AD_____

Award Number: DAMD17-02-1-0364

TITLE: Genome-Wide Nucleic Acid/Protein Interaction in Breast Cancer

PRINCIPAL INVESTIGATOR: Pamela A. Silver, Ph.D.

CONTRACTING ORGANIZATION: Dana-Farber Cancer Institute
Boston, Massachusetts 02115

REPORT DATE: May 2004

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

# REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188



| 1. AGENCY USE ONLY (Leave blank) | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | May 2004 | Annual (15 Apr 2003 - 14 Apr 2004) |

**4. TITLE AND SUBTITLE**
Genome-Wide Nucleic Acid/Protein Interaction in Breast Cancer

**5. FUNDING NUMBERS**
DAMD17-02-1-0364

**6. AUTHOR(S)**
Pamela A. Silver, Ph.D.

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Dana-Farber Cancer Institute
Boston, Massachusetts 02115

*E-Mail:* pamela_silver@dfci.harvard.edu

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**

**11. SUPPLEMENTARY NOTES**

**12a. DISTRIBUTION / AVAILABILITY STATEMENT**
Approved for Public Release; Distribution Unlimited

**12b. DISTRIBUTION CODE**

**13. ABSTRACT (Maximum 200 Words)**

Since many types of breast cancer remain untreatable, the research proposal aims to develop novel genomic technology to identify potential therapeutic targets and to aid in diagnosing various types of breast cancer at the molecular level. The overarching goal of the proposal is to develop a technology to screen nucleic-acid protein interactions on a genome scale with a focus on understanding complexes involved in breast cancer. In order to identify the regulatory networks of interactions between RNAs and proteins, we propose to develop a rapid genome-scale method to determine the specific RNA targets and RNA binding sites of proteins. The aims were to 1) discover RNA targets of specific RNA binding proteins and protein complexes focusing on those involved in breast cancer, and 2) define the RNA sequences recognized by proteins using novel technologies. In the current funding period, we have made significant progress on using a genome-wide approach to assess the association of RNA binding proteins with alternatively spliced genes involved in cancer. In addition, we have generated preliminary data suggesting that a regulator of RNA processing affects the expression of certain exons in response to estrogen.

**14. SUBJECT TERMS**
RNA-protein interactions, DNA-protein interactions, genomic screening, regulation of gene expression, technology development

**15. NUMBER OF PAGES**
14

**16. PRICE CODE**

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| Unclassified | Unclassified | Unclassified | Unlimited |

## Table of Contents

## INTRODUCTION:

Since many types of breast cancer remain untreatable, the research proposal aims to develop novel genomic technology to identify potential therapeutic targets and to aid in diagnosing various types of breast cancer at the molecular level. The overarching goal of the proposal is to develop a technology to screen nucleic-acid protein interactions on a genome scale with a focus on understanding complexes involved in breast cancer. In order to identify the regulatory networks of interactions between RNAs and proteins, we proposed to develop a rapid genome-scale method to determine the specific RNA targets and RNA binding sites of proteins. The aims were to 1) discover RNA targets of specific RNA binding proteins and 2) define the RNA sequences recognized by proteins using novel nanotechnologies including development of optically encoded beads containing both a unique optical signature and a specific oligonucleotide. This technology is being complemented by genome-wide chromatin immunoprecipitation. Progress during the past year has been made on Aims 2 and 4 of the original grant as detailed below.

## BODY:

Aim 1 was successfully completed as marked by the publication of a paper describing the assay (Brodsky and Silver, 2002). Thus, we have laid the groundwork for genomic and small molecule screening using the microbead assay. The assay has generated some interest in the community as we have recently written an invited review discussing the microbead assay we have developed (Brodsky et al., 2003). Although the goals remain the same, we are now using a new and potentially more powerful method than originally proposed in Aim 2. The research accomplishments associated with each task outlined in the approved Statement of Work are detailed below.

### Technical Objective 2: Identification of target RNAs of clinically important proteins

### Task 2: Search for RNA targets by comparing complexes between breast cancer and normal cells.

While the new microbead technology is being developed, current microarray technologies could be used to determine candidate binding targets. Recently, Chromatin Immunoprecipitation (CHiP) has emerged as a powerful method to identify where on a gene and - in combination with microarrays (ChiP-chip) - on which genes chromatin associated proteins are binding (Lei et al., 2001; Lei and Silver, 2003). Briefly, cells are cross-linked and chromatin is sheared to approximately 400 bp average size. The protein of interest is immunoprecpitated and the DNA is isolated for quantitative PCR analysis or microarrays. Because our lab, as well as others, has shown that many RNA binding proteins bind co-transcriptionally, we can take advantage of this approach. In addition to representing a more *in vivo* situation, a potentially significant advantage of CHiP is that the cells are formaldehyde cross-linked allowing the capture of dynamic interactions.

We have modified and improved the ChIP approach to localize RNA Binding Proteins (RBPs) on various genes. Some of our modifications include the use of a second protein-protein crosslinker in addition to the commonly used formaldehyde. We also use

modern log-linear fitting of real-time PCR data which enhances our sensitivity and dynamic range of the analysis.

Initially, we are focusing on multifunctional RNA binding proteins such as Polypyrmidine Binding Protein (PTB) and U2AF65. These proteins are known to regulate the alternative splicing and expression, respectively, of genes critical to cell growth and apoptosis such as caspases and fibroblast growth factors. We have tested a number of antibodies and have found that PTB and U2AF65 can be co-immunoprecipitated with their target genes as shown in Figure 1. This work used suspended HeLa cells which allowed us to prepare large quantities of chromatin to develop the approach. We observe these RBPs differentially localizing to a variety of genes including for example MDM2.

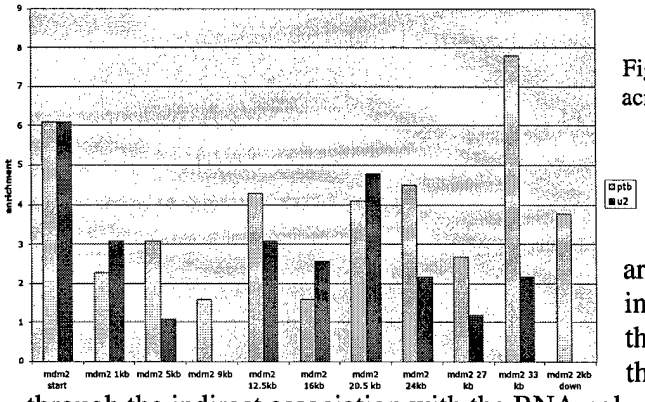



Figure 1. Enrichment of PTB and U2AF65 across mdm2.

Interestingly, these factors are associating with promoter regions in at least some genes. We expect that these RBPs are binding the RNA and that the association with DNA is through the indirect association with the RNA polymerase II complex. To determine if the observed localizations are transcription dependent, we treated the cells with the RNA Polymerase II inhibitor α-amanitin. As expected, the observed enrichments within the gene regions is reduced when transcription is inhibited. Surprisingly, the observed RBP enrichment around the transcription start site appears to be independent of RNA Polymerase II. This suggests that these RBPs are recruited to the genes by some other mechanism such as binding to initiation factors. This results may also imply that these RBPs are regulating expression by other pre-transcriptional processes. Future experiments will explore these possibilities.

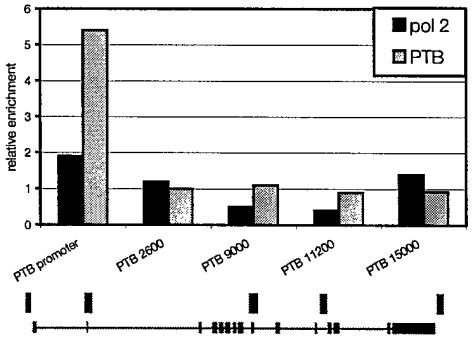



Figure 2. A) Polymerase and PTB across the ptb gene. B) α-amanitin reduces the polymerase and ptb signal except the PTB signal near the transcription start site.

**Task 3. Screen RNA-protein interactions of a targeted set of genes.**

To determine the binding profile of a larger number of genes we designed and constructed our own microarrays to probe a larger number of genes. Open Reading Frames (ORFs) were cloned from cDNA libraries, sequence verified and spotted onto

slides. This array includes a number of genes relevant to breast cancer including MDM2 and numerous kinases. A representative array is shown in

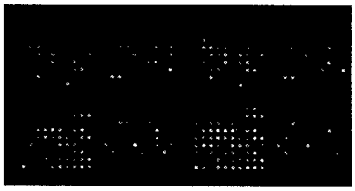


Figure 3. Representative ORF array. The range of colors suggest that some genes are highly bound and others are not.

ChIP-chip experiments with these ORF microarrays identified a number of new putative targets for PTB including SNK, DAPK3 and MDM2. These genes are rich in alternative splicing. We have verified DAPK3 and MDM2 (see Figure 1) by PCR analysis. Interestingly, DAPK3 shows approximately 10-fold stronger enrichment at its 3' end than we have observed for any other region for PTB.

**Task 4. Analyze RNA-protein interactions on a genomic scale.**

We are part of an early access program to use Affymetrix tiled arrays. These arrays include a 25mer oligonucleotide probe every 35 base pairs across regions of chromosomes. We have performed tiled array experiments with a number of antibodies with both chromosome 21/22 and ENCODE arrays as listed in Table 1.

Table 1. Antibodies used to probed various tiled arrays.

| Antibody | Target | Arrays probed to date |
|---|---|---|
| 4H8 | Hyper- and hypo-phosphorylated RNA Polymerase II | ENCODE |
| 8WG16 | Hypophosphorylated RNA Polymerase II | ENCODE |
| Anti-SM | SM core splicing factors | ENCODE |
| Anti-PTB | Polypyrmidine Tract Binding Protein splicing factor | ENCODE Chr21/22 |

Similar to the directed PCR data, we observe signal near various exons for the hyper/hypo forms of Pol II and SM, but not the hypo form.

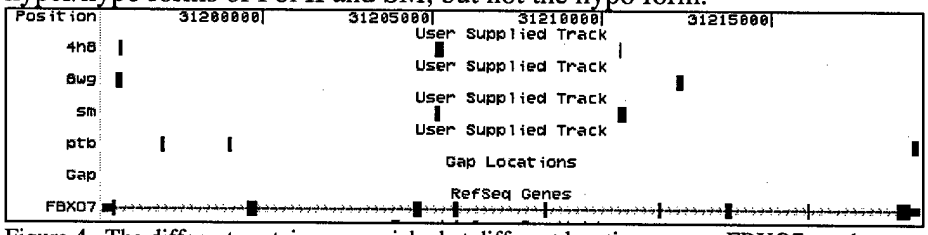



Figure 4. The different proteins are enriched at different locations across FBXO7 on chromosome 22. ON this gene, the 4H8 and SM are at the same locations across the gene suggesting that the polymerase is accumulating at places where there is co-transcriptional alternative splicing.

Across genes, the 4H8 antibody to PolII shows interesting patterns of putative enrichment locations; often around annotated exons. One interpretation is that these enrichments represent places where the RNA Polymerase is stalled and/or paused during transcription. Many of these locations are alternative exons or places where there is an alternative start or stop site for a gene. Preliminary analysis across chromosomes 21 and 22 suggest that approximately half the SM and 4H8 sites overlap. Across the same chromosomes, about

one third of the PTB sites overlap with 4H8 ones. We are analyzing these overlaps using statistical methods to determine how common and significant RNA polymerase, SM and PTB are enriched at the same locations.

PTB enrichments also are found in poorly annotated or unannotated regions. PTB is often a repressor, preventing the inclusion of exons in a transcript. Interestingly, we observe putative PTB enrichments in regions near only a genscan predicted exon. Thus this prediction may be correct and the condition that the exon is being expressed has not yet been identified. In addition, other PTB locations have no annotation, suggesting that these regions may also be expressed in some yet to be found condition.

During the development of the use of Affymetrix tiled arrays for ChIP-chip experiments, we have worked closely with Myles Brown's lab who are performing similar experiments with estrogen receptor. We are now applying our approach to investigate the post-transcriptional cellular response and RBP binding profiles induced by estrogen. Early analysis correlating novel estrogen responsive transcripts and putative PTB sites on the chromosome suggests that PTB may be helping to repress certain exons until they are induced by estrogen. Continuing the early access program with Affymetrix will allow us to be among the first to use Affymetrix genome-wide tiled arrays for these ChIP-chip experiments and to further define the effects of estrogen on a genome-wide scale.

## KEY RESEARCH ACCOMPLISHMENTS:

Successful development of chromatin IP methods to analyze RNA binding proteins.
Identification of new targets such as MDM2 and p53 for PTB and U2AF65.
Determination of a potential new regulatory role of RNA binding proteins by binding upstream of transcription start sites in a RNA polymerase II independent manner.
Observation of a large number of alternative splicing locations.
Prediction of new alternative exons.
Early indications that PTB may be involved in regulating the expression of specific exons in response to estrogen.

## REPORTABLE OUTCOMES:

Brodsky, A. S., Swinburne, I., Keenan, B., and Silver, P. A. Splicing Factors Localize to both Promoter and Alternative Exon Regions on Chromosomes. *In preparation.*

## CONCLUSIONS:

Post-transcriptional regulation is crucial for proper growth regulation as evidenced by alternative mRNA splicing and 3'end formation, regulated export of RNAs out of the nucleus, repression at the level of translation and RNA degradation (Lei and Silver, 2002). It is clear that all of these processes impact on disease yet remain poorly

understood. Our results begin to allow for a broader look at how these processes are altered in breast cancer. In particular, we have developed powerful new technologies to assess the behavior of key RNA binding proteins in both normal and cancer cells. We are identifying potential new functions of these RNA binding proteins and how transcription is organized. Furthermore, we are exploring how the transcriptome and binding profiles may change in response to estrogen. Defining the binding profiles of these RNA binding proteins and how they change in response to cancer may lead to novel therapeutic strategies.

**REFERENCES:**

Brodsky, A. and Silver, PA. A microbead-based system for identifying and characterizing RNA-protein interactions by flow cytometry. 2002, Mol Cell Proteomics, 12:922-9

Brodsky, AS, Johnston, APR, Trau, M. and Silver, PA. Analysis of RNA-protein interactions by flow cytometry. Curr Op in Molecular Therapeutics. 2003, 235-240.

Lei, EP, Krebber, H. and Silver, PA Messenger RNAs are recruited for nuclear export during transcription. 2001. Genes & Dev. 15, 1771-82.

Lei, E. and Silver, PA. Protein and RNA export out of the nucleus. Developmental Cell, 2002, 2:1-20.

Lei, E. and Silver, PA. Intron status and 3' end formation control cotranscriptional export ofmRNA. Genes & Dev., 2002. 16:2761-2766.

**APPENDICES:**

Brodsky, AS, Johnston, APR, Trau, M. and Silver, PA. Analysis of RNA-protein interactions by flow cytometry. Curr Op in Molecular Therapeutics. 2003, 235-240.

# Analysis of RNA-protein interactions by flow cytometry

Alexander S Brodsky[1]*, Angus PR Johnston[2], Matt Trau[2] & Pamela A Silver[1]

**Addresses**
[1]Dana-Farber Cancer Institute and Department of Biological Chemistry and Molecular Pharmacology
Harvard Medical School
1 Jimmy Fund Way
SM 922
Boston
MA 02115
USA
Email: alex_brodsky@dfci.harvard.edu

[2]Centre for Nanotechnology and Biomaterials
Department of Chemistry
University of Queensland
St Lucia 4072
Brisbane
QLD
Australia

*To whom correspondence should be addressed

**Current Opinion in Molecular Therapeutics** 2003 5(3):235-240
 ISSN 1464-8431

*Flow cytometry, in combination with advances in bead coding technologies, is maturing as a powerful high-throughput approach for analyzing molecular interactions. Applications of this technology include antibody assays and single nucleotide polymorphism mapping. This review describes the recent development of a microbead flow cytometric approach to analyze RNA-protein interactions and discusses emerging bead coding strategies that together will allow genome-wide identification of RNA-protein complexes. The microbead flow cytometric approach is flexible and provides new opportunities for functional genomic studies and small-molecule screening.*

***Keywords*** Flow cytometry, functional genomics, microbeads, microspheres, RNA-protein interactions

## Introduction

The determination of RNA-protein regulatory networks is critical for understanding biological pathways. The role of RNA and RNA-protein interactions in regulating gene expression is becoming more appreciated with each new discovery. RNA-protein interactions are the backbone of many post-transcriptional processes, including mRNA stability, splicing, translation and localization. Determining which RNAs and proteins interact remains a challenging goal in the post-genomics era. Many human diseases such as fragile X [1] and HIV [2] are controlled by proteins interacting with RNAs. Proteins also form complexes with both large and small (eg, 7SK) non-coding RNAs [3,4] and microRNAs [5] to regulate gene expression. Understanding how RNA-protein interactions shape gene expression pathways on genome-wide levels remains unclear.

This review highlights recent advances in technologies to study RNA-protein interactions using genomic and high-throughput methods. In particular, we will focus on the use of microbeads to explore RNA-protein complexes by flow cytometry. These methods could evolve into diagnostic assays and high-throughput screens of pharmacological agents targeted to RNA-protein interactions. In this review, the assay will be introduced, and aspects of microbead technology important for the assay, such as microbead multiplexing and surface chemistry, will be discussed.

## RNA-protein screening approaches

Many assays have been developed to examine nucleic acid protein interactions *in vitro*, including gel mobility shift, footprinting and filter binding. Hazbun and Fields performed a large-scale electrophoretic gel mobility shift assay (EMSA) to monitor DNA binding proteins from pools of glutathione-S-transferase (GST) yeast protein libraries [6]. However, similar to the other biochemical strategies, EMSA requires many manipulations making genome-wide screening labor intensive and complicated. In addition, these approaches require labeling of the RNA to monitor binding, making it difficult to pick a particular protein and determine the specifically binding RNAs.

A number of genetic methods have been developed for the analysis of RNA-protein interactions. One system that can screen for either RNA binding proteins or for RNA sequences is the three-hybrid assay [7]. However, long RNA sequences cannot be analyzed and certain sequences cause transcription termination [7]. A second genetic strategy is the Translational Repression Assay Procedure (TRAP) in yeast. This strategy works well with hairpin-containing RNA binding sites but has yet to be tested with a variety of RNA structures [8]. More recently, phage display methods have been developed with a model system to clone candidate proteins binding to a specific RNA sequence [9]. Genetic methods in mammalian cell lines, such as the Tat-fusion transcriptional activation system [10] and frameshifting assay [11], offer the ability to screen in the presence of potential binding partners. One drawback of these methods is that the complexes are forced to form in particular cellular compartments that may not be the native location. In addition, they often depend upon the generation of cDNA libraries that may be biased towards the most abundant messages and would also miss non-coding RNAs such as microRNAs.

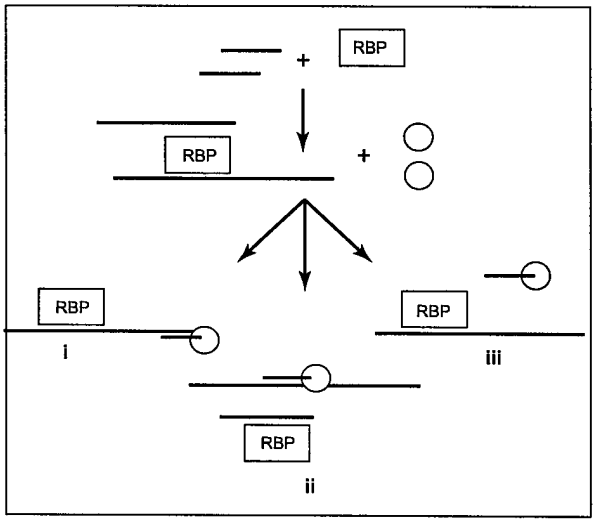
Recently, DNA chips have been used to identify RNAs bound to proteins [1,12,13•,14••]. This approach is promising for the investigation of RNA-protein interactions on a genome-wide scale. Typically, RNA-protein complexes are immunoprecipitated and the RNA is isolated and analyzed on DNA chips. Alternatively, protein can be prepared on beads and cell extract can be bound to the bead [15]. However, these approaches rely on the ability to preserve stable interactions during immunoprecipitation; many potentially weak interactions may be lost. In addition, RNA binding proteins typically have high non-specific binding constants leading to the isolation of a mixture of specific and non-specific 'bound' species, complicating the analysis. Other experiments such as systematic evolution of ligands by exponential amplification (SELEX) may be necessary to help determine the specifically binding RNAs [16].

## The microbead assay

A new approach to RNA biochemistry uses flow cytometry and oligonucleotides attached to microbeads (Figure 1)

[17••]. A fluorescently labeled protein is bound to fluorescently labeled RNA before being challenged with oligonucleotides attached to microbeads. After hybridization, under conditions that do not disrupt the RNA-protein interaction, the microbeads are sorted and analyzed by flow cytometry. The presence or absence of RNA and protein signals provides binding interaction information. RNA-protein interactions can be specifically identified from complex mixtures while simultaneously characterizing binding properties, such as the dissociation constant ($K_d$). In addition, by probing with a high-density oligonucleotide library against RNAs of interest, the binding site could be determined.

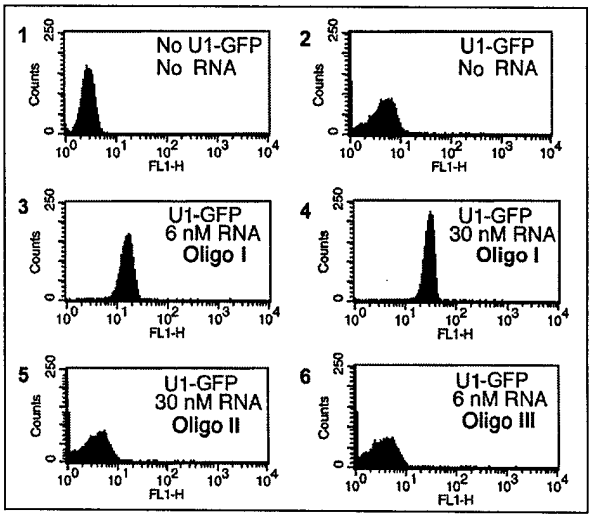
**Figure 1. Schematic of the microbead assay.**




A fluorescently labeled RNA binding protein (RBP)-RNA complex is formed and subsequently challenged with oligonucleotide beads. After reaching equilibrium, RNA and protein fluorescence on each microbead is determined by flow cytometry. The experiment can be performed with or without fluorescently labeled RNA. Three scenarios are possible: (i) RNA and protein fluorescence signal is observed indicating the bead is coupled to an oligonucleotide complementary to a RNA molecule that is binding the RBP; (ii) no protein fluorescence signal is observed but the oligonucleotide is hybridizing to the RNA. With labeled RNA, the RNA-oligonucleotide hybridization is detected. These oligonucleotides may be complementary to the RBP binding site and compete for RBP binding; and (iii) beads with neither protein nor RNA fluorescence suggesting that these oligonucleotides do not hybridize to the RNA. These sequences may be non-complementary to the RNA.

The RNA-protein microbead assay was developed with the U1 snRNP model system. U1-green fluorescent protein (GFP) was purified and bound to a 150mer RNA; binding is indicated by GFP fluorescence on the bead population. (representative flow cytometry data are shown in Figure 2). RNA mutations, oligonucleotide mismatches and dissociation constants were measured to demonstrate the specificity of the assay. Single mismatch discrimination of short oligonucleotides was possible when the signal was monitored through the protein binding. Importantly, RNAs could be specifically detected in total RNA isolated from cells. The sensitivity is in the range of other common flow cytometry assays, since picomolar RNA concentrations could be detected. In this format, the assay is accessible to most molecular biology laboratories as it uses common reagents, and many facilities have access to flow cytometers.

**Figure 2. Flow cytometry data showing GFP fluorescence on microbeads.**




U1-GFP is bound to a 150mer RNA which includes a stem-loop binding site. Histograms show the number of beads at different GFP fluorescence intensities. In the presence of RNA and U1-GFP, the fluorescence intensity of the bead population increases and a more homogeneous bead population is observed, as shown by comparison of panels **2**, **3** and **4**. The different expected outcomes, as outlined in Figure 1, are shown: (i) oligonucleotide I is complementary to the RNA distant from the binding site; (ii) oligonucleotide II is complementary to the loop of the stem-loop and competes with U1-GFP binding; and (iii) oligonucleotide III is a non-complementary oligonucleotide. Oligonucleotides II and III show non-specific binding similar to the background, as shown by comparison of panels **2**, **5** and **6**.

The microbead assay is an equilibrium binding assay that offers some distinct advantages for the biochemical characterization of RNA-protein complexes. Firstly, protein binding to large RNAs can be examined. In fact, larger RNAs offer more hybridization targets for the antisense oligonucleotide probes. Using oligonucleotides targeting different regions of the RNA, binding can be monitored across the whole RNA molecule. Also, binding reactions could be performed in the presence of potential co-operative binding partners by using cell lysates or partially purified cell fractions. Since binding can be monitored at different locations across the RNA molecule, similar to footprinting assays, specific and non-specific sites may be differentiated. This may allow non-specific binding sites to be differentiated in genomic screens as the assay can monitor interactions from the picomolar to nearly micromolar dissociation constant range. Thus, weak interactions can be monitored and potentially discriminated from non-specific interactions.

The assay requires fluorescent labeling of the RNA and/or protein. A number of strategies have been, and are continuing to be developed to label proteins with minimal disruption to their structure and function. The first generation of the RNA-protein bead assay uses a GFP fusion. Other strategies include the use of antibodies, where a fluorescently labeled antibody against a protein is used to

monitor binding [18]. Screening in yeast has found that many proteins are functional when either their N- or C-terminal are tagged [19]. Advances in *in vitro* translation may also allow site-specific labeling at the beginning or end of the protein as seen in both Promega and Roche systems. These strategies will increase the probability of obtaining functionally active, fluorescently labeled protein.

## Flow cytometry

Flow cytometry is a powerful, sensitive and quantitative technology used to measure molecular interactions. Flow cytometry has been successfully applied to examine various protein-protein, protein-DNA and DNA-DNA interactions [20,21]. As it is fluorescence based, it can also be adapted to monitor real-time kinetics and rapid quench studies. Very high sensitivity can be obtained with $10^5$ to $10^6$ particles/ml and target concentrations in the 10 to 100 pM range, well below the $K_d$ of most RNA-protein interactions. Since flow cytometry can focus on just the signal on the microbeads and not the unbound molecules in solution, typically, no washing is required, saving significant effort. Also, recent advances in coding microbeads are bringing the power of multiplexing dozens of samples simultaneously to these assays. With the ability to use automatic sample loaders running at two to three samples/min, high-throughput plate reading is now feasible. Recent reviews highlight the latest technical advances in flow cytometry, allowing for high throughput and sensitivity [20-22].

## Microbeads versus microarrays and hybridization

Binding to microbeads instead of microarrays offers a number of potentially significant advantages. Microbeads have proven to be useful for sensitive and rapid bioanalytical assays. Companies such as Luminex Corp, Lynx Therapeutics and Quantum Dot Corp have taken advantage of these properties to devise high-throughput approaches to immunoassays [23], sequencing [24,25] and single nucleotide polymorphism mapping [26,27••]. Importantly, microbead assays are typically cost-effective, fast and require minimal sample quantities.

An important advantage of microbead for the RNA-protein binding assay is the ability to perform binding on a surface that more closely resembles solution conditions. Hybridization on large planar surfaces is limited by mass transport. On the other hand, microbeads offer better diffusion characteristics, leading to significantly improved hybridization kinetics and thermodynamics [28•,29•].

The basic approach of the microbeads assay is also applicable to microarrays. However, non-specific hybridization at physiological conditions is a requirement of the assay. Due to the demanding hybridization requirements and the relative ease in synthesizing oligonucleotides with long linkers to readily available microbeads, the microbead approach offers a simple alternative to microarrays.

## Encoding strategies

### *Bead libraries*

Most high-throughput bead-based libraries use the optical properties of the support as the library code. The exception to this is the approach from Lynx Therapeutics, who utilize non-encoded support beads, and a series of molecular markers and identifiers [24]. Optical encoding of supports falls into two broad categories. The first (Luminex Corp, Quantum-dot Corp, Illumina and Nanoplex) is based on separately coding each bead and separately synthesizing the target DNA sequence (or other analyte such as RNA or peptide), then attaching each target to a coded bead. The alternative technique is to directly synthesize the target molecule on a coded bead in a combinatorial manner and track every synthetic step each individual bead experiences [30••].

### *Separate encoding*

Methods that use the separate encoding strategy employ a similar strategy to encode the beads. In each case, fluorochromes, fluorescent dyes (Luminex [31] and Illumina [32]) or fluorescent nanocrystals (Quantum-dot Corp [27••,33]) are incorporated into polystyrene beads by swelling the polystyrene in a solvent and absorbing dyes or nanocrystals into the particles. The bead is then placed in a different solvent to shrink the bead, trapping the fluorochrome in the bead. The code is formed by varying the concentration and the combination of fluorochromes present in each bead. The code can be read either by a flow cytometer (Luminex Corp, Quantum-dot Corp) or by optic fiber array (Illumina). DNA sequences are synthesized remotely (either separately in an automated DNA synthesizer or *in vivo*) and attached to the beads using standard ethylenediamine carbodiimide (EDC) coupling chemistry [34]. The separate encoding techniques are useful for small libraries since it is easy to separately synthesize hundreds of different beads and hundreds of target molecules (Table 1). However, there are limitations for larger libraries [35]. To synthesize a library of 100,000 compounds requires 100,000 separate coded beads and 100,000 separately synthesized DNA sequences combined in 100,000 coupling reactions (Figure 3). Automation of this process is possible; however, the size of the library is still limited by the number of coded beads that can be formed.

Nanoplex uses metallic rods (instead of spherical particles) with bands of material with different refractive index to form coding system, which is similar to traditional barcodes but on a microscopic scale [36]. The difference in refractive index is achieved by incorporating different metals into the rods as they are synthesized. Similar to the other separate encoding strategies, library size is limited by the number of separate reaction vessels required to synthesize the coded support and the analyte. However, unlike the fluorescent coding approach, the barcode can be incorporated over a large number of steps, so the coding system can code for many more sequences than it would be possible to synthesize in the library. At this time, there is no automated, high-throughput method of reading these barcodes.

### *Combinatorial encoding/synthesizing*

In the combinatorial method, a set of optically diverse, but distinguishable particles are used as the support for synthesizing the target DNA (or other target molecules) [37]. The optically diverse set of particles are synthesized using a combinatorial process where beads are split into a number

**Table 1. Comparison of microbead coding strategies.**

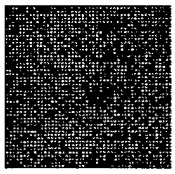
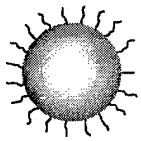
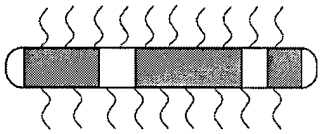
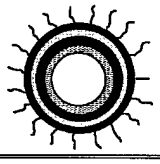
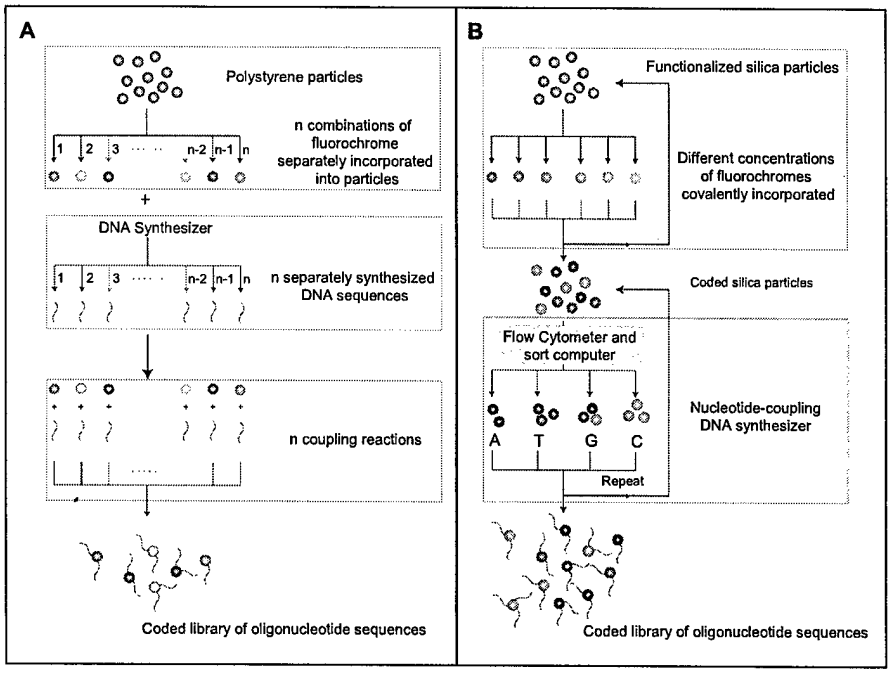
| | Encoding method | Decoding method | Library size |
|---|---|---|---|
|  DNA microarray | Positional encoding. Probes are immobilized in spatially resolved sites on a two-dimensional support. | Via position in array. | < $10^6$ probes [42]. |
|  Probe attached to bead | Non-permanently stained polymer beads with up to four fluorochromes. | Flow cytometry [31,33], optical fiber arrays [30••], digital imaging. | 100 to 270,000 probes. |
|  Probe attached to rod | Layered metallic strips on rod shaped particles. | Microscopic imaging (not automated). | > 100 probes [36] (can potentially code $10^{13}$ but library size is limited by decoding rate and library synthesis). |
|  Probe synthesized on bead | The unique optical signature of each multi-fluorescent support bead is tracked by a flow cytometer during the combinatorial synthesis of the probe. | The optical signature is analyzed by flow cytometry and the reaction history of the bead is determined by recalling data stored by the flow cytometer software during probe synthesis. | >$10^8$ probes [30••]. |

**Figure 3. Comparison of encoding techniques.**




(**A**) Separate encoding strategy where beads are individually coded in separate reaction vessels and the oligonucleotides are individually synthesized remotely. The oligonucleotide is coupled to the bead using standard EDC chemistry.
(**B**) Combinatorial encoding strategy where silica particles are coded using a split and mix process with varying concentrations of dyes. Using a customized flow cytometer, the particles are sorted into four reaction vessels (one for each base) according to predetermined parameters. The process is repeated until oligonucleotides of the required length are synthesized.

of reaction vessels and varying concentrations of fluorophores such as organic fluorescent dyes or nanoparticles, are covalently incorporated into the beads. The beads are then mixed together and the process is repeated for each subsequent dye; thus, it is not necessary to synthesize each coded bead individually [30••]. Using this method with six fluorophores and eight levels of intensity for each dye, a library of over 250,000 signatures can be constructed. Still using only six fluorophores, but with 16 levels of intensity, a library of over 16 million sequences can be generated (Table 1).

Using a flow cytometer and custom designed electronics, beads can be analyzed and sorted according to the particular optical signatures [30••]. Each bead has a predetermined sequence that is uploaded to the modified flow cytometer and sort decisions are made according to the sequences that are required for the particular library. The flow cytometer can sort into four directions, with each direction corresponding to a different nucleoside. After each 'sort' the nucleosides are coupled to the corresponding beads and once coupling is complete, the beads are mixed together and the process is repeated until the oligonucleotide sequences of the required length are synthesized. At the end of the process, beads with a known optical signature are synthesized with each unique signature corresponding to a different oligonucleotide sequence (Figure 3).

Synthesizing libraries in this way requires beads that can withstand the relatively harsh conditions of DNA synthesis. Polystyrene beads are typically not suitable for this process because they swell and leach dye during the synthesis procedure, thus catastrophically altering the optical signature. Therefore, specially synthesized silica particles are required [30••].

## Surface chemistry

The surface chemistry of the beads plays an important role in the assay. Non-specific binding of proteins to the beads is a larger problem than the non-specific binding of oligonucleotides to the beads, as generally, most surfaces with a large negative charge (eg, silica and polystyrene surfaces) have relatively low non-specific binding of oligonucleotides (due to repulsion of the negatively charged phosphate backbone). However, as proteins have positively and negatively charged regions, it is necessary to have a surface which has little or no surface charge to minimize non-specific electrostatic binding and molecule adsorption [38•]. Coating the surfaces with hydrophobic chains (such as alkyl chains) is also not ideal, since many proteins have hydrophobic regions that will also non-specifically bind to the beads [39]. One solution is to add a large excess of a protein (eg, inexpensive and abundant proteins such as bovine serum albumin) that non-specifically bind to the surface of the beads, limiting the non-specific binding of the fluorescently labeled protein. However, there is a limit to the effectiveness of this procedure and it is desirable to have a 'biologically silent' surface that limits the non-specific binding of the proteins.

Much of the surface chemistry developed for protein-chips can be applied to bead surfaces. Polyethylene glycol and oligoethylene glycol surfaces have been used to minimize the non-specific binding of proteins [39] to silica substrates.

Surface density of the probes also plays an important role in the assay. Clearly, the higher the number of probes on the beads, the higher the resultant signal; however, overloading the surface introduces problems. It is possible to load in excess of 100 million target sequences on a single bead, but at this very high surface density, steric hindrance can affect the hybridization of target DNA to the beads. In addition, false hybridization events may occur, where one target DNA strand hybridizes to multiple probe strands on the bead [40]. Similar findings have been observed on arrays [41].

## High-throughput screening and genomics

This review describes the recent development of a versatile flow cytometry approach to examine RNA-protein interactions. The emerging bead-coding and surface chemistry technologies, in combination with novel assays such as the microbead RNA-protein assay will lead to new small molecule and genomic screens. Due to the versatility and flexibility of flow cytometry and the RNA-protein assay, many variations are possible, including defining the binding spectrum of a particular RNA-binding protein, screening a protein library for binding to a specific RNA, or discovering small molecules that inhibit an RNA-protein interaction. With the increasing understanding of the importance of RNA-protein interactions in human disease and development, the contribution of these promising technologies is expected to be significant.

## Acknowledgements

Alexander S Brodsky is supported by the Claudia Adams Barr Young Investigator program and an NHGRI Genome Scholar award.

## References

•• of outstanding interest
• of special interest

1. Brown V, Jin P, Ceman S, Darnell JC, O'Donnell WT, Tenenbaum SA, Jin X, Feng Y, Wilkinson KD, Keene JD, Darnell RB *et al*: **Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome.** *Cell* (2001) **107**(4):477-487.

2. Frankel AD, Young JA: **HIV-1: Fifteen proteins and an RNA.** *Annu Rev Biochem* (1998) **67**:1-25.

3. Nguyen VT, Kiss T, Michels AA, Bensaude O: **7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes.** *Nature* (2001) **414**(6861):322-325.

4. Yang Z, Zhu Q, Luo K, Zhou Q: **The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription.** *Nature* (2001) **414**(6861):317-322.

5. Dostie J, Mourelatos Z, Yang M, Sharma A, Dreyfuss G: **Numerous microRNPs in neuronal cells containing novel microRNAs.** *RNA* (2003) **9**(2):180-186.

6. Hazbun TR, Fields S: **A genome-wide screen for site-specific DNA-binding proteins.** *Mol Cell Proteomics* (2002) **1**(7):538-543.

7. Zhang B, Kraemer B, SenGupta D, Fields S, Wickens M: **Yeast three-hybrid system to detect and analyze RNA-protein interactions.** *Methods Enzymol* (2000) **318**:399-419.

8. Paraskeva E, Hentze MW: **Translational repression assay procedure: A method to study RNA-protein interactions in yeast.** *Methods Enzymol* (2000) **318**:374-384.

9. Danner S, Belasco JG: **T7 phage display: A novel genetic selection system for cloning RNA-binding proteins from cDNA libraries.** *Proc Natl Acad Sci USA* (2001) **98**(23):12954-12959.

10. Landt SG, Tan R, Frankel AD: **Screening RNA-binding libraries using Tat-fusion system in mammalian cells.** *Methods Enzymol* (2000) **318**:350-363.

11. Kollmus H, Hauser H: **Frameshifting assay to characterize RNA-protein interactions in eukaryotic cells.** *Methods Enzymol* (2000) **318**:363-374.

12. Andrews LG, Keene JD: **Identification of specific protein-RNA target sites using libraries of natural sequences.** *Methods Mol Biol* (1999) **118**:233-244.

13. Tenenbaum SA, Carson CC, Lager PJ, Keene JD: **Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays.** *Proc Natl Acad Sci USA* (2000) **97**(26):14085-14090.
• *The first report of microarrays to analyze co-immunoprecipitating mRNAs.*

14. Hieronymus H, Silver PA: **Genome-wide analysis of RNA-protein interactions illustrates specificity of the mRNA export machinery.** *Nat Genet* (2003) **33**(2):155-161.
•• *A thorough microarray analysis of co-immunoprecipitating bound mRNAs.*

15. Trifillis P, Day N, Kiledjian M: **Finding the right RNA: Identification of cellular mRNA substrates for RNA-binding proteins.** *RNA* (1999) **5**(8):1071-1082.

16. Darnell JC, Jensen KB, Jin P, Brown V, Warren ST, Darnell RB: **Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function.** *Cell* (2001) **107**(4):489-499.

17. Brodsky AS, Silver PA: **A microbead-based system for identifying and characterizing RNA-protein interactions by flow cytometry.** *Mol Cell Proteomics* (2002) **1**(12):922-929.
•• *Description of the novel microbead assay.*

18. Yang X, Li X, Prow TW, Reece LM, Bassett SE, Luxon BA, Herzog NK, Aronson J, Shope RE, Leary JF, Gorenstein DG: **Immunofluorescence assay and flow-cytometry selection of bead-bound aptamers.** *Nucleic Acids Res* (2003) **31**(10):e54.

19. Phizicky E, Bastiaens PI, Zhu H, Snyder M, Fields S: **Protein analysis on a proteomic scale.** *Nature* (2003) **422**(6928):208-215.

20. Iannone MA: **Microsphere-based molecular cytometry.** *Clin Lab Med* (2001) **21**(4):731-742.

21. Sklar LA, Edwards BS, Graves SW, Nolan JP, Prossnitz ER: **Flow cytometric analysis of ligand-receptor interactions and molecular assemblies.** *Annu Rev Biophys Biomol Struct* (2002) **31**:97-119.

22. Nolan JP, Sklar LA: **The emergence of flow cytometry for sensitive, real-time measurements of molecular interactions.** *Nat Biotechnol* (1998) **16**(7):633-638.

23. Vignali DA: **Multiplexed particle-based flow cytometric assays.** *J Immunol Methods* (2000) **243**(1-2):243-255.

24. Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R *et al*: **Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays.** *Nat Biotechnol* (2000) **18**(6):630-634.

25. Brenner S, Williams SR, Vermaas EH, Storck T, Moon K, McCollum C, Mao JI, Luo S, Kirchner JJ, Eletr S, DuBridge RB *et al*: ***In vitro* cloning of complex mixtures of DNA on microbeads: Physical separation of differentially expressed cDNAs.** *Proc Natl Acad Sci USA* (2000) **97**(4):1665-1670.

26. Taylor JD, Briley D, Nguyen Q, Long K, Iannone MA, Li MS, Ye F, Afshari A, Lai E, Wagner M, Chen J *et al*: **Flow cytometric platform for high-throughput single nucleotide polymorphism analysis.** *Biotechniques* (2001) **30**(3):661-666.

27. Xu H, Sha MY, Wong EY, Uphoff J, Xu Y, Treadway JA, Truong A, O'Brien E, Asquith S, Stubbins M, Spurr NK *et al*: **Multiplexed SNP genotyping using the Qbead system: A quantum dot-encoded microsphere-based assay.** *Nucleic Acids Res* (2003) **31**(8):e43.
•• *Description of the use and advantages of a quantum dot-based optical code.*

28. Henry MR, Wilkins Stevens P, Sun J, Kelso DM: **Real-time measurements of DNA hybridization on microparticles with fluorescence resonance energy transfer.** *Anal Biochem* (1999) **276**(2):204-214.
• *Describes experiments comparing microbead and solution phase hybridization kinetics.*

29. Wilkins Stevens P, Henry MR, Kelso DM: **DNA hybridization on microparticles: Determining capture-probe density and equilibrium dissociation constants.** *Nucleic Acids Res* (1999) **27**(7):1719-1727.
• *Describes experiments comparing microbead and solution phase hybridization kinetics.*

30. Battersby BJ, Lawrie GA, Johnston AP, Trau M: **Optical barcoding of colloidal suspensions: Applications in genomics, proteomics and drug discovery.** *Chem Commun (Camb)* (2002) **14**:1435-1441.
•• *Description of the combinatorial coding technique.*

31. Fulton RJ, McDade RL, Smith PL, Kienker LJ, Kettman JR Jr: **Advanced multiplexed analysis with the FlowMetrix system.** *Clin Chem* (1997) **43**(9):1749-1756.

32. Walt DR: **Techview: Molecular biology. Bead-based fiber-optic arrays.** *Science* (2000) **287**(5452):451-452.

33. Han M, Gao X, Su JZ, Nie S: **Quantum-dot-tagged microbeads for multiplexed optical coding of biomolecules.** *Nat Biotechnol* (2001) **19**(7):631-635.

34. Walsh MK, Wang X, Weimer BC: **Optimizing the immobilization of single-stranded DNA onto glass beads.** *J Biochem Biophys Methods* (2001) **47**(3):221-231.

35. Battersby BJ, Grondal L, Lawrie GA, Trau M: **Colloids for encoding chemical libraries: Applications in biological screening.** In: *Colloids and Colloidal Assemblies.* Caruso F (Ed), Wiley, New York, NY, USA (2002).

36. Nicewarner-Pena SR, Freeman RG, Reiss BD, He L, Pena DJ, Walton ID, Cromer R, Keating CD, Natan MJ: **Submicrometer metallic barcodes.** *Science* (2001) **294**(5540):137-141.

37. Trau M, Battersby BJ: **Novel colloidal materials for high-throughput screening applications in drug discovery and genomics.** *Advanced Materials* (2001) **13**(12-13):975-979.

38. Song J, Choi H, Moeser G, Laibinis P: **Preparation of protein- and cell-resistant surfaces on $SiO_2$ surfaces by polyther-grafted poly(acrylic acid) thin films.** *Polymer Preprints* (2003) **44**:189-190.
• *Describes the advantages of using ethylene glycol surfaces to minimize non-specific binding.*

39. Lee SW, Laibinis PE: **Protein-resistant coatings for glass and metal oxide surfaces derived from oligo(ethylene glycol)-terminated alkytrichlorosilanes.** *Biomaterials* (1998) **19**(18):1669-1675.

40. Guo Z, Gatterman MS, Hood L, Hansen JA, Petersdorf EW: **Oligonucleotide arrays for high-throughput SNPs detection in the MHC class I genes: HLA-B as a model system.** *Genome Res* (2002) **12**(3):447-457.

41. Peterson AW, Heaton RJ, Georgiadis RM: **The effect of surface probe density on DNA hybridization.** *Nucleic Acids Res* (2001) **29**(24):5163-5168.

42. Hacia JG: **Resequencing and mutational analysis using oligonucleotide microarrays.** *Nat Genet* (1999) **21**(1 Suppl):42-47.