Title
Identification and characterization of bud-localized mRNAs in Saccharomyces cerevisiae

Permalink
https://escholarship.org/uc/item/8wd1r2j9

Author
Jambhekar, Ashwini R

Publication Date
2005

Peer reviewed|Thesis/dissertation

Identification and Characterization of Bud-Localized mRNAs in *Saccharomyces cerevisiae*

by

Ashwini R. Jambhekar

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOCHEMISTRY

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Date   University Librarian

Degree Conferred:...

Dedication

For my parents

Acknowledgements

I would like to thank my advisor, Joseph DeRisi, for his support of and enthusiasm for this work, and for giving me the independence to puruse my research in the lab in any direction it took. His input at all stages of my work was invaluable. I am also grateful to members of the lab for their insightful comments on everything from design of experiments to polishing the final drafts of publications. My colleagues' eagerness to help in any way possible exceeded my expectations.

Kelly Shepard has been a wonderful friend and colleague. Her willingness to share information, reagents, and labor with me has made this project a special experience. Her eagerness to learn every detail of my work and to read innumerable drafts of manuscripts was remarkable. I am also grateful to Ron Vale for taking the time to follow my progress so closely and for never losing sight of the broader scientific context of this work. I also thank other members of the Vale lab for making me feel welcome in their lab.

I am indebted to Pete Takizawa and Kim McDermott for performing numerous gel-shift experiments and sharing many of their other results prior to publication.

Finally, I would like to thank the other members of my thesis committee, Anita Sil and Alan Frankel. Their interest in all aspects of my work and wonderful suggestions throughout the process have helped guide me towards the successful completion of my graduate studies.

Identification and Characterization of Bud-Localized mRNAs in Saccharomyces cerevisiae

Ashwini R. Jambhekar

Abstract

RNA localization is an important regulatory mechanism utilized by many cells and organisms for ensuring proper development. Although this process has been studied for many years in *Drosophila* and *Xenopus*, the single-celled yeast *S. cerevisiae* also provides examples of RNA transport. A core complex of 3 proteins—She2, She3, and Myo4—transports *ASH1* and *IST2* mRNAs to bud tips of growing cells by an actin-myosin based mechanism. In this work, 22 additional mRNAs have been identified as She-complex targets and have been shown to localize to bud tips *in vivo*. In contrast to other organisms, *S. cerevisiae* was found to contain localization elements in the coding regions of transported RNAs. Unbiased selection of localization elements from the open reading frames of several She-complex targets revealed a short single-stranded RNA motif containing a core CG dinucleotide which was essential for She2/3 binding and transport. Further analysis also revealed additional context-dependencies affecting the function of the motif. These results establish yeast as a model system for studying RNA transport, and provide a foundation for elucidating the mechanisms of motor-cargo interactions and defining the range of regulatory functions fulfilled by localization of mRNA. Furthermore, the identification of a consensus motif and accessory features regulating She-complex recognition suggests a general mechanism by which mRNAs can simultaneously mediate protein production and binding of regulatory factors.

v

Table of Contents

List of Tables

List of Figures

Appendix

Chapter 1:

Introduction

Multiple mechanisms exist to ensure that proteins are produced at the appropriate time and location within a cell or organism. Localization of mRNA is one post-transcriptional method of achieving proper spatial and temporal regulation. There are three general mechanisms for localizing RNA molecules within a cell: 1) localized stabilization of a transcript coupled with degradation at other positions, 2) capture and anchoring of diffusible RNAs at specific locations, or 3) direct transport of RNAs via motor proteins and cytoskeletal networks (1).

Direct transport of RNA is commonly utilized for generating cell asymmetry, and it has been most extensively studied in *Drosophila* and *Xenopus* oocytes and embryos, as well as in neurons. RNA localization in oocytes and embryos specifies body axes and defines body segments of the developing organisms. RNA transport in oligodendrocytes ensures that proteins involved in synaptic activity are produced in dendrites, far from the cell body where the RNAs are transcribed. In the case of myelin basic protein (MBP), RNA localization targets the protein (which itself is membrane-localized) exclusively to myelin membranes. The motor-protein complexes responsible for transport of the various RNAs do not overlap extensively, making it difficult to elucidate general mechanisms of RNA recognition and transport (2, 3).

The yeast *Saccharomyces cerevisiae* also utilizes motor-dependent RNA transport to generate cell asymmetry. Although the progeny of a mitotic division in yeast are often considered identical, differences in cell size, length of the subsequent cell cycle, distribution of damaged proteins, and gene expression programs are observed between "mother" and "daughter" cells (4-6). One difference in gene expression patterns is the

transcription of *HO* exclusively in mother cells (7); this asymmetry is achieved by myosin-dependent transport of *ASH1* mRNA to bud tips prior to cell division (8, 9). Subsequently, Ash1, a transcriptional repressor of *HO*, accumulates in daughter nuclei, leading to *HO* expression only in mother cells (10, 11).

A minimal protein complex responsible for *ASH1* RNA transport was elucidated (see below) (12-14)and a second transported RNA, *IST2*, was identified based on its association with the complex (15). Both *ASH1* and *IST2* depend upon RNA transport for segregating the newly-synthesized protein products to daughter cells. The identification of a minimal complex responsible for transporting more than one RNA to a single location provides a model system for studying trans-acting factors which contribute to efficient transport and the cis-acting RNA elements recognized by the transport complex (shown in Fig. 1). In addition, the facile genetic analyses possible in yeast also provide an opportunity to investigate the physiological purpose of RNA transport in a single-celled organism.

Trans-acting factors:

The trans-acting factors responsible for transport of *ASH1* into the bud were initially identified in a genetic screen for mutants defective in *HO* regulation (16). This screen isolated a myosin motor (Myo4), two proteins involved in regulating the actin cytoskeletion (She4 and She5), and two other previously uncharacterized proteins (She2 and She3). Further biochemical studies revealed that Myo4, She2, and She3 are part of a

complex that includes RNA(12-14), while She4 and She5 play more indirect roles. None of the transport complex proteins was essential for survival.

She2: She2 is a novel protein that displays no homology to other known proteins outside of the yeast family, and is hypothesized to bind RNA. She2 exists in both nuclear and cytoplasmic pools, and association with RNA is necessary for its nuclear export (17). Like hnRNPI in *Xenopus* (18), She2 may "mark" RNAs for cytoplasmic localization prior to nuclear export. Based on early coimmunoprecipitation studies, She2 was hypothesized to bind RNA directly (12-14). *ASH1* RNA coimmunoprecipated with She2 in *myo4Δ*, *she3Δ*, or *she4Δ* strains, and She2 was required for *ASH1* to coimmunoprecipitate with Myo4 and She3 (14, 19). Nevertheless, the protein displays no characterized RNA-binding motifs. The recently-solved crystal structure (20) revealed a dimeric, gobular protein consisting of 5 α-helices; the structure did not reveal any known structural homologs or domains. Mutation of residues at the dimer interface disrupted binding to RNA in an *in vitro* filter-binding assay (20), suggesting that the protein binds RNA in dimeric form. Additional lysine and arginine residues were shown to be essential for binding *in vitro* and in the 3-Hybrid assay *in vivo* (20, 21), but identification of these residues has been insufficient to define an RNA-binding site. Furthermore, residues that confer specificity to the binding interaction have not been elucidated. Ectopic linkage of RNA to She2 is not sufficient to direct its localization (22), suggesting that precise RNA-protein contacts are necessary for successful transport.

4

In contrast to the early coimmunoprecipitation studies, some recent evidence suggests that She2 is not sufficient for binding RNA. The purified protein displays low (micromolar) affinity for RNA *in vitro* by gel-shift assay, and the binding is largely non-specific (23). One group reported that She2 facilitated the formation of RNA dimers (21), suggesting that She2 might play an indirect role as a chaperone in promoting the proper folding of RNA; other groups, however, did not detect RNA dimers (23). Addition of She3 to RNA and She2 has been shown to increase the affinity of She2 for RNA and to create larger molecular weight complexes (12), raising the possibility that a heteromeric complex of She2 and She3 is the active RNA binding species. Given these results, it is not clear why She2 binds to RNA *in vivo* in *she3*Δ strains, but requires She3 for binding *in vitro*.

She3: Like She2, She3 has no known homologs outside of the yeast species. The N-terminus of She3 interacts with the tail of Myo4 *in vitro* and *in vivo* via a coiled-coil domain, while the C-terminus interacts with She2 (12, 13). Although ectopic linkage of RNA to She2 cannot mediate bud-localization of RNA, a similar linkage to She3 can bypass the requirement for She2 for RNA localization (13). Based on these studies, She3 was proposed to link the She2-RNA complex to Myo4 (12-14), although recent evidence suggests that She3 may play a more direct role in RNA binding. She3 and Myo4 colocalize to bud tips in the absence of She2 (albeit at lower levels than in WT strains) (16, 17), and both She3 and Myo4 (but not She2) play a role in transporting cortical ER into buds (24).

5

Myo4: Myo4 is a Type V unconventional myosin motor. To date, its only known cargoes are RNA and cortical ER (24).

Other trans-acting factors: The remaining trans-acting factors fall into three categories: 1) those that regulate the actin cytoskeleton, 2) translation regulators, and 3) nuclear proteins.

Cytoskeletal regulators: She4 and She5 were the first two cytoskeletal regulators implicated in RNA localization, but unlike the other She-proteins, they do not coimmunoprecipitate with *ASH1* RNA (14). She4 is required for proper distribution of actin patches and interacts with the motor domains of various myosins, facilitating their binding to the actin cytoskeletion (25, 26). She5 is bud-localized and required for proper polarization of the actin cytoskeletion (16, 27, 28). It interacts with Bud6, which also regulates actin polarization and was shown by targeted mutation to play a role in RNA localization. Deletion of She5 or Bud6 does not affect transport of RNA into the bud, but impairs anchoring of the RNA to the bud tip (29).

Recently, Arf1, which is involved in formation of COPI vesicles, was also shown to bind poly-A binding protein (Pab1) and facilitate *ASH1* localization (30). Some alleles of *ARF1* cause defects in the actin cytoskeletion and mislocalize *ASH1* indirectly, while other alleles indicated that Arf1 plays a more direct role in *ASH1* localization. Because

Arf1 also interacts with some translation regulators, it is possible that RNA localization defects in *arf1* mutants arise from translation misregulation.

Translation regulators: Translation repressors Mpt5, Puf6, and Khd1 play a minor role in directing *ASH1* localization (31-33): deletion of any of these proteins reduces localization efficiency by 2- to 3-fold. Puf6 coimmunoprecipitates with the She-complex (31) and Khd1 colocalizes with the complex *in vivo* (32), implying that both proteins function directly in the localization process. Translation of *ASH1* is necessary for proper anchoring of the RNA to the bud tip (34); accordingly, overexpression of Khd1 decreases anchoring efficiency, presumably by repressing translation (32). In contrast, the localization defect of a *puf6*Δ mutant (which overexpresses Ash1) can be partially suppressed by repressing translation with ectopic RNA stem-loops that provide a barrier to ribosome passage (31). It is possible that translation rates must be properly modulated to achieve efficient RNA transport. Whether the translation repressors play distinct roles in both RNA localization and translation, or whether the localization defects stem from misregulation of translation in the mutants, remains to be determined.

Nuclear factors: To date, the only other trans-acting factor involved in RNA localization is Loc1 (35). The extent of RNA delocalization is greater in the *loc1*Δ mutant than in any of the translational repressor mutants described above. Loc1 binds sequences from *ASH1* as well as other double-stranded RNAs, suggesting that its specificity for RNA is more degenerate than that of the core She-complex. Unlike the other trans-acting factors which

are present in the cytoplasm, Loc1 is exclusively nuclear. Along with She2, it may serve to "mark" RNAs for bud-localization prior to nuclear export.

Given the existence of a protein complex devoted solely to RNA transport, as well as additional mechanisms for subtle regulation of transport and anchoring, it was quite surprising that only two transported mRNAs had been identified. In this work (see Ch. 2), twenty-two additional transported mRNAs have been identified by immunoprecipitation of core She-complex proteins and detection of bound RNAs by microarray analysis, followed by visualization of candidate RNA distribution in live cells. Many of the proteins encoded by the localized RNAs are involved in cell wall and cell membrane regulation. Unlike Ash1 and Ist2, none of the other protein products encoded by the localized transcripts requires RNA transport for accurate protein localization, and some RNAs encode proteins that are not bud-localized.

Despite the existence of at least 24 mRNAs localized by the She-complex, mutations in the complex have no reported phenotypes other than mating-type switching defects. Because many of the proteins encoded by the localized RNAs are properly localized in the absence of RNA transport, it is apparent that the RNA transport system functions redundantly with other, unidentified localization systems which act at the protein level. To elucidate any redundancies, a genetic screen was carried out to identify non-essential genes that showed synthetic lethality in combination with mutations in the She-complex (see Appendix). Although none of the candidates generated by the screen showed genetic interactions with the She-complex upon retesting, many of the candidates

8

were involved in various aspects of budding, mother-daughter differentiation, and maintenance of homeostasis, suggesting that localization of some transcripts is essential for regulating cell fate and/ or for maintaining cell wall and membrane structure under conditions used for the initial screen. Additionally, it appears that mutations in the She-complex cause subtle survival defects that are apparent only in some mutant backgrounds under sub-optimal growth conditions.

Cis-acting elements:

Cis-acting elements within transported RNAs function in concert with trans-acting factors to mediate localization. In most systems, one or more fragments of the localized RNA are recognized and bound by the transport complex; these RNA localization elements, or "zipcodes," are usually able to function out of context. Traditionally, zipcodes have been identified by deletion mapping. Fragments of the localized RNA are assessed, by direct visualization of RNA distribution *in vivo*, for their ability to localize a reporter RNA (e.g. (36)); or, less frequently, to complement in *cis* the phenotype of a localization-defective RNA. In many systems, the specific RNA-binding protein of the transport complex remains unknown; therefore, the exact zipcode bases recognized by the complex cannot be determined by *in vitro* binding assays. A complementary approach to deletion mapping involves searching for sequence or structural similarities between RNAs transported to the same location or by the same motor complex, and subsequent verification of the candidate features by mutational analysis (37). The former approach generally identifies sequences sufficient for

9

localization, but the importance of each base within the element is not assessed. The latter approach identifies necessary features which are often not sufficient for transport (38).

Zipcodes have been identified in various RNAs, yet it has been difficult to elucidate general principles governing RNA recognition by the transport complex. The known zipcodes range from 50 to several hundred nt in length, and are usually found in 3' UTRs. In most cases, a combination of both sequence and structural features are necessary for directing localization. Many zipcodes function redundantly (39) or synergistically(40) in directing localization and some are even bipartite (41), thus obscuring the exact features recognized by the transport complex. To complicate the situation, some zipcodes (e.g. the 3' UTR of *oskar*) require endogenous, full-length RNA for localization as well as successful completion of other processing events such as splicing (42). Some zipcodes also overlap with sequences responsible for translation regulation, and the two functions sometimes cannot be separated (43). Furthermore, each localized RNA within a cell often employs a specialized transport complex, making it difficult to identify critical zipcode features by comparison.

In yeast, zipcodes have been identified in *ASH1* based on their ability to localize a reporter RNA. Gonzalez et al. identified three zipcodes: N, C, and U. N and C lie within the coding regions, while U overlaps the last 7 bases of the ORF and the following 70 bases of the 3'UTR (34). Independently, Chartrand et al. mapped four zipcodes in ASH1, termed E1, E2A, E2B, and E3. E1 lies within zipcode N, and E2A and B are within C. E3 overlaps with U but contains additional bases at the 5' and 3' ends (44). These zipcodes localized efficiently even when full-length *ASH1* RNA is not experessed, indicating that

the zipodes are not transported by a "piggy-back" mechanism. Furthermore, mRNA

processing was not required for transport, as zipcodes under the control of a PolIII

promoter also localized efficiently (45). Despite the evidence that the ASH1 zipcodes

were directly recognized by the She-complex, no extensive sequence or structural

similarities were observed between the zipcodes. Disruption of base-pairing, however,

abolished localization. These results led to the initial hypothesis that She2 was a double-

stranded RNA binding protein (34, 44).

The identification of 22 additional bud-localized mRNAs in this work provided a

model system to study zipcode recognition. This set of RNAs represents the largest

collection of sequences known to be localized by a single core transport complex;

therefore numerous zipcodes could be isolated and key determinants identified by

comparative analysis. A degenerate 7-base, single-stranded motif has been identified

which is essential for localization and conserved in all She-complex targets known to

date. In addition, non-conserved sequences outside of the motif also facilitated She-

complex recognition, suggesting that complex RNA features are necessary for zipcode

activity (see Ch. 3).

Figure 1: Schematic of subcellular localizations of and interactions between trans-acting factors and cis-acting elements responsible for *ASH1* mRNA localization. Only 2 (N and U) of the 4 zipcodes in *ASH1* are shown.

References:

1. Chartrand, P., Singer, R. H. & Long, R. M. (2001) *Annu Rev Cell Dev Biol* **17**, 297-310.

2. Johnstone, O. & Lasko, P. (2001) *Annu Rev Genet* **35**, 365-406.

3. Smith, R. (2004) *Neuroscientist* **10**, 495-500.

4. Aguilaniu, H., Gustafsson, L., Rigoulet, M. & Nystrom, T. (2003) *Science* **299**, 1751-3.

5. Colman-Lerner, A., Chin, T. E. & Brent, R. (2001) *Cell* **107**, 739-50.

6. Laabs, T. L., Markwardt, D. D., Slattery, M. G., Newcomb, L. L., Stillman, D. J. & Heideman, W. (2003) *Proc Natl Acad Sci U S A* **100**, 10275-80.

7. Nasmyth, K. (1983) *Nature* **302**, 670-6.

8. Takizawa, P. A., Sil, A., Swedlow, J. R., Herskowitz, I. & Vale, R. D. (1997) *Nature* **389**, 90-3.

9. Long, R. M., Singer, R. H., Meng, X., Gonzalez, I., Nasmyth, K. & Jansen, R. P. (1997) *Science* **277**, 383-7.

10. Bobola, N., Jansen, R. P., Shin, T. H. & Nasmyth, K. (1996) *Cell* **84**, 699-709.

11. Sil, A. & Herskowitz, I. (1996) *Cell* **84**, 711-22.

12. Bohl, F., Kruse, C., Frank, A., Ferring, D. & Jansen, R. P. (2000) *Embo J* **19**, 5514-24.

13. Long, R. M., Gu, W., Lorimer, E., Singer, R. H. & Chartrand, P. (2000) *Embo J* **19**, 6592-601.

14. Takizawa, P. A. & Vale, R. D. (2000) *Proc Natl Acad Sci U S A* **97**, 5273-8.

13

15.   Takizawa, P. A., DeRisi, J. L., Wilhelm, J. E. & Vale, R. D. (2000) *Science* **290,** 341-4.

16.   Jansen, R. P., Dowzer, C., Michaelis, C., Galova, M. & Nasmyth, K. (1996) *Cell* **84,** 687-97.

17.   Kruse, C., Jaedicke, A., Beaudouin, J., Bohl, F., Ferring, D., Guttler, T., Ellenberg, J. & Jansen, R. P. (2002) *J Cell Biol* **159,** 971-82.

18.   Kress, T. L., Yoon, Y. J. & Mowry, K. L. (2004) *J Cell Biol* **165,** 203-11.

19.   Munchow, S., Ferring, D., Kahlina, K. & Jansen, R. P. (2002) *Curr Genet* **41,** 73-81.

20.   Niessing, D., Huttelmaier, S., Zenklusen, D., Singer, R. H. & Burley, S. K. (2004) *Cell* **119,** 491-502.

21.   Gonsalvez, G. B., Lehmann, K. A., Ho, D. K., Stanitsa, E. S., Williamson, J. R. & Long, R. M. (2003) *Rna* **9,** 1383-99.

22.   Gonsalvez, G. B., Little, J. L. & Long, R. M. (2004) *J Biol Chem* **279,** 46286-94.

23.   Olivier, C., Poirier, G., Gendron, P., Boisgontier, A., Major, F. & Chartrand, P. (2005) *Mol Cell Biol* **25,** 4752-66.

24.   Estrada, P., Kim, J., Coleman, J., Walker, L., Dunn, B., Takizawa, P., Novick, P. & Ferro-Novick, S. (2003) *J Cell Biol* **163,** 1255-66.

25.   Toi, H., Fujimura-Kamada, K., Irie, K., Takai, Y., Todo, S. & Tanaka, K. (2003) *Mol Biol Cell* **14,** 2237-49.

26.   Wesche, S., Arnold, M. & Jansen, R. P. (2003) *Curr Biol* **13,** 715-24.

27.   Evangelista, M., Blundell, K., Longtine, M. S., Chow, C. J., Adames, N., Pringle, J. R., Peter, M. & Boone, C. (1997) *Science* **276**, 118-22.

28.   Kohno, H., Tanaka, K., Mino, A., Umikawa, M., Imamura, H., Fujiwara, T., Fujita, Y., Hotta, K., Qadota, H., Watanabe, T., Ohya, Y. & Takai, Y. (1996) *Embo J* **15**, 6060-8.

29.   Beach, D. L. & Bloom, K. (2001) *Mol Biol Cell* **12**, 2567-77.

30.   Trautwein, M., Dengjel, J., Schirle, M. & Spang, A. (2004) *Mol Biol Cell* **15**, 5021-37.

31.   Gu, W., Deng, Y., Zenklusen, D. & Singer, R. H. (2004) *Genes Dev* **18**, 1452-65.

32.   Irie, K., Tadauchi, T., Takizawa, P. A., Vale, R. D., Matsumoto, K. & Herskowitz, I. (2002) *Embo J* **21**, 1158-67.

33.   Tadauchi, T., Matsumoto, K., Herskowitz, I. & Irie, K. (2001) *Embo J* **20**, 552-61.

34.   Gonzalez, I., Buonomo, S. B., Nasmyth, K. & von Ahsen, U. (1999) *Curr Biol* **9**, 337-40.

35.   Long, R. M., Gu, W., Meng, X., Gonsalvez, G., Singer, R. H. & Chartrand, P. (2001) *J Cell Biol* **153**, 307-18.

36.   Kim-Ha, J., Webster, P. J., Smith, J. L. & Macdonald, P. M. (1993) *Development* **119**, 169-78.

37.   Betley, J. N., Frith, M. C., Graber, J. H., Choo, S. & Deshler, J. O. (2002) *Curr Biol* **12**, 1756-61.

38.   Huang, Y. S., Carson, J. H., Barbarese, E. & Richter, J. D. (2003) *Genes Dev* **17**, 638-53.

15

39.   Macdonald, P. M. & Kerr, K. (1998) *Mol Cell Biol* **18,** 3788-95.

40.   Gautreau, D., Cote, C. A. & Mowry, K. L. (1997) *Development* **124,** 5013-20.

41.   Ainger, K., Avossa, D., Diana, A. S., Barry, C., Barbarese, E. & Carson, J. H. (1997) *J Cell Biol* **138,** 1077-87.

42.   Hachet, O. & Ephrussi, A. (2004) *Nature* **428,** 959-63.

43.   Crucs, S., Chatterjee, S. & Gavis, E. R. (2000) *Mol Cell* **5,** 457-67.

44.   Chartrand, P., Meng, X. H., Singer, R. H. & Long, R. M. (1999) *Curr Biol* **9,** 333-6.

45.   Beach, D. L., Salmon, E. D. & Bloom, K. (1999) *Curr Biol* **9,** 569-78.

Chapter 2:

Widespread Cytoplasmic mRNA Transport in *S. cerevisiae*: Identification of 22 New Bud-Localized Transcripts Using DNA Microarray Analysis

Author contributions:

This chapter is a reprint from the following reference:

Kelly A. Shepard, Andre P. Gerber, Ashwini Jambhekar, Peter.A.Takizawa, Patrick O. Brown, Daniel Herschlag, Joseph L. DeRisi, and Ronald D. Vale. (2003) Widespread Cytoplasmic mRNA Transport in *S. cerevisiae*: Identification of 22 New Bud-Localized Transcripts Using DNA Microarray Analysis. PNAS **100** (20): 11429-11434.

Kelly Shepard, Peter Takizawa, and Andre Gerber performed the microarray experiments described in Fig. 1 and Table 2. Kelly Shepard generated the yeast strains, plasmids, and obtained the data shown in figures 2-4. Ashwini Jambhekar generated yeast strains, plasmids, and performed microscopy on half of the full-length and coding region RNAs listed and depicted in figures 2 and 4, and generated half of the yeast strains described in figure 3. Patrick Brown, Daniel Herschlag, Joseph DeRisi, and Ronald Vale supervised research.

Joseph L. DeRisi, Thesis Advisor

Abstract:

Cytoplasmic mRNA localization provides a means of generating cell asymmetry and segregating protein activity. Previous studies have identified two mRNAs that localize to the bud tips of the yeast *Saccharomyces cerevisiae*. To identify additional localized mRNAs, we immunoprecipitated the RNA transport components She2p, She3p, and Myo4p and performed DNA microarray analysis of their associated RNAs. A secondary screen, utilizing a GFP-tagged RNA reporter assay, identified 22 new mRNAs that are localized to bud tips. These messages encode a wide variety of proteins, including several involved in stress responses and cell wall maintenance. Many of these proteins are asymmetrically localized to buds. However, asymmetric localization also occurs in the absence of RNA transport, suggesting the existence of redundant protein localization mechanisms. In contrast to findings in metazoans, the untranslated regions are dispensable for mRNA localization in yeast. This study reveals an unanticipated widespread use of RNA transport in budding yeast.

19

Introduction:

Localization of mRNA in eukaryotic cells constitutes an important mechanism for sequestering protein activity, regulating gene expression, and establishing or maintaining cell polarity (1, 2). Studies of mRNA localization in a variety of organisms have suggested the following sequence of events: Initially, an RNA molecule with specialized targeting information, or "zipcode," is recognized by a protein or protein complex that recruits a cytoskeletal motor protein (3). Next, the resulting RNP complex is transported to a specific subcellular location along actin filaments or microtubules (4). Finally, the transcript becomes anchored to its final destination, where translation occurs only at the targeted location (5). In many cases, translational repression is coupled with mRNA transport to ensure protein expression does not occur while in transit to the appropriate site of activity (5).

In *Saccharomyces cerevisiae*, RNA encoding the daughter-cell specific transcription factor, Ash1p, was discovered to be localized to the bud tip (6, 7). This represented the first description of RNA localization in a single-celled eukaryote. *ASH1* transcripts are recognized by the She2p protein, which becomes associated with the Myo4p myosin motor via the She3p adapter protein (8-10). This RNP complex travels along actin cables to the emerging bud where the transcript is anchored and translated. The Ash1p protein then is transported into the bud nucleus where it represses the HO locus and inhibits mating type switching. Two other proteins are also important for efficient *ASH1* localization in yeast: Loc1p, a nuclear protein (11), and Khd1p, a KH (hnRNP K homology) domain containing protein that is thought to link translational repression to the localization process (12).

20

(hnRNP K homology) domain containing protein that is thought to link translational repression to the localization process (12).

A second localized mRNA, *IST2*, was identified by a microarray-based strategy involving immunoprecipitation of She proteins, amplification of associated RNAs, and subsequent hybridization on DNA microarrays to determine their enrichment compared to those from a control immunoprecipitation (13). *IST2* encodes a plasma membrane protein that is enriched in the bud. Along with bud-localized expression, Ist2p protein is prevented from diffusing into the mother cell by the septin barrier at the mother-bud junction (13). In addition to *IST2*, microarray analysis identified 10 other mRNAs that were associated with the She complex (13). *In situ* hybridization procedures, however, revealed only *IST2* and *ASH1* to be asymmetrically localized, while the remainder were inconclusive or ambiguous (13). These same eleven transcripts were immunoprecipitated by each of She proteins independently, suggesting that these results reflect *bona fide* associations. However, the utility of this approach for genome-wide identification of localized mRNAs remained to be established. In this study, we have further refined various microarray approaches and developed improved methods for screening candidate transcripts for localization. With these improved methodologies, we have identified a family of messages that are localized to the tips of buds. Along with *ASH1* and *IST2*, these bring the total number of mRNAs known to be transported by the She protein machinery to 24.

Materials and Methods:

Strains/ Plasmids and Microscopy:

Myc-tagged strains for microarray experiments were derived from W303 (8).

TAP-tagged She3p and Myo4p strains were obtained from Cellzome (14). The protein

A-tagged She2 strain (APG32) was generated by transforming BY4741 (15) with a PCR

fragment generated from plasmid pFA6-TEVzz-kanMX6 (16), inserting two tandem IgG-

binding domains downstream of *SHE2* by the method of Longtine et al. (17). Strains

harboring carboxyl-terminal GFP protein tags were a gift from Erin O'Shea and derived

from ATCC 201388. For examination of these proteins in a mutant background, *SHE2*

was disrupted by the method of Longtine et al. (17).

For GFP-tagging of mRNA, the pGAL-U1A plasmid was created by inserting the

GAL1 promoter and four copies of the U1A aptamer site upstream of a unique NotI site

and a *CYC1* terminator sequence in the unique SacII site. To test sequences for

localization, PCR products were amplified from genomic DNA using primers with

terminal NotI sites. PCR products were cloned into the NotI site of pGAL-U1A. All

constructs were confirmed by sequencing. To assess RNA localization, induction and

visualization of pGAL-U1A constructs was preformed as described (8). In general,

approximately 50-67% of pre-mitotic cells within a population displayed visible green

RNA particles after 2 hrs. of induction.

To quantify RNA localization, >100 pre-mitotic cells with small to medium buds

were identified and scored for GFP-RNA localized selectively to the bud versus a random

distribution throughout both mother and bud. Most strains were analyzed by independent observers in a double-blind fashion.

For amino-terminal GFP tagging of proteins, pAG36 was constructed by replacing the *MET25* promoter in pUG36 (unpublished; a gift from J.H. Hegemann, Heinrich Heine University, Germany) with a *GAL1* promoter via *SacI/XbaI* restriction sites. Genes encoding the protein to be visualized were cloned into either *Eco*RI or *Hind*III/*Xho*I sites. Plasmid pHS20 was constructed by Sesaki and Jensen (18) and was a gift from Michael P. Yaffe.

To visualize proteins, strains with carboxyl-terminal GFP tags were grown to mid-log phase in YPD and examined by fluorescence microscopy. To visualize amino-terminally tagged proteins, cells were grown overnight in SD-URA, diluted to 0.5 OD/ml, and induced for 1-2 hr with 0.2% galactose prior to examination by fluorescence microscopy.

Immunoprecipitation and Microarray Analysis:

Two different protocols were used in this study to identify She protein-associated RNAs (Figure 1). First, immunoprecipitaton of She proteins, amplification of associated RNAs and hybridization to microarrays was performed as described by Takizawa et al. (13). For the second method, one liter of cells were cultured at 30°C in YPD medium and collected during exponential growth by centrifugation. Cells were washed twice in 20 mM Tris-HCl (pH 8.0), 140 mM KCl, 1.8 mM MgCl₂, 0.1 % NP-40, 0.02 mg/ml heparin and resuspended in the above buffer containing 0.5 mM DTT, 1 mM PMSF, 0.5 µg/ml

leupeptin, 0.8 µg/ml pepstatin, 20 U/ml DNase I, 100 U/ml RNasin (Promega), and 0.2 µg/ml heparin. Purification of tagged proteins and isolation of associated RNA was essentially performed as described in (19) and (20). Briefly, cells were broken mechanically with glass beads, and extracts were incubated with IgG-agarose beads (Sigma). The beads were washed four times, and She proteins were released from the beads by cleavage with TEV-protease (Invitrogen). RNA was isolated by phenol/chloroform extraction and isopropanol precipitation from TEV eluates, which corresponds to the purified fraction, and from extracts (input). Both RNA samples, input and purified, were reverse transcribed and amino-allyl labeled with the fluorescent dyes Cy3 and Cy5 (Amersham), respectively. The samples were mixed and competitively hybridized to yeast cDNA microarrays containing all yeast genes as previously described (21).

Data Analysis and Retrieval:

Microarray data were extracted and analyzed essentially as described (22). For microarray Method 1, data for each experiment were extracted into Microsoft Excel after filtering out array elements that yielded weak signal intensities: the sum of median intensities for the two channels was required to be greater than 150, and the regression correlation value ($r2$) for the two channels was required to be greater or equal to 0.6.

For Method 2, data were stored and extracted from the Stanford Microarray Database (23). Similar to Method 1, low intensity signals were removed from the data set by filtering those elements for which the ratio of the signal to the background was greater

24

than 1.5 for the Cy3 channel (total RNA control), and greater than 1.0 for the Cy5

channel (pull-down RNA). In addition, the regression correlation value ($r2$) for the two

channels was required to be greater or equal to 0.6. Furthermore, array elements were

removed from the data set if more than five of the ten experiments in this method did not

satisfy the requirements described above. Since actual ratio values for enriched RNAs in

individual pull-downs were not comparable due to differences in pull-down

methodology, amplification and immunoprecipitation efficiencies, we used a method

identical to Lieb at al (22) whereby the median percentile rank for each array element

was calculated using ranked relative enrichment values for each experiment.

Data Interpretation:

Two of the 5 "Method 1" experiments were performed with microarrays

containing coding and intergenic sequences, whereas the remaining 3 were performed

with microarrays containing only yeast coding sequences. To compare experiments, all

intergenic loci were eliminated before percentile ranks were assigned. Median ranks from

all 5 experiments were then compared directly to determine median rankings.

The experiments performed with intergenic arrays suggested that several RNAs

from noncoding regions potentially interact with the She proteins. However, only two

experiments were performed with these types of arrays, and we are thus hesitant to

attribute significance to these results at this stage. Some of these intergenic regions are

adjacent to localized coding sequences (*ASH1*, *IST2*), while others are derived from

overlapping sequences (as the arrays are derived from PCR products, a positive

25

hybridization signal does not distinguish strands). Several, however, are unrelated to tested transcripts.

In addition to intergenic sequences, several transposable elements were identified as positives by our microarray screens. Because of the large number of candidates to investigate, we chose to focus our efforts on RNA derived from coding sequences. The relevance of She-interactions with intergenic and transposon-derived RNA remains to be determined.

Results:

Identification of She-dependent transport candidates using microarray-based approaches:
To identify mRNAs transported by the She machinery, we performed two different microarray-based experiments. First, using the method described by Takizawa et al. (8), we immunoprecipitated myc-tagged She proteins from either tagged or untagged extracts using a monoclonal anti-myc antibody. RNAs associated with the immunoprecipitates were amplified by random-primed RT-PCR, fluorescently labeled by further PCR, and hybridized to yeast microarrays to determine which transcripts were enriched in the tagged versus untagged immunoprecipitates (Method 1, Figure 1A).

Method 1 involved amplification of immunoprecipitated RNAs prior to microarray analysis. One advantage of this strategy is that rare or transiently expressed RNAs could be identified, even if initial amounts were miniscule. However, variation in the quantities of mRNA in the immunoprecipitates can lead to differential amplification during PCR. Thus, enrichment values, as estimated by intensity of signals on

26

microarrays, are largely nonquantitative. To address this limitation, a second method was employed whereby the She-proteins were either protein A or TAP (tandem affinity)-tagged and affinity purified (14, 21). She-associated RNAs were then directly labeled by reverse transcription and compared to total RNA by competitive hybridization on yeast microarrays (Method 2, Figure 1B).

Within each of the two experimental methodologies, an overlapping set of transcripts were enriched in the She2p, She3p and Myo4p immunoprecipitates, consistent with previous studies indicating that these proteins interact as a complex for mRNA transport (8-10). In addition, 13 transcripts were identified by both methodologies (see Table 2 for details of the microarray results). Although a number of transposable elements emerged as positives, preliminary *in situ* hybridization analysis indicated that these RNAs are not selectively enriched in the bud (Peter Takizawa, personal communication). Thus, for the remainder of this study, we focused our efforts on only those candidates that encode predicted or known proteins.

At least 24 mRNAs are transported to the tips of emerging buds by the She proteins:

Of the 24 She protein-associated transcripts listed in Table 1, eleven were also identified by Takizawa et al. (13) and described in their supplementary material. However, their further studies using *in situ* hybridization identified only *IST2* and *ASH1* as localized RNAs. The remainder yielded ambiguous results due to low or variable signals, problematic background from the hybridization procedure, or poor reproducibility (8). To improve the localization assay and determine which transcripts

27

were *bona fide* She-protein transport substrates, we used a U1A aptamer-based GFP

tagging system described by Takizawa and Vale (13) that allows mRNA visualization by

fluorescence microscopy. In this procedure, a yeast strain is transformed with two

plasmids. The first expresses GFP fused to U1A, an RNA-binding protein that recognizes

a specific sequence, the U1A aptamer. The second plasmid harbors a galactose-inducible

promoter and four copies of the U1A aptamer fused to the 5' end of a transcript to be

analyzed. To aid in visualization, the U1A-GFP fusion carries a nuclear localization

signal to direct excess, unbound protein to the nucleus. This GFP tagging procedure

proved more sensitive and reproducible than *in situ* hybridization procedures, as

expression from an inducible promoter allows overexpression and thus visualization of

transient or rare transcripts that might have been difficult to detect. It is possible that

U1A-GFP tagging might interfere with a transcript's localization, and overexpression

might alter stoichiometry or anchoring properties of the resulting RNP complexes.

However, these limitations did not detract from the utility and ease of this GFP-tagging

system as a powerful secondary screen for RNA localization.

To assess localization, we utilized the GFP-tagging strategy to examine mRNA

candidates that were identified in the top 98th percentile of one or both microarray

methodologies (Table 2). To define criteria for mRNA localization, we quantified the

fraction of pre-mitotic cells with GFP-RNA concentrated in bud tips compared to those

with GFP-RNA randomly distributed throughout mother and bud. Pre-mitotic cells were

chosen, because after the nuclei divide, we and others (24) observed that bud tip-

localized RNAs redistribute first to the mother-daughter junction and then throughout the

cytoplasm, making localization problematic to establish in larger-budded cells. We defined a "localized" mRNA as one that forms green particles that associate with the tips of small and emerging buds in >50% of pre-mitotic cells after 1-2 hr of induction, similar to what has been described for GFP-tagged *ASH1* (8, 25-27). We defined an "unlocalized" mRNA as one that forms green particles that are randomly distributed between mother and bud, as has been reported for nonlocalized transcripts such as *ADH1* (8, 25-27). Figure 2 shows examples of these two types of distribution.

For the top ten (Method 1) and nine of the top ten (Method 2) highest-ranking candidates from the array experiments (Table 2), GFP-RNA was associated with bud tips in >90% of the small-budded cells that contained visible RNA particles. In total, out of 38 tested, we found a total of 17 mRNAs that were clearly localized to bud tips, 15 that were unlocalized, and one that did not form visible GFP-RNA (*SHE3*). A full list of all mRNAs tested can be viewed in Table 2.

Although the majority of RNAs displayed clear localized or unlocalized phenotypes, several transcripts exhibited partial localization. Three of the mRNAs, *MET4*, *LCB1*, and *KSS1*, displayed small-bud-localized GFP-RNA in ~50% of cells while the remaining half of the cells displayed randomly distributed particles (Table 1). Two other candidates, *MTL1* and *YPL066w* exhibited an apparently random distribution of GFP-RNA in most cells; however, 15-30% of cells displayed buds that were enriched for GFP-RNA, something that was never observed in "unlocalized" candidates. This variability in localization efficiency might result from attenuation of anchoring, translational, or localization signals in the transcripts due to an artifact of the U1A-GFP

tagging strategy. Alternatively, it might reflect a genuine difference in affinity of these transcripts for the She machinery *in vivo*. Because of their weak localization activity, we chose not to pursue *MTL1* and *YPL066W* for further analysis.

To determine whether bud-specific localization of mRNA requires the She protein machinery, we examined mRNA localization in *she2Δ* cells. As described previously for *ASH1* and *IST2*, all of the localized RNAs (Table 1), as well as the 5 partially localized mRNAs described above, were delocalized from bud tips and appeared randomly distributed throughout both mothers and buds (Figure 2). Similar delocalization was observed in *myo4Δ* or *she3Δ* mutants for three transcripts that were tested: *MID2*, *IST2*, and *TPO1* (data not shown). These data indicate that, like *ASH1* and *IST2*, all of the new localized transcripts require the She-complex to localize to the tips of growing buds.

Many, but not all, proteins encoded by localized mRNAs show asymmetric distributions: Messenger RNA localization is essential for the asymmetric distribution of Ash1p protein to bud nuclei and Ist2p protein to the bud plasma membrane (6, 7, 13). To determine the normal localization of proteins encoded by the RNA candidates, we tagged each protein with GFP and assessed its subcellular distribution by fluorescence microscopy (Table 1). Because *ASH1* and most localized RNAs from metazoan cells have been shown to contain important RNA localization signals in or near their 3' UTRs (1, 6, 7), we initially tagged most of the proteins at the amino-terminus and placed their expression under control of the inducible *GAL1* promoter. By this strategy, we also hoped

to see expression of rare or transient transcripts that might have been difficult to detect at endogenous levels.

GFP-tagged proteins Tpo1p, YJL051p, and Srl1p were localized specifically to the periphery of buds, similar to what was observed for Ist2p (13) (Figure 3a). However, some of the proteins, including several that were predicted to be membrane-associated, such as Mid2p and Wsc2p, appeared as amorphous green particles in the cytoplasm. To determine whether this localization was genuine or perhaps an artifact of overexpression or tag position, we obtained strains harboring versions of the proteins that were chromosomally tagged with GFP at their carboxyl-termini and expressed from their endogenous promoters. Many of these GFP-tagged proteins showed asymmetric localization to the bud (Figure 3a), including Srl1p, YJL051p, YML072p, YNL087p, and Wsc2p, while GFP-Mmr1p (YLR190w) localized to discrete punctae at future bud sites, small buds, and mother-bud junctions in post-mitotic cells. The remaining proteins were segregated symmetrically to various subcellular destinations (Table 1, Fig. 3b).

Asymmetric protein localization can be achieved in the absence of mRNA localization: She-dependent RNA sorting is essential for asymmetric localization of Ash1p and Ist2p to the bud nucleus and bud membrane respectively (6, 7, 13). To determine whether RNA transport to the bud tip is essential for normal protein distribution, we expressed each of the GFP-tagged proteins in *she2Δ* cells and examined their distributions *in vivo*. Surprisingly, protein localizations, including those that were asymmetric, were unaltered in *she2Δ* mutants (Figure 3b). Thus, unlike Ash1p and Ist2p

31

(6, 7, 13), the majority of the She protein-RNA transport substrates encode proteins with redundant targeting information such that they are distributed correctly in the absence of RNA transport.

The coding regions of RNA transport substrates are sufficient for targeting to the bud: Zipcodes are regions of RNA sequence that are sufficient to confer localization of a reporter RNA (1). In metazoans, all published zipcodes to date have mapped to 3'- or (rarely) to 5'-untranslated regions of a transcript (1). However, the yeast *ASH1* mRNA contains zipcodes in both the coding sequence and 3'-untranslated region (6, 7, 26). To determine whether the other She-transported mRNAs contain zipcodes, we created U1A-GFP tagged versions of each localized message that included only the coding sequence and the *CYC1* terminator rather than endogenous termination sequences. Surprisingly, the coding sequence alone was sufficient to enable transport to the tips of growing buds for the all of the localized mRNAs (Table 1). In most cases, the efficiency of localization was similar to that observed for the full-length constructs (Figure 4). However, the *EGT2* coding transcript conferred bud localization in only 15-20% of the cells. For 12 of the localized mRNAs, we assessed the localization of the 3'-UTR alone (here defined as the five hundred base-pair region immediately downstream of the stop codon). The 3' UTRs of *ERG2*, *CLB2*, and *EGT2* conferred partial localization in the U1A–GFP reporter system, while the other tested 3'-UTRs appeared unlocalized (*IST2, YGR046w, MMR1, TPO1, YML072c, SRL1, MID2, YMR171c*) or did not form detectable GFP-RNA (*WSC2*)

(data not shown). These data indicate that zipcode information within the coding region is usually sufficient for the recognition and transport of mRNA by the She-complex.

Discussion:

We have identified 22 new substrates for the yeast She-protein RNA transport machinery through a combination of protein immunoprecipitation, DNA microarray analysis, and GFP-RNA visualization. The GFP-RNA tagging strategy, as opposed to *in situ* hybridization, proved crucial for documenting the localization of additional RNAs compared with the original study by Takizawa et al. (13). This study reveals that RNA localization in yeast is more widespread than previously appreciated. It is possible that these newly-identified mRNAs represent the majority of the She-protein cargoes, given that the lower-ranking candidates in the screen were less likely to localize to bud tips. However, our experiments were performed on cells that were growing exponentially in rich media. Any transcripts that are expressed only transiently or under a limited set of circumstances would likely have been overlooked. In addition, some weakly-binding RNAs may dissociate from the She complex *in vitro* prior to immunoprecipitation. Finally, it is possible that alternative RNA localization mechanisms exist in yeast that are independent of the She-complex. Indeed, RNA targeting to the mitochondria has been described in budding yeast (28). Although the mechanism for this process has yet to be established, none of the mitochondrial-targeted mRNAs were identified in our microarray experiments as substrates for the She machinery.

33

Predicted functions of proteins encoded by localized mRNAs:

The proteins encoded by localized mRNAs are diverse, although some appear to participate in common pathways related to sensing or responding to stress. *WSC2* encodes a heat shock sensor that transduces signals via a *MPK1* pathway, while *MID2* transmits a similar signal in response to α-factor. Mutations in either of these proteins can lead to osmotic sensitivity (29, 30), and a *mid2Δ* mutation is suppressed by overexpression of Wsc2p (30). *MTL1* is thought to encode a cell wall sensor and displays significant sequence homology with both *WSC2* and *MID2* (30, 31). Two other localized RNAs, *BRO1* and *KSS1*, encode components of MAP kinase signaling pathways that regulate cellular responses to various environmental stresses (32, 33), while *IST2* and *TPO1* encode membrane transporters that may modulate intracellular ion concentrations (34, 35). The functional significance of this potentially interconnected group of localized messages requires further elucidation. However, there may exist interactions or interdependencies amongst She-complex targets at the RNA or protein level that do not readily reveal themselves through traditional genetic methods.

Several localized RNAs encode proteins that play roles in synthesis and modification of the plasma membrane and cell wall. *ERG2* and *LCB1* encode enzymes involved in lipid synthesis (36), while *EGT2* encodes a cellulase that may be involved in cell separation (37). Localization of these proteins to the bud would concentrate their activities at the most active site of cell growth and remodeling.

Ten of the 24 She-localized mRNAs are transcribed from cell-cycle regulated genes (Table 1), a significantly higher proportion than the genomic representation of such

genes (10%) (34). Seven of these transcripts have peak expressions at M or M/G1 (Table 1). She2 transcription also peaks at M phase, suggesting that RNA transport machinery might be maximally expressed at the time of its greatest usage (34). Of the other M-phase regulated transcripts, *CLB2* is of particular interest, as the mRNAs for cyclin B homologs are localized in a variety of metazoan organisms (1). Although we only were able to detect Clb2p in nuclei and spindle pole bodies, domain analysis has revealed that there might be a subpopulation of Clb2p present at the bud tip and mother-bud junction (38). A bud-specific pool of Clb2p might be important for initiating a daughter-specific genetic program. Perhaps the She-complex is responsible for localizing a subset of *CLB2* mRNA which resides outside the nucleus and is expressed only transiently or at levels that are difficult to detect by conventional means. Indeed, it is possible that asymmetric subpopulations could exist for other proteins encoded by localized messages whose overall distributions appeared random in our visualization assay.

Many of the localized RNAs encode proteins that are known or predicted to associate with membranes. Several of these are asymmetrically enriched in the bud, and many encode proteins of unknown function (e.g. *YML072c, YJL051c, YNL087w)*. Others (*YMR046w, YLR434c, DNM1, ERG2, LCB, MET4, CLB2*) encode proteins that are symmetrically distributed to subcellular membranous structures such as mitochondria, endoplasmic reticulum, or nuclei. Despite their apparent symmetry, it is possible that these proteins have asymmetric distributions that were disrupted by introduction of GFP tags and/or overexpression. Alternatively, these proteins may be synthesized at the site of

the localized transcript, but the newly-translated proteins may rapidly equilibrate between mother and bud, thereby creating a symmetric steady-state distribution.

The role of RNA transport:

Messenger RNA transport is generally thought to play a primary role in creating an asymmetric protein distribution in yeast (6, 7). However, other than *ASH1*, there is no evidence to suggest that asymmetry is important for the function of proteins encoded by localized transcripts. Indeed, the abolition of RNA transport by deletion of the She machinery has no obvious defects in growth or fitness (39). However, we found that most asymmetrically-distributed proteins encoded by RNA transport substrates maintained their localization in the absence of RNA transport. Thus, these proteins must themselves contain targeting information that allows for their appropriate post-translational segregation. For the plasma membrane proteins, this redundancy may involve the secretory pathway, which includes actin-based transport of vesicles to the bud. The ability of these proteins to sort to their appropriate locations post-translationally is consistent with observations in other organisms (39). For example, both Prospero mRNA and protein are asymmetrically localized in *Drosophila* neuroblasts, yet the protein can achieve asymmetry even when its RNA is symmetrically dispersed (40). These results might imply the necessity of alternative targeting mechanisms for important biological functions.

The observation that several She substrates encode proteins that are not asymmetrically distributed raises the possibility that She proteins and mRNA transport

might also serve functions that are not directly related to protein localization. One possibility is that the She complex may provide buds with a "start-up package" of mRNA, so that a daughter cell can respond to stimuli without initiating a round of its own transcription. This notion is analogous to the transfer of maternal mRNAs into developing *Drosophila* oocytes. Such a feature might give fitness advantages to progeny under certain environmental stresses, perhaps some that are too subtle to have been detected by standard competitive growth assays. Alternatively, She proteins might affect the anchoring or translation of their substrates, either by direct interaction with RNA or through interactions with accessory molecules, such as Khd1 (11) or Loc1 (12). Studies in *Drosophila* have identified several molecules that integrate RNA transport with related processes such as splicing, nuclear export and translational control (41, 42). It is likely that similar relationships in yeast will emerge upon further elucidation of the functional contribution of each component of the She mRNA transport complex.

The coding region contains mRNA localization elements:

The identification of 24 substrates for the She machinery provides new resources for identifying sequences that are important for recognition by She2p. Determination of zipcode sequences in yeast as well as higher organisms has been challenging, as this information is thought to reside in secondary or tertiary structures that are difficult to predict. In contrast to other organisms such as *Drosophila*, where multiple RNA localization pathways exist and few zipcode-binding adapter molecules have been identified, these yeast RNAs interact with the same adapter molecule, She2p. In this

37

study, we have determined that the majority of yeast zipcode domains lie within the coding sequences of the genes. A recent study from Chartrand et al. (43) reported that localization elements in the *ASH1* coding sequence serve to decelerate translation, presumably so that *ASH1* mRNA is not prematurely expressed while in transit to the bud tip. The number and position of zipcode elements in the coding sequences of other localized yeast mRNAs may prove similarly important for timing of protein expression. Indeed, it is even possible that RNA localization to the bud tip might be a secondary consequence of translational regulation by certain zipcode elements, which could explain why several of the proteins encoded by localized RNAs are not found specifically in the bud.

With this new repertoire of localized RNAs, mapping of zipcode motifs and analysis of their contribution to protein expression will yield insight into how these sequences mediate both motility and translation. In addition, future functional and computational analyses of these regions could aid in the development of predictive tools for zipcode identification in yeast as well as other organisms.

Applications in other systems:

We have demonstrated that a microarray approach based on purification of RNA-binding proteins, combined with a robust reporter system as a secondary screen, provides a powerful technique for identifying localized RNAs. This strategy could be used to discover new RNA cargoes in a variety of organisms. Any molecule that is known or suspected to play a role in transport, whether it is a motor protein or an adapter, could be

used as a handle to purify and identify associated RNAs. An aptamer-GFP binding strategy has now been described that allows visualization of RNAs in mammalian cells (44), but other RNA visualization techniques, such as *in situ* hybridization (13, 45) and microinjection (46, 47) also can be utilized as secondary screens. Given the relative ease of these procedures and the increasing availability of genomic tools, it is possible that analogous experiments will lead to the rapid identification of large repertoires of localized RNAs in a wide variety of organisms.

Table 1. Localized Transcripts

GENE	RNA Localization: full	coding	she2Δ	Cell Cycle Regulation	Predicted Function	Protein Localization	Tag Position
ASH1	yes	yes	no	M	transcription	bud nucleus (51-53)	N,C
BRO1	yes	yes	no	none	stress transduction	punctae on vacuole	C
CLB2	yes	yes	no	M	cyclin B	nuclei, spindle poles	C
CPS1	yes	yes	no	none	carboxypeptidase	cytoplasmic punctae	N
DNM1	yes	yes	no	S	mitochondrial fission	mitochondrial periphery (18, 54)	C
EGT2	yes	yes	no	M	cellulase	membranes, large-bud enriched	C
ERG2	yes	partial	no	M	sterol isomerase	endoplasmic reticulum	N
IST2	yes	yes	no	none	tranporter	bud plasma membrane (13)	N
MID2	yes	yes	no	none	membrane receptor	cell periphery, mother-bud junction	C
MMR1	yes	yes	no	M	unknown	bud sites and tips, mother-bud junction	N
SRL1	yes	yes	no	G1	unknown	periphery of small buds	C
TPO1	yes	yes	no	M	polyamine transport	bud plasma membrane	N
WSC2	yes	yes	no	S	membrane receptor	membranes, bud-enriched	C
YGR046W	yes	yes	no	none	unknown	mitochondria	N
YJL051C	yes	yes	no	M	unknown	membranes, bud-enriched	N
YLR434C	yes	yes	no	none	unknown	mitochondria	N
YML072C	yes	yes	no	G2	unknown	membranes, bud-enriched	C
YMR171C	yes	yes	no	none	unknown	endoplasmic reticulum	N
YNL087W	yes	yes	no	none	unknown	membranes, bud-enriched	C
KSS1	partial	n.d.	no	none	MAP kinase	n.d.	n.d.
LCB1	partial	partial	no	none	ER, lipid synthesis	endoplasmic reticulum	C
MET4	partial	partial	no	none	transcription	nuclei	C
MTL1	weak	n.d.	no	none	MID2-like	n.d.	n.d.
YPL066C	weak	n.d.	no	none	unknown	n.d.	n.d.

List of She-localized mRNAs. Also indicated is whether or not each transcript is cell cycle regulated and its peak stage (48, 49), the predicted or known function, and protein localization of each. Proteins that are asymmetrically enriched in the bud are indicated by shading. The full list of transcripts assayed for localization as well as their percentile rankings from microarray analysis can be viewed in Table 2. C=carboxyl-terminus, N= amino-terminus. "yes" = > 90% bud localization; "partial" = 50-60% localization; "weak" = 15-30% localization; "no" = unlocalized (< 5% localization).

Table 2. Whole Genome Analysis of She-Associated mRNAs

Method 1 Previously Published

GENE ID	NAME	MEDIAN RANK
YMR202W	ERG2	1
YBR086C	IST2	0.999
YNL283C	WSC2	0.999
YGR046W	YGR046W	0.999
YBL005W-A	YBL005W-A	0.998
YLR190W	YLR190W	0.998
YLL028W	TPO1	0.998
YJL051W	YJL051W	0.998
YOR247W	SRL1	0.997
YOR248W	YOR248W	0.997
YOR247W	SRL1	0.997
YKL185W	ASH1	0.996

rankings from Takizawa et al. (2000)
1 X Myo4-myc Immunoprecipitation
1 X She3-myc Immunoprecipitation
1 X She2-myc Immunoprecipitation

*See Supplementary Material
www.sciencemagazine.org/feature/data/1053179.sh

- localized
- unlocalized
- not expressed
- not tested

Method 1:

GENE ID	NAME	MEDIAN RANK	
		method 1	method 2
YBR086C	IST2	0.999	0.999
YNL283C	WSC2	0.999	1.000
YMR202W	ERG2	0.999	1.000
YGR046W		0.999	0.994
YKL185W	ASH1	0.998	1
YLR190W	MMR1	0.998	0.875
YLL028W	TPO1	0.998	1
YOR247W	SRL1	0.997	0.964
YJL051W		0.997	0.985
YML072C		0.996	0.998
YMR171C		0.995	0.984
YPL066W		0.993	0.872
YJL172W	CPS1	0.994	0.642
YDR158W	HOM2	0.992	0.939
YLR332W	MID2	0.992	0.997
YNL327	EGT2	0.992	0.616
YPL084W	BRO1	0.992	0.621
YNL087W		0.991	0.487
YGR023W	MTL1	0.99	0.986
YKR012C		0.99	0.88
YPR119W	PLB1	0.989	n.d.
YMR296C	LCB1	0.989	0.324
YBR286W	APE3	0.988	0.368
YML015C	TAF11	0.988	0.791
YPR119W	CLB2	0.988	0.994
YBR130C	SHE3	0.987	0.673
YNL103W	MET4	0.983	0.789

data from Takizawa et al. (2000) ranked with data
from two additional experiments:
2 X She2-myc Immunoprecipitation
2 X She3-myc Immunoprecipitation
1 X Myo4-myc Immunoprecipitation

Method 2:

GENE ID	NAME	MEDIAN RANK	
		method 2	method 1
YMR202W	WSC2	1.000	0.999
YKL185W	ASH1	1.000	0.998
YLL028W	TPO1	1.000	0.998
YNL283C	ERG2	0.999	0.999
YBR086C	IST2	0.999	0.999
YML072C		0.998	0.996
YLR332W	MID2	0.997	0.992
YBL005W	PDR3	0.996	0.882
YPR119W	CLB2	0.994	0.988
YGR046W		0.994	0.999
YMR295C		0.992	0.889
YML109W	ZDS2	0.991	0.229
YMR294W-A		0.990	0.605
YGR022C	MTL1	0.986	0.990
YMR171C		0.984	0.995
YJL051W		0.985	0.997
YMR023C	MSS1	0.985	0.941
YKL187C		0.984	0.726
YCR047C	BUD23	0.983	0.534
YBL034C	STU1	0.981	0.859
YDR317W		0.981	0.311
YLR434C		0.980	0.485
YGL116W	CDC20		0.326
YML097C	VPS9	0.978	0.370
YML113W	DAT1	0.977	0.633
YLL001W	DNM1	0.974	0.859
YDR340W		0.974	0.961
YHR177W		0.973	0.091
YHR076W	PTC7	0.971	0.262
YOR247W	SRL1	0.964	0.997
YGR040w	KSS1	0.69	0.937

data from 10 experiments:
4 X She2-ZZ Immunopurification
3 X She3-TAP Immunopurification
3 X Myo4-TAP Immunopurification

We calculated percentile ranks for each gene within individual experiments. We then calculated median percentile ranks for each gene within each methodology. For Method 1, the first data set is derived from the supplemental data of Takizawa et al. (13). The second data set combines values from the Takizawa immunoprecipitations with data from two additional immunoprecipitations (1 x She3 and 1 x She2) for improved rank assignments. The third data set (Method 2) summarizes a total of 10 immunopurifications: 4 replicates for She2, and 3 each for She3 and Myo4. Top mRNA candidates are listed for each methodology. Rankings from both methodologies are designated for comparative purposes. Percentile ranks cannot be compared directly because the number of species analyzed per array differs between Methods 1 and 2. Shadings indicate whether RNA was localized, unlocalized, or was not tested.

Figure 1: Schematic representation of microarray-based screens for localized RNAs.

(A) Immunoprecipitations were performed with an anti-myc antibody from cellular extracts harboring either myc-tagged or untagged She proteins. RNAs enriched in pellets were amplified by RT-PCR and PCR, then labeled with either Cy5 (tagged) or Cy3 (untagged) nucleotides. Microarrays were probed with labeled PCR products and enrichment for Cy5 was assessed. (B) Each She protein was protein A or TAP-tagged and affinity purified. Associated RNAs were labeled directly by RT with Cy5, then competitively hybridized on microarrays with total RNA from wild-type cells that had been labeled with Cy3. (C) Candidates from microarray analyses were tested for localization *in vivo* with a GFP-RNA tagging strategy.
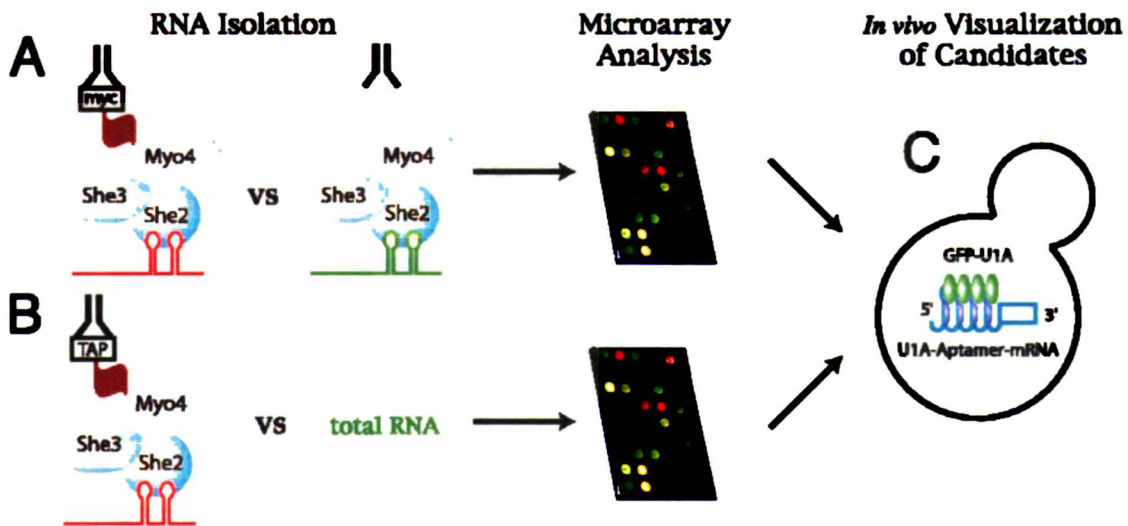
Figure 2: Localized and unlocalized RNAs identified through microarray analyses.
(A) Wild-type (*SHE2*, upper) or *she2Δ* (lower) cells expressing GFP-RNA for indicated
transcript were visualized by fluorescence microscopy. (B) Representative images for all
localized RNAs in WT cells. Bar = 2 μm.

A)
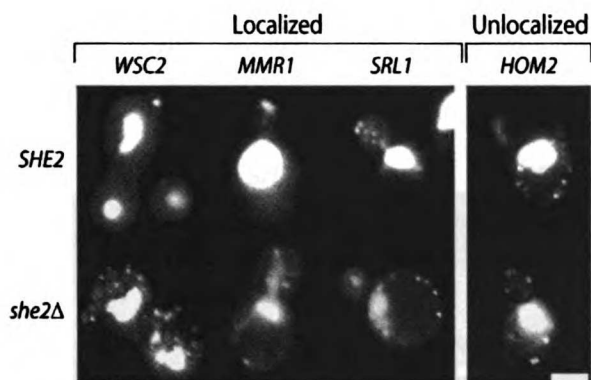
B)

Figure 3: Some She-complex targets encode asymmetrically localized proteins, and protein localization is independent of RNA transport. (A) Asymmetrically localized GFP-tagged proteins encoded by She-complex targets in wild-type (*SHE2*, upper) or *she2Δ* (lower) cells. (B) Symmetrically localized GFP-tagged proteins encoded by She-complex targets in WT cells. Bar = 2 μm.

A)

B)
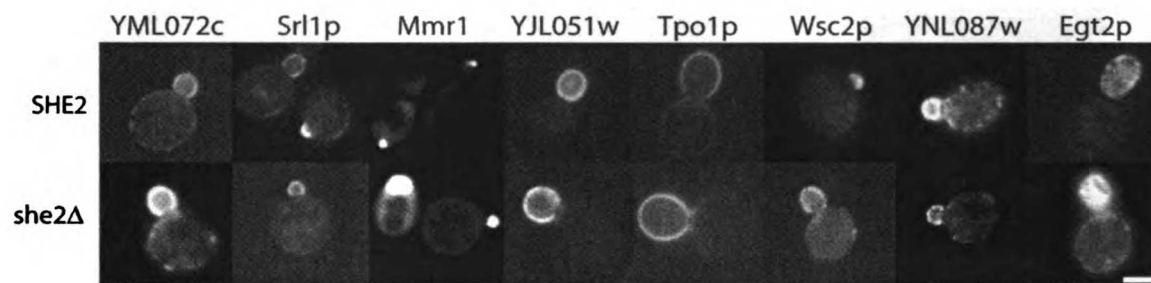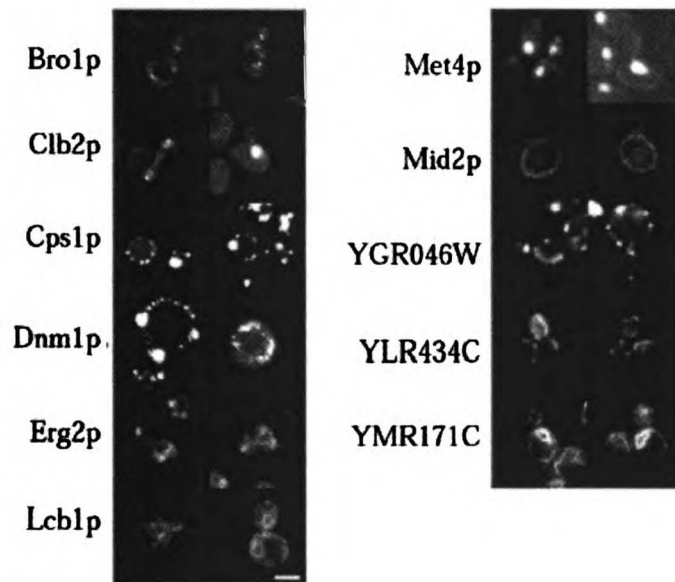
Figure 4: Coding sequences of She-transport substrates are largely sufficient for RNA localization. Wild-type cells expressing GFP-RNA for coding regions of indicated transcripts were visualized by fluorescence microscopy. Three representative images are shown for each RNA. Bar = 2 μm.
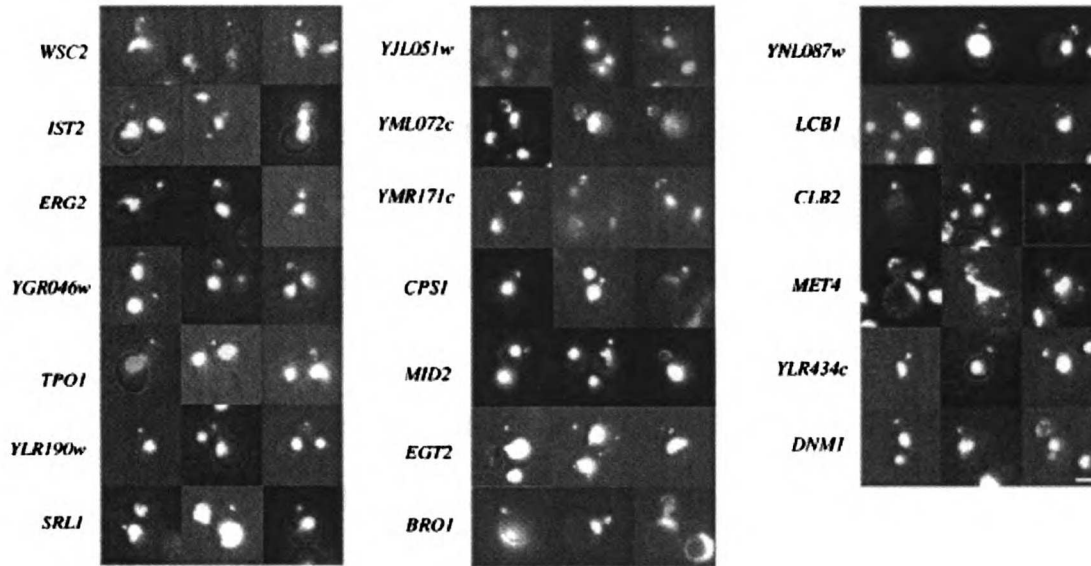
References:

1. Bashirullah, A., Cooperstock, R. L. & Lipshitz, H. D. (1998) *Annu Rev Biochem* **67,** 335-94.

2. Tekotte, H. & Davis, I. (2002) *Trends Genet* **18,** 636-42.

3. Kislauskis, E. H. & Singer, R. H. (1992) *Curr Opin Cell Biol* **4,** 975-8.

4. Oleynikov, Y. & Singer, R. H. (1998) *Trends Cell Biol* **8,** 381-3.

5. Johnstone, O. & Lasko, P. (2001) *Annu Rev Genet* **35,** 365-406.

6. Long, R. M., Singer, R. H., Meng, X., Gonzalez, I., Nasmyth, K. & Jansen, R. P. (1997) *Science* **277,** 383-7.

7. Takizawa, P. A., Sil, A., Swedlow, J. R., Herskowitz, I. & Vale, R. D. (1997) *Nature* **389,** 90-3.

8. Takizawa, P. A. & Vale, R. D. (2000) *Proc Natl Acad Sci U S A* **97,** 5273-8.

9. Long, R. M., Gu, W., Lorimer, E., Singer, R. H. & Chartrand, P. (2000) *Embo J* **19,** 6592-601.

10. Bohl F, K. C., Frank A, Ferring D, Jansen RP (2000) *Embo J* **19,** 5514-24.

11. Long, R. M., Gu, W., Meng, X., Gonsalvez, G., Singer, R. H. & Chartrand, P. (2001) *J Cell Biol* **153,** 307-18.

12. Irie, K., Tadauchi, T., Takizawa, P. A., Vale, R. D., Matsumoto, K. & Herskowitz, I. (2002) *Embo J* **21,** 1158-67.

13. Takizawa, P. A., DeRisi, J. L., Wilhelm, J. E. & Vale, R. D. (2000) *Science* **290,** 341-4.

14. Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M. A., Copley, R. R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G. & Superti-Furga, G. (2002) *Nature* **415,** 141-7.

15. Brachmann, C. B., Davies, A., Cost, G. J., Caputo, E., Li, J., Hieter, P. & Boeke, J. D. (1998) *Yeast* **14,** 115-32.

16. Preker, P. J., Kim, K. S. & Guthrie, C. (2002) *Rna* **8,** 969-80.

17. Longtine, M. S., McKenzie, A., 3rd, Demarini, D. J., Shah, N. G., Wach, A., Brachat, A., Philippsen, P. & Pringle, J. R. (1998) *Yeast* **14,** 953-61.

18. Sesaki, H. & Jensen, R. E. (1999) *J Cell Biol* **147,** 699-706.

19. Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M. & Seraphin, B. (1999) *Nat Biotechnol* **17,** 1030-2.

20. Gerber, A. P., Herschlag, D. & Brown, P. O. (2004) *PLoS Biol* **2,** E79.

21. DeRisi, J. L., Iyer, V. R. & Brown, P. O. (1997) *Science* **278,** 680-6.

22. Lieb, J. D., Liu, X., Botstein, D. & Brown, P. O. (2001) *Nat Genet* **28,** 327-34.

23. Gollub, J., Ball, C. A., Binkley, G., Demeter, J., Finkelstein, D. B., Hebert, J. M., Hernandez-Boussard, T., Jin, H., Kaloper, M., Matese, J. C., Schroeder, M., Brown, P. O., Botstein, D. & Sherlock, G. (2003) *Nucleic Acids Res* **31,** 94-6.

24. Beach, D. L., Salmon, E. D. & Bloom, K. (1999) *Curr Biol* **9,** 569-78.

25. Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S. M., Singer, R. H. & Long, R. M. (1998) *Mol Cell* **2,** 437-45.

26. Chartrand, P., Meng, X. H., Singer, R. H. & Long, R. M. (1999) *Curr Biol* **9,** 333-6.

27. Bloom, K. S., Beach, D. L., Maddox, P., Shaw, S. L., Yeh, E. & Salmon, E. D. (1999) *Methods Cell Biol* **61,** 369-83.

28. Corral-Debrinski, M., Blugeon, C. & Jacq, C. (2000) *Mol Cell Biol* **20,** 7881-92.

29. Zu, T., Verna, J. & Ballester, R. (2001) *Mol Genet Genomics* **266,** 142-55.

30. Ketela, T., Green, R. & Bussey, H. (1999) *J Bacteriol* **181,** 3330-40.

31. Rajavel, M., Philip, B., Buehrer, B. M., Errede, B. & Levin, D. E. (1999) *Mol Cell Biol* **19,** 3969-76.

32. Nickas, M. E. & Yaffe, M. P. (1996) *Mol Cell Biol* **16,** 2585-93.

33. Courchesne, W. E., Kunisawa, R. & Thorner, J. (1989) *Cell* **58,** 1107-19.

34. Dolinski, K., Balakrishnan, R., Christie, K. R., Costanzo, M. C., Dwight, S. S., Engel, S. R., Fisk, D. G., Hirschman, J. E., Hong, E. L.,, Issel-Tarver, L., Sethuraman, A., Theesfeld, C. L., Binkley, G., Lane, C., Schroeder, M., Dong, S., Weng, S., Andrada, R., Botstein, D., and & Cherry, J. M. *http: genome-www.stanford.edu Saccharomyces* *03/01/03.*

35. Tomitori, H., Kashiwagi, K., Sakata, K., Kakinuma, Y. & Igarashi, K. (1999) *J Biol Chem* **274,** 3265-7.

36.   Ashman, W. H., Barbuch, R. J., Ulbright, C. E., Jarrett, H. W. & Bard, M. (1991)
        Lipids **26,** 628-32.

37.   Kovacech, B., Nasmyth, K. & Schuster, T. (1996) *Mol Cell Biol* **16,** 3264-74.

38.   Hood, J. K., Hwang, W. W. & Silver, P. A. (2001) *J Cell Sci* **114,** 589-97.

39.   Serano, T. L. & Cohen, R. S. (1995) *Development* **121,** 3013-21.

40.   Broadus, J., Fuerstenberg, S. & Doe, C. Q. (1998) *Nature* **391,** 792-5.

41.   Mansfield, J. H., Wilhelm, J. E. & Hazelrigg, T. (2002) *Development* **129,** 197-209.

42.   Dreyfuss, G., Kim, V. N. & Kataoka, N. (2002) *Nat Rev Mol Cell Biol* **3,** 195-205.

43.   Chartrand, P., Meng, X. H., Huttelmaier, S., Donato, D. & Singer, R. H. (2002)
        Mol Cell **10,** 1319-30.

44.   Fusco, D., Accornero, N., Lavoie, B., Shenoy, S. M., Blanchard, J. M., Singer, R.
        H. & Bertrand, E. (2003) *Curr Biol* **13,** 161-7.

45.   Femino, A. M., Fay, F. S., Fogarty, K. & Singer, R. H. (1998) *Science* **280,** 585-90.

46.   Pederson, T. (2001) *Nucleic Acids Res* **29,** 1013-6.

47.   Wilkie, G. S. & Davis, I. (2001) *Cell* **105,** 209-19.

48.   Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M.
        B., Brown, P. O., Botstein, D. & Futcher, B. (1998) *Mol Biol Cell* **9,** 3273-97.

49.   Cho, R. J., Campbell, M. J., Winzeler, E. A., Steinmetz, L., Conway, A.,

Wodicka, L., Wolfsberg, T. G., Gabrielian, A. E., Landsman, D., Lockhart, D. J.

& Davis, R. W. (1998) *Mol Cell* **2,** 65-73.

Chapter 3:

Unbiased Selection of Localization Elements Reveals

cis-acting Determinants of mRNA Bud-Localization in

Saccharomyces cerevisiae

Author contributions:

This chapter is a reprint from the following reference:

Ashwini Jambhekar, Kimberly McDermott, Katherine Sorber, Kelly A. Shepard, Ronald D. Vale, Peter A. Takizawa, Joseph L. DeRisi. (2005). Unbiased selection of localization elements reveals cis-acting determinants of mRNA bud-localization in *Saccharomyces cerevisiae*. PNAS **102**(50): 18005-18010.

Ashwini Jambhekar obtained all data in figures 1, 2, 3a, b, d, 4, 5, 6, 7, 8, 11. Kimberly McDermott and Peter Takizawa obtained data in figures 3c and 9. Katherine Sorber obtained data shown in figure 11, and assisted with the experiment in figure 5a. Ashwini Jambhekar obtained data listed in Tables 1-4, 6-8. Katherine Sorber obtained all data in Table 5, and collected 20% of the data listed in Table 4. Kelly Shepard tested 2 of the candidate zipcodes in mentioned in the Results. Ronald D. Vale and Joseph L. DeRisi supervised research.

Joseph L. DeRisi, Thesis Advisor

Abstract:

Cytoplasmic mRNA localization is a mechanism used by many organisms to generate asymmetry and sequester protein activity. In the yeast *Saccharomyces cerevisiae*, mRNA transport to bud tips of dividing cells is mediated by the binding of She2p, She3p, and Myo4p to coding regions of the RNA. To date, twenty-four bud-localized mRNAs have been identified, yet the RNA determinants that mediate localization remain poorly understood. Here, we utilized Nonhomologous Random Recombination (NRR) to generate libraries of sequences that could be selected for their ability to bind She-complex proteins, thereby providing an unbiased approach for minimizing and mapping localization elements in several transported RNAs. Analysis of the derived sequences and predicted secondary structures revealed short sequence motifs that mediate binding to the She-complex as well as RNA localization to the bud tip *in vivo*. A predicted single-stranded core CG dinucleotide appears to be an important component of the RNA-protein interface although other nucleotides contribute in a context-dependent manner. Our findings further our understanding of RNA recognition by the She-complex, and the methods employed here should be applicable for elucidating minimal RNA motifs involved in many other types of interactions.

Introduction:

Localization of mRNA is commonly employed to target proteins to specific regions within a cell. In most cases, this process requires recognition by RNA-binding protein(s) and linkage of the resulting RNA-protein complex directly or indirectly to molecular motors (1). The determinants of recognition, transport factor binding, and subsequent targeting are cis-acting sequences often found in untranslated regions. Precise characterization of these RNA "zipcodes" has proven to be cumbersome for several reasons. The reported length of the minimal sequence requirements for transport ranges from 50 nucleotides (nt) to several hundred, and this apparent complexity is compounded by functional redundancy among zipcodes and a diversity of cellular recognition components (2-5).

The yeast *Saccharomyces cerevisiae* provides a tractable model system to characterize the determinants of zipcode recognition. To date, twenty-four bud-localized mRNAs have been identified, and coding regions were shown to mediate transport (6). Localization is dependent on the She-complex, which comprises She2p, a putative RNA binding protein, Myo4p, a type V myosin motor, and She3p, which interacts directly with both Myo4p and She2p (7-9).

Independent studies of one transported RNA, *ASH1*, identified three (N, C, U) (10) or four (E1, E2A, E2B, E3) (11) zipcodes based on their ability to mediate localization of a reporter. Only one of the elements lies in the 3' UTR; the remaining are located within the coding region. These elements bear no obvious primary sequence or secondary structural similarity to each other, and mutational analysis suggested that

54

secondary structure was required for activity (10, 11). Recently, Olivier et al. (12) reported that a CGA triplet in a loop, along with a single-stranded cytosine six bases away and opposite to the triplet, was necessary for bud-localization of *ASH1* and two other RNAs. However, these criteria are insufficient to identify zipcodes in other RNAs localized by the She-complex (6).

To extend our understanding of the She-complex-RNA interaction, we employed an unbiased approach to select zipcode-containing fragments from pools of known localized RNAs. The fragments were tested for localization *in vivo* and *bona fide* zipcodes were subjected to further analysis, which revealed a highly-degenerate motif predicted to lie in single-stranded regions and necessary for She-complex-dependent transport. Highlighting the complexity of the She2/3p-RNA interaction, we also found that the precise sequences mediating recognition and transport depend upon the context of the adjacent sequence and structural features in the mRNA.

Materials and Methods:

Nonhomologous Random Recombination (NRR):

NRR was carried out as described (13). Briefly, 1-4 µg of DNA encoding She-complex targets was PCR amplified from S288c genomic DNA or plasmid clones using Pfu DNA polymerase (Stratagene) and digested with 0.1-1 U DNaseI (Gibco) at room temperature for 5 min. DNA fragments 20-200 bp in length were size selected by agarose gel electrophoresis and purified by electroelution. Fragments were blunted with 9U T4 DNA Polymerase (NEB) and 200 µM dNTPs at 12°C for 30 min as described by

manufacturer. DNA was purified by phenol-chloroform extraction and sepharose gel

filtration. 5-15% of resulting DNA was used for ligation with 30 pmol of 5'

phosphorylated T7hairpin

(AAACCCTATAGTGAGTCGTATTAGTTTAAACGGCCCGCGCGGGCCGTTTAAA

CTAATACGACTCACTATAGGGTTT), or with a combination of 15 pmol each of 5'

phosphorylated XmaHairpin1

(AAACCCGGGCCTGACTCCGAAGTCGTTTAAACGGCCGCGCGGCCGTTTAAAC

GACTTCGGAGTCAGGCCCGGGTTT) and SphHaripin1

(AAACGCATGCCTGACTCCGAAGTCGTTTAAACGGCCGCGCGGCCGTTTAAAC

GACTTCGGAGTCAGGCATGCGTTT). Ligated DNA was digested with *Pme*I to

remove hairpin ends and hairpin dimers. 10% of the restriction digest was used for PCR

with 1 μM of XmaT7 primer (TCGACCCGGGTAATACGACTCACTATAGGG) or 1

μM of NRRprimer1 (AAACGACTTCGGAGTCAGG) using Pfu DNA polymerase

(Stratagene). Reactions were denatured at 94°C for 1 min and then cycled 35 to 40 times

as follows: 94°C 20 sec, 52°C 30 sec, 68°C 1:30. PCR products were digested with 20 U

*Xma*I (XmaT7 products) or 20 U *Xma*I plus 20 U *Sph*I (NRRprimer1 products) overnight.

NRR was carried out with T7hairpin and *ASH1, YLR434c, ERG2,* or *MID2* sequences

separately. XmaHairpin1 and SphHairpin1 were used for separate NRR reactions with

CPS1, DNM1, WSC2, MMR1, or *YGR046w*, or with a pool of *ERG2, MID2*, and bp 1-

1000 and 1500-1761 of *TPO1*. Bp 1000-1500 of *TPO1* were excluded because they do

not to contain any localization sequences (AJ and JLD, unpublished results). The full

coding region of each gene was used for NRR (unless noted), except for *ASH1*, which

56

included the coding region plus 99 bp of downstream sequence which contains a localization signal (10, 11).

3-Hybrid RNA-Expression Library Construction:

NRR products were ligated into the *Xma*I site (for XmaT7 products) or asymmetrically into *Xma*I and *Sph*I sites (for NRRprimer1 products) of pIIIΔA/ MS2.2. This vector consists of pIIIA/MS2.2 (14) with a deletion of the *Aat*II-*Tth*111-I fragment encoding *ADE2*. In all cases, the library size was sufficient to ensure that every sequence was represented at least once. Separate libraries were constructed for *YLR434c* and *ASH1*. Another library contained NRR products derived from a pool of *ERG2, MID2*, and bases 1-500 and 1500-1761 of *TPO1*. A fourth library contained *ERG2, MID2, WSC2, DNM1, CPS1, YGR046W, MMR1, YMR171C, SRL1* NRR products. Each library was screened separately by 3-Hybrid analysis (15).

For randomization experiments, complementary oligonucleotides fully degenerate at the indicated positions were annealed and cloned into the *Not*I and *Xho*I sites of pAJ232, which consists of pIIIΔA/MS2.2 with an insertion of *Not*I and *Xho*I in the *Xma*I site. Plasmid library members were selected at 10mM 3-AT as described below.

3-Hybrid selection:

DNA encoding the carboxyl-terminus of She3 (bp 706-1278) was cloned into *Xma*I/ *Sac*I sites of Gal4-AD expression vector pACT2. The *ADE2* ORF with 500 bp of upstream sequence and 300 bp downstream sequence was cloned in the *Not*I site of

57

pACT2, providing an additional marker. The resulting plasmid was introduced into the 3-Hybrid L40 coat host strain (15). Where indicated, *SHE2* was deleted in L40 coat as described (16).12-20 µg of RNA plasmid libraries was transformed into the She3-L40 coat strain using the lithium acetate method (17). Transformants were plated on SD-HIS-URA medium containing 6.67mg/ L adenine and 0, 0.5, 1, 5, 10 or 15mM 3-aminotriazole (3-AT). White transformants represent candidates which require She3 for expression of the *HIS3* reporter. Red transformants represent candidates which have lost the *SHE3* plasmid but express *HIS3* in the absence of any RNA-protein interaction, and accounted for <15% of the transformants selected at or above 5mM 3-AT. 30-80 white colonies growing at the highest 3-AT concentrations were tested for expression of LacZ by X-gal filter assay (14). All tested candidates expressed LacZ (data not shown). Plasmids were rescued and inserts fully sequenced from the 5' end.

Quantitative ß-galactosidase assays were performed as described (14), except that cells were lysed with Yeast Protein Extraction Reagent (Pierce).

Visualization of RNA:

The U1A-GFP system was used for visualizing RNA localization *in vivo* (6, 8). RNAs longer than 150 nt were cloned directly into the pGAL-U1A vector (6) containing *NotI* and *XhoI* cloning sites. Shorter RNAs were assayed by fusing to the 3' end of the unlocalized *ADH1* gene. For RNAs shorter than 75 nt, a linker containing a 13 bp inverted repeat separated by *NotI* and *XhoI* sites was inserted downstream of *ADH1*. Synthetic oligos (Operon) encoding the target RNA sequences were ligated into the *NotI*

and *XhoI* sites, so that the RNA was expressed with flanking inverted repeats which formed a stable helix.

For visualization of RNA, the pGAL-U1A plasmid containing the RNA of interest was introduced into a W303 yeast strain harboring the U1A-GFP plasmid (6, 8). >50 premitotic cells expressing RNA were counted from 2 independent transformants for each RNA as described (6).

RNA structure predictions:

All RNA structure predictions were computed using MFOLD (18, 19)

Protein purification and gel shifts:

She2p-HA contains a single HA epitope at its C-terminus. She2p-HA was overexpressed in *S. cerevisiae* and isolated from cell extracts with anti-HA antibodies coupled to protein A sepharose (Sigma). She2p-HA was eluted from the resin with excess HA peptide, dialyzed to remove free peptide and concentrated in a Microcon YM-10 (Millipore). His-She3p 251-425 contains a His_6 tag at the N-terminus of amino acids 251 -425 of She3p. His-She3p 251-425 was expressed in BL21 RIPL (Stratagene) and purified with Ni-NTA agarose (Qiagen) according to the manufacture's instructions. To generate ^{32}P -labeled RNAs for mobility shifts, annealed oligos containing a T7 promoter followed by a particular zipcode sequence were used as templates in an *in vitro* transcription reaction. The oligo templates were added to a Maxiscript T7 (Ambion) reaction containing UTP-^{32}P (Amersham). Full length RNAs were gel purified from the

59

reactions. Each gel shift reaction contained 0.5 nM labeled RNA, 0.1 mg/mL tRNA in 25 mM Hepes-KOH pH 7.5, 100 mM KCl, 2 mM MgCl$_2$, 1 mM DTT. Purified She2p-HA and His-She3 251- 425 were added at varying concentrations. Reactions were incubated at room temperature for 30 min and then run on a 5% acrylamide gel (37.5:1) in TBE at 4°C. The gel was fixed, dried and exposed to film.

Results:

Identification of She-complex dependent localization sequences:

We sought to identify short zipcodes from known transported RNAs in a high-throughput manner without making assumptions about exact zipcode length, orientation, or connectivity. For this reason, we utilized Nonhomologous Random Recombination (NRR) (13) to generate libraries of sequences that could be selected for their ability to bind to She-complex proteins. We reasoned that the region of overlap of multiple, independently-selected clones would define a short zipcode.

To generate a library by NRR, DNA encoding a target RNA was digested with DNaseI, and 20-200 bp fragments were isolated and ligated in the presence of hairpin linkers to generate products containing 1-3 tandem fragments of various sizes and connectivities flanked by hairpins. The products were PCR amplified with primers complementary to the linker sequence and selected for interaction with the She-complex by 3-Hybrid assay (Fig. 1a). As bait, we used the carboxyl-terminus of She3p, which interacts with She2p (7) and displays proper specificity for RNA targets (9) (vector and IRE controls, Fig. 2). For the two RNAs tested, the 3-Hybrid interaction also required

endogenous She2p (*she2* WSC2N and *she2* Umin, Fig. 2), indicating the formation of a tripartite RNA-protein complex.

To validate this approach, we subjected *ASH1* to NRR and 3-Hybrid selection. Sequencing of NRR-generated clones prior to selection revealed fragments derived from various parts of the gene (Fig. 1b). After selection, almost all clones fell within previously identified localization elements (Fig. 1c). Although no sequences were recovered from E2A, this zipcode is active in the 3-Hybrid system (12); therefore its absence in our selection most likely resulted from insufficient sequencing of positive transformants. Only one selected clone did not contain a fragment overlapping known localization elements and was not pursued further. In all cases, the sequences defined by selected overlapping clones were shorter than the zipcodes from which they were derived (10, 11). To verify that the shorter sequences localized *in vivo*, we used the U1A-GFP system (6, 8) to visualize RNA distribution in live cells. Sequences shorter than 150 nt in length were fused to the 3' end of *ADH1* and assayed for their ability to direct bud localization of the RNA. All *ASH1* sequences defined by the NRR/ 3-Hybrid selection localized to bud tips in >90% of cells (Table 1, Fig. 3 c-e insets).

Ten other genes encoding localized RNAs were screened in this manner individually (*YLR434c*) or in pools (*ERG2, MID2, TPO1, WSC2, MMR1, SRL1, CPS1, DNM1, YGR046w*), and ten more putative zipcodes were identified ranging from 50 (YLR434-1) to 201 (DNM1N) nt in length (Table 1). All sequences defined by overlapping clones were tested for localization *in vivo*. Although the control *ADH1* reporter was localized in only 20% of cells, our experience with testing various constructs

has revealed that, in rare cases, unlocalized RNAs can produce dim, bud-localized particles in up to 60% of cells in a She2p-independent manner. Thus, we classified any RNA that was localized in fewer than 60% of cells as unlocalized. Only one selected RNA, CPS1CR, failed to localize by this criterion. Of the remainder, nine sequences localized in >90% of cells in a She2p-dependent manner (Table 1, data not shown). Two others, TPO1N and DNM1N, localized less efficiently (in 70-80% of cells). In general, sequences recovered multiple times at high 3-AT concentrations were more likely to localize than those recovered once or only at low 3-AT concentrations (Table 1, 2). Although some zipcodes were recovered numerous times, we failed to recover any zipcodes from CPS1, MID2, MMR1, or YGR046w, suggesting that the screen was not saturating.

Identification of a conserved She2/3p-dependent localization motif:

We used MEME analysis, which identifies statistically over-represented sequence motifs within a data set (20), to find any motifs shared by the newly-identified zipcodes. The data set consisted of the nine zipcodes displaying >90% localization activity, including two (WSC2N and YLR434-2) which had been minimized by deletion mapping (Fig. 3). Of several candidates, one degenerate motif (RCGAADA) was present in all input sequences and mapped almost exclusively (in seven out of eight cases) to single-stranded regions of the secondary structures predicted by MFOLD (18, 19) (Fig. 3). One zipcode, WSC2N, displayed two copies of the motif--a more degenerate version in the terminal loop and a consensus sequence in the 3' bulge (Fig. 3a). Additionally, seven

zipcodes contained an adenosine six bases upstream of the motif. This sequence pattern was observed in 3 other zipcodes not included in the MEME analysis (E2A in *ASH1* and zipcodes in *IST2* and *YMR171c* (12)).

Five zipcodes were selected for further analysis based on the fact that the 7-base motif could be mutated or deleted in these RNAs without affecting the predicted structure of the remainder of the molecule (Fig. 3a-e, Table 3). Wild-type (WT) zipcodes localized in >90% of budded cells (Fig. 4), and displayed ß-galactosidase activities above 200 Miller Units (Fig. 2). All zipcodes required the motif for localization and LacZ expression (Fig. 4a, b, d). Deletions or mutations of the motif in E1min, E2Bmin, and YLR434-2 abolished activity in both assays. Deletion of the motif in Umin also abolished localization, but decreased ß-galactosidase activity by only 65% (Fig. 4a, b). WSC2N, which contains two copies of the motif, required mutations in both to abolish localization and ß-galactosidase activity (Figs. 4a, b, 5).

The ability of purified She2p and the carboxyl terminus of She3p (251-425) to bind WT and mutant zipcodes directly was also tested by RNA mobility shift. Nanomolar concentrations of She2p and She3p retarded the mobility of all WT zipcodes, indicating that She2/3 bind directly to each zipcode (Fig. 4c, 6). Furthermore, the protein complex displayed sequence-specific binding, as mutations of the motif in Umin, YLR434-2, E2Bmin, and E1min decreased or abolished the shift (Fig. 4c, 6). Although a large amount of WT RNAs remained unbound at the highest protein concentrations, it is unlikely that additional proteins facilitate She-complex binding to RNA *in vivo*, as a limited number of proteins, like She2p (21), are present in both the nucleus and

63

cytoplasm to facilitate bud-localization and 3-Hybrid activity. It is more likely that some

of the RNA misfolds and cannot bind She2/3 *in vitro*. Nevertheless, we conclude that the

degenerate motif is essential for RNA binding of She2/3, and that activity in localization

and ß-galactosidase assays reflects binding of the RNA to the She-complex.

In addition to the recognition motif, MEME analysis identified an adenosine six

bases upstream in seven zipcodes. Mutation or deletion of this base caused varying

effects on 3-Hybrid activity ranging from an increase (YLR434-2min) to a 5-fold

reduction (E2Bmin) (Fig 7a). Some base substitutions may be more favorable than others

at this position, resulting in the range of phenotypes displayed by the mutations in

different zipcodes. Although this adenosine was highly conserved among the zipcodes, its

contribution to binding was context-dependent.

While the mutational analyses revealed that the primary sequence of the motif

was essential for zipcode activity, they did not address the structural requirements for

She2/3 recognition. To determine whether the single-stranded nature of the motif was

necessary for She-complex recognition, the 5' end of YLR434-2 was changed to

complement the motif at the 3' end, thus placing the motif in a predicted duplex. The

resulting RNA (YLR434-2 double-stranded motif) failed to localize *in vivo* and did not

display significant 3-Hybrid ß-galactosidase activity (Fig. 4a, b), indicating that the She-

complex cannot bind its recognition site in a stable helix. We also observed that the

recognition motifs bordered predicted helices in most zipcodes. To test whether this

juxtaposition was essential, two nucleotides were inserted between the stems and motifs

of four zipcodes. The resulting mutant phenotypes ranged from no decrease in ß-

galactosidase activity (Umin) to a complete abolition of She-complex interaction (E1min, YLR434-2) (Fig. 7a).

Although the above results implied that the stems of zipcodes were important for She2/3 binding, no primary sequence similarities were observed in these regions. The current models (10-12) proposed that stems play only a structural role in the RNA-protein interaction. In support of the model, compensatory mutations in the stem of YLR434-2 preserved zipcode function (Fig. 8); but similar mutations in E2Bmin abolished 3-Hybrid activity (Fig. 7b). Therefore, each base pair in the E2Bmin stem was individually mutated in order to identify essential bases. Mutation of each of the two base pairs adjacent to the loop decreased ß-galactosidase activity 2- to 4-fold, while mutating the pair at the base of the stem had no effect. (The $C_{1284} \bullet G_{1300}$ pair was not tested because substitutions were predicted to disrupt the entire stem). Surprisingly, no single base-pair mutation decreased activity to the extent that mutation of the entire stem did. These results indicated that the primary sequence of the stem contributes to She2/3 binding in some cases, and that bases in the stem of E2Bmin contribute in an additive manner. Collectively, these results support the role of the degenerate, single-stranded motif in mediating She-complex recognition; however, the precise sequence and topological requirements appear to be context-dependent.

Analysis of base contributions within a single zipcode:

Because it appeared that conserved bases in the recognition motif as well as other, less-conserved bases contributed to She2/3 binding, we investigated in detail the

65

sequence requirements for She2/3 binding to a single zipcode. Four- to seven-base regions of a further-minimized E2Bmin zipcode were fully randomized, and the resulting sequences were selected for She-complex binding by the 3-Hybrid system.

The contribution of each base in the loop of E2Bmin was determined via two separate, overlapping randomization/ selection experiments. One position in the loop (1288) displayed no base preferences for She-complex recognition (Fig. 9b). In contrast, 6 out of 7 bases in the WT motif were significantly over-represented upon selection (Fig 9a), but the importance of each base within the motif for She-complex recognition appeared to vary. The 5' $A_{1291}CG$ triplet was highly over-represented in the selected clones, while a lesser bias towards adenosines at the 3' end was detected (Fig. 9a, Table 4). In support of these observations, mutation of the guanosine (G1293C) in the context of a selected E2Bmin clone ($A_{1291}CGUUUU \rightarrow ACCUUUU$) decreased activity 10-fold (data not shown). The motif randomization was repeated in zipcode YLR434-2, and although similar results were obtained, the strength of the base preferences varied at some positions (Fig. 10, Table 5). Surprisingly, the strength of the bias for $C_{1292}G$ varied even between the two overlapping E2B experiments (Fig. 9b, Table 6), indicating that the requirements for She-complex binding are influenced by the variability of the surrounding region. We noticed that most selected sequences were predicted to form the same secondary structure as WT E2Bmin. While the observed sequence biases may have resulted from structural constraints, the recovered clones represented only a small fraction of sequences predicted to form the same structure as the natural zipcode (data not shown), suggesting that secondary structure alone cannot mediate She2/3p recognition.

66

In addition to the over-representation of bases in the recognition motif, we also detected a bias towards the adenosine at the 5' end of the loop (A_{1287}) and a stronger requirement for the $C_{1289}G$ dinucleotide upstream of the recognition motif (Fig. 9b, Table 6). Olivier et al. recently reported that the $C_{1289}GA$ triplet was essential for She2p binding (12); our results supported the importance of these bases as well as the downstream $C_{1292}G$. Taken together, our results show that a repeated CG dinucleotide promotes She-complex binding: the consensus sequence, by base frequency, of positions 1289-93 of E2Bmin was CGACG, and CGACGA was most frequently selected in the context of YLR434-2. However, the CG dinucleotide followed by adenosines occurs most frequently in natural zipcodes, and this pattern is sufficient for bud-localization.

The sequence and structural requirements in the stem of E2Bmin were also analyzed by randomization and selection. The bias towards base-pairing was strongest at the second position from the top of the loop, whereas the base of the stem was paired only somewhat more often than was expected at random (Fig. 9c). Although targeted mutatgenesis had revealed weak sequence preferences in the two loop-proximal base pairs, no biases were observed by randomization/ selection (Fig. 9c, Table 7), possibly because 3-AT selection does not discriminate between modest differences in 3-Hybrid activity (22). Surprisingly, we recovered a bias towards the $C_{1283}C$ dinucleotide in the 5' strand of the stem and a weaker bias for G_{1300} (Fig. 9c, Table 8). The bias towards this guanosine likely results from the need to base-pair with C_{1284}. These results further support our conclusion that stems can contribute both sequence and structural information for She-complex recognition.

Sequence requirements in the 3' tail were also revealed. Although Olivier et al. reported that C_{1302} was essential for She2 binding (12), only a modest bias towards this cytosine was detected (Fig. 9c, Table 8). 11 out of 12 clones that contained substitutions at this position had a UC dinucleotide immediately upstream, even though this pattern was not observed in native zipcodes lacking an analogous cytosine. It is apparent that the requirements for She-complex recognition are flexible, and that the cytosine described by Olivier et al. is not essential for all zipcodes.

Using the requirements elucidated by the mutational and randomization analyses, we sought to identify zipcodes in other localized RNAs. One candidate zipcode (bases 798-839 of *MID2*), which contains a single-stranded ACGAAAU motif adjacent to a stem and an adenosine 6 bases upstream, was localized above background levels (in 65-70% of budded cells), but less efficiently than other zipcodes isolated by 3-Hybrid assay. Candidate zipcodes in *IST2* and *BRO1*, however, failed to be localized above background levels (data not shown). Additionally, WSC2C was the only isolated zipcode that did not contain the recognition motif in a single-stranded region and required two stem-loops for WT activity (Fig. 11). These results suggest that RNA recognition by She2/3 is complex, and that the current knowledge of the binding requirements and/ or the prediction tools are insufficient for accurately identifying new zipcodes.

Discussion:

We have employed a high-throughput selection for mapping She-complex binding sites in RNA targets. This methodology uses NRR to prepare DNA encoding localized RNAs, followed by 3-Hybrid selection to identify small fragments containing binding sites. Unlike other *in vitro* evolution techniques, NRR does not alter WT binding sites, making it easier to deconvolute the sequences after selection. Secondly, NRR covers sequence space efficiently because every starting pool contains a She2/3-binding site, eliminating the need to sample every nucleotide at every position and thus generating positive results from low-complexity libraries. Unlike conventional deletion mapping approaches, NRR samples all orientations and connectivities of input sequences.

By subjecting the NRR-derived pool to an *in vivo* 3-Hybrid selection, we could recover potentially lower-affinity and lower-abundance library members which may be missed by *in vitro* SELEX-style selection or candidate mutagenesis approaches. At the same time, the 3-Hybrid selection resulted in a low rate of false positives, since higher-abundance library members did not have a significant selective advantage. Finally, the *in vivo* selection ensured that the She proteins retained any post-translational modifications that may be necessary for WT activity.

Complex sequence and structural features mediate She2/3 binding:

Initial analysis of the NRR-derived zipcodes revealed a conserved single-stranded, 7-base motif lying proximal to a duplex region. Targeted mutagenesis confirmed that the motif sequence was necessary in different zipcodes for RNA transport

and for direct binding to She2/3. The structural context of the motif was also important for She-complex recognition: positioning the motif in a duplex abolished activity, and increasing the distance between the motif and adjacent stem decreased activity in three out of four zipcodes. A simple sequence motif stabilized by surrounding secondary structure appears to be a common theme of many protein binding sites in mRNAs, e.g. the Smg binding site in *nos* RNA (23). The She2/3 recognition site defined in this work expands on the CGA triplet reported by Olivier et al. (12) by virtue of a larger set of zipcodes which allowed us to identify the more degenerate bases downstream of the triplet as part of the recognition site. An additional single-stranded cytosine defined by Olivier et al. does not appear to be essential for She-complex recognition, since several natural zipcodes do not contain this nucleotide.

Quantitative analysis (by randomization/ selection) of the nucleotide requirements for She-complex binding contributed to a more thorough description of the RNA-protein interaction. Nucleotides at the 5' end of the motif, particularly a CG dinucleotide, were most important for binding, while the 3' adenosines made a weaker contribution. All natural zipcodes contained an adenosine following the CG dinucleotide, and this base was strongly favored in 2 out of 3 randomization experiments, suggesting that it too plays a major role in binding. Bases outside of the conserved motif also facilitated She-complex binding: some bases in the stem and 3' tail of E2Bmin were over-represented in the selected clones even though these sequences were not present in other zipcodes and one zipcode (YLR434-2) did not contain essential stem sequences.

The randomization/ selection experiments revealed an unexpected plasticity in the sequence requirements for She-complex recognition. When the four adenosines at the 3' end of the E2Bmin motif were held constant, there was only a weak bias for the upstream CG dinucleotide; but when these adenosines were allowed to vary, the CG dinucleotide was strongly required, suggesting that some motif bases can bypass the requirement for others. Surprisingly, the two CG dinucleotides in E2Bmin do not to function redundantly, as the requirement for the downstream CG was strongest when the upstream CG was invariable. A second example of sequence flexibility is that a UC dinucleotide can suppress mutations of a downstream cytosine identified by Olivier et al. (12) as essential for She2p binding. Some of these context-dependent effects may result from the RNA adopting a sub-optimal fold upon binding She2/3. The extensive sequence and structural plasticity, however, suggests that the She-complex recognizes a precise three-dimensional structure in its target RNAs—the complex may bind specifically to the key CG dinucleotide, with the surrounding bases simply maintaining the required structure.

One goal of defining a minimal RNA motif is to generate a predictive model whereby zipcodes could be identified in other RNAs *in silico*. We found that the core motif appears in She2/3 targets as well as in other RNAs known not to be localized, confirming that the motif alone does not confer specificity to the RNA-protein interaction. When the motif as well as other accessory features (e.g. an upstream adenosine and/ or a cytosine six nucleotides away from the motif) was used to identify new zipcodes, many localized RNAs did not contain any sequences that fit these criteria.

71

From our analyses of known zipcodes, we conclude that RNA recognition likely involves complex structural features which cannot be appreciated using current tools of searching linear sequences and prediction of secondary structures. Thus, accurate prediction of zipcodes in other localized RNAs awaits a three-dimensional structure of the She-complex bound to a target RNA as well as methods for predicting this structural fold in other RNAs. Meanwhile, the combination of NRR and 3-Hybrid selection provides a rapid and accurate way to isolate *bona fide* localization signals, and additional minimized zipcodes will aid in elucidating the range of sequences/ structures bound by the She-complex.

Table 1: Summary of elements identified by NRR/ 3-Hybrid selection.

Zipcode	Coordinates	Length (nt)	3-Hybrid Activity	#times recovered	%Localized
*E1min	635-683	49	+++	8	>90
*E2Bmin	1279-1314	36	+++	11	>90
*Umin	1766-1819	54	+	2	>90
*other	1684-1719R	36	++	1	N/D
WSC2N	418-71	54	++	14	>90
WSC2C	1313-84	72	++	6	>90
ERG2N	180-250	71	++	24	>90
DNM1N	605-805	201	+	1	70-80
DNM1C	1656-1752	97	+	1	>90
SRL1C	419-596	178	+	6	>90
YLR434-1	[21-55][195-209]	50	+	15	70-80
YLR434-2	[138-186][56-90]	76	+	11	>90
TPO1N	2-178	177	±	6	70-80
CPS1CR	1305-1456R	152	+	1	<60

Coordinates indicate the smallest overlapping fragment common to all sequences isolated for each zipcode. Nucleotides are numbered with from the adenosine of the start codon as +1. * sequences derived from *ASH1*. When multiple fragments were contained in one clone, the fragments are listed in 5' to 3' order. Fragments in italics were cloned in the antisense orientation. The length of each clone is given in nucleotides. Activity in the 3-Hybrid assay was assessed by highest 3-AT concentration at which the sequence was recovered. ±= 1mM, + = 5mM, ++ = 10mM, +++ = 15mM 3-AT. Also shown is the number of recovered clones containing the indicated sequence. %Localized refers to the percent of cells with exclusively bud-localized RNA. N/D: not determined.

Table 2: Coordinates of all clones isolated by 3-Hybrid selection, and number of times each clone was recovered from independent yeast transformants. Nucleotides are numbered with the adenosine of the start codon as +1. All isolates of SRL1C contained a deletion of base 458. All isolates of TPO1N contained a T80C mutation. "R" indicates that the fragment was recovered in the antisense orientation.

Gene	Element Name	Coordinates	# times recovered
ASH1	E1	611-87	1
		614-739	1
		620-87	2
		620-91	1
		624-87	1
		[1327-42R]-[624-87]	1
		635-83	1
	E2B	1210-1323	1
		1218-1314	1
		[1266-1314]-[354-402R]	1
		1267-1332	1
		1270-1332	1
		1273-1328	2
		1273-1338	1
		[1689-1727]-[746-788R]-[1276-1326]	1
		1279-1323	1
		1279-1332	1
	U	1750-1853	1
		[1-5]-[1766-1819]-[859-911R]	1
	other	1684-1719R	1
ERG2	ERG2N	133-299	1
		138-267	2
		139-367	1
		146-271	2
		146-328	1
		158-250	2
		158-269	1
		158-271	1
		158-289	1

74

Table 2, continued

Gene	Element Name	Coordinates	# times recovered
ERG2	ERG2N cont'd	158-291	1
		158-299	4
		158-328	1
		158-355	1
		160-256	1
		160-271	1
		160-303	1
		171-256	1
		180-328	1
WSC2	WSC2N	418-520	2
		412-486	1
		[1121-1161R]-[418-510F]	6
		415-486	1
		409-486	1
		415-484	1
		394-471	1
		388-487	1
	WSC2C	1313-1384	1
		1278-1384	2
		1278-1418	1
		1278-1391	1
		1354-1512	1
SRL1	SRL1C	419-596	2
		419-599	1
		419-597	1
		419-633	1
		419-598	1
DNM1	DNM1C	1656-1752	1
	DNM1N	605-805	1
CPS1	CPS1CR	1305-1456R	1
TPO1	TPO1N	2-178	6
YLR434	YLR434-1	[21-55R]-[194-209]	15
	YLR434-2	[137-186R]-[56-80R]	11

Table 3: Sequences of all WT and engineered mutant zipcodes analyzed in this work.

Bases identified by MEME analysis are in green, mutations in red, and deletions are indicated by dashes.

RNA	Sequence
E1min (WT)	AAAUACGCGAAGAAGUGGCUCAUUUCAAGCCAUUAAGUAUACCCAAACUC
E1Δmotif	----------------------------GUGGCUCAUUUCAAGCCAUUAAGUAUACCCAAACUC
E1shifted motif	AAAUACGCGAAGACCAGUGGCUCAUUUCAAGCCAUUAAGUAUACCCAAACUC
E1mutA	--AAUACGCGAAGAAGUGGCUCAUUUCAAGCCAUUAAGUAUACCCAAACUC
Umin (WT)	GAUACAUGGAUAACUGAAUCUCUUUCAACUAAUAAGAGACAUUAUCACGAAACA
UΔmotif	GAUACAUGGAUAACUGAAUCUCUUUCAACUAAUAAGAGACAUUAUC----------------
Ushifted motif	GAUACAUGGAUAACUGAAUCUCUUUCAACUAAUAAGAGACAUUAUCAUACGAAACA
UmutA	GAUACAUGGAUAACUGAAUCUCUUUCAACUAAUAAGAGACGUUAUCACGAAACA
WSC2Nmin (WT)	AGUUCAAAAACGUCCACGAAAUUGGACACGAAAACU
WSC2Nmut1	AGUUCAAAAACGUCCACGAAAUUGGACACCCCGGCU
WSC2Nmut2	AGUUCAAAAACGUCCACCUCUUUGGACACGAAAACU
WSC2Nmut3	AGUUCAAAAACGUCCACCUCUUUGGAC----CCCGGCU
YLR434-2min (WT)	GAUAUAGAUCCAAAGAAAUCUGCGAAAAAUUUU
YLR434-2Δmotif	GAUAUAGAUCCAAAGAAAUCUGAUAG----------------
YLR434-2mut stem1	GAUAGUCUACCAAAGAAUAGAUCGAAAAAUUUU
YLR434-2mut stem2	GAUAGUCUACCAAAGAAAUCUGCGAAAAAUUUU
YLR434-2mut stem3	GAUAUAGAUCCAAAGAAUAGAUCGAAAAAUUUU
YLR434-2 double stranded motif	AAAAUUUUUCGUAGAUCCAAAGAAAUCUGCGAAAAAUUUU
YLR434-2shifted motif	GAUAUAGAUCCAAAGAAAUCUGCCGCGAAAAAUUUU
YLR434-2mutA	GAUAUAGAUCCAAAGUAAUCUGCGAAAAAUUUU
E2Bmin (WT)	CCCTCCACACCGACGAAAAGUGGCAAGAUGAGAUCA
E2Bmut motif	CCCTCCACACCGUGCGUUCGUGGCAAGAUGAGAUCA
E2B flip stem	CCGAGGUGACCGACGAAAACACCCAUCAUGAGAUCA
E2B flip bp1	CCTGCACACCGACGAAAAGUGCCAAG
E2B flip bp3	CCTCCUCACCGACGAAAAGACGCAAG
E2B flip bp4	CCTCCAGACCGACGAAAACUGGCAAG
E2B shifted motif	CCTCCACACCGACGAAAAUAGUGGCAAG
E2BmutA	CCTCCGCACCGACGAAAAGUGGCAAG

76

Table 4: Sequences recovered after randomization of bases 1291-97 of E2Bmin and selection for interaction with She3p. Sequences of bases 1291-97 are shown; those recovered more than once are indicated.

Sequence	# times recovered	Sequence	# times recovered
ACGCTAA	x2	ACGTAGA	
AGAGTAC		ACGCGAT	x2
ACGCATT		AAGCACT	
ACGTCAC		ACCAGAA	x2
ACGAAGA	x2	ACGTAAT	x2
ACGCTTT		ACGCAAA	x2
ACGCAAC		ACGTGAC	x2
ACCTACG		ACGCACA	x2
ACGAAAC		ACGCGTA	
AAGTCTT		ACGAGAA	
ATGTGAA		ACGTCTT	
ACGAATG	x2	AAGTACA	
ACGCTTC	x2	ACGCGAC	
ACGTCAA	x2	AAGTCAA	
ACGAAAT		ACGTATC	
ACGTCTC		ACGCTCA	
ACGCAAT	x2	ACGATTC	
ACGCTAT		AAGCAAA	
ACGTATA		ATACTAA	
ACGTATT	x2	AAGTAAC	
AAGTAAA		ACGCCTT	
ACGTTTT			

77

Table 5: Sequences recovered after randomization of bases 23-28 of YLR434-2 and selection for interaction with She3p. Sequences of bases 23-28 are shown; those recovered more than once are indicated.

Sequence	# times recovered
CGAATA	
CGATAC	
CGAAGC	x2
CGATGA	
CGAAGT	x2
CGAACT	
CGTATC	x2
CGACGC	x2
CGACGA	x4
CAAATC	
CTACGT	x2
CGACTT	
CGAAGA	x2
CGCGAT	
CACATC	
ACAGAT	
CGAACA	
CGAGAT	x2
CACAAT	x2
CTGAAT	
CGACGT	
CTTAAT	
CAACGT	
CGAGGC	
CGTTAT	
CGATAA	

Table 6: Sequences recovered after randomization of bases 1287-93 of E2Bmin and selection for interaction with She3p. Sequences of bases 1287-93 are shown; sequences recovered more than once are indicated. WT sequence is also indicated.

Sequence	# times recovered
TCCGATT	x3
AACGACG	x6
GTCGATG	
ACCGATT	x2
CCCGACG	x3
ACGGAAT	
ACCGACG	x5 WT
AGCGAAG	x3
ATCGATG	
TGCGATC	
TGCGAAT	
AACGAAG	
GACGATT	
AGCGAAA	
AACGATG	x4
GTCGAAG	x2
ATCGACG	x3
AGCGGAA	
GACGAAT	
GACGACG	
GACGAAG	
AACGTAT	
AGCGAAT	
TGCGAAC	x2
GACGGCG	
AGCGACG	
AACGAAT	
TCCGACT	
ACCGAAA	
ATCGAAG	

Table 7: Sequences recovered after randomization of bases 1285-6 and 1298-9 of E2Bmin and selection for interaction with She3p. Each row represents a single selected clone; clones recovered more than once are indicated. Headings indicate the coordinate of each base; only bases at the randomized positions are shown

| 1285	1286	1298	1299	#times recovered
G	C	A	C	
A	T	T	T	
T	C	T	G	
C	T	A	G	x3
T	T	A	A	x3
C	A	T	G	x2
G	C	G	G	x3
C	A	C	T	
C	G	C	G	
G	C	G	C	
T	T	G	A	x2
C	T	G	T	
A	A	T	T	
G	G	C	C	
T	A	T	A	x2
A	C	A	C	
C	A	C	G	
C	T	T	G	
T	A	A	A	
T	A	C	A	
C	C	G	G	
G	C	G	T	
C	T	C	G	
A	G	C	T	
G	T	A	C	

Table 8: Sequences recovered after randomization of bases 1283-4 and 1300-03 of E2Bmin and selection for interaction with She3p. Each row represents a single selected clone; clones recovered more than once are indicated. Headings indicate the coordinate of each base; only bases at the randomized positions are shown. WT sequence is indicated.

1283	1284	1300	1301	1302	1303	#times recovered
A	G	C	T	C	A	
C	C	T	C	A	T	
C	C	T	C	G	C	x2
A	C	G	G	C	A	
A	C	G	G	C	T	
C	C	T	C	T	G	x2
C	T	A	G	C	A	x2
C	T	A	G	C	T	
T	C	G	A	C	A	
C	C	T	C	A	G	x2
C	C	T	T	C	T	
A	C	G	T	C	A	x2
A	G	G	G	G	G	
A	C	T	C	T	C	
A	G	A	T	C	G	
C	C	G	G	C	A	x4 WT
T	C	G	A	C	G	
C	C	T	T	C	C	
C	T	T	C	T	A	
C	C	C	T	C	T	
A	C	G	A	C	G	x2
C	C	T	C	T	T	
C	C	T	C	T	C	
C	C	G	G	C	G	x3
A	T	A	T	C	G	
G	A	T	C	C	G	
A	T	A	T	C	T	
T	T	A	A	C	A	
C	T	A	G	C	C	

81

Figure 1: 3-Hybrid scheme for selection of She3-interacting RNA fragments. (A)
Schematic of 3-Hybrid assay and representation of *ASH1* NRR library members (B) prior
to and (C) following 3-Hybrid selection. Each arrow represents a fragment from *ASH1*.
The direction of the arrowhead indicates whether the fragment is expressed in the sense
(right) or antisense (left) orientation from the 3-Hybrid RNA expression vector. The
position of each arrow corresponds to the location of the fragment within the gene, and
arrow colors indicate the connectivity of the fragments in the clone. Clones recovered in
more than one independent yeast transformant are indicated.
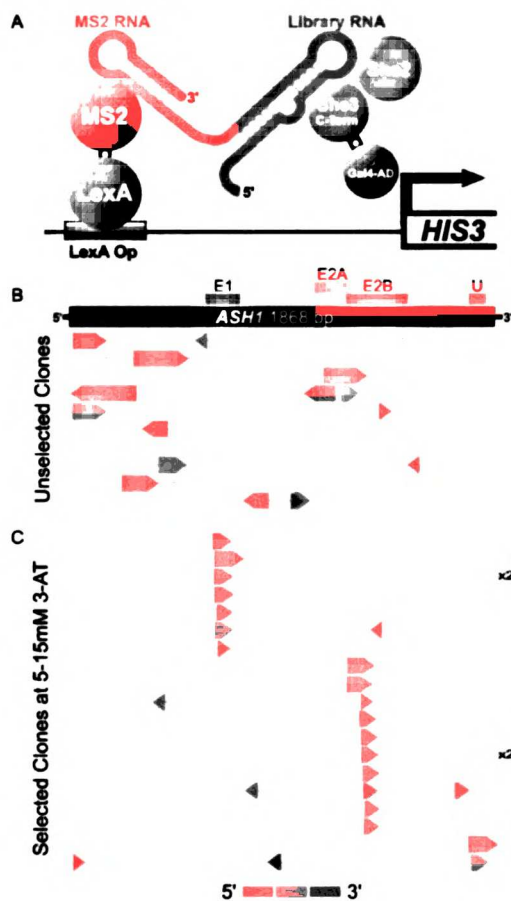
Figure 2: Interaction of RNAs with the carboxyl-terminus of She3p in the 3-Hybrid system. ß-galactosidase activities are shown for WT RNAs depicted in Fig. 2, as well as empty vector and IRE controls. *she2* indicates that the *SHE2* ORF was deleted in the 3-Hybrid host strain.
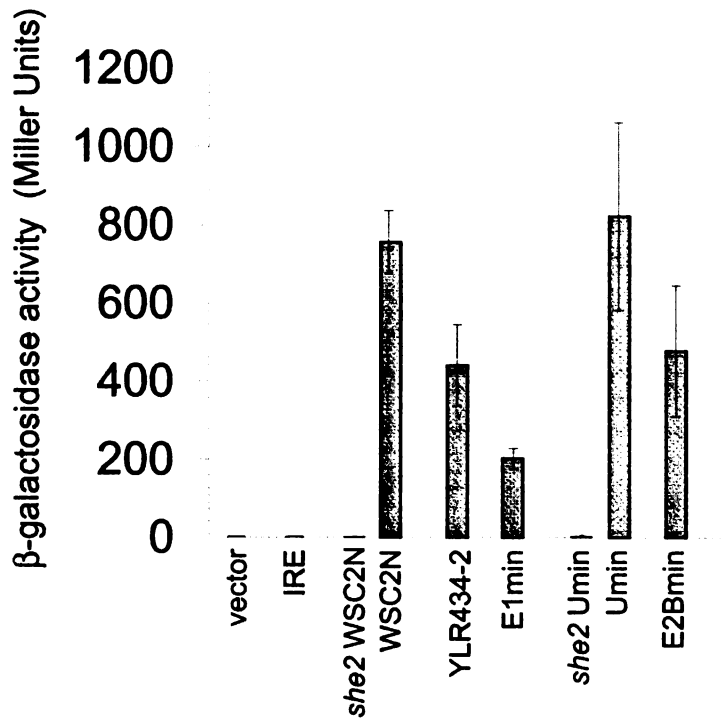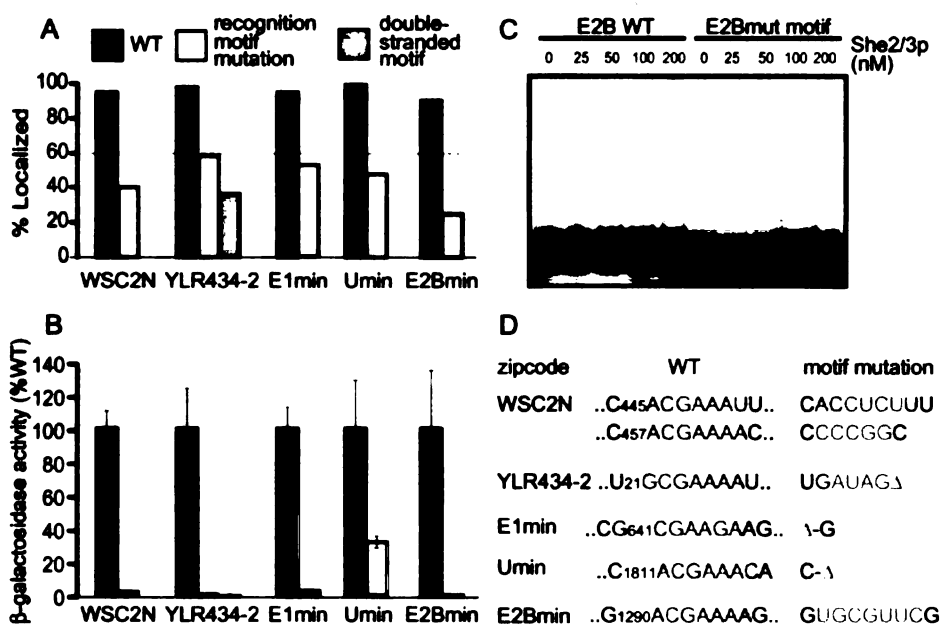
Figure 3: Sequences and predicted structures of fragments shorter than 100 nt isolated by NRR/ 3-Hybrid analysis. Bases identified by MEME analysis are green. (A) WSC2N, (B) YLR434-2, (C) E2Bmin, (D) E1min, (E) Umin, (F) ERG2N, (G) YLR434-1, (H) WSC2C, (I) DNM1C. Bases are numbered with the adenosine of the start codon as +1, with the exception of YLR434-1 and YLR434-2, which are numbered with the 5' base as +1. Insets contain representative GFP-RNA localization images. RNA particles are cytoplasmic; excess, unbound U1A-GFP is sequestered in the nucleus.
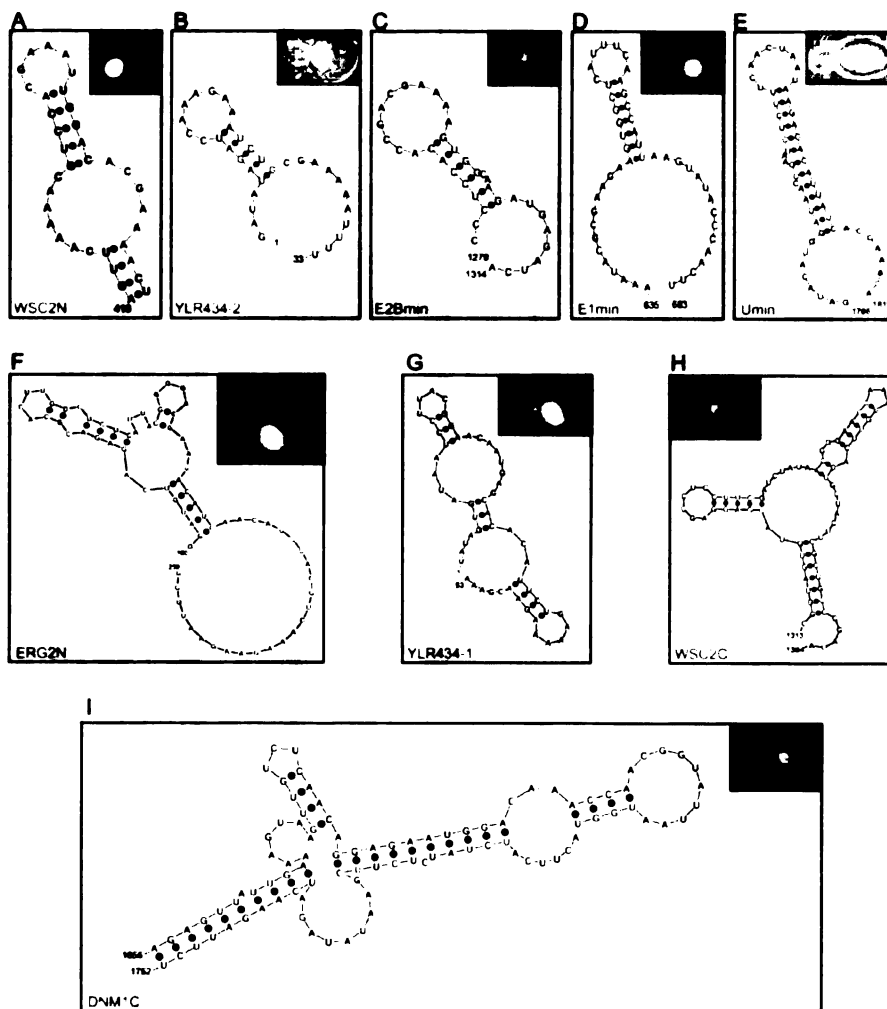
Figure 4: The recognition motif mediates She2/3-dependent localization and binding. (A) Localization ability and (B) 3-Hybrid ß-galactosidase activity of zipcodes in Fig. 2 containing WT bases or mutations in the recognition motif. "double-stranded motif" indicates that the recognition motif is in an ectopic duplex. Dashed line in (A) indicates the threshold below which RNAs were considered unlocalized. (C) *In vitro* binding of She2p and She3p to E2Bmin. RNA mobility shift assay consists of WT or mutant RNA lacking the recognition motif with increasing concentrations of purified She2p-HA and His-She3p carboxyl terminus. (D) WT and mutant motifs sequences used in a-c. Motif bases are green and mutations red.

D

| zipcode	WT	motif mutation
WSC2N	..C$_{445}$ACGAAAUU..	CACCUCUUU
	..C$_{457}$ACGAAAAC..	CCCCGGC
YLR434-2	..U$_{21}$GCGAAAAU..	UGAUAGΛ
E1min	..CG$_{641}$CGAAGAAG..	Λ-G
Umin	..C$_{1811}$ACGAAACA	C-Λ
E2Bmin	..G$_{1290}$ACGAAAAG..	GUGCGUUCG

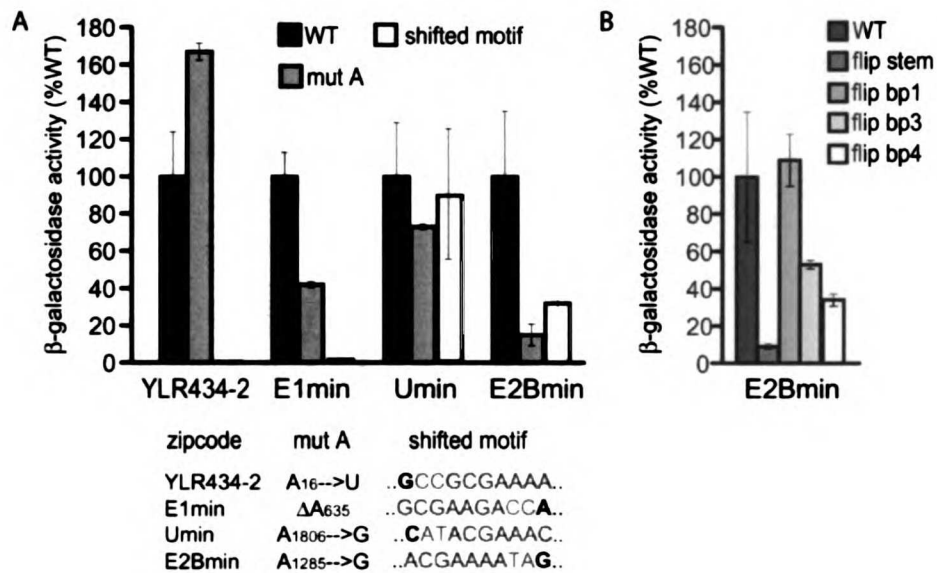**Figure 5: Two copies of the recognition motif in WSC2N are partially redundant.**

(A) Sequences and predicted secondary structures for WT and mutant WSC2N RNAs tested for (B) localization and (C) 3-Hybrid ß-galactosidase activity. In (A), bases identified by MEME analysis are green and mutations in red. Insets contain representative GFP-RNA localization images. RNA particles are cytoplasmic; excess, unbound GFP is sequestered in the nucleus.
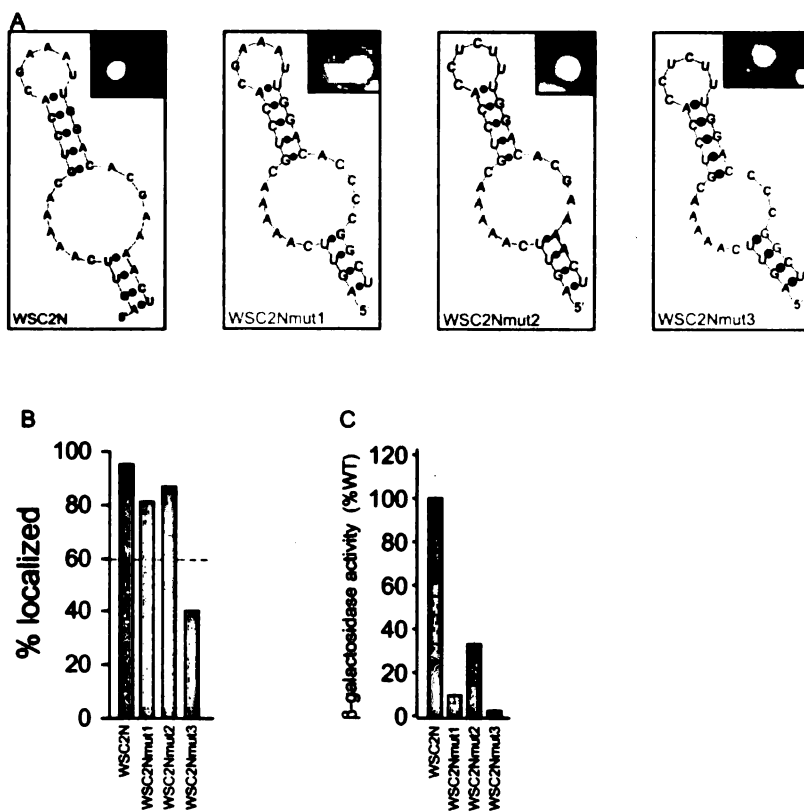
Figure 6: Gel mobility shift assays as described in Figure 6 using WT and recognition motif mutant RNAs. (A) E1min, (B) Umin, (C) YLR434-2.
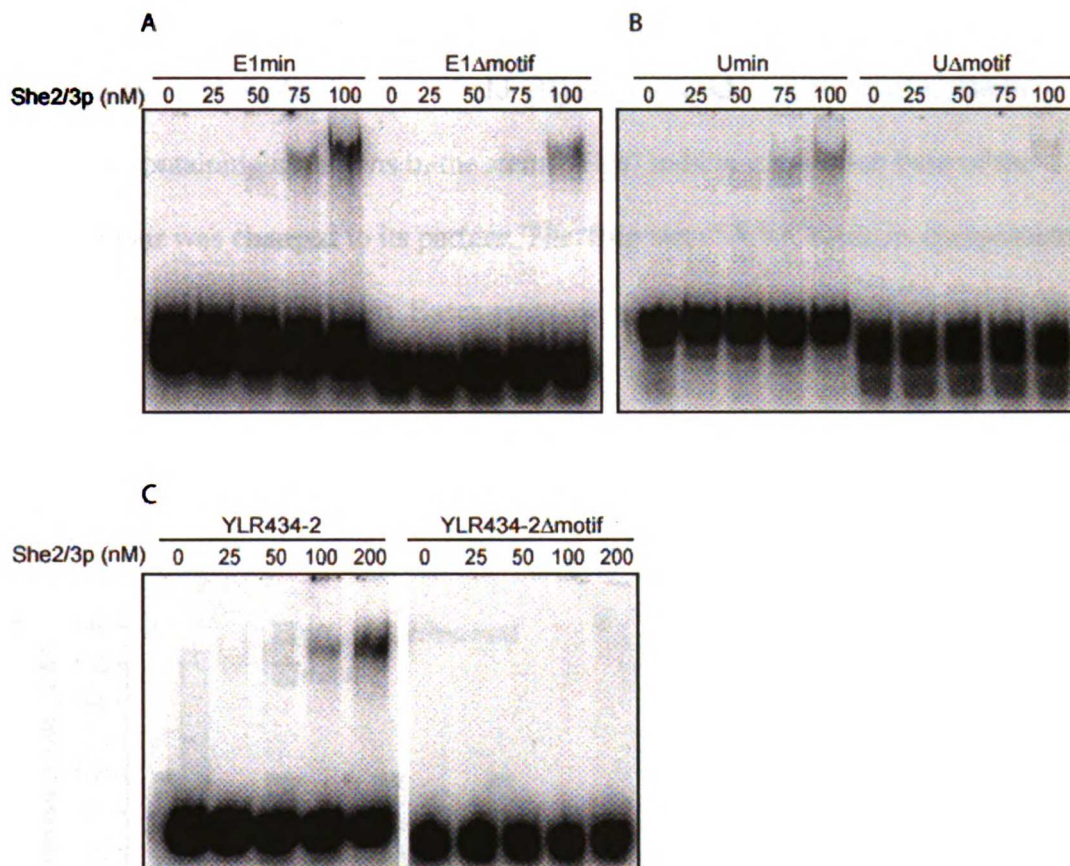
Figure 7: Context-dependency of recognition motif is revealed by mutational analysis. (A) 3-Hybrid ß-galactosidase activity of zipcodes bearing mutations in the upstream adenosine (mutA) or 2 nt insertions between the motif and adjacent helix (shifted motif). Mutations are defined for each zipcode. Motif bases are in green, insertions red, and duplex bases are bold. (B) ß-galactosidase activity of E2Bmin sequences containing mutations in the stem. "Flip" indicates that each base of the indicated pair was changed to its partner. The "flip stem" RNA contains compensatory mutations along the entire stem. Bases-pairs are numbered with the bottom of the stem as 1.

| zipcode	mut A	shifted motif
YLR434-2	$A_{16} \rightarrow U$	..**G**CCGCGAAAA..
E1min	ΔA_{635}	..GCGAAGACC**A**..
Umin	$A_{1806} \rightarrow G$	..**C**ATACGAAAC..
E2Bmin	$A_{1285} \rightarrow G$	..ACGAAAATA**G**..

Figure 8: ß-galactosidase activities of WT YLR434-2 and sequences containing mutations in the stem. In YLR434-2mut stem1, the sequence of each strand was exchanged with that of the opposite strand, preserving all base pairs. In YLR434-2mut stem2, the 5' strand of the stem was changed to its complement. The 3' strand was similarly mutated in YLR434-2mut stem3 (see Table 3).
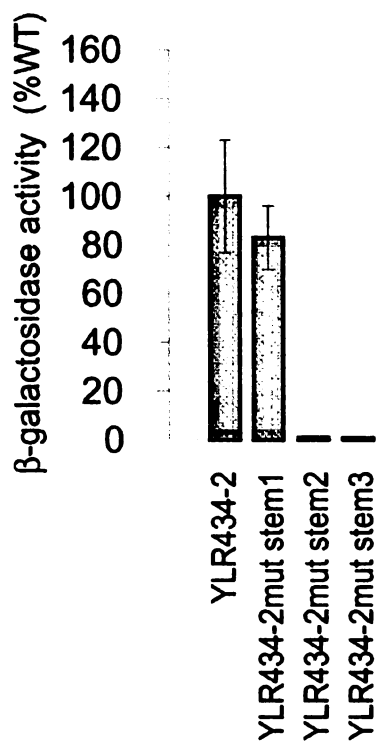
**Figure 9: Predicted secondary structure for E2Bmin and sequence logos derived from randomization and 3-Hybrid selection** of bases (A) 1291-7, (B) 1287-93, or (C) 1283-6 and 1298-1303. The height of each letter is proportional to the fraction of the observed frequency relative to the expected frequency at each position (24, 25). The color of each dot in (C) indicates the frequency of base-pairing among the selected clones.
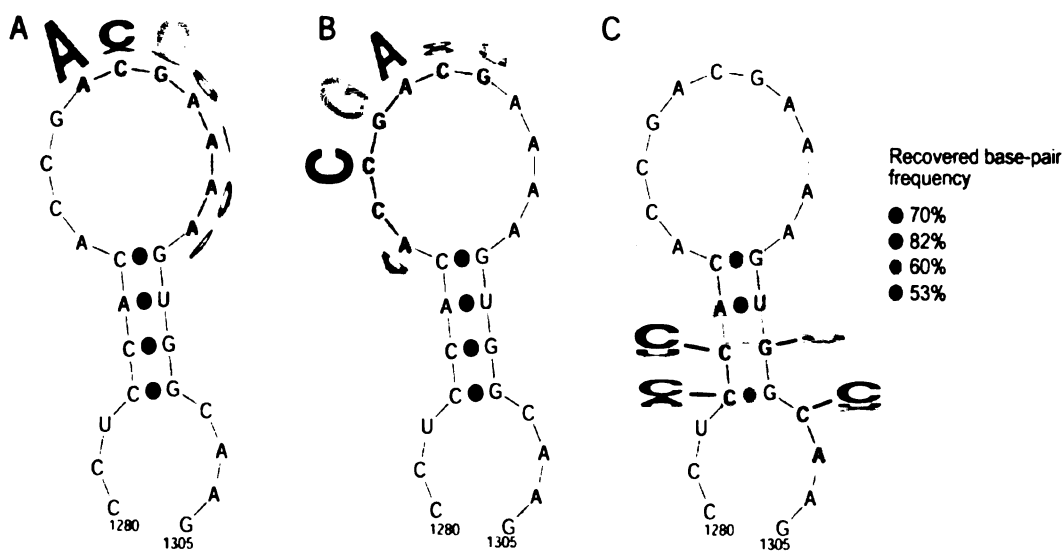
Figure 10: Randomization and 3-hybrid selection of recognition motif in YLR434-2 reveals sequence requirements for She2/3 recognition. Sequence and predicted structure of YLR434-2, and sequence logo derived from randomization and selection of bases 23-28 for interaction with She3p.
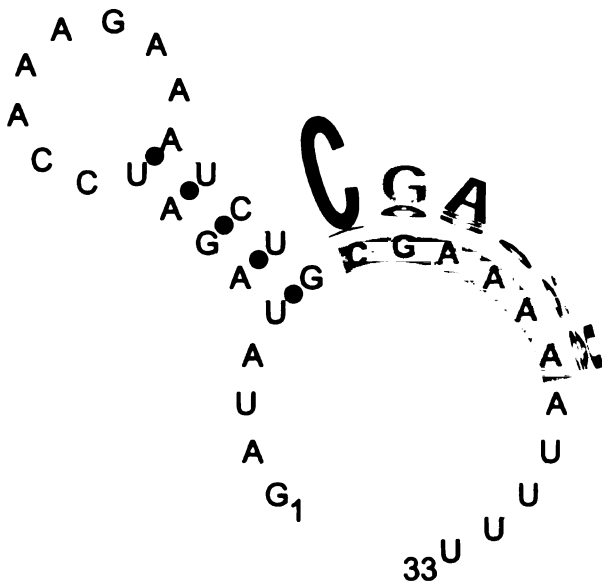
Figure 11: Two stem-loops are required for full activity of zipcode WSC2C. (A) Sequences and predicted structures of full-length WSC2C and WSC2C fragments tested for (B) localization *in vivo*.
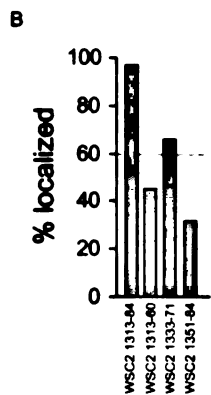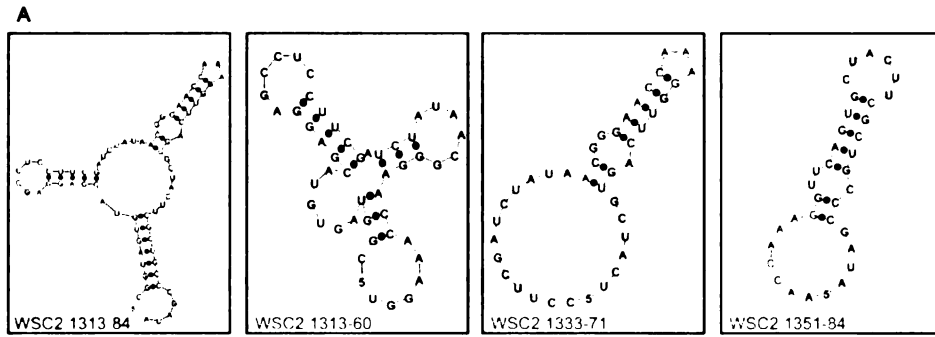
References:

1. Oleynikov, Y. & Singer, R. H. (1998) *Trends Cell Biol* **8,** 381-3.

2. Betley, J. N., Frith, M. C., Graber, J. H., Choo, S. & Deshler, J. O. (2002) *Curr Biol* **12,** 1756-61.

3. Kim-Ha, J., Webster, P. J., Smith, J. L. & Macdonald, P. M. (1993) *Development* **119,** 169-78.

4. Macdonald, P. M. & Kerr, K. (1998) *Mol Cell Biol* **18,** 3788-95.

5. Gautreau, D., Cote, C. A. & Mowry, K. L. (1997) *Development* **124,** 5013-20.

6. Shepard, K. A., Gerber, A. P., Jambhekar, A., Takizawa, P. A., Brown, P. O., Herschlag, D., DeRisi, J. L. & Vale, R. D. (2003) *Proc Natl Acad Sci U S A* **100,** 11429-34.

7. Bohl, F., Kruse, C., Frank, A., Ferring, D. & Jansen, R. P. (2000) *Embo J* **19,** 5514-24.

8. Takizawa, P. A. & Vale, R. D. (2000) *Proc Natl Acad Sci U S A* **97,** 5273-8.

9. Long, R. M., Gu, W., Lorimer, E., Singer, R. H. & Chartrand, P. (2000) *Embo J* **19,** 6592-601.

10. Gonzalez, I., Buonomo, S. B., Nasmyth, K. & von Ahsen, U. (1999) *Curr Biol* **9,** 337-40.

11. Chartrand, P., Meng, X. H., Singer, R. H. & Long, R. M. (1999) *Curr Biol* **9,** 333-6.

12. Olivier, C., Poirier, G., Gendron, P., Boisgontier, A., Major, F. & Chartrand, P. (2005) *Mol Cell Biol* **25,** 4752-66.

13. Bittker, J. A., Le, B. V. & Liu, D. R. (2002) *Nat Biotechnol* **20**, 1024-9.

14. Bernstein, D. S., Buter, N., Stumpf, C. & Wickens, M. (2002) *Methods* **26**, 123-41.

15. SenGupta, D. J., Zhang, B., Kraemer, B., Pochart, P., Fields, S. & Wickens, M. (1996) *Proc Natl Acad Sci U S A* **93**, 8496-501.

16. Longtine, M. S., McKenzie, A., 3rd, Demarini, D. J., Shah, N. G., Wach, A., Brachat, A., Philippsen, P. & Pringle, J. R. (1998) *Yeast* **14**, 953-61.

17. Gietz, R. D. & Woods, R. A. (2002) *Methods Enzymol* **350**, 87-96.

18. Mathews, D. H., Sabina, J., Zuker, M. & Turner, D. H. (1999) *J Mol Biol* **288**, 911-40.

19. Zuker, M. (2003) *Nucleic Acids Res* **31**, 3406-15.

20. Bailey, T. L. & Elkan, C. (1994) *Proc Int Conf Intell Syst Mol Biol* **2**, 28-36.

21. Kruse, C., Jaedicke, A., Beaudouin, J., Bohl, F., Ferring, D., Guttler, T., Ellenberg, J. & Jansen, R. P. (2002) *J Cell Biol* **159**, 971-82.

22. Hook, B., Bernstein, D., Zhang, B. & Wickens, M. (2005) *Rna* **11**, 227-33.

23. Crucs, S., Chatterjee, S. & Gavis, E. R. (2000) *Mol Cell* **5**, 457-67.

24. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. (2004) *Genome Res* **14**, 1188-90.

25. Schneider, T. D. & Stephens, R. M. (1990) *Nucleic Acids Res* **18**, 6097-100.

Chapter 4:

Conclusion

Generation of cell asymmetry by active transport of mRNA is important for the survival on many different cell types and organisms. We have shown that RNA transport occurs commonly in the single-celled yeast: at least 24 mRNAs are transported to the bud tips of dividing cells by a single core motor complex. It is likely that additional mRNAs are transported by the She-complex, since only the mRNAs falling in the top 3 percentile of the list of candidates generated by microarray analysis were tested. Other transcripts falling lower in the list may also be localized by the She-complex. Recently, Jeffrey Gerst reported at the 2005 FASEB meeting that mRNAs encoding proteins involved in polarized growth and secretion are also transported to the bud by the She-complex (1).

The yeast RNA transport system provides an opportunity to study both the general principles of motor-driven transport and novel aspects of RNA localization. Like most transport systems, yeast utilizes a molecular motor which interacts with linker proteins that bind RNA. These interactions are likely to be conserved in a variety of organisms. Unlike most transported RNAs in other organisms, however, the RNAs in yeast are recognized by the motor protein complex on the basis of sequence and structural features of the coding regions rather than the 3' UTRs. Furthermore, all of the bud-localized mRNAs identified to date in yeast are transported by the same core complex, whereas other organisms utilize a variety of transport complexes for localizing RNAs. And unlike other organisms, yeast generally does not require RNA transport to ensure bud- or organelle-specific localization of the proteins encoded by She2/3 target mRNAs.

She2/3-binding is sufficient for localizing RNA to buds:

Selection of zipcodes by NRR/ 3-Hybrid selection revealed that She2/3 binding

was sufficient for mediating RNA transport: >95% of the sequences selected by the

engineered *in vivo* RNA-protein binding assay localized to buds, indicating that binding

to She2/3 is sufficient to confer localization. Accordingly, all tested zipcodes bound to

She2/3 *in vitro*, suggesting that no other factors were needed to facilitate the interaction

of She3 with zipcodes in the 3-Hybrid system. Although we used the carboxyl-terminus

of She3 as bait in the 3-Hybrid selection, She2 is also required for binding RNA, since it

is required for binding to the zipcodes *in vitro* (Peter Takizawa, personal communication)

and She3 does not interact with its targets in a *she2Δ* 3-Hybrid host strain (Ch. 3 Fig. 2

and (2)). In support of this finding, Kruse et al report the existence of a nuclear pool of

She2, which could, in conjunction with She3-Gal4AD, activate the *HIS3* and LacZ

reporters in the 3-Hybrid assay. Despite evidence that the translation inhibitors Khd1 and

Puf6 (3, 4), and other factors such as Loc1 (5) aid in localization, there does not appear to

be any additional requirement for these factors to bind zipcodes directly, because all

sequences identified on the basis of binding ability are functional for localization. The

She-complex is the first example of a minimal protein complex whose binding is to RNA

is sufficient for mediating transport.

A degenerate motif is essential for She2/3-dependent mRNA transport:

A degenerate sequence motif with some secondary structural constraints was

found to be essential for She-complex mediated mRNA transport. The low complexity of

the She-complex recognition site is compatible with its location in open reading frames: maintaining a complex sequence or structural protein recognition site would not have been feasible given the sequence constraints placed upon the RNAs by their protein-coding requirements. It is hypothesized that mRNAs are largely unstructured due to their association with various processing factors following transcription (6); accordingly, we have not found extensive structure in the zipcode sequences. Some zipcodes also serve a third function in regulating translation either directly (7) or indirectly by providing translation factor binding sites (4). Because zipcodes mediate multiple aspects of mRNA regulation, it is important that features outside of the core motif can augment the sequence diversity recognized by the She-complex. This strategy would allow the complex to bind specifically to a variety of RNA targets, and may represent a mechanism employed by diverse mRNA-binding proteins.

The localization motif is likely separable from other mRNA processing signals:

We have shown that a single zipcode at the 3' end of a reporter RNA is sufficient to localize the RNA to >90% of buds. Interestingly, we find that localized RNAs often contain multiple zipcodes: four have been identified in *ASH1* and two in *WSC2*, and preliminary deletion mapping indicates the presence of two zipcodes in *TPO1* as well (data not shown). It appears likely that multiple zipcodes are necessary for anchoring of the RNA to the bud tip, since Chartrand et al (7) report that mutation of 3 out of the 4 zipcodes in *ASH1* compromises anchoring only and does not affect transport into the bud. In contrast, *Drosophila* (8) and *Xenopus* (9) require multiple zipcodes for efficient RNA

98

transport. Our results generalize the finding that one zipcode is sufficient for bud-localization of RNA in yeast, suggesting that formation of higher-order protein complexes on a single RNA target is not necessary for transport in this system.

Chartrand et al (7) have shown that the *ASH1* zipcodes also serve as translation repressors. Their work suggests that the extensive secondary structure present in the previously mapped zipcodes provides a barrier to the passage of ribosomes. Because the zipcodes identified in this work are short and contain minimal secondary structure, it is unlikely that these sequences present insurmountable barriers to translation. While the secondary structure surrounding the She2/3 binding sites appears necessary for localization in the context of the native mRNA, it might play a more direct role in translation control. Mutating the seven base She-binding sites in *ASH1* and analyzing the effects on translation and localization will indicate whether these two functions are separable.

Possible mechanisms of RNA binding:

Because various non-conserved sequence and structural features outside of the core motif contribute to She-complex binding, it is likely that the complex recognizes a combination of primary sequence, and secondary and tertiary structure in its RNA targets. Several mechanisms could explain how the complex binds its targets. One mechanism, exemplified by *bicoid*, is RNA dimerization (10). While an RNA dimer would provide a symmetric binding site for dimeric She2p, dimerization is unlikely because She2p binds RNA in a 2:1 ratio *in vitro* (11), and the zipcodes lack self-complementary regions

99

needed for dimerization. Another possibility is that the RNA adopts a sub-optimal secondary structure upon protein binding. In a multi-step binding event, the She-complex may initially recognize some RNA sequence/ structural features but make critical contacts to other bases after RNA remodeling. For example, the adenosine upstream of the core motif, despite being in different predicted structural contexts, may adopt identical conformations in native zipcodes upon She2/3 binding. Finally, the She-complex may recognize limited tertiary structural features of a few bases, with other bases simply maintaining the three-dimensional architecture. The CG dinucleotide is an attractive candidate for mediating specific recognition, as it was highly conserved in native zipcodes and over-represented in all randomization experiments. Specific She-complex recognition of a complicated RNA structure would account for the variations in sequence and structural preferences between zipcodes.

Future directions:

The identification of 22 new bud-localized mRNAs and the sequence determinants responsible for directing their transport has opened up many avenues of study. Firstly, the purpose of widespread RNA transport has remained elusive; no phenotypes were identified for She-complex mutants under a variety of conditions tested by Giaever et al. (12) or by synthetic lethal screens and/ or direct testing for temperature-dependent effects (see Appendix). It is possible that the She-complex has additional cargoes which are expressed and/ or transported under conditions different from those used for the microarray analyses and phenotypic screens. Identification of other

transported mRNAs under a variety of growth conditions may elucidate why yeast harbors a system dedicated to specific recognition and transport of mRNAs to bud tips. Additionally, analyzing phenotypes for She-complex mutations in other yeast species may also prove to be valuable. In *Candida albicans*, Ash1 protein localizes to filament tips and is essential for virulence (13, 14). It is likely that mutations in the She-complex of this pathogenic yeast (which contains 2 homologs of She3 and none of She2) will affect its virulence and survival in mammalian hosts.

Secondly, yeast presents an unusual paradox in that the proteins encoded by bud-localized mRNAs are properly localized to buds or organelles in the absence of RNA transport. This finding suggests that one or more back-up protein localization mechanisms exist, and that proper localization of these proteins is essential for viability. Screening for secondary mutations which fail to correctly localize these proteins in She-complex mutants may reveal novel mechanisms for targeting proteins to subcellular destinations post-translationally. Additionally, such mutations may display phenotypes which would indicate the role of RNA transport in cell survival.

Just as the processing and sorting of proteins encoded by transported RNAs remain an open for investigation, many questions still remain about the processing of mRNAs themselves. Many of the stages in the life of an mRNA outlined in Fig. 1 (Introduction) have yet to be verified. For example, does She2 associate with its targets in the nucleus, and does nuclear Loc1 protein aid in loading She2 onto mRNAs as has been proposed (15)? The dynamics of translational regulation and transport following nuclear export also remain unsolved. It is not clear whether She-complex binding and translation

(ribosome binding) are mutually exclusive. Evidence that translational repressors Khd1 and Puf6 are necessary for efficient transport of translation-competent mRNAs (3, 4), but not for direct binding of the She-complex to its cognate zipcodes (Ch. 3), suggests that She2/3 and ribosomes may compete for binding to mRNAs *in vivo*. How this competition is regulated in a spatial and/ or temporal manner has not been determined. Finally, it has been shown that *ASH1* mRNA is anchored at the bud tip following transport, and that successful anchoring requires translation of the mRNA (16). It remains to be seen whether other She-complex targets are similarly anchored in a translation-dependent manner. Arp1 has been suggested to play a role in the anchoring process (17), but additional factors responsible for anchoring the RNA at the plasma membrane remain to be identified.

The biochemical and biophysical properties of the She-complex mRNP also remain to be investigated. The crystal structure of She2 suggests that it acts as a dimer, and mutations that disrupt the dimer interface interfere with RNA binding *in vitro* (11). However, it is not clear whether the She2 dimer is the active RNA binding species *in vivo* when complexed with She3. Furthermore, the stoichiometry of She3 in the mRNP complex remains unknown. Our work, combined with preliminary results from P. Takizawa as well as previously published work, suggests that She2 contacts single-stranded RNA specifically but with low affinity (18, 19), while She3 binds double-stranded RNA. The affinities of She2 and She3, either alone or in complex, for RNA have not been determined, nor has the conformation of the mRNP been elucidated. Precise definition of the nucleotides contacted by each of these proteins, by nuclease

102

mapping or cross-linking, will shed light on the substrate preferences of the proteins. Finally, the mechanism of Myo4 motility also remains to be investigated. Studies have shown that, *in vitro*, Myo4 acts non-processively (20) whereas *in vivo* its velocity is consistent with the those measured for other (processive) typeV myosin motors (21). Currently, it is not known what effects linkage of an RNA cargo has on the processivity, velocity, and step size of Myo4, and whether multiple zipcodes in mRNAs function redundantly or synergistically in mediating transport.

Finally, our results have provided a basis for studying specific recognition of mRNAs by mRNA-binding proteins. Messenger RNAs have to employ unique strategies to provide binding sites for processing factors, as they (unlike noncoding RNAs) must simultaneously provide information for protein synthesis. Identifying additional zipcodes from the localized RNAs will broaden our knowledge of the range of sequences and structures that are recognized by the She-complex. Our results suggest that the She-complex recognizes at least 2 classes of zipcodes: one class bearing a seven-base single-stranded motif investigated in detail in Ch. 3, and a second class consisting of 2 stem-loops exemplified by WSC2C (see Ch. 3 Fig. 11) and the 3' end of *TPO1* (data not shown). Three-dimensional structures of the She-complex bound to various target RNA zipcodes will elucidate how the She-complex maintains both specificity and flexibility in target recognition. The strategy employed by the She-complex for RNA recognition will likely apply to various types of mRNA-protein interactions.

References:

1. Jansen, R. P. & Kiebler, M. (2005) *Nat Struct Mol Biol* **12,** 826-829.

2. Long, R. M., Gu, W., Lorimer, E., Singer, R. H. & Chartrand, P. (2000) *Embo J* **19,** 6592-601.

3. Gu, W., Deng, Y., Zenklusen, D. & Singer, R. H. (2004) *Genes Dev* **18,** 1452-65.

4. Irie, K., Tadauchi, T., Takizawa, P. A., Vale, R. D., Matsumoto, K. & Herskowitz, I. (2002) *Embo J* **21,** 1158-67.

5. Long, R. M., Gu, W., Meng, X., Gonsalvez, G., Singer, R. H. & Chartrand, P. (2001) *J Cell Biol* **153,** 307-18.

6. Buratti, E. & Baralle, F. E. (2004) *Mol Cell Biol* **24,** 10505-14.

7. Chartrand, P., Meng, X. H., Huttelmaier, S., Donato, D. & Singer, R. H. (2002) *Mol Cell* **10,** 1319-30.

8. Gavis, E. R., Curtis, D. & Lehmann, R. (1996) *Dev Biol* **176,** 36-50.

9. Chan, A. P., Kloc, M. & Etkin, L. D. (1999) *Development* **126,** 4943-53.

10. Ferrandon, D., Koch, I., Westhof, E. & Nusslein-Volhard, C. (1997) *Embo J* **16,** 1751-8.

11. Niessing, D., Huttelmaier, S., Zenklusen, D., Singer, R. H. & Burley, S. K. (2004) *Cell* **119,** 491-502.

12. Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., Arkin, A. P., Astromoff, A., El-Bakkoury, M., Bangham, R., Benito, R., Brachat, S., Campanaro, S., Curtiss, M., Davis, K., Deutschbauer, A., Entian, K. D., Flaherty, P., Foury, F., Garfinkel, D.

J., Gerstein, M., Gotte, D., Guldener, U., Hegemann, J. H., Hempel, S., Herman, Z., Jaramillo, D. F., Kelly, D. E., Kelly, S. L., Kotter, P., LaBonte, D., Lamb, D. C., Lan, N., Liang, H., Liao, H., Liu, L., Luo, C., Lussier, M., Mao, R., Menard, P., Ooi, S. L., Revuelta, J. L., Roberts, C. J., Rose, M., Ross-Macdonald, P., Scherens, B., Schimmack, G., Shafer, B., Shoemaker, D. D., Sookhai-Mahadeo, S., Storms, R. K., Strathern, J. N., Valle, G., Voet, M., Volckaert, G., Wang, C. Y., Ward, T. R., Wilhelmy, J., Winzeler, E. A., Yang, Y., Yen, G., Youngman, E., Yu, K., Bussey, H., Boeke, J. D., Snyder, M., Philippsen, P., Davis, R. W. & Johnston, M. (2002) *Nature* **418,** 387-91.

13. Inglis, D. O. & Johnson, A. D. (2002) *Mol Cell Biol* **22,** 8669-80.

14. Munchow, S., Ferring, D., Kahlina, K. & Jansen, R. P. (2002) *Curr Genet* **41,** 73-81.

15. Beach, D. L. & Bloom, K. (2001) *Mol Biol Cell* **12,** 2567-77.

16. Gonzalez, I., Buonomo, S. B., Nasmyth, K. & von Ahsen, U. (1999) *Curr Biol* **9,** 337-40.

17. Trautwein, M., Dengjel, J., Schirle, M. & Spang, A. (2004) *Mol Biol Cell* **15,** 5021-37.

18. Gonsalvez, G. B., Lehmann, K. A., Ho, D. K., Stanitsa, E. S., Williamson, J. R. & Long, R. M. (2003) *Rna* **9,** 1383-99.

19. Olivier, C., Poirier, G., Gendron, P., Boisgontier, A., Major, F. & Chartrand, P. (2005) *Mol Cell Biol* **25,** 4752-66.

20.   Reck-Peterson, S. L., Tyska, M. J., Novick, P. J. & Mooseker, M. S. (2001) *J Cell Biol* **153,** 1121-6.

21.   Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S. M., Singer, R. H. & Long, R. M. (1998) *Mol Cell* **2,** 437-45.

Appendix

Contributions:

Kelly Shepard co-conducted automated SGA screens whose results are listed in Tables 2 and 3. Ashwini Jambhekar constructed all strains, obtained all data in Tables 1 and 4, co-conducted SGA screens whose results are listed in Tables 2 and 3, performed all tetrad dissections and random spore analyses indicated in Tables 2-4, performed experiments shown in Figures 2-3, and (unless noted otherwise) performed the growth assays described in Results.

Joseph L. DeRisi, Thesis Advisor

Abstract:

Subcellular localization of mRNA is an essential process in many organisms. In the yeast *S. cerevisiae*, an actin-myosin based RNA transport mechanism was elucidated on the basis its role in transporting *ASH1* mRNA which regulates mating-type switching; subsequently, 23 additional mRNAs were found to be transported to bud tips of dividing cells by the same core complex. However, the purpose of localizing these additional RNAs to bud tips remains unclear, and RNA transport in this organism is not essential for survival. Here, we have used synthetic genetic array (SGA) analysis to attempt to elucidate a phenotype for RNA transport in yeast. Strains mutant for *she2* or *myo4*, genes encoding two of the core transport complex members, were assayed for their ability to produce viable haploid double mutants when crossed to the yeast deletion collection, which represents deletions in all non-essential genes. No genuine interactions were identified between *she2* or *myo4* and any of the deletions assayed, suggesting that RNA transport does not function redundantly with any of the process represented in the collection under the conditions tested. Many sources of false positives and false negatives inherent in SGA analysis may have contributed to the failure to identify true genetic interactions. It remains possible that RNA transport is essential under certain growth conditions or in certain mutant backgrounds not represented in the deletion collection. It is also possible that mutations in RNA transport cause subtle defects in other processes which would be elucidated only by direct testing and not by SGA analysis. Despite the tight specificity of the She-complex for its RNA targets, a phenotype for widespread mRNA transport in yeast remains elusive.

Introduction:

Subcellular localization of mRNA is an essential process in many different cell types and organisms. It is often used to ensure that protein products are located at the site of their activity. Although the proteins themselves can be (and often are) localized to appropriate organelles or subcellular regions, transporting the mRNA precursors is more efficient solution, as one transcript can give rise to many protein products. In *Drosophila* oocytes, localization of *bicoid* and *nanos* mRNAs to opposite ends of the cell ensures that proteins which act later to produce the head and tail regions of the embryo will function in the appropriate locations. Translational control mechanisms often act in concert with the RNA trafficking systems to prevent messages which fail to be localized or are en route to their final destination from being translated ectopically (1). RNA transport complexes are essential in most systems, and inactivation of transport results either in protein production in inappropriate locations or (if translation is tightly coupled to RNA localization) a complete absence of the protein product. Because many of the transported RNAs in Drosophila oocytes and embryos are essential for embryonic development, mislocalization of these transcripts often causes embryonic lethality or gross morphological defects.

In *S. cerevisiae*, a core transport complex consisting of the myosin motor Myo4 and linkers She2 and She3 transports at least 24 mRNAs to bud tips of dividing cells. These transcripts encode a variety of proteins, although several function directly or indirectly in maintenance of cell wall or membrane structures (2). 10 of the 24 bud-localized RNAs encode proteins that are also bud-localized, and only one of the

transported mRNAs, *YJL051c*, is essential for viability. In contrast to other systems, the transport complex in yeast is not essential, and mutation of any of the core components is not reported to cause fitness defects under many different growth conditions tested (3). One possible reason for the non-essential nature of the She-complex is that many of the protein products encoded by the localized transcripts are properly localized even in the absence of a functional transport complex. Only two of the 24 She-complex targets (*ASH1* and *IST2*) require RNA transport to ensure bud-localization of their protein products (2).

These initial findings suggested that the RNA transport pathway functions redundantly with a mechanism for protein localization, and possibly with other mechanisms for cell wall/ membrane maintenance. One approach to dissecting overlapping or redundant mechanisms to is to identify synthetic genetic interactions: mutants in the pathway of interest are screened for secondary mutations which enhance the phenotype of the original mutation. Usually, the phenotype assayed for is a growth defect of the double mutant: synthetic interactions are evinced as double mutations which compromise growth to a greater extent than either single mutation alone does. Mutations identified by this method usually lie in pathways that operate parallel to that of the query mutation. Therefore, identification of synthetic interactions can help to elucidate the functions of the query gene product.

Traditionally, synthetic lethal screens are conducted by deleting the non-essential gene of interest and introducing a WT copy of the gene on a plasmid. The strain is then randomly mutagenized and screened for mutants which have become dependent on the

plasmid for survival. To be successful, this method requires that the initial mutagenesis be saturating; however saturation is difficult to achieve because of the large number of mutations needed to ensure that every gene has been inactivated. Furthermore, identification of the isolated secondary mutations can be cumbersome.

Recently, an automated technique (synthetic genetic array (SGA) analysis) has been developed for identifying synthetic interactions (4). This method involves crossing the query mutation to a yeast deletion collection of 4700 single-deletion strains. The resulting diploids are sporulated and screened for their ability to produce viable haploid double mutants. The advantages of this approach are two-fold: it ensures that all viable single deletions are screened for synthetic interactions with the query mutation, and it allows immediate identification of the synthetically-interacting mutations.

SGA analysis was carried out with strains deleted for either Myo4, the myosin motor responsible for RNA bud-transport, or She2, a linker connecting RNA to Myo4. Although several candidate synthetic interaction partners were identified, none of the double mutants displayed significant growth defects upon retesting. Therefore, it appears that, under the growth conditions tested, RNA transport does not act in concert with any other mechanism in order to perform an essential cellular function.

Materials and Methods:

Yeast strains and media:

All strains used were derivatives of S288c. Gene disruptions were performed in the SGA host strain (4) as described using the NatMX marker (5). The *myo4Δ*::Nat allele

was generated by transforming the *myo4*Δ::Kan strain from the deletion collection with plasmid p4339 digested with *Eco*RI, resulting in replacement of the Kan marker with Nat. Query *myo4*Δ::Nat and *she2*Δ::Nat strains were backcrossed to the isogenic WT strain to ensure >90% spore viability. The *lyp1*Δ deletion was introduced by PCR amplifying 1.2kb surrounding the *LYP1* locus from a *lyp1*Δ deletion strain and transforming the PCR product into LYS+ versions of the SGA query strains (obtained by back-crossing). Unless noted, all experiments were conducted on solid medium containing 2% agar. YPD consists of 2% peptone, 1% yeast extract, 2% glucose, 120mg/l adenine. Pre-spo medium consists of 1% yeast extract, 3% nutrient broth, 5% glucose. Sporulation medium consists of 0.05% amino acid –his –ura dropout mix, 0.05 mM uracil, 0.05 mM tryptophan, and 10g/l filter-sterilized potassium acetate. Synthetic drop-out medium consists of 0.67% yeast nitrogen base without amino acids and ammonium sulfate, 0.2% amino acid dropout powder, and 2% glucose. Where indicated, G418 was added at a concentration of 200mg/l, Nat at 100mg/l, canavanine at 50mg/l, S-AEC at 50mg/l. In cases where G418 and/ or Nat were added to synthetic medium, yeast nitrogen base was replaced with 0.1% monosodium glutamate (MSG) and 0.17% yeast nitrogen base without amino acids and without ammonium sulfate. Unless noted, all strains were incubated at 30°C.

SGA analysis:

SGA analysis was performed as described (4). Deletion collection strains grown on YPD+G418 were mated to query stains growing on YPD+Nat on YPD. Diploids were

selected on YPD+G418+Nat and replicated to pre-spo medium at 30°C and incubated

overnight. Strains were replicated to sporulation medium and incubated for 7-10 days at

25°C. Sporulation was verified by microscopic analysis of at least 5 strains. Cells were

replicated to SD-his-arg+canavanine (or SD-his-arg-lys+canavanine+S-AEC for *lyp1*Δ

query strains) and incubated for two days, then transferred onto a fresh plate of the same

medium for 1 day to select for haploid meiotic products. The resulting cells were

transferred to SD[MSG]-his-arg+canavanine+G418 or SD[MSG]-his-arg-

lys+canavanine+S-AEC+G418 medium (for *lyp1*Δ query strains) to select for haploid

cells containing the deletion collection mutations. These haploid single mutants were then

simultaneously replicated to a fresh plate of the same medium as well as to medium

selective for double mutants bearing the additional query mutation (SD[MSG]-his-

arg+canavanine+G418+Nat, or SD[MSG]-his-arg-lys+canavanine+S-AEC+G418+Nat

for *lyp1*Δ query strains). After 1 day of incubation, the sizes of the patches on the two

media were compared. Any patches which appeared smaller or inviable on the double-

mutant selective medium as compared to the single-mutant selective medium were

identified as candidates showing synthetic interactions between the query and deletion-

collection alleles. Candidate single mutant cells were transferred from the single-mutant-

selection medium onto YPD+Nat to confirm that the query mutation was not present in

these cells.

Verification of SGA candidates:

Heterozygous diploids bearing both the query mutation and the candidate deletion collection mutation were generated as described above and sporulated. Tetrad analysis was performed as described (6). Random spore analysis was performed by streaking spores onto SD[MSG]-his-arg+canavanine+G418 or SD[MSG]-his-arg-lys+canavanine+S-AEC+G418 medium for *lyp1*Δ strains to select for haploids containing the deletion collection mutation. Plates were incubated at 30°C for 2-3 days. 5 small and 5 large colonies from each streak were patched onto YPD and replicated onto YPD+Nat to test for the presence of the additional query mutation. Diploids which were found not to produce double mutant haploids (or produced only slow-growing double mutants) by random spore analysis were subjected to tetrad analysis.

Analysis of growth rates:

Strains were grown overnight in liquid YPD culture to saturation, then diluted in fresh YPD to OD_{600} of 0.3. Strains were grown to log phase for 3.5 hours to an OD_{600} of 0.65-0.8. 5 serial 10-fold dilutions of each strain were made, and 2μl of each dilution was spotted on YPD or 1/4x YPD and incubated at 30°C or 37°C as indicated.

Results:

Rationale of experiments:

Synthetic genetic array (SGA) analysis entailed crossing a query mutation into the yeast deletion collection and isolating double mutants in an automated manner (Fig 1).

Strains containing auxotrophic or drug-resistance markers were sequentially pinned onto media selective for diploids, haploid meiotic progeny, haploid progeny bearing the deletion collection mutation, and finally haploids containing both the deletion collection and the query mutations. Specifically, a MATα query strain was first created by replacing the gene of interest with the nourseothricin resistance (Nat^R) cassette. This strain also contains an Mfa1pr-*HIS3* construct integrated in the *CAN1* locus; the *CAN1* disruption confers recessive resistance to canavanine as well histidine prototrophy in MATa, but not diploid or MATα, cells (therefore, the query strain itself is a histidine auxotroph). One version of the strain provides more stringent selection against diploids by virtue of an additional *lyp1*Δ deletion, which confers recessive resistance to the toxic lyine analog S-aminoethyl-L-cysteine (S-AEC). The query strain was crossed to the deletion collection, in which each non-essential gene is replaced with the Kan resistance cassette. Each deletion collection member was present in duplicate. Diploids were selected on the basis of their resistance to both Nat and G418 (a kanamycin analog). Following sporulation, meiotic progeny were selected in the presence of canavanine (and S-AEC for *lyp1*Δ query strains) on medium lacking histidine. This selection eliminated unsporulated diploids (phenotypically his- can1^S) as well as unmated parental haploids (can1^S and/ or his- cells) which may have survived Nat/ G418 selection. Subsequently, deletion collection mutants were selected in the presence of G418 while maintaining haploid selection. Finally, double mutants were isolated by introducing an additional selection for the Nat marker. Double mutant strains which grew more slowly than the corresponding Kan-marked single mutants harbored deletion-collection mutations which were candidate

116

synthetic interaction partners with the query mutation. Although each deletion collection strain was tested in duplicate, any mutation which showed a synthetic interaction in one of the two replicates was considered a candidate.

Unlike conventional synthetic lethality screens, the SGA screen requires that all strains be able to mate, sporulate, and germinate efficiently. In principle, deletion collection members defective in these processes would be eliminated because of their inability to produce viable single mutants; however, it was possible that such mutants would appear as candidate synthetic interaction partners with the query mutation. Additionally, genes linked to the selectable markers in the query strain were expected to score as candidates, since the proportion of spores containing both linked markers would be limited by the map distance separating them. Therefore, to determine which deletions were likely to appear as false positives by SGA analysis, a WT strain bearing all necessary markers for selection (including *lyp1*Δ) was tested. We expected that *can1* would be identified in the screen because the WT strain contains the Nat[R] marker at the *CAN1* locus, and both *can1*::Kan and *can1*::Nat alleles would not be present in the same haploid. As expected, no *can1* "double mutants" were recovered. Only 9 other mutations appeared synthetic lethal/ sick with the WT strain. Of these, *cbc2* was also reported by Tong et al. (4) to show synthetic lethality with a WT strain, and 4 fell in the same pathways as other genes in the Tong et al. dataset (Table 1). Our set of mutations showing synthetic interactions with the WT strain was much smaller than that reported by Tong et al., most likely because of differences in scoring methods. We relied only on visual detection of growth differences between single and double mutants. Tong et al.

performed the screen in triplicate, quantitated the diameter of each single and double

mutant strain, and scored as synthetic lethal/ sick all double mutants with statistically

significant size differences when compared to the corresponding single mutant strain.

Despite the differences in scoring method, we reasoned that any candidates from the

Tong et al. dataset showing synthetic interactions with query mutations were likely to

represent non-specific interactions; therefore, we considered the union of our dataset and

that of Tong et al. as the background set of false positive candidates generated by SGA

analysis.

After using the datasets from the WT screens to filter the list of candidates

generated by screening a query mutation, subsequent tests were conducted to confirm the

observed genetic interactions. First, the single (Kan-marked) mutant patches generated by

the SGA screen were streaked on YPD+Nat to confirm that the presence of the additional

query (Nat-marked) mutation caused cells to be inviable or grow slowly; single deletion

patches that appeared to contain healthy double mutant cells were eliminated from the list

of candidates. Synthetic interactions between the remaining candidate deletions and the

query mutation were confirmed by one of two methods. Both began by generating and

sporulating the doubly heterozygous diploid; the resulting spores were then tested by

tetrad or random spore analysis. Tetrad analysis was performed by dissecting tetrads on

standard rich medium (YPD). In this assay, a true synthetic interaction could be identified

by double mutant spores which produced inviable or slow-growing colonies, while single

mutant spore clones (of either genotype) appeared similar to WT clones. Random spore

analysis was performed by streaking the spores for single colonies on medium selective

for haploid meiotic progeny bearing the deletion collection mutation. If the candidate

mutation were synthetic lethal with the query mutation, none of the selected colonies

would bear the query mutation; if the mutation caused synthetic slow growth (i.e. a

"synthetic sick" phenotype), the double mutant colonies would be smaller than the single

mutants. Any strain that failed to produce large double mutants was retested by tetrad

dissection.

SGA analysis of *she2*Δ:

A query strain containing a *she2*Δ::Nat deletion was subjected to SGA analysis

and 84 candidates were identified (Table 2). Of these, 15 were reported to show synthetic

interactions with a WT strain, and 13 were linked to *SHE2*. Testing the remaining 56

haploid single (deletion collection) mutants on Nat-containing media revealed that all of

the single mutants patches contained cells harboring the second *she2*::Nat deletion.

Therefore, 20 candidates which appeared clearly lethal on the SGA double mutant

selection plates were chosen for random spore analysis. As expected based on the fact

that many transported RNAs encode proteins involved in cell wall maintenance, this set

of 20 contained candidates involved in related pathways: two were involved in cell wall

regulation (*CHS1*, *DSE4*), and one (*YKL162c-a*) was homologous to known cell

membrane proteins. Of the remainder, 7 were uncharacterized ORFs, 5 were involved in

nutrient uptake/ metabolism (*ADH4*, *SNZ2*, *TKL1*, *GAP1*, *GTR1*), and 4 played regulatory

roles in processes such as sumoylation (*WSS1*), RNA PolII phosphorylation (*CTK2*),

histone acetylation (*HAT1*), and transcription (*YAP3*). One was involved in mating

(*FUS1*). All 20 were tested by random spore analysis and were found to produce viable double mutants with no obvious growth defects.

All single mutants generated by the SGA screen were pinned onto fresh plates selective for single or double mutants and incubated at 37°C to screen for synthetic temperature sensitivity. Although several candidate deletion collection mutants appeared to have growth defects at 37°C in a *she2*Δ background, none of them showed similar defects upon retesting by random spore analysis (Kelly Shepard, personal communication).

SGA analysis of *myo4*Δ:

SGA analysis was initially carried out in a *myo4*Δ::Nat strain lacking the *lyp1Δ* marker, and 124 candidates were identified (Table 3). 10 were linked to *MYO4* and 5 were reported to show spurious lethality in the SGA screen. After eliminating the SGA-derived candidate Kan^R patches containing cells with the *myo4*Δ deletion, 91 candidates remained. None of these candidates overlapped with the initial set of 84 candidates generated by the *she2* SGA screen. Of the 91 candidates, 33 were uncharacterized ORFs, 7 of the gene products were involved in transcription, 11 were involved in basic metabolism, 5 were involved in regulating metal ion homeostasis, 2 in cell wall assembly, 4 in budding or other asymmetric cellular processes, and 3 in translation regulation. The remainder functioned in various cellular processes. 43 of the *myo4*Δ candidates were tested by tetrad dissection, and the remainder by random spore analysis. All candidate deletions produced healthy double mutants in combination with *myo4*Δ, indicating that

they showed up as false positives in the screen. To confirm that the double mutant strains displayed no growth defects, 17 unique double mutants generated by tetrad dissection were streaked on YPD in parallel with a *myo4Δ* single mutant, and growth rates of the double mutants were visually compared to that of the *myo4* strain. No significant differences were discernible (data not shown).

Tetrad analysis revealed several possible reasons for the high rate of false positives. Some strains produced double mutants at a frequency lower than expected, even though the double mutants grew normally and the candidate mutation was unlinked to *MYO4*. Other strains displayed generally low spore viability: while the spore viability of healthy strains was >95%, 21 of the candidate deletion library members displayed spore viability <70% when crossed to the *myo4Δ*::Nat strain and only 7 displayed spore viability >90%. Thus, the inability of a strain to produce sufficient numbers of viable haploid meiotic progeny correlated with its appearance as a synthetic lethal candidate in the SGA screen.

Because we expected that genes involved in regulation of cell asymmetry might function redundantly with the RNA transport system, two candidate genes, *ace2Δ* and *mub1Δ*, were tested more rigorously for growth defects in a *myo4* background. Ace2 is a daughter-specific transcription factor, and Mub1 restricts bud formation to one bud per cell cycle. WT, *myo4Δ*, *ace2Δ*, *mub1Δ*, *myo4Δace2Δ*, and *myo4Δmub1Δ* strains were isolated by tetrad dissection of heterozygous diploids. Each strain was grown to log phase in rich medium, and five 10-fold serial dilutions of each were spotted on rich (YPD) medium and incubated at 30°C. Because many of the localized RNAs encode proteins

involved in cell wall maintenance, we hypothesized that any synthetic growth defects would be more easily detected under conditions compromising cell wall integrity. Therefore, dilutions of the strains were also spotted under conditions of low osmolarity (1/4x YPD) and incubated at 37°C. All single and double mutant strains displayed growth comparable to the WT strain under both conditions. Additionally, no significant growth defects were observed in any of the strains when growing from saturation to log phase. These results indicate that Myo4 does not function redundantly with Ace2 or Mub1 in promoting cell growth or maintaining a functional cell wall.

The single and double mutants generated by SGA analysis were tested at 37°C to identify any synthetic temperature sensitivities in the double mutants. Similar to the results with *she2*Δ, no synthetic temperature sensitivity was observed (Kelly Shepard, personal communication).

Because no true synthetic lethal interactions were detected with *myo4*Δ, the SGA screen was repeated with a *myo4*Δ strain containing the additional *lyp1*Δ marker which selects against diploids. We reasoned that any true synthetic interactions which were obscured by the presence of contaminating diploids would be revealed upon more stringent selection with *lyp1*Δ. We also reasoned that candidates which appeared in replicate experiments would be worth studying more exhaustively, as they would be likely to display growth defects in a *myo4* background under conditions other than those used for tetrad or random spore analysis. This screen generated 128 candidates (Table 4). As in the previous screen, 10 of the candidates were linked to *MYO4*, although the sets of linked genes identified in the two screens did not entirely overlap. The *lyp1*Δ marker

appeared to confer more stringent selection for haploids, since 23 of the candidates were identified by Tong et al. as showing spurious lethality in the WT SGA screen, whereas only 5 such candidates were identified in the previous *myo4* screen. Initially, it appeared that only one candidate, *EST3*, overlapped between the two datasets; but upon re-examination of the SGA plates from the first screen, 3 other candidates from the second screen (*PSP2*, *YAK1*, *YER135c*) were found to show slight growth defects in the first screen as well. Most of the candidates generated in one screen did not show any synthetic interactions in the other. However, due to pinning errors or other technical problems, in a few cases some candidates identified in one screen had not been propagated to the final stages of the other screen, and therefore had been tested directly in only one of the experiments. The set of candidates generated by the second screen was not filtered by testing the SGA-derived Kan^R single mutants on YPD+Nat because it was possible that relaxing the selection for haploid Kan^R cells allowed diploids or Nat^R Kan^S haploids (which had escaped selection) to grow robustly, thus obscuring genuine candidates. Instead, all candidates were tested directly by random spore analysis.

Three candidates—*caf40Δ*, *ilm1Δ*, and *uaf30Δ*-- appeared to produce double mutants with growth defects. Caf40 is a member of the Ccr40/Not4 transcription complex (7), Ilm1 is required for mitochondrial genome maintenance (8)and has also been implicated in regulating chitin levels in the cell wall (9), and Uaf30 is an RNA PolI transcription factor (10). Each of these candidates was tested by tetrad dissection for synthetic interactions with *myo4*. Viable double mutants were recovered after sporulating the *caf40Δ* x *myo4Δ* heterozygous diploid. Dissection of *ilm1Δ* x *myo4Δ* tetrads revealed

a slow-growth phenotype that segregated 2:2 and was linked to the *ilm1*Δ mutation; however, there was no growth defect linked to the *myo4* mutation (Fig. 2). Accordingly, *ilm1*Δ is reported to grow more slowly than WT strains (11). The cross with *uaf30*Δ was interesting in that 3 distinct colony sizes (small, medium, and large) were visible among the meiotic progeny (Fig. 3); the colony sizes segregated in a 1:1:2 ratio in most tetrads. As expected, all the large colonies were WT or *myo4*Δ single mutants. We expected that the medium colonies would be *uaf30*Δ single mutants (10), and the small ones *uaf30*Δ*myo4*Δ double mutants (due to synthetic growth defects caused by the two mutations). While the small and medium colonies always contained the *uaf30* mutation, there was no correlation between the presence of the *myo4* mutation and small vs. medium colony size. Contrary to the expectations, some of the small colonies were *uaf30*Δ single mutants, and some of the medium colonies were *uaf30*Δ*myo4*Δ double mutants. These results indicated that an additional heterozygous allele was segregating in a Mendelian fashion and causing a synthetic growth defect in a *uaf30*Δ background (see discussion). Thus, it is clear that various types of growth/ viability defects can appear as synthetic interactions in the SGA screen.

Taken together, our results reveal that the RNA transport complex in yeast does not functionally overlap with any other non-essential pathways promoting cell growth under standard conditions or at high temperature. In support of our findings, Haarer et al. tested several candidate genes (including including alleles of actin, *myo2*, and secretory pathway components) for synthetic lethality with *myo4* and did not find any synthetic

124

interactions (12). However, the fact that far more candidate synthetic interactions were generated by screening the mutant strains than WT suggests that mutations in the RNA transport system may cause slight fitness defects in some mutants which could not be detected under the various conditions used for confirming the genetic interactions. Nevertheless, any fitness defects of the mutant strains are likely not specific to any particular cellular process, as the sets of candidates generated by independent screens did not overlap significantly.

Discussion:

Automated SGA analysis provides a rapid method for screening the entire yeast deletion collection for mutations showing genetic interactions with a deletion of interest. This method allows one to discern the extent of saturation of the screen, as it is straightforward to identify which mutations survived the early selection steps and were directly tested for synthetic interactions. In contrast, the extent of saturation cannot be unambiguously determined in a conventional synthetic lethal screen. Additionally, mutations showing genetic interactions with the query mutation can be easily and accurately identified, as each deletion collection member is expected to harbor only one defined mutation. Because the She-complex transports a variety or RNAs to the bud tips of dividing yeast cells, it was expected that one or more genuine synthetic interactions would be identified between *she2*Δ or *myo4*Δ, thereby elucidating the physiological purpose of widespread mRNA transport in yeast.

Expected outcomes of *myo4Δ* and *she2Δ* SGA screens:

We expected that SGA analysis would reveal synthetic interactions between She-complex mutants and two different classes of mutations: those compromising cell wall or membrane integrity and those involved in protein transport. Many of the proteins encoded by She-complex targets function in cell wall/ membrane maintenance, particularly in response to environmental stress (2). Since the bud is the most active site of cell wall and membrane growth, it seemed reasonable that proteins involved in those processes would be synthesized at the site of their activity. Because formation of a robust cell wall is essential for survival, we expected that delocalization of transcripts encoding proteins involved in cell wall regulation, when combined with other mutations compromising cell wall/ membrane function, would cause defects in cell growth, as both the protein inactivated by the deletion-collection mutation and the those encoded by the localized RNAs would be absent or in low supply at the bud tip. Although several of the candidates generated by SGA analysis were involved in cell wall or membrane regulation, none of these mutations displayed a significant growth defect upon retesting in *myo4Δ* or *she2Δ* backgrounds.

We also expected that SGA analysis of *myo4Δ* and *she2Δ* would identify mutations in genes that regulate protein localization, since we had observed that proteins encoded by the transported the RNAs are bud-localized even in the absence of RNA transport (2). In fact, our failure to identify mutations in cell wall/ membrane components may have arisen from the fact that proteins encoded by She-complex targets were not significantly depleted in the bud even in mutants. Two possible mechanisms for

localizing proteins are 1) making novel use of previously-characterized components of cellular transport systems or 2) regulating translation such that only bud-localized RNAs are translated. One transported RNAs, *IST2*, encodes a protein which is localized to the bud membrane by a non-canonical transport system (13), suggesting that the former mechanism may be relevant for protein localization. Although Ist2 is unusual in that the protein is delocalized in *she2Δ* mutants, we hypothesized that other proteins encoded by transported RNAs might also be localized by the same novel system utilized by Ist2. Juschke et al. have reported that Ist2 is localized independently of Myo2 and does not traverse through the compartments utilized by other secreted proteins but may, like Vg1 mRNA and protein in *Xenopus* (14), be localized by transport on cortical endoplasmic reticulum (13). An independent study found that a fraction of *ASH1* RNA associates with COPI vesicles, and it was suggested that COPI-mediated transport delivers the RNA to ER for bud localization (15). Cortical ER transport into the bud requires She3 and Myo4 (16), and, although most proteins encoded by She2/3 targets are bud-localized in *she2* mutants, it is possible that their localization requires Myo4, She3, and cortical ER transport. However, mutations that compromise cortical ER transport do not cause RNA delocalization, and Tpo1 localization was found not to be She3/ Myo4 dependent (Kelly Shepard, personal communication). Nevertheless, it was possible that some of the RNAs are translated on COPI vesicles and delivered to the bud via cortical ER. Surprisingly, we did not recover any mutations in cortical ER transport pathways or vesicle-based protein trafficking systems in the SGA analyses.

127

Another possible mechanism for protein bud-localization is restriction of protein synthesis to the bud: the mRNAs may be translated only in the bud, or they may be degraded in the mother. *Drosophila* uses both localized translation and degradation mechanisms to ensure proper protein localization: *nanos* RNA is symmetrically distributed but translated only at the posterior of the oocyte (17), while *hsp83* is degraded everywhere except at the posterior end (18). Despite the precedent that translation regulation may mediate bud-localization of proteins, we did not identify any known translation repressors, activators, or mRNA decay factors by SGA analysis.

There are several possible explanations for our failure to identify any true synthetic interactions with *myo4Δ* or *she2Δ*: robust cell wall/ membrane synthesis or protein localization may not be essential for viability under the conditions used for testing, or other proteins involved in these systems may themselves be essential and therefore mutations not represented in the deletion collection. Unlike a traditional synthetic lethal screen SGA analysis cannot identify alleles of essential genes that interact with the query mutation. Secondly, it is possible that any defects conferred by She-complex mutations are specific to a particular stage of the cell cycle, as the expression of a significant percentage of the localized mRNAs is cell-cycle regulated (2). A subtle, cell-cycle dependent phenotype may not be evident when analyzing asynchronous populations of cells. Furthermore, the low sporulation rate of the S288c strain background, additional mutations in some deletion collection strains, and the selection conditions all contribute to both false positives and false negatives as described below.

Sources of false positives:

A 100% false positive rate was revealed in all three screens (2 with *myo4Δ* and 1 with *she2Δ*); this rate of false positives likely results from characteristics of the strains as well as the selection strategy.

Both the medium used for selection and the selection protocol itself contributed to the high false positive rate. Generation of double mutant strains involved four successive rounds of replication following sporulation. This strategy amplifies subtle growth defects exponentially, since SGA analysis generates mixed populations of cells at each round of selection that are in competition with each other—the ratio of WT to mutant cells with slight growth defects increases during each round of growth. In contrast, tetrad or random spore assays involve analyzing the growth of a single colony directly from a spore, which does not exponentially amplify differences in growth rates. Secondly, the selection was performed on medium lacking 2 or 3 amino acids and containing 3 or 4 drugs for the standard or *lyp1Δ* strains, respectively. Finally, the strains were pinned at high density, creating competition for nutrients. These sub-optimal growth conditions likely amplified subtle growth defects of some mutant strains and may explain the recovery of many candidates involved in amino acid and sugar metabolism. In contrast to the SGA selection conditions, tetrad analysis was performed at low density on rich medium, which could obscure condition-dependent phenotypes. Random spore analysis, while conducted on medium analogous to that used for the SGA selections, nevertheless maintained colonies at low density. In theory, verification of SGA candidates by tetrad analysis is more rigorous than random spore analysis because it allows direct comparison of the growth

rates of all possible genotypes. However, random spore analysis is more rapid and likelier to reveal mutations that cause subtle defects, as haploids are selected under more stringent conditions. Performing tetrad dissections on the same medium used for random spore selection would combine the advantages of both techniques; however, any synthetic interaction which is evinced only under such stringent conditions will be difficult to follow in subsequent experiments aimed at detailed characterization of the genetic interaction.

Tetrad analysis of 43 candidates generated by the *myo4* screen revealed additional sources of false positives. Low spore viability was a common occurrence in many of the candidate strains. Because the S288c strain sporulated at a rate of <1% and only 1/16 of the spores (1/32 spores from the *lyp1*Δ crosses) contained the markers necessary to survive selection, further (non-specific) reduction in spore viability may have led to insufficient numbers of viable double mutants being propagated through SGA analysis. The low spore viability may have resulted from aneuploidy in the deletion collection strains, as 8% of strains are estimated to be aneuploid (19). Alternately, some mutations such as *ecm4* (which displayed 34% spore viability when crossed to *myo4*) may affect spore wall formation and thus play a more direct role in decreasing spore viability.

Another cause of apparent synthetic lethality was the presence of mutations in the deletion colletion strains unlinked to the Kan[R] marker. This situation was apparent in the *myo4*Δ x *uaf30*Δ tetrads: a synthetic phenotype was observed between *uaf30*Δ and a factor that did not co-segregate with any of the markers. Oakes et al. report that the slow growth of *uaf* mutants can be suppressed by expanded rDNA repeats on ChrXII (20).

Therefore, it is likely that the deletion collection *uaf30Δ* strain had acquired these expanded repeats, while the *myo4* strain had not. After sporulation, *uaf30Δ* alleles which cosegregated with the expanded rDNA repeats likely produced medium sized colonies, while *uaf30Δ* mutants that inherited a WT ChrXII from the *myo4Δ* parent produced small colonies. While such interactions should theoretically not have influenced the results of the SGA screen, when coupled with the sub-optimal growth conditions and small sampling of meiotic progeny, they may have lead to apparent synthetic interactions between the query and deletion collection mutations.

Sources of false negatives:

The rate of false negative interactions for these screens is unclear, as no genetic interactions have been reported for either *she2* or *myo4*. Similar to our experience, Luo et al. failed to recover any synthetic interactions with *mlc2* in an SGA screen, although upon further probing they found that Mlc2 plays a role in contractile ring disassembly following cytokinesis (21).

Several factors could result in the failure to recover genuine synthetic interactions. Firstly, the SGA screens were not saturating, as approximately 5% of the deletion collection members were not propagated to the final rounds of selection (data not shown) either due to pinning errors or inability to pass the early rounds of selection. For example, *ste* mutants cannot be assayed by this method because they fail to mate to the query strain, yet they may be interact with pathways, such as RNA transport, which generate cell asymmetry. Secondly, just as additional mutations in the deletion collection strains

131

could lead to false positives, they could have also generated false negatives. Some of the collection members are reported to harbor ectopic WT copies of the deleted gene, thus making the strains functionally WT. Finally, the markers used for SGA selections were "leaky." For example, although only MATa strains were expected to express the Mfa1pr-*HIS3* gene, MATα and diploids were also weak histidine prototrophs (data not shown). Contaminating diploids or parental haploids on the final selection plates would obscure any genuine synthetic interactions.

Despite our failure to uncover any synthetic phenotypes for the RNA transport system, its tight specificity for RNA targets suggests that the system plays an important role in cell survival and is not a gratuitous phenomenon. It is possible that RNA transport is essential under certain growth conditions or in certain strain backgrounds. Testing She-complex mutants under a variety of conditions, particularly those known to compromise cell wall integrity, may reveal essential functions of the transport system. Other approaches can also be taken to identify any processes that function redundantly with RNA localization. Analyzing the differences in expression profiles of WT and She-complex mutant strains may reveal genes that function in processes redundant with RNA transport. Alternately, a conventional synthetic lethal screen may recover alleles of genes not represented in the deletion collection, which covers only 80% of the genome. It is likely that we did not recover any of the expected synthetic interactions because proteins participating in overlapping mechanisms were essential and therefore their deletions not represented in the deletion collection. Elucidation of a phenotype for RNA transport mutants (other than

Table 1: SGA candidates showing synthetic slow growth or lethality in combination with a WT strain bearing *can1*::Nat and *lyp1*Δ markers. Genes overlapping with the Tong et al. WT synthetic lethal dataset are in green, those linked to selected markers are red, and those closely related to other genes in the Tong et al. WT dataset are purple.

Name
NUP60
DUR1,2
MET8
CHA1
MSH6
GCV3
YDR067c
NPR2
CAN1
CBC2

Table 2: SGA candidates showing synthetic slow growth or lethality in combination with _she2Δ._ Genes showing synthetic interactions with a WT strain are in green, and those linked to _SHE2_ are in red. Candidates in blue were tested by random spore analysis.

Name	Name	Name
FUS1	RLM1	YJR192c
CDC10	SPO19	HBT1
YAL053w	YPL180w	BCK1
RPL12B	LGE1	YCL076w
RIM101	YCR090c	DSE4
YAP3	HEX3	SNT309
ADH4	TYR1	GTR1
YIL029c	YIR014c	IES1
YHR130c	YGL118c	
WSS1	YDL119c	
YGL220w	YBR184w	
RTT107	YNL034w	
VMA7	SFB2	
YIL012w	YBR293w	
RMA1	YNL050c	
YKL133c	UMP1	
AAT2	MET2	
PFD1	GUP2	
OAC1	RAD4	
YKL121w	CTK2	
YKL123w	CST9	
YKL136w	YOL160w	
YKL137w	PRS3	
SSH4	YIR020w-B	
PGM1	YKL162c-A	
PMU1	YER153c	
MYO3	CCW14	
SHE2	YNL276c	
YKL131w	GAP1	
TOS4	CWH36	
RPL37a	RPS1B	
CHS1	SPT4	
LSC1	TKL1	
YNL165w	YPR077c	
YOR175c	YNL068c	
SNZ2	ERV41	
HAT1	YPR098c	
BUD7	YLL007c	

Table 3: SGA candidates showing synthetic slow growth or lethality in combination with _myo4Δ._ Genes showing synthetic interactions with a WT strain are in green, and those linked to _MYO4_ are in red. Candidates in black were viable after the SGA-derived single mutants were re-selected for presence of the _myo4Δ_::Nat allele and were discarded from subsequent analysis. Candidates in blue were tested by random spore analysis and those in purple by tetrad dissection.

Name	Name	Name	Name
RPL19B	SKY1	YBR203w	**EST3**
MET8	PLB1	**LYP1**	MGA2
PYC2	MRP144	PCL1	YOL099c
YAL031c	**SPO77**	TFP1	YJR044c
SNC1	FRE1	YNL046w	UGA4
MYO4	TSA1	YNL034w	YHK115c
YAL028w	YLR414c	**CTS1**	YHL041w
YAL026c	YLR152c	APT1	RIM101
YAL020c	YLR296w	**MAG1**	PEX8
YAL019w	YER143w	ECM4	GRE1
YAL034c	ACE2	YHR180w	GIC1
YAL036c	ZRT2	HBS1	UBR1
YBR242w	IMP2	**EDS1**	TDH3
YBR235w	YLR111w	**GRS1**	YHR049c-a
YDR433w	CUS2	YPR076w	YGL261c
CPR5	IBD2	YSA1	YHR176w
RPL12B	YOR088w	YLR101c	LAG1
DLD3	YNL179c	SIF2	YGR137w
YGL036w	RTT106	GAT4	ECM14
YEL064c	CRS5	**SER33**	INP51
PGM1	YRR1	HAC1	YGR151c
GCN3	RPA49	**YFL032w**	PDE1
YKR030w	BAG7	**RPL13a**	TOK1
LAP4	YPR014	YGR285c	PRY1
YJL178c	**PHR1**	COT1	YGR025w
ASG7	**PDE2**	YBR224w	YJL064w
MNN11	**CVT19**	YGL235w	**IME1**
AQY2	**YLR304c-a**	CYC3	LTE1
YKL187c	**UBP14**	VPS41	HSE1
MNR2	**SOL1**	YML090w	
YJL216c	DER1	MUB1	

Table 4: SGA candidates showing synthetic slow growth or lethality in combination with *myo4*Δ. The query strain contained the additional *lyp1*Δ marker for more stringent haploid selection. Genes showing synthetic interactions with a WT strain are in green, and those linked to *MYO4* are in red. Candidates in purple were tested by tetrad dissection, all others were tested by random spore analysis.

Name	Name	Name	Name
NUM1	INO1	YPL184c	YER113
CDC10	TIF2	KTR6	YKR073c
YAR037c	VRP1	SPS4	YGL235w
YAR040c	YMR196w	MAM3	YGL241w
YAR044w	YMR009w	ECM2	YKL003w-a
YBR147c	MRE11	YDL012c	CYS3
YBL083c	PEX12	YGR223c	YAF9
UIP2	RPS18B	APD1	MRP21
DEP1	YML035c-a	CDC50	SPO7
AGP1	YLR415c	MRF1	TBS1
YDR119w	ALD2	YBR014c	PBS2
SNC1	YLR416c	SSN2	MLF3
AST1	SEC22	YDL071c	CCW14
MYO4	PSP2	BUD13	ERP1
YAL028w	YNL155w	RPS16B	TOF1
YAL027w	KIN4	ZUO1	VIK1
YDR100w	CAF40	UAF30	YKR033c
PMT2	STE13	WSC2	SMF1
YAL049c	YOR082c	YER163c	SKM1
YDR063w	ODC2	YMR160w	EST3
FUN19	FIG4	RPS21B	NGL3
RVS167w	YVC1	YJL028w	YGR210c
REF2	ISU2	YAK1	HIT1
RPB9	BUD21	YKL005C	LTE1
ARG4	YMR299c	ARC18	SNT209
VMA7	YOR175c	PMD1	PHO13
SKI8	YNL228w	YLR346c	YCL075
MDR1	PDR17	YER143c	RAD57
SDH1	SAP30	YER135c	RAD2
YKL118w	ARP8	ILM1	ICL2
YLL023c	IST1	YFR039c	ENO1
RPS0B	MDM12	YER119c-a	SLH1

137

Figure 1: Schematic of SGA analysis. Markers selected at each stage are in red; the medium used for selection is indicated. The growth of haploid single and double mutants was compared, and candidate single mutant patches were manually retested on YPD+Nat to verify that these cells did not contain the additional *she* mutation.
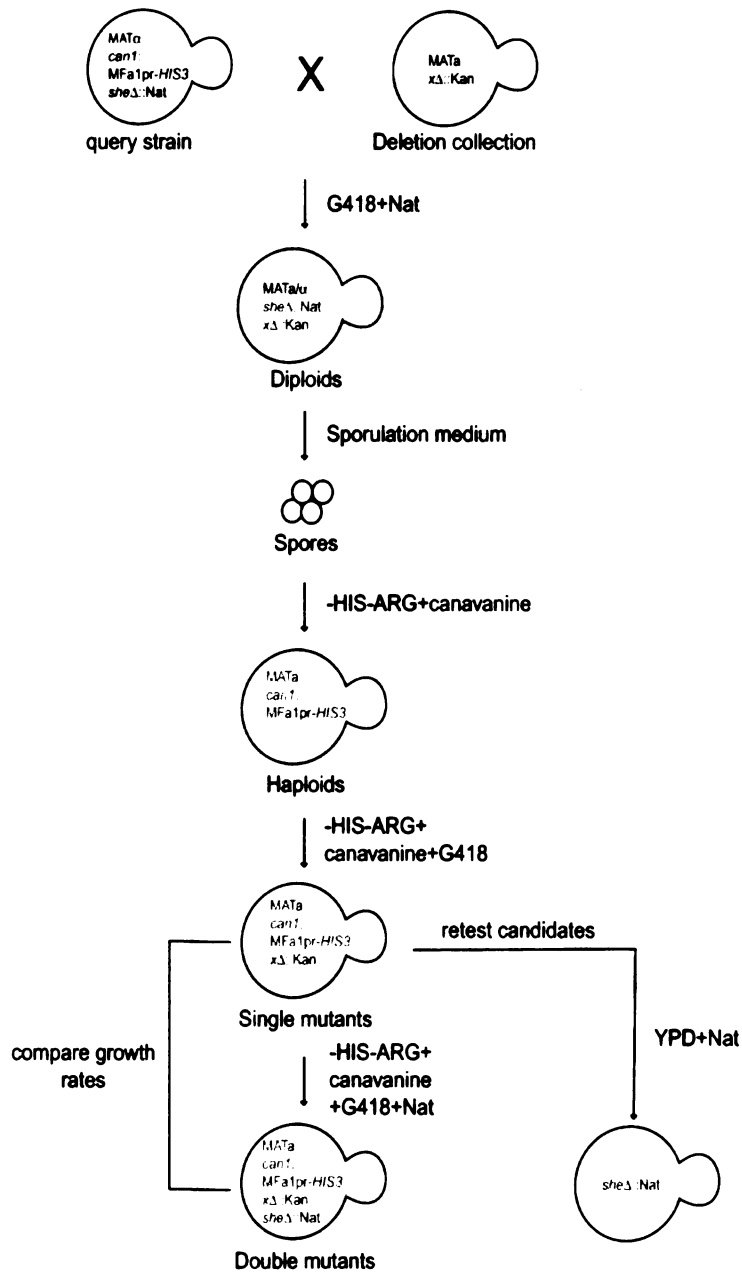
Figure 2: A slow-growth phenotype is linked to *ilm1Δ* and does not correlate with *MYO4* genotype. Tetrad dissection of a *myo4Δ*, *ilm1Δ* heterozygous diploid. Each row across is a single tetrad. Colored dots indicate the genotype of each spore clone, and tetrad type is indicated (T= tetratype, PD= parental ditype).
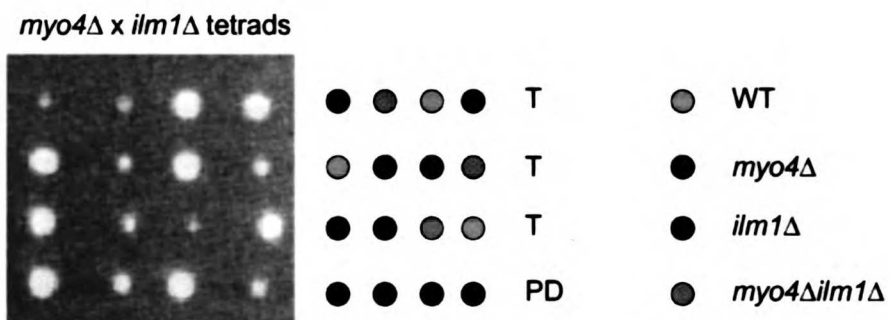
Figure 3: A factor unlinked to *MYO4* causes synthetic slow growth in a *uaf30Δ* background. Tetrad dissection of a *myo4Δ*, *uaf30Δ* heterozygous diploid. Each row across is a single tetrad. Colored dots indicate the genotype of each spore clone, and tetrad type is indicated (PD= parental ditype, T= tetratype, NPD= non-parental ditype).
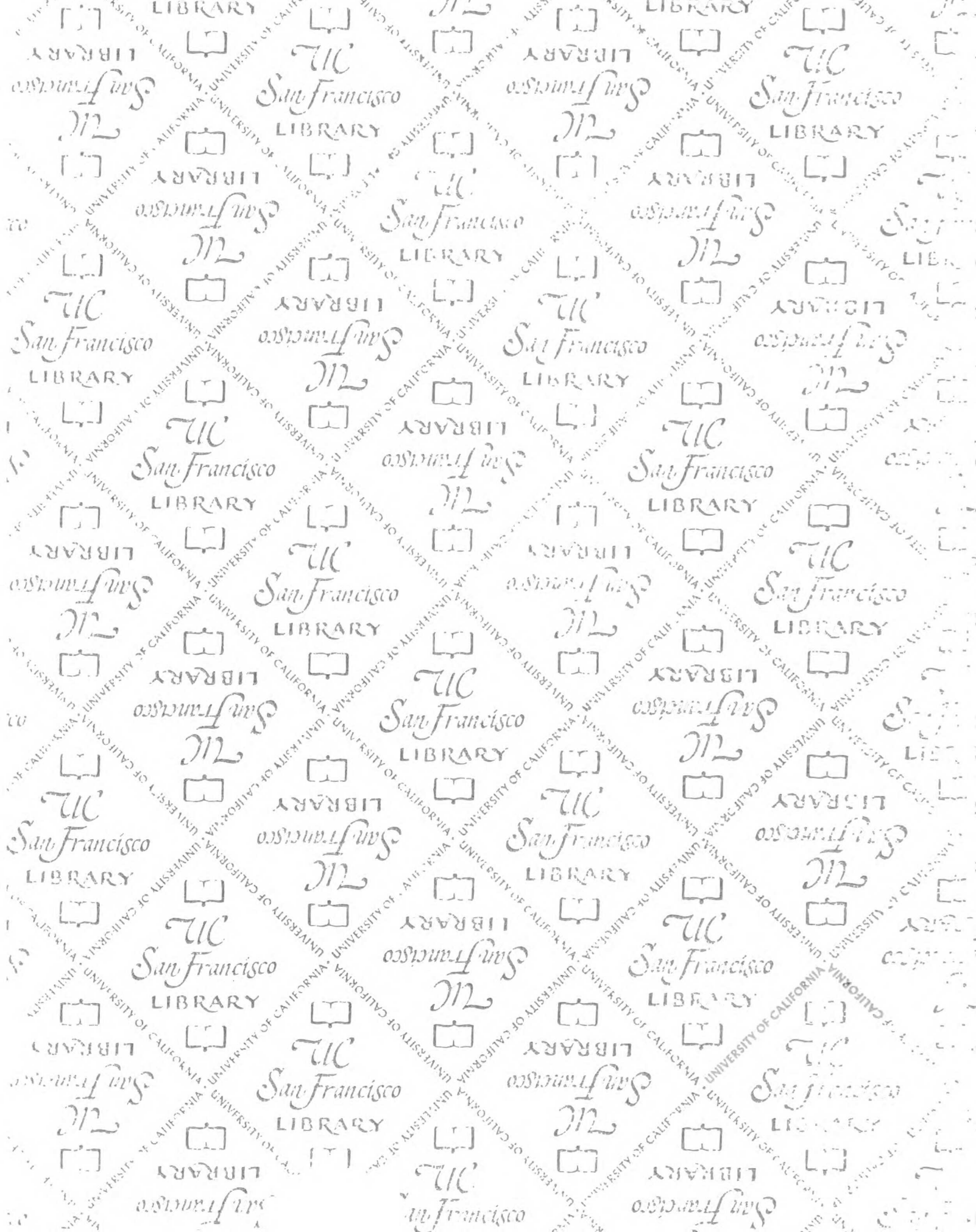
.

myo4Δ x *uaf30Δ* tetrads
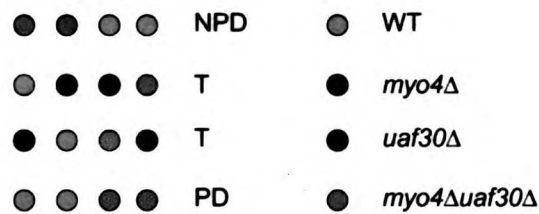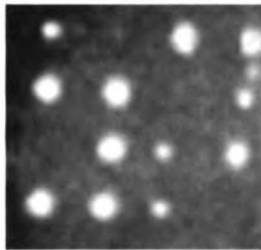
References:

1. Johnstone, O. & Lasko, P. (2001) *Annu Rev Genet* **35,** 365-406.

2. Shepard, K. A., Gerber, A. P., Jambhekar, A., Takizawa, P. A., Brown, P. O., Herschlag, D., DeRisi, J. L. & Vale, R. D. (2003) *Proc Natl Acad Sci U S A* **100,** 11429-34.

3. Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., Arkin, A. P., Astromoff, A., El-Bakkoury, M., Bangham, R., Benito, R., Brachat, S., Campanaro, S., Curtiss, M., Davis, K., Deutschbauer, A., Entian, K. D., Flaherty, P., Foury, F., Garfinkel, D. J., Gerstein, M., Gotte, D., Guldener, U., Hegemann, J. H., Hempel, S., Herman, Z., Jaramillo, D. F., Kelly, D. E., Kelly, S. L., Kotter, P., LaBonte, D., Lamb, D. C., Lan, N., Liang, H., Liao, H., Liu, L., Luo, C., Lussier, M., Mao, R., Menard, P., Ooi, S. L., Revuelta, J. L., Roberts, C. J., Rose, M., Ross-Macdonald, P., Scherens, B., Schimmack, G., Shafer, B., Shoemaker, D. D., Sookhai-Mahadeo, S., Storms, R. K., Strathern, J. N., Valle, G., Voet, M., Volckaert, G., Wang, C. Y., Ward, T. R., Wilhelmy, J., Winzeler, E. A., Yang, Y., Yen, G., Youngman, E., Yu, K., Bussey, H., Boeke, J. D., Snyder, M., Philippsen, P., Davis, R. W. & Johnston, M. (2002) *Nature* **418,** 387-91.

4. Tong, A. H., Evangelista, M., Parsons, A. B., Xu, H., Bader, G. D., Page, N., Robinson, M., Raghibizadeh, S., Hogue, C. W., Bussey, H., Andrews, B., Tyers, M. & Boone, C. (2001) *Science* **294,** 2364-8.

5. Goldstein, A. L. & McCusker, J. H. (1999) *Yeast* **15,** 1541-53.

6.   Amberg, D. C., Burke, Daniel J., Strathern, Jeffrey N. (2005) *Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor).

7.   Chen, J., Rappsilber, J., Chiang, Y. C., Russell, P., Mann, M. & Denis, C. L. (2001) *J Mol Biol* **314,** 683-94.

8.   Entian, K. D., Schuster, T., Hegemann, J. H., Becher, D., Feldmann, H., Guldener, U., Gotz, R., Hansen, M., Hollenberg, C. P., Jansen, G., Kramer, W., Klein, S., Kotter, P., Kricke, J., Launhardt, H., Mannhaupt, G., Maierl, A., Meyer, P., Mewes, W., Munder, T., Niedenthal, R. K., Ramezani Rad, M., Rohmer, A., Romer, A., Hinnen, A. & et al. (1999) *Mol Gen Genet* **262,** 683-702.

9.   Lesage, G., Shapiro, J., Specht, C. A., Sdicu, A. M., Menard, P., Hussein, S., Tong, A. H., Boone, C. & Bussey, H. (2005) *BMC Genet* **6,** 8.

10.   Siddiqi, I. N., Dodd, J. A., Vu, L., Eliason, K., Oakes, M. L., Keener, J., Moore, R., Young, M. K. & Nomura, M. (2001) *Embo J* **20,** 4512-21.

11.   Deutschbauer, A. M., Jaramillo, D. F., Proctor, M., Kumm, J., Hillenmeyer, M. E., Davis, R. W., Nislow, C. & Giaever, G. (2005) *Genetics* **169,** 1915-25.

12.   Haarer, B. K., Petzold, A., Lillie, S. H. & Brown, S. S. (1994) *J Cell Sci* **107 (Pt 4),** 1055-64.

13.   Juschke, C., Ferring, D., Jansen, R. P. & Seedorf, M. (2004) *Curr Biol* **14,** 406-11.

14.   Deshler, J. O., Highett, M. I. & Schnapp, B. J. (1997) *Science* **276,** 1128-31.

15. Trautwein, M., Dengjel, J., Schirle, M. & Spang, A. (2004) *Mol Biol Cell* **15,** 5021-37.

16. Estrada, P., Kim, J., Coleman, J., Walker, L., Dunn, B., Takizawa, P., Novick, P. & Ferro-Novick, S. (2003) *J Cell Biol* **163,** 1255-66.

17. Gavis, E. R. & Lehmann, R. (1994) *Nature* **369,** 315-8.

18. Ding, D., Parkhurst, S. M., Halsell, S. R. & Lipshitz, H. D. (1993) *Mol Cell Biol* **13,** 3773-81.

19. Hughes, T. R., Roberts, C. J., Dai, H., Jones, A. R., Meyer, M. R., Slade, D., Burchard, J., Dow, S., Ward, T. R., Kidd, M. J., Friend, S. H. & Marton, M. J. (2000) *Nat Genet* **25,** 333-7.

20. Oakes, M., Siddiqi, I., Vu, L., Aris, J. & Nomura, M. (1999) *Mol Cell Biol* **19,** 8559-69.

21. Luo, J., Vallen, E. A., Dravis, C., Tcheperegine, S. E., Drees, B. & Bi, E. (2004) *J Cell Biol* **165,** 843-55.