

Alcuni mRNA vengono localizzati in posizioni specifiche della cellula.

Quanti?

circa il 10% degli mRNA di origine materna negli oociti di *Drosophila*

circa 400 mRNA vengono localizzati nei dendriti/assoni dei neuroni, nei mammiferi

Recent estimates brought to up to 70% of early *drosophila* mRNAs showing some degree of preferential oocyte localization

Some genes have localization signals in both mRNA and protein (sometimes redundant)

Why to localize mRNA instead of proteins?

- 1) to localize proteins correctly but preventing their presence elsewhere
- 2) regulate gene expression differentially in different cell localizations
- 3) immediateness of specific local responses

In most cases, *cis*-acting elements in RNA are found in 3'-UTR  
(cases of elements within the coding sequence are known, however)

This is interesting because at least 50% of alternative 3'-UTR are tissue-specific:

Large number of AS events involving the 3' UTR

- Alternative Poly(A) site
- Alternative last exon
- Introns within 3' UTR

This may imply that differential regulation of RNA fate (localization, stability) is an important option for many genes that is controlled during RNA maturation.

**See again: Wang et al. 2008 – Alternative isoform regulation... (Fig. 2)**

Problem: if a transcript contains a intron within the 3'UTR (not that uncommon) ... does it mean that an EJC remains there after splicing and that, therefore, the RNA is always targeted to the NMD pathway ?

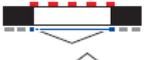
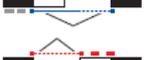
## ARTICLES

# Alternative isoform regulation in human tissue transcriptomes

Eric T. Wang<sup>1,2\*</sup>, Rickard Sandberg<sup>1,3\*</sup>, Shujun Luo<sup>4</sup>, Irina Khrebtkova<sup>4</sup>, Lu Zhang<sup>4</sup>, Christine Mayr<sup>5</sup>, Stephen F. Kingsmore<sup>6</sup>, Gary P. Schroth<sup>1</sup> & Christopher B. Burge<sup>1</sup>

Through alternative processing of pre-messenger RNAs, individual mammalian genes often produce multiple mRNA and protein isoforms that may have related, distinct or even opposing functions. Here we report an in-depth analysis of 15 diverse human tissue and cell line transcriptomes on the basis of deep sequencing of complementary DNA fragments, yielding a digital inventory of gene and mRNA isoform expression. Analyses in which sequence reads are mapped to exon-exon junctions indicated that 92–94% of human genes undergo alternative splicing, ~86% with a minor isoform frequency of 15% or more. Differences in isoform-specific read densities indicated that most alternative splicing and alternative cleavage and polyadenylation events vary between tissues, whereas variation between individuals was approximately twofold to threefold less common. Extreme or 'switch-like' regulation of splicing between tissues was associated with increased sequence conservation in regulatory regions and with generation of full-length open reading frames. Patterns of alternative splicing and alternative cleavage and polyadenylation were strongly correlated across tissues, suggesting coordinated regulation of these processes, and sequence conservation of a subset of known regulatory motifs in both alternative introns and 3' untranslated regions suggested common involvement of specific factors in tissue-level regulation of both splicing and polyadenylation.

470

Alternative transcript events	Total events ( $\times 10^3$ )	Number detected ( $\times 10^3$ )	Both isoforms detected	Number tissue-regulated	% Tissue-regulated (observed)	% Tissue-regulated (estimated)
Skipped exon 	37	35	10,436	6,822	65	72
Retained intron 	1	1	167	96	57	71
Alternative 5' splice site (A5SS) 	15	15	2,168	1,386	64	72
Alternative 3' splice site (A3SS) 	17	16	4,181	2,655	64	74
Mutually exclusive exon (MXE) 	4	4	167	95	57	66
Alternative first exon (AFE) 	14	13	10,281	5,311	52	63
Alternative last exon (ALE) 	9	8	5,246	2,491	47	52
Tandem 3' UTRs 	7	7	5,136	3,801	74	80
Total	105	100	37,782	22,657	60	68



# Subcellular mRNA Localization in Animal Cells and Why It Matters

Christine E. Holt<sup>1</sup> and Simon L. Bullock<sup>2\*</sup>

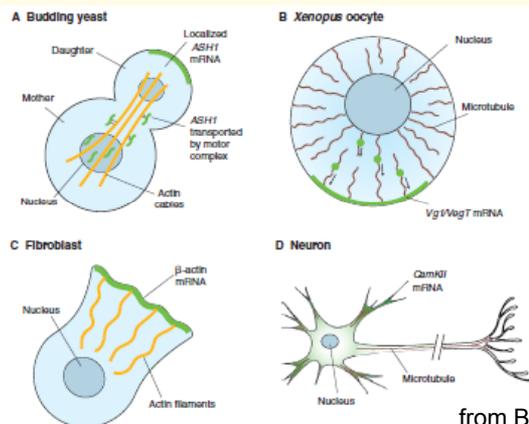
Subcellular localization of messenger RNAs (mRNAs) can give precise control over where protein products are synthesized and operate. However, just 10 years ago many in the broader cell biology community would have considered this a specialized mechanism restricted to a very small fraction of transcripts. Since then, it has become clear that subcellular targeting of mRNAs is prevalent, and there is mounting evidence for central roles for this process in many cellular events. Here, we review current knowledge of the mechanisms and functions of mRNA localization in animal cells.

1212

27 NOVEMBER 2009 VOL 326 SCIENCE www.sciencemag.org

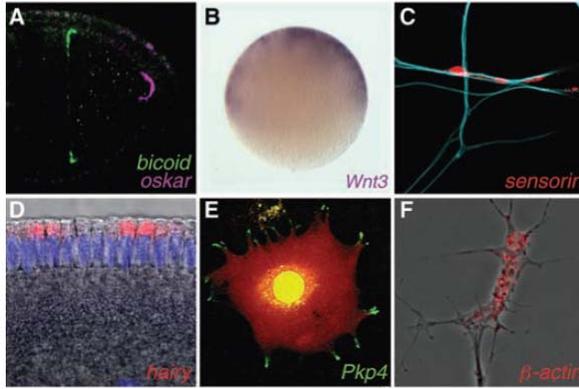
## REVIEW

### Box 1. Key examples of localized mRNAs

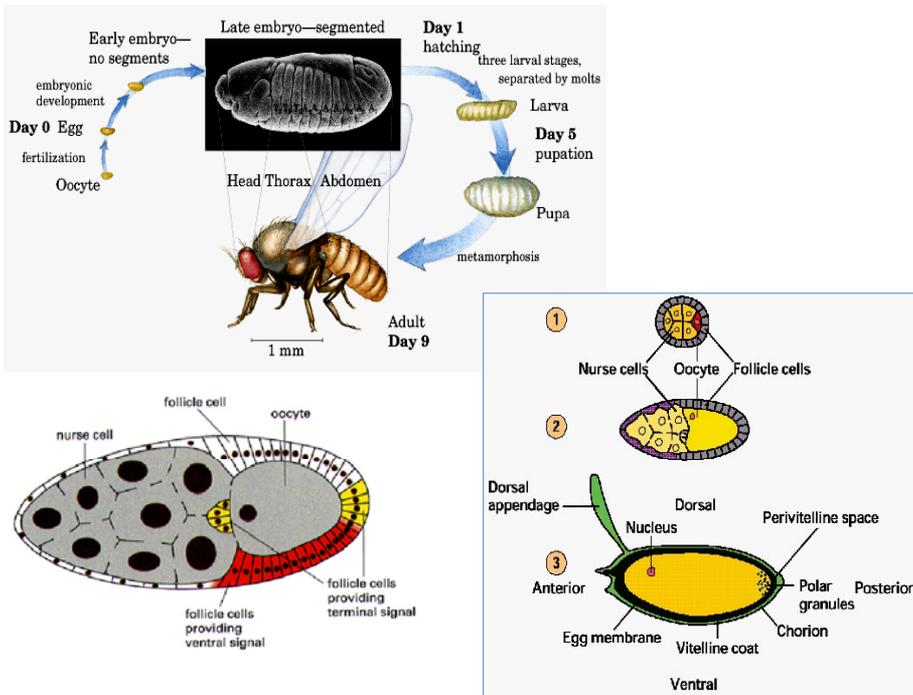


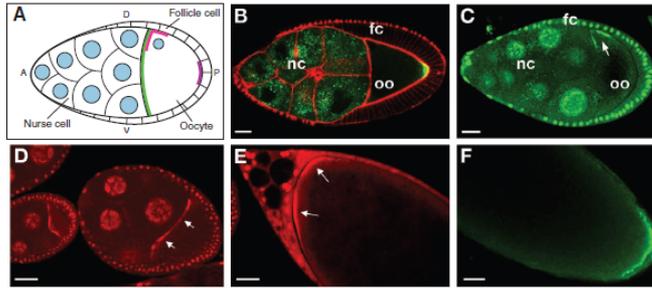
from Becalska & Gavis, 2009

In budding yeast, localization of *ASH1* mRNA to the bud tip by myosin-mediated transport on actin cables (see A in figure) targets Ash1p to the daughter cell, where it is required to repress mating type switching (reviewed by Gonzalez et al., 2005). Thus, mating type switching occurs only in the mother cell, thereby ensuring that both yeast mating types are present in the population. Germ layer specification in *Xenopus* embryos relies on localized mRNAs, including *Vg1* and *VegT* (reviewed by King et al., 2005) (see B in figure), which are transported to the vegetal pole of the oocyte by kinesin motors and anchored to the cortex by an actin-dependent mechanism. After fertilization, *Vg1* and *VegT* RNAs are inherited by the embryonic vegetal cells, where *Vg1* protein, a TGF $\beta$  homolog, participates in mesoderm specification, while *VegT*, a transcription factor, regulates endoderm specification and mesoderm induction. mRNA localization plays an important role in the polarization of somatic cells, such as fibroblasts and neurons (see C,D in figure). Localization of  $\beta$ -actin mRNA to the leading edge of migrating fibroblasts provides a high local concentration of actin monomers that drives assembly of the actin filaments needed for forward movement. Similarly,  $\beta$ -actin mRNA localization to growth cones in developing axons promotes the motility required for axon guidance (reviewed by Condeelis and Singer, 2005). Dendritic localization of RNAs like calcium/calmodulin-dependent protein kinase II $\alpha$  (*CaMKII $\alpha$* ) mRNA in hippocampal neurons facilitates a rapid response to synaptic activity in the form of local protein translation and contributes to learning and memory-related synaptic plasticity (reviewed by Martin et al., 2000).



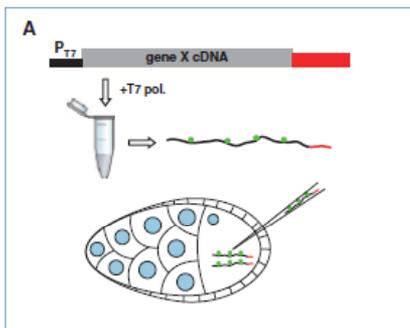
- (A) Differential localization of mRNA determinants within the *Drosophila* oocyte.  
 (B) Animal localization of a transcript encoding a signaling molecule required for axis development in the egg of a cnidarian, *Clytia*.  
 (C) mRNA enrichment in synapses of an *Aplysia* sensory neuron in response to contact with a target motor neuron (blue).  
 (D) Apical localization of an mRNA in the *Drosophila* embryo, which facilitates entry of its transcription factor product into the nuclei (purple).  
 (E) mRNA localization in pseudopodial protrusions of a cultured mammalian fibroblast (red signal indicates the cell volume).  
 (F) mRNA enrichment within a *Xenopus* axonal growth cone.  
 mRNAs were visualized by means of in situ hybridization except in (E), in which the MS2-green fluorescent protein (GFP) system was used. *Drosophila* images are reproduced from (50) with permission.





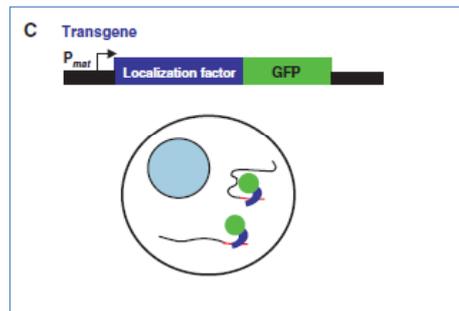
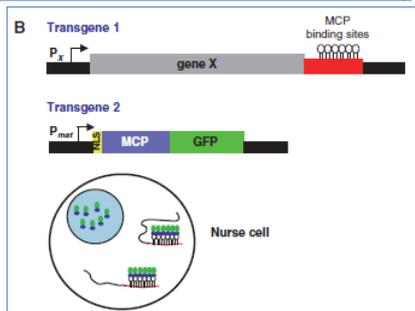
**Fig. 1. Localized distributions of *grk*, *bcd*, *osk* and *nos* mRNAs.** (A) Schematic showing *grk* (pink), *bcd* (green) and *osk* (purple) mRNA localization in mid-oogenesis (stage 9). *nos* mRNA is not yet localized at this stage. The anteroposterior (AP) and dorsoventral (DV) axes are indicated. (B) GFP-Stau (green), as proxy for *osk* mRNA, at the posterior pole of the oocyte (oo) during mid-oogenesis. GFP-Stau is also detected in the nurse cell (nc) cytoplasm. The actin cytoskeleton is highlighted in red with phalloidin. fc, follicle cells. Orientation is the same as in A. (C-F) Visualization of endogenous mRNAs using the MS2 system: (C) *grk* and (D) *bcd* during mid-oogenesis; (E) *bcd* and (F) *nos* in late oocytes. Owing to the promoter used, the MCP-GFP and MCP-RFP fusion proteins are expressed in both the nurse cells and follicle cells, whereas the MS2-tagged mRNAs are produced only in the nurse cells. MCP-GFP/RFP that is not bound to mRNA enters both the nurse cell and follicle cell nuclei. Scale bars: 20  $\mu$ m. Image in B was modified, with permission, from Huynh et al. (Huynh et al., 2004); image in C was modified, with permission, from Jaramillo et al. (Jaramillo et al., 2008); images in D and E are reproduced, with permission, from Weil et al. (Weil et al., 2006). Image in F is courtesy of K. Sinsimer (Princeton University, Princeton, NJ, USA). *bcd*, *bicoid*; *grk*, *gurken*; GFP, green fluorescent protein; MCP, MS2 coat protein; *nos*, *nanos*; *osk*, *oskar*; RFP, red fluorescent protein.

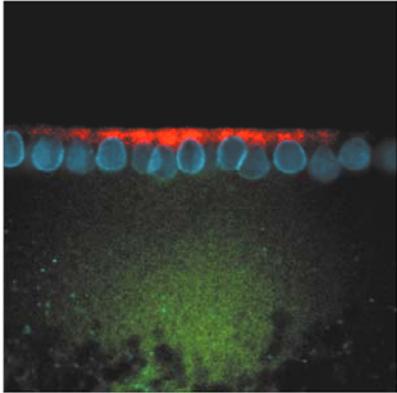
from Becalska & Gavis, 2009



How to study localization of specific mRNAs

from Becalska & Gavis, 2009

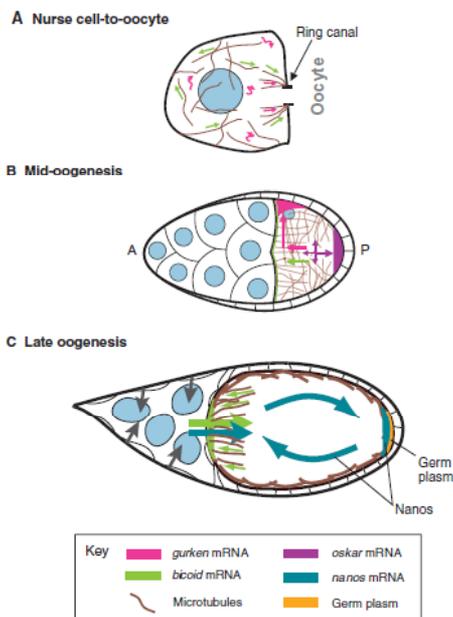




*Hairy* mRNA localization in syncytial *Drosophila* blastocyst, as determined by microinjection of fluorochrome-labeled mRNAs.

**Red** = wild-type *hairy* mRNA

**Green** = mRNA from a mRNA lacking 3'-UTR



**Fig. 3. Models for mRNA localization.** In all panels, microtubules are shown in brown, nurse cell and follicle cell nuclei in blue.

**(A) Movements of *grk* and *bcd* mRNAs within the nurse cells during mid-oogenesis.**

Straight arrows indicate directed movement on microtubules, squiggly arrows indicate movement of *grk* with cytoplasmic flows. **(B) Microtubule-dependent transport of *grk*, *bcd* and *osk* mRNAs within the oocyte during mid-oogenesis.** The oocyte nucleus is shown in gray. Colored arrows show the directions of RNA movements.

**(C) Localization of *bcd* and *nos* at late stages of oogenesis.**

Contraction of the nurse cells for dumping is indicated by gray arrows pointing inward; entry of *bcd* and *nos* into the oocyte is indicated by large straight arrows. Small green arrows depict transport of *bcd* on anterior microtubules, curved dark green arrows depict diffusion of *nos* facilitated by ooplasmic streaming.

## Possibili meccanismi di localizzazione dell'RNA

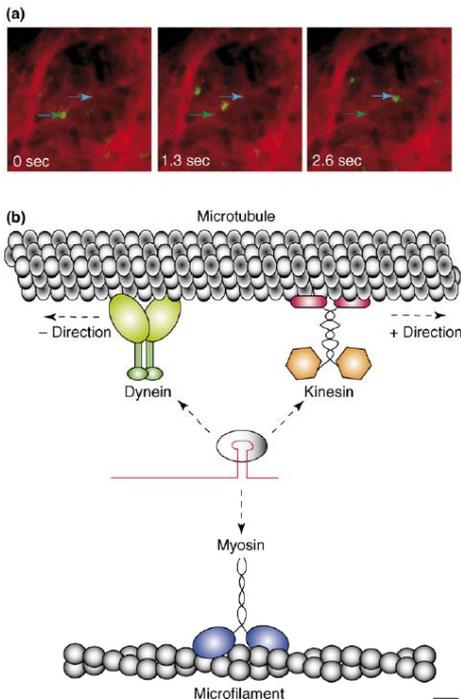
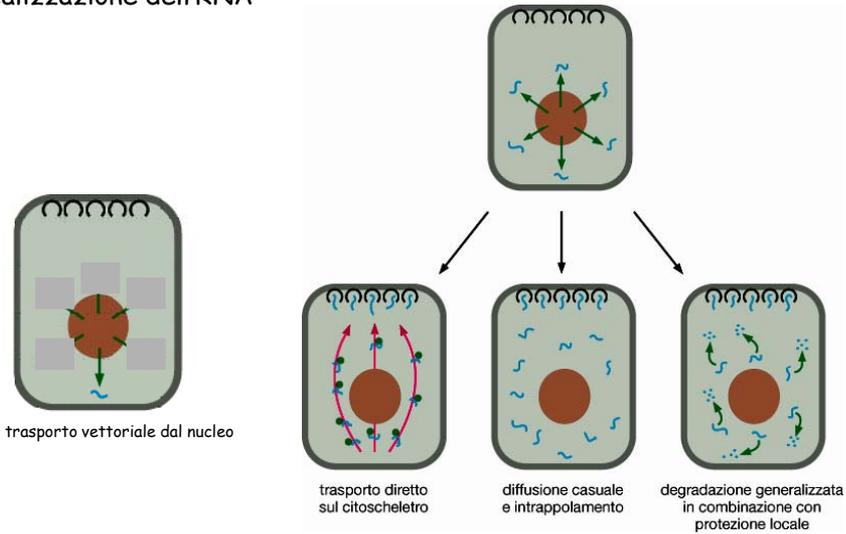
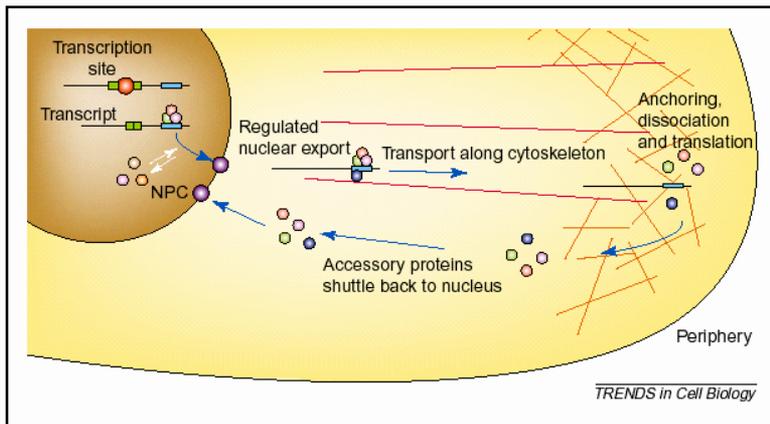
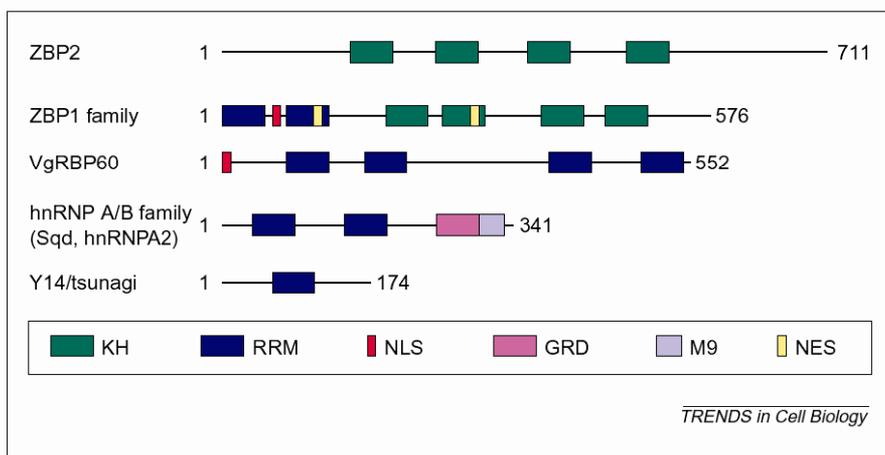


Figure 4. Interaction with cytoskeletal motors during mRNA localization. (a) A LEcontaining mRNA labeled in vivo with a GFP tag shows tracking along a microtubule in time-lapsed images. Red, tubulin; green arrow, starting position; blue arrow, ending position [70]. (b) The interactions of L-RNPs with microtubule motor complexes or microfilament motor complexes have been implicated in localization. All three types of cytoskeleton-dependent motors (kinesin, and dynein for microtubules and myosin for microfilaments) have been suggested to have roles in either directed movement or local anchoring. Based on the identified examples of L-RNP-motor protein complexes, localizing mRNA is depicted here as interacting with motors indirectly through LE-binding proteins as a RNP complex (represented for simplicity as a gray oval bound to a localizing mRNA, but can in fact be extremely large with multiple RNA-binding proteins and mRNAs). In metazoans, the molecular details of these interactions are not known.



**Fig. 2.** Model of localization factors shuttling between the nucleus and the cytoplasm. Factors associate with the RNA at the transcription site to form a localization complex or 'locasome' - a particle that is designated to localize associated RNAs, by riding with the RNA through the nuclear pore complex (NPC) where additional factors join and assemble motors. The locasome then travels to the cell periphery directed by the polarity of the cytoskeletal filaments and anchors on the isotropic microfilament network, where it releases the shuttling factors and becomes competent for translation. The shuttling factors then return to the nucleus. The cell periphery might be the growth cone of a neuron, the lamellipod of a fibroblast, the vegetal pole of a *Xenopus* oocyte or the posterior pole of a *Drosophila* embryo.

from: Farina & Singer 2002, Trends in Cell Biol, 12: 466.



**Fig. 1.** Domain structures of RNA-binding proteins with a nuclear role in cytoplasmic RNA localization. The hnRNP K homology (KH) domain is coloured green, the RNA recognition motif (RRM) dark blue, the nuclear localization signal (NLS) red, the glycine-rich domain (GRD) purple, the M9 nucleocytoplasmic shuttling domain light blue, and the putative nuclear export signal yellow.

from: Farina & Singer 2002, Trends in Cell Biol, 12: 466.

**Table 1. Properties of *trans*-acting factors associated with localized mRNAs in the nucleus and the cytoplasm<sup>a</sup>**

Protein	Organism	Cell type	mRNA	Homologues	Associated functions
<b>hnRNP A/B family</b>					
HnRNP A2	Human	Oligodendrocyte	MBP		Splicing
Sqd (hrp40)	<i>Drosophila</i>	Oocyte	<i>gurken</i>	Human hnRNP A1, <i>S. cerevisiae</i> Np13, Hrp1, <i>C. tentans</i> hrp36	mRNA processing and transport, NMD
<b>ZBP1 family</b>					
ZBP1	Chicken	Fibroblast, neuron	$\beta$ -Actin	CRD-BP, IMP1, IMP2, IMP3, KOC, HCC, dIMP	mRNA stability, translational regulation, overexpressed in cancer
Vg1RBP	<i>Xenopus</i>	Oocyte	<i>Vg1</i> , <i>Veg T</i>	See above	See above
ZBP2	Chicken	Fibroblast, neuron	$\beta$ -Actin	KSRP, MARTA	Splicing
She2p	<i>S. cerevisiae</i>	Budding yeast	<i>ASH1</i>	NK	NK
Loc1p	<i>S. cerevisiae</i>	Budding yeast	<i>ASH1</i>	NK	NK
Tsunagi	<i>Drosophila</i>	Oocyte	<i>Oskar</i>	Human Y14	EJC, NMD

<sup>a</sup>Abbreviations: EJC, exon-exon junction complex; NK, not known; NMD, nonsense-mediated mRNA decay.

from: Farina & Singer 2002, Trends in Cell Biol, 12: 466.

The *trans*-acting factor **hnRNP A2** is involved in the localization of myelin basic protein (MBP) mRNA to the myelin-forming processes of mammalian oligodendrocytes.

Like many of the proteins involved in RNA localization, hnRNP A2 has several other functions, including splicing, nuclear export, translational regulation and RNA stabilization.

As a *trans*-acting factor of MBP mRNA localization, hnRNP A2 binds to a 21-nucleotide (nt) *cis* element called the hnRNP A2 response element (A2RE), formerly known as the 'RNA trafficking sequence' (RTS), that is located in the 3' untranslated region (UTR) of the mRNA.

Although it is a predominantly nuclear protein, hnRNP A2 shuttles to the cytoplasm, by means of its M9 nucleocytoplasmic shuttling domain, where it localizes in cytoplasmic granules [16].

### Sqd localizes gurken mRNA

The *Drosophila* protein Sqd (hs hnRNPA1) is an RNA-binding protein of 42 kDa that is needed for the proper localization of *gurken* (*grk*) transcripts during oogenesis. Like hnRNPA2, Sqd is a member of a class of the hnRNPs that shuttle between the nucleus and the cytoplasm through an M9 shuttling motif.

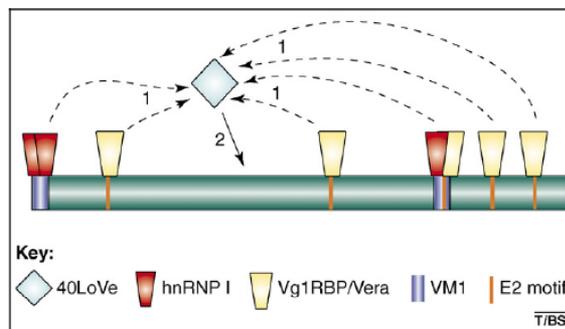
Dorsoventral patterning in *Drosophila* requires localization of *grk* mRNA to the dorsoanterior corner of the oocyte. In *sqd* mutants, *grk* mRNA is mislocalized at the anterior of the oocyte, and does not accumulate anterodorsally.

Sqd protein is thought to associate with *grk* mRNA in the nucleus and to deliver it to cytoplasmic anchors at the dorsoanterior of the oocyte.

Sqd is known to bind directly to the 3' UTR of *grk* mRNA.

How are cis-elements made ?

50-350 nt long, sometimes very complex

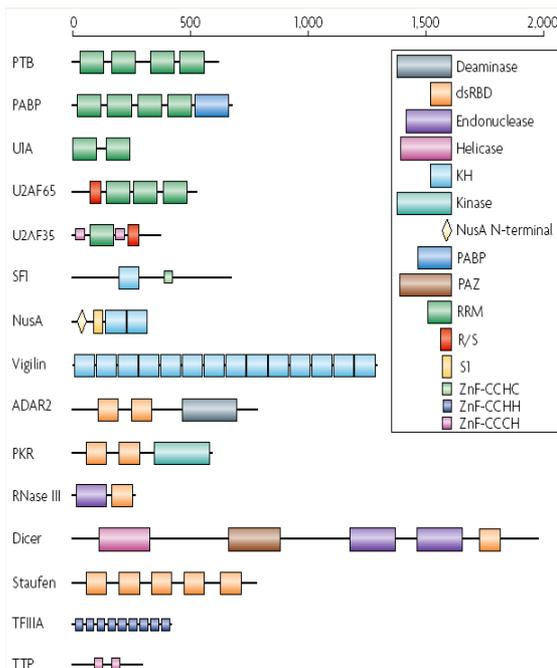


**Figure 1.** The LE of Vg1 mRNA. Vg1 is a TGF- $\beta$  superfamily protein involved in mesoderm induction during *Xenopus* embryogenesis. The Vg1 mRNA localizes to the vegetal pole of oocytes through a 360-nucleotide LE in the 3' untranslated region. The LE is depicted in green; the motifs that have been identified as required for localization are highlighted, as are the *trans*-acting factors that are known to bind these sequences. The binding of 40LoVe (pale blue diamond) to an unidentified site occurs after previous recognition of the Vg1 LE by the RNA-binding proteins hnRNP I and Vg1RBP/Vera (as indicated by the numbered arrows).

# RNA-binding proteins: modular design for efficient function

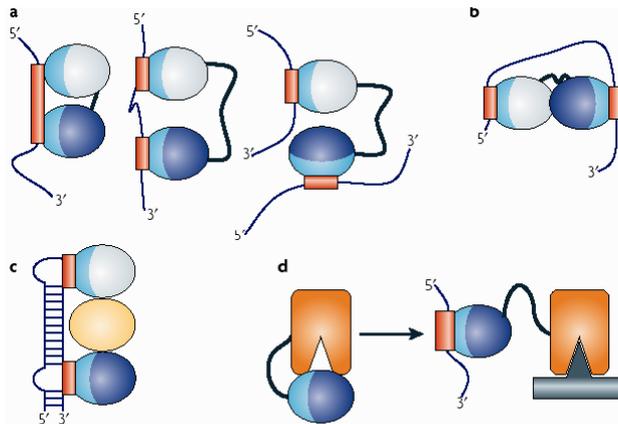
Bradley M. Lunde<sup>1,5</sup>, Claire Moore<sup>1</sup> and Gabriele Varani<sup>\*†</sup>

**Abstract** | Many RNA-binding proteins have modular structures and are composed of multiple repeats of just a few basic domains that are arranged in various ways to satisfy their diverse functional requirements. Recent studies have investigated how different modules cooperate in regulating the RNA-binding specificity and the biological activity of these proteins. They have also investigated how multiple modules cooperate with enzymatic domains to regulate the catalytic activity of enzymes that act on RNA. These studies have shown how, for many RNA-binding proteins, multiple modules define the fundamental structural unit that is responsible for biological function.

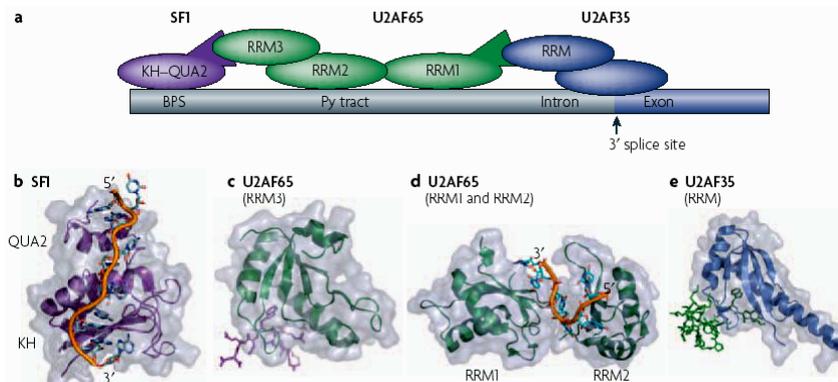


**Many RNA-binding proteins have a modular structure.** Representative examples from some of the most common RNA-binding protein families, as illustrated here, demonstrate the variability in the number of copies (as many as 14 in vigilin) and arrangements that exist. This variability has direct functional implications. For example, Dicer and RNase III both contain an endo-nuclease catalytic domain followed by a double-stranded RNA-binding domain (dsRBD). So, both proteins recognize dsRNA, but Dicer has evolved to interact specifically with RNA species that are produced through the RNA interference pathway through additional domains that recognize the unique structural features of these RNAs. Different domains are represented as coloured boxes.

These include the RNA-recognition motif (RRM; by far the most common RNA-binding protein module), the K-homology (KH) domain (which can bind both single-stranded RNA and DNA), the dsRBD (a sequence-independent dsRNA-binding module) and RNA-binding zinc-finger (ZnF) domains. Enzymatic domains and less common functional modules are also shown. PABP, poly(A)-binding protein; PTB, polypyrimidine-tract binding; R/S, Arg/Ser-rich domain; SFI, splicing factor-1; TTP, tristetraprolin; U2AF, U2 auxiliary factor.



**Figure 2 | RNA-binding modules are combined to perform multiple functional roles.** RNA-binding domains (RBDs) function in various ways. **a** | They recognize RNA sequences with a specificity and affinity that would not be possible for a single domain or if multiple domains did not cooperate. Multiple domains combine to recognize a long RNA sequence (left), sequences separated by many nucleotides (centre), or RNAs that belong to different molecules altogether (right). **b** | RBDs can organize mRNAs topologically by interacting simultaneously with multiple RNA sequences. **c** | Alternatively, they can function as spacers to properly position other modules for recognition. **d** | They can combine with enzymatic domains to define the substrate specificity for catalysis or to regulate enzymatic activity. The RNA-binding modules are represented as ellipses with their RNA-binding surfaces coloured in light blue, and the corresponding binding sites in the RNA coloured in red; individual domains are coloured differently.



**Figure 4 | Protein–protein interactions and protein–RNA interactions define the site of spliceosomal assembly.** **a** | Schematic of the interactions between various proteins and RNA at the splicing site. The structures of some of the key domains that are involved in these interactions are shown in panel **b**. In the RNA, the branch-point sequence (BPS), pyrimidine tract (Py tract), and the 3' splice site are labelled with the intron shown in grey and the exon in dark blue. **b** | Splicing factor-1 (SF1) recognizes the BPS through its K-homology (KH)–Quaking homology-2 (QUA2) domains, which creates an extended KH domain that can recognize the full BPS sequence RNA<sup>42</sup>. **c** | This interaction is strengthened by protein–protein interactions between the N terminus of SF1 and the non-canonical RNA-recognition motif-3 (RRM3) of U2 auxiliary factor subunit-65 (U2AF65)<sup>35</sup>. **d** | RRM3 is bound to the pyrimidine tract through its first two canonical RRM, RRM1 and RRM2 (REF. 10). **e** | Last, the U2AF65 interaction is also aided by protein–protein interactions between its N terminus and the non-canonical RRM of U2AF35 bound at the 3' splice site<sup>35</sup>. The protein and peptide structures are colour coded as in panel **a**.

Methods to identify RNA elements:

RNA SELEX

RNA recombination and selection

Comparison of regulatory sequences

Immunoprecipitation or CLIP followed by microarrays or RNA-Seq

RNAcompete

## Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins

Debashish Ray<sup>1,4</sup>, Hilal Kazan<sup>2,4</sup>, Esther T Chan<sup>3</sup>, Lourdes Peña Castillo<sup>1</sup>, Sidharth Chaudhry<sup>3</sup>, Shaheynoor Talukder<sup>1</sup>, Benjamin J Blencowe<sup>1,3</sup>, Quaid Morris<sup>1-3</sup> & Timothy R Hughes<sup>1,3</sup>

Metazoan genomes encode hundreds of RNA-binding proteins (RBPs) but RNA-binding preferences for relatively few RBPs have been well defined<sup>1</sup>. Current techniques for determining RNA targets, including *in vitro* selection and RNA co-immunoprecipitation<sup>2-5</sup>, require significant time and labor investment. Here we introduce RNAcompete, a method for the systematic analysis of RNA binding specificities that uses a single binding reaction to determine the relative preferences of RBPs for short RNAs that contain a complete range of k-mers in structured and unstructured RNA contexts. We tested RNAcompete by analyzing nine diverse RBPs (HuR, Vts1, FUSIP1, PTB, U1A, SF2/ASF, SLM2, RBM4 and YB1). RNAcompete identified expected and previously unknown RNA binding preferences. Using *in vitro* and *in vivo* binding data, we demonstrate that preferences for individual 7-mers identified by RNAcompete are a more accurate representation of binding activity than are conventional motif models. We anticipate that RNAcompete will be a valuable tool for the study of RNA-protein interactions.

7-base (7loop) and 8-base loop (8loop) sequences in the context RNA hairpins containing unique 10-base pair stems. RNAs in the unstructured category should be either linear or contain weak secondary structures under our assay conditions; most RNAs interact intramolecularly under physiological conditions, and therefore it is not possible to design a large and diverse population of entirely linear RNAs. Moreover, constraints were introduced to minimize folding of unstructured RNAs, misfolding of the structured RNAs, extensive base-pairing among any two RNAs and microarray cross-hybridization. This resulted in reduction of desired coverage. The two sets of unstructured RNAs in the final array design each contain 81% of possible 10-mers, but contain many instances of shorter k-mers. For example, each set contains at least 12 copies of all possible 8-mers, at least 64 copies of all possible 7-mers, with the exception of the one containing a SapI restriction site (GCTCTC/GAAGAGC) used for pool synthesis (see below). It also contains 59% of all possible 8loops (75% of which are in duplicate), 99.4% of all possible 7loops (99.3% of which are in duplicate) and all possible loops of six bases or less (100% of which are in duplicate). Thus, the pool contains indepe



Protein	Domain(s)	Known motif/ binding site	Our highest- correlating motif	Top five 7-mers	Correlation 7-mers motif
Vts1	One SAM domain			CCUGCCC CCGGCCC CCGGCCC CCGGCCC CCUGCCC	0.53 0.50
SLM2	One KH domain	NA		AGAAAA AUAUAAC TAAMAA AAAMAA AUAUAAC	0.66 0.35
YB1	Full-length; one cold-shock domain			CCUGCCG CCUGCCG CCUGCCG CCUGCCG CCUGCCG	0.32 0.14
RBM4	Full-length; two RRM domains	NA		CCGGCCC CCGGCCC CCGGCCC CCGGCCA CCGGCCG	0.56 0.48
SF2/ASF	Two RRM domains			CCGAGCC CCGGCCG CCGGCCG CCGGCCG AGCCGCG	0.62 0.48
FUSIP1	One RRM domain			AAAGAAA CAAGAAA AAAGAAC AAAGAAA ACAGAAA	0.32 0.43
HuR	Full-length; three RRM domains			UUUUUUU UUUUUUU UUUUUUU UUUUUUU UUUUUUU	0.62 0.56
U1A	One of two RRM domains			AUUGCAC UUGCACCA UUGCACG UUGCACU UUGCACU	0.54 0.20
PTB	Full-length; four RRM domains			CUUUUUU UUUUUUU UUUUUUU UUUUUUU UUUUUUU	0.39 0.30

## Unbiased selection of localization elements reveals cis-acting determinants of mRNA bud localization in *Saccharomyces cerevisiae*

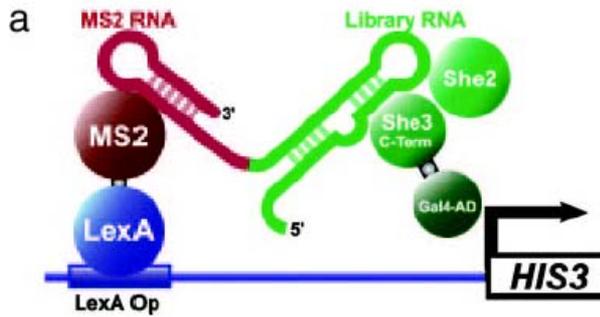
Ashwini Jambhekar\*, Kimberly McDermott<sup>1</sup>, Katherine Sorber\*, Kelly A. Shepard<sup>2</sup>, Ronald D. Vale<sup>1,2</sup>, Peter A. Takizawa<sup>1</sup>, and Joseph L. DeRisi<sup>1,2</sup>

Departments of <sup>1</sup>Biochemistry and Biophysics and <sup>2</sup>Cellular and Molecular Pharmacology, and <sup>3</sup>Howard Hughes Medical Institute, University of California, San Francisco, CA 94107; and <sup>4</sup>Department of Cell Biology, Yale University School of Medicine, New Haven, CT 06520

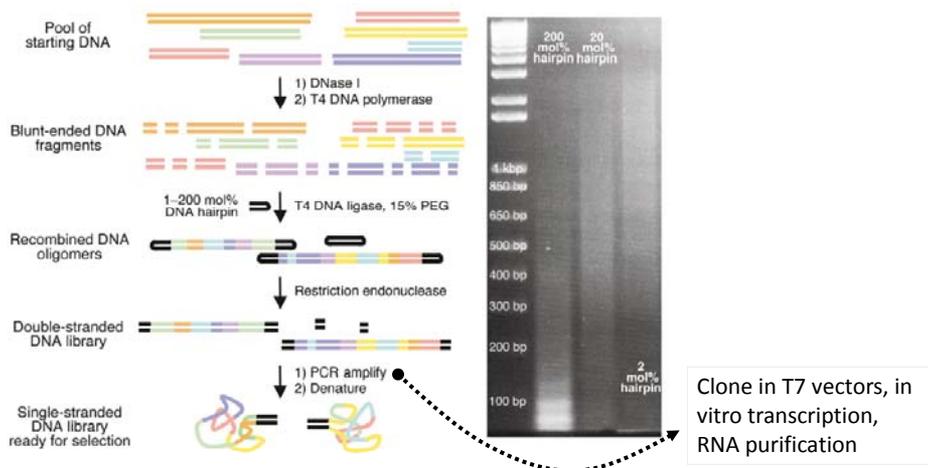
Cytoplasmic mRNA localization is a mechanism used by many organisms to generate asymmetry and sequester protein activity. In the yeast *Saccharomyces cerevisiae*, mRNA transport to bud tips of dividing cells is mediated by the binding of She2p, She3p, and Myo4p to coding regions of the RNA. To date, 24 bud-localized mRNAs have been identified, yet the RNA determinants that mediate localization remain poorly understood. Here, we used nonhomologous random recombination 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 and 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 used here should be applicable for elucidating minimal RNA motifs involved in many other types of interactions.

Article

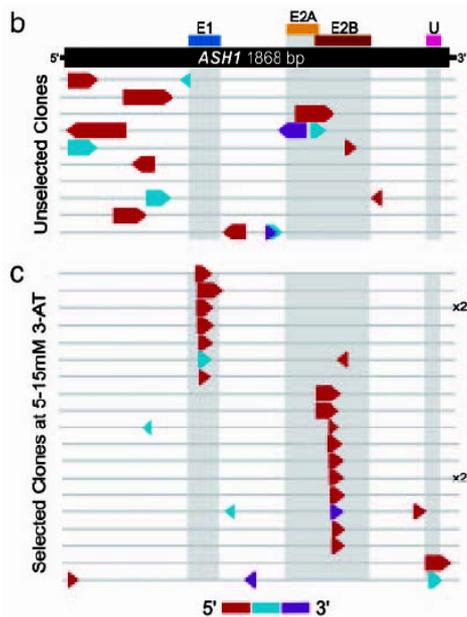
From: Jambhekar et al. (2005),  
PNAS 102: 18005-10.



**Fig. 1.** Three-hybrid scheme for selection of She3p-interacting RNA fragments. (a) Schematic of three-hybrid assay



**Figure 2.** Overview of the nonhomologous random recombination (NRR) method. (A) Starting DNA sequences are randomly digested with DNase I, blunt-ended with T4 DNA polymerase, and recombined with T4 DNA ligase under conditions that strongly favor intermolecular ligation over intramolecular circularization. (B) A defined stoichiometry of hairpin DNA added to the ligation reaction controls the average length of the recombined products. The completed ligation reaction is digested with a restriction endonuclease to provide a library of double-stranded recombined DNA flanked by defined primer-binding sequences.

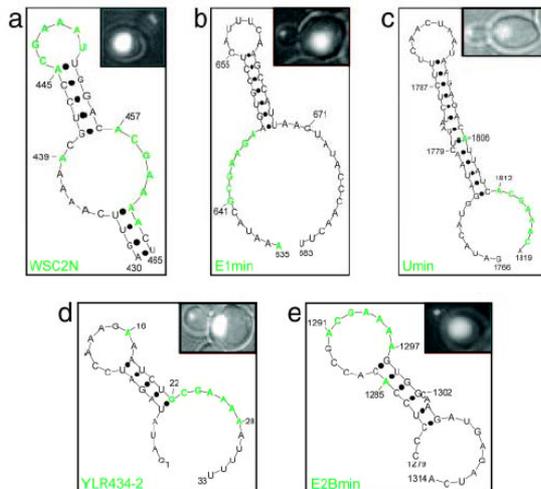


**Fig. 1.** Representation of *ASH1* NRR library members before (b) and after (c) three-hybrid selection. Each arrow represents a fragment from *ASH1*. The direction of the arrowhead indicates whether the fragment is expressed in the sense (right) or antisense (left) orientation from the three-hybrid 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.

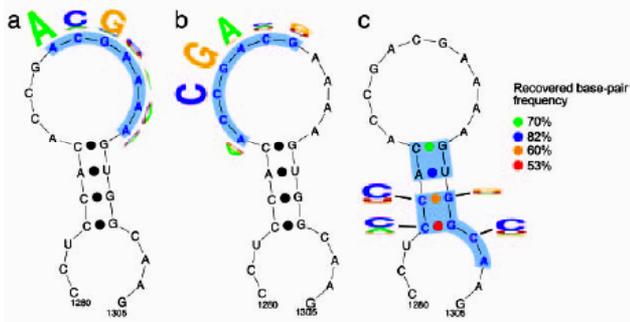
**Table 1. Summary of elements identified by NRR/three-hybrid selection**

Zipcode	Coordinates	Length, nt	Three-hybrid activity	No. of times recovered	Percent 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	418–471	54	++	14	>90
WSC2C	1313–1384	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 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 three-hybrid assay was assessed by the 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. Percent localized refers to the percent of cells with exclusively bud-localized RNA. N/D, not determined.



**Fig. 2.** Sequences and predicted structures of WT zipcodes. Bases identified by MEME analysis are green. (a) WSC2N. (b) E1min. (c) Umin. (d) YLR434-2. (e) E2Bmin. Bases are numbered with the adenosine of the start codon as 1, with the exception of YLR434-2, which is 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.



**Fig. 5.** Predicted secondary structure for E2Bmin and sequence logos derived from randomization and three-hybrid selection of bases 1291–1297 (a), 1287–1293 (b), or 1283–1286 and 1298–1303 (c). The height of each letter is proportional to the fraction of the observed frequency relative to the expected frequency at each position (22, 23). The color of each dot in c indicates the frequency of base-pairing among the selected clones.