Drosha regulates gene expression independently of RNA-cleavage function

Natalia Gromak[1*], Martin Dienstbier[1], Sara Macias[2], Mireya Plass[3,5], Eduardo Eyras[3,4] Javier F. Caceres[2] and Nicholas J. Proudfoot[1*]

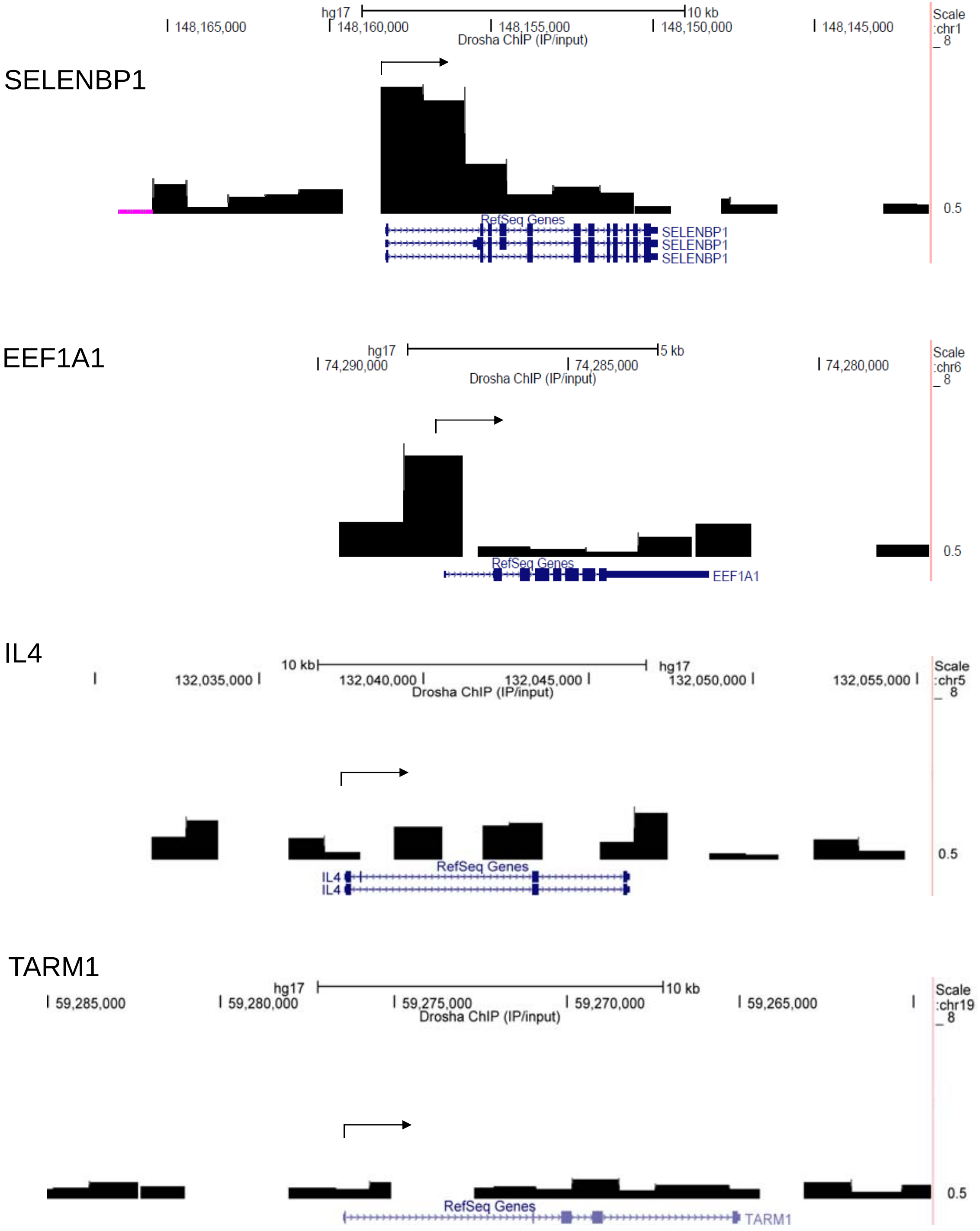

Figure S1, related to Figure 1A.

Drosha ChIP-on-chip profile across SELEBP1, EEF1A, IL4 and TARM1 genes.

Drosha ChIP-on-chip profile across SELEBP1 and EEF1A1 genes showing the signal enrichment over the TSSs. No enrichment of Drosha signal was observed over IL4 and TARM1 genes, not expressed in HeLa cells. Arrow above Chip-on-chip profile indicates the position of TSS for each gene.

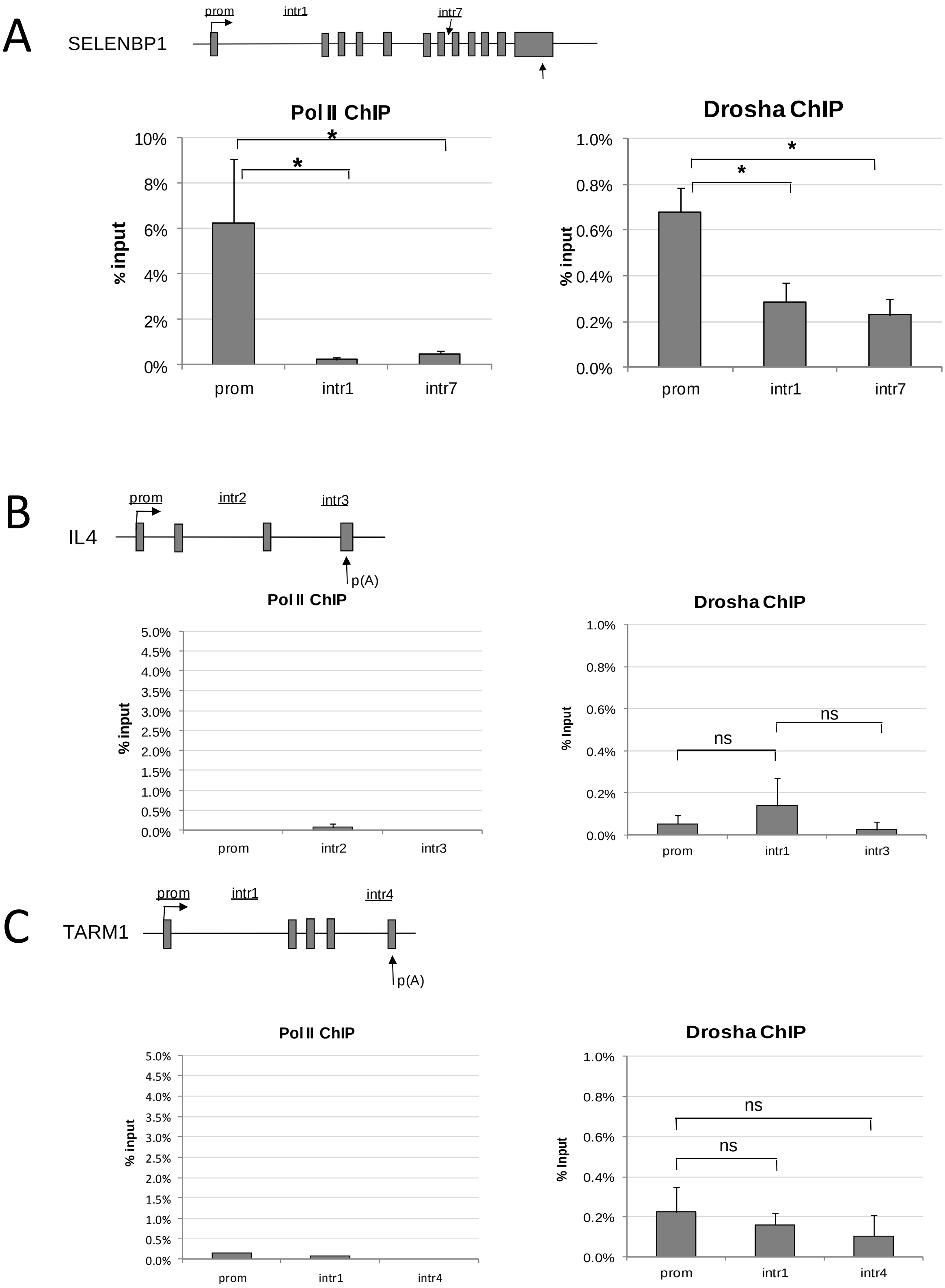

Figure S2, related to Figure 1A.
Drosha and Pol II profiles of SELEBP1, IL4 and TARM1 genes

A) Pol II (left panel) and Drosha (right panel) ChIP analysis on SELENBP1 gene. Bars represent average values from at least three independent experiments +/- SD. Position of PCR amplicons are shown above the gene diagram.
B, C) Pol II (left panels) and Drosha (right panels) ChIP profiles over IL4 (**B**) and TARM1 (**C**) genes, not expressed in HeLa cells. Amplicon positions are shown above gene diagram.

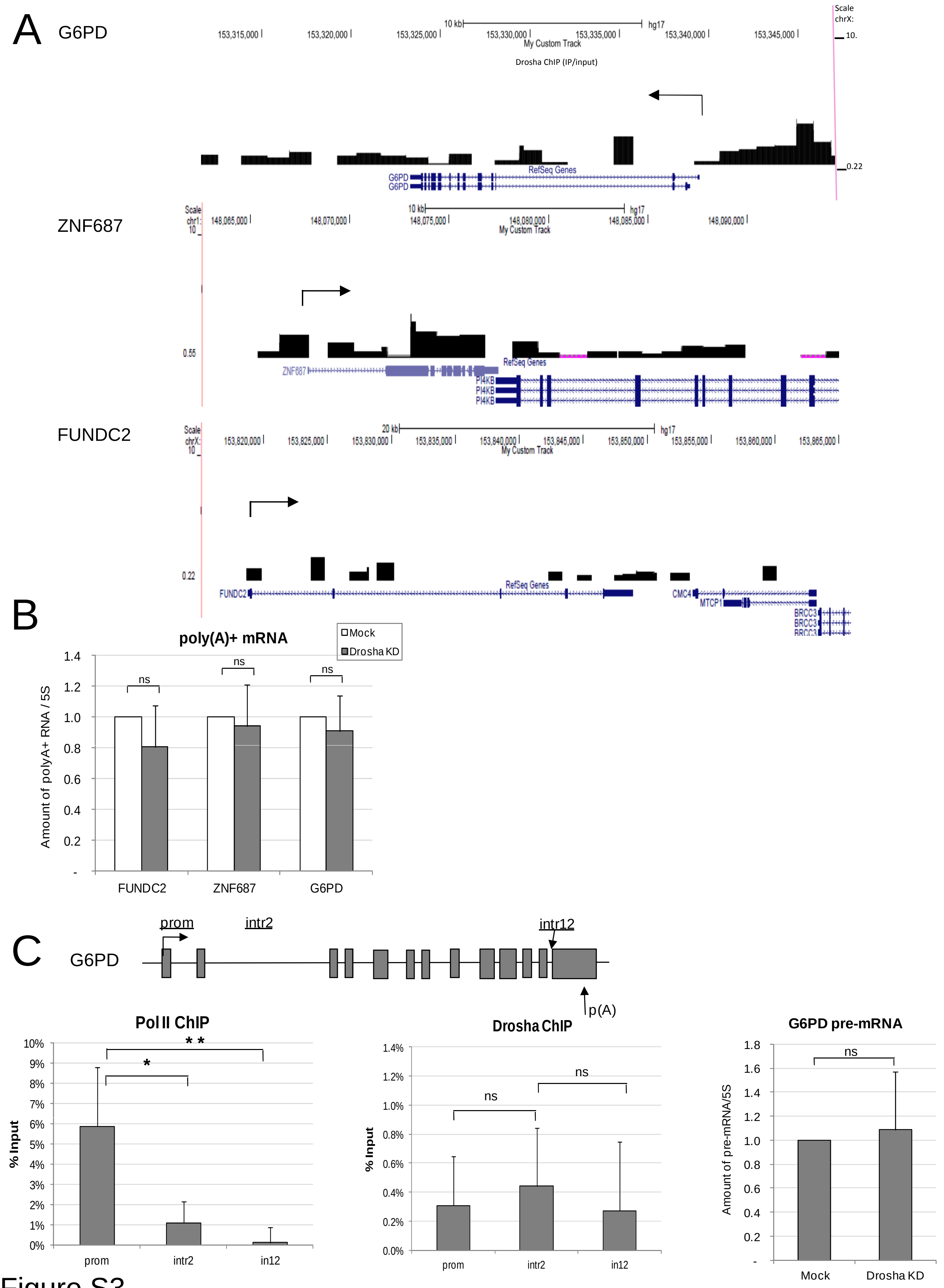
A
G6PD
153,315,000
153,320,000
153,325,000
10 kb
153,330,000
153,335,000
hg17
153,340,000
153,345,000
Scale
chrX:
10.
My Custom Track
Drosha ChIP (IP/input)
0.22
RefSeq Genes
G6PD
G6PD
ZNF687
Scale
chr1:
10
148,065,000
148,070,000
10 kb
148,075,000
148,080,000
148,085,000
hg17
148,090,000
My Custom Track
0.55
ZNF687
RefSeq Genes
PI4KB
PI4KB
PI4KB
FUNDC2
Scale
chrX:
10
153,820,000
153,825,000
20 kb
153,830,000
153,835,000
153,840,000
153,845,000
153,850,000
hg17
153,855,000
153,860,000
153,865,000
My Custom Track
0.22
RefSeq Genes
FUNDC2
CMC4
MTCP1
BRCC3
BRCC3
BRCC3
B
poly(A)+ mRNA
□Mock
▨Drosha KD
Amount of polyA+ RNA / 5S
1.4
1.2
1.0
0.8
0.6
0.4
0.2
-
ns
ns
ns
FUNDC2
ZNF687
G6PD
C
G6PD
prom
intr2
intr12
p(A)
Pol II ChIP
% Input
10%
9%
8%
7%
6%
5%
4%
3%
2%
1%
0%
* *
*
prom
intr2
in12
Drosha ChIP
% Input
1.4%
1.2%
1.0%
0.8%
0.6%
0.4%
0.2%
0.0%
ns
ns
prom
intr2
in12
G6PD pre-mRNA
Amount of pre-mRNA/5S
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
-
ns
Mock
Drosha KD

Figure S3

Figure S3, related to Figure 1A.

Analysis of G6PD, ZNF687 and FUNDC2 genes

A) Drosha ChIP-on-chip profile across G6PD, ZNF687 and FUNDC2 genes. Arrow above Chip-on-chip profile indicates the position of TSS.

B) The levels of poly(A)+ mRNA for FUNDC2, ZNF687 and G6PD genes detected in mock-treated and Drosha knocked-down HeLa cells. RNA levels were measured using qRT-PCR and normalised to 5S rRNA. RNA levels in mock-treated cells were taken as 1.

C) Top panel: Diagram of G6PD gene with positions of PCR amplicons used for RT-qPCR.

Pol II (left panel) and Drosha (midlle panel) ChIP analysis on G6PD gene. Position of PCR amplicons are shown above the gene diagram. Right panel: The levels of pre-mRNA G6PD transcripts detected in mock-treated and Drosha knocked-down HeLa cells. RNA levels were measured using qRT-PCR and normalised to 5S rRNA. RNA levels in mock-treated cells were taken as 1.

Bars represent average values from at least three independent experiments +/- SD.

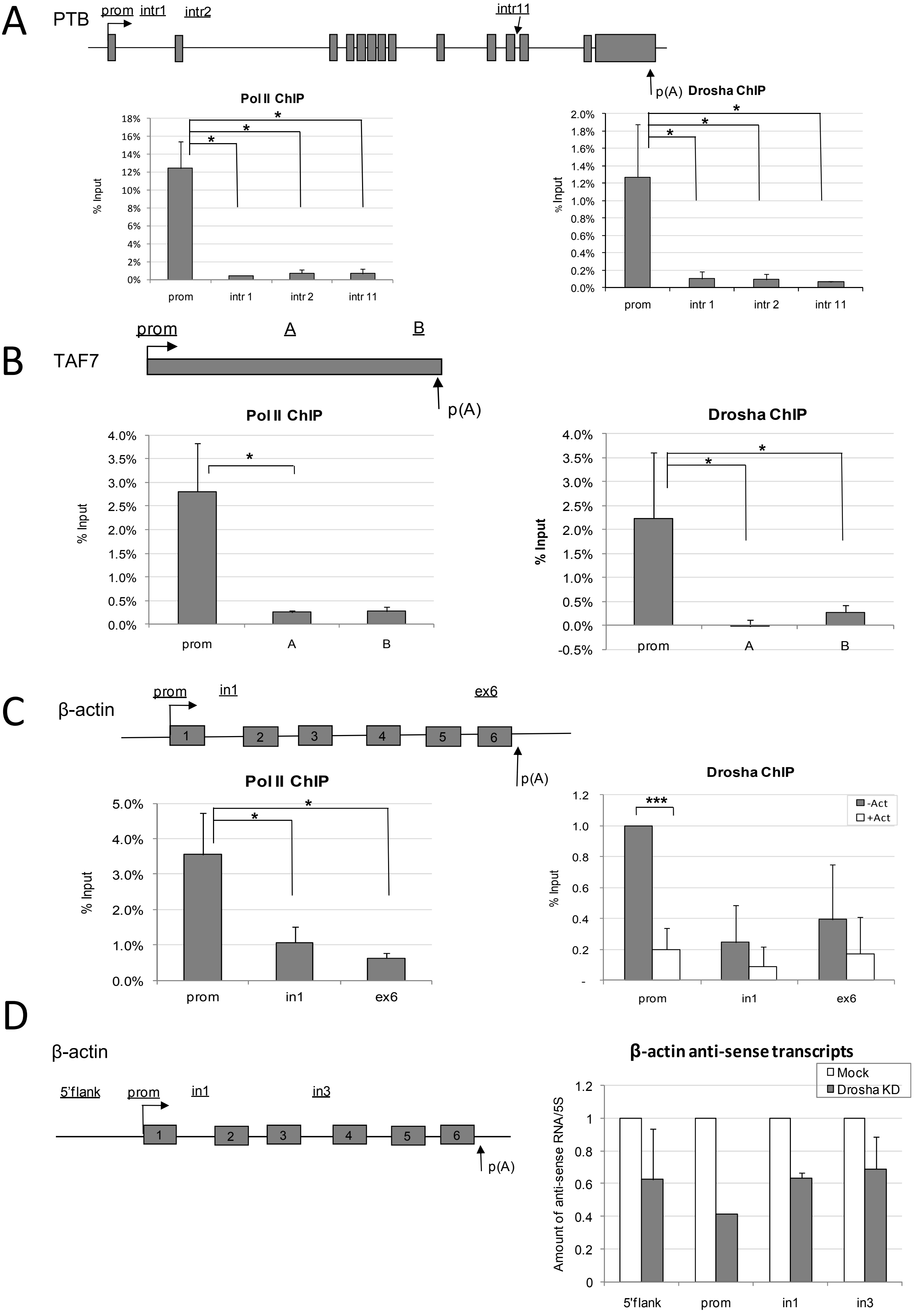
A
PTB
prom
intr1
intr2
intr11
p(A)
Pol II ChIP
% Input
18%
16%
14%
12%
10%
8%
6%
4%
2%
0%
*
*
*
prom
intr 1
intr 2
intr 11
Drosha ChIP
% Input
2.0%
1.8%
1.6%
1.4%
1.2%
1.0%
0.8%
0.6%
0.4%
0.2%
0.0%
*
*
*
prom
intr 1
intr 2
intr 11
B
TAF7
prom
A
B
p(A)
Pol II ChIP
% Input
4.0%
3.5%
3.0%
2.5%
2.0%
1.5%
1.0%
0.5%
0.0%
*
prom
A
B
Drosha ChIP
% Input
4.0%
3.5%
3.0%
2.5%
2.0%
1.5%
1.0%
0.5%
0.0%
-0.5%
*
*
prom
A
B
C
β-actin
prom
in1
ex6
1
2
3
4
5
6
p(A)
Pol II ChIP
% Input
5.0%
4.0%
3.0%
2.0%
1.0%
0.0%
*
*
prom
in1
ex6
Drosha ChIP
% Input
1.2
1.0
0.8
0.6
0.4
0.2
-
***
-Act
+Act
prom
in1
ex6
D
β-actin
5'flank
prom
in1
in3
1
2
3
4
5
6
p(A)
β-actin anti-sense transcripts
Mock
Drosha KD
Amount of anti-sense RNA/5S
1.2
1
0.8
0.6
0.4
0.2
0
5'flank
prom
in1
in3

Figure S4

Figure S4, related to Figure 1E.

Drosha binds human genes in a transcription-dependent manner

A,B,C) Drosha binds 5' ends of human PTB, TAF7 and β-actin genes. Chromatin-immuno-precipitation (ChIP) experiments in HeLa cells with Pol II (left panel) and Drosha (right panel) antibodies on PTB (**A**), TAF7 (**B**) and β-actin (**C**) genes.

C) Right panel: Drosha ChIP experiments across β-actin gene were carried out in mock-treated HeLa cells (grey bars) or cells, treated with 5ug/ml of Actinomycin D for 6 hours (white bars). Primers used for ChIP analysis are shown above the gene diagrams. Bars represent the average values from at least three independent experiments +/- SD.

D) Level of β-actin anti-sense transcripts detected in Drosha-depleted HeLa cells.

Left panel: Diagram of β-actin gene with positions of PCR amplicons used for RT-qPCR.
Right panel: The levels of anti-sense transcripts detected over the 5'gene region of endogenous β-actin gene in mock-treated (white bars) and Drosha knocked-down (grey bars) HeLa cells. RNA levels were measured using qRT-PCR and normalised to 5S rRNA. RNA levels in mock-treated cells were taken as 1. Bars represent average values from at least three independent experiments +/- SD.

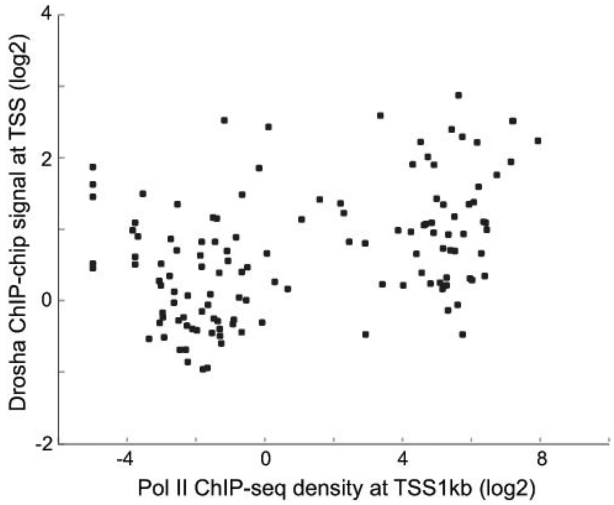

Figure S5, related to Figure 1A.

Drosha enrichment at the TSS correlates with Pol II density.

Drosha signal for each gene is shown as avarage log2(IP/input) for chip probes overlaping the TSS (extended 250bp on both sides). Average Pol II density was calculated in 1kb region around the TSS using data from HeLa Pol II ChIP-seq analysis from ENCODE/Stanford/Yale/USC/Harvard Transcription Factor Binding Sites dataset (track downloaded from the UCSC depository: http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeSydhTfbs/wgEncodeSydhTfbsHelas3Pol2StdSig.bigWig). Only refseq transcripts covered by at least one probe over the TSS and the gene body in Drosha array are shown. There is weak (Pearson R=0.41), but significant (p=1e-6) positive correlation between the Drosha probe signal at TSS and PolII ChIP density.

DGCR8 CLIP

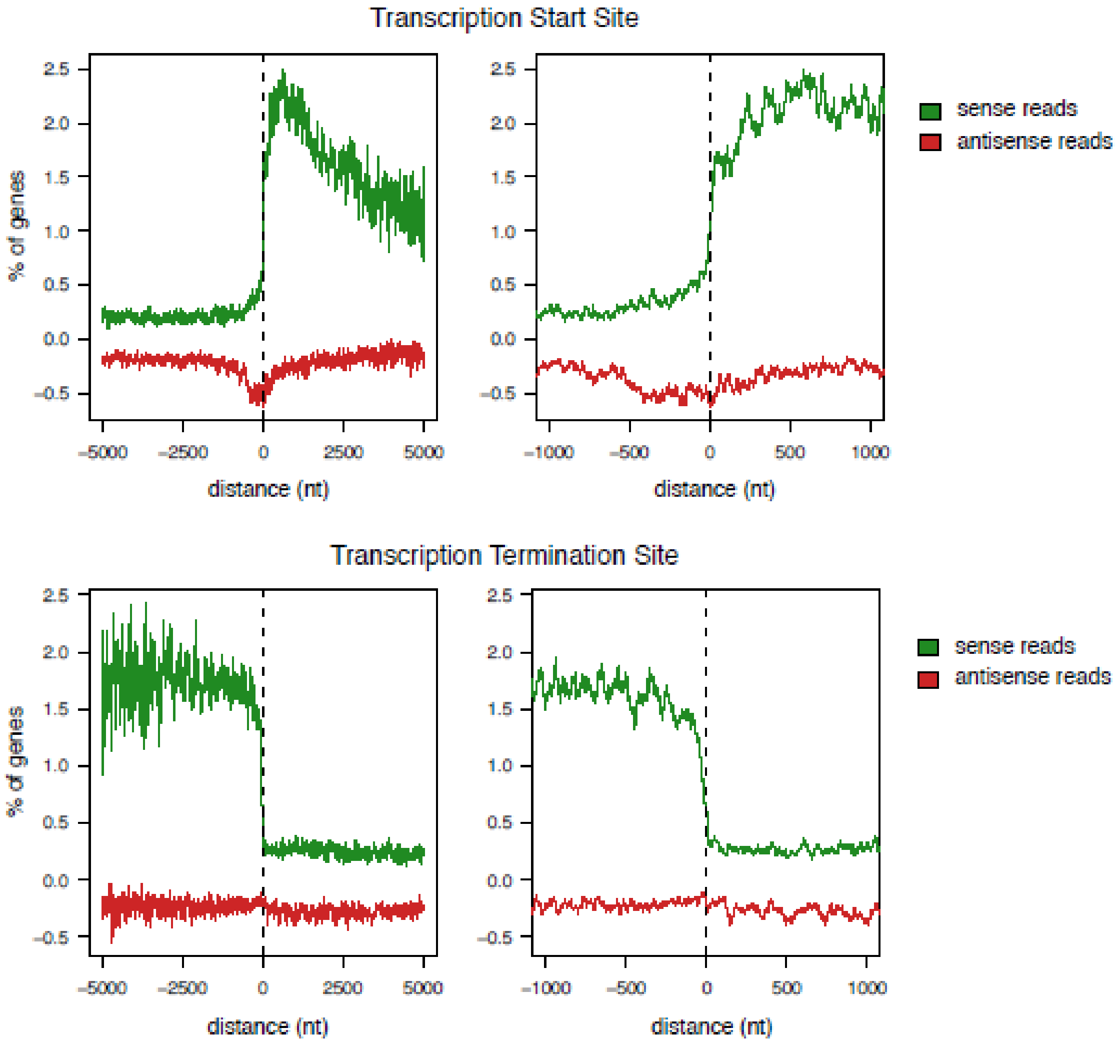

Figure S6, related to Figure 2B.

DGCR8 HITS-CLIP.

Distribution of significant sense (green) and anti-sense (red) DGCR8 CLIP reads mapping around transcription start sites (TSS; top panels) and transcription termination regions (TTS; bottom panels) of protein coding genes using mRNA transcripts (cDNA) rather than pre-mRNA (Fig. 2B), as determined by R software. The Y axis shows the number of genes detected as a percentage, whereas the X axis represents the location in relation to the position of TSS and TTS.

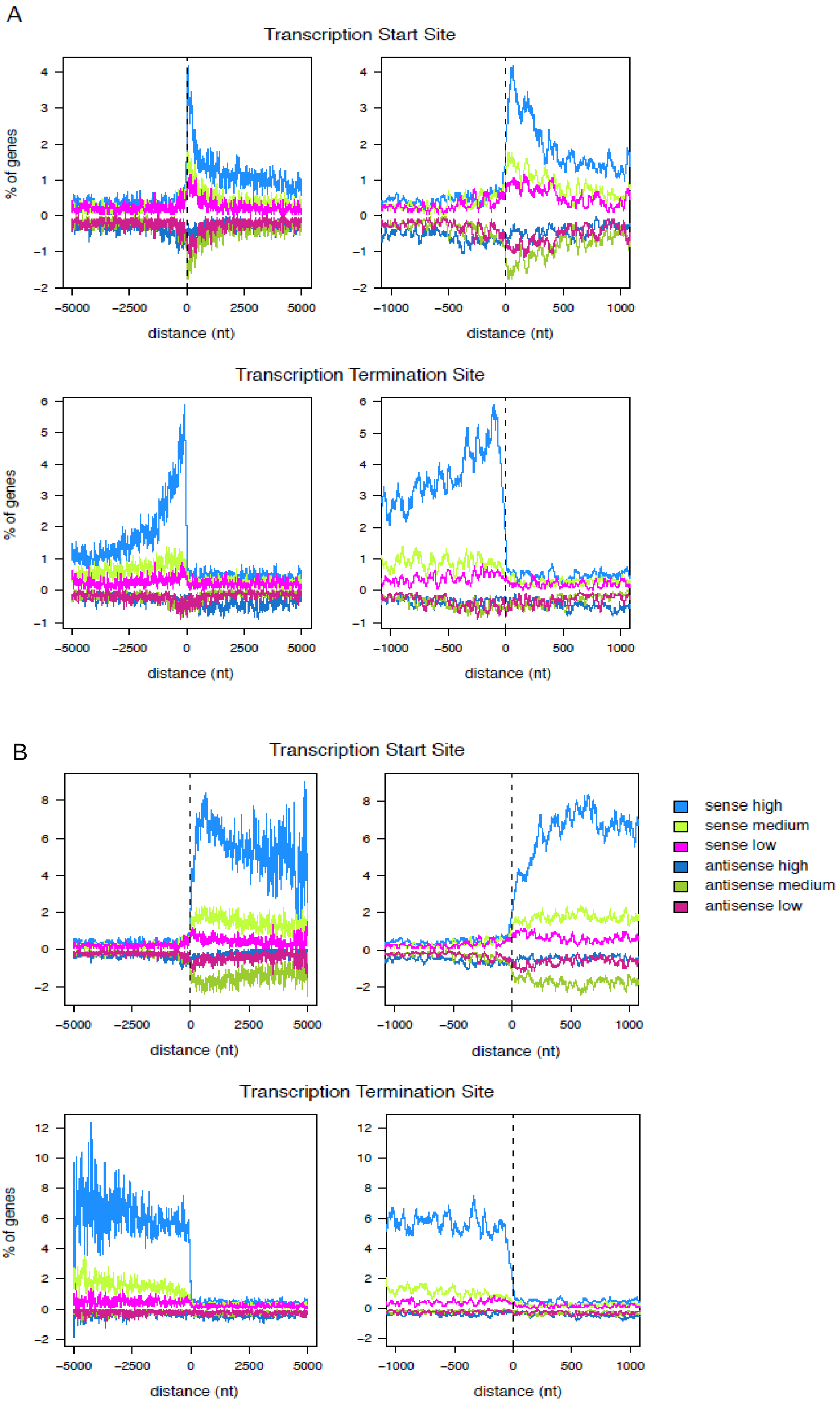
A
Transcription Start Site
% of genes
4
3
2
1
0
−1
−2
−5000
−2500
0
2500
5000
distance (nt)
−1000
−500
0
500
1000
distance (nt)
Transcription Termination Site
% of genes
6
5
4
3
2
1
0
−1
−5000
−2500
0
2500
5000
distance (nt)
−1000
−500
0
500
1000
distance (nt)
B
Transcription Start Site
8
6
4
2
0
−2
−5000
−2500
0
2500
5000
distance (nt)
−1000
−500
0
500
1000
distance (nt)
Transcription Termination Site
% of genes
12
10
8
6
4
2
0
−2
−5000
−2500
0
2500
5000
distance (nt)
−1000
−500
0
500
1000
distance (nt)
sense high
sense medium
sense low
antisense high
antisense medium
antisense low

Figure S7

Figure S7, related to Figure 2B.

Distribution of significant reads mapping around transcription start sites (TSS; top panels) and transcription termination regions (TTS; bottom panels) of protein coding genes on the genomic DNA (A) and cDNA (B) for genes with high, medium or low expression levels both in sense (blue, green and magenta) and in antisense (dark blue, dark green and dark magenta) orientation. The Y axis shows the number of genes detected as a percentage, whereas the X axis represents the location in relation to the position of TSS and TTS. Number of genes analysed in each category: high expression level (2727); low expression level (2727); medium expression level (2726).

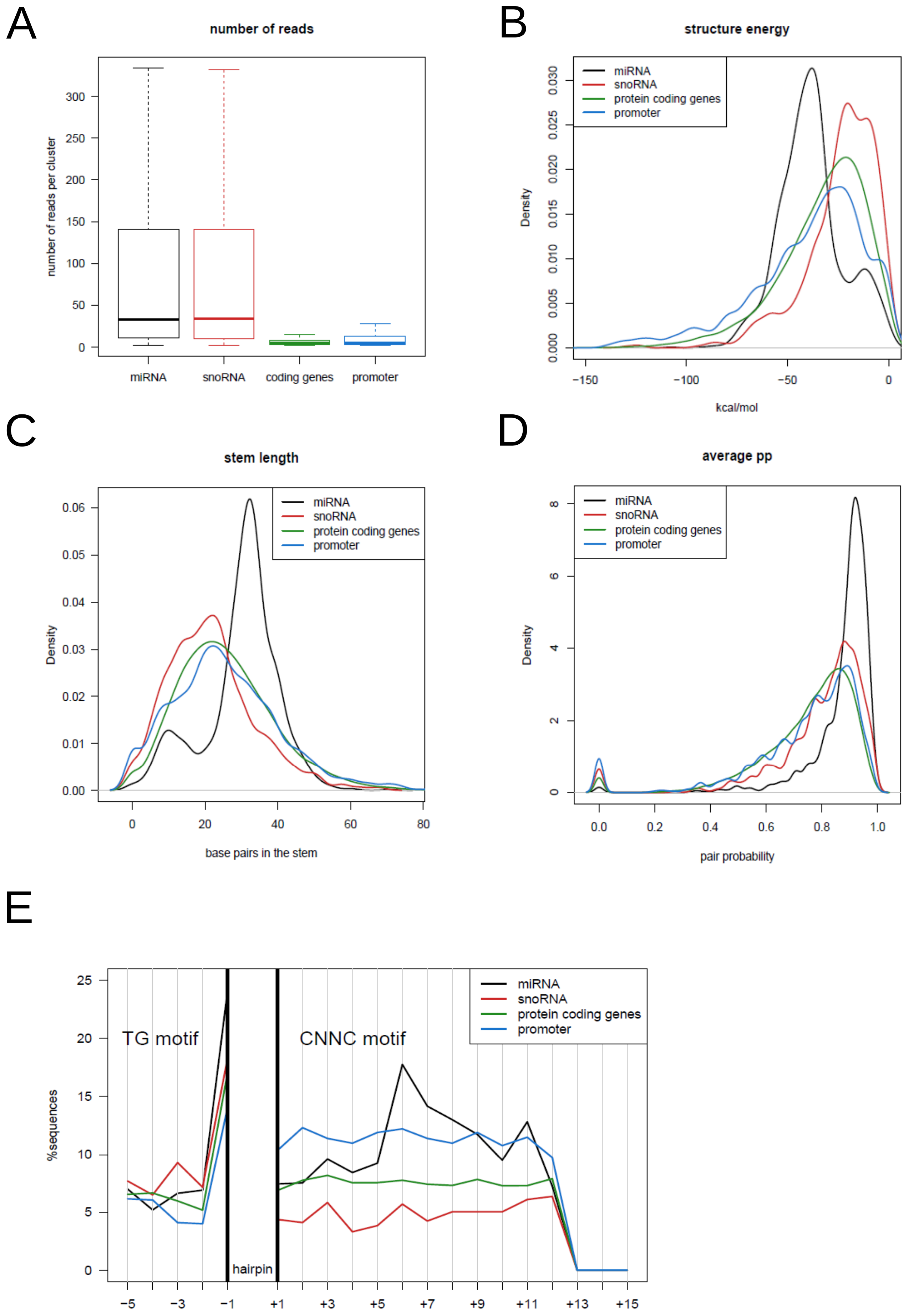
A
number of reads
number of reads per cluster
0
50
100
150
200
250
300
miRNA
snoRNA
coding genes
promoter
B
structure energy
Density
0.000
0.005
0.010
0.015
0.020
0.025
0.030
−150
−100
−50
0
kcal/mol
miRNA
snoRNA
protein coding genes
promoter
C
stem length
Density
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0
20
40
60
80
base pairs in the stem
miRNA
snoRNA
protein coding genes
promoter
D
average pp
Density
0
2
4
6
8
0.0
0.2
0.4
0.6
0.8
1.0
pair probability
miRNA
snoRNA
protein coding genes
promoter
E
%sequences
0
5
10
15
20
25
TG motif
CNNC motif
hairpin
−5
−3
−1
+1
+3
+5
+7
+9
+11
+13
+15
miRNA
snoRNA
protein coding genes
promoter

Figure S8

Figure S8, related to Discussion section.

Bioinformatic analysis of the structures formed by Microprocessor targets

A. Number of reads detected for each class of structures (promoters, snoRNA, miRNAs and coding genes) in DGCR8 CLIP experiments. Clusters in promoters have a significantly lower amount of reads that miRNAs and snoRNAs (Kolmogorov-Smirnov test p-value < 2.2e-16 and 9.691e-08 respectively).

B. Secondary structure energy. The optimal secondary structures in promoters have a higher minimum free energy than miRNAs.

C. Stem length. Structures predicted in promoters are shorter than those in miRNAs, suggesting their potential less efficient structure recognition by the Microprocessor complex.

D. Pair probability. The pair probability gives a measure of structure stability by evaluating the probability of the predicted base pairs. Pair probability of 1 means that the nucleotides will always be basepairing with each other, whereas a pair probability of 0.5 means that they will only be basepaired in 50% of the cases. Structures predicted in promoters have a lower pair probability than miRNA regions, pointing towards their less efficient recognition by the Microprocessor.

E. Presence of motifs important for the processing of the stem. These motifs are TG dinucleotide at the base of the stem at positions -14 and -13 and CNNC motif downstream of the stem at position +6-+8, relative to the cleavage site, as described in (Auyeung et al., 2013). Since the position of cleavage site is not defined in such structures, the presence of the motifs was analysed in relation to the base of the stem predicted. 20% of the structures predicted to overlap with pri-miRNAs had a TG motif at the base of the stem and a high CNNC bias at positions +6-+8 after the stem. In the case of promoter structures, TG bias upstream of the stems is the lowest of all the groups analyzed, and that there is no positional bias of the CNNC motif after the stem.

Supplemental Experimental Procedures

RNA analysis

OligodT primer was used for reverse transcription to detect the polyA+ RNA. For the detection of nascent RNA, reverse PCR primer was used for the reverse transcription reaction. The qPCR primers for amplification of pre-mRNAs were the following: PTB intr2F/R, GAPDH intr1F/R, TAF7 BF/R. SELENBP1 intr1F/R, IL4 intr 1F/R, TARM1 intr4F/spliced R. qPCR primers for the amplification of polyA+ RNAs were the following: β-actin ex5F/ex6R, PTB ex13F/ex14R, GAPDH F/R3, γ-actin spliced F/R, TAF7 BF/R, SELENBP1 ex7F/ex8R, Drosha F/R. For quantitative real-time PCR, 2 μl of cDNA was analyzed using a Rotorgene 3000 real-time PCR machine (Corbett Research) in the presence of Quantitec SYBR green (Qiagen). Cycling parameters were 95°C for 15 min, followed by 45 cycles of 95°C for 15 s, 58°C for 20 s, and 72°C for 20 s. Fluorescence intensities were plotted against the number of cycles by using an algorithm provided by the manufacturer. Primer sequences are shown in Supplemental Table 1.

Drosha ChIP-on-chip analysis

Four hundred and fifty nanograms of purified ChIP DNA was labelled with Cy3-2′-deoxycytosine 5′-triphosphate (dCTP, for IP DNA; Amersham PA53021) or Cy5-dCTP (for input DNA; Amersham PA55021) using BioPrime DNA labelling System (18094-011, Invitrogen) according to manufactures instructions. Labelled DNA was purified through Amersham G50 columns. Cy3-labelled IP DNA and Cy5-labelled input DNA were combined, precipitated in ethanol with human Cot-1 DNA (Invitrogen 15279-011) and yeast tRNA (Invitrogen 15401-011) and resuspended in hybridisation buffer (50%[v/v] formamide, 2x standard saline citrate (SSC), 10 mM Tris pH 7.4, 5% [w/v] dextran sulphate, 0.1% [v/v] Tween 20). Hybridisation solution containing the labelled DNAs was hybridised to the human ENCODE 5.1.1 array (Sanger Centre, Cambridge) in a HS 400 Pro Hybridisation Station (Tecan Austria, Groding/Salzburgh, Austria) for 45 hours. ENCODE array is represented by 44 separate tiled regions covering 30 Mb and ~481 human genes (~1% coverage). Median length of region covered by each probeset is 1143 bp. The arrays were then scanned in a Scan Array Gx Plus scanner (Perkin Elmer, Shelton, CT), faulty spots excluded and the spot intensities were quantified using Scan Array Express Version 3.0 (Perkin Elmer) with background subtraction. The background-corrected ChIP signal divided by the background-corrected input signal (both globally normalised using Lowess normalisation) were used for the analysis. The data were then plotted and visualised on a UCSC genome browser (version NCBI35/hg17).

For the calculation of metagene profile and average statistics for different annotated regions (Figure 1E i, ii), NCBI RNA reference sequence (RefSeq) annotation collection (Pruitt et al., 2005) was downloaded from UCSC genome browser track (NCBI35/hg17). Annotated transcript start and end sites were extended 250 bp on both sides before calculating probe overlaps. Gene body was defined as a region between 1.5 kb downstream of the transcript start and the transcript end. Metagene profiles of Drosha binding around annotated TSS and TTS were generated by calculating average probe signal (log2(IP/input)) for all probes covering corresponding position in annotated RefSeq genes. Only genes longer than 3 kb were used to generate these profiles. Each position represents an average signal of at least 120 probes.

DGCR8 HITS-CLIP analyses

To investigate whether the binding of DGCR8 correlates with gene expression, we divided protein coding genes from Ensembl54 into three groups according to their expression level: high, medium and low. First, we mapped mRNA-seq data from HEK 293 cells (GSM 936076; (Baltz et al., 2012) on the human genome (hg18) using BWA (Li and Durbin, 2009). Next, we estimated the expression level of genes as the number of reads mapping to the longest mRNA divided by the mRNA length. Only transcripts with at least 5 reads per 1000 nt of mRNA were kept. We ordered the remaining genes according to their expression and selected the low expressed genes (bottom 20%; quantile 0.2), the medium expressed genes (quantile 0.4-0.6) and the highly expressed genes (top 20%; quantile 0.8). To build the profiles, only reads belonging to significant clusters FDR < 0.01 (Macias et al., 2012) were considered.

Table 3, related to Experimental Procedures. Sequences of PCR primers

Name	**Sequence (5'→3')**
PTB	
prom (F)	GTC TCC GCC ATT TTG TGA GT
prom (R)	ATG GCA CAC AGA GCA GAC C
intr1 (F)	GAG AGG CTG GAG ATG TTT CG
intr1 (R)	GCC TCT TCT GAC ACC ACA GA
intr2 (F)	GCC CCT TAG GAA TGG AAA AG
intr2 (R)	GCC GTT GGT ACA AAG GTA GG
intr11(F)	CAG TGG CCG ATA AAG CAA AC
intr11(R)	CTG CAC TAG GGC GTT CTC CTT C
PTB ex13 (F)	GGC TCC AAG AAC TTC CAG AAC
PTB ex14 (R)	ATT GCT GGA AAA CAG GAC CTT
γ-actin	
prom (F)	GGA AAG ATC GCC ATA TAT GGA C
prom (R)	TCA CCG GCA GAG AAA CGC GAC
intr1 (F)	CCG CAG TGC AGA CTT CCG AG
intr1 (R)	CGG GCG CGT CTG TAA CAC GG
ex5(F)	GTG ACA CAG CAT CAC TAA GG
ex5 (R)	ACA GCA CCG TGT TGG CGT
γ-actin spliced (F)	AAT CTT GCG GCA TCC ACG AG
γ-actin spliced (R)	TCG TAC TCC TGC TTG CTA ATC C
GAPDH	
prom (F)	AGC TCA GGC CTC AAG ACC TTG GGC T
prom (R)	GGC TGA CTG TCG AAC AGG AGC
intr1 (F)	CCA CTA GGC GCT CAC TGT TC
intr1 (R)	TCG TAG ACG CGG TTC GG
intr2a (F)	TGC GGG GTC ACG TGT CGC AG
intr2a (R)	AAC GGC TGC CCA TTC ATT TC
intr2b (F)	GGT CAA CGC TAG GCT GGC AG
intr2b (R)	GTG GCA TGG TGC CAA GCC GG
ex8(F)	GGT GGT GAA GCA GGC GTC GGA GGG
ex8 (R)	GAG CCA GTC TCT GGC CCC AGC CAC
GAPDH spliced (F)	ACA TCA AGA AGG TGG TGA AG
GAPDH spliced (R3)	GGG TCT TAC TCC TTG GAG GC
TAF7	

prom (F)	CTT CCG TTT TTG CTG GGT AG
prom (R)	GAG CGT CTC CTT GTC GTT TC
A (F)	AGA CCT GCC CTG TGT TAT GG
A (R)	CTA GCA ACT GGC TCC TCC AC
B (F)	TTG CAG GAT TAG CGG AAT GT
B (R)	ACG TCA CAA AGC AGC AA
β-actin	
β-actin ex4 (F)	TCGTGCGTGACATTAAGGAG
β-actin ex4 (R)	GTCAGGCAGCTCGTAGCTCT
β-actin ex5 (F) spliced	GGA CAT CCG CAA AGA CCT GTA
β-actin ex6 (R) spliced	CTC CAA CCG ACT GCT GTC ACC
prom (F)	GAG GGG AGA GGG GGT AAA
prom (R)	AGC CAT AAA AGG CAA CTT TCG
In1 (F)	CGG GGT CTT TGT CTG AGC
In1 (R)	CAG TTA GCG CCC AAA GGA C
Ex 6 (F)	GGA GCT GTC ACA TCC AGG GTC
Ex6 (R)	TGC TGA TCC ACA TCT GCT GG
IL4	
prom (F)	TAT CTT TGT CAG CAT TGC ATC
prom(R)	ACA AAG TTG CCG GCA CAT GCT
Intr1 (F)	TGA AGG GTT TCT TGG GTG GA
Intr1(R)	GAG GTT CAT TAT GGA ACT CTC TG
intr3(F)	TTC AGG TGA CAA GTG CCA CAG
IL4 spliced (R)	CTG GTT GGC TTC CTT CAC AG
IL4 spliced (F)	TCC TGA AAC GGC TCG ACA G
T**ARM1**	
prom(F)	TGT TAC TGC CCA CAC TCT GG
prom(R)	GGG ATC ATG ATG GCT CCT TAG
intr1(F)	GCT CAC GCC TGT AAT CCC AG
intr1(R)	TAG TAG AAA TGG GGT TTC ACC
intr4 (F)	CAT TAA GTG GGT GAT GAT AGT CAG
TARM1 spliced (F)	CAG CTT GAG ATA TTG GTG ACA G
TARM1 spliced (F)	GTC GTA CGA AGT TAC CCA GG
SELENBP1	
prom(F)	CAG CTG GTT GTA TAA ATT CCC
prom(R)	TTG CTG TGC TGG TGT CAG AG
intr1 (F)	CTG CAC CCT CGT GTT TAC TTC
intr1(R)	AGC TTG GGC AAA GTA CTG AAG
intr7(F)	TTC TGA TAG GCT GAC CTC TCT G
Ex7 spliced (F)	CCA TCC AGC GCT TCT ACA AG

Ex8 spliced (R)	CAG CCA GCC CTT CAC TTT C
G6PD	
prom (F)	GGA AAC GGT CGT ACA CTT CG
prom (R)	CAA ACA GCG TGT ATT TTA CCG
intr (F)	AGG AGA GGT ACC AGG TGG AG
intr2(R)	TGT CAA ACT CCT GAC CTC AG
intr12 (F)	GCC TCC CAA GCC ATA CTA TG
intr12 (R)	CCT GCC ATA AAT ATA GGG GAT G
G6PD spliced (F)	AAC GTG AAG CTC CCT GAC G
G6PD spliced (R)	CCT GCC ATA AAT ATA GGG GAT G
ZNF687 F	TAG AGA AAC ATG TCC AGG TCC
ZNF687 R	TCA CTG CAA GAG TCA GAA GAC
FUNDC2 F	CAA AGA GCA GCT GAA GAT CCG
FUNDC2 R	CCC CCA GTT ACT AGA ACA TTC
5S (F)	AGC GTC TAC GGC CAT ACC
5S (R)	GGT ATT CCC AGG CGG TCT C
miR-330 (F)	CCT TCT TCC AGG ATC GCG TC
miR-330 (R)	GAG GTC TCC GAT GAA AAC GG
T7+β-actin (F)	TTA TCG AAA TTA ATA CGA CTC ACT ATA GGG AGAC CTT CGC CCG TGC AGA GCC
T7+β-actin (R)	CAG ATT GGG GAC AAA GGA AG
T7+miR-17-19 (F)	TTA TCG AAA TTA ATA CGA CTC ACT ATA GGG AGAC GAA TTC TTA AGG CAT AAA TAC G
miR-17-19 (R)	GAT AAC TAA ACA CTA CCT GC
Drosha (F)	GAG ACC TAG CCT AGT TTT CCTG
Drosha (R)	AAT GCA CAT TCA CCA AAG TCAA
EGFP (F)	GAC GTA AAC GGC CAC AAG TTC
EGFP (R)	TGG TGC AGA TGA ACT TCA G

Table S1, related to Figure 1A. Drosha ChIP-on-chip analysis

A) List of RefSeq genes with at least one probe covering both the TSS region and the gene body. Both number of probes and their average signal (log2(IP/input)) are represented. Last column indicates the level of enrichment of the Drosha signal at TSS compared to the gene body.

B) Drosha ChIP-on-chip data.

List of NCBI35/hg17 genomic coordinates together with background-corrected Drosha-ChIP signal divided by the background-corrected input signal for each ENCODE 5.1.1 array probe (faulty spots exlcluded) .

Table S2, related to Figure 2B, C. DGCR8 HITS-CLIP data.

List of anti-sense clusters mapping to promoter regions of Ensembl genes. List of DGCR8 sense clusters is presented in (Macias et al., 2012). Promoters were defined as the 1000nt upstream of the TSS plus 200nt downstream of the TSS.

Supplemental references

Auyeung, V.C., Ulitsky, I., McGeary, S.E., and Bartel, D.P. (2013). Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. Cell *152*, 844-858.

Baltz, A.G., Munschauer, M., Schwanhausser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M.*, et al.* (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol Cell *46*, 674-690.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.

Macias, S., Plass, M., Stajuda, A., Michlewski, G., Eyras, E., and Caceres, J.F. (2012). DGCR8 HITS-CLIP reveals novel functions for the Microprocessor. Nat Struct Mol Biol *19*, 760-766.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res *33*, D501-504.