Supplementary Information

Total RNA-Seq to identify pharmacological effects on specific stages of mRNA synthesis

Sarah A. Boswell[1,2], Andrew Snavely[3], Heather M. Landry[3], L. Stirling Churchman[3], Jesse M. Gray[3*], and Michael Springer[1,2*]

---

[1] Department of Systems Biology, Harvard Medical School, Boston MA
[2] Laboratory of Systems Pharmacology, Harvard Medical School, Boston MA
[3] Department of Genetics, Harvard Medical School, Boston MA
* Contributed equally and corresponding authors gray@genetics.med.harvard.edu and michael_springer@hms.harvard.edu

Supplementary Results

Supplementary Table 1. GEO submission metadata.

Supplementary Figures

Supplementary Figure 1. Total RNA-Seq profiles for IsoG, DRB, FP, Meayamycin (MY), SSA, and siRNA targeting RRP40.

Supplementary Figure 2. Definition of metrics used to classify small molecule action on transcription and splicing.

Supplementary Figure 3. Meayamycin inhibits splicing, and effects of IsoG on the RNA polymerase CTD are reversible.

Supplementary Figure 4. NET-Seq reads around polyadenylation sites recapitulate the termination defect in IsoG treated cells first observed by RNA-Seq.

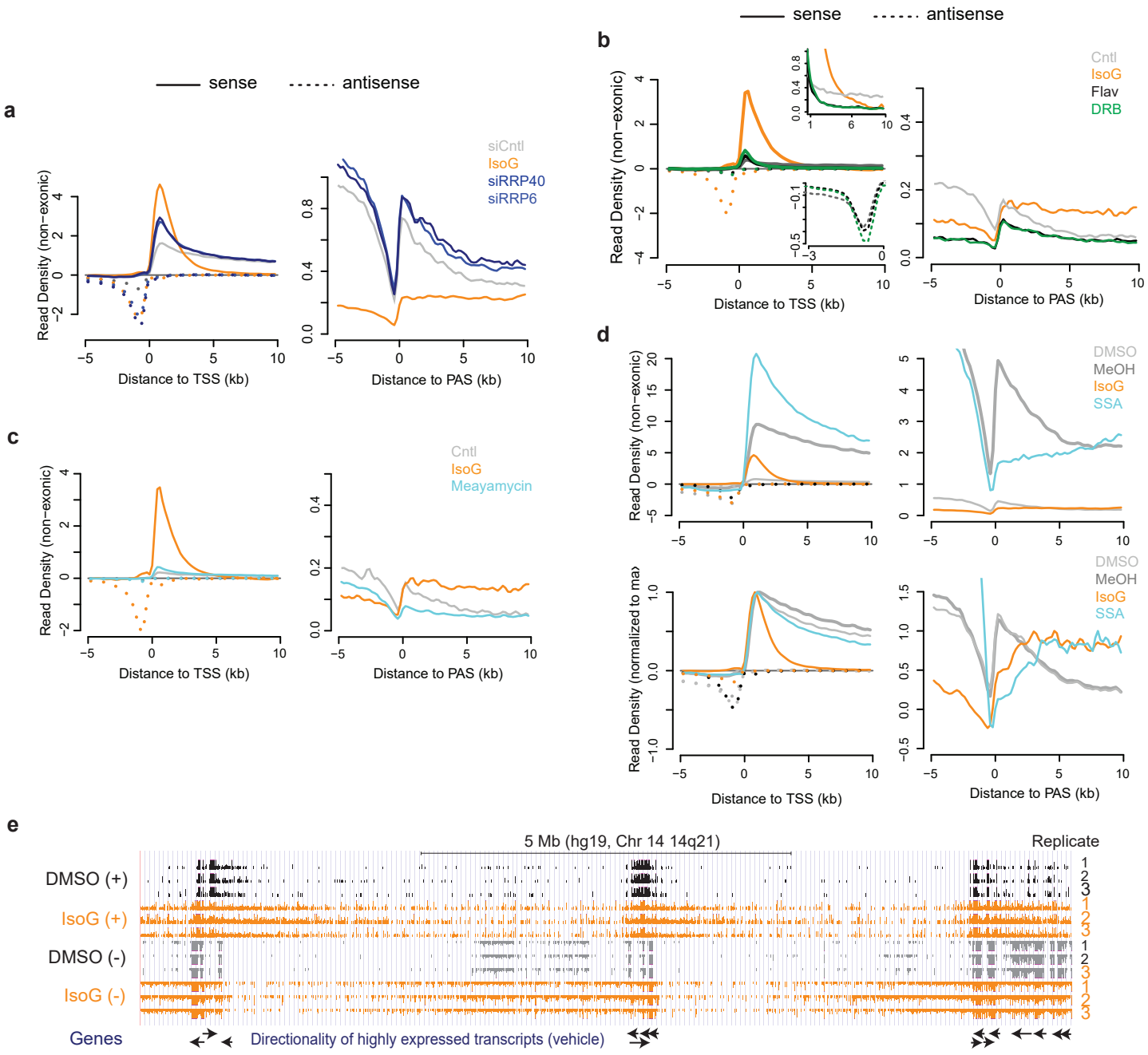
Supplementary Figure 5. Sensitivity of metrics to downsampling of RNA-Seq read depth.

Supplementary Reference:

55. Kaida, D. et al. Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. Nat. Chem. Biol. 3, 576–583 (2007).

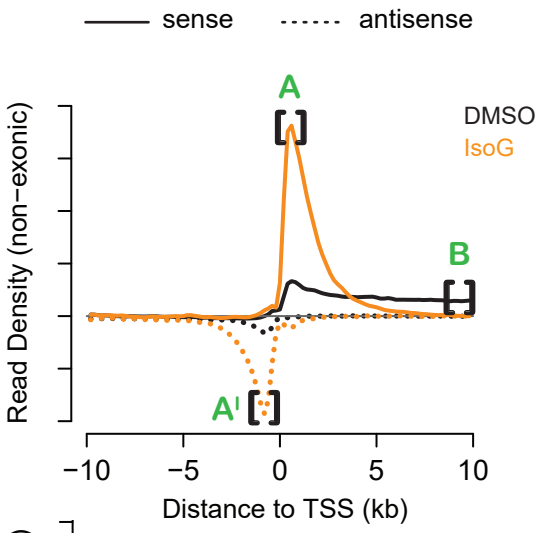
Supplementary Table 1. GEO submission metadata. Table of all the experiments and samples for this paper as submitted to GEO.

file name	instrument model	read length	single or paired-end	raw read counts
HiSeq_IsoG-CPT__D6-26_CPT__UNMAPPED-NOPAIRS__50bp.fq	Illumina HiSeq 2000	50	single	29886918
HiSeq_IsoG-CPT__D6-26_IsG__UNMAPPED-NOPAIRS__50bp.fq	Illumina HiSeq 2000	50	single	46022294
HiSeq_IsoG-CPT__D6-26_c10__UNMAPPED-NOPAIRS__50bp.fq	Illumina HiSeq 2000	50	single	37818611
HiSeq_IsoG-CPT__D6-26_c60__UNMAPPED-NOPAIRS__50bp.fq	Illumina HiSeq 2000	50	single	34972053
HiSeq_IsoG-CPT__D7-10_CPT__UNMAPPED-NOPAIRS__50bp.fq	Illumina HiSeq 2000	50	single	34080550
HiSeq_IsoG-CPT__D7-10_c10__UNMAPPED-NOPAIRS__50bp.fq	Illumina HiSeq 2000	50	single	31323474
HiSeq_IsoG-CPT__D8-01_IsG__UNMAPPED-NOPAIRS__50bp.fq	Illumina HiSeq 2000	50	single	39106256
HiSeq_IsoG-CPT__D8-01_c60__UNMAPPED-NOPAIRS__50bp.fq	Illumina HiSeq 2000	50	single	35467531
HiSeq_IsoG-CPT__D9-19_IsG__UNMAPPED-NOPAIRS__50bp.fq	Illumina HiSeq 2000	50	single	41004036
HiSeq_IsoG-CPT__D9-19_c60__UNMAPPED-NOPAIRS__50bp.fq	Illumina HiSeq 2000	50	single	32706937
C_polyA_1.fastq	Illumina HiSeq 2000	50	single	16044027
C_polyA_2.fastq	Illumina HiSeq 2000	50	single	19003387
C_total_1.fastq	Illumina HiSeq 2000	50	single	10322814
C_total_2.fastq	Illumina HiSeq 2000	50	single	14668030
DRB_polyA_1.fastq	Illumina HiSeq 2000	50	single	14339784
DRB_polyA_2.fastq	Illumina HiSeq 2000	50	single	16616184
DRB_total_1.fastq	Illumina HiSeq 2000	50	single	19846938
DRB_total_2.fastq	Illumina HiSeq 2000	50	single	11039001
IsoG_polyA_1.fastq	Illumina HiSeq 2000	50	single	15697483
IsoG_polyA_2.fastq	Illumina HiSeq 2000	50	single	22254735
IsoG_total_1.fastq	Illumina HiSeq 2000	50	single	19222840
IsoG_total_2.fastq	Illumina HiSeq 2000	50	single	14355621
C_18hr_B1__AS05_ACAGTG.R1.fastq	Illumina HiSeq 2000	50	single	11267263
C_18hr_B2__AS06_TAGCTT.R1.fastq	Illumina HiSeq 2000	50	single	9006933
C_6hr_B1__AS01_TTAGGC.R1.fastq	Illumina HiSeq 2000	50	single	5608642
C_6hr_B2__AS02_CAGATC.R1.fastq	Illumina HiSeq 2000	50	single	1445249
MY_18hr_B1__AS07_GCCAAT.R1.fastq	Illumina HiSeq 2000	50	single	11814071
MY_18hr_B2__AS08_GGCTAC.R1.fastq	Illumina HiSeq 2000	50	single	25596331
MY_6hr_B1__AS03_TGACCA.R1.fastq	Illumina HiSeq 2000	50	single	9638975
MY_6hr_B2__AS04_ACTTGA.R1.fastq	Illumina HiSeq 2000	50	single	6657706
C_total_18h_B1.fq	Illumina HiSeq 2000	50	single	9120119
C_total_18h_B2.fq	Illumina HiSeq 2000	50	single	10040863
C_total_6h_B1.fq	Illumina HiSeq 2000	50	single	10447786
C_total_6h_B2.fq	Illumina HiSeq 2000	50	single	8571269
Drb_total_18h_B1.fq	Illumina HiSeq 2000	50	single	9601847
Drb_total_18h_B2.fq	Illumina HiSeq 2000	50	single	10508965
Drb_total_6h_B1.fq	Illumina HiSeq 2000	50	single	10481620
Drb_total_6h_B2.fq	Illumina HiSeq 2000	50	single	9519695
Flav_total_18h_B1.fq	Illumina HiSeq 2000	50	single	10272661
Flav_total_18h_B2.fq	Illumina HiSeq 2000	50	single	8786717
Flav_total_6h_B1.fq	Illumina HiSeq 2000	50	single	10904424
Flav_total_6h_B2.fq	Illumina HiSeq 2000	50	single	7865542
IsoG_total_18h_B1.fq	Illumina HiSeq 2000	50	single	6005841
IsoG_total_18h_B2.fq	Illumina HiSeq 2000	50	single	8970784
IsoG_total_6h_B1.fq	Illumina HiSeq 2000	50	single	10278058
IsoG_total_6h_B2.fq	Illumina HiSeq 2000	50	single	9025001
DMSO_Rep1_160623.fastq	Illumina NEXTSeq	75	single	98229966
DMSO_Rep2.fastq	Illumina NEXTSeq	75	single	205709503
IsoG_Rep1.fastq	Illumina NEXTSeq	75	single	261046299
IsoG_Rep2_160623.fastq	Illumina NEXTSeq	75	single	108051980

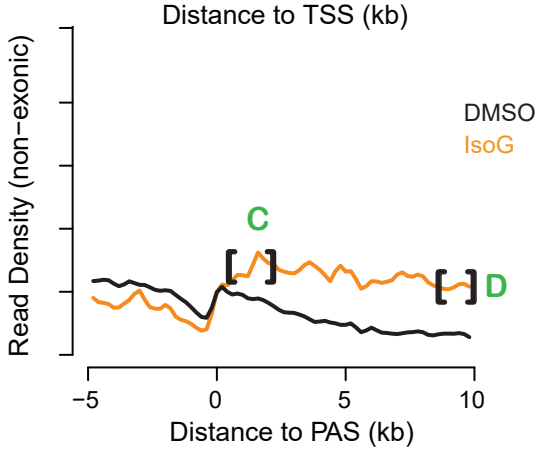

a
sense
antisense
siCntl
IsoG
siRRP40
siRRP6
Read Density (non−exonic)
Distance to TSS (kb)
Distance to PAS (kb)
b
sense
antisense
Cntl
IsoG
Flav
DRB
Read Density (non−exonic)
Distance to TSS (kb)
Distance to PAS (kb)
c
Cntl
IsoG
Meayamycin
Read Density (non−exonic)
Distance to TSS (kb)
Distance to PAS (kb)
d
DMSO
MeOH
IsoG
SSA
Read Density (non−exonic)
Read Density (normalized to ma
Distance to TSS (kb)
Distance to PAS (kb)
e
5 Mb (hg19, Chr 14 14q21)
Replicate
DMSO (+)
IsoG (+)
DMSO (-)
IsoG (-)
Genes
Directionality of highly expressed transcripts (vehicle)

Supplementary Figure 1. Total RNA-Seq profiles for IsoG, DRB, FP, Meayamycin (MY), SSA, and siRNA targeting RRP40. (**a-c**) Non peak-normalized metaplots corresponding to the peak-normalized metaplots in Fig. 2. Whereas the Fig. 2 plots are designed to show promoter-proximal versus promoter-distal RNA abundance, these plots are normalized to ERCC spike in controls to show absolute abundance. Insets are zoomed in to show that in IsoG-treated cells read density drops below that of vehicle-treated cells by ~6 kb downstream of TSSs (top) and that peak antisense read densities are similar among Flav, DRB, and control. In these and all other ERCC normalized TSS metaplots, the background level of transcription as measured in the –5 to –10 kb region upstream of the plotted TSSs is subtracted (Online Methods). (**d**) Comparison of IsoG with the splicing inhibitor Spliceostatin A (SSA) (processed identically to similar plots in Fig 2). Top panel is the non peak-normalized data as in (a-c); bottom panel is the peak-normalized data. Both IsoG and SSA data in (d) were from Tseng et al[21]. (**e**) Representative example locus showing that IsoG causes pervasive transcription beyond and through gene boundaries. Hela cells were treated for 18h with 30 uM IsoG in (e). Arrows at bottom show the annotated directionality of transcripts expressed highly in the vehicle-treated cells.

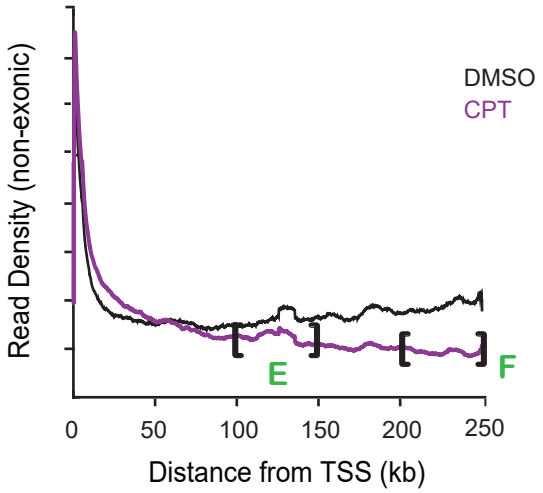
a



b



c



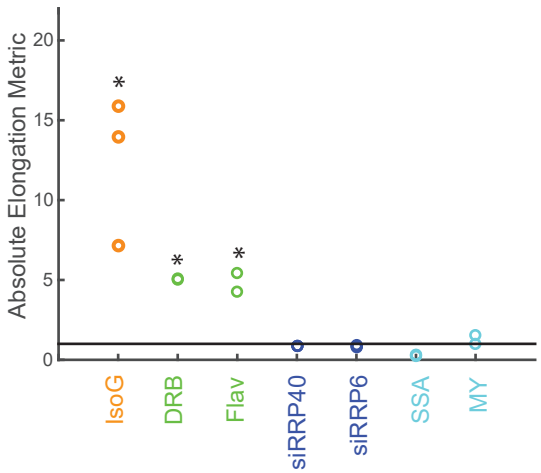
d

Metric Name	Metric
Antisense TSS / Sense TSS	A^I/A
TSS / Gene Body	A/B
Absolute Gene Body	B
Intron / Exon	B_{exon}/B_{intron}
PAS-downstream / PAS	D/C
Density Change Across Long Gene Bodies	E/F

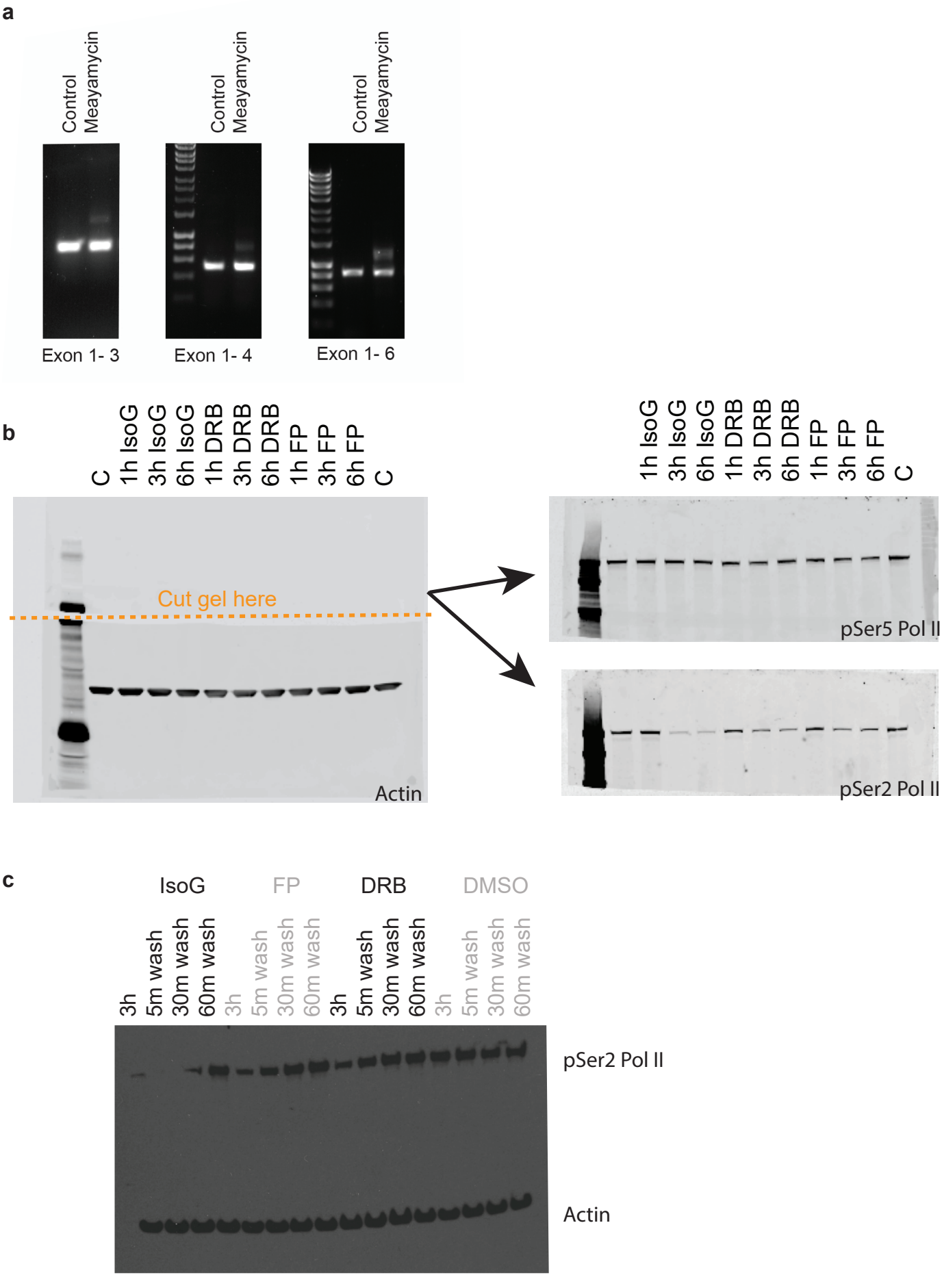
e

Metric Variables	
A^I	-600 to -1000bp
A	+600 to +1000bp
B	+9 to +10 kb
C	+200 to +1000bp
D	+9.2 to +10kb
E	+100 to +150kb
F	+200 to +250kb

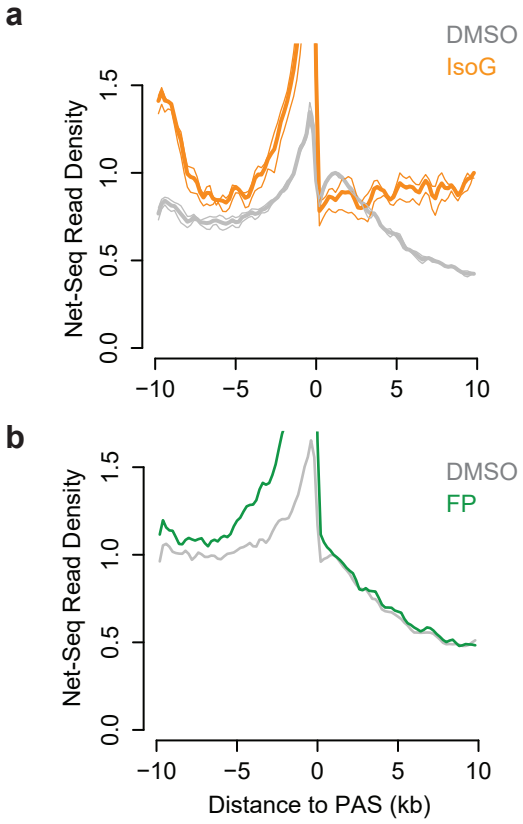
f



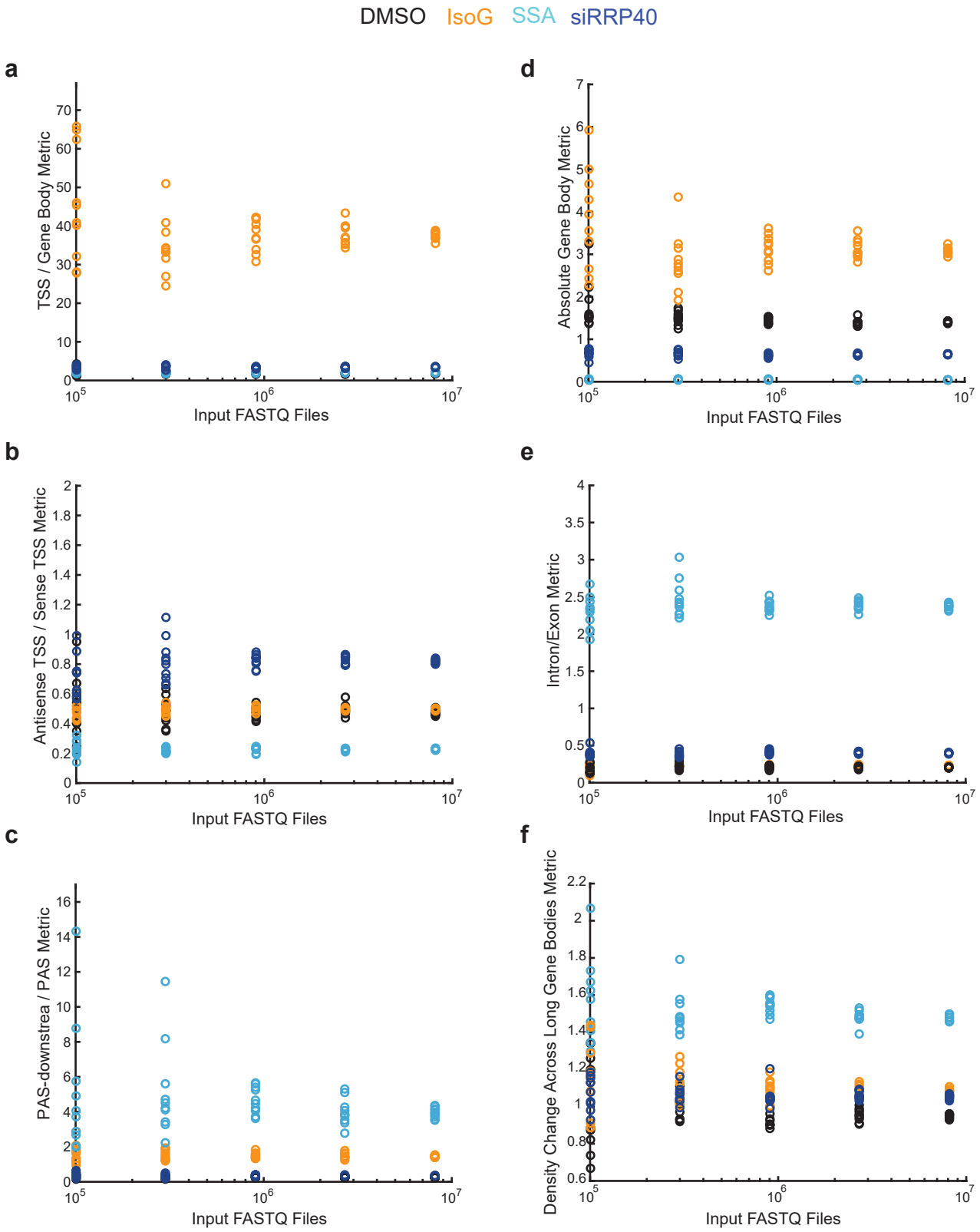
Supplementary Figure 2. Definition of metrics used to classify small molecule action on transcription and splicing. (**a**) At top, TSS metaplots annotated to show regions used for exosome, promoter-proximal elongation and splicing metric calculations. (**b**) PAS metaplots annotated to show regions used for the termination metric calculation. (**c**) TSS metaplots annotated to show regions used for the gene body elongation defect metric calculation. TSS metaplots were made here as described in Online Methods. (**d**) Table showing the formulas used to compute each metric. (**e**) Table showing the regions of the metaplots used for each variable in the formulas shown in (d). (**f**) Absolute gene body metric, a measure of decreased elongation based on the ERCC spike-in-normalized non-exonic RNA-Seq signal from +9 to +10 kb relative to the TSS. Plotted values represent vehicle controls divided by matched drug treatments, such that positive values represent decreased RNA-Seq signal upon drug treatment. NOTE: this metric will also vary based on rates of splicing and transcription initiation. Samples and treatments are the same as in Fig. 2. Stars indicate significant increases ($p<0.05$) based on a one-tailed *t*-test with Bonferonni correction.


a
Control
Meayamycin
Exon 1- 3
Exon 1- 4
Exon 1- 6
b
C
1h IsoG
3h IsoG
6h IsoG
1h DRB
3h DRB
6h DRB
1h FP
3h FP
6h FP
C
Cut gel here
Actin
pSer5 Pol II
pSer2 Pol II
c
IsoG
FP
DRB
DMSO
3h
5m wash
30m wash
60m wash
pSer2 Pol II
Actin

Supplementary Figure 3. Meayamycin inhibits splicing, and effects of IsoG on the RNA polymerase CTD are reversible. (**a**) Accumulation of unspliced introns in the IkB gene after 6 hours of Meayamycin (MY) treatment, as detected by PCR using previously validated primers[55]. (**b**) Full gel images from western blots in Fig. 4a with a dashed line marking where the gel was cut for probing with different antibodies. (**c**) Serine 2 phosphorylation on the CTD recovers after washout of IsoG, DRB, or flavopiridol (FP). Cells in (c) were treated with 30uM IsoG, 100uM DRB, or 300nM FP for 3 hours and washed out for 5–60 minutes as indicated above each lane of the gel.



Supplementary Figure 4. **NET-Seq reads around polyadenylation sites recapitulate the termination defect in IsoG treated cells first observed by RNA-Seq.** Metagene analysis of NET-Seq data from HeLa S3 cells aligned at polyadenylation sites (PASs) and normalized to peak expression downstream of the PAS. (**a**) Cells treated with IsoG for 6 hours at 30uM; the solid line is the average of two replicates and the lighter lines are the replicate data sets. (**b**) Cells treated with flavopiridol (FP) for 1 hour as previously published[36].


DMSO IsoG SSA siRRP40
a
TSS / Gene Body Metric
Input FASTQ Files
b
Antisense TSS / Sense TSS Metric
Input FASTQ Files
c
PAS-downstrea / PAS Metric
Input FASTQ Files
d
Absolute Gene Body Metric
Input FASTQ Files
e
Intron/Exon Metric
Input FASTQ Files
f
Density Change Across Long Gene Bodies Metric
Input FASTQ Files

Supplementary Figure 5. Sensitivity of metrics to downsampling of RNA-Seq read depth. To estimate the relationship between read depth and the accuracy of our metrics, we downsampled four of our experimental samples (DMSO, IsoG, SSA, and siRRP40). Downsampling was performed by randomly choosing sequencing reads without replacement from the initial fastq file to create a new fastq file. The resulting fastq file was processed in a manner identical to that of the original sample. Each sample was down sampled ten times at five different downsampling depths (10^5, $3x10^5$, $9x10^5$, $2.7x10^6$, and $8.1x10^6$). (**a-f**) For each downsampling depth, the metrics described in Supplementary Fig. 2a-e and Online Methods were computed.