: Example of a divergent mRNA-mRNA constellation (FARSA and CALR genes) with TSSs separated by >1000bp, and organized as in (h).

Supplementary Figure 2k-n



Supplementary Figure 2: PROMPT biogenesis, fate and properties within divergent mRNA TSSs constellations

k: RT-qPCR analysis showing that the PROMPT of the COL4A6 mRNA TSS extends into the nearby COL4A5 gene to create a 5'extended COL4A5 mRNA. Schematic on top shows the gene constellation and distance between annotated mRNA TSSs. Left panel shows a genome-browser zoom-in of data on the same strand as the PROMPT. CAGE summits for PROMPT and annotated mRNA TSSs are indicated by yellow and green highlights, respectively. TIF-seq-RRP40+ZC-CHC8 blocks associated to PROMPT or annotated mRNA are colored yellow and green, respectively. qPCR amplicons measuring 5'extended mRNA and annotated mRNA are outlined in yellow and green, respectively. GENCODE transcript annotations on the same strand are shown as in (h). Right panel shows qPCR amplification curves of HeLa cell derived cDNA for the two amplicons depicted in the left panel, colored accordingly. Y-axis shows qPCR signal. X-axis shows qPCR cycle. Dotted lines indicate means across biological triplicates. Standard error of the mean (SEM) estimates are shown as shaded ribbons.

l: RT-qPCR analysis showing that the PROMPT of the UBAP2L mRNA TSS extends into the nearby C1orf43 gene, organized as in (k).

m: RT-qPCR analysis showing that the PROMPT of the CBX3 mRNA TSS extends into the nearby HNRNPA2B1 gene, organized as in (k).

n: RT-qPCR analysis showing that the PROMPT of the MEF2BNB mRNA TSS extends into the nearby RFXANK gene, organized as in (k).

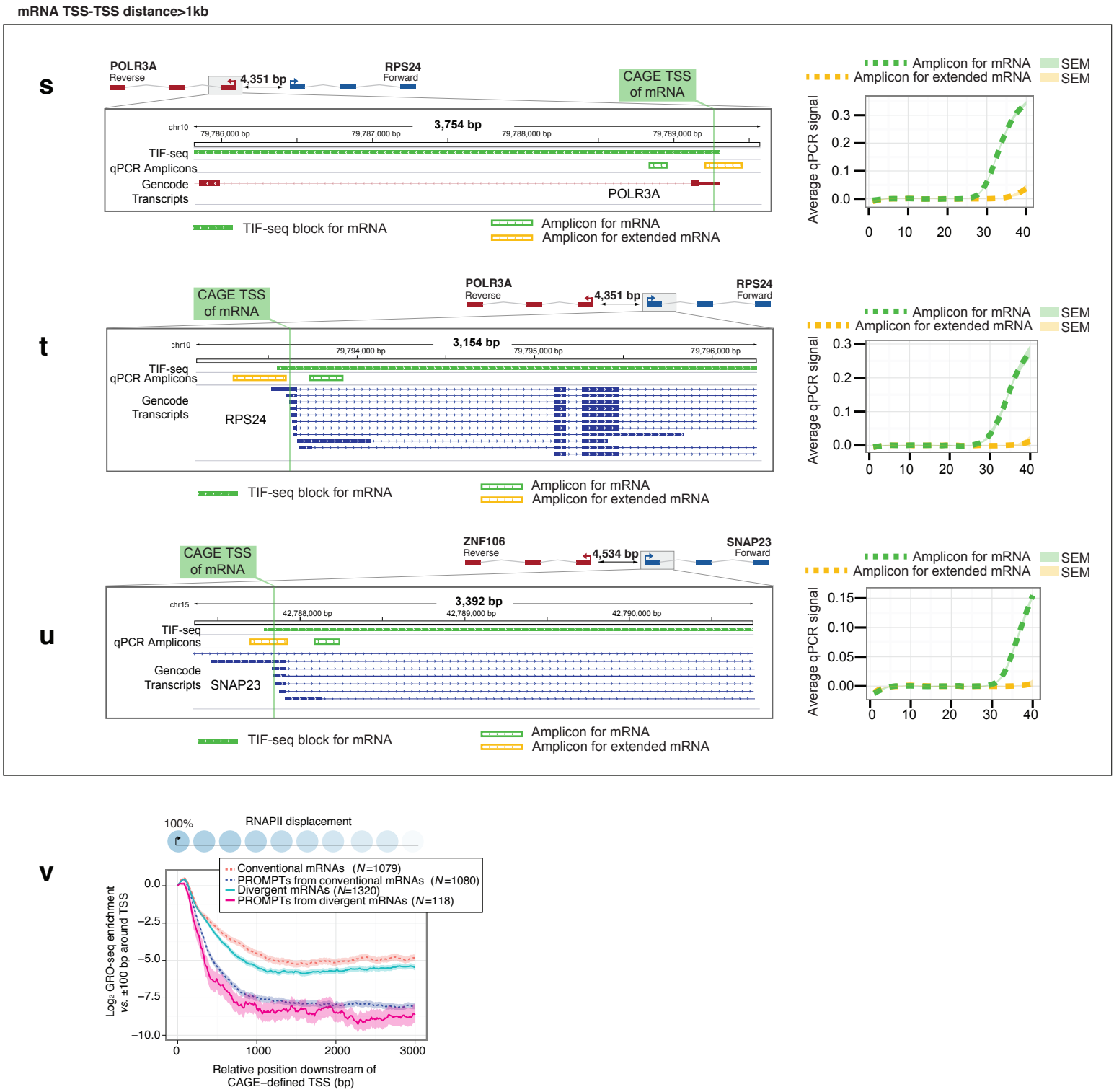
Supplementary Figure 2: PROMPT biogenesis, fate and properties within divergent mRNA TSSs constellations

o: RT-qPCR analysis showing that the PROMPT of the UFD1L mRNA TSS extends into the nearby CDC45 gene, organized as in (k).

p: RT-qPCR analysis showing that the PROMPT of the CCDC115 mRNA TSS extends into the nearby IMP4 gene, organized as in (k).

q: RT-qPCR analysis showing that the PROMPT of the FARSA mRNA TSS does not extend into the distal CALR gene, organized as in (k).

r: RT-qPCR analysis showing that the PROMPT of the CALR mRNA TSS does not extend into the distal FARSA gene, organized as in (k).



Supplementary Figure 2: PROMPT biogenesis, fate and properties within divergent mRNA TSSs constellations

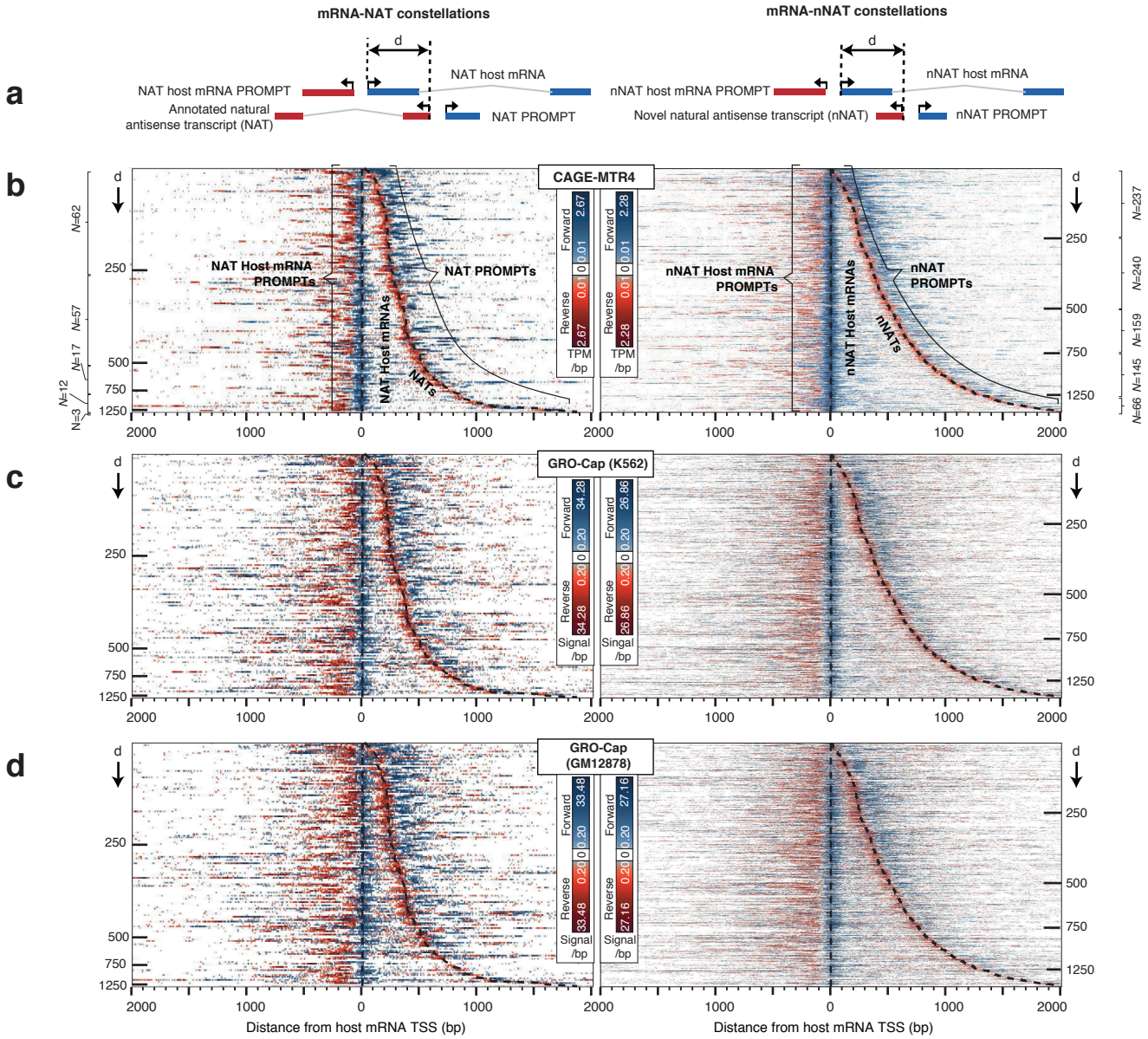
s: RT-qPCR analysis showing that the PROMPT of the RPS24 mRNA TSS does not extend into the distal POLR3A gene, organized as in (k).

t: RT-qPCR analysis showing that the PROMPT of the POLR3A mRNA TSS does not extend into the distal RPS24 gene, organized as in (k).

u: RT-qPCR analysis showing that the PROMPT of the ZNF106 mRNA TSS does not extend into the distal SNAP23 gene, organized as in (k).

v: Same as Fig. 3d but showing Global nuclear run-on sequencing (GRO-seq)[8].

Supplementary Figure 3a-d



Supplementary Figure 3: Organization and properties of TSS pairs within NAT and nNAT constellations

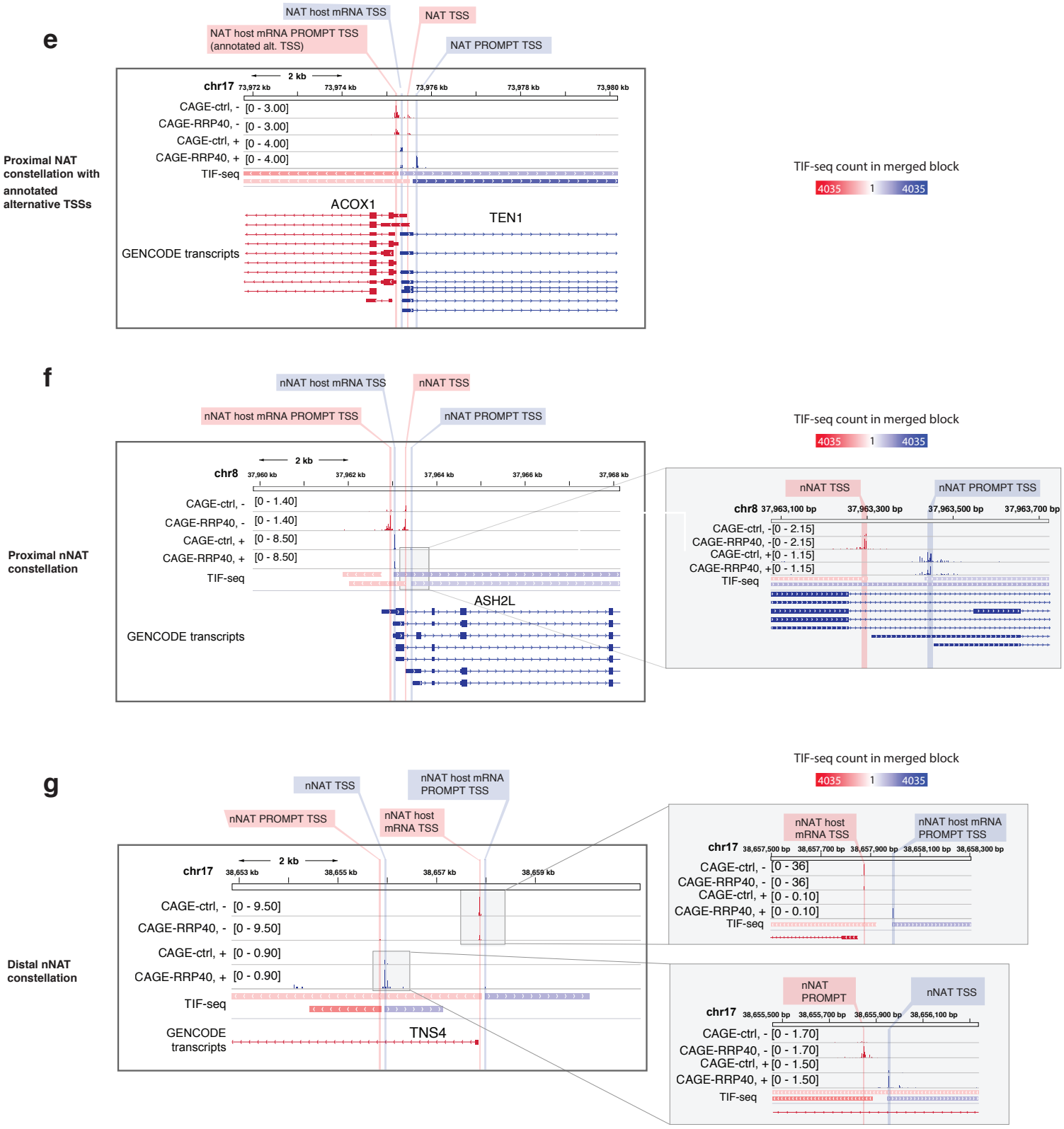
a: Schematic overview of analyzed loci (as in Fig. 4a).

b: Heat maps organized as in Fig. 4b, showing CAGE-MTR4 data.

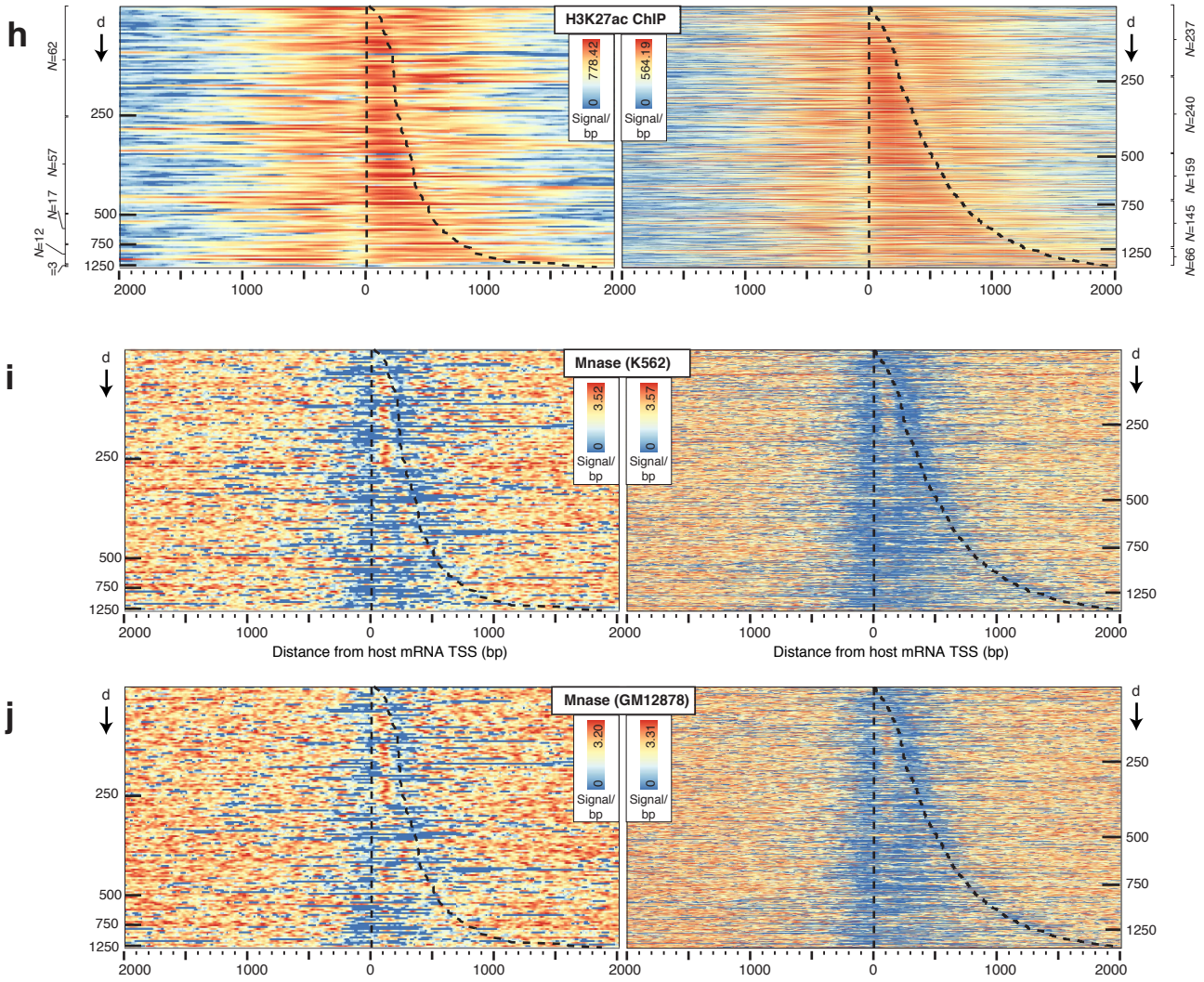
c: Heat maps organized as in Fig. 4b, showing GRO-cap(K562)[1] data.

d: Heat maps organized as in Fig. 4b, showing GRO-cap(GM12878)[1] data.

Supplementary Figure 3e-g



Supplementary Figure 3: Organization and properties of TSS pairs within NAT and nNAT constellations
e: Genome-browser example of an mRNA-NAT constellation (ACOX and TEN1 genes), organized as in Supplementary Fig. 2h.
f: Genome-browser example of a proximal mRNA-nNAT constellation (ASHL as host gene), organized as in (e).
g: Genome-browser example of a distal mRNA-nNAT constellation (TNS4 as host gene), organized as in (e).

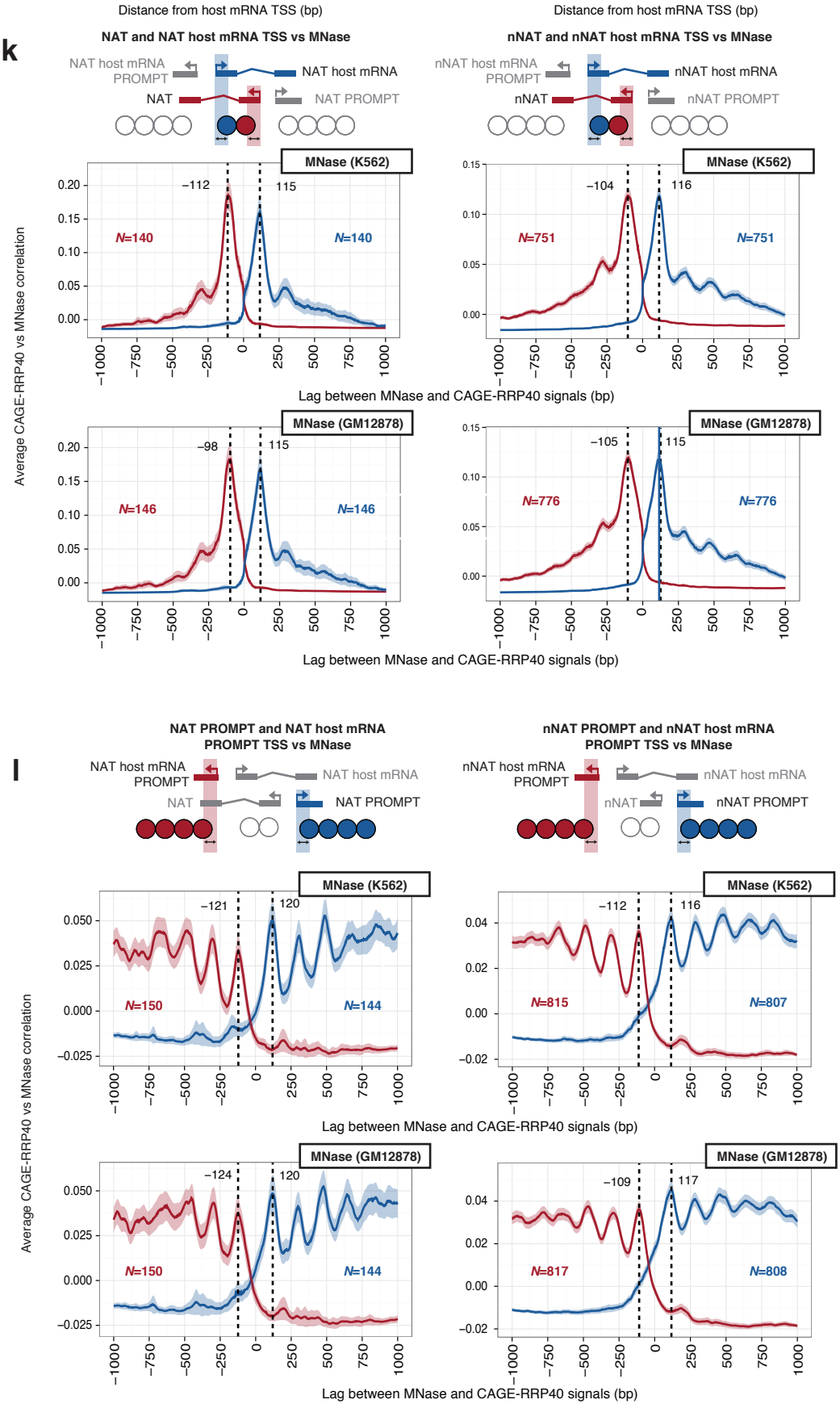


Supplementary Figure 3: Organization and properties of TSS pairs within NAT and nNAT constellations

h: Heat map organized as in (b), showing ENCODE H3K27ac ChIP data[2].

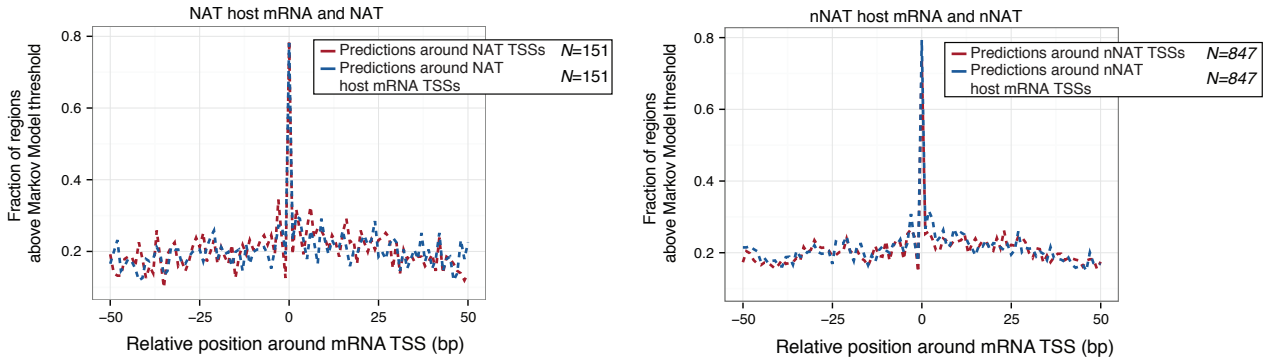
i: Heat map organized as in (b), showing MNase (K562) data[2].

j: Heat map organized as in (b), showing MNase (GM12878) data[2].



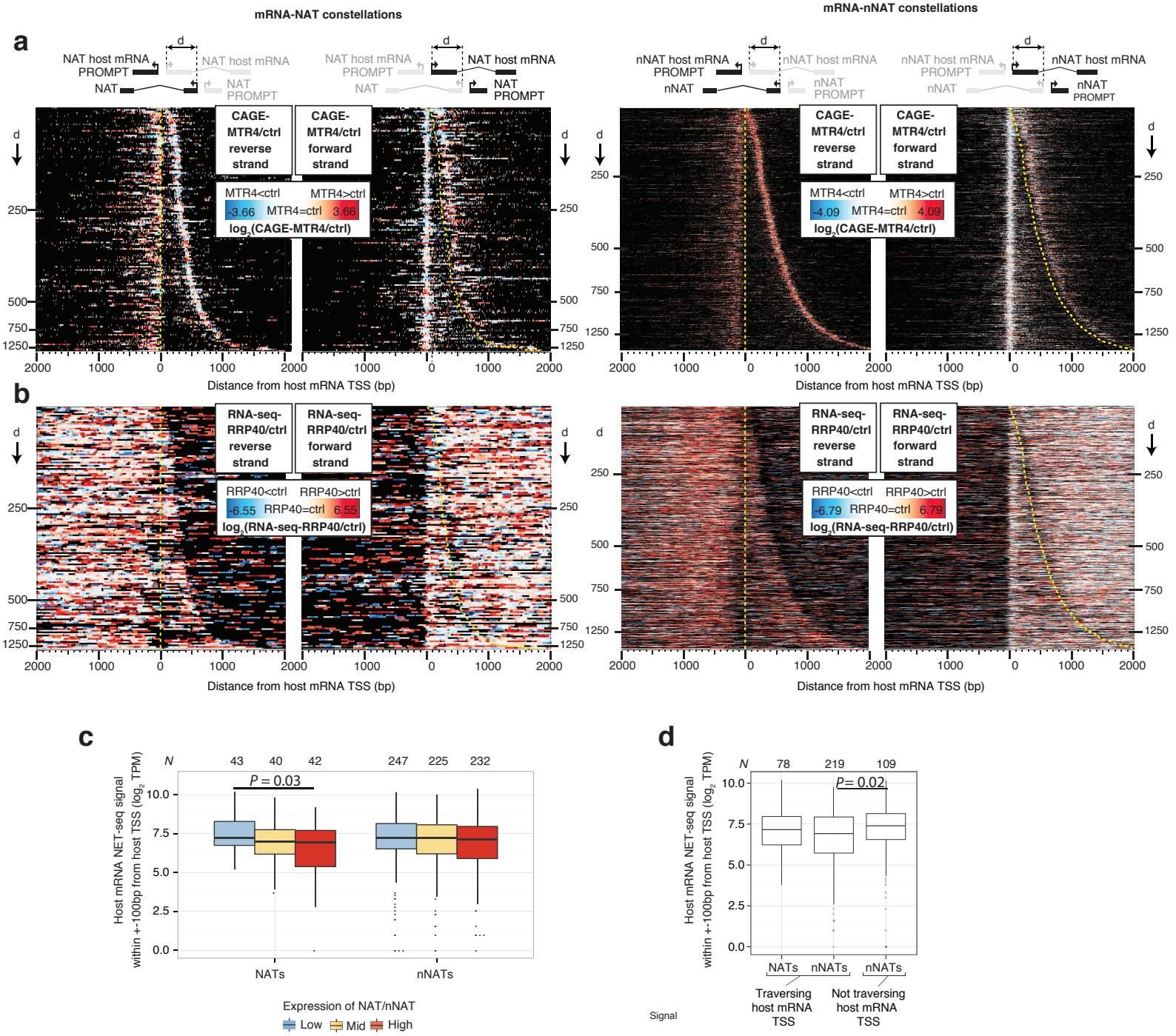
Supplementary Figure 3: Organization and properties of TSS pairs within NAT and nNAT constellations
k: Cross-correlation analysis as in Supplementary Fig. 2f, but correlating CAGE-RRP40 at NAT/nNAT (red) and NAT/nNAT host mRNA (blue) TSSs with MNase (K562 as top panel and GM12878 as lower panel)[2] signal downstream of respective TSSs. Left and right panels show NAT and nNAT constellations, respectively. Schematics on top show which TSSs, strands and regions that were analyzed.
l: Cross-correlation analysis as in (k), but correlating CAGE-RRP40 at NAT/nNAT PROMPT (blue) and NAT/nNAT host mRNA PROMPT (red) TSSs with MNase (K562/GM12878)[2] signal downstream of respective TSSs, as indicated by the schematics on top.

m



Supplementary Figure 3: Organization and properties of TSS pairs within NAT and nNAT constellations

m: TSS position predictability as in Supplementary Fig. 1k based on DNA sequence around NAT/NAT host mRNA (left panel) and nNAT/nNAT host mRNA (right panel) TSSs.



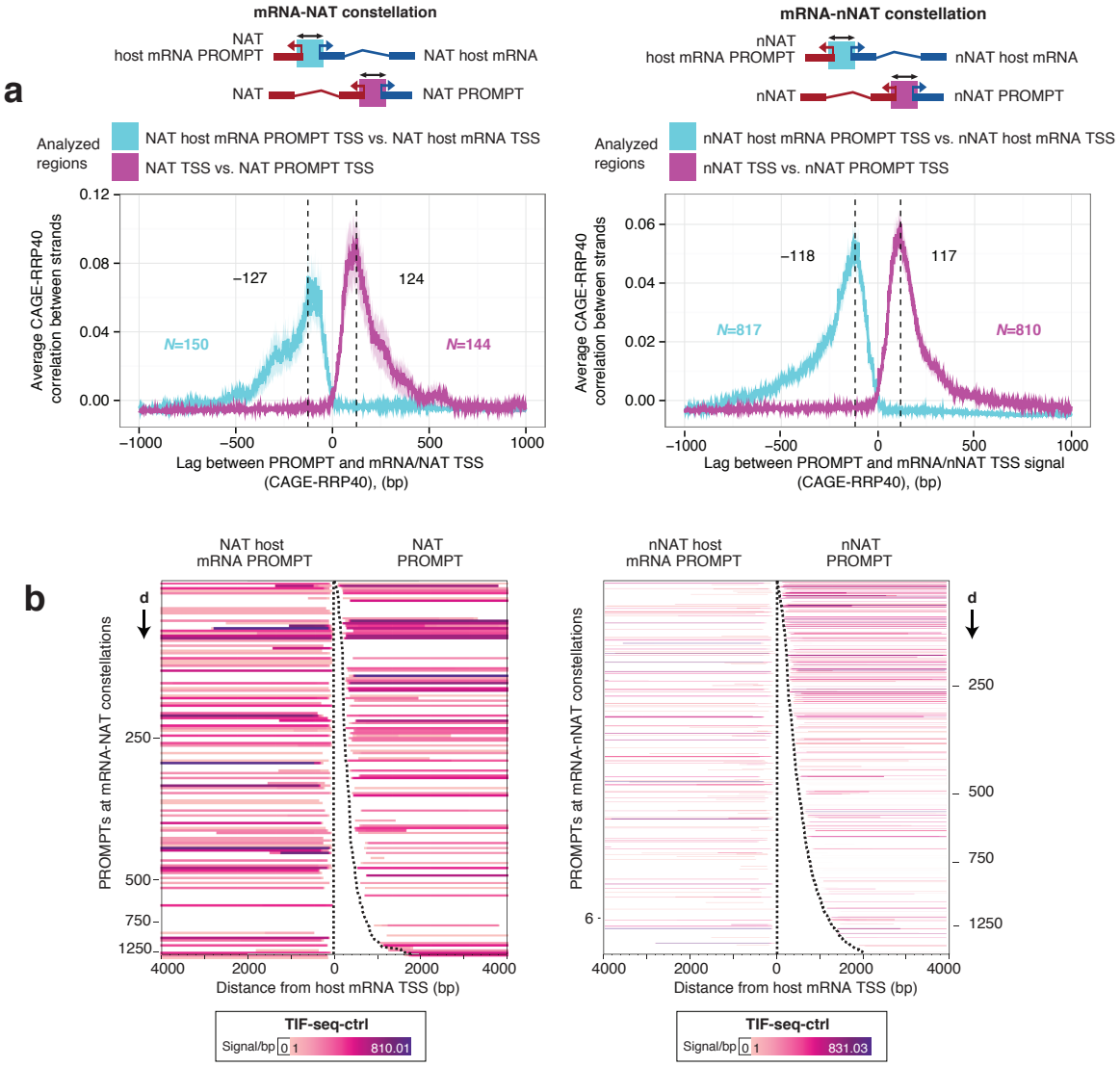
Supplementary Figure 4: Properties of NATs and nNATs

a: Heat maps organized as in Fig. 5a, showing $\log_2$ CAGE-MTR4/ctrl ratios.

b: Heat maps organized as in (a), showing $\log_2$ RNA-seq-RRP40/ctrl ratios.

c: Relation between host mRNA and NAT/nNAT levels as in Fig. 5d, but displaying NET-seq signals measured within the +/-100bp region around host mRNA TSSs. Boxplots were split by CAGE-RRP40 levels of NAT (left) and nNAT (right) TSSs as in Fig. 5d. P-value indicates Mann-Whitney two-sided tests between distributions.

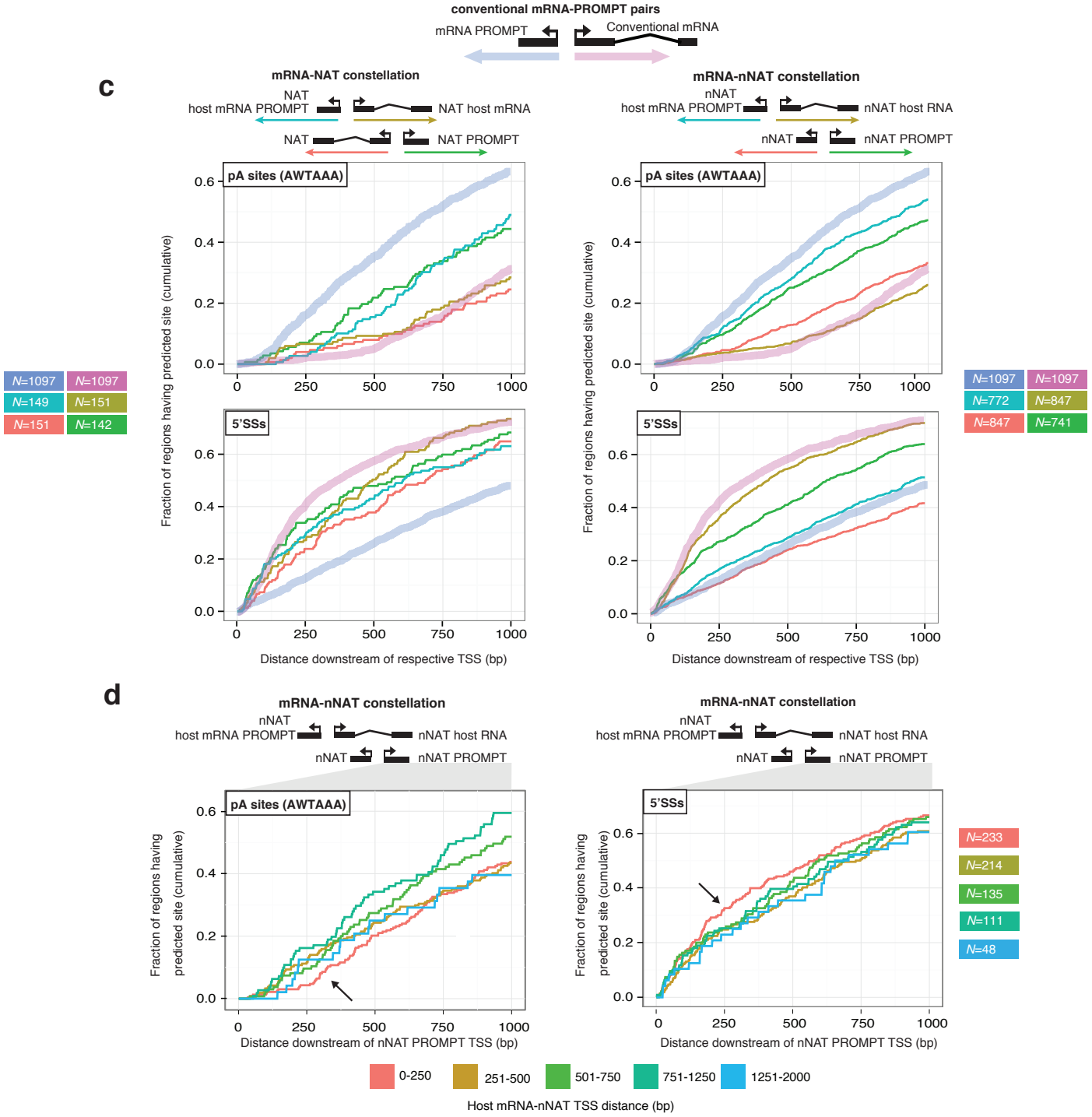
d: Relation between host mRNA levels and NAT/nNAT 'traversal' of the host mRNA TSS as in Fig. 5e, but displaying NET-seq signals measured within the +/-100bp region around host mRNA TSSs. Significance levels as in (c).



Supplementary Figure 5: PROMPT generation and properties within convergent loci constellations

a: Cross-correlation analyses as in Supplementary Fig. 2e, but showing the analyses between CAGE-RRP40 signals corresponding to TSSs of host mRNAs and their PROMPTs (cyan) as well as NATs/nNATs and their respective PROMPTs (pink). Schematics on top show analyzed TSSs, strands and regions. Left and right panels show NAT and nNAT constellations, respectively.

b: Heat maps showing length distributions of PROMPTs within NAT/nNAT constellations as in the bottom panel of Fig. 6b, but using TIF-seq-ctrl reads.



Supplementary Figure 5: PROMPT generation and properties within convergent loci constellations

c: Occurrences of predicted pA sites (top panels) and 5′SSs (bottom panels), as in Fig. 3f but downstream of the indicated TSSs within mRNA-NAT (left) or mRNA-nNAT (right) constellations as indicated by the color code at the top. Broad lines show corresponding analyses downstream of conventional mRNAs (pink) and their PROMPTs (blue) for reference.

d: Occurrences of predicted pA sites (left panel) and 5′SSs (right panel), analyzed as in (c) but downstream of nNAT PROMPT TSSs, and split by host mRNA-nNAT TSS-TSS distance as indicated by the color legend below the plots. Note that PROMPTs that are most proximal to the host mRNA TSS harbor the lowest and highest incidences of downstream pA sites and 5'SSs, respectively (red lines, indicated by black arrows).

Supplementary figure legend references

1. 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).
2. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012).
3. Thurman, R. E. et al. The accessible chromatin landscape of the human genome. Nature 489, 75–82 (2012).
4. Frith, M. C. et al. A code for transcription initiation in mammalian genomes. Genome Research 18, 1–12 (2007).
5. Mayer, A. et al. Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. Cell 161, 541–554 (2015).
6. Ntini, E. et al. Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. Nat Struct Mol Biol 20, 923–928 (2013).
7. Robinson, J. T. et al. Integrative genomics viewer. Nature Biotechnology 29, 24–26 (2010).
8. Andersson, R. et al. Nuclear stability and transcriptional directionality separate functionally distinct RNA species. Nat Comms 5, 5336–5336 (2014).

Supplementary Table 1: Data sources

Each row shows metadata for each data set employed in this study. This includes, from left to right columns, experimental method, dataset name, original assembly the dataset was mapped to, cell line origin, GEO accession number(s)/Short Read Archive accession numbers/ENCODE experiment id, and whether the data set was produced by pooling two or more libraries. If the latter was true, the Pearson correlation (R^2) between libraries on each strand is indicated. If more than two libraries were compared, the average R^2 value is shown.

Method	Dataset name	Originally Mapped Assembly	Cell Line	Data source	Pooled	R^2
CAGE	CAGE-ctrl	hg19	HeLa	GSE48286-GSM1174214	F	NA
CAGE	CAGE-RRP40	hg19	HeLa	GSE48286-GSM1174215	F	NA
CAGE	CAGE-MTR4	hg19	HeLa	GSE49834	F	NA
RNA-seq	RNA-seq-ctrl	hg19	HeLa	GSE48286-GSM1174216	F	NA
RNA-seq	RNA-seq-RRP40	hg19	HeLa	GSE48286-GSM1174217	F	NA
TIF-seq	TIF-seq-ctrl_rep1	hg19	HeLa	GSE75183	T	Plus: 0.88; Minus:0.84;
TIF-seq	TIF-seq-ctrl_rep2	hg19	HeLa	GSE75183		
TIF-seq	TIF-seq-RRP40+ZCCH8	hg19	HeLa	GSE75183	F	NA
GRO-cap	GRO-cap K562	hg19	K562	GSE60456-GSM1480321	F	NA
GRO-cap	GRO-cap GM12878	hg19	GM12878	GSE60456-GSM1480323	F	NA
Dnase-seq	Dnase-seq	hg19	HeLa	ENCSR959ZXU	processed as in Mayer et al, Cell 2015	NA
NET-seq	NET-seq	hg19	HeLa	GSE61332-GSM1505438 and GSM1505439	processed as in Mayer et al,	NA

					Cell 2015	
GRO-seq	GRO-seq	hg19	HeLa	GSE62046	processed as in Andersson et al, Nat Commun 2014	NA
ChIP-seq	H3K27ac	hg19	HeLa	GSE29611-GSM733684	F	NA
ChIP-seq	H3K3me1	hg19	HeLa	GSE29611-GSM798322	F	NA
ChIP-seq	H3K3me3	hg19	HeLa	GSE29611-GSM733682	F	NA
Mnase	Mnase K562	hg19	K562	GSE35586-GSM920557	F	NA
Mnase	Mnase GM12878	hg19	GM12878	GSE35586-GSM920558	F	NA
small RNA-seq	HeLa18-30-ctrl	hg18	HeLa	GSE29116-GSM721073	T (pooled libraries from different conditions)	Average Plus:0.63; Average Minus:0.63
small RNA-seq	HeLa18-30-RRP40	hg18	HeLa	GSE29116-GSM721074		
small RNA-seq	HeLa18-30-XRN1/2	hg18	HeLa	GSE29116-GSM721077		
small RNA-seq	HeLa18-30-N	hg18	HeLa	GSE29116-GSM721078		
ChIP-exo	RNAPII	raw reads	K562	SRR770759 and SRR770760	processed as in Andersson et al, Molecular Cell, 2015	NA
ChIP-exo	TBP	raw reads	K562	SRR770743 and SRR770744	processed as in Andersson et al, Molecular	NA

					Cell, 2015	
ChIP-exo	TFIIB	raw reads	K562	SRR770745 and SRR770746	processed as in Andersson et al, Molecular Cell, 2015	NA

Supplementary Table 2: qPCR primers and genomic locations of amplicons

qPCR primers sequences used in Supplementary Fig. 2k-u.
Primers corresponding to the extended isoform (i.e. the PROMPT of the upstream gene TSS) is denoted ‘extended’.

Primer name	Primer sequence
FARSA_extended_for	ATGACTCCTTCCAGTGTGCTC
FARSA_extended_rev	GTCCCATGCACCTTTTGCG
FARSA_mRNA_for	CCTCGTCCATCATCGTACCC
FARSA_mRNA_rev	ACGTCCAAGCCAAATTCCCT
CALR_extended_for	GAGGCGCAGCCTAACATAGT
CALR_extended_rev	TTTAAAACGACCCTCCGGCA
CALR_mRNA_for	CGTTTAGAGGTCCAACACGGT
CALR_mRNA_rev	AGGAAAGGAGATCCCCGACA
SNAP23_extended_for	GTGAATTCCCAGACTCTCCGT
SNAP23_extended_rev	TGGAGACTCACCAAGCGGG
SNAP23_mRNA_for	GGAGAACATCGAGGGACCAG
SNAP23_mRNA_rev	CTGTGGTCACCATACAGGCA
POLR3A_extended_for	TAGAGAAACGATGCCCCCAG
POLR3A_extended_rev	AAATACAGCACACGCCTCGG
POLR3A_mRNA_for	AGGAGAATGCGAGCTTTGGC
POLR3A_mRNA_rev	GGTCAGGCTGAGGGGATTTC
RPS24_extended_for	ACCCACCTCTAACTCGGAAGA
RPS24_extended_rev	CATATCAACGCGCACGGAAG
RPS24_mRNA_for	TGAGCTATAGGCACGCGAAG
RPS24_mRNA_rev	CGAAGTCAACAAGCGCAACA

COL4A5_extended_for	AAGTTTCCCTCCCCCAAATA
COL4A5_extended_rev	TTCCCTTCTGTCCAGACTTGC
COL4A5_mRNA_for	TCTTCACCCAAGCCTCACTG
COL4A5_mRNA_rev	AGGGCCAGTAAGAACAAGCC
C1orf43_extended_for	CTCCCCGGCCATTTCCTTAC
C1orf43_extended_rev	CCGCTTCCTCTTACTGTCGT
C1orf43_mRNA_for	TCCCCGGCCATTTCCTTACG
C1orf43_mRNA_rev	GAAGGCGTCGTTCTCCTTTCC
HNRNPA2B1_ext_for	TGGCGCTGTAGTGAGAACTG
HNRNPA2B1_ext_rev	CGTAAGGTGGGGCGTAGAG
HNRNPA2B1_mRNA_for	TTTGTCCTACGGCTCGCATT
HNRNPA2B1_mRNA_for	CTGCGCACCTCATTAAAGGC
CDC45_extended_for	TAAAGATGTCCGCCCAGCG
CDC45_extended_rev	GGTCAAGACTCCCGCCAAAT
CDC45_mRNA_for	GGTGTGGGGTGGGATAAAGG
CDC45_mRNA_rev	AGGCCACGACAGATCCAAAG
RFXANK_extended_for	ATTCCCTCCGAGTCGGTGAGGA
RFXANK_extended_rev	ATAGCAGGAGGCAAACTGGA
RFXANK_mRNA_for	TGGGACCAATCGCACCTTTG
RFXANK_mRNA_rev	TGTCTGTACATAGCTTATCGGGG
IMP4_extended_for	AATCGCCTTGAGATGTCCCC
IMP4_extended_rev	CCACATCCCCAGCACTTCAC
IMP4_mRNA_for	CCGAGCGTCTTTCCAAATGC
IMP4_mRNA_rev	CAAAACCCGGGTACTGCAAC

Genomic location of amplicons

Amplicon name	hg19 coordinates
FARSA_mRNA	chr19:13044121-13044295
FARSA_extended	chr19:13044509-13044682
CALR_extended	chr19:13049265-13049461
CALR_mRNA	chr19:13049667-13049856
SNAP23_extended	chr15:42787699-42787922
SNAP23_mRNA	chr15:42788089-42788238
POLR3A_mRNA	chr10:79788840-79788953
POLR3A_extended	chr10:79789211-79789453
RPS24_extended	chr10:79793306-79793599
RPS24_mRNA	chr10:79793735-79793919
COL4A5_extended	chrX:107682929-107683120
COL4A5_mRNA	chrX:107683224-107683405
C1orf43_extended	chr1:154192982-154193169
C1orf43_mRNA	chr1:154192983-154193078
HNRNPA2B1_extended	chr7:26240284-26240496
HNRNPA2B1_mRNA	chr7:26239410-26239583
CDC45_extended	chr22:19467180-19467436
CDC45_mRNA	chr22:19467927-19468055
IMP4_extended	chr2:131100331-131100566
IMP4_mRNA	chr2:131100834-131100962
RFXANK_extended	chr19:19303546-19303816
RFXANK_mRNA	chr19:19303953-19304072

Supplementary note 1

Example of heat map visualization

As an example of heat map visualizatoin, let us consider the generation of the CAGE-RRP40 heat map displaying mRNA-mRNA TSS pairs in Supplementary Fig. 2b. The underlying data are mRNA TSS pairs and CAGE TPM intensity data per bp on both strands, in the +-3,500 bp region relative to the midpoint between mRNA TSSs. Forward and reverse strands were defined by the plus and minus strands of the hg19 assembly. First, mRNA pairs were ordered by their increasing TSS-TSS distances. Each row, corresponding to one mRNA TSS pair, originally consisting of 7,001bp, was split into non-overlapping 701 (=7,001-1/10+1) bins (~10bp each bin), and the average CAGE TPM signal/bp in each bin was calculated separately for each strand. For each strand, each such bin was then assigned a color based on the $\log_2$ (TPM/bp + pseudo-counts) as described below on a white background, producing two heat maps. For composite images like Fig. 2a, these two heat maps were overlaid.