

Published in final edited form as:

*Mol Cell*. 2012 September 14; 47(5): 734–745. doi:10.1016/j.molcel.2012.06.021.

## Promoters Recognized by Forkhead Proteins Exist for Individual 21U-RNAs

Germano Cecere<sup>1</sup>, Grace X.Y. Zheng<sup>2,†</sup>, Andres R. Mansisidor<sup>1</sup>, Katherine E. Klymko<sup>1,3</sup>, and Alla Grishok<sup>1,\*</sup>

<sup>1</sup>Department of Biochemistry and Molecular Biophysics, College of Physicians & Surgeons, Columbia University, New York, New York 10032, USA

<sup>2</sup>Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

<sup>3</sup>Department of Biology, Columbia College, Columbia University, New York, New York 10027, USA

### Summary

*C. elegans* 21U-RNAs are equivalent to the piRNAs discovered in other metazoans and have important roles in gametogenesis and transposon control. The biogenesis and molecular function of 21U-RNAs and piRNAs are poorly understood. Here, we demonstrate that transcription of each 21U-RNA is regulated separately through a conserved upstream DNA motif. We use genomic analysis to show that this motif is associated with low nucleosome occupancy, a characteristic of many promoters that drive expression of protein-coding genes, and that RNA polymerase II is localized to this nucleosome-depleted region. We establish that the most conserved 8-mer sequence in the upstream region of 21U-RNAs, CTGTTTCA, is absolutely required for their individual expression. Furthermore, we demonstrate that the 8-mer is specifically recognized by Forkhead family (FKH) transcription factors and that 21U-RNA expression is diminished in several FKH mutants. Our results suggest that thousands of small non-coding transcription units are regulated by FKH proteins.

### Introduction

There are multiple endogenous RNA interference (RNAi) pathways in *C. elegans* (Fischer, 2010). The three major ones involve different types of short RNAs discovered in the nematode: microRNAs (Lau et al., 2001; Lee and Ambros, 2001), endogenous-siRNAs (endo-siRNAs) (Ambros et al., 2003) and 21U-RNAs (Ruby et al., 2006). 21U-RNAs are 21-nt-long RNAs bearing a 5' terminal uridine residue. Thousands of 21U-RNAs are

© 2012 Elsevier Inc. All rights reserved.

\*Contact: ag2691@columbia.edu.

†Present address: Howard Hughes Medical Institute and Program in Epithelial Biology, Stanford School of Medicine, Stanford, CA94305.

#### Accession Numbers

GEO accession number for histone H3 and histone H2B ChIP-chip data sets is GSE38253.

#### Supplemental Information

The Supplemental Information includes seven Supplemental Figures with Supplemental Figure Legends, Supplemental Experimental Procedures and References.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

produced from two regions on chromosome IV spanning several megabases (Ruby et al., 2006) (Figure 1, Figure S1), and they map to introns of protein-coding genes as well as intergenic regions (Ruby et al., 2006). A specific 34-nt sequence motif with the 8-mer core sequence CTGTTTCA is located approximately 20-nt upstream of each 21U-RNA sequence (Ruby et al., 2006). Interestingly, this motif is conserved in *C. briggsae* and *C. remanei*, but the sequences of the 21U-RNAs are not (Ruby et al., 2006; de Wit et al., 2009). 21U-RNAs generally do not correspond to repetitive elements and their sequence complexity is similar to that of the *C. elegans* genome (Ruby et al., 2006).

21U-RNAs are similar to the *Piwi*-interacting RNAs (piRNAs) found in other animals (Lau, 2010). First, both 21U-RNAs and piRNAs are produced from large continuous regions of chromosomes, although no conserved motifs have been identified with piRNAs. Second, both types of short RNAs interact with the PIWI sub-family of Argonaute family proteins (Batista et al., 2008; Das et al., 2008; Wang and Reinke, 2008). 21U-RNAs exist in a complex with the PRG-1 protein (Batista et al., 2008; Wang and Reinke, 2008), which is localized to the P granules that specify the germline (Batista et al., 2008; Wang and Reinke, 2008). This complex is produced in the germline and is maternally contributed to the embryos and possibly early larvae (Batista et al., 2008; Das et al., 2008; Wang and Reinke, 2008). In *prg-1* mutant worms, the level of 21U-RNAs is greatly reduced, and the mutants are sterile at elevated temperatures (Batista et al., 2008; Das et al., 2008; Wang and Reinke, 2008).

Interestingly, while hundreds of closely located piRNAs in other animals are collinear and match the same DNA strand (Aravin et al., 2007), neighboring 21U-RNAs in *C. elegans* are often separated by only a few base pairs and come from different DNA strands (Ruby et al., 2006). This lack of collinearity and the presence of a conserved upstream motif along with a putative TATA box suggest that 21U-RNAs may represent independent transcription units, as was proposed by Ruby and colleagues at the time of 21U-RNA discovery (Ruby et al., 2006).

Here, we demonstrate that the conserved upstream sequence serves as a promoter for individual 21U-RNAs as it contains a DNA motif recognized by Forkhead transcription factors imbedded in poly(dA:dT) sequences that repel nucleosomes. We show that the Forkhead-binding DNA motif is absolutely required for 21U-RNA expression and that several proteins of this family contribute to 21U-RNA biogenesis. Consistent with these results, we find that RNA polymerase II is enriched at 21U-RNA loci in the germline and initiates transcription of nascent 21U-RNA at the -2 position relative to the first U in mature 21U-RNA.

## Results

### Nucleosome-depleted regions characteristic of promoters exist upstream of each 21U-RNA

The promoters of protein-coding genes often contain Nucleosome-Depleted Regions (NDR), which in many cases are due to the underlying DNA sequences (Vinces et al., 2009; Khoueiry et al., 2010). DNA sequences are known to have intrinsic properties that either favor or hinder interactions with nucleosomes (Kunkel and Martinson, 1981; Lowary and Widom, 1998). Although the extent to which these intrinsic properties affect nucleosome positioning *in vivo* is still a matter of debate, it is clear that in many cases underlying DNA sequences do play a role (Vinces et al., 2009; Khoueiry et al., 2010). For example, poly(dA:dT) sequences are known to repel nucleosomes and are often found in nucleosome-free regions of promoters (Struhl, 1985; Rando and Chang, 2009; Arya et al., 2010). If nucleosome-depleted regions exist upstream of each 21U-RNA, this may indicate their

independent transcriptional regulation. Indeed, we observed that in the conserved 34-nt upstream motif, there is a prominent poly(dA:dT) stretch (Ruby et al., 2006) that is likely to repel nucleosomes (Figure 7).

To establish a relationship between nucleosome occupancy and 21U-RNA sequences, we performed ChIP-chip experiments looking at the distribution of histone H3 across the two 21U-RNA- rich regions of chromosome IV. These experiments revealed a decreased number of histone H3 peaks in the areas producing 21U-RNAs (Figure S1). To describe this observation quantitatively, we analyzed the correlation between H3 enrichment peaks and 21U-RNA density on chromosome IV (Figure 1). The H3 enrichment peaks and 21U-RNA locations were binned, and a correlation coefficient was calculated between the total number of H3 peaks and 21U-RNAs in each bin. This analysis revealed that H3 enrichment was negatively correlated with the presence of 21U-RNAs (Figure 1A). We also observed a similar anti-correlation between ChIPchip peaks and 21U-RNA locations when we examined H2B::GFP expressed in the germline (Figure 1A) or endogenous H3 expressed in *glp-4* mutant worms that do not have germline tissue (Figure 1B), and similar results were obtained in *prg-1* mutant worms deficient in 21U-RNA accumulation (Figure 1B). To make sure that the low nucleosome occupancy found in 21U-RNA- rich regions was not due to an experimental artifact of the employed ChIP-chip technique, we analyzed additional nucleosome occupancy data obtained by micrococcal nuclease digestion and deep sequencing of the protected fragments (Valouev et al., 2008). This analysis also revealed a similar anti-correlation between nucleosome occupancy and 21U-RNA-rich regions (Figure 1A). Taken together, these results suggest that the DNA sequences located at 21U-RNA-rich regions may play a role in maintaining these chromosomal loci in a nucleosome-depleted state through a mechanism that is independent of transcriptional activity. Consistently, a theoretical model of nucleosome occupancy predicted for the *C. elegans* genome (Kaplan et al., 2009), which relies heavily on the nucleosome-repelling properties of poly(dA:dT), showed an anti-correlation between predicted nucleosome locations and 21U-RNA-rich regions similar to that found experimentally (Figure 1A).

To more precisely define the nucleosome occupancy around 21U-RNAs, we performed metagene analysis using published nucleosome occupancy data (Valouev et al., 2008), (Figure 2A). This analysis demonstrated that there are NDRs corresponding to the 34-nt sequence motif found upstream of individual 21U-RNAs, which is reminiscent of the NDRs found at the promoters of protein-coding genes in yeast (Mavrigh et al., 2008; Shivaswamy et al., 2008) and *C. elegans* (Valouev et al., 2008; Ercan et al., 2010; Ooi et al., 2010). The A +T content of DNA sequences upstream of 21U-RNAs is high (Figure 2B), suggesting that poly(dA:dT) tracks present in the 34-nt sequence motif contribute to the observed signature.

### RNA polymerase II is localized to the NDRs upstream of 21U-RNAs

The above analysis suggests that the upstream motif may function as a nucleosome-free promoter element to direct the transcription of individual 21U-RNAs. To test this, we analyzed the *in vivo* localization of RNA polymerase II (Pol II) across 21U-RNA loci using Pol II ChIP-seq data sets available from the modENCODE project. As shown in Figure 3A, Pol II peaks were indeed evident at the putative promoter corresponding to the NDR and the 21U-RNA upstream motif. Since NDRs upstream of 21U-RNAs are present both in the soma and in the germline, but 21U-RNAs are enriched in the latter, it is possible that transcriptional machinery is recruited to the NDRs upstream of each 21U-RNA mostly in the germline tissue. To test for specific recruitment of Pol II to the 21U-RNA loci in the germline, we performed ChIP-qPCR experiments using a well-established Pol II antibody (Baugh et al., 2009) in wild type adult worms and *glp-4* mutant worms that lack germline tissue. Indeed, we were able to detect a low but significant Pol II enrichment over the IgG control at both tested 21U-RNA loci in wild type worms but not in *glp-4* mutant worms

(Figure 3B). Similar results were obtained analyzing the Pol II enrichment over a germline-specific protein-coding gene (Figure S2). These results strongly indicate that Pol II is recruited to 21U-RNA promoters in the germline.

### **The core 8-mer CTGTTTCA upstream sequence is required for individual 21U-RNA expression**

If the upstream motif of 21U-RNAs is functionally required for their transcription, then a deletion in the core consensus sequence upstream of a particular 21U-RNA should compromise its expression. To address this question, we created an *in vivo* model where we could manipulate the consensus upstream 21U-RNA sequence. For this purpose, we used the natural *C. elegans* isolate strain JU258, which differs from the standard N2 Bristol strain by the existence of a number of characterized DNA deletions, some of which are located on chromosome IV in the 21U-RNA-rich regions (Figure 4A). We indeed found that JU258 worms lack specific 21U-RNAs normally present in N2 (Figure 4B), and when we complemented the niDf199 ~ 4kb deletion present in JU258 with a 35kb fosmid containing this region (Figure 4A) we were able to restore the expression of the missing 21U-RNAs, as measured by RT-qPCR (Figure 4B).

We confirmed that the expression profile of 21U-RNAs produced from the fosmid was identical to published results, with more 21U-RNAs present in L4 and adult (Figure S3A), and with a dependence on *prg-1* for their expression (Batista et al., 2008; Das et al., 2008; Wang and Reinke, 2008) (Figure S3B). We then used this transgenic *in vivo* model to evaluate the requirement of the upstream consensus motif for 21U-RNA production. The core 8-mer CTGTTTCA sequence is the most conserved within the 21U-RNA motif (Ruby et al., 2006), and it has been demonstrated that the abundance of individual 21U-RNAs correlates positively with the “consensus score” of their 8-mers (Batista et al., 2008). We therefore generated transgenic worms using a fosmid with a deletion of the 8-mer motif next to 21U-3372. Strikingly, the expression of this 21U-RNA was completely lost in the transgenic animals, but the neighboring 21U-RNAs located within 1.2 kb of 21U-3372 were produced at normal levels (Figure 4C). The loss of 21U-3372 expression correlated with an increase in nucleosome occupancy at its modified promoter (Figure 4D). This experiment demonstrates the requirement of the consensus 8-mer sequence for 21U-3372 expression and suggests that each 21U-RNA could be expressed as an independent transcriptional unit.

### **The 5' end of a nascent 21U-RNA maps two nucleotides upstream of the first uridine in 21U-RNA**

Multiple lines of evidence described above strongly suggest that the conserved upstream element may serve as a promoter for nascent 21U-RNA transcripts. Consistently, we were not able to detect transcription through the 21U-RNA upstream region by 5' RACE when we used a downstream RT primer antisense to the mature 21U-RNA (Figure S4). These negative results suggest that the 5' end of the nascent 21U-RNA transcript is located very close to the 5' end of the mature 21U-RNA. We repeated 5' RACE experiments with RT primers downstream of the mature 21U-RNA sequence, which revealed that the nascent 21U-RNA transcript is two nucleotides longer than the mature 21U-RNA at the 5' end (Figure 5A). These results are consistent with genome-wide deep sequencing of capped transcripts, which identified 21U-RNA transcription start sites globally (Gu and Mello, personal communication). Importantly, the 21U-RNA precursor transcript was enriched in the germline (Figure 5B), indicating that 21U-RNA genes are transcribed in germline tissue, which is consistent with the Pol II ChIP data shown in Figure 3B.

## Forkhead transcription factors recognize the CTGTTTCA DNA motif, localize to 21U-RNA loci *in vivo*, and promote 21U-RNA expression

Mapping the 5' end of the nascent 21U-RNA transcript just upstream of the first U, not including the upstream consensus motif, further solidified the possibility that the 8-mer motif could serve as a transcription factor binding site. Consistently, we find that *C. elegans* nuclear extracts contain factors that are able to specifically recognize the CTGTTTCA DNA motif, as shown by electrophoretic mobility gel shift assays (EMSA) using a dsDNA probe containing the 8-mer motif (Figure 5C). The binding was specific to the 8-mer motif since it could be competed with an excess of non-biotinylated dsDNA oligonucleotide (see Experimental Procedures), but not with a dsDNA oligonucleotide mutated in the conserved CTGTTTCA motif (CGCCCGCA). Also, no gel shift was observed with the mutated probe in independent experiments (Figure S5A). Together, our results suggest that the core CTGTTTCA sequence upstream of each 21U-RNA could serve as a binding site for a germline-enriched transcription factor that would allow Pol II-mediated transcription of 21U-RNAs instead of being involved in the processing of a putative 21U-RNA precursor transcript.

The CTGTTTCA sequence in the 21U-RNA motif bears a strong resemblance to the consensus binding site, TTGTTTAC, for the conserved DAF-16/FOXO family of transcription factors (Calnan and Brunet, 2008). Notably, the TGT TTT sequence common to the 21U-RNA motif and the FOXO binding site is essential for DNA recognition by the FOXO proteins according to structural studies (Obsil and Obsilova, 2010). We analyzed 21U-RNA accumulation in a *daf-16* null mutant but did not find any significant change in 21U-RNA accumulation compared to wild type worms (Figure S6B). Since there are 15 Forkhead (FKH) transcription factors in *C. elegans* (Hope et al., 2003), we considered the possibility that one or more of them could have an affinity to the 8-mer motif in the germline and regulate 21U-RNA transcription. We examined the transcript levels of all 15 FKH factors in wild type and *glp-4* mutant adults to identify proteins with preferential germline expression, and we selected eleven candidates that showed significant depletion in *glp-4(-/-)* (Figure S6A).

Next, we analyzed 21U-RNA levels in available mutants for some of these genes and used RNAi treatment for others. We found a significant decrease in the 21U-RNA expression in the *unc-130(ev505)* null mutant (Figure 6A, left). This reduction in 21U-RNAs is unlikely to be an indirect consequence of a germline defect or due to a positive regulation of *prg-1* transcription by UNC-130, because the mRNA and the protein levels of *prg-1* do not change in the *unc-130* mutant (Figure 6A, right). Interestingly, *unc-130(ev505)* exhibits reduced fertility, particularly at 25°C (Figure S6C), which is a known *prg-1* phenotype (Cox et al., 1998; Batista et al., 2008; Wang and Reinke, 2008).

21U-RNA levels were also significantly reduced upon simultaneous inactivation by RNAi of the closely related genes *fkh-3*, *fkh-4* and *fkh-5* (Figure 6B), a treatment that did not compromise germline function or *prg-1* expression. On the contrary, a decrease in 21U-RNA levels upon depletion of *fkh-1/pha-4* by RNAi (Figure S6D) is likely due to sterility associated with a decrease in *prg-1* mRNA (Figure S6E). However, these negative results do not allow us to exclude a possible role for PHA-4 and some other essential FKH factors, such as FKH-6 and LET-381, in the regulation of 21U-RNA expression. Instead, viable mutants of FKH proteins not enriched in the germline, such as *lin-31* and *fkh-9*, did not affect 21U-RNA accumulation (Figure S6B).

To test whether the Forkhead proteins found to specifically affect 21U-RNA production were able to bind the 21U-RNA upstream motif, we expressed the UNC-130, FKH-3 and FKH-5 proteins in bacteria and performed gel-shift experiments. Our results indicate that

these proteins can specifically recognize the DNA motif present upstream of 21U-RNAs (Figure 6D and Figure S5B–C) and can therefore play a direct and redundant role in the regulation of 21U-RNA transcription.

Next, we performed ChIP experiments to confirm the binding of FKH proteins to the 21U-RNA promoters *in vivo*. The occupancy of Pol II at the 21U-RNA promoters is very low (Figure 3B), suggesting that initiation of transcription is not very efficient and that FKH proteins may not be abundant in the germline. To obtain the high levels of FKH expression in the germline required for ChIP, we generated transgenic lines where UNC-130::GFP was expressed from a germline-specific promoter, *mex-5*. We used both bombardment (Praitis et al., 2001) and Mos1-mediated Single Copy transgene Insertion (*MosSCI*) (Frokjaer-Jensen et al., 2008) techniques for making transgenic strains and observed germline expression in several lines (Figure S7A). ChIP experiments with anti-GFP antibody using two transgenic strains detected an enrichment in UNC-130::GFP binding to DNA at multiple 21U-RNA loci but not at the control loci, including the chromosome IV region located between individual 21U-RNA genes, the *prg-1* promoter or the 18S RNA coding region (Figure 6C and S7B–C). These results confirm the specific interaction between UNC-130 and 21U-RNA genes *in vivo*.

## Discussion

Here we demonstrate that the DNA motif located upstream of each 21U-RNA serves as a promoter for its individual transcription by RNA Polymerase II. We show that this promoter is recognized by Forkhead transcription factors, which implicates this family of proteins in the regulation of small non-coding RNAs. Our findings open new directions for further studies of the mechanisms governing expression of small 21U-RNA genes and their biological functions.

### 21U-RNA Biogenesis

We have accumulated evidence supporting the independent transcription of individual 21U-RNAs and demonstrated germline-specific enrichment of RNA polymerase II at 21U-RNA loci. We propose a model for 21U-RNA biogenesis that relies on the conserved DNA motif present upstream of each 21U-RNA (Figure 7). In this model, the poly(dA:dT) sequence in the upstream motif helps to define a nucleosome-free region in chromatin, and the CTGTTTCA sequence allows binding of germline-enriched FKH transcription factors that promote initiation of Pol II-directed transcription 2nt upstream of the mature 21U-RNA sequence. Since FKH proteins are also expressed in somatic cells, it is possible that the germline expression of PRG-1, together with the germline enrichment of specific FKH proteins, leads to the accumulation of 21U-RNAs only in germline tissue. It is also possible that 21U-RNAs are expressed in additional specific cell types, such as neurons, and that this has not been uncovered due to the high abundance of germline tissue in adult worms. Moreover, we cannot exclude the possibility that a competitor present in somatic tissues, e.g. a protein with affinity to AT-rich DNA sequences, may bind to the 21U-RNA promoters and prevent their somatic transcription.

We believe that our transgenic system will be very useful for future studies addressing regulation of 21U-RNA transcription and the coupling between the transcription and the further modifications and processing of 21U-RNAs. Although mature 21U-RNAs are 21 nucleotides in length, the nascent transcripts are likely to be longer, since we have cloned a 21U-RNA precursor that included at least 38 nucleotides downstream of the 21U-RNA sequence.

RNA polymerase II transcribes protein-coding mRNA and also a variety of shorter non-coding RNAs, most notably spliceosomal U1 and U2 snRNAs (Lykke-Andersen and Jensen, 2007; Egloff et al., 2008). Regulation of transcription termination of the non-coding RNAs transcribed by Pol II has been best studied in yeast and involves the RNA-binding Nrd1p complex that interacts with Pol II and also recognizes specific RNA sequences (Lykke-Andersen and Jensen, 2007). The nuclear exosome trims the 3' ends of these RNAs until it reaches RNA elements protected by interacting proteins (Vasiljeva and Buratowski, 2006). It is possible that 21U-RNA transcription and processing are tightly coupled and that the nascent 21U-RNA transcripts are bound by PIWI protein PRG-1, protecting the first 21 nucleotides from nuclear exosome trimming. In the absence of PRG-1, the nascent 21U-RNA transcripts are likely to be completely degraded by the nuclear exosome. The involvement of a nuclease responsible for 3' end generation of *Drosophila* piRNAs has been proposed (Brennecke et al., 2007; Gunawardane et al., 2007), and, most recently, a 3' to 5' exonuclease activity has been implicated in piRNA biogenesis in a silkworm cell-free system (Kawaoka et al., 2011). The trimming of the 3' end has also been demonstrated in the biogenesis of primary siRNAs in *S. pombe* (Halic and Moazed, 2010). It would be very interesting to investigate a coupling between transcriptional regulation and 3' end trimming in the biogenesis of 21U-RNAs.

### 21U-RNA function

The best known function of piRNAs in *Drosophila* and mammals is control of repetitive elements (Siomi et al., 2011). Although some germline defects in piRNA-related mutants in *Drosophila* are secondary to mobilization of transposons (Klattenhoff and Theurkauf, 2008), the fact that less than 50% of vertebrate piRNAs map to repetitive regions (Lau, 2010) suggests that piRNAs may affect gametogenesis by other, yet undiscovered, means. There are few examples of 21U-RNAs initiating the silencing of transposons in *C. elegans* (Das et al., 2008), since the well-developed system of endogenous siRNAs interacting with worm-specific Argonautes (WAGO) is largely dedicated to genome surveillance in the nematode (Gu et al., 2009). Therefore, *C. elegans* 21U-RNAs represent a good model for addressing the role of piRNAs in fertility and beyond.

The PIWI-subfamily Argonaute protein PRG-1 was the first factor associated with 21U-RNA function. In addition, the worm ortholog of methyltransferase HEN1 required for methylation of *C. elegans* 21U-RNAs has been described recently (Billi et al., 2012; Montgomery et al., 2012). Now, we implicate several Forkhead proteins in 21U-RNA biogenesis, and it would be interesting to further investigate how they affect germline function. Specifically, we would be interested in finding similarities in phenotypes and gene expression changes between the *prg-1* mutant and available *unc-130* mutant or *fkh-5(RNAi)* worms.

One exciting direction for future work is the possibility of 21U-RNA function in the nervous system. Recent studies have reported expression of piRNAs in many tissues, including neurons (Lee et al., 2011; Yan et al., 2011). A limited set of piRNAs was shown to be expressed in the mouse hippocampus (Lee et al., 2011), and piRNAs have recently been discovered in the nervous system of *Aplysia* (Rajasethupathy et al., 2012). The involvement of UNC-130 in 21U-RNA regulation provides a possible link to the neuronal function of 21U-RNA. We found that *unc-130* is expressed in the germline, but this Forkhead protein has been previously implicated in the development of chemosensory neurons (Sarafi-Reinach and Sengupta, 2000) and axon guidance (Nash et al., 2000). Some targets of UNC-130 have been identified, but the mutant phenotype cannot be fully explained by the regulation of the known targets. It is especially intriguing that all known alleles of *unc-130*, even the nulls, display temperature sensitivity in that the defects are more pronounced at 25°C. Temperature sensitivity is also a key feature of *prg-1* mutants in *C. elegans*, and we

have shown that *unc-130(ev505)* partially phenocopies the temperature sensitive reduction in brood size characteristic of the *prg-1* mutant. Therefore, it would be very interesting to investigate whether *prg-1* mutant worms display *unc-130*-specific neuronal phenotypes.

In conclusion, our work provides a foundation for a number of research directions aimed at 1) understanding the coupling between 21U-RNA transcription and biogenesis, and 2) elucidating the roles of 21U-RNAs in the germline and nervous system. Since the biological role of *C. elegans* piRNAs (21U-RNAs) in fertility and potentially in direct gene expression regulation is more clearly separated from transposon control compared to other animals, future research about 21U-RNAs is likely to shed light on piRNA biology.

## Experimental Procedures

### *C. elegans* Strains

Strains were maintained at 20°C unless otherwise noted, using standard methods (Brenner, 1974). Bristol N2 was the wild-type strain used. All other strains used in this study are listed in the Supplemental Experimental Procedures.

### Chromatin Immunoprecipitation

Chromatin Immunoprecipitation was performed as described in the Supplemental Experimental Procedures.

### Preparation of DNA samples and ChIP-chip

Preparation of DNA samples for ChIP-chip analysis is described in the Supplemental Experimental Procedures.

### ChIP-chip data processing

We used data normalized by NimbleGen to perform the correlation analysis shown in Figure 1 (see below). We have also used raw data to generate similar correlation results. Raw data were median normalized for each channel.

### Nucleosome occupancy data

Published (Valouev et al., 2008) nucleosome occupancy data was downloaded from UCSC (<http://hgdownload.cse.ucsc.edu/goldenPath/ce4/database/nucleosomeStringency.txt.gz> and <http://hgdownload.cse.ucsc.edu/goldenPath/ce4/nucleosome/nucleosomeAdjustedCoverage.wigAscii.gz>). The data from Chromosome IV was used for the correlation analysis described in the Supplemental Experimental Procedure.

### Data processing of predicted nucleosome occupancy data

The raw data was downloaded from the Segal lab website ([http://genie.weizmann.ac.il/software/data/ce4\\_avg\\_occ.txt.gz](http://genie.weizmann.ac.il/software/data/ce4_avg_occ.txt.gz)), and was converted to chromosome format by running the perl script, `nucleo08_chv2chr.pl`, provided by the Segal lab. The data from Chromosome IV was used for the correlation analysis described in the Supplemental Experimental Procedures.

### Metagene analysis of nucleosome occupancy, Pol II occupancy, and sequence composition around 21U-RNAs

The metagene analyses were performed as described in the Supplemental Experimental Procedures.



## RNA extraction

Synchronous populations of animals were grown at 20°C on NGM plates seeded with OP50 *E. coli* at a density of approximately 100,000 animals per 15 cm Petri dish, and harvested at L4-Young Adult stage. The harvested animals were washed three times with M9 buffer and the pellet was frozen in dry ice with TRI Reagent (MRC, Inc.). After five times of freeze and thaw total RNA was isolated according to the TRI Reagent protocol. Ten micrograms of RNA was treated with 2U of Turbo DNase (Ambion) at 37°C for 1hr followed by phenol-extraction and isopropanol-precipitation.

## Quantitative RT-PCR

Short RNA RT-PCR was carried out as described previously (Chen et al., 2005; Das et al., 2008), except that 500ng of total RNA was used for each reverse transcription reaction in a final volume of 20µl. One tenth of RT product was used for qPCR reaction. The sequence of primers for RT and qPCR reactions was that described by (Das et al., 2008), additional primer information is available upon request. The reactions were performed in triplicate.

## 5' RACE

In order to evaluate putative precursor 21U-RNA transcripts, 5' rapid amplification of cDNA ends (RACE) was done using the Ambion First Choice™ RLM RACE kit according to the manufacturer's instructions. The 5' RACE was performed using 20 microgram of total RNA from adult *C. elegans* that was treated with DNase (Ambion) and enriched for mRNA using the Terminator 5'-Phosphate-Dependent Exonuclease (EpiBio).

## Recombinant fosmid construction

The WRM0611aH08 fosmid containing the niDf199 locus was obtained from *C. elegans* fosmid library generated by *C. elegans* Reverse Genetics Core Facility, Vancouver, B.C., Canada. ([http://www.geneservice.co.uk/products/clones/Celegans\\_Fos.jsp](http://www.geneservice.co.uk/products/clones/Celegans_Fos.jsp))

We generated a derivative fosmid construct lacking the 8-mer motif upstream of 21U-3372 by fosmid recombineering method as described by (Dolphin and Hope, 2006).

## RNAi experiments

We used RNAi-sensitized *eri-1* mutant genetic background (Kennedy et al., 2004) for all RNAi experiments. We put L1 larvae on plates seeded with bacteria producing dsRNA of interest and picked ten adult worms from each RNAi-treated plate, we washed worms with M9 and put them in TRI Reagent (MRC, Inc.). After five times of freeze and thaw total RNA was isolated accordingly to the TRI Reagent protocol, except than the final isopropanol-precipitated RNA pellets were resuspended in 10µl of water and used directly for RT reaction as described above.

## Protein expression and purification

We cloned the cDNAs of *unc-130*, *fkh-3*, and *fkh-5* in the multicloning site of the expression vector pMAL™-p2X Vector (New England BioLabs), which encodes maltose-binding protein (MBP), to create MBP fusion proteins. We used the BL21 Competent *E. coli* cells (Invitrogen) to express the recombinant MBP proteins following the manufacturer's instructions.

### Nuclear protein extraction

We prepared nuclear and cytoplasmic protein extracts as described by (Chen et al., 2000), except that we resuspended the nuclear pellet in 50mM Tris-HCl, pH7.5, 400mM KCl, 10mM MgCl<sub>2</sub> in order to extract the proteins from the pellet.

### Electroforetic Mobility Shift Assay (EMSA)

We performed the EMSA assay using biotinylated dsDNA oligonucleotide probes (synthesized by IDT) and the lightShift<sup>®</sup> Chemiluminescent EMSA Kit (Thermo Scientific, 20148) following the manufacturer's instructions. We used between 0.2 and 0.5μg of recombinant protein or 20μg of nuclear extract for each binding reaction.

### Western Blotting

Western blotting was performed as described in (Mansidor et al., 2011) using anti-Actin (Millipore, MAB1501R), anti-H3 (Millipore, 05-928), anti-Mouse IgG HRP labeled (Perkin Elmer), anti-Rabbit IgG HRP labeled (Perkin Elmer) antibodies, and anti-PRG-1 antibody (Batista et al., 2008), (a gift from the Mello lab).

### Fertility Assay

Gravid adults were grown at 20°C or 25°C for two generations. Their synchronized L1 progeny were single-picked and grown to adulthood at either 20°C or 25°C, respectively. Animals were transferred to fresh plates daily during the period of egg-laying and their progeny counted as larvae.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgments

We thank P. Sharp for advice, support of early stages of this work and comments on the manuscript, T. Maniatis and E. Greene for comments on the manuscript, W. Gu and C. Mello for communicating results before publication, L. Neal and P. Batista for performing preliminary experiments not included in the manuscript, B. Tursun and L. Cochella for technical advice, O. Rando for bioinformatic advice, R. Ruiz and S. Nicholis for technical assistance, I. Greenwald for equipment, J. Culotti for providing *unc-130* mutant strains, O. Hobert for reagents, C. Burge, G. Ruby, M. Gallio and members of the Sharp and Grishok labs for discussions. Some of the strains used in this study were provided by the Caenorhabditis Genetics Center, which is supported by the National Institutes of Health-National Center for Research Resources. This work was supported by 3260-07 Special Fellow Award from The Leukemia and Lymphoma Society, The Arnold and Mabel Beckman Foundation Young Investigator Award and NIH Director's New Innovator Award (1 DP2 OD006412-01) to A.G., United States Public Health Service grant P01-CA42063 from the National Cancer Institute to Phillip A. Sharp and partially by MIT Cancer Center Support (core) grant P30-CA14051 from the National Cancer Institute.

### References

- Ambros V, Lee RC, Lavanway A, Williams PT, Jewell D. MicroRNAs and other tiny endogenous RNAs in *C. elegans*. *Curr Biol*. 2003; 13:807–818. [PubMed: 12747828]
- Aravin AA, Hannon GJ, Brennecke J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science*. 2007; 318:761–764. [PubMed: 17975059]
- Arya G, Maitra A, Grigoryev SA. A structural perspective on the where, how, why, and what of nucleosome positioning. *J Biomol Struct Dyn*. 2010; 27:803–820. [PubMed: 20232935]
- Batista PJ, Ruby JG, Claycomb JM, Chiang R, Fahlgren N, Kasschau KD, Chaves DA, Gu W, Vasale JJ, Duan S, et al. PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in *C. elegans*. *Mol Cell*. 2008; 31:67–78. [PubMed: 18571452]

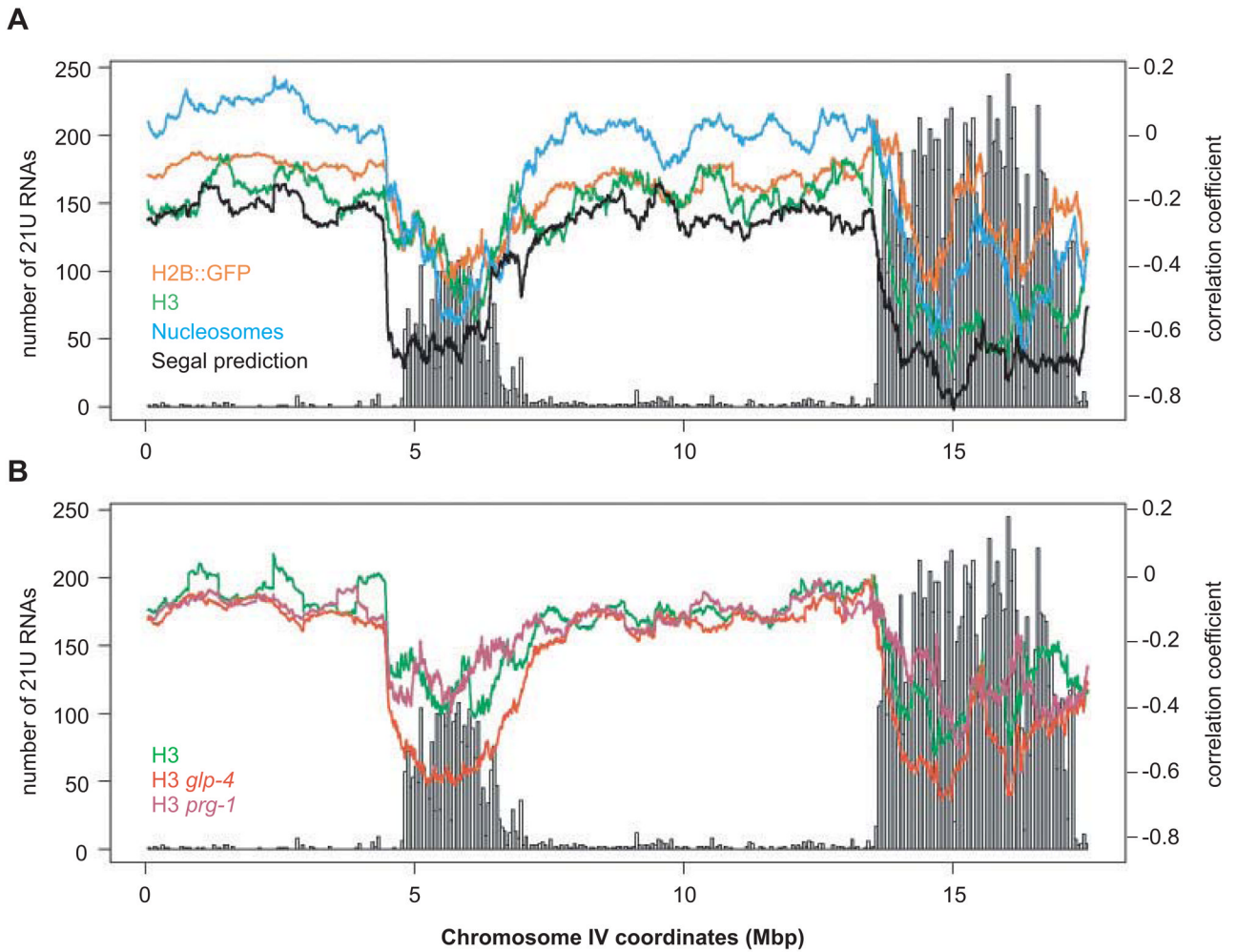
- Baugh LR, Demodena J, Sternberg PW. RNA Pol II accumulates at promoters of growth genes during developmental arrest. *Science*. 2009; 324:92–94. [PubMed: 19251593]
- Billi AC, Alessi AF, Khivansara V, Han T, Freeberg M, Mitani S, Kim JK. The *Caenorhabditis elegans* HEN1 Ortholog, HENN-1, Methylates and Stabilizes Select Subclasses of Germline Small RNAs. *PLoS Genet*. 2012; 8:e1002617. [PubMed: 22548001]
- Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell*. 2007; 128:1089–1103. [PubMed: 17346786]
- Brenner S. The genetics of *Caenorhabditis elegans*. *Genetics*. 1974; 77:71–94. [PubMed: 4366476]
- Calnan DR, Brunet A. The FoxO code. *Oncogene*. 2008; 27:2276–2288. [PubMed: 18391970]
- Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, Nguyen JT, Barbisin M, Xu NL, Mahuvakar VR, Andersen MR, et al. Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res*. 2005; 33:e179. [PubMed: 16314309]
- Chen F, Hersh BM, Conradt B, Zhou Z, Riemer D, Gruenbaum Y, Horvitz HR. Translocation of *C. elegans* CED-4 to nuclear membranes during programmed cell death. *Science*. 2000; 287:1485–1489. [PubMed: 10688797]
- Cox DN, Chao A, Baker J, Chang L, Qiao D, Lin H. A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes Dev*. 1998; 12:3715–3727. [PubMed: 9851978]
- Das PP, Bagijn MP, Goldstein LD, Woolford JR, Lehrbach NJ, Sapetschnig A, Buhecha HR, Gilchrist MJ, Howe KL, Stark R, et al. Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the *Caenorhabditis elegans* germline. *Mol Cell*. 2008; 31:79–90. [PubMed: 18571451]
- de Wit E, Linsen SE, Cuppen E, Berezikov E. Repertoire and evolution of miRNA genes in four divergent nematode species. *Genome Res*. 2009; 19:2064–2074. [PubMed: 19755563]
- Dolphin CT, Hope IA. *Caenorhabditis elegans* reporter fusion genes generated by seamless modification of large genomic DNA clones. *Nucleic Acids Res*. 2006; 34:e72. [PubMed: 16717278]
- Egloff S, O'Reilly D, Murphy S. Expression of human snRNA genes from beginning to end. *Biochem Soc Trans*. 2008; 36:590–594. [PubMed: 18631122]
- Ercan S, Lubling Y, Segal E, Lieb JD. High nucleosome occupancy is encoded at X-linked gene promoters in *C. elegans*. *Genome Res*. 2010
- Fischer SE. Small RNA-mediated gene silencing pathways in *C. elegans*. *Int J Biochem Cell Biol*. 2010; 42:1306–1315. [PubMed: 20227516]
- Frokjaer-Jensen C, Davis MW, Hopkins CE, Newman BJ, Thummel JM, Olesen SP, Grunnet M, Jorgensen EM. Single-copy insertion of transgenes in *Caenorhabditis elegans*. *Nat Genet*. 2008; 40:1375–1383. [PubMed: 18953339]
- Gu W, Shirayama M, Conte D Jr, Vasale J, Batista PJ, Claycomb JM, Moresco JJ, Youngman EM, Keys J, Stoltz MJ, et al. Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the *C. elegans* germline. *Mol Cell*. 2009; 36:231–244. [PubMed: 19800275]
- Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi MC. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science*. 2007; 315:1587–1590. [PubMed: 17322028]
- Halic M, Moazed D. Dicer-independent primal RNAs trigger RNAi and heterochromatin formation. *Cell*. 2010; 140:504–516. [PubMed: 20178743]
- Hope IA, Mounsey A, Bauer P, Aslam S. The forkhead gene family of *Caenorhabditis elegans*. *Gene*. 2003; 304:43–55. [PubMed: 12568714]
- Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J, et al. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*. 2009; 458:362–366. [PubMed: 19092803]
- Kawaoka S, Izumi N, Katsuma S, Tomari Y. 3' End Formation of PIWI-Interacting RNAs In Vitro. *Mol Cell*. 2011; 43:1015–1022. [PubMed: 21925389]
- Kennedy S, Wang D, Ruvkun G. A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*. *Nature*. 2004; 427:645–649. [PubMed: 14961122]

- Khoueiry P, Rothbacher U, Ohtsuka Y, Daian F, Frangulian E, Roure A, Dubchak I, Lemaire P. A cis-regulatory signature in ascidians and flies, independent of transcription factor binding sites. *Curr Biol*. 2010; 20:792–802. [PubMed: 20434338]
- Klattenhoff C, Theurkauf W. Biogenesis and germline functions of piRNAs. *Development*. 2008; 135:3–9. [PubMed: 18032451]
- Kunkel GR, Martinson HG. Nucleosomes will not form on double-stranded RNA or over poly(dA).poly(dT) tracts in recombinant DNA. *Nucleic Acids Res*. 1981; 9:6869–6888. [PubMed: 7335494]
- Lau NC. Small RNAs in the animal gonad: guarding genomes and guiding development. *Int J Biochem Cell Biol*. 2010; 42:1334–1347. [PubMed: 20227517]
- Lau NC, Lim LP, Weinstein EG, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*. 2001; 294:858–862. [PubMed: 11679671]
- Lee EJ, Banerjee S, Zhou H, Jammalamadaka A, Arcila M, Manjunath BS, Kosik KS. Identification of piRNAs in the central nervous system. *RNA*. 2011; 17:1090–1099. [PubMed: 21515829]
- Lee RC, Ambros V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science*. 2001; 294:862–864. [PubMed: 11679672]
- Lowary PT, Widom J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol*. 1998; 276:19–42. [PubMed: 9514715]
- Lykke-Andersen S, Jensen TH. Overlapping pathways dictate termination of RNA polymerase II transcription. *Biochimie*. 2007; 89:1177–1182. [PubMed: 17629387]
- Mansidor AR, Cecere G, Hoersch S, Jensen MB, Kawli T, Kennedy LM, Chavez V, Tan MW, Lieb JD, Grishok A. A Conserved PHD Finger Protein and Endogenous RNAi Modulate Insulin Signaling in *Caenorhabditis elegans*. *PLoS Genet*. 2011; 7:e1002299. [PubMed: 21980302]
- Mavrich TN, Ioshikhes IP, Venters BJ, Jiang C, Tomsho LP, Qi J, Schuster SC, Albert I, Pugh BF. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res*. 2008; 18:1073–1083. [PubMed: 18550805]
- Montgomery TA, Rim YS, Zhang C, Downen RH, Phillips CM, Fischer SE, Ruvkun G. PIWI Associated siRNAs and piRNAs Specifically Require the *Caenorhabditis elegans* HEN1 Ortholog henn-1. *PLoS Genet*. 2012; 8:e1002616. [PubMed: 22536158]
- Nash B, Colavita A, Zheng H, Roy PJ, Culotti JG. The forkhead transcription factor UNC-130 is required for the graded spatial expression of the UNC-129 TGF-beta guidance factor in *C. elegans*. *Genes Dev*. 2000; 14:2486–2500. [PubMed: 11018016]
- Obsil T, Obsilova V. Structural basis for DNA recognition by FOXO proteins. *Biochim Biophys Acta*. 2010
- Ooi SL, Henikoff JG, Henikoff S. A native chromatin purification system for epigenomic profiling in *Caenorhabditis elegans*. *Nucleic Acids Res*. 2010; 38:e26. [PubMed: 19966274]
- Praitis V, Casey E, Collar D, Austin J. Creation of low-copy integrated transgenic lines in *Caenorhabditis elegans*. *Genetics*. 2001; 157:1217–1226. [PubMed: 11238406]
- Rajasethupathy P, Antonov I, Sheridan R, Frey S, Sander C, Tuschl T, Kandel ER. A Role for Neuronal piRNAs in the Epigenetic Control of Memory-Related Synaptic Plasticity. *Cell*. 2012; 149:693–707. [PubMed: 22541438]
- Rando OJ, Chang HY. Genome-wide views of chromatin structure. *Annu Rev Biochem*. 2009; 78:245–271. [PubMed: 19317649]
- Ruby JG, Jan C, Player C, Axtell MJ, Lee W, Nusbaum C, Ge H, Bartel DP. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell*. 2006; 127:1193–1207. [PubMed: 17174894]
- Sarafi-Reinach TR, Sengupta P. The forkhead domain gene *unc-130* generates chemosensory neuron diversity in *C. elegans*. *Genes Dev*. 2000; 14:2472–2485. [PubMed: 11018015]
- Shivaswamy S, Bhinge A, Zhao Y, Jones S, Hirst M, Iyer VR. Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS Biol*. 2008; 6:e65. [PubMed: 18351804]
- Siomi MC, Sato K, Pezic D, Aravin AA. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol*. 2011; 12:246–258. [PubMed: 21427766]

- Struhl K. Naturally occurring poly(dA-dT) sequences are upstream promoter elements for constitutive transcription in yeast. *Proc Natl Acad Sci U S A*. 1985; 82:8419–8423. [PubMed: 3909145]
- Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, Zeng K, Malek JA, Costa G, McKernan K, et al. A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res*. 2008; 18:1051–1063. [PubMed: 18477713]
- Vasiljeva L, Buratowski S. Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol Cell*. 2006; 21:239–248. [PubMed: 16427013]
- Vinces MD, Legendre M, Caldara M, Hagihara M, Verstrepen KJ. Unstable tandem repeats in promoters confer transcriptional evolvability. *Science*. 2009; 324:1213–1216. [PubMed: 19478187]
- Wang G, Reinke V. A *C. elegans* Piwi, PRG-1, Regulates 21U-RNAs during Spermatogenesis. *Curr Biol*. 2008; 18:861–867. [PubMed: 18501605]
- Yan Z, Hu HY, Jiang X, Maierhofer V, Neb E, He L, Hu Y, Hu H, Li N, Chen W, et al. Widespread expression of piRNA-like molecules in somatic tissues. *Nucleic Acids Res*. 2011; 39:6596–6607. [PubMed: 21546553]

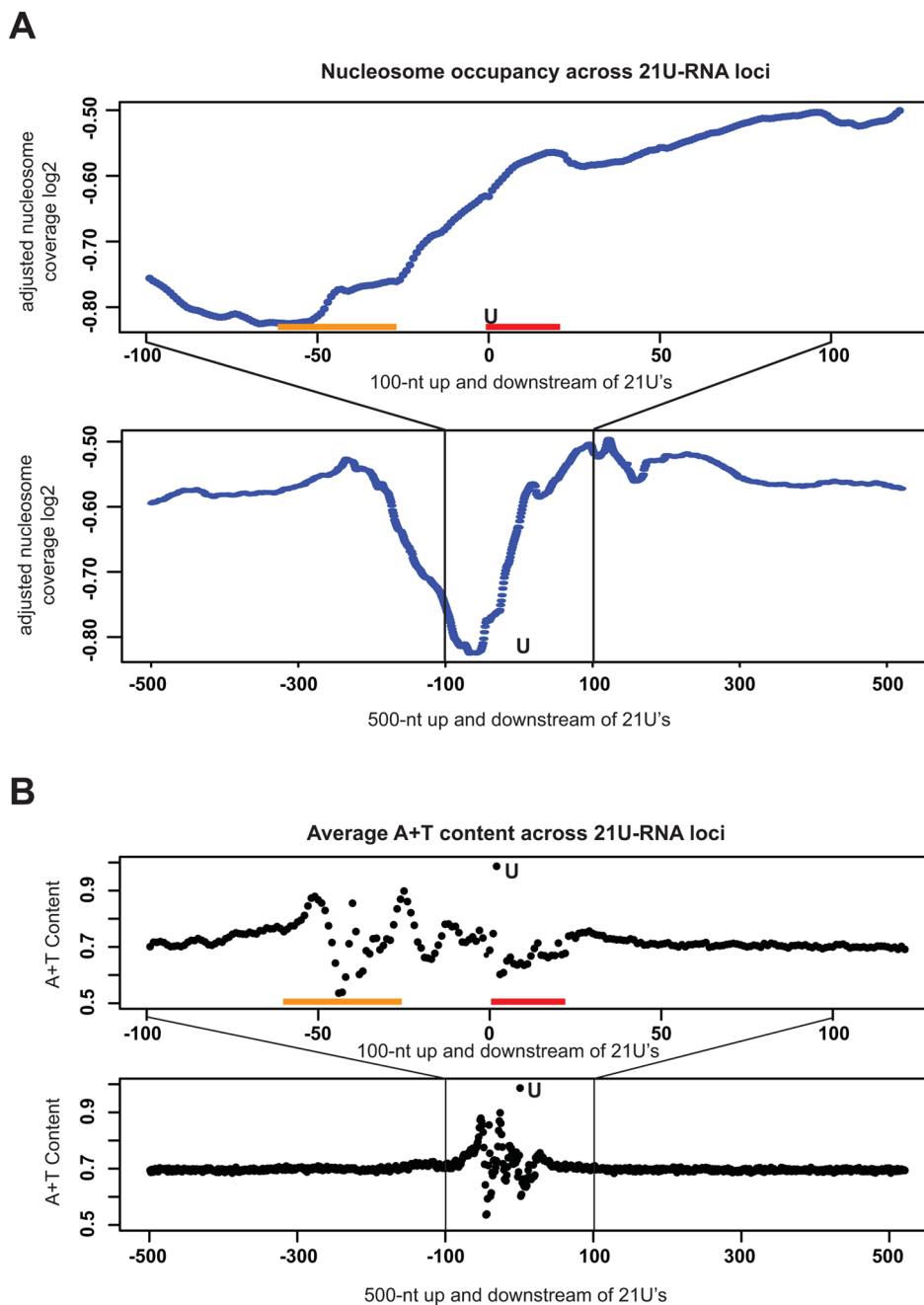
### Highlights

- A conserved 21U-RNA upstream motif is depleted of nucleosomes
- The CTGTTTCA core of the upstream motif is required for 21U-RNA expression
- Forkhead proteins bind the CTGTTTCA motif and stimulate 21U-RNA production
- Pol II initiates 21U-RNA transcription downstream of the CTGTTTCA motif



**Figure 1. 21U-RNA-rich loci on chromosome IV are depleted of nucleosomes**

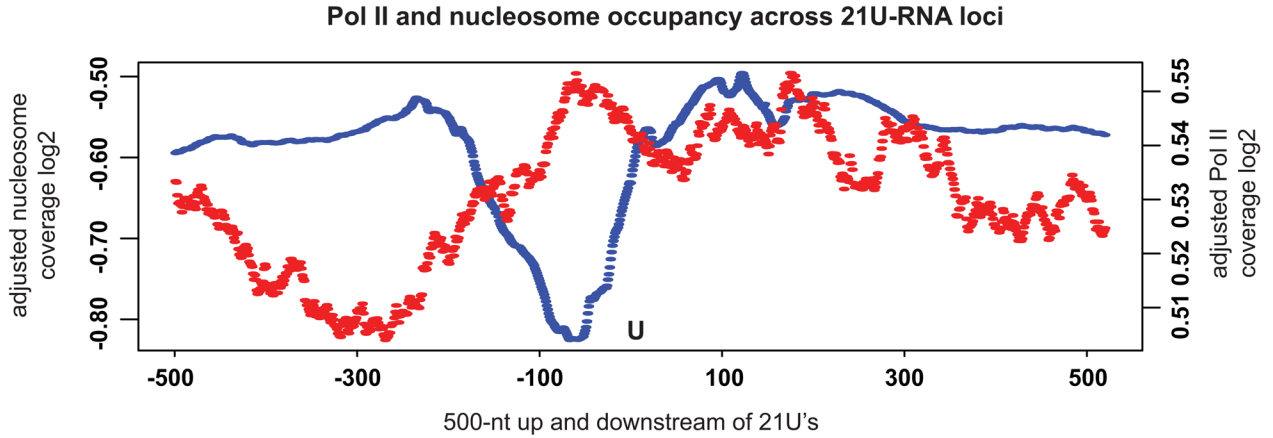
(A) Abundance of 21U-RNAs (Batista et al., 2008) (bars) (left y-axis) at the indicated regions of chromosome IV and correlation plots (lines, right y-axis) between 21U-RNA loci and histone H3 enrichment peaks obtained by ChIP-chip (green), germline-specific histone H2B::GFP ChIP-chip enrichment peaks (dark yellow), nucleosome-protected regions in micrococcal nuclease digest of chromatin (Valouev et al., 2008) (blue) and predicted nucleosome occupancy based on underlying DNA sequence (black); see also Figure S1. (B) Correlation plots (lines, right y-axis) between 21U-RNA loci and histone H3 enrichment peaks obtained by ChIP-chip performed on young adult animals with germline (green), lacking germline - *glp-4(bn2)*- (orange) or lacking 21U-RNAs - *prg-1(tm872)*- (purple). All the correlations in 21U-RNA rich regions are statistically significant ( $p < 0.05$ ).



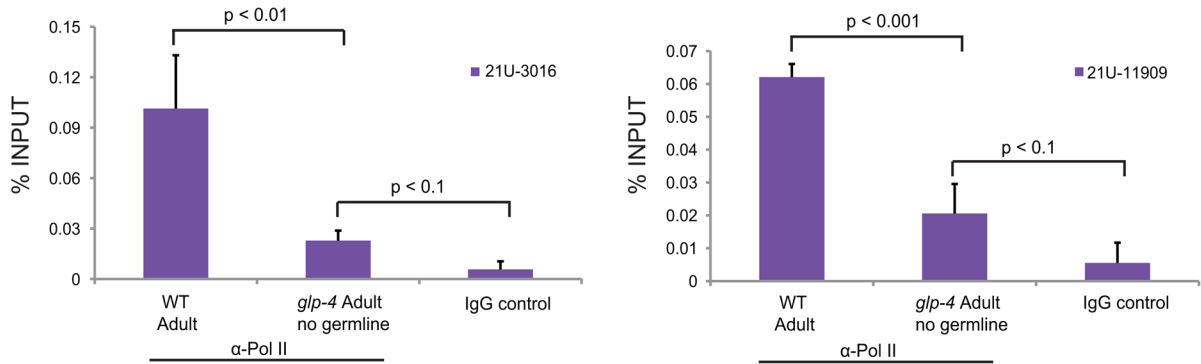
**Figure 2. Nucleosome-depleted regions exist upstream of individual 21U-RNA loci**  
 (A) Nucleosome occupancy per base pair averaged across all 21U-RNA loci and shown 100 nucleotides (top) or 500 nucleotides (bottom) upstream and downstream of the first U in 21U-RNA. The position of 21U-RNA (+1 to 21) on the X-axis is indicated in red, position of the conserved upstream motif (−59 to −25) is indicated in orange. (B) Average A+T content plots corresponding to regions shown in (A).



A

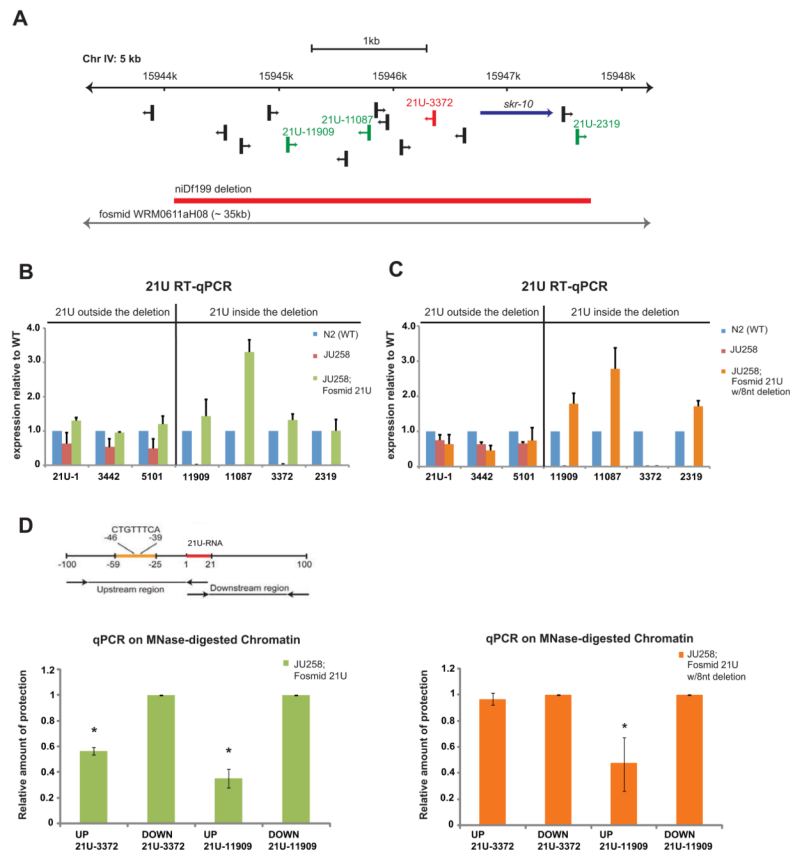


B

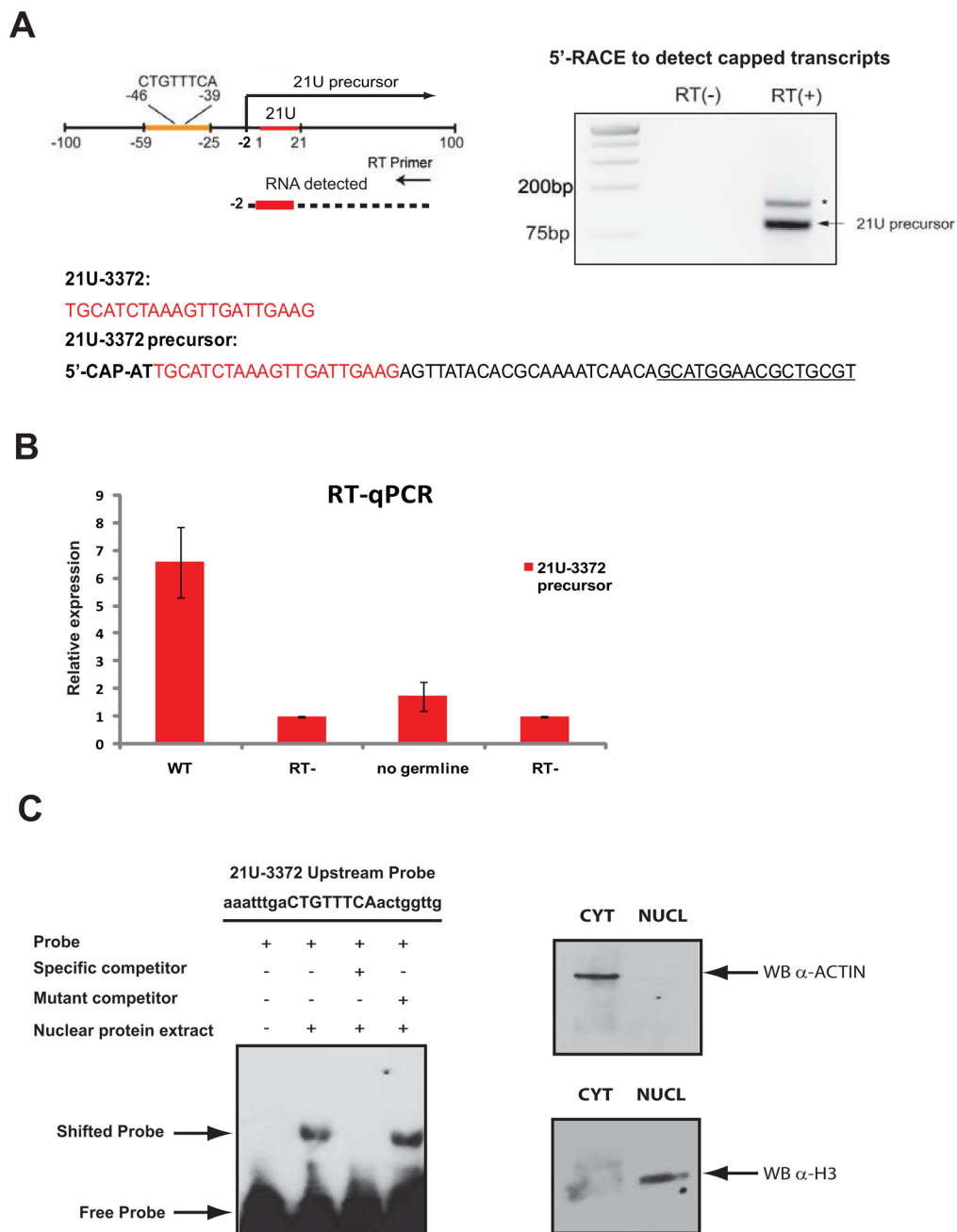


**Figure 3. RNA Polymerase II is localized to the nucleosome-depleted regions upstream of 21U-RNA-producing loci**

(A) Nucleosome occupancy per base pair averaged across all 21U-RNA loci and shown 500 nucleotides upstream and downstream of the first U in 21U-RNA is shown in blue and the Pol II occupancy per base pair averaged across all 21U-RNA loci is shown in red. (B) ChIP-qPCR with 8WG16 anti-Pol II antibody shows a significant Pol II enrichment over 21U-3016 and 21U-11909 loci in wild type adult worms, but not in *glp-4* mutant worms lacking a germline compared to the IgG control, see Figure S2 for additional control genes. Error bars represent standard deviation from the mean from three independent experiments.

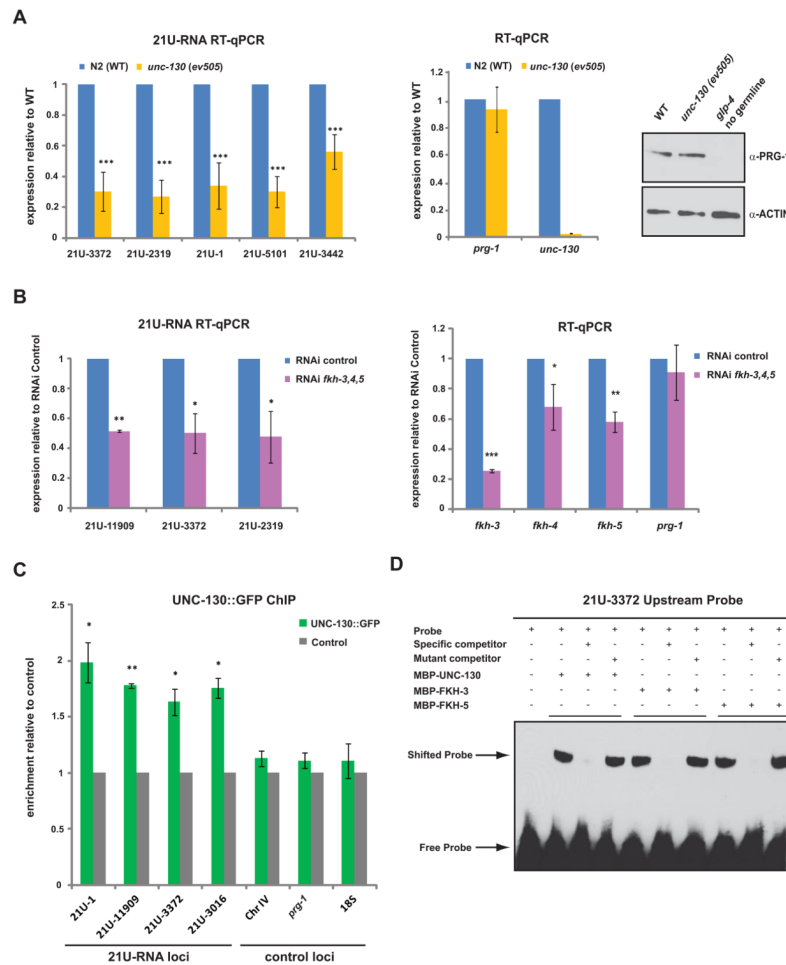


**Figure 4. The CTGTTTCA 8-mer sequence upstream 21U-3372 is required for its expression** (A) Schematic of a 5kb region on chromosome IV indicating: positions of 21U-RNAs (vertical bars with arrows showing direction of their expression), deletion niDf199 present in the JU258 strain, and fosmid WRM0611aH08 used for complementation of the deletion. Expression of 21U-RNAs colored in red and green has been detected by RT-qPCR (B, C); the CTGTTTCA motif upstream of 21U-3372 (red) has been deleted in C. Additional annotated 21U-RNAs overlapping with those shown in (A) are omitted for clarity. (B, C) Quantification of 21U-RNA expression levels by RT-qPCR in the wild type N2 strain (blue), the JU258 strain with niDf199 (red), the transgenic strain carrying a fosmid covering the region deleted in JU258 (green), and the transgenic strain containing the fosmid with a deletion of the 8nt conserved motif next to 21U-R-3372 (orange). RNA was extracted from adult worms. Expression levels shown are relative to the N2 strain. miR-52 expression was used as an internal control. Error bars represent standard deviation from the mean from three independent experiments. See Figure S3 for additional control experiments. (D) qPCR on MNase-digested chromatin of the transgenic strain carrying a fosmid covering the region deleted in JU258 (green) and the transgenic strain containing the fosmid with a deletion of the 8nt conserved motif next to 21UR- 3372 (orange). The schematic indicates primers used to detect the upstream or downstream regions of 21U-3372 and 21U-11909. Error bars represent standard deviation from the mean from two independent experiments. See Figure S3 for MNase-digested chromatin preparations.



**Figure 5. Transcription of the precursor for 21U-3372 starts 2bp upstream of the first T while the CTGTTTCA upstream sequence represents a DNA motif bound by nuclear proteins** (A) (left panel) Schematic representation of a 21U-3372 locus showing the position of the RT primer used for the detection of a precursor transcript and the transcription start site 2bp upstream of the 21U-RNA sequence. (right panel) Results of a 5' RACE experiment showing the detection of a band corresponding to a 21U-RNA precursor on an agarose gel. \* Indicates non-specific amplification. The sequence of 21U-3372 (red) and its precursor are shown. See Figure S4 for additional 5' RACE experiments that suggest a lack of transcription through the upstream motif. (B) Quantification by RT-qPCR of the level of 21U-3372 precursor expression in young adult animals with germline (WT) and lacking germline, *glp-4* (no germline). Error bars represent standard deviation from the mean from

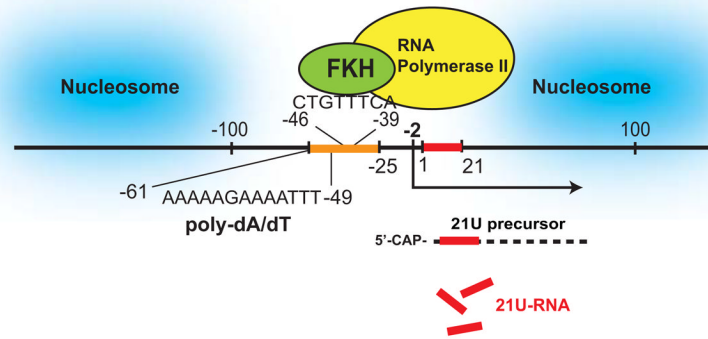
two independent experiments. (C) Electrophoretic mobility gel shift assay (EMSA) with nuclear protein extract from wild type adult worms and a 26bp dsDNA biotinylated oligonucleotide probe surrounding the 8-mer CTGTTTCA motif upstream of 21U-3372. The specificity of the binding was determined by incubating the reaction with 200x excess of non-biotinylated 26bp probe as a specific competitor or with 200x non-biotinylated 26bp probe with a mutated 8-mer motif (CGCCCGCA). Western blots on the right using anti-histone H3 and anti-actin antibodies show the nuclear (NUCL) and cytoplasmic (CYT) protein fractions prepared for EMSA. As expected, histone H3 is enriched in the nuclear fraction and actin in the cytoplasmic fraction. See Figure S5 for additional EMSA control experiments.



**Figure 6. Germline-enriched Forkhead proteins bind to the CTGTTTCA 8-mer sequence and promote 21U-RNA production**

(A, B) Quantification of the expression of 21U-RNAs by RT-qPCR in the *unc-130(ev505)* mutant (A, left panel) relative to WT, and upon simultaneous inactivation by RNAi of *fkh-3*, *fkh-4* and *fkh-5* (B, left panel) in the RNAi-sensitized *eri-1* mutant background relative to control RNAi. RNA was extracted from L4 worms (A) or by picking 10 individual adult worms (B). The expression of 21U-RNAs was normalized to miR-52. Error bars represent standard deviation from the mean from at least three independent experiments and the number of asterisks indicates the p values as follows: \* < 0.05, \*\* < 0.001, \*\*\* < 0.0001. mRNA quantification by RT-qPCR for *prg-1* and *unc-130* (A, right panel) or *prg-1* and *fkh-3*, *fkh-4* and *fkh-5* (B, right panel) is shown for the experiments presented in A and B (left panels). *act-3* mRNA was used as an internal control. Error bars represent standard deviation from the mean from at least two biological replicates. In addition, PRG-1 protein levels in WT, *unc-130 (ev505)* and *glp-4* (no germline) worms are shown by western (A, right panel). Anti-actin western was used as a loading control. See Figure S6 for 21U-RNA expression quantification in *daf-16*, *lin-31* and *fkh-9* mutants or *pha-4(RNAi)*. (C) ChIP-qPCR detects specific enrichment of UNC-130 at the 21U-RNA loci. ChIP-qPCR was performed using anti-GFP antibodies in transgenic strains generated by bombardment and expressing UNC-130::GFP from the germline promoter *mex-5* (green bars). A transgenic line with no UNC-130::GFP germline expression was used as control (grey bars). The enrichment of UNC-130::GFP at different 21U-RNA loci was calculated relative to the control line. Three chromosomal regions were used as controls: a chromosome IV region

located between individual 21U-RNA genes, the promoter region of the *prg-1* gene and the coding sequence of ribosomal RNA (18S). Error bars represent standard deviation from the mean from three biological replicates and asterisk indicates p values < 0.05. Similar results were obtained with a MosSCI transgenic UNC-130::GFP line, see Figure S6. (D) Electrophoretic mobility gel shift assays (EMSA) with immunopurified bacterial recombinant FKH proteins fused with N-terminal Maltose Binding Protein (MBP) and a 26bp dsDNA biotinylated oligonucleotide probe surrounding the 8-mer CTGTTTCA motif upstream of 21U-3372. The specificity of the binding was determined by incubating the reaction with 200x excess non-biotinylated 26bp probe as a specific competitor or with 200x non-biotinylated 26bp probe with a mutated 8-mer motif (CGCCCGCA). See Figure S5 for additional EMSA control experiments.



**Figure 7. Nucleosome-depleted promoters recognized by Forkhead proteins exist for individual 21U-RNAs**  
Schematic model of the 21U-RNA promoter.