Cell Reports
Report

Drosha Regulates Gene Expression Independently of RNA Cleavage Function

Natalia Gromak,[1,*] Martin Dienstbier,[1] Sara Macias,[2] Mireya Plass,[3,4] Eduardo Eyras,[3,5] Javier F. Cáceres,[2] and Nicholas J. Proudfoot[1,*]
[1]Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK
[2]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK
[3]Computational Genomics Group, Universitat Pompeu Fabra, Dr. Aiguader 88, 08003 Barcelona, Spain
[4]The Bioinformatics Centre, Department of Biology, University of Copenhagen, 2200 Copenhagen, Denmark
[5]Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain
*Correspondence: natalia.gromak@path.ox.ac.uk (N.G.), nicholas.proudfoot@path.ox.ac.uk (N.J.P.)
http://dx.doi.org/10.1016/j.celrep.2013.11.032

SUMMARY

Drosha is the main RNase III-like enzyme involved in the process of microRNA (miRNA) biogenesis in the nucleus. Using whole-genome ChIP-on-chip analysis, we demonstrate that, in addition to miRNA sequences, Drosha specifically binds promoter-proximal regions of many human genes in a transcription-dependent manner. This binding is not associated with miRNA production or RNA cleavage. Drosha knockdown in HeLa cells downregulated nascent gene transcription, resulting in a reduction of polyadenylated mRNA produced from these gene regions. Furthermore, we show that this function of Drosha is dependent on its N-terminal protein-interaction domain, which associates with the RNA-binding protein CBP80 and RNA Polymerase II. Consequently, we uncover a previously unsuspected RNA cleavage-independent function of Drosha in the regulation of human gene expression.

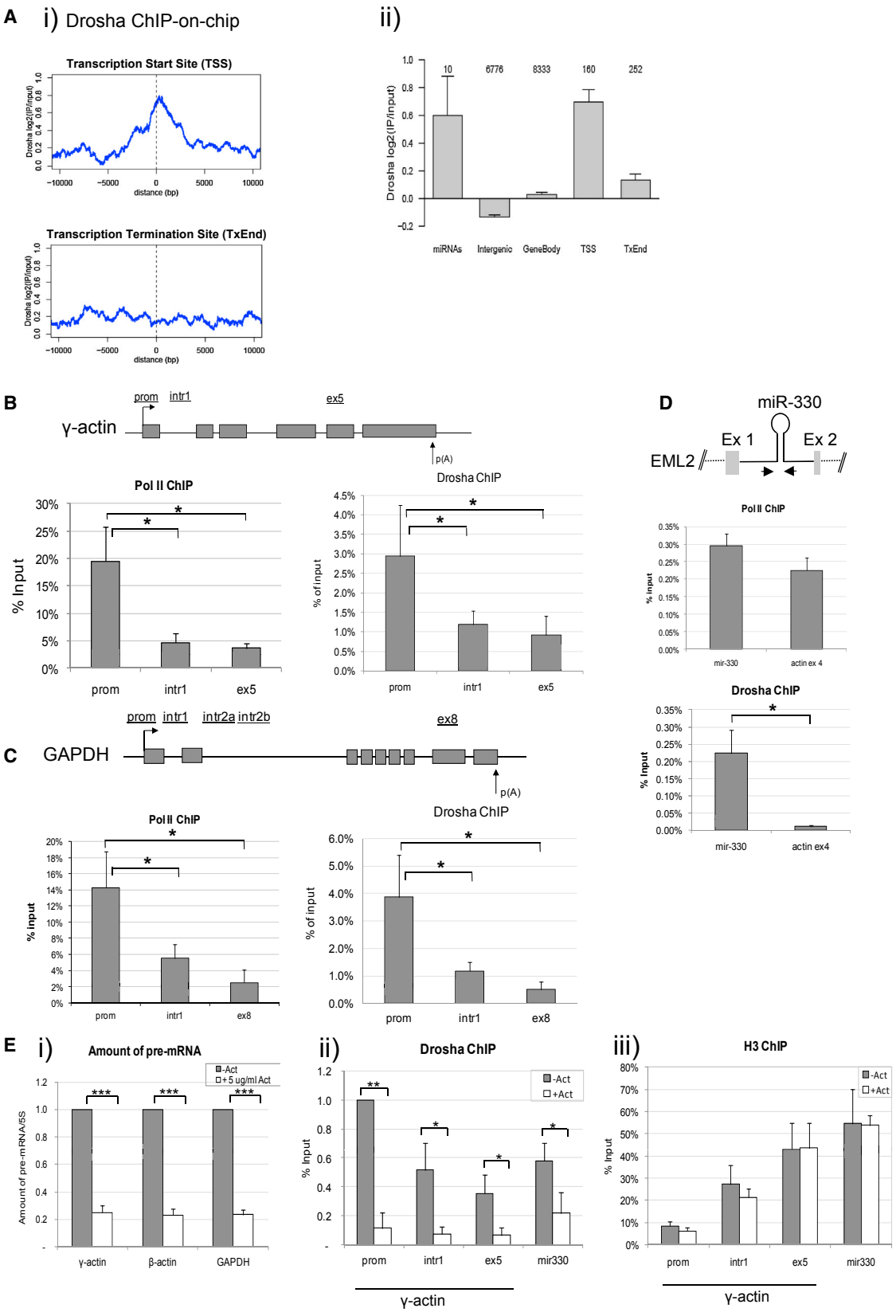
INTRODUCTION

MicroRNAs (miRNAs) are short noncoding regulatory molecules, involved in diverse biological processes. Biogenesis of miRNAs involve a nuclear phase, where the Microprocessor complex, comprising Drosha, an RNase III-like enzyme and its cofactor DGCR8, process primary miRNAs (pri-miRNAs) into a 70 nt pre-miRNA (Han et al., 2004; Lee et al., 2003; Zeng et al., 2005). This occurs cotranscriptionally from both independently transcribed and intron-encoded miRNAs (Ballarino et al., 2009; Kim and Kim, 2007; Morlando et al., 2008). Following Drosha-mediated RNA cleavage and pre-miRNA release from the nascent RNA, 5′ and 3′ nascent RNA ends are trimmed by 5′-3′ Xrn2 and 3′-5′ exosome (Morlando et al., 2008), and the pre-miRNA precursor is exported to the cytoplasm (Lund et al., 2004; Yi et al., 2003). Here, a second RNase III enzyme, Dicer, further processes the pre-miRNA into the mature miRNA duplex (Bernstein et al., 2001) that targets specific mRNAs for degradation or translational inactivation (reviewed in Bartel, 2009).

MiRNA levels are tightly regulated at the posttranscriptional level by a number of RNA-binding proteins (Siomi and Siomi, 2010). Furthermore, Drosha can directly regulate levels of Microprocessor complex by cleaving hairpin structures in DGCR8 mRNA, thereby decreasing DGCR8 protein levels (Han et al., 2009; Triboulet et al., 2009). Along the same lines, Drosha knockdown in *Drosophila* leads to upregulation of some mRNAs containing conserved RNA hairpins, potentially recognized by the Microprocessor complex (Kadener et al., 2009). Several recent studies demonstrated the ability of Microprocessor complex to cleave mRNAs, thus regulating their expression. Many Drosha-dependent mRNA cleavage events were identified in mESCs, consistent with Microprocessor regulation of coding mRNAs through direct cleavage (Karginov et al., 2010). Drosha can also cleave the TAR hairpin of the HIV-1 transcript, resulting in premature termination of RNA polymerase II (Pol II) (Wagschal et al., 2012). A recent DGCR8 HITS-CLIP analysis extended these observations and revealed general noncanonical functions of the Microprocessor complex (Macias et al., 2012). Transcriptome and proteome studies of mice lacking Drosha and Dicer suggest that both enzymes have nonredundant functions, as their deficiency can induce different phenotypes (Chong et al., 2010). Although many RNAs were stabilized by Drosha depletion, some were downregulated, consistent with Drosha possessing independent functions to its role in canonical miRNA biogenesis.

In human cells Drosha exists in two distinct multiprotein complexes (Gregory et al., 2004). The smaller complex, containing just Drosha and DGCR8, is necessary and sufficient for miRNA processing. The larger complex, displaying only weak pre-miRNA processing activity in vitro, contains DEAD-box RNA helicases, double-stranded RNA-binding proteins, hnRNP proteins, members of FUS/TLS family of proteins, and the SNIP1 protein, implying additional functions in gene expression. Thus, DEAD

Cell PRESS
Open ACCESS



(legend on next page)

box helicases p68/p72 increase Drosha processing efficiency for a subset of miRNAs and at gene-specific promoters interact with transcriptional coactivators and Pol II and regulate alternative splicing (Fuller-Pace and Ali, 2008). Nuclear scaffolding protein hnRNPU and members of FUS/TLS family are also associated with regulation of transcription (Wang et al., 2008). SNIP1, a component of a large SNIP1/SkIP-associated complex, involved in transcriptional regulation and cotranscriptional processing, interacts with Drosha and plays a role in miRNA biogenesis (Fujii et al., 2006; Yu et al., 2008). Ars2 is implicated in RNA silencing that functions in antiviral defense in flies and cell proliferation in mammals (Gruber et al., 2009; Sabin et al., 2009). It interacts with the nuclear cap-binding complex (CBP20/CBP80) and is involved in miRNA biogenesis, suggesting a link between RNA silencing and RNA-processing pathways. CBP20/CBP80 proteins are also implicated in miRNA biogenesis in plants (Kim et al., 2008). Overall, the existence of this large Drosha-complex with only weak miRNA-processing activity suggests that Drosha may play multiple roles in miRNA-independent gene regulation.

Using genome-wide and gene-specific approaches we now show that Drosha binds to the promoter-proximal regions of many human genes in a transcription-dependent manner. Similarly, DGCR8 binds promoter-proximal regions of many human genes, suggesting that the whole Microprocessor is recruited at promoter regions. We also find that Drosha interacts with Pol II and its depletion from human cells causes transcriptional downregulation with a concomitant decrease in nascent and mature mRNA levels. This positive function of Drosha in gene expression is mediated through its interaction with the RNA-binding protein CBP80 and dependent on the N-terminal protein-interaction domain of Drosha. Thus, results presented in this paper demonstrate an miRNA- and cleavage-independent function of Drosha.

RESULTS

Drosha Binds 5′ Ends of Human Genes in a Transcription-Dependent Manner

Pre-miRNA processing occurs cotranscriptionally. Consequently binding of Drosha to nascent miRNA sequences can be detected using chromatin immunoprecipitation (ChIP) analysis (Morlando et al., 2008). To investigate genome-wide binding of Drosha, we employed a ChIP-on-chip approach using human 5.1.1 ENCODE array. Our results show that, in addition to miRNA regions, Drosha binds many human genes. This binding is enriched at the transcriptional start sites (TSSs), compared to the ends of transcripts (TxEnd), other gene body, or intergenic regions (Figure 1A; full gene list in Table S1). In particular, we observed that, out of 160 TSS-proximal probes, 55 (34%) had more than 2-fold enrichment of Drosha-ChIP signal over the input signal. This is significantly more than 1,669 (10%) out of 15,998 total probes or 953 (11%) out of 8,333 probes in the gene bodies ($p < 0.0001$, Fisher's exact test). The same level of enrichment was observed for four (40%) out of ten probes overlapping with annotated miRNA regions (Figure 1Aii). By way of illustration, we present SELENPB1 and EEF1A1 genes, which scored positively in our genomic analysis, where Drosha selectively binds to 5′-proximal gene regions (Figure S1). Genes that are not expressed in HeLa cells such as IL4 and TARM1 demonstrated only background Drosha signal (Figure S1). We also identified genes, such as G6PD, ZNF687, and FUNDC2, which are expressed in HeLa cells but did not demonstrate any significant Drosha enrichment (Figure S3A). We further confirmed these array data using gene-specific primers in ChIP experiments. In particular, we observed an enrichment of Drosha binding to promoter-proximal region of SELENBP1 gene, colocalizing with Pol II peak (Figure S2A). We did not detect any significant binding of Drosha or Pol II to IL4 and TARM1 genes, further confirming ChIP-on-chip results (Figures S2B and S2C). We also did not detect any significant Drosha binding over promoter-proximal region, enriched for Pol II, on G6PD gene (Figure S3C).

To investigate the function of Drosha at the beginning of human genes in more detail, we compared Pol II and Drosha binding profiles for specific human genes. Both Pol II and Drosha were enriched over the promoter-proximal regions of human γ-actin, GAPDH, PTB, β-actin, and intronless TAF7 genes (Figures 1B, 1C, and S4A–S4C). The enrichment of Drosha observed over these genes is higher than over the miR-330 locus within the EML2 gene, used as a positive control for Drosha binding (Morlando et al., 2008) (Figure 1D). As a negative control, we used ChIP for histone H3, which was found to be depleted from promoter-proximal regions and enriched in the body of these genes (Figure 1Eiii). To test if Drosha binding to these genes is transcription dependent, we treated HeLa cells with Pol II transcriptional inhibitor actinomycin D. Following this treatment, we observed ~80% reduction in pre-mRNA levels and a substantial reduction in Drosha binding to miR-330 locus, γ-actin, and β-actin genes (Figures 1E and S4C). In contrast, we saw no effect on histone H3 binding (Figure 1Eiii). These results indicate that, in addition to miRNA sequences, Drosha binds 5′ end regions of

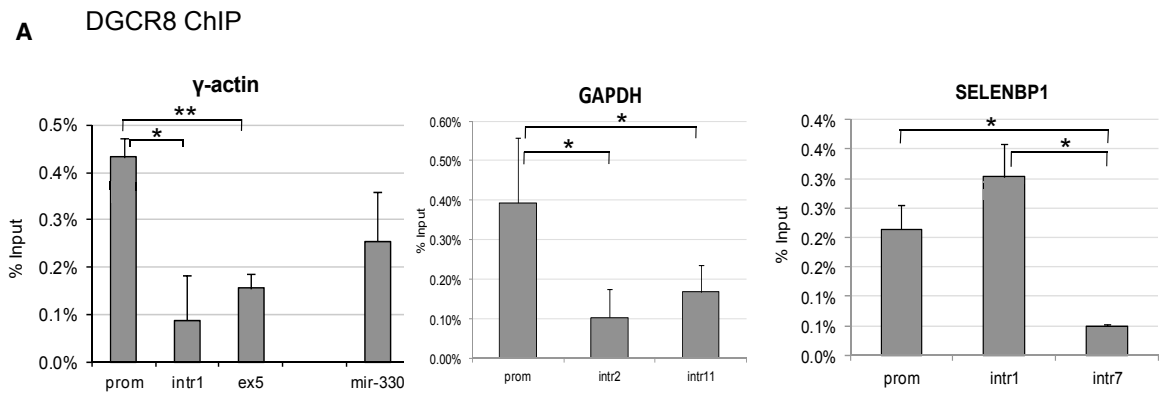
Figure 1. Drosha Binds 5′ Ends of Human Genes in a Transcription-Dependent Manner

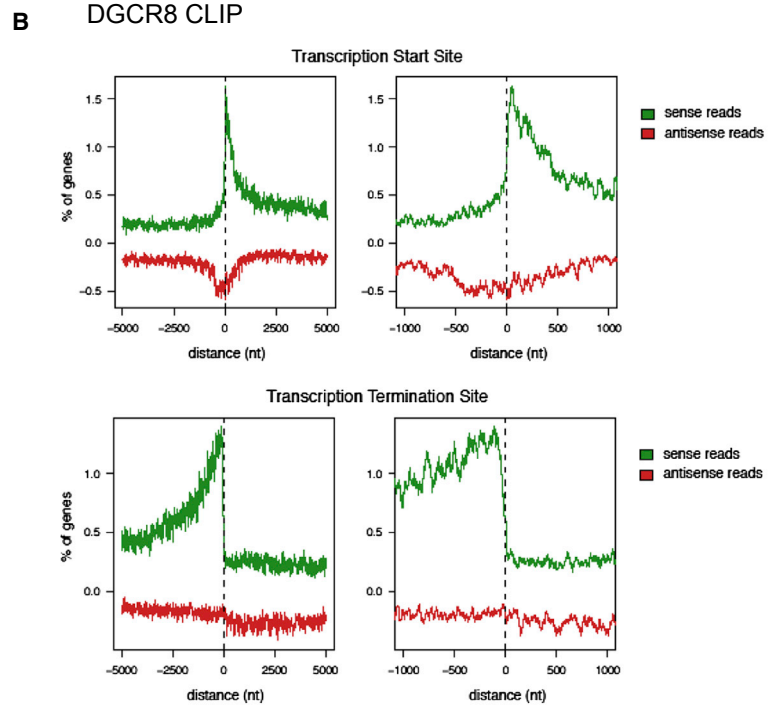
(A) (i) Metagene profiles of Drosha binding around transcriptional start sites (TSS, top) and transcriptional termination sites (TxEnd, bottom) based on Drosha ChIP-on-chip analysis in HeLa cells using ENCODE array. (ii) Average Drosha ChIP enrichment over input (log2) for probes overlapping miRNAs, intergenic regions, gene bodies, TSSs, and TxEnds. Error bars represent SEM. Number of probes covering each region is indicated above the bars. TSS signal is significantly enriched ($p < 0.001$, t test) compared to all intergenic gene bodies or TxEnd regions.

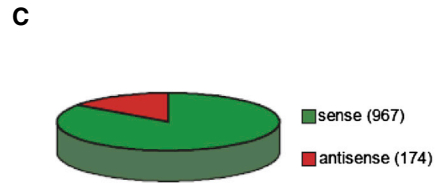
(B–D) Pol II and Drosha ChIP analyses on γ-actin (B) and GAPDH (C) genes and the miR-330 genomic locus, compared to β-actin ex 4 (D) in HeLa cells. ChIP amplicons are shown above the gene diagrams.

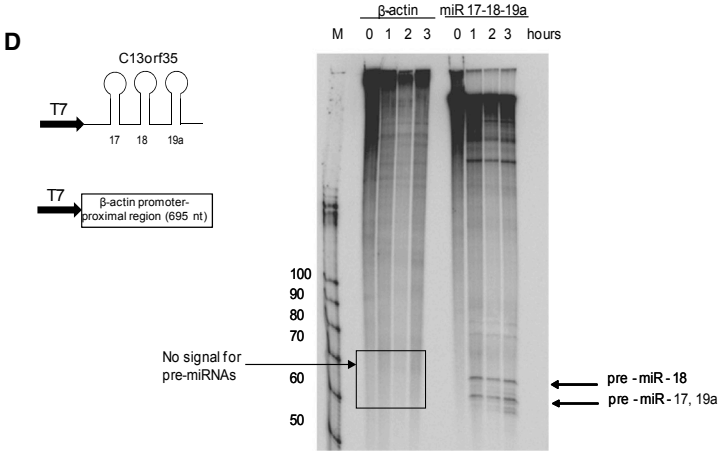
(E) (i) qRT-PCR analysis of γ-actin, β-actin and GAPDH pre-mRNAs from HeLa cells mock-treated or treated with 5 μg/ml actinomycin D for 6 hr, normalized to 5S rRNA. The amount of RNA in the untreated cells was taken as 1. Drosha (ii) and H3 (iii) ChIPs on γ-actin gene and miR-330 gene locus in HeLa cells, mock-treated or treated with 5 μg/ml of actinomycin D for 6 hr.

Bars in (B)–(E) represent the average values from at least three biological experiments ±SD. See Figures S1–S5.









(legend on next page)

human protein-coding genes in a transcription-dependent manner. Furthermore, we also searched for potential correlation between Drosha recruitment using our Drosha ChIP-array data and level of Pol II signal determined from available chromatin immunoprecipitation sequencing (ChIP-seq) experiments in UCSC depository. We have calculated Pol II signal in 1 kb region around the TSS for Refseq transcripts covered by probes in the Drosha array. Following this analysis, we identified a weak (Pearson R = 0.41) but significant (p = 1e-6) correlation between the Drosha probe signal at TSS and Pol II ChIP density (Figure S5), indicating that Drosha is preferentially recruited to the Pol II-rich regions. Overall, these results suggest that Drosha binding may be required for enhanced gene expression.

DGCR8 Binds Promoter-Proximal Regions of Human Genes

Drosha is known to bind miRNA sequences as part of the Microprocessor complex, which also contains DGCR8, a double-stranded RNA-binding protein that is deleted in the DiGeorge syndrome (Landthaler et al., 2004). To investigate whether Drosha binds 5′ regions of human genes together with DGCR8, we carried out ChIP experiments in HeLa cells (Figure 2A). DGCR8 was found to be enriched over the miR-330 locus, γ-actin, GAPDH, and SELENBP1 promoter-proximal gene regions (Figure 2A). These regions overlap Drosha binding regions (Figures 1, S2, and S4), suggesting that Drosha and DGCR8 may form a complex at promoter-proximal regions of human genes.

To substantiate these findings in a genome-wide fashion, we analyzed a DGCR8 HITS-CLIP experiment that was performed with endogenous and overexpressed DGCR8 protein in HEK293T cells to identify DGCR8 endogenous RNA targets (Macias et al., 2012). In particular, we observed 1,101 DGCR8 clusters, corresponding to 967 genes, in the sense orientation (see supplemental material in Macias et al., 2012) and 196 DGCR8 clusters, corresponding to 174 genes, in antisense orientation (Figure 2C; Table S2 for gene identity). Interestingly, significant clusters of DGCR8 binding were seen at −1,000/+200 bp from the TSSs and just upstream of the transcription termination regions (TTSs) (false discovery rate [FDR] < 0.01) (Figure 2B). In particular, we observed significant enrichment of DGCR8 binding in the region within 200 nt downstream of the TSS (sense versus antisense ks test p < 2.2e-16; Figure 2B). DGCR8 binding in this region is most pronounced in genes expressed at medium and high levels when compared to genes expressed at lower levels (ks test p < 2.2e-16 in both cases) (Figure S7A). These results clearly indicate that DGCR8, similar to Drosha, binds to promoter-proximal regions of human genes. However, it should be noted that when we analyzed DGCR8 binding to spliced mRNA transcripts (cDNAs) (Figure S6), we also see an enrichment over exonic sequences immediately downstream of the TSS. Again, this is seen in genes with medium and high expression (ks test p < 2.2e-16 in both cases) (Figure S7B). This may be related to the reported binding of DGCR8 to multiple exonic regions (Macias et al., 2012), potentially reflecting an additional function.

Promoter-Proximal Binding of Drosha Is Not Related to miRNA Synthesis or RNA Cleavage

Binding of Drosha to promoter-proximal regions of human genes may be related to its potential role in processing miRNAs from these genomic regions, even though such elements are not annotated in the human genome databases. To test if Drosha binds and cleaves miRNAs present in promoter-proximal regions, we performed in vitro pri-miRNA cleavage assays in HeLa nuclear extracts (Guil and Cáceres, 2007). Using in vitro transcription reactions, we generated radiolabeled RNAs from gene regions, identified by Drosha ChIP. We employed miR-17-92 cluster within the C13orf25 human genomic region as a positive control for miRNA production. Following in vitro cleavage assay in HeLa extracts, miRNAs were generated from the control miR-17-19a sequences but not from the promoter-proximal region of the β-actin gene (Figure 2D). We also failed to detect pre-miRNAs or RNA cleavage products in vitro for promoter-proximal regions of other genes, identified in Drosha ChIP assays (data not shown). These findings suggest that Drosha binding over promoter-proximal regions is not associated with miRNA synthesis and RNA cleavage.

Drosha Knockdown Causes Transcriptional Downregulation and a Decrease in Pre-mRNA and Poly(A)+ RNA

To investigate the promoter-associated function of Drosha, we performed RNAi-induced Drosha knockdown in HeLa cells (Figure 3A). Following Drosha depletion, we observed a small increase in Pol II binding over the promoters of γ-actin and GAPDH genes, using ChIP analysis (Figure 3A). We also observed an ∼20%–40% decrease in pre-mRNA levels and ∼30%–60% decrease in poly(A)+ RNA levels in HeLa cells depleted for Drosha (Figure 3B). In contrast, we did not detect any significant change in the levels of poly(A)+ and pre-mRNA corresponding to G6PD, ZNF687, and FUNDC2 genes, for which we observed lack of promoter-proximal Drosha binding (Figures S3B and S3C). This suggests the specificity of observed transcriptional effects, following Drosha depletion.

Figure 2. DGCR8 Binds Promoter-Proximal Regions of Human Genes

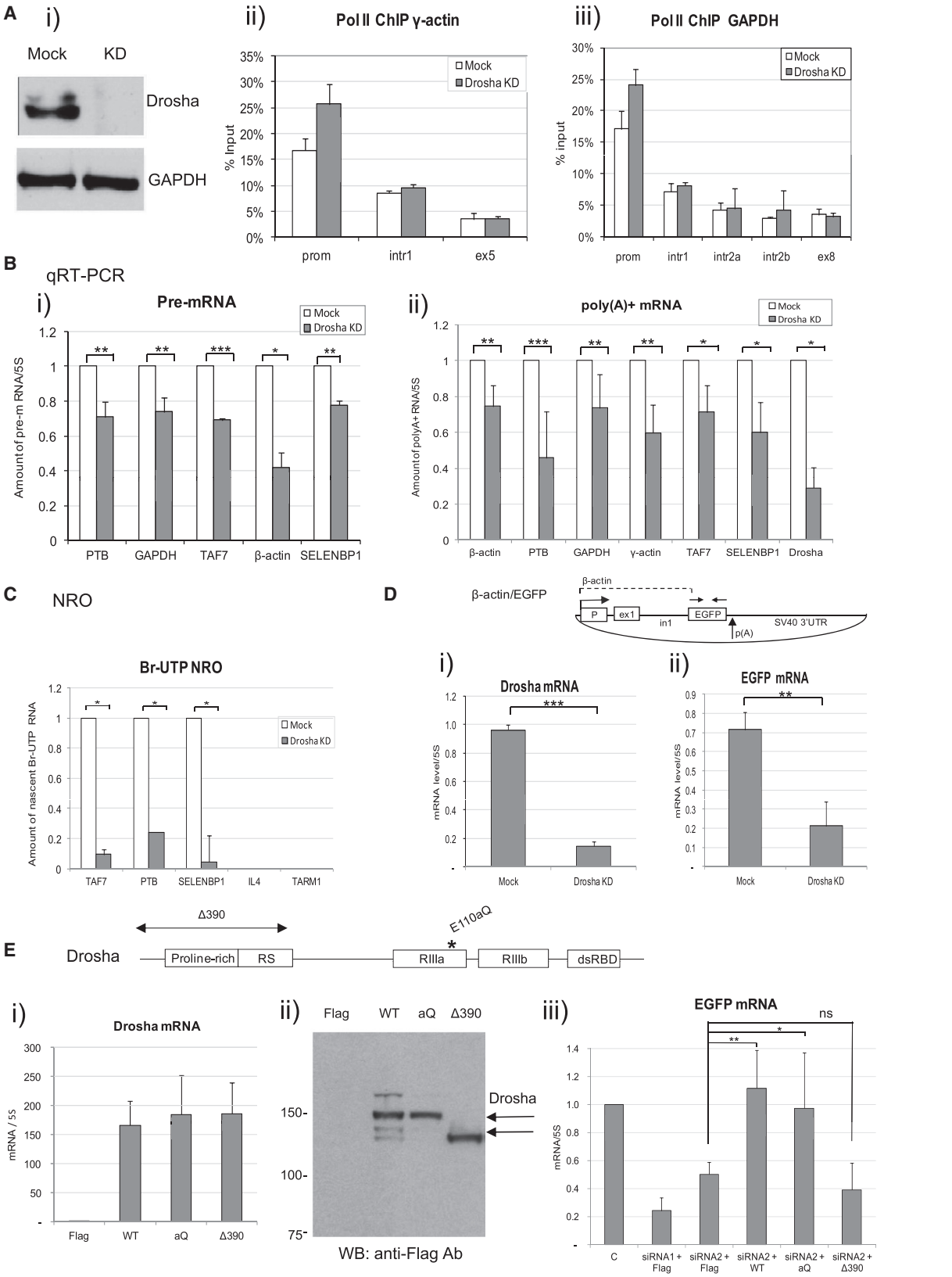
(A) DGCR8 ChIP analysis on γ-actin, GAPDH, and SELENBP1 genes and the miR-330 genomic locus in HeLa cells. SELENBP1 gene diagram is in Figure S2A. Bars represent the average values from at least three biological experiments ±SD.

(B) DGCR8 HITS-CLIP. Distribution of significant sense (green) and antisense (red) reads mapping around transcription start sites (TSSs; top panels) and transcription termination regions (TTSs; bottom panels) of protein coding genes, as determined by CEAS software, using FDR < 0.01. 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.

(C) Distribution of sense and antisense DRCG8 CLIP binding sites. 1,101 DGCR8 binding sites in the sense orientation correspond to 967 genes, and 196 DGCR8 antisense binding sites correspond to 174 genes.

(D) Drosha in vitro processing reactions in HeLa nuclear extracts with β-actin promoter-proximal region and miR17-18-19a sequences. Arrows indicate the positions of pre-miRNAs.

See also Figures S6 and S7.



(legend on next page)

To evaluate if Drosha has a direct effect on transcription, we employed nuclear run-on (NRO) using Br-UTP as the labeled nucleotide (Core and Lis, 2008; Lin et al., 2008). The advantage of Br-UTP NRO over Pol II ChIP is the ability of Br-UTP NRO to detect actively transcribing polymerases, as opposed to total Pol II levels detected by ChIP. As demonstrated in Figure 3C, we observed a substantial decrease in the level of nascent transcription for TAF7, PTB, and SELENBP1 genes in Drosha-knocked-down cells. As a negative control for this experiment, we used IL4 and TARM1 genes. These genes are not expressed in HeLa cells; hence, we did not observe any detectable transcription signal in Br-UTP NRO reactions. These results suggest that Drosha has a direct positive effect on transcription, consistent with the genome-wide expression data, where highly expressed genes correlated with increased Drosha binding on TSSs (Figure 1A). Furthermore, the antisense transcripts detected over the promoter-proximal region of the β-actin gene were also decreased upon Drosha depletion (Figure S4D). These results further indicate that Drosha is not involved in the cleavage of nascent sense or antisense RNA transcripts, pointing toward a cleavage-independent function of Drosha in human gene expression.

Drosha Regulates the Expression of a Heterologous Reporter

We next employed a β-actin/EGFP reporter plasmid, where transcription of EGFP is driven by the β-actin promoter, fused to β-actin exon 1 and intron 1 regions (Qin and Gunning, 1997) (Figure 3D). As described above, Drosha binds the promoter-proximal region of β-actin gene (Figure S4C). Interestingly, in HeLa cells depleted for Drosha we observed ~75% reduction in expression of EGFP mRNA from the plasmid (Figure 3D), recapitulating downregulation of endogenous β-actin mature and pre-mRNAs (Figure 3B). These results suggest that Drosha plays a positive role in the regulation of β-actin/EGFP expression.

Next, we cotransfected β-actin/EGFP reporter plasmid with RNAi-resistant Flag-tagged Drosha (WT), a catalytic mutant E110aQ (aQ), unable to process miRNAs, and Δ390 construct into 293 cells, depleted for Drosha, using siRNA2 (Han et al., 2004). The Δ390 Drosha construct is active in miRNA processing but lacks the N-terminal proline-rich and RS-rich domains, proposed to be important for Drosha protein-protein interactions (Han et al., 2004) (Figure 3E). Both mRNAs and proteins corresponding to WT, aQ, and Δ390 Drosha were expressed at a high level in 293T cells, depleted for endogenous Drosha (Figure 3Ei and 3Eii). Overexpression of both WT and aQ catalytic Drosha mutant resulted in the increase of EGFP mRNA (Figure 3Eiii). In contrast, overexpression of Δ390 construct had no significant effect on the expression of EGFP mRNA. These results suggest that Drosha plays a positive role in the regulation of the β-actin promoter. Importantly, this Drosha function is independent of its ability to cleave RNA but does require the N-terminal protein-protein interaction domain.

Drosha Regulates Human Gene Expression through Interaction with CBP80 and RNA Polymerase II

We hypothesize that the cleavage-independent stimulatory role of Drosha on gene expression may be mediated through its association with protein cofactors. We therefore performed coimmunoprecipitation experiments and found that transiently overexpressed Flag-tagged Drosha interacts with both unphosphorylated (IIA) and phosphorylated (IIO) forms of Pol II in 293T cells, supporting its role in transcriptional regulation. Drosha also interacts with CBP80 and Ars2 proteins, involved in miRNA biogenesis and RNA processing pathways, respectively (Gruber et al., 2009; Sabin et al., 2009) (Figure 4A). Furthermore, by ChIP analysis we found that CBP80 protein binds promoter-proximal regions of γ-actin and GAPDH genes, colocalizing with Drosha binding (Figures 4B and 4C). We next observed that Drosha-CBP80 interaction is not affected by the aQ mutation in the catalytic domain of Drosha but is significantly reduced when the N-terminal domain of Drosha is deleted in Δ390 construct or when both Δ390 and aQ mutations are combined (Figure 4D, top panel). These results strongly support the view that Drosha interacts with CBP80 through its N-terminal RS domain. We also studied the interaction of Drosha with Pol II (Figure 4D, middle). Similar to CBP80, a catalytic mutation in Drosha (aQ) does not affect its interaction with Pol II. However, the Drosha Δ390 construct showed reduced interaction with Pol II, and specifically with its phosphorylated form (IIO). A similar effect was observed with Δ390 aQ construct. This suggests that the N-terminal domain of Drosha is particularly important for Drosha-mediated transcriptional effects in human cells because it mediates interaction of Drosha with CBP80 and Pol II. We also

Figure 3. Drosha Depletion Causes Transcriptional Downregulation and Decreases Pre-mRNA and Poly(A)+ RNA Levels

(A) Western blot analysis of protein extracts (50 μg) from mock-treated and Drosha siRNA-treated HeLa cells, probed with Drosha and GAPDH antibodies (i). Pol II ChIP across γ-actin (ii) and GAPDH (iii) genes in mock-treated and Drosha-depleted HeLa cells.

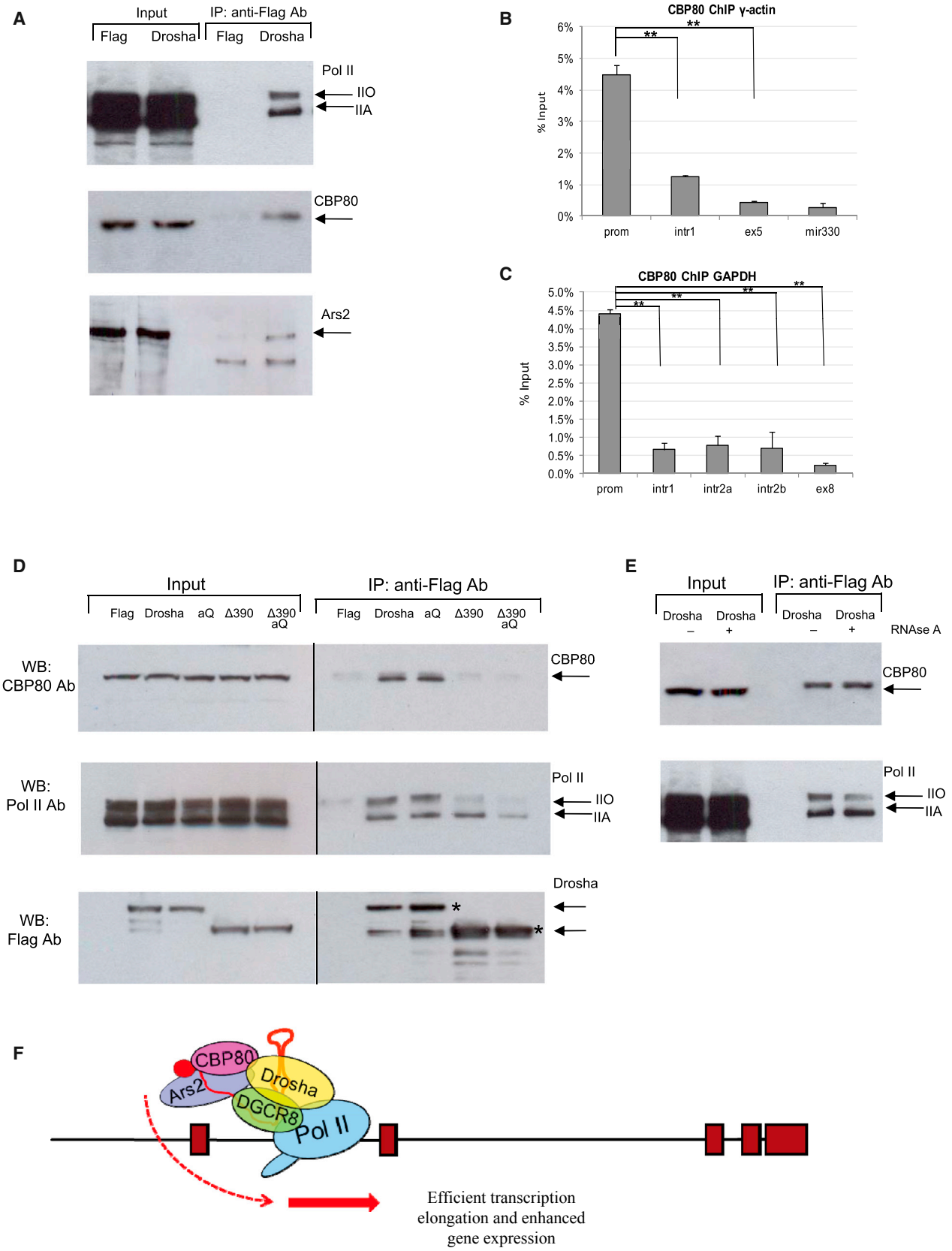
(B) qRT-PCR analysis of pre-mRNA (i) and poly(A)+ RNA (ii) in mock-treated and Drosha-depleted cells. The level of RNA in mock-treated samples was taken as 1.

(C) Br-UTP NRO analysis carried out in mock-treated and Drosha-depleted cells. Amount of nascent Br-UTP RNA was calculated by subtracting the background of U-RNA produced over a specific gene probe. The level of Br-UTP RNA in mock-treated samples was taken as 1.

(D) Diagram of β-actin/EGFP plasmid. Promoter, exon 1, and intron 1 sequences are derived from the human β-actin gene, indicated by dashed line. qRT-PCR primers are shown above the diagram. SV40 3′ UTR contains a poly(A) signal, indicated with an arrow. qRT-PCR analysis of Drosha (i) and EGFP (ii) mRNAs in mock-treated and Drosha-depleted HeLa cells.

(E) Top, domain structure of Drosha protein. Double-stranded RNA-binding (dsRBD), RNaseIII (RIIIa, RIIIb), proline-rich, and serine-arginine (RS) domains are shown. Positions of catalytic E110aQ (aQ) mutation and Δ390 deletion of the N-terminal proline-rich and RS-rich domains are indicated above the diagram. Bottom, qRT-PCR analysis of Drosha mRNA (i) and EGFP mRNA (iii) in Drosha-depleted 293T cells, overexpressed with RNAi-resistant Flag, Drosha WT, aQ, and Δ390 constructs. The level of Drosha (i) or EGFP (iii) mRNA in cells, overexpressed with Flag was taken as 1. siRNA2, targeting 3′ UTR of endogenous Drosha mRNA, was used for Drosha depletion in overexpression experiments. (ii) Western blot of total protein extracts (50 μg) from 293T cells, overexpressed with Flag, Drosha WT, aQ, and Δ390 plasmids, probed with Flag antibody. The positions of the marker are indicated on the left.

Bars in (A)–(E) represent average values from at least three independent experiments ±SD. In (B), (D), and (E), the levels of RNA were normalized to 5S rRNA.



(legend on next page)

observed that interaction of Drosha with CBP80 and Pol II is RNA independent (Figure 4E). Consequently, this result provides a molecular explanation for the ability of Drosha to regulate gene expression in a miRNA- and cleavage-independent manner.

DISCUSSION

We demonstrate the existence of a miRNA- and cleavage-independent function of Drosha in the regulation of human gene expression. Combining whole-genome analysis with gene-specific approaches, we demonstrate that Drosha binds to the 5′ ends of human genes in a transcription-dependent manner (Figure 1). This binding is not associated with miRNA biogenesis or RNA cleavage but correlates with the level of gene expression (Figures 1 and 2). Depletion of Drosha from HeLa cells led to direct downregulation of transcription based on nuclear run-on and Pol II accumulation at the promoters of these genes, resulting in reduced levels of pre-mRNA and poly(A)$^+$ mRNA (Figure 3). This suggests a positive effect of Drosha on gene transcription. ChIP and HITS-CLIP analysis of DGCR8 also revealed binding sites at promoter-proximal regions, enriched at genes expressed at a higher level, suggesting that Drosha and DGCR8 may bind to promoter regions as a complex (Figures 2 and S7). A positive function of Drosha in the regulation of gene expression was further confirmed using a heterologous reporter construct, where Drosha overexpression influences reporter expression (Figure 3). This function of Drosha is independent of its catalytic activity and mediated through its N-terminal domain, which interacts with the RNA-binding protein CBP80 and Pol II, which also bind to the 5′ end of human genes (Figures 3 and 4). Taken together, these results suggest a model whereby Drosha promotes gene expression in a cleavage-independent manner by binding to RNA hairpins formed at the 5′ ends of the nascent RNAs and interacting with CBP80 and Pol II, through its N-terminal RS domain (Figure 4F).

Initially, we hypothesized that this positive function of Drosha may be associated with generation/regulation of short promoter-associated transcripts (PROMPTs), originally predicted to regulate transcription through changes in chromatin structure, promoter methylation, or Pol II recycling (Preker et al., 2008). PROMPTs may direct Drosha to the beginning of the gene through interaction with its binding factors. In our studies, both sense and antisense transcripts, detected over the promoter-proximal regions, were destabilized in Drosha-depleted cells (Figure S4D), suggesting that Drosha does not cleave these transcripts and its function is not mediated through PROMPTs.

Drosha binding at the beginning of human genes may be also related to Pol II release from promoter-proximal pausing events and be required for the recruitment of transcription/RNA processing factors. Interestingly, Drosha binding coincides with the binding site of the negative elongation factor NELF, which defines the "late" elongation checkpoint, ensuring conversion to productive elongation mediated by phosphorylation of Pol II CTD, NELF, and DSIF by the positive transcription elongation factor b (P-TEFb) (Egloff et al., 2009). As both nucleosome positioning and chromatin modifications play an important role in the elongation functions of P-TEFb, we also checked whether Drosha knockdown causes a change in chromatin modifications. However, we did not observe any significant changes (data not shown). This suggests that Drosha regulates gene expression through a different mechanism. GC-rich sequences are known to be enriched in the promoter regions of human genes (Calistri et al., 2011). Transcripts for these sequences are likely to form RNA hairpin structures, predicted to be suitable for Microprocessor binding mediated by DGCR8 (Macias et al., 2012; Table S1). If Drosha were to cleave such RNAs, the consequences of Drosha cleavage across the whole human genome would be detrimental to cellular survival. We hypothesize that Drosha does not cleave promoter-associated RNA but is instead utilized as a positive regulator of gene expression. We further analyzed the structures found in promoter transcripts and compared them to miRNA, snoRNAs, and protein-coding region structures, found to be bound by DGCR8 in CLIP analysis. Our bioinformatic analysis confirmed that structures in promoter regions have shorter stems, higher minimum free energy, and lower base-pairing probability (Figure S8). In addition, promoter transcripts also lack motifs important for the processing of the stem, such as TG dinucleotide at the base of the stem and CNNC motif downstream of the stem, as described in Auyeung et al. (2013). Taken together, this analysis suggests that promoter structures may be less stable, and their recognition and processing by the Microprocessor complex may be less efficient compared to miRNAs or other sequences cleaved by Drosha.

Evidence that components of the miRNA machinery may have additional functions comes from several previous studies. In early-stage thymocytes, Drosha recognizes and cleaves mRNAs harboring secondary stem-loop structures (Chong et al., 2010). Drosha can also regulate viral gene expression of Kaposi's

Figure 4. Drosha Regulates Human Gene Expression through Its Interaction with CBP80 and Pol II

(A) Pull-down experiments performed with anti-Flag in 293 cells, overexpressed with Flag and Drosha WT-Flag constructs. Western blots were probed with Pol II, Ars2, and CBP80 antibodies. Input represents 1% of material added to IP.

(B and C) CBP80 ChIP analyses across γ-actin (B) and GAPDH (C) genes and miR-330 gene locus in HeLa cells. Bars represent average values from at least three independent experiments ±SD.

(D) Pull-down experiments performed with anti-Flag in 293 cells, overexpressed with Flag, Drosha WT, aQ, Δ390, and Δ390 aQ constructs. Input represents 1% of material added to IP. Western blot was probed with CBP80 (top panel) and Pol II (middle panel) and FLAG (bottom panel) antibodies. Asterisks indicate full-length proteins.

(E) Pull-down experiments performed with anti-Flag in 293 cells, overexpressed with Drosha WT construct. Cell extracts were nontreated (−) or treated (+) with 50 μg/ml of RNase A. Input represents 1% of material added to IP. Western blot was probed with CBP80 (top panel) and Pol II (bottom panel) antibodies.

(F) Model for the function of Drosha in the regulation of human gene expression. Following transcription of nascent RNA by Pol II, Drosha, most probably in a complex with DGCR8, binds GC-rich hairpins at the promoter-proximal regions of human genes. Drosha does not cleave such hairpins, but it interacts with Pol II and CBP80 through its N-terminal proline-rich domain. Drosha can enhance gene expression, independently of its RNA cleavage function. Depletion of Drosha results in decrease of nascent transcription and subsequent decrease in the amount of nascent and poly(A)+ mRNA.

Sarcoma-associated Herpesvirus (KSHV) (Lin and Sullivan, 2011). DGCR8 together with Drosha controls the abundance of many cellular RNAs, including noncoding RNAs, mRNAs, and alternatively spliced isoforms (Macias et al., 2012). Finally, Microprocessor can regulate transcription of some cellular mRNA and HIV provirus, causing Pol II premature termination through the cleavage of the TAR stem-loop structure (Wagschal et al., 2012). These regulatory activities of Drosha each involve recognition and cleavage of target RNAs in an analogous manner to its function in miRNA biogenesis in contrast to the positive function of Drosha on transcription described here. Possibly this positive function of Drosha is related to the action of RNA-binding factors, known to modulate the function of Drosha in human cells. Such factors are likely to interact with the N-terminal proline-rich and RS domains, proposed to be the main sites of Drosha protein-protein interactions. We confirmed the previously described interaction between Drosha and CBP80 and Ars2 proteins (Gruber et al., 2009; Sabin et al., 2009). Furthermore, our results also show that Drosha interacts with the C-terminal domain (CTD) of Pol II (Figure 4). This interaction may be direct or mediated via proteins, containing RS domains and interacting with the CTD, potentially modulating Drosha's cleavage-independent function. Interestingly, in *Drosophila* the key RNAi components Dicer 2 and Argonaute 2 associate with chromatin and interact with the core transcription machinery, affecting Pol II dynamics (Cernilogar et al., 2011), further supporting the role of RNAi machinery in transcriptional regulation. Because both Drosha and Dicer may have additional roles to miRNA biogenesis future studies will be necessary to evaluate the full spectrum of molecular functions that these RNase-III-like enzymes play in human cells. This will represent an exciting avenue for future research in gene expression.

EXPERIMENTAL PROCEDURES

Plasmids

pCK-Flag WT, aQ, and Δ390 Drosha constructs (Han et al., 2004) and β-actin/EGFP plasmid (Qin and Gunning, 1997) were described previously. pflag-CMV-2 (Flag) was purchased from Sigma-Aldrich (E7398).

RNA Analysis

Total RNA was harvested using TRIZOL reagent (Invitrogen) followed by DNase I treatment (Roche). Total RNA (2 μg) was reverse-transcribed using SuperScript Reverse Transcriptase (Invitrogen) and random hexamer primers (Invitrogen) or gene-specific primers, as described in the Supplemental Experimental Procedures and Table S3, followed by qPCR. Gene-specific primers were used in all figures, apart from Figure 3Eiii, where random hexamers were used for quantitative RT-PCR (qRT-PCR), following the manufacturer's description (Invitrogen).

In Vitro Processing Assays

MiR-17-19 and γ-actin gene regions were amplified from genomic DNA using T7+β-actin (F/R) and T7+miR-17-19 (F/R) primers accordingly. In vitro processing was performed as in Guil and Cáceres (2007).

RNAi and Protein Analysis

The RNAi was carried out as described (Wollerton et al., 2004). mRNA target sequence for Drosha small interfering RNA (siRNA) duplex was 5′-CGA GUAGGCUUCGUGACUU-3′ (siRNA1) and 5′-GAGUAUUUACUUGCUCAG UAA-3′ (siRNA2). siRNA1 was used in all figures apart from Figure 3E, where siRNA2 was used. Pull-downs were carried out in RNase A nontreated extracts, except Figure 4E where 50 μg/ml RNase A was used to treat cell extracts during the pull-down procedure as described in Gruber et al. (2009). The immunoprecipitated proteins were detected by western blotting. Western blots were probed with Drosha (AbCam), actin (Sigma-Aldrich), GAPDH (Sigma-Aldrich), Pol II (Covance), and CBP80 (Sigma) antibodies.

ChIP and HITS-CLIP Analyses

ChIP analysis was carried out as previously described (West et al., 2004). Five micrograms of each antibody was used per ChIP. The immunoprecipitated DNAs were used as templates for qPCR. Drosha Chip-on-chip experiments were carried out as described in De Gobbi et al. (2007) and in the Supplemental Information. HITS-CLIP for DGCR8 was based on a published protocol (Wang et al., 2009) with minor modifications, as described in Macias et al. (2012) and the Supplemental Information.

Br-UTP Nuclear Run-on Analysis

The Br-UTP NRO was carried out largely as described (Lin et al., 2008; Skourti-Stathaki et al., 2011) with some modifications. Nuclear pellets were resuspended in transcription buffer (40 mM Tris-HCl [pH 7.9], 300 mM KCl, 10 mM $MgCl_2$, 40% glycerol, 2 mM DTT) and 10 mM mix of rATP, rCTP, rGTP, and Br-UTP or rUTP (in the control samples). The NRO reaction was performed at 30°C for 30 min. Total RNA was isolated using TRIzol reagent (Invitrogen) according to manufacturer's instructions and treated with RNase-free DNase I (Roche). Two microliters of anti-BrU antibody (Sigma-Aldrich) was preincubated with 30 μl of Protein G Dynabeads (Upstate) and 10 μg tRNA per sample for 1 hr at 4°C. The beads were washed three times with RSB-100 buffer (10 mM Tris-HCl [pH 7.4], 100 mM NaCl, 2.5 mM $MgCl_2$, 0.4% Triton X-100) and resuspended in 150 μl RSB-100 with 40 U RNase-OUT (Invitrogen) and 5 μg of glycogen. Total RNA was added to the beads and incubated for additional 1 hr at 4°C. The beads were washed three times with RSB-100 buffer. RNA bound to the beads was extracted with TRIzol reagent followed by DNase I treatment. The RT reaction was performed using SuperScript III Reverse Transcriptase (Invitrogen) following the manufacturers' instructions. The real-time quantitative PCR was performed using a Corbett Research Rotor-Gene GG-3000 machine. The PCR mixture contained QuantiTest SYBR green PCR master mix (Qiagen), 2 μl of template cDNA, and primers from Table S3. 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. Amount of nascent Br-UTP RNA was calculated by subtracting the background of U-RNA produced over a specific gene probe.

Statistical Analysis

Unless otherwise stated, results are shown as the average values from at least three independent biological experiments ±SD. The asterisk indicates statistical significance (*$p < 0.05$; **$p < 0.05$; ***$p < 0.05$), on the basis of an unpaired, two-tailed distribution determined with a Student's t test.

ACCESSION NUMBERS

The NCBI Gene Expression Omnibus accession number for the sequencing raw data for endogenous and overexpressed DGCR8 HITS-CLIP reported in this paper is GSE39086.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, eight figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2013.11.032.

AUTHOR CONTRIBUTIONS

N.G. and N.J.P. conceived the study. N.G. performed all experiments and analyzed the data. M.D. performed bioinformatic analysis of Drosha ChIP-on-chip data. S.M., E.E., and M.P. analyzed DGCR8 binding at promoter

regions by HITS-CLIP and carried analysis in Figures S6–S8. N.G., J.F.C., and N.J.P. wrote the paper.

ACKNOWLEDGMENTS

We thank N. Kim for providing pCK-Flag WT, E110aQ, and ΔN390 Drosha constructs, M. de Gobbi and D. Higgs for the help with ChIP-on-chip experiments, and Natalie Braun for the help with Figure S3C. The work was supported by a Wellcome Trust Programme Grant to N.J.P. and Royal Society University Research fellowship and MRC NIRG MR/J007870/1 to N.G. J.F.C. and S.M. were supported by the Medical Research Council and Wellcome Trust (grant 095518/Z/11/Z). E.E. was supported by grants from the Spanish Ministry of Science and by the Sandra Ibarra Foundation (BIO2008-01091, BIO2011-23920, and CSD2009-00080). M.P. is supported by the Novo Nordisk Foundation.

Received: May 21, 2013
Revised: September 13, 2013
Accepted: November 15, 2013
Published: December 19, 2013

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.

Ballarino, M., Pagano, F., Girardi, E., Morlando, M., Cacchiarelli, D., Marchioni, M., Proudfoot, N.J., and Bozzoni, I. (2009). Coupled RNA processing and transcription of intergenic primary microRNAs. Mol. Cell. Biol. *29*, 5632–5638.

Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. Cell *136*, 215–233.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363–366.

Calistri, E., Livi, R., and Buiatti, M. (2011). Evolutionary trends of GC/AT distribution patterns in promoters. Mol. Phylogenet. Evol. *60*, 228–235.

Cernilogar, F.M., Onorati, M.C., Kothe, G.O., Burroughs, A.M., Parsi, K.M., Breiling, A., Lo Sardo, F., Saxena, A., Miyoshi, K., Siomi, H., et al. (2011). Chromatin-associated RNA interference components contribute to transcriptional regulation in Drosophila. Nature *480*, 391–395.

Chong, M.M., Zhang, G., Cheloufi, S., Neubert, T.A., Hannon, G.J., and Littman, D.R. (2010). Canonical and alternate functions of the microRNA biogenesis machinery. Genes Dev. *24*, 1951–1960.

Core, L.J., and Lis, J.T. (2008). Transcription regulation through promoter-proximal pausing of RNA polymerase II. Science *319*, 1791–1792.

De Gobbi, M., Anguita, E., Hughes, J., Sloane-Stanley, J.A., Sharpe, J.A., Koch, C.M., Dunham, I., Gibbons, R.J., Wood, W.G., and Higgs, D.R. (2007). Tissue-specific histone modification and transcription factor binding in alpha globin gene expression. Blood *110*, 4503–4510.

Egloff, S., Al-Rawaf, H., O'Reilly, D., and Murphy, S. (2009). Chromatin structure is implicated in "late" elongation checkpoints on the U2 snRNA and beta-actin genes. Mol. Cell. Biol. *29*, 4002–4013.

Fujii, M., Lyakh, L.A., Bracken, C.P., Fukuoka, J., Hayakawa, M., Tsukiyama, T., Soll, S.J., Harris, M., Rocha, S., Roche, K.C., et al. (2006). SNIP1 is a candidate modifier of the transcriptional activity of c-Myc on E box-dependent target genes. Mol. Cell *24*, 771–783.

Fuller-Pace, F.V., and Ali, S. (2008). The DEAD box RNA helicases p68 (Ddx5) and p72 (Ddx17): novel transcriptional co-regulators. Biochem. Soc. Trans. *36*, 609–612.

Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. Nature *432*, 235–240.

Gruber, J.J., Zatechka, D.S., Sabin, L.R., Yong, J., Lum, J.J., Kong, M., Zong, W.X., Zhang, Z., Lau, C.K., Rawlings, J., et al. (2009). Ars2 links the nuclear cap-binding complex to RNA interference and cell proliferation. Cell *138*, 328–339.

Guil, S., and Cáceres, J.F. (2007). The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. Nat. Struct. Mol. Biol. *14*, 591–596.

Han, J., Lee, Y., Yeom, K.H., Kim, Y.K., Jin, H., and Kim, V.N. (2004). The Drosha-DGCR8 complex in primary microRNA processing. Genes Dev. *18*, 3016–3027.

Han, J., Pedersen, J.S., Kwon, S.C., Belair, C.D., Kim, Y.K., Yeom, K.H., Yang, W.Y., Haussler, D., Blelloch, R., and Kim, V.N. (2009). Posttranscriptional crossregulation between Drosha and DGCR8. Cell *136*, 75–84.

Kadener, S., Rodriguez, J., Abruzzi, K.C., Khodor, Y.L., Sugino, K., Marr, M.T., 2nd, Nelson, S., and Rosbash, M. (2009). Genome-wide identification of targets of the drosha-pasha/DGCR8 complex. RNA *15*, 537–545.

Karginov, F.V., Cheloufi, S., Chong, M.M., Stark, A., Smith, A.D., and Hannon, G.J. (2010). Diverse endonucleolytic cleavage sites in the mammalian transcriptome depend upon microRNAs, Drosha, and additional nucleases. Mol. Cell *38*, 781–788.

Kim, Y.K., and Kim, V.N. (2007). Processing of intronic microRNAs. EMBO J. *26*, 775–783.

Kim, S., Yang, J.Y., Xu, J., Jang, I.C., Prigge, M.J., and Chua, N.H. (2008). Two cap-binding proteins CBP20 and CBP80 are involved in processing primary MicroRNAs. Plant Cell Physiol. *49*, 1634–1644.

Landthaler, M., Yalcin, A., and Tuschl, T. (2004). The human DiGeorge syndrome critical region gene 8 and Its D. melanogaster homolog are required for miRNA biogenesis. Curr. Biol. *14*, 2162–2167.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O., Kim, S., and Kim, V.N. (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 415–419.

Lin, Y.T., and Sullivan, C.S. (2011). Expanding the role of Drosha to the regulation of viral gene expression. Proc. Natl. Acad. Sci. USA *108*, 11229–11234.

Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S., and Fu, X.D. (2008). The splicing factor SC35 has an active role in transcriptional elongation. Nat. Struct. Mol. Biol. *15*, 819–826.

Lund, E., Güttinger, S., Calado, A., Dahlberg, J.E., and Kutay, U. (2004). Nuclear export of microRNA precursors. Science *303*, 95–98.

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

Morlando, M., Ballarino, M., Gromak, N., Pagano, F., Bozzoni, I., and Proudfoot, N.J. (2008). Primary microRNA transcripts are processed co-transcriptionally. Nat. Struct. Mol. Biol. *15*, 902–909.

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. Science *322*, 1851–1854.

Qin, H., and Gunning, P. (1997). The 3′-end of the human beta-actin gene enhances activity of the beta-actin expression vector system: construction of improved vectors. J. Biochem. Biophys. Methods *36*, 63–72.

Sabin, L.R., Zhou, R., Gruber, J.J., Lukinova, N., Bambina, S., Berman, A., Lau, C.K., Thompson, C.B., and Cherry, S. (2009). Ars2 regulates both miRNA- and siRNA- dependent silencing and suppresses RNA virus infection in Drosophila. Cell *138*, 340–351.

Siomi, H., and Siomi, M.C. (2010). Posttranscriptional regulation of microRNA biogenesis in animals. Mol. Cell *38*, 323–332.

Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Mol. Cell *42*, 794–805.

Triboulet, R., Chang, H.M., Lapierre, R.J., and Gregory, R.I. (2009). Post-transcriptional control of DGCR8 expression by the Microprocessor. RNA *15*, 1005–1011.

Wagschal, A., Rousset, E., Basavarajaiah, P., Contreras, X., Harwig, A., Laurent-Chabalier, S., Nakamura, M., Chen, X., Zhang, K., Meziane, O., et al. (2012). Microprocessor, Setx, Xrn2, and Rrp6 co-operate to induce premature termination of transcription by RNAPII. Cell *150*, 1147–1157.

Wang, X., Arai, S., Song, X., Reichart, D., Du, K., Pascual, G., Tempst, P., Rosenfeld, M.G., Glass, C.K., and Kurokawa, R. (2008). Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature *454*, 126–130.

Wang, Z., Tollervey, J., Briese, M., Turner, D., and Ule, J. (2009). CLIP: construction of cDNA libraries for high-throughput sequencing from RNAs cross-linked to proteins in vivo. Methods *48*, 287–293.

West, S., Gromak, N., and Proudfoot, N.J. (2004). Human 5′ —> 3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature *432*, 522–525.

Wollerton, M.C., Gooding, C., Wagner, E.J., Garcia-Blanco, M.A., and Smith, C.W. (2004). Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. Mol. Cell *13*, 91–100.

Yi, R., Qin, Y., Macara, I.G., and Cullen, B.R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev. *17*, 3011–3016.

Yu, B., Bi, L., Zheng, B., Ji, L., Chevalier, D., Agarwal, M., Ramachandran, V., Li, W., Lagrange, T., Walker, J.C., and Chen, X. (2008). The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. Proc. Natl. Acad. Sci. USA *105*, 10073–10078.

Zeng, Y., Yi, R., and Cullen, B.R. (2005). Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. EMBO J. *24*, 138–148.

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*]

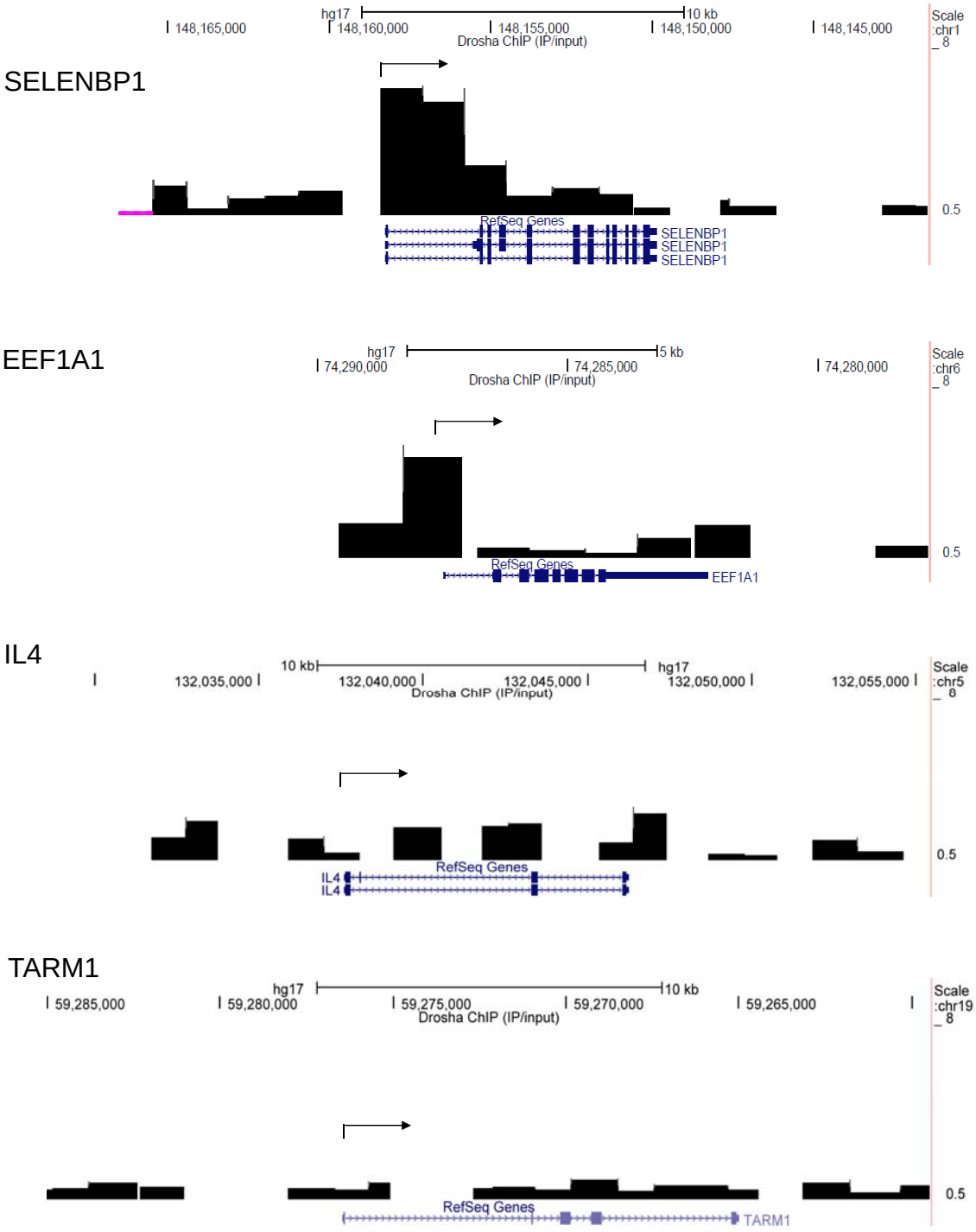


Figure S1, related to Figure 1A.

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.

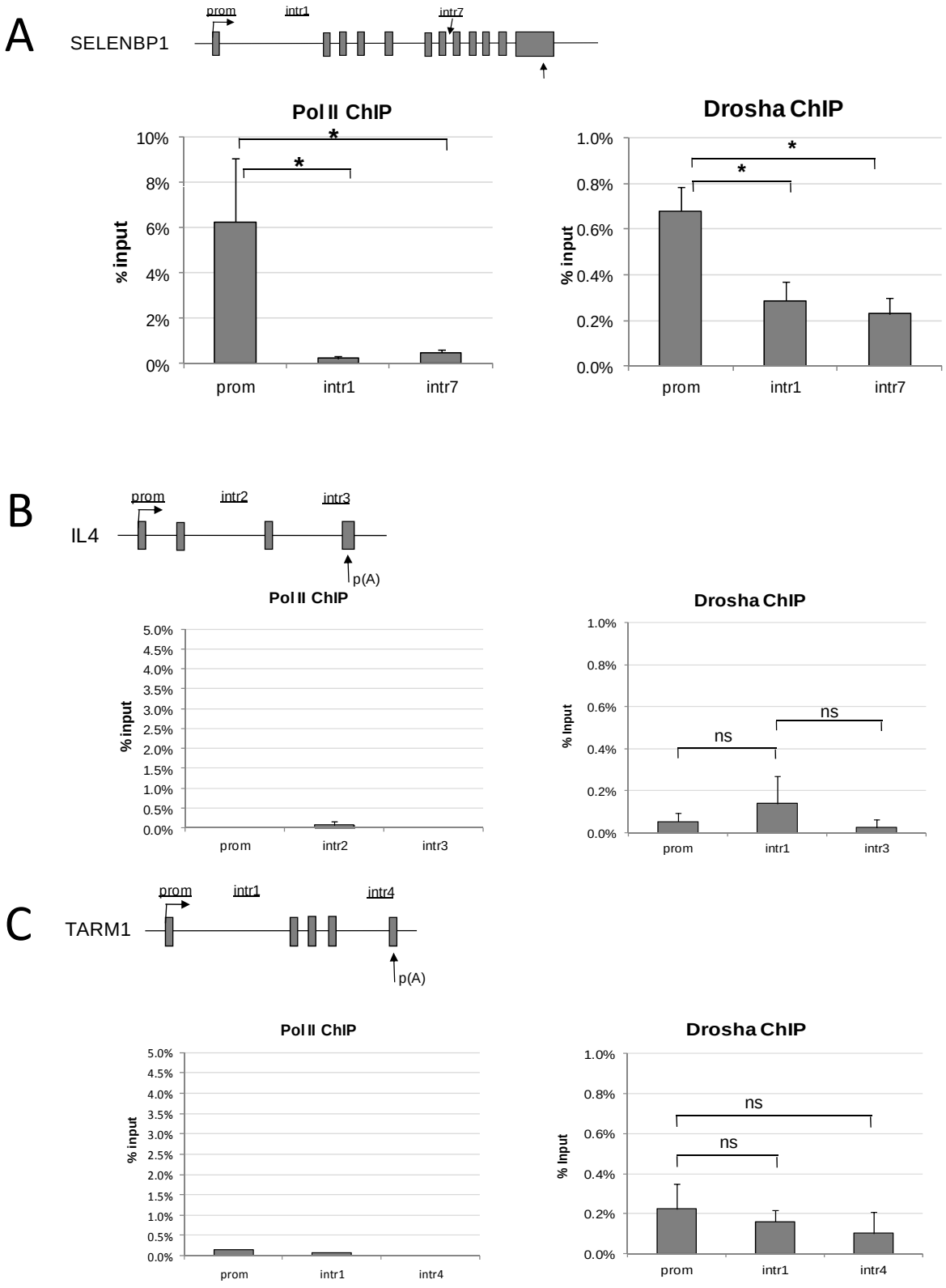


Figure S2, related to Figure 1A.
Drosha and Pol II profiles of SELEBP1, IL4 and TARM1 genes

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.

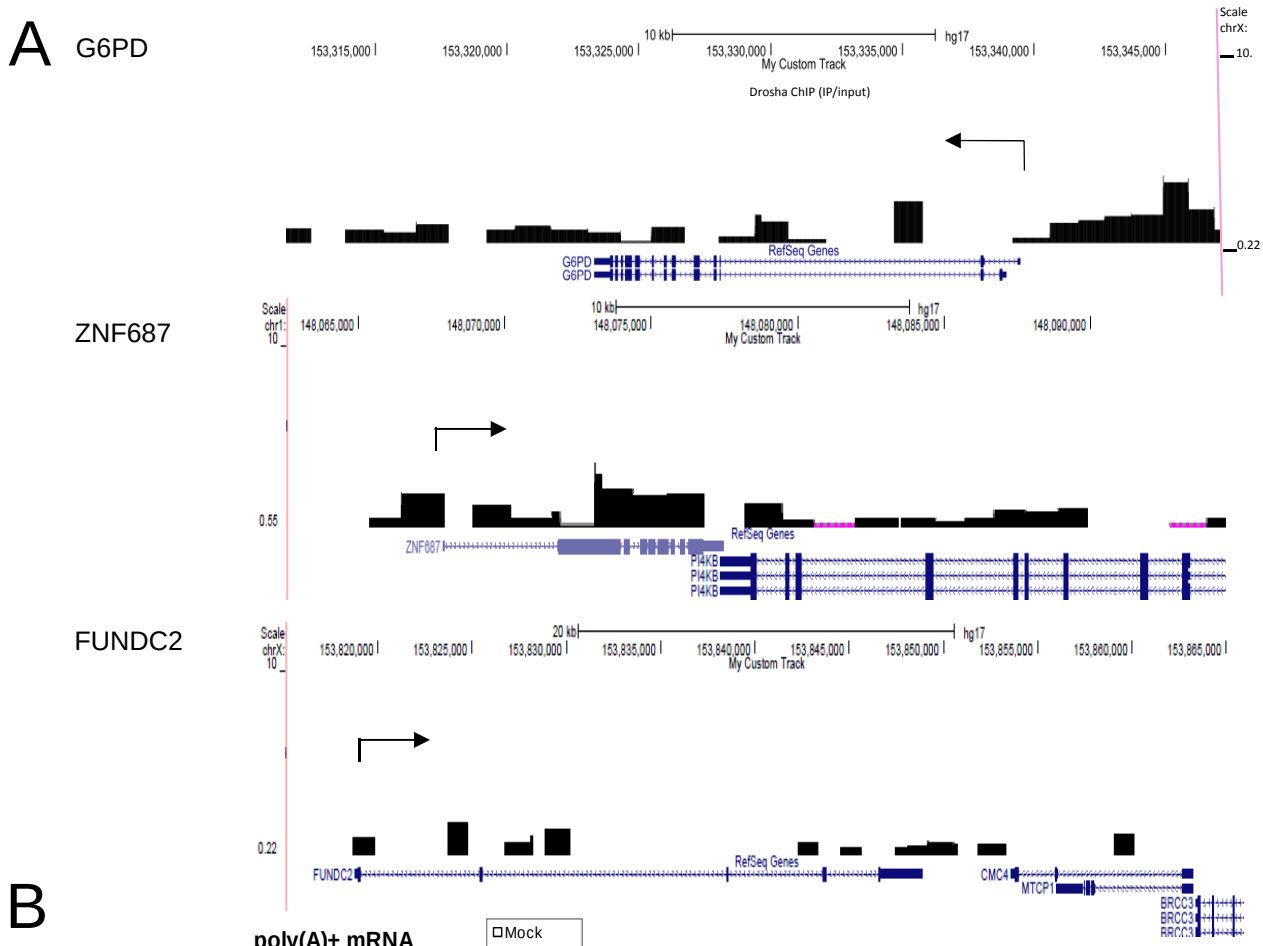

A
G6PD
ZNF687
FUNDC2
My Custom Track
Drosha ChIP (IP/input)
RefSeq Genes
B
poly(A)+ mRNA
Mock
Drosha KD
Amount of polyA+ RNA / 5S
ns
FUNDC2
ZNF687
G6PD
C
G6PD
prom
intr2
intr12
p(A)
Pol II ChIP
% Input
prom
intr2
in12
Drosha ChIP
G6PD pre-mRNA
Amount of pre-mRNA/5S
Mock
Drosha KD

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.

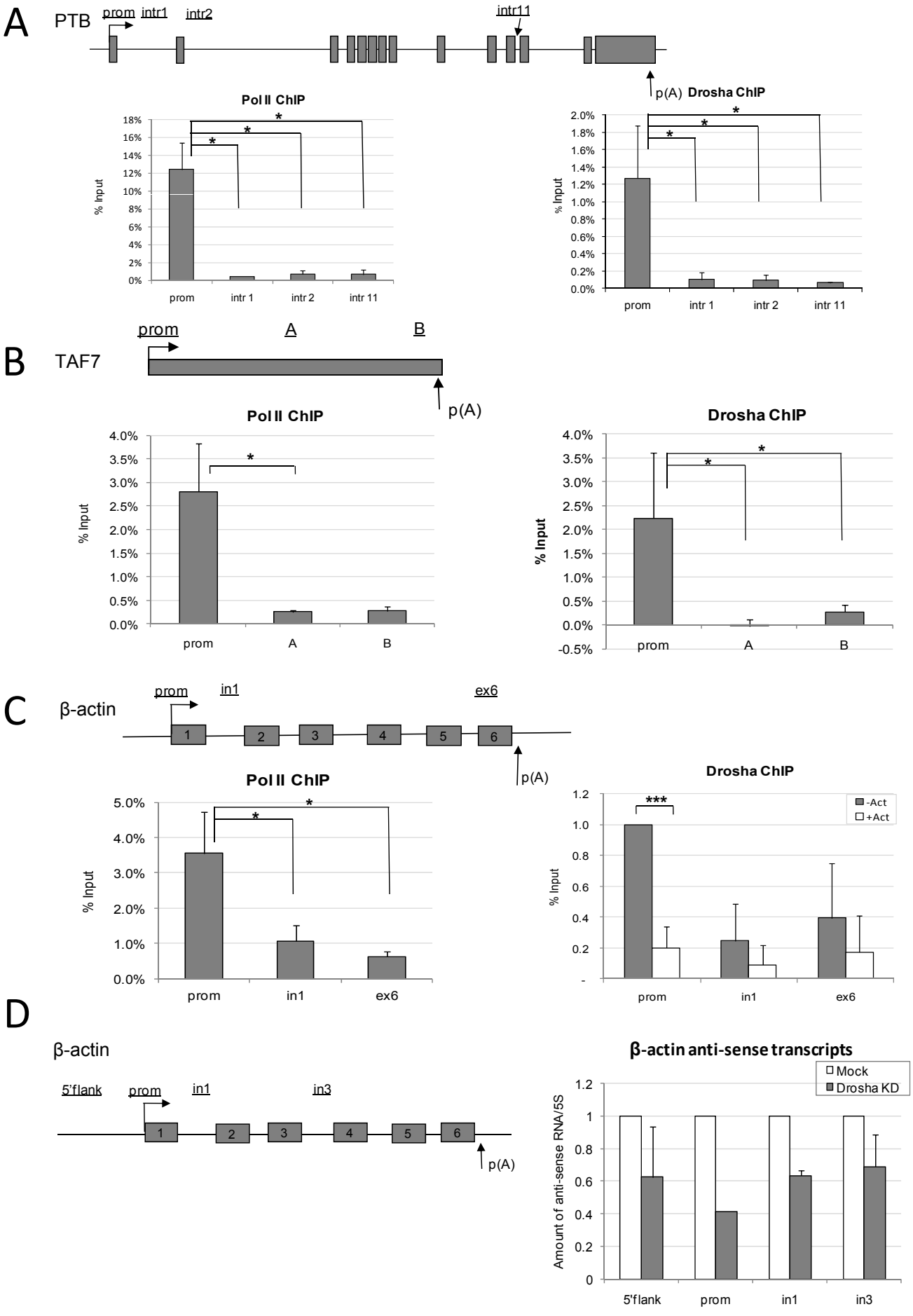

A
PTB
prom
intr1
intr2
intr11
p(A)
Pol II ChIP
Drosha ChIP
% Input
prom
intr 1
intr 2
intr 11
B
TAF7
prom
A
B
p(A)
Pol II ChIP
Drosha ChIP
C
β-actin
prom
in1
ex6
p(A)
Pol II ChIP
Drosha ChIP
-Act
+Act
D
β-actin
5'flank
prom
in1
in3
p(A)
β-actin anti-sense transcripts
Mock
Drosha KD
Amount of anti-sense RNA/5S

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.

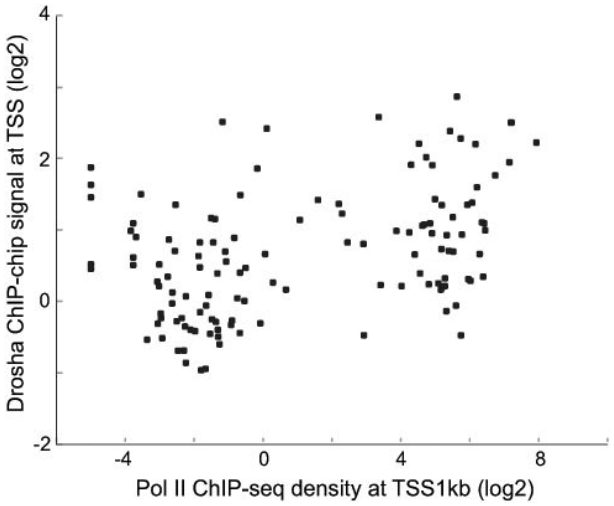


Figure S5, related to Figure 1A.

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

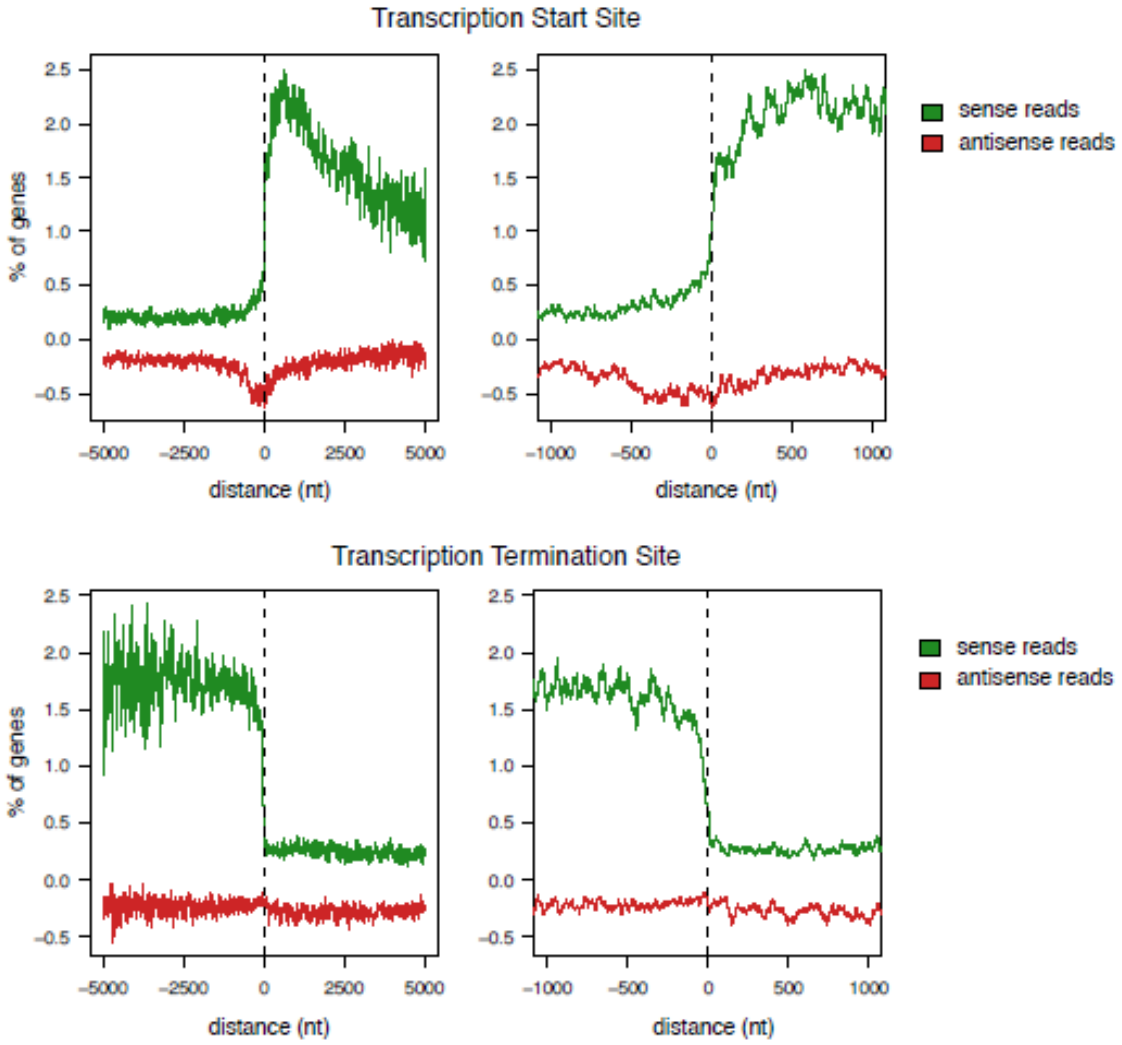


Figure S6, related to Figure 2B.

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.

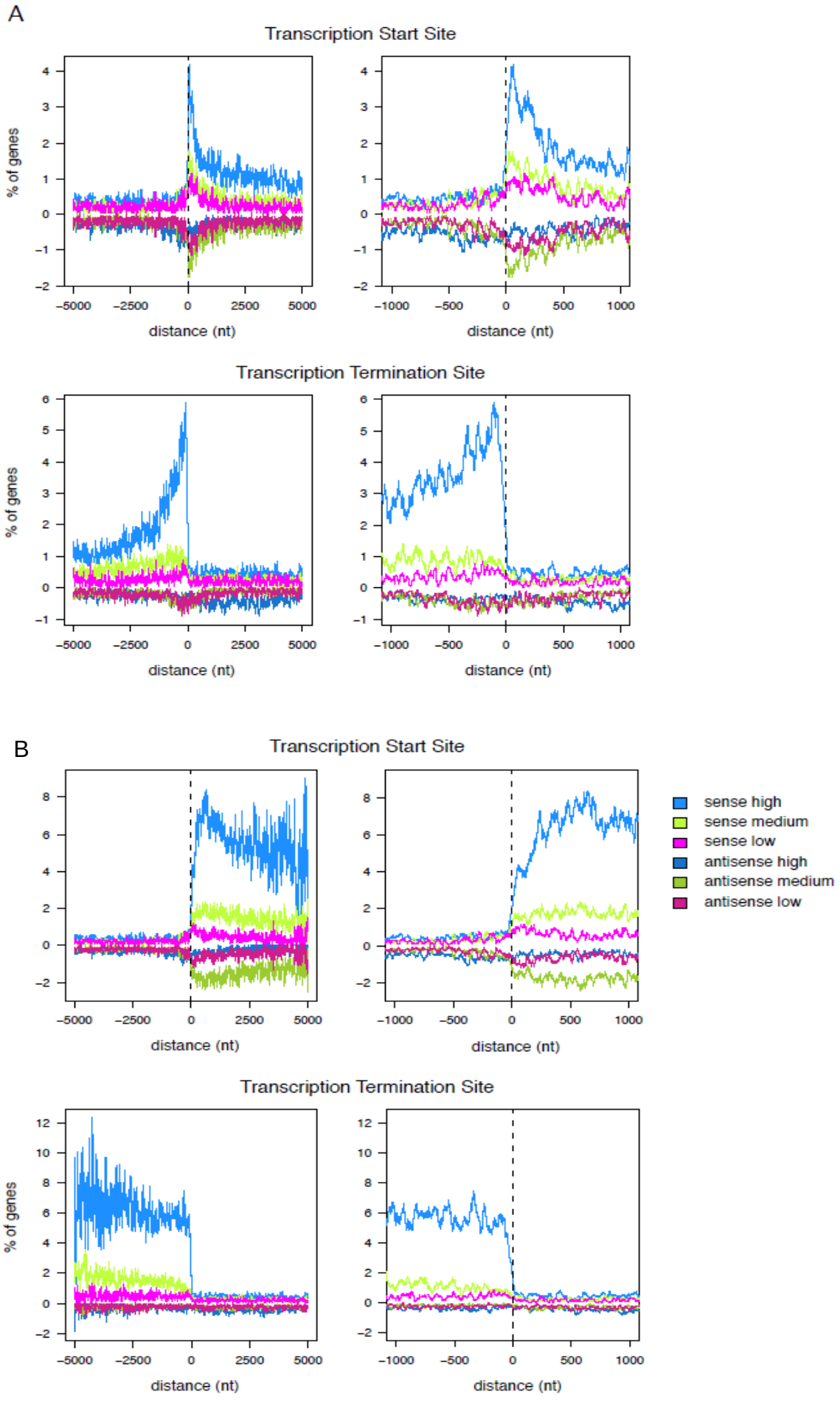

A
Transcription Start Site
% of genes
distance (nt)
Transcription Termination Site
% of genes
distance (nt)
B
Transcription Start Site
distance (nt)
Transcription Termination Site
% of genes
distance (nt)
sense high
sense medium
sense low
antisense high
antisense medium
antisense low

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).

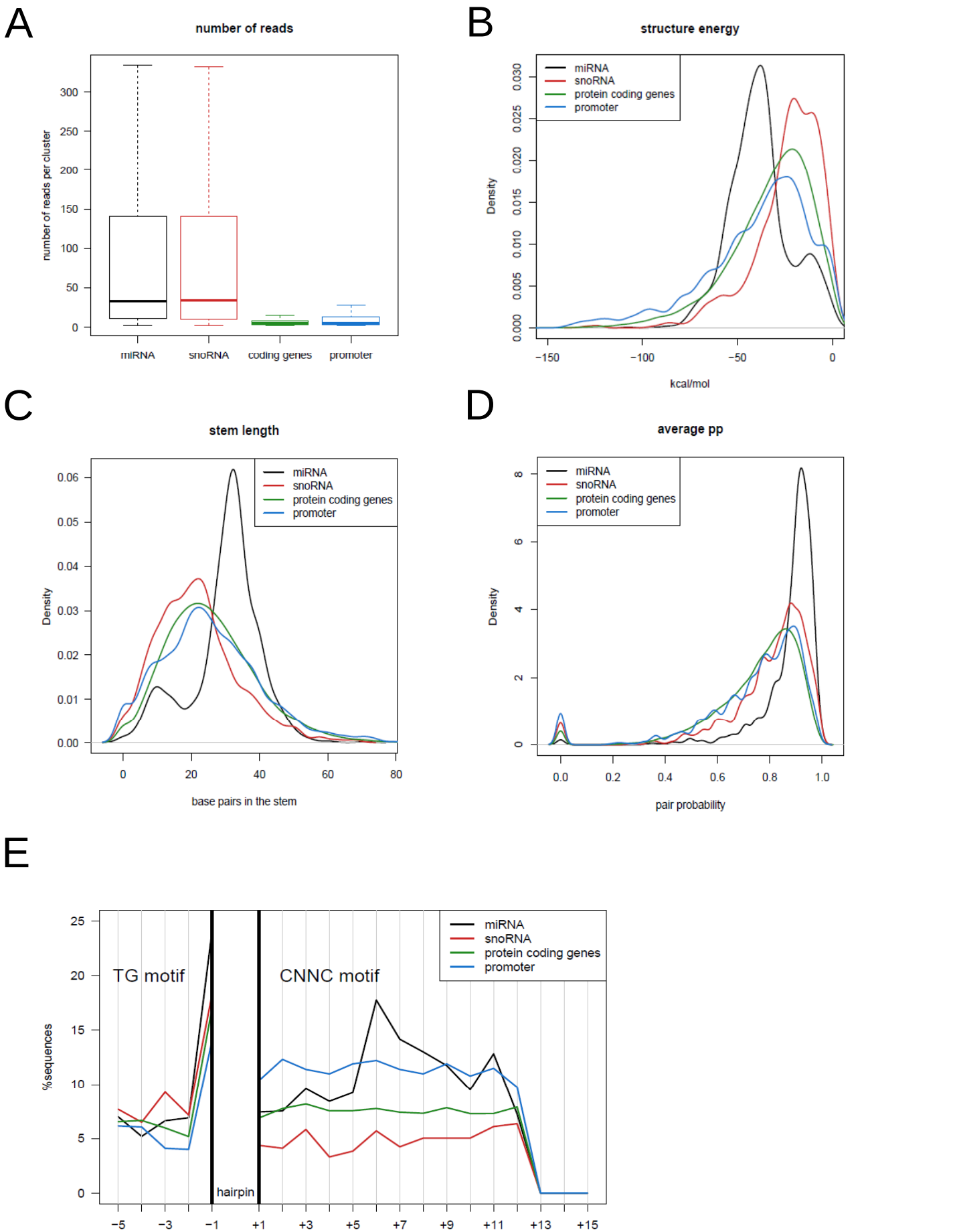

A
number of reads
number of reads per cluster
300
250
200
150
100
50
0
miRNA
snoRNA
coding genes
promoter
B
structure energy
miRNA
snoRNA
protein coding genes
promoter
Density
0.030
0.025
0.020
0.015
0.010
0.005
0.000
−150
−100
−50
0
kcal/mol
C
stem length
miRNA
snoRNA
protein coding genes
promoter
Density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

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.