iScience, Volume *26*

Supplemental information

Rapid nuclear deadenylation of mammalian messenger RNA

Jonathan Alles, Ivano Legnini, Maddalena Pacelli, and Nikolaus Rajewsky

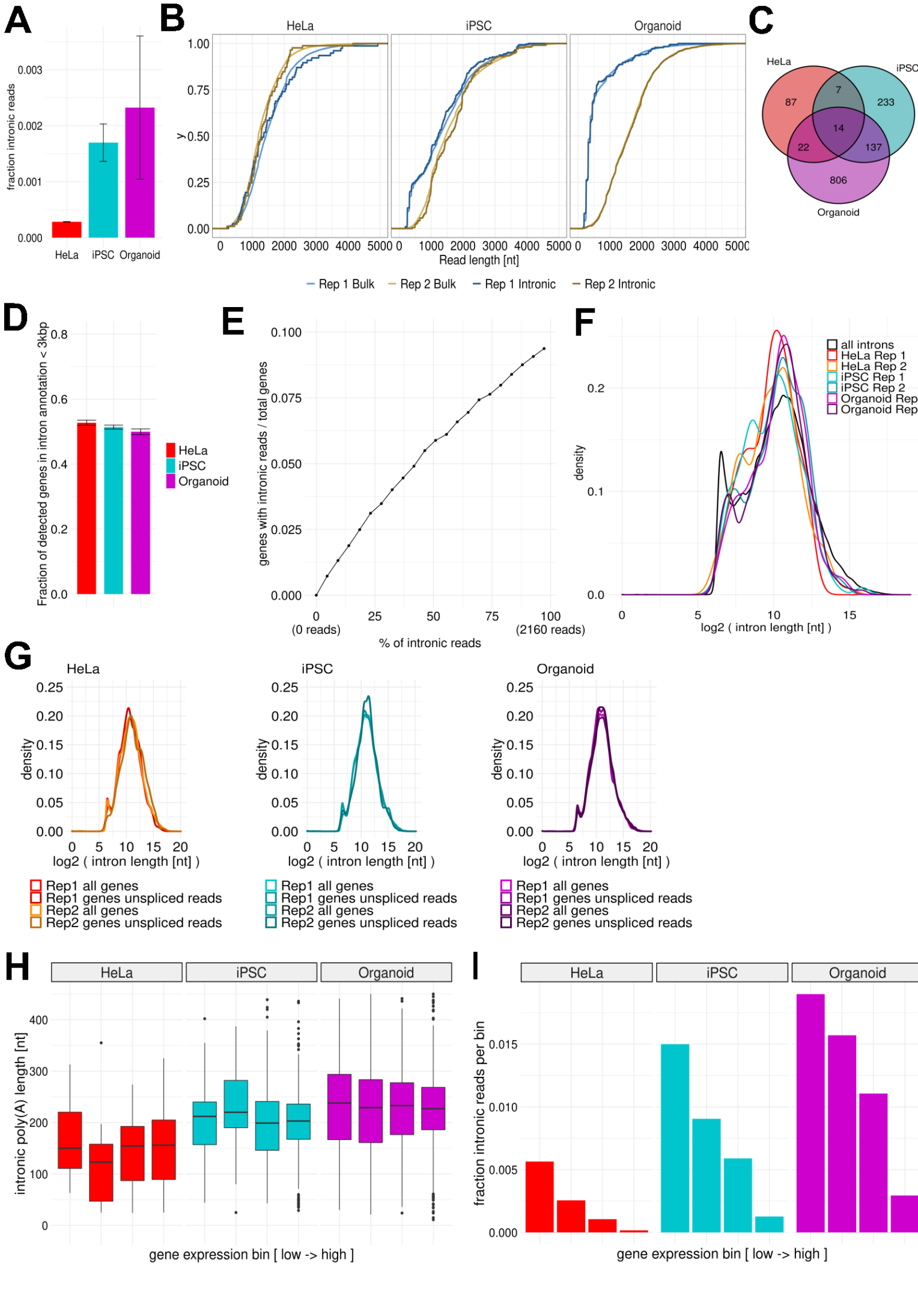
A
fraction intronic reads
0.003
0.002
0.001
0.000
HeLa
iPSC
Organoid
B
HeLa
iPSC
Organoid
1.00
0.75
0.50
0.25
0.00
y
0
1000
2000
3000
4000
5000
Read length [nt]
Rep 1 Bulk
Rep 2 Bulk
Rep 1 Intronic
Rep 2 Intronic
C
HeLa
iPSC
Organoid
87
7
233
14
22
137
806
D
Fraction of detected genes in intron annotation < 3kbp
0.8
0.6
0.4
0.2
0.0
HeLa
iPSC
Organoid
E
genes with intronic reads / total genes
0.100
0.075
0.050
0.025
0.000
0
25
50
75
100
(0 reads)
(2160 reads)
% of intronic reads
F
density
0.2
0.1
0.0
0
5
10
15
log2 (intron length [nt])
all introns
HeLa Rep 1
HeLa Rep 2
iPSC Rep 1
iPSC Rep 2
Organoid Rep 1
Organoid Rep 2
G
HeLa
iPSC
Organoid
density
0.25
0.20
0.15
0.10
0.05
0.00
0
5
10
15
20
log2 (intron length [nt])
Rep1 all genes
Rep1 genes unspliced reads
Rep2 all genes
Rep2 genes unspliced reads
H
HeLa
iPSC
Organoid
intronic poly(A) length [nt]
400
300
200
100
0
gene expression bin [low -> high]
I
HeLa
iPSC
Organoid
fraction intronic reads per bin
0.015
0.010
0.005
0.000
gene expression bin [low -> high]

Figure S1. Features of unspliced mRNAs from PacBio datasets, Related to Figure 1

A Fraction of unspliced, intronic reads in HeLa S3, iPSC, Organoids FLAM-seq datasets. Error bars indicate standard error of the mean for 2 replicates.

B Cumulative distributions of raw read length for intronic and bulk reads for HeLa S3, iPSC and Organoid datasets.

C Venn diagram of genes with detected unspliced (intronic) reads per dataset.

D Fraction of detected genes in FLAM-seq HeLa S3, iPSC, and Organoids dataset represented in the intron annotation database at a maximum intron distance of 3kb from gene end. Error bars indicate standard error of the mean for 2 replicates.

E Downsampling analysis: Intronic reads were downsampled to respective fraction of total reads and the fraction of genes with intronic reads was compared to the total of detected genes.

F Intron length of unspliced introns compared to distribution of all annotated introns for identification of unspliced molecules.

G Intron length distribution from Gencode annotation comparing all genes detected in a dataset with those genes with associated unspliced reads.

H Binning of intronic reads poly(A) length distributions by gene expression for merged datasets. Number of reads per bin is indicated below the boxplot.

I Fraction of intronic reads by gene expression bin for merged datasets.

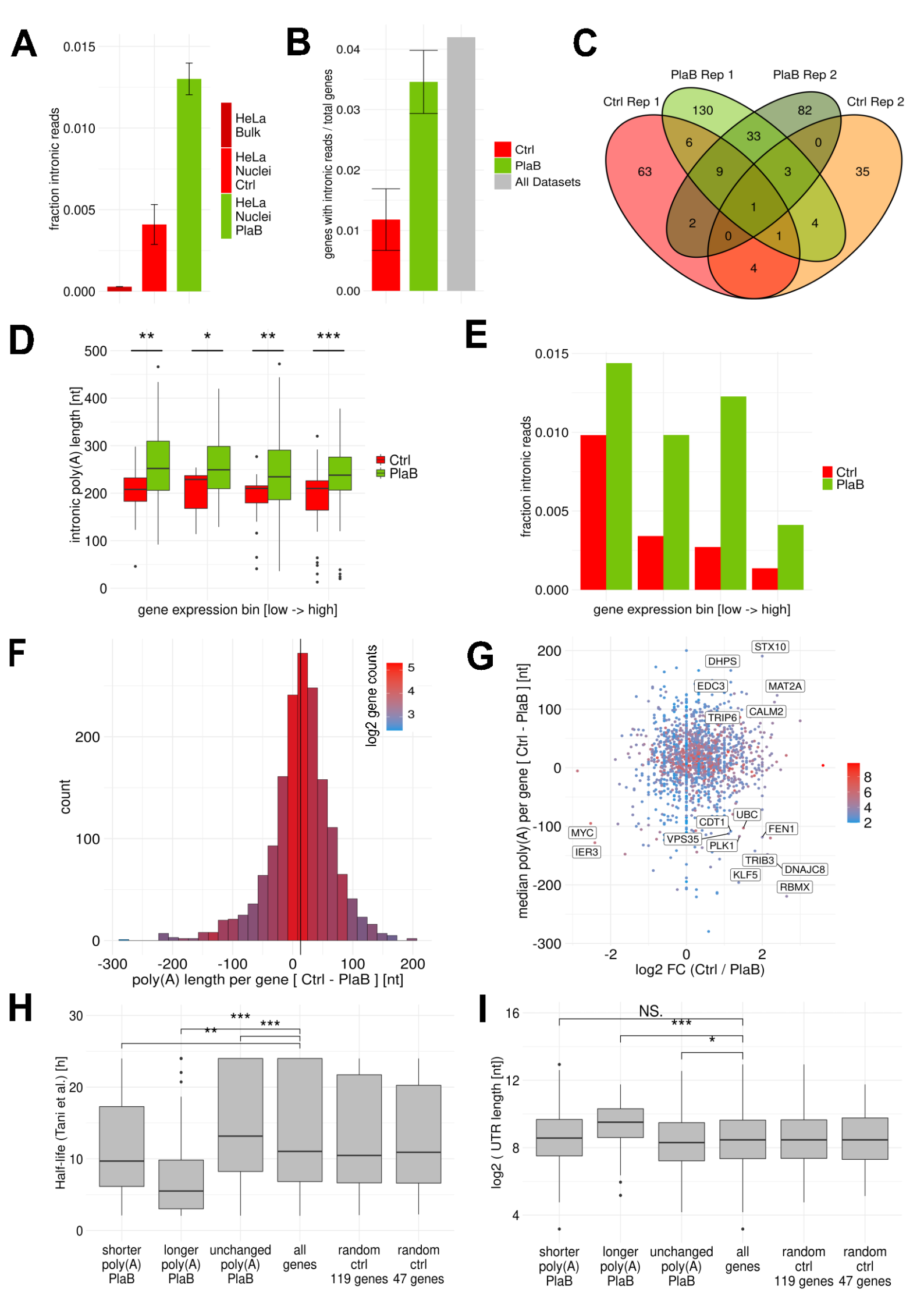
A
fraction intronic reads
0.015
0.010
0.005
0.000
HeLa Bulk
HeLa Nuclei Ctrl
HeLa Nuclei PlaB
B
genes with intronic reads / total genes
0.04
0.03
0.02
0.01
0.00
Ctrl
PlaB
All Datasets
C
PlaB Rep 1
PlaB Rep 2
Ctrl Rep 1
Ctrl Rep 2
130
82
6
33
0
63
9
3
35
1
2
4
0
1
4
D
**
*
**
***
intronic poly(A) length [nt]
500
400
300
200
100
0
Ctrl
PlaB
gene expression bin [low -> high]
E
fraction intronic reads
0.015
0.010
0.005
0.000
Ctrl
PlaB
gene expression bin [low -> high]
F
count
200
100
0
log2 gene counts
5
4
3
-300
-200
-100
0
100
200
poly(A) length per gene [Ctrl - PlaB] [nt]
G
median poly(A) per gene [Ctrl - PlaB] [nt]
200
100
0
-100
-200
-300
STX10
DHPS
EDC3
MAT2A
CALM2
TRIP6
UBC
CDT1
FEN1
MYC
VPS35
PLK1
IER3
TRIB3
DNAJC8
KLF5
RBMX
8
6
4
2
-2
0
2
log2 FC (Ctrl / PlaB)
H
Half-life (Tani et al.) [h]
30
20
10
0
***
**
***
shorter poly(A) PlaB
longer poly(A) PlaB
unchanged poly(A) PlaB
all genes
random ctrl 119 genes
random ctrl 47 genes
I
log2 (UTR length [nt])
16
12
8
4
NS.
***
*
shorter poly(A) PlaB
longer poly(A) PlaB
unchanged poly(A) PlaB
all genes
random ctrl 119 genes
random ctrl 47 genes

Figure S2. Analysis of long reads upon splicing inhibition, Related to Figure 1

A Fraction of intronic reads for HeLa Bulk, Nuclei Ctrl and Nuclei PlaB conditions. Error bars denote standard deviation of 2 replicates.

B Fraction of genes with detected unspliced, intronic reads normalized to total number of genes detected in each dataset. Error bars denote standard deviation of 2 replicates.

C Venn diagram of genes with detected unspliced (intronic) reads per dataset.

D Binning of intronic reads poly(A) length distributions by gene expression for merged datasets. Number of reads per bin is indicated below the boxplot. Asterisks indicate significance level in difference between Ctrl and PlaB (two-sided t-Test; ns : P value > 0.05, * : P ≤ 0.05, ** : P ≤ 0.01, *** : P ≤ 0.001).

E Fraction of intronic reads by gene expression bin for merged datasets.

F Median poly(A) length difference per gene comparing HeLa S3 nuclei control versus PlaB treatment. Color indicates average log2 gene counts. Vertical line indicates mean.

G Difference in median poly(A) length per gene between PlaB and control conditions compared to fold change in gene expression. Highlighted are genes with most prominent differences.

H Distributions of gene HeLa S3 half-life (Tani et al.) for bins of genes with change in poly(A) length upon PlaB splicing inhibition and control distributions. Asterisks indicate significance level in difference between Ctrl and PlaB (two-sided t-Test; ns : P value > 0.05, * : P ≤ 0.05, ** : P ≤ 0.01, *** : P ≤ 0.001).

I Distributions of transcript 3'UTR length for bins of genes with respective change in poly(A) length upon PlaB splicing inhibition and control distributions. Asterisks indicate significance level in difference between Ctrl and PlaB (two-sided t-Test; ns : P value > 0.05, * : P ≤ 0.05, ** : P ≤ 0.01, *** : P ≤ 0.001).

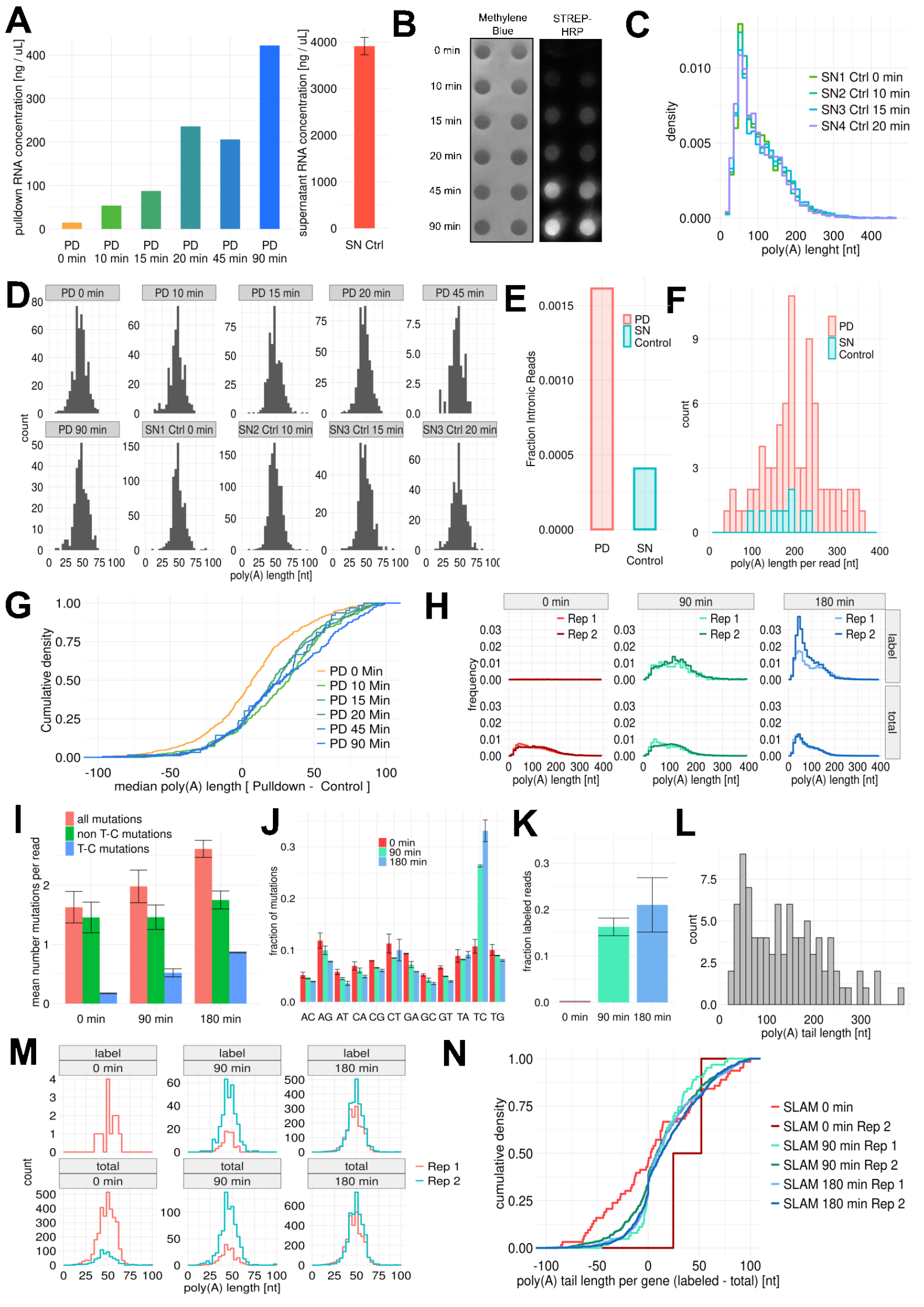
A
pulldown RNA concentration [ng / uL]
PD 0 min
PD 10 min
PD 15 min
PD 20 min
PD 45 min
PD 90 min
supernatant RNA concentration [ng / uL]
SN Ctrl
B
Methylene Blue
STREP-HRP
0 min
10 min
15 min
20 min
45 min
90 min
C
density
poly(A) lenght [nt]
SN1 Ctrl 0 min
SN2 Ctrl 10 min
SN3 Ctrl 15 min
SN4 Ctrl 20 min
D
PD 0 min
PD 10 min
PD 15 min
PD 20 min
PD 45 min
PD 90 min
SN1 Ctrl 0 min
SN2 Ctrl 10 min
SN3 Ctrl 15 min
SN3 Ctrl 20 min
count
poly(A) length [nt]
E
Fraction Intronic Reads
PD
SN
Control
F
count
poly(A) length per read [nt]
G
Cumulative density
median poly(A) length [Pulldown - Control]
PD 0 Min
PD 10 Min
PD 15 Min
PD 20 Min
PD 45 Min
PD 90 Min
H
0 min
90 min
180 min
Rep 1
Rep 2
label
total
frequency
poly(A) length [nt]
I
all mutations
non T-C mutations
T-C mutations
mean number mutations per read
0 min
90 min
180 min
J
0 min
90 min
180 min
fraction of mutations
AC AG AT CA CG CT GA GC GT TA TC TG
K
fraction labeled reads
0 min
90 min
180 min
L
count
poly(A) tail length [nt]
M
label
total
0 min
90 min
180 min
count
poly(A) length [nt]
Rep 1
Rep 2
N
cumulative density
poly(A) tail length per gene (labeled - total) [nt]
SLAM 0 min
SLAM 0 min Rep 2
SLAM 90 min Rep 1
SLAM 90 min Rep 2
SLAM 180 min Rep 1
SLAM 180 min Rep 2

Figure S3. Analysis of metabolic labeling experiments, Related to Figure 2

A RNA concentrations obtained from pulldown (PD) fractions after indicated time points. Error bars denotes standard error of the mean.

B Dot blot with Methylene-Blue staining and Strep-HRP signal against biotinylated RNA before pulldown.

C Poly(A) length distribution of supernatant (SN) fractions for 0 min to 20 min labeling timepoints.

D Poly(A) length distributions of mitochondrial transcripts in PD and SN fractions.

E Fraction of intronic reads in merged datasets of PD and SN fractions.

F Poly(A) tail length distributions of intronic reads from merged datasets of PD and SN fractions.

G Cumulative density distribution of median poly(A) tail length difference per gene between PD and SN fractions.

H Poly(A) tail length profiles of labeled reads and all reads in SLAM-Seq experiments for 0 min, 90 min, and 180 min labeling.

I Average number of mutations per read in SLAM-Seq datasets considering all mismatches per read, T-to-C mismatches or non-T-C mismatches. Error bars denote standard deviation of 2 replicates.

J Observed nucleotide conversions in SLAM-Seq datasets normalized to all possible conversions. Error bars denote standard deviation of 2 replicates.

K Fraction of reads assigned as ‘labeled’ in SLAM-Seq datasets. Error bars denote standard deviation of 2 replicates.

L Poly(A) tail length distribution of intronic reads in merged SLAM-Seq datasets.

M Poly(A) tail length distributions of mitochondrial transcripts in labeled or total poly(A) tail length bin.

N Cumulative density distribution of median poly(A) tail length difference per gene between SLAM-Seq total and labeled reads.

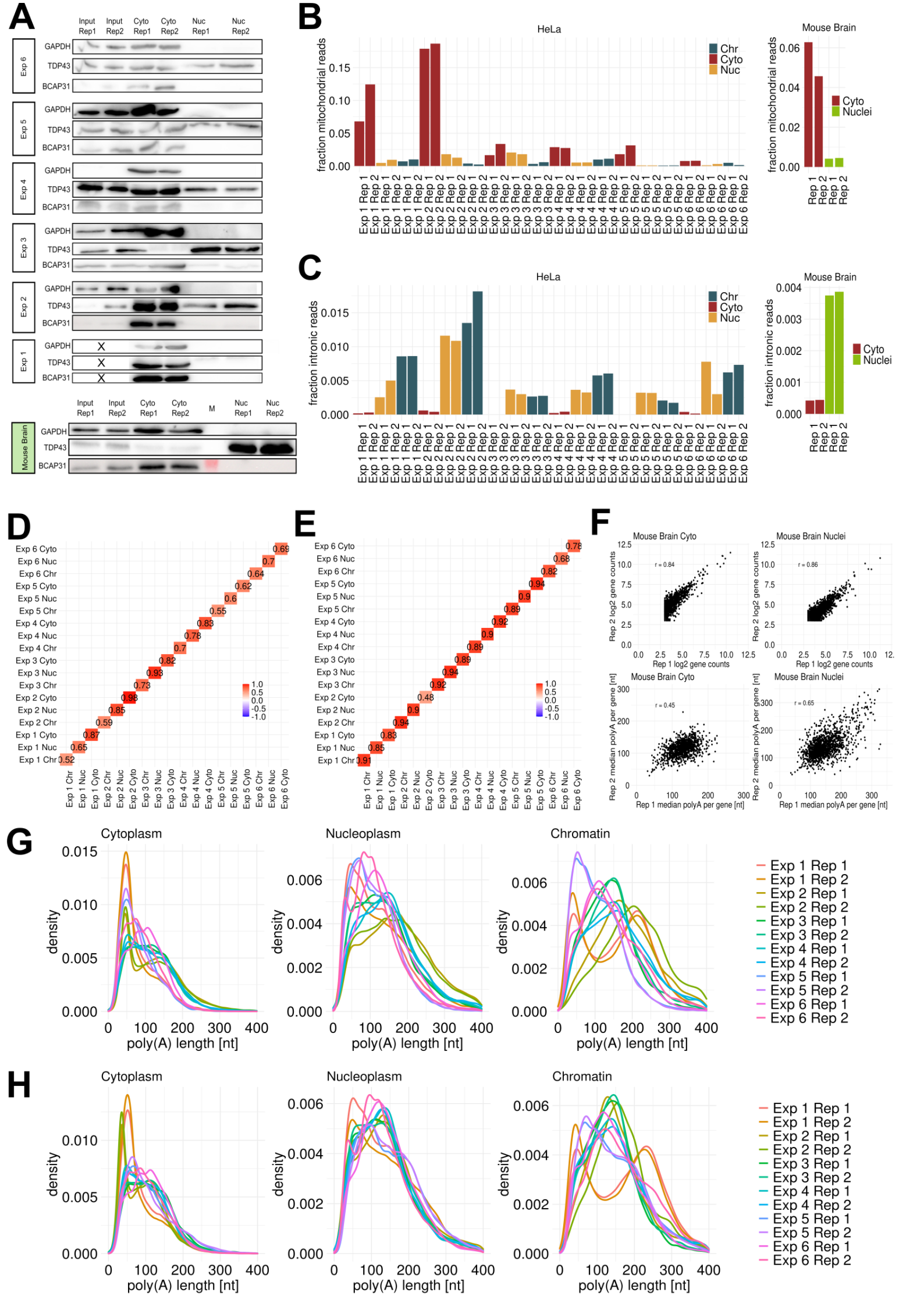
A
Input Rep1
Input Rep2
Cyto Rep1
Cyto Rep2
Nuc Rep1
Nuc Rep2
Exp 6
Exp 5
Exp 4
Exp 3
Exp 2
Exp 1
GAPDH
TDP43
BCAP31
Mouse Brain
M
B
HeLa
Mouse Brain
fraction mitochondrial reads
Chr
Cyto
Nuc
Nuclei
C
fraction intronic reads
D
E
F
Mouse Brain Cyto
Mouse Brain Nuclei
Rep 2 log2 gene counts
Rep 1 log2 gene counts
Rep 2 median polyA per gene [nt]
Rep 1 median polyA per gene [nt]
G
Cytoplasm
Nucleoplasm
Chromatin
density
poly(A) length [nt]
H

Figure S4. Validation of biochemical fractionations, Related to Figure 3

A Western Blot analysis of fractions obtained from biochemical fractionation of HeLa S3 cell lines and mouse brain samples. Markers used are GAPDH as cytoplasmic marker, TDP43 as cytoplasmic and nuclear marker and BCAP31 as ER marker.

B Fraction of mitochondrial reads in each replicate for cytoplasmic, nucleoplasmic and chromatin fractions. Left: HeLa S3 cell lines Right: Mouse brain.

C Fraction of intronic reads in each replicate for cytoplasmic, nucleoplasmic and chromatin fractions. Left: HeLa S3 cell lines Right: Mouse brain.

D Correlation coefficients for median poly(A) tail length per gene between HeLa S3 technical replicates and corresponding fractions for genes with more than 10 counts.

E Correlation coefficients for gene expression counts between HeLa S3 technical replicates and corresponding fractions for genes with more than 10 counts.

F Comparison of gene expression counts per gene (top) and median poly(A) tail length per gene (bottom) for HeLa S3 mouse brain samples for genes with more than 8 counts.

G Poly(A) tail length densities for HeLa S3 samples by fraction.

H Poly(A) tail length densities for HeLa S3 samples by fraction scaled by average poly(A) tail length of sample over all fractions.

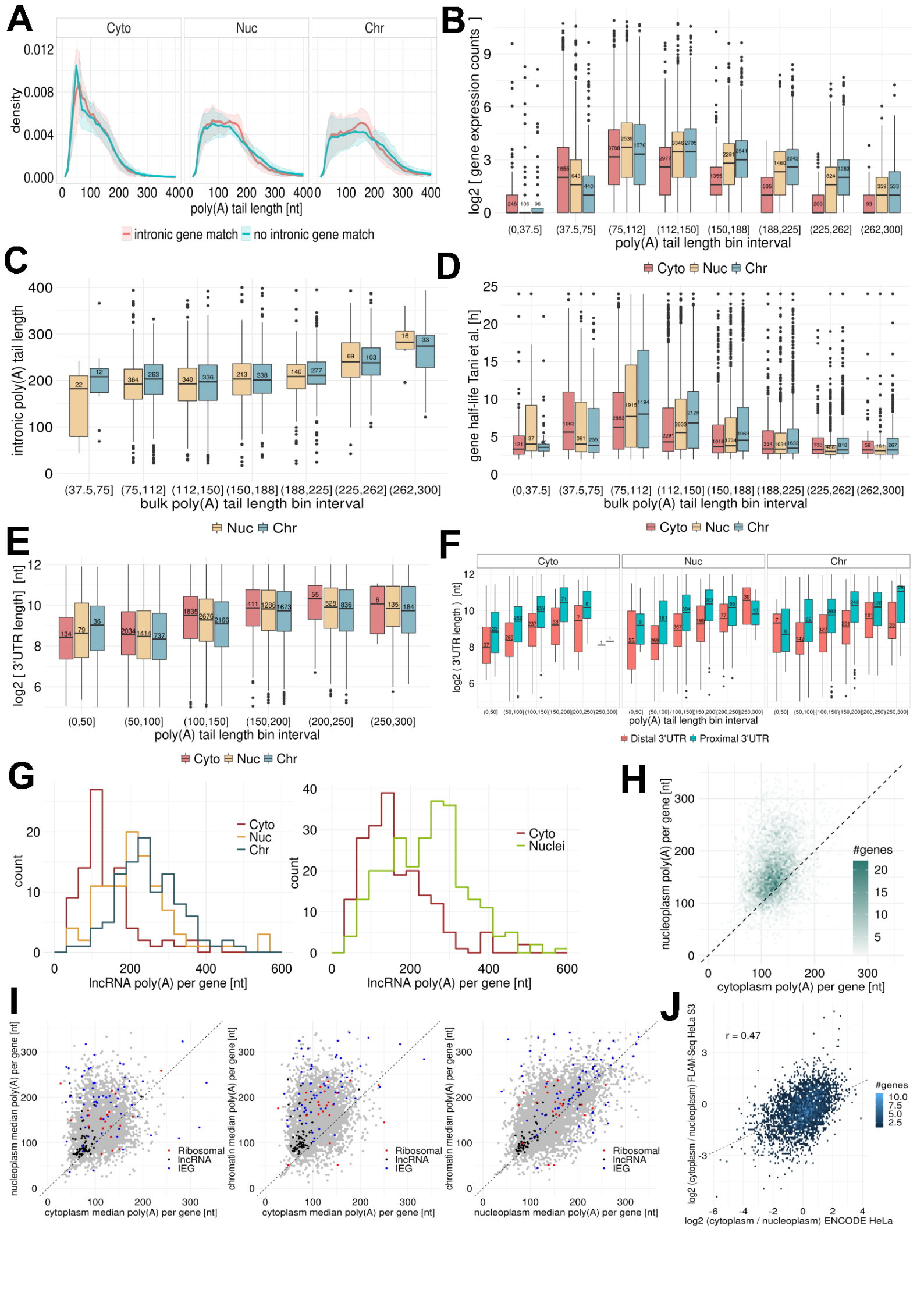
A
Cyto
Nuc
Chr
density
poly(A) tail length [nt]
intronic gene match
no intronic gene match
B
log2 [gene expression counts]
poly(A) tail length bin interval
(0,37.5]
(37.5,75]
(75,112]
(112,150]
(150,188]
(188,225]
(225,262]
(262,300]
C
intronic poly(A) tail length
bulk poly(A) tail length bin interval
D
gene half-life Tani et al. [h]
bulk poly(A) tail length bin interval
E
log2 [3'UTR length] [nt]
poly(A) tail length bin interval
F
log2 (3'UTR length) [nt]
poly(A) tail length bin interval
Distal 3'UTR
Proximal 3'UTR
G
count
lncRNA poly(A) per gene [nt]
Cyto
Nuc
Chr
Nuclei
H
nucleoplasm poly(A) per gene [nt]
cytoplasm poly(A) per gene [nt]
#genes
I
nucleoplasm median poly(A) per gene [nt]
chromatin median poly(A) per gene [nt]
cytoplasm median poly(A) per gene [nt]
nucleoplasm median poly(A) per gene [nt]
Ribosomal
lncRNA
IEG
J
r = 0.47
log2 (cytoplasm / nucleoplasm) FLAM-Seq HeLa S3
log2 (cytoplasm / nucleoplasm) ENCODE HeLa
#genes

Figure. S5 Poly(A) tail analysis for biochemical fractionations, Related to Figure 3

A Bulk poly(A) tail length distribution of genes with detected intronic reads ('intronic gene match') versus genes without detected intronic reads for HeLa S3 cell fractions. Ribbons indicate standard deviation of 12 replicate samples.

B Molecule counts for genes binned by average poly(A) tail length per gene. Numbers indicate the number of genes in each bin. Datasets were merged by biochemical fraction.

C Poly(A) tail length of intronic reads binned by median poly(A) tail length of associated genes. Numbers indicate the number of intronic reads in each bin. Datasets were merged by biochemical fraction.

D Half-lives for genes binned by median poly(A) tail length. Numbers indicate the number of genes in each bin. Datasets were merged by biochemical fraction.

E 3'UTR length for genes binned by median poly(A) tail length. Numbers indicate the number of genes in each bin. Datasets were merged by biochemical fraction.

F 3'UTR length of proximal and distal 3'UTR isoforms for genes with 2 annotated 3'UTR isoforms binned by median poly(A) tail length for each isoform. Numbers indicate the number of 3'UTR isoforms in each bin. Datasets were merged by biochemical fraction.

G Median poly(A) tail length per gene for lncRNA genes in HeLa S3 fractionation experiments (left) and mouse brain fractionations (right). Datasets were merged by biochemical fraction.

H Median poly(A) tail length per gene between cytoplasmic and nuclear fractions for mouse brain fractionation experiments. Datasets were merged by biochemical fraction.

I Median poly(A) tail length per gene compared across fractions from HeLa S3 cell lines for ribosomal protein genes ('Ribosomal'), immediate early genes ('IEGs') and lncRNAs. Datasets were merged by biochemical fraction.

J Cytoplasmic-to-nuclear enrichment of genes quantified by ENCODE for HeLa cell lines versus FLAM-seq quantification. Datasets were merged by biochemical fraction.

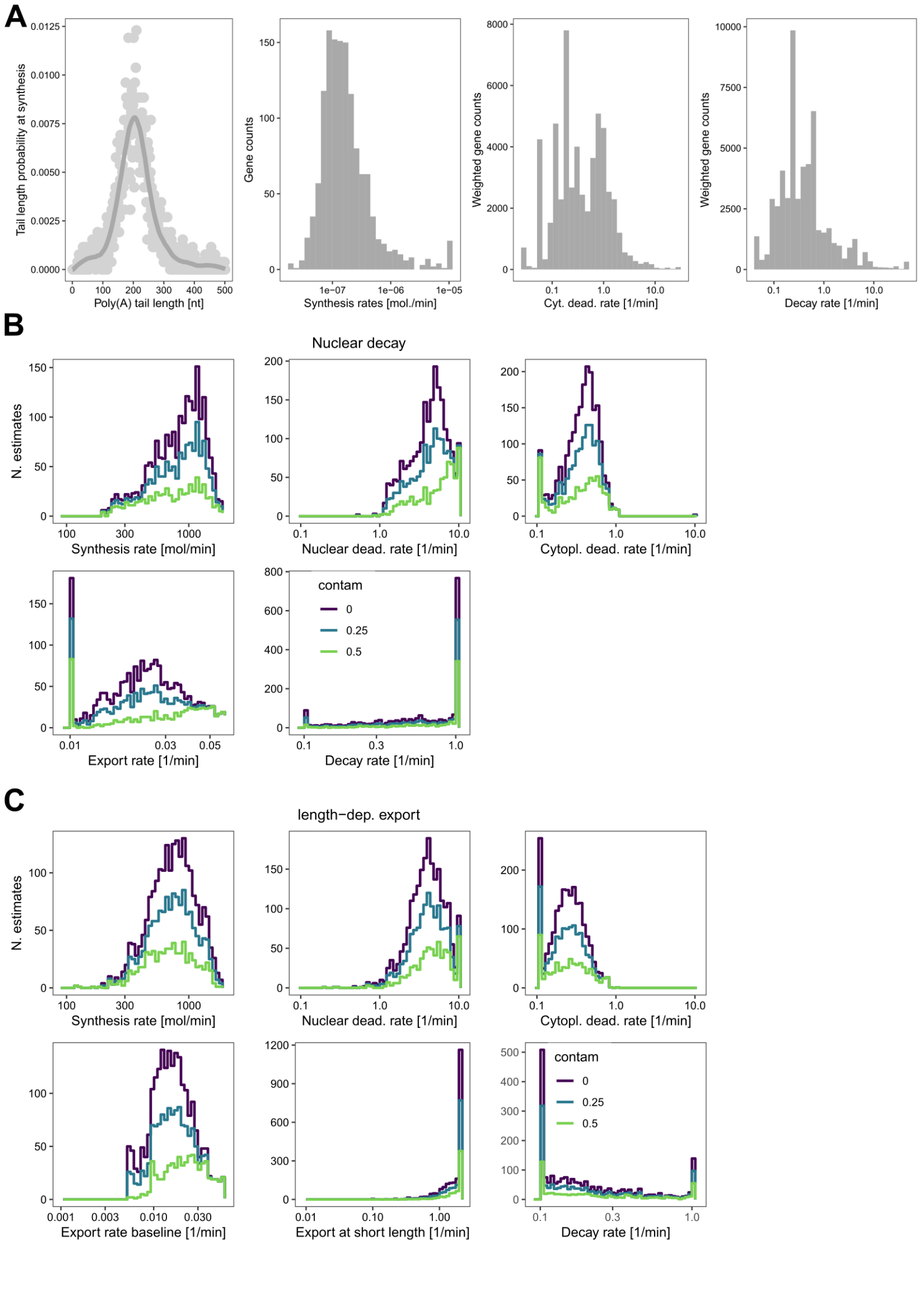
A
Tail length probability at synthesis
Poly(A) tail length [nt]
Gene counts
Synthesis rates [mol./min]
Weighted gene counts
Cyt. dead. rate [1/min]
Weighted gene counts
Decay rate [1/min]
B
Nuclear decay
N. estimates
Synthesis rate [mol/min]
Nuclear dead. rate [1/min]
Cytopl. dead. rate [1/min]
Export rate [1/min]
Decay rate [1/min]
contam
0
0.25
0.5
C
length−dep. export
N. estimates
Synthesis rate [mol/min]
Nuclear dead. rate [1/min]
Cytopl. dead. rate [1/min]
Export rate baseline [1/min]
Export at short length [1/min]
Decay rate [1/min]
contam
0
0.25
0.5

Figure S6. Rate estimates for deadenylation models, Related to Figure 4

A Data used for initial parameter guesses in the optimization: slp parameter (left) is estimated from intronic reads of 12 FLAM-seq fractionation replicates; synthesis rates are estimated by summing up all transcription rates estimated from Eisen et al. (2020) (distribution per gene shown in middle left), nuclear and cytoplasmic deadenylation rates are initialized according to the mean of the deadenylation rates measured by Eisen et al. (2020) (distribution weighted for gene expression in FLAM-seq data shown in middle right); decay rate: same as before (right).

B From top left to bottom right: final distributions from optimization of synthesis rate, nuclear and cytoplasmic deadenylation rate, export rate, decay rate. Parameters were estimated with cytoplasmic contamination of nuclear RNA set at 0, 25 and 50% (see color legend), for the nuclear decay - constant export (ndce) model.

C Same as B., for the no nuclear decay, tail length dependent export (nndle) model.

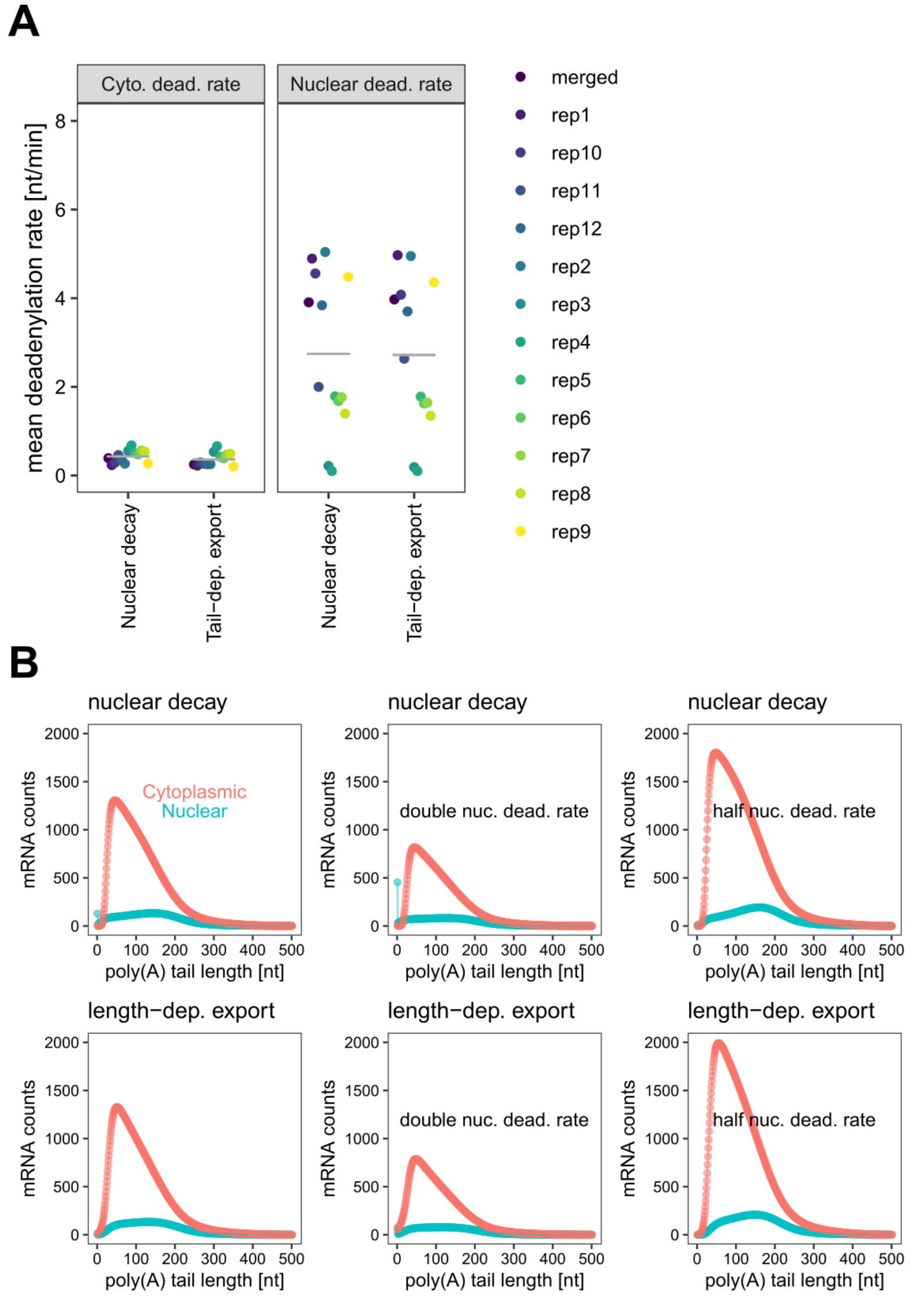
A
Cyto. dead. rate
Nuclear dead. rate
mean deadenylation rate [nt/min]
0
2
4
6
8
Nuclear decay
Tail−dep. export
Nuclear decay
Tail−dep. export
merged
rep1
rep10
rep11
rep12
rep2
rep3
rep4
rep5
rep6
rep7
rep8
rep9
B
nuclear decay
nuclear decay
nuclear decay
length−dep. export
length−dep. export
length−dep. export
mRNA counts
poly(A) tail length [nt]
Cytoplasmic
Nuclear
double nuc. dead. rate
half nuc. dead. rate
0
500
1000
1500
2000
0
100
200
300
400
500

Figure. S7 Deadenylation rates and poly(A) profiles for deadenylation models, Related to Figure 4

A Values of k_1 and k_2 (nuclear and cytoplasmic deadenylation rate) estimated independently from 12 fractionation experiments for both nuclear decay constant export (ndce) and no nuclear decay, tail length dependent export (nndle). k_1 values are consistently higher than k_2, except for two replicates, which were performed in the same batch, and where tail lengths were overall longer than for all other replicates (possibly a technical artifact of the poly(A) selection step in the FLAM-seq protocol as discussed in the results section).

B Simulation of mRNA levels in the nucleus (green) and cytosol (red) with the model assuming constant export and nuclear decay (ndce model; top) and the model assuming tail-length dependent export (nndle model; bottom), using the average of the estimated parameters distributions as input (left), or increasing and decreasing the nuclear deadenylation k_1 rate 2-fold (middle and right). For this specific example, we assumed a cytoplasmic contamination in the nuclear fraction of 0%.

Process	Parameter	Definition	Initial values	Final values (ndce/nndle)	Description
Data input	a	mRNA molecules per cell	fixed at 200000	fixed	total number of mRNA molecules per cell (at steady state)
Data input	b	nuc/cyt ratio	0.21	fixed	Ratio of mRNA molecules in the nucleus vs in the cytoplasm, estimated from RNA abundance in 12 fractionation experiments
Data input	c	chr/nuc ratio	0.69	fixed	Ratio of mRNA molecules in the chromatin vs in the nucleoplasm estimated from RNA abundance in 12 fractionation experiments; used to weight FLAM-seq data from cell fractionations
Data input	contam	cytoplasmic contamination	0.00, 0.25, 0.50	fixed (at three different values)	Assumed contamination fo cytoplasmic mRNA in the nuclear fraction; three different scenarios tested
Data input	l (lengths)	tail lengths [nt]	from 1 to 501	fixed	allowed tail lengths (from 1 to 501)
Data input	nuc_data	nuclear tail length distribution	empirical distribution from FLAM-seq data	fixed (scaled with contam and for individual replicates)	Distribution of tail lengths for nuclear mRNA as observed in FLAM-seq data from nucleoplasm and chromatin fractions; function of a, b, c, contam.
Data input	cyt_data	cytoplasmic tail length distribution	empirical distribution from FLAM-seq data	fixed (scaled with contam and for individual replicates)	Distribution of tail lengths for cytoplasmic mRNA as observed in FLAM-seq data from cytoplasmic fraction; function of a, b, c, contam.
Data input / transcription	slp	tail length probability at synthesis	empirical distribution from FLAM-seq data	fixed	probabiliy function for tail length at synthesis, fixed from FLAM-seq data (intronic reads)
Transcription	α	synthesis rate [molecules/min]	719	914 +/- 372, 782 +/- 314	multiplier for slp, represent the total number of mRNA molecules produced per minute in a cell; recomputed from Eisen et al., 2020.
Nuclear deadenylation	k_1	nuclear deadenylation rate [min^{-1}]	0.25	4.23 +/- 2.05, 4.06 +/- 2.04	contant deadenylation rate in the nucleus; initial values provided as the weighted mean from Eisen et al., 2020
Export (ndce)	k_E	export rate [min^{-1}]	0.02	0.024 +/- 0.009	costant export rate (does not vary with tail length); if export constant, decay is allowed also in the nucleus
Export (nndle)	s_E	export rate [min^{-1}]	0.02	0.016 +/- 0.01	constant baseline for logistic export rate (does not vary with tail length; added to a logistic function defined by aE, mE, vE); if export is logistic, decay not allowed in the nucleus
Export (nndle)	a_E	export rate [min^{-1}]	1.2	1.60 +/- 0.50	logistic export rate (varies with tail length)
Export (nndle)	m_E	tail length scaling	5	6 +/- 3	tail length mean for logistic export rate
Export (nndle)	v_E	tail length variance	10	12 +/- 4	tail length variance for logistic export rate
Cytoplasmic deadenylation	k_2	cytoplasmic deadenylation rate [min^{-1}]	0.25	0.39 +/- 0.32, 0.26 +/- 0.13	contant deadenylation rate in the cytoplasm; initial value provided as weighted mean from Eisen et al., 2020
Decay	a_d	decay rate [min^{-1}]	0.31	0.63 +/- 0.33, 0.26 +/- 0.25	logistic decay rate: allowed only in cytoplasm, or also in the nucleus in case of constant export; initial value provided as weighted mean from Eisen et al., 2020
Decay	m_d	tail length scaling	16	10 +/- 5, 11 +/- 6	tail length mean for logistic decay rate; initial value provided from Eisen et al., 2020
Decay	v_d	tail length variance	11	6 +/- 2, 9 +/- 3	tail length variance for logistic decay rate; initial value provided from Eisen et al., 2020

Supplementary Table 1, Related to Figure 4.

Definition and description for all parameters used in the ndce and nndle models optimization. Means of initial guesses are shown, while estimated values are shown as mean and standard deviation of merged distribution from 1,000 iterations for 0, 25 and 50% cytoplasmic contamination of nuclear fraction, for ndce and nndle (comma separated).