DGCR8 HITS-CLIP reveals novel functions for the Microprocessor

Sara Macias1, Mireya Plass2,3, Agata Stajuda1, Gracjan Michelewski1,5, Eduardo Eyras2,4 & Javier F Cáceres1

The Drosha–DGCR8 complex (Microprocessor) is required for microRNA (miRNA) biogenesis. DGCR8 recognizes the RNA substrate, whereas Drosha functions as the endonuclease. Using high-throughput sequencing and cross-linking immunoprecipitation (HITS-CLIP) we identified RNA targets of DGCR8 in human cells. Unexpectedly, miRNAs were not the most abundant targets. DGCR8-bound RNAs also comprised several hundred miRNAs as well as small nuclear RNAs (snRNAs) and long noncoding RNAs. We found that the Microprocessor controlled the abundance of several miRNAs as well as of MALAT1. By contrast, DGCR8-mediated cleavage of snRNAs was independent of Drosha, suggesting the involvement of DGCR8 in cellular complexes with other endonucleases. Binding of DGCR8 to cassette exons is a new mechanism for regulation of the relative abundance of alternatively spliced isoforms. These data provide insights in the complex role of DGCR8 in controlling the fate of several classes of RNAs.

As DGCR8 is the Microprocessor component that recognizes and directly binds RNA, it is the best candidate to study new targets for this complex and to elucidate whether the Microprocessor is involved in the processing of cellular RNAs other than miRNAs. For this reason, we performed HITS-CLIP of DGCR8 in HEK 293T cells. This technique has been used to map in vivo RNA–protein interactions for several RNA-binding proteins, such as the neuronal splicing factor NOVA21 and the SR protein family member SRSF1 (ref. 22), as well as to define miRNA–mRNA interactions mediated by Argonaute proteins in both mouse brain23 and in nematodes24. More recently, several variations of this protocol have been introduced (individual-nucleotide-resolution CLIP and photoactivatable ribonucleoside–enhanced CLIP)25,26.

As expected, we identified most of the miRNAs known to be expressed in HEK 293T cells as well as the DGCR8 mRNA, which was known to be a target of the Microprocessor complex27–29. However, these were not the most abundant targets. We found that DGCR8 binds to many different RNAs, including several hundred miRNAs, snRNAs and long noncoding RNAs (lncRNAs). Whereas DGCR8 together with Drosha controls the abundance of several miRNAs and specific alternatively spliced isoforms, the DGCR8-mediated cleavage of snRNAs is independent of Drosha. This suggests the involvement of DGCR8 in cellular complexes with other endonucleases. Altogether, these data suggest an expanded role for the Microprocessor and/or DGCR8 in the control of RNA processing.

**RESULTS**

**HITS-CLIP identifies new targets of the Microprocessor**

To determine whether the Microprocessor is involved in the processing of cellular RNAs (other than miRNAs) that contain hairpin structures,
we focused on the identification of endogenous RNA targets of DGCR8. For this, we used HITS-CLIP, which relies on UV-light irradiation to induce protein covalent cross-links in protein–RNA complexes in situ. This allows the use of highly stringent immunoprecipitation and wash conditions to select only the RNAs that are directly bound to the protein of interest (reviewed in refs. 30,31). We performed two replicate HITS-CLIP experiments on UV light–irradiated HEK293T cells. Each replicate comprised two independent experiments: one involved immunoprecipitation of endogenous DGCR8 protein (Supplementary Fig. 1a), and the second one involved immunoprecipitation of a transiently transfected epitope-tagged DGCR8 protein (pCG-T7-DGCR8) (Supplementary Fig. 1b). We focused our analyses on the second replicate, as it gave higher depth of coverage and better mapping to the genome. Results from the first replicate are shown in Supplementary Figure 1c–e. In the second HITS-CLIP experiment we obtained 37 million reads for the endogenous DGCR8 protein, and 36 million reads for the epitope-tagged T7-DGCR8; we mapped 93% and 98% of the reads to the genome, respectively (Supplementary Table 1). After removing read duplicates, uniquely mapped reads from each of the data sets were clustered according to their genomic positions. To identify DGCR8 binding sites (clusters), we applied a modified false discovery rate (mFDR) cutoff to determine each cluster, as previously described24. We assessed the reproducibility of the DGCR8 targets identified in CLIP experiments for both endogenous and overexpressed DGCR8 proteins. Analysis of the reads of these two samples revealed a strong correlation (Pearson correlation coefficient, R > 0.8) (Fig. 1a).

To describe DGCR8 bona fide targets, we focused on overlapping clusters between these two samples (a complete list is available at http://regulatoygenomics.upf.edu/Data/DGCR8). In this way, we found 8,196 common targets (harboring one or more clusters) at the genomic level, which we mapped to noncoding regions with the aim of describing the targets of DGCR8 by the comparison of the cluster landscape generated by endogenous and overexpressed DGCR8. The binding sites allowed the identification of 8,196 targets.

Figure 2 DGCR8 binds and controls the stability of the long intergenic noncoding RNA MALAT1. (a) Schematic representation of DGCR8 binding clusters on MALAT1 and overlap of the binding sites with small RNA libraries. (b) Immunoprecipitation followed by RT-PCR analysis for overexpressed (left) and endogenous (right) DGCR8 binding endogenous MALAT1 RNA. (c) siRNA–mediated depletion of Drosha (siRNA to DROSHA) compared to mock-transfected cell situation (control siRNA). (d) mESCs lacking DGCR8 (Dgcr8 knockout cells) had upregulated Malat1 expression compared to wild-type cells (v6.5) that is partially reversed upon reintroduction of mouse Dgcr8 (mDgcr8). (e) Malat1 expression in mESCs lacking Dicer and in the parental mESC line (where Dicer was flanked by loxp sites (flxox) but not removed). For c–e, values are averages of at least three independent experiments; * P < 0.05 (t test). Errors bars, s.d. (f) A construct containing 500 base pairs of the 5′ end of MALAT1 RNA (boxed region in a) was fused to firefly luciferase open reading frame (LUC) and transfected into Drosha-depleted (siRNA to DROSHA) or mock-depleted (control siRNA) HEK cells. Levels of luciferase were analyzed relative to the cotransfected luciferase reporter. An SV40 construct containing a single miR-18a target site (SV40-miR18a) in its 3′ UTR was used to monitor miRNA activity. Error bars, s.d. (n = 3, except for LUC siRNA control, n = 2 with two values plotted separately), and * P < 0.05 (t test).
that were distributed as follows: 45% mapped to intergenic regions, 43% to protein coding genes and 5% to lncRNAs (Fig. 1b). We also found that 30% of the genomic clusters mapped to repetitive elements, mainly long interspersed element (LINE)-1 and long terminal repeat (LTR) retroelements, but were excluded these from the graphical representation. When analyzing protein-coding genes, we observed that the majority of the clusters were located in introns (26%, including miRNAs), with 12% in coding exons, and also found clusters in 3' and 5' untranslated regions (UTRs) (Fig. 1b). We also observed CLIP tags mapping to other noncoding RNAs, such as rRNA, small nucleolar RNAs (snRNAs) and snoRNAs (Fig. 1c). As expected, because of the canonical function of the Microprocessor, we identified 117 miRNAs (Fig. 1d), with the let-7 family being the most abundant family. We also identified the only so far described noncanonical RNA substrate for the Microprocessor, the DGCR8 mRNA27 (Fig. 1d). For each of the clusters, an RNA secondary structure was predicted to select those clusters overlapping with structures resembling that of pri-miRNAs (see http://regulatorygenomics.upf.edu/Data/DGCR8 and Supplementary Note). We found that more than 60% of the DGCR8 binding sites overlapped stable RNA secondary structures (Supplementary Table 1). We observed that 11–20% of the DGCR8 binding sites overlapped with small RNAs containing a 5'-phosphorylated end, which are proposed to be generated by endonucleolytic activity32.

The Microprocessor regulates the abundance of MALAT1
MALAT1 (metastasis-associated lung adenocarcinoma transcript 1, also known as NEAT2) is a lncRNA that is abundant in mammals and that had been initially identified owing to its elevated expression in several cancers33. We found that DGCR8 bound to MALAT1 and several other lncRNAs (Figs. 1b, 2a and http://regulatorygenomics.upf.edu/Data/DGCR8). We confirmed this interaction by immuno-precipitating overexpressed and endogenous DGCR8 protein followed by an analysis of the associated RNA by semiquantitative reverse-transcription (RT)-PCR (Fig. 2b). To study the effect of Microprocessor binding to the MALAT1 RNA, we determined its steady-state levels upon siRNA-mediated depletion of Drosha in HeLa cells. We observed a twofold increase in the levels of MALAT1 RNA when Drosha levels were reduced (Fig. 2c), with the let-7 family being the most abundant family. We also identified the only so far described noncanonical RNA substrate for the Microprocessor, the DGCR8 mRNA27 (Fig. 1d). For each of the clusters, an RNA secondary structure was predicted to select those clusters overlapping with structures resembling that of pri-miRNAs (see http://regulatorygenomics.upf.edu/Data/DGCR8 and Supplementary Note). We found that more than 60% of the DGCR8 binding sites overlapped stable RNA secondary structures (Supplementary Table 1). We observed that 11–20% of the DGCR8 binding sites overlapped with small RNAs containing a 5'-phosphorylated end, which are proposed to be generated by endonucleolytic activity32.
DGCR8 regulates snoRNA stability independently of Drosha

Another class of RNAs bound by DGCR8 was snoRNAs (Fig. 1c). These small RNAs are responsible for guiding many chemical modifications of other RNAs, such as rRNA, tRNA and snRNAs. First, we confirmed the binding of endogenous and overexpressed DGCR8 to snoRNAs observed by CLIP using immunoprecipitation followed by quantitative RT-PCR in HEK 293T cells (Fig. 3a). This reported binding to snoRNAs has a physiological consequence as seen by the accumulation of U17a, U16 and U92, which represent different subclasses of snoRNAs, in mouse cells lacking DGCR8, in HeLa cells in which DGCR8 was knocked down (Fig. 3b) or when overexpressing a dominant negative form of DGCR8 in HEK293T cells (Supplementary Fig. 2a–c). Mature snoRNA levels did not change in the absence of Dicer (Fig. 3b). Thus, DGCR8 can affect the stability of the mature forms of an H/ACA box snoRNA (U17a), a C/D box snoRNA (U16) and a small Cajal body–specific RNA (U92) (Fig. 3b). SnoRNAs are mostly located in introns and transcribed as part of the host gene. Their biogenesis involves trimming of the host introns from the 5′ and 3′ end after pre-mRNA splicing and release of the mature snoRNA that is protected from degradation by the core snoRNP components. We determined that DGCR8 affects the levels of mature snoRNAs but not of their precursor forms (Fig. 3c), indicating that DGCR8 acts after synthesis. To examine the role of the Microprocessor in snoRNA stability in human cells, we studied the effect of either depleting Drosha or overexpressing a dominant negative form of Drosha. To our surprise we could not detect any significant change in mature snoRNA levels in these situations (Fig. 3d and Supplementary Fig. 2d). These data suggested a new DGCR8 function that operates independently of Drosha, strongly indicating that DGCR8 might associate with other nucleases to cleave and destabilize snoRNAs. To test this hypothesis we incubated radiolabeled mature snoRNAs with immunoprecipitated T7-DGCR8 and observed the appearance of smaller bands, corresponding to digested products (Fig. 3e). Addition of immunoprecipitated Flag-Drosha did not result in formation of these subproducts, despite Drosha being active in the processing of a pri-miRNA precursor (pri-miR-24-2) (Fig. 3e). This DGCR8–mediated effect was specific for snoRNAs, as U1 snRNA was not cleaved by the DGCR8 complex (Fig. 3e). In vitro cleavage analyses revealed the precise location of the snoRNA cleavages, which occur toward the snoRNA 3′ end, potentially generating 3′-end fragments of 20-nucleotide small RNAs for C/D box snoRNAs (U16) and of 40-nucleotides for H/ACA (U92) (Supplementary Fig. 3a,b). We observed stabilization of the 70-nucleotide fragment when overexpressing dominant negative forms of DGCR8 in human cells (Supplementary Fig. 2e,f). This observation agrees with previous bioinformatic analyses of DGcr8 knockout and Dicer knockout small RNA libraries whereby snoRNA-encoded small RNAs (sdRNAs) are affected by the loss of DGCR8 and Dicer. Here we observed that the loss of DGCR8 decreased the amount of small RNAs originated from both the 5′ and the 3′ ends for both C/D box and H/ACA snoRNAs (Supplementary Fig. 3c,d), and these sites did not coincide with the cleavage sites previously described for H/ACA box snoRNAs (at the H box), which are required for the production of miRNA-like molecules in a Drosha-dependent and Dicer-dependent pathway. Collectively, these data demonstrate a role for DGCR8 in controlling the abundance of mature snoRNAs, in a Drosha-independent manner. This strongly suggests that DGCR8 associates with other endonuclease(s) to specifically destabilize snoRNAs and/or to produce new types of small RNAs.

The Microprocessor cleaves mRNA of protein-coding genes

We obtained DGCR8 CLIP tags mapping to 2,256 different mRNAs, comprising 3,610 clusters, indicating that some mRNAs harbor more than one DGCR8 binding site. We also found DGCR8 binding sites at the 5′ UTR and first exon of the DGCR8 mRNA and, in agreement with previous data, these clusters overlapped hsa-miR-1306 (ref. 27)
Modulation of alternative spliced isoforms by DGCR8

Analysis of HITS-CLIP data revealed DGCR8 binding to 241 predicted cassette exons (Fig. 5a). This would suggest the interesting possibility that Microprocessor-mediated cleavage of these cassette exons would affect the relative abundance of alternatively spliced isoforms. A global analysis of alternative splicing by exon junction arrays in mouse cells lacking Dgcr8 revealed 318 changes in cassette exons, as well as 40 alternative 5’ or 3’ splice site events and 18 changes in mutually exclusive exons (Supplementary Fig. 5a and Supplementary Table 2), suggesting a role for this complex in regulation of alternative splicing.

We first focused on four cassette exons shown to be bound by DGCR8 in the CLIP experiments that also have an RNA secondary structure resembling that of pri-miRNAs (Supplementary Fig. 5b–e). An UV light cross-linking assay revealed that DGCR8 binding to each of these cassette exons was affected when the predicted secondary structures were disrupted by mutation (Fig. 5b). Isoforms bound by DGCR8 were stabilized in human cells depleted of Drosha or in mouse cells lacking Dgcr8 (Fig. 5c–f and Supplementary Fig. 5f), strongly suggesting that the Microprocessor specifically cleaves and destabilizes mRNA isoforms containing DGCR8 binding sites, leading to an altered ratio of the alternatively spliced isoforms. Accordingly, we observed an overlap of these DGCR8 alternatively spliced exons with small RNA libraries (Fig. 5c,d), confirming that cleavage can occur at these locations, which could help to explain the change in the relative abundance of those isoforms containing the DGCR8-bound exon. In addition, in those cases that we tested, we could confirm that no changes in alternative splicing were observed in the absence of Dicer, indicating that these effects are independent from the presence of miRNAs (Tcf3 and Csnk1d) (Fig. 5f and Supplementary Fig. 5g). Altogether these data suggest that DGCR8 binding to cassette exons may influence the relative abundance of alternatively spliced isoforms.

DISCUSSION

Our analysis revealed that DGCR8 bound more than 2,000 mRNAs and in many cases depletion of Microprocessor components led to an upregulation of these mRNAs, suggesting that they are direct targets of the Microprocessor. For some of these mRNAs that were upregulated upon depletion of Drosha or DGCR8, we confirmed Drosha-dependent cleavage in vitro. This extends a previous observation showing that a hairpin localized in the 5’ UTR of the DGCR8 pre-mRNA is targeted by the Microprocessor27–29. Nevertheless, it remains elusive whether any of these cleavage products are also Dicer substrates that would lead to the generation of miRNA-like small RNAs. However, we observed that a large proportion of DGCR8 binding sites also overlap with small RNAs harboring a 5’-phosphorylated tail.
end, which have been proposed to be produced by the action of endonucleases. The fact that the Microprocessor destabilizes mRNAs containing DGCR8 binding sites in cassette exons suggested that the Microprocessor might have a role in the processing of these alternative exons, which would impact regulation of alternative splicing. We confirmed that DGCR8 can modulate the relative abundance of alternatively spliced isoforms and that binding to these sequences depends on RNA secondary structure.

Functional studies of the miRNA processing machinery have shown that Drosha, DGCR8 and Dicer deficiencies generate similar phenotypes, resulting in early embryonic lethality, strongly suggesting that miRNAs are essential for normal development38,39. Furthermore, several lines of evidence indicated additional functions for Drosha and/or Dicer, although initial global studies suggested DGCR8 mRNA as the only noncanonical substrate for the Microprocessor40. First, microarray profiling of mouse embryonic fibroblasts derived from Drosha and Dicer knockout mice showed poor correlation between the populations of mRNAs affected by these two endonucleases. Second, Drosha was shown to recognize and cleave many protein-coding mRNAs with secondary stem-loop structures in early-stage thymocytes41 as well as viral RNAs to directly regulate viral expression42. Finally, Drosha-dependent mRNA cleavage events that functionally regulate mRNA levels in mESCs have been recently described43. A simple explanation for these observations is that noncanonical functions for the Microprocessor and/or Dicer could be responsible for these effects. The comparison of these reports with our data only resulted in a partial overlap, probably due to the use of different cell lines for these studies (T cells versus HEK 293T cells). It also remains possible that binding of DGCR8 might not always lead to changes in relative mRNA expression.

Here we provided a genome-wide view of DGCR8 binding sites in HEK 293T cells. Our studies revealed more functions for the Microprocessor component DGCR8, implicating this protein in the control of noncoding RNAs stability (IncRNAs and snoRNAs) and of the relative abundance of alternatively spliced isoforms. In vitro processing assays with immunoprecipitated DGCR8 showed cleavage of the mature snoRNA species in a Drosha-independent manner, suggesting the involvement of DGCR8 in cellular complexes with other endonucleases (Fig. 3 and Supplementary Figs. 2 and 3). A DGCR8-independent role for Drosha in pre-rRNA processing also has been suggested by recent evidence. Whereas Drosha is required for processing of 125 pre-rRNA13,44, Dgcr8 knockout mESCs did not display evident defects in ribosomal RNA processing45, suggesting alternative complexes to the canonical Microprocessor. The presence of miRNA-like molecules encoded by H/ACA snoRNAs independent of Drosha function but involving a Dicer-dependent cleavage activity has been reported37. Bioinformatic analyses of the abundance of small RNAs encoded from snoRNAs showed that are dependent on both DGCR8 and Dicer18, although the role of Drosha was not assessed. From an evolutionary point of view, some miRNAs have been shown to evolve from snoRNAs. Their structure conservation as well as the retained capacity to bind dsykerin and fibrillarin, two of the core components from snoRNP particles, suggests a common origin45–48. DGCR8 has been reported to localize to the nucleolus49. Our work proposes a new complex between DGCR8 and another endonuclease or other endonucleases that would specifically cleave snoRNAs and probably other classes of cellular RNAs. Although it is well known how snoRNAs are synthesized, very little is known about their turnover and decay, and it would be of great interest to study the implication of DGCR8 in this process. These findings raise the possibility that there could be an alternate DGCR8 complex or complexes using different nucleases to process a variety of cellular RNAs.

All the new functions described here for the Microprocessor complex could be useful to reinterpret the phenotypes observed for Drosha- or DGCR8-deficient cells that were mainly attributed to the deficiency of miRNAs. DGCR8 is located in the genomic region (22q11.2) that is deleted in DiGeorge syndrome patients that expands around 30 genes11,12. Mouse models for this deletion showed a minor alteration of miRNA abundance but still exhibit behavioral and cognitive deficits, and cardiac abnormalities that resemble the human condition47. The Dgcr8 heterozygous mouse still showed some of the traits present in DiGeorge syndrome patients, such as altered short-term plasticity that has been linked to schizophrenia48. Although it is clear that in this situation that the miRNA biogenesis pathway is affected, it would be of great interest to determine the importance of the new DGCR8 functions for the origin and development of the disease.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Gene Expression Omnibus database: GSE39086 (sequencing raw data for endogenous and overexpressed DGCR8 HITS-CLIP).

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We are grateful to S. Heras for discussions and critical reading of the manuscript; N. Kim (Seoul National University) for Flag-Drosha, dominant negative DGCR8 and Drosha expression vectors; B. Seraphin (Institut de Génétique et de Biologie Moléculaire et Cellulaire, Strasbourg) for the Dcp1 antibody and R. Bellot (University of California, San Francisco) for Dicer knockout and Dicer flox/flox cell lines. This work was supported by the Medical Research Council and by the Wellcome Trust (084057/Z/07/Z to J.F.C.; A.S., G.M. and S.M. were partially funded by this). E.E. and M.P. were supported by grants from the Spanish Ministry of Science and by the Sandra Ibarra Foundation (BIO2008-01091, BIO2011-23920 and CSD2009-00080). S.M. was the recipient of a European Molecular Biology Organization long-term postdoctoral fellowship. M.P. is supported by the Novo Nordisk Foundation. J.F.C. is a recipient of a Wellcome Trust Senior Investigator Award (grant 095518/Z/11/Z).

AUTHOR CONTRIBUTIONS

S.M. and J.F.C. conceived, designed and interpreted the experiments. S.M., G.M. and CSD2009-00080). S.M. was the recipient of a European Molecular Biology Organization long-term postdoctoral fellowship. M.P. is supported by the Novo Nordisk Foundation. J.F.C is a recipient of a Wellcome Trust Senior Investigator Award (grant 095518/Z/11/Z).

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/nsmb.2344. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.
9. Bernstein, E., Caudy, A.A., Hammond, S.M. & Hannon, G.J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).

10. Hultqvist, G. et al. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**, 834–838 (2001).

11. Shiohama, A., Sasaki, T., Noda, S., Minoshima, S. & Shimizu, N. Molecular cloning and expression analysis of a novel gene DGC88 located in the DiGeorge syndrome chromosomal region. *Biochem. Biophys. Res. Commun.* **304**, 184–190 (2003).

12. Landthaler, M., Yalcin, A. & Tuschi, T. The human DiGeorge syndrome critical region gene 8 and its D. melanogaster homolog are required for miRNA biogenesis. *Curr. Biol.* **14**, 2164–2167 (2004).

13. Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F. & Hannon, G.J. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**, 231–235 (2004).

14. Gregory, R.I. et al. The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**, 235–240 (2004).

15. Davis, B.N. & Hata, A. Regulation of microRNA biogenesis: a miRiad of mechanisms. *Mol. Cell* **7**, 537–545 (2009).

16. Winter, J., Jung, S., Keller, S., Gregory, R.I. & Diederichs, S. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat. Cell Biol.* **11**, 228–234 (2009).

17. Kim, Y., Loedige, I. & Filipowicz, W. The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.* **11**, 597–610 (2010).

18. Han, J. et al. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* **125**, 887–901 (2006).

19. Faller, M., Matsunaga, M., Yin, S., Loo, J.A. & Guo, F. Heme is involved in microRNA processing. *Nat. Struct. Mol. Biol.* **14**, 23–29 (2007).

20. Faller, M. et al. DGCR8 recognizes primary transcripts of microRNAs through highly cooperative binding and formation of higher-order structures. *RNA* **16**, 1570–1583 (2010).

21. Licatalosi, D.D. et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464–469 (2008).

22. Sanford, J.R. Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res.* **19**, 381–394 (2009).

23. Chi, S.W., Zhang, J.B., Mele, A. & Darnell, R.B. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* **460**, 479–486 (2009).

24. Zeng, S.W., Zeng, J.B., Mele, A. & Darnell, R.B. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* **460**, 479–486 (2009).

25. Zeng, S.W., Zeng, J.B., Mele, A. & Darnell, R.B. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* **460**, 479–486 (2009).

26. Winter, J. et al. Post-transcriptional crossregulation between Drosha and DGCR8. *Cell* **141**, 479–486 (2009).

27. Krol, J., Loedige, I. & Filipowicz, W. The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.* **11**, 597–610 (2010).

28. Shiohama, A., Sasaki, T., Noda, S., Minoshima, S. & Shimizu, N. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* **125**, 887–901 (2006).

29. Faller, M., Matsunaga, M., Yin, S., Loo, J.A. & Guo, F. Heme is involved in microRNA processing. *Nat. Struct. Mol. Biol.* **14**, 23–29 (2007).

30. Faller, M. et al. DGCR8 recognizes primary transcripts of microRNAs through highly cooperative binding and formation of higher-order structures. *RNA* **16**, 1570–1583 (2010).

31. Licatalosi, D.D. et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464–469 (2008).

32. Sanford, J.R. et al. Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res.* **19**, 381–394 (2009).

33. Chi, S.W., Zeng, J.B., Mele, A. & Darnell, R.B. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* **460**, 479–486 (2009).

34. Zeng, S.W., Zeng, J.B., Mele, A. & Darnell, R.B. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* **460**, 479–486 (2009).

35. Winter, J. et al. Post-transcriptional crossregulation between Drosha and DGCR8. *Cell* **141**, 479–486 (2009).

36. Kadener, S. et al. Genome-wide identification of targets of the drosha-pasha/DGCR8 complex. *RNA* **15**, 537–545 (2009).

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

38. Shiohama, A., Sasaki, T., Noda, S., Minoshima, S. & Shimizu, N. Nucleolar localization of DGCR8 and identification of eleven DGCR8-associated proteins. *Exp. Cell Res.* **313**, 129–141 (2010).

39. Han, J. et al. Posttranscriptional crosstalk between Drosha and DGC88. *Cell* **141**, 479–486 (2009).

40. Faller, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).

41. Han, J. et al. Posttranscriptional crossregulation between Drosha and DGCR8. *Cell* **141**, 479–486 (2009).

42. Kadener, S. et al. Genome-wide identification of targets of the drosha-pasha/DGCR8 complex. *RNA* **15**, 537–545 (2009).

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

RNA extraction and quantitative RT-PCR. Total RNA was isolated from cells by Trizol, following manufacturer's instructions. RNA was DNase-treated (RQ1 DNase, Promega) and checked for DNA contamination. Five hundred nanograms of total RNA was used for quantitative RT-PCR analyses with the SuperScript III Platinum SYBR Green One-Step qRT-PCR kit (Invitrogen, 11736-051). Primers for quantitative RT-PCR analysis are listed in Supplementary Table 3. Data were analyzed with Bio-Rad CFX Manager software. For all the experiments we showed the average and s.d. of at least three independent biological replicates.

Northern blot for snoRNAs. Northern blots were performed as previously described. Briefly, 5–15 µg of total RNA was separated in 1× TBE-urea 10% polyacrylamide gel. Gel was transferred to nylon membrane (Hybond, GE Healthcare) for 1 h using Genie Blotter in 1× TBE. Membrane was crosslinked twice 1,200 J/m². Pre-hybridization was performed overnight at 40°C in hybridization buffer (1× SSC, 1% SDS and 200 µg/ml ssDNA from Sigma (D7656)). Templates and probes for hybridization were prepared using miRNA Probe Construction kit (Ambion, AM1530). Probe hybridization was performed overnight at 40°C and washed 3 times, each with wash buffer (0.2× SSC and 0.2% SDS). Radioactive signals were analyzed using a Phosphorimager.

Radioactive RNA labeling and in vitro processing. Templates for RNA synthesis and labeling were obtained by genomic PCR (primers are listed in Supplementary Table 3) adding a T7 promoter sequence at the 5’ end. Primer-24-2 was cloned in pGEM-Easy (Promega) and NdeI digested overnight to be transcribed. Transcription reactions were performed with T7 polymerase from Ambion (AM2082) in the presence of 40 µmol of [32P]UTP. Probes were gel-purified, phenol-extracted and ethanol precipitated. We included 50,000 c.p.m. of each probe either with immunoprecipitated T7-DGCR8, Flag-Drosha or control immunoprecipitates in the presence of buffer A (0.5 mM ATP, 20 mM creatine phosphate and 3.2 mM MgCl₂). Reactions were incubated for 30 min at 30°C, followed by standard phenol-chloroform extraction and ethanol precipitation. RNAs were separated in an 8–10% 1× TBE-polyacrylamide urea gel. Gels were analyzed using a Phosphorimager.

Cell culture, antibodies, plasmids and transfection experiments. HEK 293T and HeLa cells were cultured in DMEM supplemented with 10% FCS and pen-strep and grown in standard conditions. For mESCs, gelatin-coated plates were used without growth factor. HeLa cells were cultured in DMEM supplemented with 10% FCS and pen-strep and the following commercial antibodies were used: mouse monoclonal (Novagen, 69522), anti–α-tubulin (Sigma, T4026). Anti-Dicer1 antibody was provided by B. Seraphin (Institut de Génétique et de Biologie Moléculaire et Cellulaire, Strasbourg). A mammalian expression vector for Dgcr8 was created by inserting a Dgcr8 ORF (ENST00000351989, Ensembl) using XbaI and DpnII sites into the mammalian expression vector, pCGT7, which has been previously described. Vectors for overexpression of wild-type Drosha and dominant-negative forms of DGCR8 and Drosha (Flag-Drosha, mDRBD1&2 DGCR8, AC114 Drosha, respectively) were provided by N. Kim (Seoul National University). Transfections were performed using Lipofectamine 2000 (Invitrogen) and cells were collected 48 h after transfection. DROSHA knockdown was performed in HeLa cells with Dharmafect 4 (Dharmacon) and two rounds of siRNA transfection. SiRNAs against Drosha, DGCR8 mRNAs and nontargeting siRNAs were purchased from Dharmacon (L-016996-00-1, L-015713-00-0 and D-001810-02, respectively). Cells were collected 24 h after the second transfection.

Immunoprecipitations and quantitative RT-PCR. Cells overexpressing DGCR8 or Drosha proteins were collected 48 h after transfection. Cells were trypsinized, centrifuged and washed with PBS. Cell pellets were resuspended in buffer D (20 mM Hepes-KOH pH 7.9, 100 mM KCl, 0.2 mM EDTA, 0.5 mM DTT, 0.2 mM PMSF, 5% glycerol and RNAse inhibitors) and sonicated with Bioruptor for 5 cycles (30 s on 30 s off). Lysates were centrifuged for 10 min at maximum speed and the supernatant was saved as a cell lysate. A T7 tag antibody agarose from Novagen was used (69026) to immunoprecipitate T7-DGCR8, whereas anti-Flag M2 affinity from Sigma (A220) was used to immunoprecipitate Flag-Drosha. Beads were washed 3 times with buffer D and sonicated lysates were added to bind overnight at 4°C. The next day, immunoprecipitated material was washed five times with buffer D (150 mM KCl). For in vitro processing reactions, beads were resuspended in 150 µl of buffer D and kept at −80°C until use. For RNA extraction, beads were resuspended in 250 µl of 0.3 M sodium acetate and 750 µl of Trizol LS (Invitrogen). RNA was purified following the manufacturer's instructions. The RNA pellet was resuspended in water (20–25 µl). As input, 1/20th part of the starting material was saved before immunoprecipitation and was resuspended in 20 µl after RNA extraction. Before RT-PCR analysis, RNA samples were treated with RQ1 DNase (Promega, M6610A) for 1 h at 37°C, phenol-chloroform–extracted and ethanol-precipitated. For RT-PCR experiments, 1 µl of input and immunoprecipitated RNA was used with SuperScript III One-Step RT-PCR with Platinum Taq (Invitrogen, 12574-026) and several cycles of amplification conditions were used (25–40) to obtain semiquantitative results. For quantitative RT-PCR, SuperScript III Platinum SYBR Green One-Step qRT-PCR Kit was used (Invitrogen, 11736-051). Quantitative RT-PCR analysis of the immunoprecipitated material was normalized to the amount of material in the input fraction. Primers used for this analysis are listed in Supplementary Table 3. RT-PCR products were analyzed in ethidium-stained agarose gels.

Identification of snoRNA cleavage sites. Transcripts were 5’ end–labeled with T4 polynucleotide kinase (Roche) and [γ-32P]ATP (4,500 Ci/mmol; Amersham Pharmacia Biotech). Transcripts were purified by electrophoresis in a 10% denaturing polyacrylamide gel, and labeled RNAs were visualized by autoradiography. RNAs were incubated in buffer A in the presence of total extracts or immunopurified T7-DGCR8 (as described above). To determine the cleavage sites, the products of the 5’-labeled RNA fragmentation reaction along with the products of alkaline hydrolysis and limited T1 nuclease digestion of the same RNA molecule were separated on 10% polyacrylamide gels containing 7.5 M urea, 90 mM Tris-borate buffer and 2 mM EDTA. The alkaline hydrolysis ladder was generated by incubation of labeled RNA in formamide containing 0.5 mM MgCl₂ at 100°C for 10 min. Partial T1 RNase digestion of RNAs was performed under semidenaturing conditions (10 mM sodium citrate pH 5.0, 3.5 M urea) with 0.2 U/µl of the enzyme and incubation at 55°C for 15 min. Electrophoresis was performed at 1,500 V and was followed by autoradiography at ~80°C with an intensifying screen or exposed to a Phosphorimager screen (Molecular Dynamics).

UV light cross-linking. Uniformly labeled RNA probes were incubated with immunopurified T7-DGCR8 from cells in 20 µl reactions containing 10 mM Tris-HCl pH 7.5, 50 mM KCl, 0.5 mM DTT and RNAase inhibitors, at 4°C for 30 min. After incubation, reactions were placed in a 96-well plate on ice and irradiated with 254 nm UV light for 5 min. One microliter of RNase A/T1 cocktail was then added to each reaction and incubated at 37°C for 15 min to degrade unprotected RNA. Samples were loaded in 10% Tris-glycine protein gels and after electrophoresis, stained with Colloidal Blue Staining kit (Invitrogen) for loading purposes, and dried and exposed overnight on a Phosphorimager screen to visualize cross-linked DGCR8 protein.

mESC DGCR8 knockout rescue experiments. Mouse DGCR8 was amplified from total mouse RNA using oligos containing EcoRI and Sal restriction sites. The cDNA from DGCR8 was cloned in the plasmid pEFtr-IREs-DsRed-Express-2 vector (Clontech, 631980). Wild-type mESCs were transfected with pEFtr-IREs-DsRed-Express-2 empty plasmid and Dgcr8 knockout with the plasmid expressing mDGCR8 cdNA using Lipofectamine 2000 (Invitrogen). Six hours after transfection the medium was replaced with fresh medium, and 1 d after transfection G418 was added at 200 µg/ml for 5 d to select for transfected cells. After selection, RNA from transfected cells was extracted to proceed with the analysis.

 Luciferase reporter assays. The assays were performed in three independent replicates. HeLa cells were first depleted of Drosha using siRNAs (as indicated) and 24 h later were co-transfected with Luciferase reporter plasmids and a second round of Drosha siRNA using Dharmafect DUO (Dharmacon). An empty plasmid (LUC) containing a TK promoter fused to firefly luciferase based on pGL3-basic vector (a gift from J. Reddington, Medical Research Council, Human Genetics Unit, Edinburgh) was used to clone a 500-bp fragment from MALAT1 sequence.
upstream of the Firefly ORF using HindIII sites. A miRNA-reporter plasmid (SV40-miR18a) was also transfected to monitor Drosha knockdown efficiency\(^51\).

In all assays, a plasmid expressing Renilla luciferase was co-transfected as an internal control. Twenty-four hours after transfection cells were lysed and the levels of firefly and Renilla luciferase activity were measured using the Promega Dual Luciferase Reaction System. The data are expressed as a ratio of firefly/Renilla luciferase activity and normalized to mock (siRNA control) value. Luminescence was measured using a Monolight 3010 luminometer (Pharmingen).

**Identification of homologous mouse exons.** The Ensembl 54 annotation was used to extract the human-mouse one-to-one orthologs and the mouse exons\(^52\). For each of the gene pairs, the best-reciprocal exon pairs were calculated by performing all against all local exon alignments with exonerate\(^53\).

**Identification of alternative exons.** Expressed sequence tag (EST) alignments to hg18 from the University of California Santa Cruz (UCSC) genome browser were used for this purpose\(^54\). For each internal exon the inclusion level \(I\) was calculated as the proportion of ESTs verifying both splice sites (\(EST_i\)) over all ESTs including or excluding (\(EST_e\)) both splice sites

\[
I = \frac{EST_i}{EST_i + EST_e}
\]

Only those exons with more than ten ESTs (\(EST_i + EST_e\)) were considered. An exon was classified as 'cassette' if the inclusion was lower than 0.9 and as 'constitutive' otherwise. For mouse, the EST alignments to mm9 from the UCSC genome browser were used\(^54\). The inclusion levels for mouse exons were calculated as above and only internal exons were considered.

**HITS-CLIP protocol and processing and mapping of sequencing reads.** See Supplementary Note.

49. Michlewski, G. & Caceres, J.F. Antagonistic role of hnRNP A1 and KSRP in the regulation of let-7a biogenesis. Nat. Struct. Mol. Biol. 17, 1011–1018 (2010).

50. Caceres, J.F., Misteli, T., Screaton, G.R., Spector, D.L. & Krainer, A.R. Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. J. Cell Biol. 138, 225–238 (1997).

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

52. Flicek, P. et al. Ensembl's 10th year. Nucleic Acids Res. 38, D557–D562 (2010).

53. Slater, G.S. & Birney, E. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6, 31 (2005).

54. Fujita, P.A. et al. The UCSC Genome Browser database: update 2011. Nucleic Acids Res. 39, D876–D882 (2011).