A bioinformatics search pipeline, RNA2DSearch, identifies RNA localization elements in *Drosophila* retrotransposons

RUSSELL S. HAMILTON,[1,2,5] EVE HARTSWOOD,[3,5] GEORGIA VENDRA,[2] CHERYL JONES,[3] VERONIQUE VAN DE BOR,[2,4] DAVID FINNEGAN,[3] and ILAN DAVIS[1,2]

[1]Department of Biochemistry, University of Oxford, South Parks Road, OX1 3QU, United Kingdom
[2]Wellcome Trust Centre for Cell Biology, School of Biological Sciences, Kings Buildings, University of Edinburgh, Edinburgh EH9 3JR, United Kingdom
[3]Institute of Cell Biology, School of Biological Sciences, Kings Buildings, University of Edinburgh, Edinburgh EH9 3JR, United Kingdom
[4]ISBDC UMR6543 CNRS, Centre de Biochimie, Parc Valrose, Faculté des Sciences, 06108 NICE cedex 2, France

ABSTRACT

mRNA localization is a widespread mode of delivering proteins to their site of function. The embryonic axes in *Drosophila* are determined in the oocyte, through Dynein-dependent transport of *gurken/TGF-α* mRNA, containing a small localization signal that assigns its destination. A signal with a similar secondary structure, but lacking significant sequence similarity, is present in the *I* factor retrotransposon mRNA, also transported by Dynein. It is currently unclear whether other mRNAs exist that are localized to the same site using similar signals. Moreover, searches for other genes containing similar elements have not been possible due to a lack of suitable bioinformatics methods for searches of secondary structure elements and the difficulty of experimentally testing all the possible candidates. We have developed a bioinformatics approach for searching across the genome for small RNA elements that are similar to the secondary structures of particular localization signals. We have uncovered 48 candidates, of which we were able to test 22 for their localization potential using injection assays for Dynein mediated RNA localization. We found that *G2* and *Jockey* transposons each contain a *gurken/I* factor-like RNA stem–loop required for Dynein-dependent localization to the anterior and dorso–anterior corner of the oocyte. We conclude that *I* factor, *G2*, and *Jockey* are members of a ''family'' of transposable elements sharing a *gurken*-like mRNA localization signal and Dynein-dependent mechanism of transport. The bioinformatics pipeline we have developed will have broader utility in fields where small RNA signals play important roles.

Keywords: transposable elements; intracellular RNA localization; bioinformatics; *Drosophila*; RNA secondary structure; Dynein

INTRODUCTION

mRNA localization is a commonly used mechanism for delivering proteins to their cytoplasmic site of function (St Johnston 2005; Muller et al. 2007). Examples are known from a wide range of eukaryotic systems. *cis*-Acting RNA signals of diverse length, sequence, and structure are recognized by *trans*-acting protein factors that in most cases facilitate the association of the RNA cargo with a molecular motor that transports it to its correct intracellular destination. Well-characterized examples range from *Ash1* mRNA in *Saccharomyces cerevisiae* (Chartrand et al. 2002) to *Vg1* mRNA localization in *Xenopus* (Mowry and Melton 1992; Deshler et al. 1997) and actin mRNA in chicken fibroblast cells (Kislauskis et al. 1994).

In *Drosophila*, mRNA localization and translational regulation play an essential role in determining the polarity of the oocyte and embryo. Localized mRNAs include *oskar (osk)* and *nanos (nos)* at the posterior, and *bicoid (bcd)* at the anterior, as well as *gurken* (*grk*) RNA, which encodes a TGFα signal, targeted to the dorso–anterior corner of the oocyte. Grk protein is required to establish both the anterior–posterior and dorso–ventral axes of the oocyte and embryo, by providing positional information to overlying follicle cells. *grk* RNA is synthesized in nurse cells, transported

[5]These authors contributed equally to this work.

Reprint requests to: Ilan Davis, Department of Biochemistry, University of Oxford, South Parks Road, OX1 3QU, UK; e-mail: Ilan.Davis@bioch.ox.ac.uk; fax: 44-0131-650-7031; or David Finnegan, Institute of Cell Biology, School of Biological Sciences, Kings Buildings, University of Edinburgh, Edinburgh EH9 3JR, United Kingdom; e-mail: David.Finnegan@ed.ac.uk; fax: 44-01865-613340.

Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.1264109.

to the oocyte along microtubules (MTs) (Caceres and Nilson 2005; Clark et al. 2007) and concentrated at the posterior, where it is translated. It is then transported to the anterior face of the oocyte, and then dorsally to form a cap between the nucleus and the adjacent follicle cells where it is again translated (MacDougall et al. 2003). The movement of *grk* RNA along MTs is mediated by Dynein and requires the *grk* localization signal (GLS), a small stem–loop within the coding region that is recognised by a number of proteins that together are presumed to specify the destination (Van De Bor et al. 2005).

The *I* factor, a non-LTR retrotransposon, encodes a transcript that is localized within the oocyte in a similar manner to *grk* mRNA. The *I* factor RNA contains a *cis*-acting RNA localization signal, the *I* factor localization signal (ILS), that is related to the GLS in secondary structure and competes for shared components of the localization machinery (Seleme et al. 2005; Van De Bor et al. 2005). Like *grk*, the *I* factor is transcribed in nurse cells (Seleme et al. 1999) and is transported into the oocyte, first to the posterior pole and then to the dorso–anterior corner, where it forms a cap overlying the nucleus.

Non-LTR retrotransposons are transcribed into a full-length RNA transposition intermediate that moves to the cytoplasm to be translated and then enters the nucleus, together with at least the reverse transcriptase and nuclease encoded by the second open reading frame. Once in the nucleus, non-LTR RNA is reverse transcribed into DNA that is integrated at a new site by a process called target primed reverse transcription (Cost et al. 2002). Transposition of the *I* factor takes place in the female germ-line and we have suggested that its RNA transposition intermediate is localized at the dorso–anterior corner of the oocyte to facilitate nuclear entry and transposition (Van De Bor et al. 2005). All non-LTR retrotransposons are believed to undergo target primed reverse transcription, raising the possibility that the transposition intermediates of other non-LTR elements in *Drosophila* could contain *grk*-like localization elements.

cis-Acting signals required for localization of other RNAs within the *Drosophila* oocyte have been mapped with varying degrees of precision. These include *bcd*, *fs(1)K10 (K10)*, *orb*, and *nos* (Macdonald and Struhl 1988; Kim-Ha et al. 1993; Macdonald et al. 1993; Serano and Cohen 1995; Gavis et al. 1996). With the exception of the *K10* and *orb* localization signals, that are very similar in sequence and structure (Cohen et al. 2005), no obvious similarities in either primary sequence or secondary structure have been found (Davis and Ish-Horowicz 1991; Francis-Lang et al. 1996; Simmonds et al. 2001; Bullock et al. 2003).

Although there is only 35% sequence identity between the GLS and ILS (Van De Bor et al. 2005), their secondary structures show striking similarities, as predicted by Mfold (Zuker 2000). This similarity suggests that for these localization signals, secondary structure may be of key importance for the interaction of *trans*-acting proteins with the mRNA cargos. Detection of localization signals that are functionally related to the GLS and ILS will therefore require the detection of sequences with the potential to form similar secondary structures.

Here, we describe a bioinformatics pipeline, which we have used to screen all the transposable elements identified in the genome sequence of *D. melanogaster* (release 5.1) for potential localization signals similar in secondary structure to the GLS and ILS. We have identified two elements, *G2* (Di Nocera 1988) and *Jockey* (Mizrokhi et al. 1988; Priimagi et al. 1988) that contain localization sequences that can direct the correct localization of RNA injected into stage 8/9 oocytes. We propose that such secondary structure motifs may play a role in the transposition of several non-LTR retrotransposons in *Drosophila* by allowing their RNA transposition intermediates to move to the nucleus of cells in which they transpose.

RESULTS AND DISCUSSION

RNA2DSearch

To assess whether the GLS and ILS share primary sequence similarity, we carried out a pairwise sequence alignment (Needleman and Wunsch 1970), using nucleotide substitution matrices. We found only limited sequence similarity between the two localization sequences (Fig. 1A). Moreover, BLAST (Altschul et al. 1990) searches in the *Drosophila* sequence databases using the GLS failed to find the ILS and vice versa. In contrast, secondary structure predictions of the GLS and ILS using Mfold (Zuker 2000), highlight obvious similarities at the secondary structural level (Fig. 1D). We also performed searches using RSEARCH (Klein and Eddy 2003), a stochastic context free grammar (SCFG) approach, which incorporates secondary structure information in the searches. While RSEARCH is a more powerful method than BLAST, it still fails to retrieve the ILS using the GLS as bait in the search and vice versa. This is because SCFGs are designed to find homologous RNAs and the GLS and ILS are not sufficiently similar in sequence and structure to be considered homologs; however, the secondary structure similarity leads us to believe there is a conserved motif required for binding the proteins of the localization machinery (Fig. 1A). This suggests the possibility of searching the genome for sequences that could be transcribed into RNA with a secondary structure motif similar to a consensus derived from the dorso–anterior localization signal of *grk* and the *I* factor. To achieve this goal we have developed a bioinformatics pipeline that we named RNA2DSearch (Fig. 2; see Materials and Methods), able to search for RNA secondary structure motifs on a genomic-wide scale and to compare them to a reference structure using several existing RNA bioinformatics resources.

FIGURE 1. Comparison of the GLS and ILS. (*A*) Sequence and 2D secondary similarity. The GLS and ILS sequences show a low sequence identity of 35% when aligned with the Needleman–Wunsch global pairwise alignment algorithm with nucleotide substitution matrices (Needleman and Wunsch 1970). A Dynalign suboptimal structure prediction and alignment shows similarity in the secondary structures (Mathews 2005). Experimental validation of the predicted RNA localization signals: gene structures and mapping of regions necessary and sufficient for localization. (*B*) *G2* structure showing position of the G2LS and the sequences tested experimentally. (*C*) *Jockey* structure showing the JLS position and the sequences tested experimentally. The green indicates localizing fragments, and red non-localizing. Comparison of the localizing hits found by RNA2DSearch to the GLS and ILS. (*D*) The predicted secondary structures, using Mfold (Zuker 2000), show the GLS and ILS share similar structures. The G2LS and JLS show a high degree of secondary structure similarity to the GLS and ILS. (*A*,*D*) Sequences are colored by secondary structure (red:loop; green:internal loop; orange:bulge).

Identification of two additional non-LTRs RNA localization signals

As the localization signal within *I* factor RNA may play a role in transpositon we wanted to determine whether other transposable elements (TEs) contain related localization signals. Release 5.1 of the *D. melanogaster* genome sequence contains 6003 copies of 155 unique transposable elements, containing over 7.5×10^6 nucleotides (www.flybase.org). Each TE family is likely to include both active and inactive members, and these cannot be distinguished a priori, making it impractical to test RNA from individual members of each family for their localization ability. In order to identify likely candidates to test we screened the entire set of TE sequences using RNA2DSearch. The extent of the transcribed region of most transposable elements has yet to be determined, although transcripts of non-LTR retrotransposons cover the complete sequence of the element. All the hits that we tested come from the transcribed region of the transposable element concerned, and correspond to the sequence of the transcribed strand.

Predicted stable stem–loop structures were ranked according to their similarity to the ILS and GLS as indicated by the "full" structure similarity score assigned by RNAdistance (Table 1; Hofacker 2003). Elements with an RNAdistance score of 11 or less, considering only those with the lowest score in the case of elements with two or more candidates. We identified 48 elements matching these criteria, and of these we tested the 22 for which we could obtain a BAC clone containing genomic DNA of the particular element concerned. Testing the other elements was not practical, as PCR strategies cannot readily amplify the unique copies of a transposable element from a genomic sequence containing highly related individual TE sequences. These comprised 11 non-LTR retrotransposons, eight LTR retrotransposons, and three transposons (cut and paste elements). We transcribed fluorescently labeled RNA from each using DNA containing the predicted stem–loops, and injected these into stage 8/9 oocytes. Two RNAs, one from the element *G2* (TE19713) and the other from *Jockey* (TE20045) localized with a pattern comparable to that of

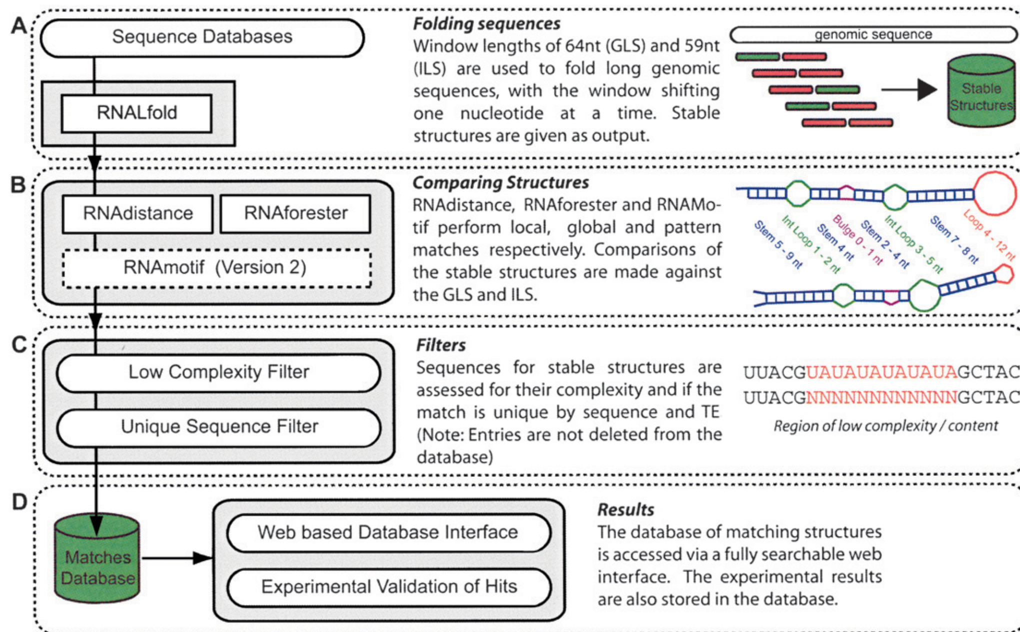


FIGURE 2. Flow diagram summarizing the RNA2DSearch pipeline. (*A*) Transposable element sequences are folded and collate using RNALfold. (*B*) Secondary structures of the stable structures are compared against known localizing signals GLS and ILS using RNAdistance (Hofacker 2003), RNAforester (Hochsmann et al. 2003) and RNAMotif (Macke et al. 2001) (Version 2). (*C*) Sequence analysis steps such as measuring sequence complexity with Dust and ensuring the matches are non-redundant by sequence and transposable element. (*D*) Structures similar to the GLS and ILS are collated in a relational database and linked to a custom web based interface for searching and annotation by the user.

I factor and *grk* RNAs (Fig. 3A–D). We have named these stem–loop structures as G2LS, for *G2* localization signal and JLS for *Jockey* localization signal.

We tested whether the G2LS and JLS are responsible for the ability of *G2* and *Jockey* RNAs to localize by injecting a deletion series of RNAs flanking the putative localization signals (Fig. 1B,C). All RNAs containing the putative *G2* stem–loop localized and this was abolished if the stem–loop sequence was deleted. Furthermore, G2LSx1, an RNA containing the *G2* stem–loop alone also localized (Fig. 1B). We conclude that the G2LS is both necessary and sufficient for localization. Interestingly, an RNA (G2LSx2) that contains two tandem copies of the G2LS localizes more efficiently than RNA with a single copy (Fig. 1B). We obtained comparable results for *Jockey* fragments with and without the JLS demonstrating that the motif is also necessary for localization. However, the JLS alone is not sufficient for localization (Fig. 1C; see Materials and Methods).

The G2LS and JLS are transported within the oocyte by Dynein

We have previously shown that movement of RNA containing either the GLS or ILS to the dorso–anterior corner of the oocyte requires the major minus end directed microtubule (MT) motor, cytoplasmic Dynein (Van De Bor et al. 2005). To test whether RNAs containing the G2LS or JLS localize by a similar mechanism, we used hypomorphic alleles of *Dhc* as well as an inhibitory anti-Dhc monoclonal antibody (Sharp et al. 2000) that we have previously shown disrupts the localization of RNA containing the GLS or ILS (Van De Bor et al. 2005). Injection of anti-Dhc antibody into oocytes 10–15 min before injection of fluorescently labeled RNA containing either the G2LS or the JLS significantly reduced the movement of these RNAs to the dorso–anterior corner (Fig. 3I,K) indicating that Dynein is required for their localization. Control injections of similar concentrations of a rabbit pre-immune serum (Fig. 3F,K), or a control mouse ascites fluid (Sigma) (data not shown), did not inhibit localization (see Materials and Methods). We also found that the efficiency of localization of injected G2LS or JLS containing RNAs was reduced in oocytes of females *trans*-heterozygous for two hypomorphic Dynein heavy chain mutants ($Dhc^{6\text{-}6}/Dhc^{6\text{-}12}$) (Fig. 3G,J). Injections of the same RNAs into control oocytes from females heterozygous for either of the mutations ($Dhc^{6\text{-}6}/+$ or $/Dhc^{6\text{-}12}/+$) localized at similar efficiencies to wild-type egg chambers (Fig. 3J). We conclude that *G2* and *Jockey* localize within the oocyte by a similar Dynein-dependent mechanism to *grk* and *I* factor RNAs and require related localization machinery.

In total, we have identified three non-LTR retrotransposons (ILS, G2LS, and JLS) that contain a sequence that can direct the localization of RNA to the same site as *grk* mRNA within the oocyte. As part of their life cycle, non-LTR retrotransposons reverse transcribe a full length RNA

TABLE 1. Summary of the transposable elements tested from the RNA2DSearch predictions

TE/Gene	Loc	ILS: *D*/*F* GLS: *D*/*F*	Copies	Start	End	Source	Database link
Non-LTR retrotransposons							
I	*Yes*	***0.02**/0.00 0.22/0.67*	*24*	—	—	—	—
X	No	**0.07**/0.34 0.23/0.61	22	69572	71130	BACR13E23	[GenBank:AC011757]
G2	*Yes*	***0.08**/0.19 0.22/0.59*	*4*	*123495*	*124995*	*BACR06L03*	*[GenBank:AC130649]*
Cr1a	No	**0.10**/0.42 0.26/0.79	28	115010	113796	BACR02D22	[GenBank:AC007587]
Rt1c	No	**0.10**/0.27 0.22/0.72	4	148628	150237	BACR34K23	[GenBank:AC154045]
R1	No	0.22/0.54 **0.10**/0.34	12	46440	49954	BACR03I06	[GenBank:AC018631]
Rt1b	No	**0.11**/0.54 0.25/0.61	32	126808	129045	BACR03A06	[GenBank:AC012166]
BS	No	**0.11**/0.52 0.25/0.65	2	96620	96570	Primer	[GenBank:AC007208]
Doc2	No	**0.11**/0.43 0.21/0.63	2	49656	52095	BACR02D22	[GenBank:AC130649]
Baggins	No	**0.11**/0.52 0.23/0.78	8	142504	142556	Primer	[GenBank:AC007477]
Jockey	*Yes*	***0.11**/0.45 0.23/0.67*	*2*	*155*	*2218*	*Genomic*	*[NCBI:M38643]*
Rt1a	No	**0.11**/0.38 0.21/0.54	20	18606	20567	BACR43K14	[GenBank:AC008356]
LTR retrotransposons							
Invader 3	No	**0.06**/0.31 0.24.0.76	16	116229	114964	BACR03A06	[GenBank:AC012166]
Diver	No	0.09/0.50 0.23/0.75	4	119522	121097	BACR17L24	[GenBank:AL077782]
Blood	No	**0.10**/0.46 0.20/0.74	44	58430	56095	BACR17O24	[GenBank:AC007694]
1731	No	**0.10**/0.39 0.22.0.65	4	122424	122930	BACR36A03	[GenBank:AC016368]
Burdock	No	**0.11**/0.39 0.21/0.69	24	126939	128815	BACR26B05	[GenBank:AC154047]
Mdg1	Weak	0.22/0.55 **0.10**/0.29	38	4257	6550	Genomic	[NCBI:X59545]
297	No	**0.11**/0.41 0.21/0.69	12	32271	34534	BACR26B05	[GenBank:AC154047]
412	No	**0.11**/0.38 0.27/0.77	102	5041	6066	BACR17O24	[GenBank:AC007694]
Cut and paste							
S	No	**0.10**/0.44 0.22/0.67	18	22211	25103	BACR03I06	[GenBank:AC018631]
1360	No	**0.10**/0.39 0.24/0.61	14	121619	123455	BACR03A06	[GenBank:AC012166]
TC1	No	**0.10**/0.32 0.26/0.72	12	133651	135203	BACR15A11	[GenBank:AC009206]
gurken							
gurken	*Yes*	*0.24/0.68 **0.00**/0.00*	—	—	—	—	*[Flybase:CG17610]*

Each transposable element tested in the screen is listed. Beside each name are data about the test sequence: *Loc* records if the RNA localized in either an oocyte or embryo injection assay, secondary structure similarity score against the GLS and ILS using RNAdistance (*D*) (Hofacker 2003) and RNAforester (*F*) (Hochsmann et al. 2003); Copies indicate how many copies of the hits appear in the genome (Kaminker et al. 2002); Start and End positions define the region of the element tested; The source of the sequence is listed and an accession number identifies the published copy of the tested sequence. Italic text indicates RNAs localizing in oocytes. Bold text indicates the score used in the cutoff for selection of a hit in the RNA2Dsearch method.

at a genomic site. This is primed by a 3′OH at a single strand break in chromosomal DNA. We propose that the RNA localization signals that we have identified, allow the non-LTR transposons concerned to target their RNA to the host nucleus during transposition. A similar RNA nuclear localization signal would not be required for either LTR retrotransposons or transposons as the former reverse transcribe their RNA in the cytoplasm using a tRNA primer, while the latter move without an RNA intermediate. Indeed the potential localization signals that we identified in LTR retrotransposons or transposons did not direct localization of RNA in stage 8/9 oocytes (Table 1).

Having defined a total of four related signals (GLS, ILS, G2LS, and JLS) for localization to the dorso–anterior corner of the oocyte we are well placed to further define primary and secondary structure features that give them their ability to localization. This can be done using the G2LS and JLS to refine the search parameters of RNA2DSearch and this should allow us to discover other non-LTR transposable elements with similar modes of mRNA localization. We should also be able to more easily identify those features that are essential for their activity. Furthermore, the bioinformatics pipeline we have developed is highly flexible and at the same time rapid, thus allowing multiple searches of genomic databases for any relatively small RNA structural motif.

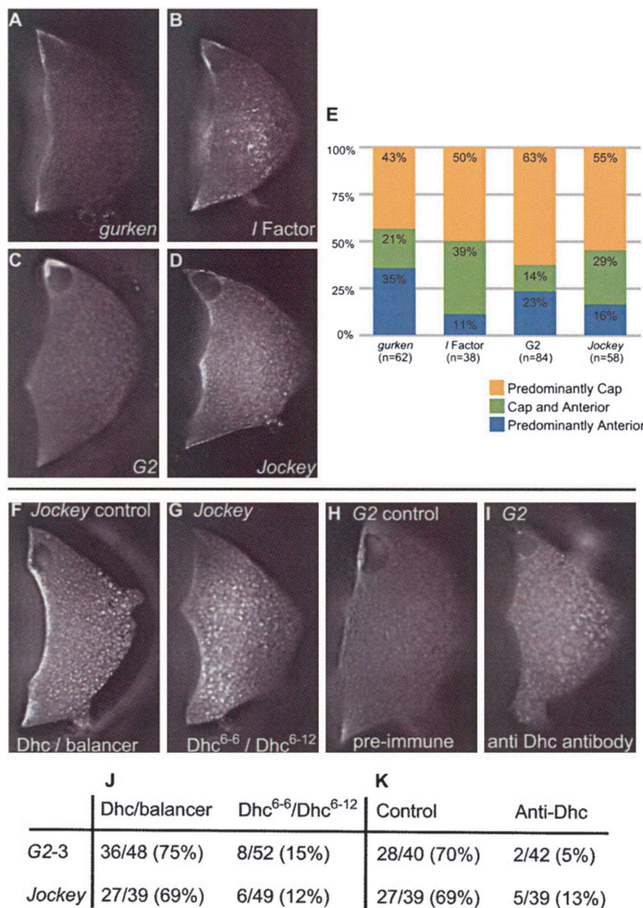
MATERIALS AND METHODS

Transposable element sequence databases

Sequences annotated as being transposable elements from FlyBase (Release 5.1; www.flybase.org) were used. Annotations are from Kaminker et al. (2002); supplementary materials were also utilized.

Folding sequences into stable secondary structures

A discussion of the bioinformatics approaches applicable to RNA motif searching can be found in Hamilton and Davis (2007).



	J		K	
	Dhc/balancer	Dhc^{6-6}/Dhc^{6-12}	Control	Anti-Dhc
G2-3	36/48 (75%)	8/52 (15%)	28/40 (70%)	2/42 (5%)
Jockey	27/39 (69%)	6/49 (12%)	27/39 (69%)	5/39 (13%)

FIGURE 3. A comparison of the localization patterns of RNA containing the GLS and ILS to the *Jockey* and *G2* hits found with RNA2DSearch. The localization patterns of labeled RNAs injected into stage 8/9 oocytes. Pictures were taken 40 min post injection. (*A*) *grk* RNA, full length transcript. (*B*) *I* factor RNA containing the ILS within a 1.7kb transcript. (*C*) G2 RNA: G2B (Fig. 1B) containing G2LS. In (*A–C*) the localized RNA can be seen forming a cap at the nucleus with some injected RNA remaining at the anterior cortex. (*D*) *Jockey* RNA J-D (Fig. 1C), containing the JLS. This RNA mainly localizes to the anterior cortex but can form a weak cap. (*E*) The proportion of oocytes showing a particular localization pattern after injection of RNA with *grk*, *I* factor, *G2* or *Jockey* localization motifs (usually within large fragments of 1.5–2.2 kb) is indicated with *n* at the base of each bar. *G2* and *Jockey* localization signals require Dynein for their localization. (*F*) Control, *Jockey* RNA injected into a *Dhc* / balancer oocyte. (*G*) *Jockey* RNA injected into *trans*-heterozygous hypomorphic *Dhc* mutant oocytes (Dhc^{6-6}/Dhc^{6-12}). (*H*) Control, *G2* RNA injected into an oocyte preinjected with a preimmune serum, showing no affect. (*I*) *G2* RNA injected into an oocyte that had been preinjected with an antibody raised against Dhc. (*J–K*) Fraction and percentage of RNA localization events over the total number of injected oocytes.

Locally stable structures of a defined length were identified by folding the input sequence, which could be an entire genome, in overlapping windows set at the defined length, progressing along the sequence, one nucleotide at a time using RNALfold (Hofacker et al. 2004). In this analysis window lengths of 64 and 59 nucleotides were selected, as these are the lengths of the GLS and ILS, respectively.

Comparing secondary structures to the reference structures

Each structure was compared to the reference structures, in this case the GLS and ILS. Separate comparisons were made using RNAdistance (Hofacker 2003) and RNAforester (Hochsmann et al. 2003). RNAdistance performs global structure comparisons, and in contrast, RNAforester was used to compare structures on a local scale, and find similarities between substructures. Both RNAdistance and RNAforester provide scores for the secondary structure similarities (scaled to be in the range 0–1), the lower the score, the more similar the structures being compared. RNA-distance scores of below 0.11, against either the GLS or ILS, were used as a cutoff for being considered as a hit. This threshold was chosen as it reflects the point at which the frequency of hits greatly increases. An additional step using RNAMotif (Macke et al. 2001) was included in a modification of the RNA2DSearch method. RNAMotif does not make direct comparisons to a reference structure, but instead uses descriptions files, based on the predicted secondary structures of the reference sequences (and primary sequences). The descriptions used incorporate all the main features common to both the GLS and ILS, accommodating small differences between the two structures (see Fig. 2B). The RNAMotif step was still able to predict the G2 and JLS, but produced less false positive hits.

Excluding low complexity sequences

The program Dust (Morgulis et al. 2006) was used to identify low complexity regions within sequences selected as having the potential to form secondary structures similar to the reference structure, the percentage of low complexity in each sequence was calculated so that these can be filtered out if desired (see Fig. 2C). A strict filter for removal of low complexity was used.

Availability of programs and supporting data

The code for RNA2DSearch is open source and freely available from the RNA2DSearch website (www.rna2dsearch.com). The code is provided as a set of Perl scripts and modules. In addition there must be working installations of RNAVienna, Mfold, and RNAMotif for RNA2DSearch to run. We have also provided the search results in raw text files for downloading, RNAMotif description files and full links to the Flybase TE database used.

Experimental testing of RNA sequences

The DNA templates used for in vitro transcription of labeled RNA were obtained in one of three ways. (1) Amplification from specific BAC clones containing the specific element insertion of interest (C.H.O.R.I. BAC Resource). (2) By direct amplification from the genome (*Jockey* and *Mdg1*). (3) Reannealing of appropriate oligonucleotides to create the motif (*BS* and *Baggins*) and subcloning this DNA into pGEM-T (Promega) for in vitro transcription. (See Table 1 for details). Amplification steps (1 and 2) incorporated a T7 primer for direct transcription from the appropriate strand (GTAATACGACTCACTATAGGGAGA). High fidelity polymerases Phusion (Finnzymes) or Platinum pfx (Invitrogen) were used for PCR amplification. Primers were designed to amplify between 800 and 1600 bases surrounding the hits of interest. The control *grk* RNA was full length as described in

Figure 3 of Neuman-Silberberg and Schupbach (1993). The *I* factor control RNA was a Kpn1 to Cla1 fragment of the *I* factor cloned into pBluescript and digested with BsmI prior to transcription.

Some potential localization motifs were tested in a less stringent assay, the ability to move to the apical side of nuclei at blastoderm, another Dynein-dependent process (Bullock and Ish-Horowicz 2001; Wilkie and Davis 2001). The stem–loop from the LTR retrotransposon *mdg1* localized weakly in this assay but failed to localize in the oocyte. We define a transcript to have not localized when it forming a diffuse signal throughout the ooplasm.

Fly strains

Flies were maintained on standard yeast–cornmeal–agar medium at 23°C. Localization of RNA was tested in embryos or oocytes from the wild-type Oregon R strain, Dhc $^{6-6}$ / *Dhc*$^{6-12}$*trans*-heterozygotes (a gift from T. Hayes, University of Minnesota, Minneapolis, MN).

Imaging.

Imaging and injection were performed on a widefield DeltaVision microscope (Applied Precision, Olympus lX70, and Roper Coolsnap HQ) using an Olympus 20x/0.75NA objective lens and a Burleigh micromanipulator from Scientifica.

Synthesis of fluorescently labeled, capped RNA

RNA was transcribed in vitro using T7 or Sp6 RNA polymerase (Wilkie and Davis 2001) labeled with cy3-UTP (Perkin Elmer). RNA was purified over miniquick RNA columns (Roche Applied Science), precipitated and resuspended in DEPC treated water to a minimum concentration of 200 ng/μL for injection.

Injection of RNA and antibodies

RNA was injected into either living *Drosophila melanogaster* blastoderm embryos or oocytes. Eggs were collected at 23°C and dechorionated using 50% bleach solution. Partially dehydrated eggs were injected at blastoderm (stages 4–5) in halocarbon oil 95. Oocytes were dissected in halocarbon oil 95 and separated into individual ovarioles directly onto coverslips as described in (Parton and Davis 2006). RNA or antibody was injected using sterile Femtotips (Eppendorf). The anti-Dhc antibody was kindly supplied by D. Sharp (Albert Einstein College of Medicine, New York) injected at 11 mg/mL. Either control mouse ascites fluid (Sigma M8273) or a rabbit preimmune serum were used as controls at equivalent concentrations to anti-Dhc. Antibody was preinjected into oocytes 10–15 min prior to the injection of RNA; all injection steps were completed within 45 min of oocyte dissection.

ACKNOWLEDGMENTS

R.S.H. was supported by a Medical Research Council Career Establishment Grant and an eDIKT grant from the Science and Technologies Facilities Council (STFC) (to I.D.). E.H. and C.J. are supported by a BBSRC project grant (to D.F. and I.D.). G.V. was supported by a studentship from the Greek Government. V.V.D.B. was supported by an HFSP postdoctoral fellowship. I.D. was supported by a Wellcome Trust Senior Research Fellowship (067413). We thank Ian Moss who provided data on *G2* variants as part of his Honours Project and Jim Brodie for help testing the *F* element.

Received July 10, 2008; accepted November 12, 2008.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Bullock, S.L. and Ish-Horowicz, D. 2001. Conserved signals and machinery for RNA transport in *Drosophila* oogenesis and embryogenesis. *Nature* **414:** 611–616.

Bullock, S.L., Zicha, D., and Ish-Horowicz, D. 2003. The *Drosophila* hairy RNA localization signal modulates the kinetics of cytoplasmic mRNA transport. *EMBO J.* **22:** 2484–2494.

Caceres, L. and Nilson, L.A. 2005. Production of gurken in the nurse cells is sufficient for axis determination in the *Drosophila* oocyte. *Development* **132:** 2345–2353.

Chartrand, P., Meng, X.H., Huttelmaier, S., Donato, D., and Singer, R.H. 2002. Asymmetric sorting of ash1p in yeast results from inhibition of translation by localization elements in the mRNA. *Mol. Cell* **10:** 1319–1330.

Clark, A., Meignin, C., and Davis, I. 2007. A Dynein-dependent shortcut rapidly delivers axis determination transcripts into the *Drosophila* oocyte. *Development* **134:** 1955–1965.

Cohen, R.S., Zhang, S., and Dollar, G.L. 2005. The positional, structural, and sequence requirements of the *Drosophila* TLS RNA localization element. *RNA* **11:** 1017–1029.

Cost, G.J., Feng, Q., Jacquier, A., and Boeke, J.D. 2002. Human L1 element target-primed reverse transcription in vitro. *EMBO J.* **21:** 5899–5910.

Davis, I. and Ish-Horowicz, D. 1991. Apical localization of pair-rule transcripts requires 3′ sequences and limits protein diffusion in the *Drosophila* blastoderm embryo. *Cell* **67:** 927–940.

Deshler, J.O., Highett, M.I., and Schnapp, B.J. 1997. Localization of *Xenopus* Vg1 mRNA by Vera protein and the endoplasmic reticulum. *Science* **276:** 1128–1131.

Di Nocera, P.P. 1988. Close relationship between non-viral retroposons in *Drosophila* melanogaster. *Nucleic Acids Res.* **16:** 4041–4052.

Francis-Lang, H., Davis, I., and Ish-Horowicz, D. 1996. Asymmetric localization of *Drosophila* pair-rule transcripts from displaced nuclei: Evidence for directional nuclear export. *EMBO J.* **15:** 640–649.

Gavis, E.R., Lunsford, L., Bergsten, S.E., and Lehmann, R. 1996. A conserved 90 nucleotide element mediates translational repression of nanos RNA. *Development* **122:** 2791–2800.

Hamilton, R.S. and Davis, I. 2007. RNA localization signals: Deciphering the message with bioinformatics. *Semin. Cell Dev. Biol.* **18:** 178–185.

Hochsmann, M., Toller, T., Giegerich, R., and Kurtz, S. 2003. Local similarity in RNA secondary structures. *Proc. IEEE Comput. Soc. Bioinform. Conf.* **2:** 159–168.

Hofacker, I.L. 2003. Vienna RNA secondary structure server. *Nucleic Acids Res.* **31:** 3429–3431.

Hofacker, I.L., Priwitzer, B., and Stadler, P.F. 2004. Prediction of locally stable RNA secondary structures for genome-wide surveys. *Bioinformatics* **20:** 186–190.

Kaminker, J.S., Bergman, C.M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D.A., Lewis, S.E., Rubin, G.M., et al. 2002. The transposable elements of the *Drosophila* melanogaster euchromatin: A genomics perspective. *Genome Biol.* **3:** 0084.1–0084.20.

Kim-Ha, J., Webster, P.J., Smith, J.L., and Macdonald, P.M. 1993. Multiple RNA regulatory elements mediate distinct steps in localization of oskar mRNA. *Development* **119:** 169–178.

Kislauskis, E.H., Zhu, X., and Singer, R.H. 1994. Sequences responsible for intracellular localization of beta-actin messenger RNA also affect cell phenotype. *J. Cell Biol.* **127:** 441–451.

Klein, R.J. and Eddy, S.R. 2003. RSEARCH: Finding homologs of single structured RNA sequences. *BMC Bioinformatics* **4:** 44. doi: 10.1186/1471-2105-4-44.

Macdonald, P.M., Kerr, K., Smith, J.L., and Leask, A. 1993. RNA regulatory element BLE1 directs the early steps of bicoid mRNA localization. *Development* **118:** 1233–1243.

Macdonald, P.M. and Struhl, G. 1988. *cis*-Acting sequences responsible for anterior localization of bicoid mRNA in *Drosophila* embryos. *Nature* **336:** 595–598.

MacDougall, N., Clark, A., MacDougall, E., and Davis, I. 2003. *Drosophila* gurken (TGFα) mRNA localizes as particles that move within the oocyte in two Dynein-dependent steps. *Dev. Cell* **4:** 307–319.

Macke, T.J., Ecker, D.J., Gutell, R.R., Gautheret, D., Case, D.A., and Sampath, R. 2001. RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Res.* **29:** 4724–4735.

Mathews, D.H. 2005. Predicting a set of minimal free energy RNA secondary structures common to two sequences. *Bioinformatics* **21:** 2246–2253.

Mizrokhi, L.J., Georgieva, S.G., and Ilyin, Y.V. 1988. Jockey, a mobile *Drosophila* element similar to mammalian LINEs, is transcribed from the internal promoter by RNA polymerase II. *Cell* **54:** 685–691.

Morgulis, A., Gertz, E.M., Schaffer, A.A., and Agarwala, R. 2006. WindowMasker: Window-based masker for sequenced genomes. *Bioinformatics* **22:** 134–141.

Mowry, K.L. and Melton, D.A. 1992. Vegetal messenger RNA localization directed by a 340-nt RNA sequence element in *Xenopus* oocytes. *Science* **255:** 991–994.

Muller, M., Heuck, A., and Niessing, D. 2007. Directional mRNA transport in eukaryotes: Lessons from yeast. *Cell. Mol. Life Sci.* **64:** 171–180.

Needleman, S.B. and Wunsch, C.D. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48:** 443–453.

Neuman-Silberberg, F.S. and Schupbach, T. 1993. The *Drosophila* dorsoventral patterning gene gurken produces a dorsally localized RNA and encodes a TGFα-like protein. *Cell* **75:** 165–174.

Parton, R. and Davis, I. 2006. Lifting the fog: Image restoration by deconvolution. In *Cell biology* (ed. J.E. Celis), pp. 187–200. Academic, New York.

Priimagi, A.F., Mizrokhi, L.J., and Ilyin, Y.V. 1988. The *Drosophila* mobile element jockey belongs to LINEs and contains coding sequences homologous to some retroviral proteins. *Gene* **70:** 253–262.

Seleme, M.C., Busseau, I., Malinsky, S., Bucheton, A., and Teninges, D. 1999. High-frequency retrotransposition of a marked I factor in *Drosophila* melanogaster correlates with a dynamic expression pattern of the ORF1 protein in the cytoplasm of oocytes. *Genetics* **151:** 761–771.

Seleme, M.C., Disson, O., Robin, S., Brun, C., Teninges, D., and Bucheton, A. 2005. In vivo RNA localization of I factor, a non-LTR retrotransposon, requires a *cis*-acting signal in ORF2 and ORF1 protein. *Nucleic Acids Res.* **33:** 776–785.

Serano, T.L. and Cohen, R.S. 1995. A small predicted stem–loop structure mediates oocyte localization of *Drosophila* K10 mRNA. *Development* **121:** 3809–3818.

Sharp, D.J., Rogers, G.C., and Scholey, J.M. 2000. Cytoplasmic Dynein is required for poleward chromosome movement during mitosis in *Drosophila* embryos. *Nat. Cell Biol.* **2:** 922–930.

Simmonds, A.J., dosSantos, G., Livne-Bar, I., and Krause, H.M. 2001. Apical localization of wingless transcripts is required for wingless signaling. *Cell* **105:** 197–207.

St Johnston, D. 2005. Moving messages: The intracellular localization of mRNAs. *Nat. Rev. Mol. Cell Biol.* **6:** 363–375.

Van De Bor, V., Hartswood, E., Jones, C., Finnegan, D., and Davis, I. 2005. gurken and the I factor retrotransposon RNAs share common localization signals and machinery. *Dev. Cell* **9:** 51–62.

Wilkie, G.S. and Davis, I. 2001. *Drosophila* wingless and pair-rule transcripts localize apically by Dynein-mediated transport of RNA particles. *Cell* **105:** 209–219.

Zuker, M. 2000. Calculating nucleic acid secondary structure. *Curr. Opin. Struct. Biol.* **10:** 303–310.