**In vitro** cloning of complex mixtures of DNA on microbeads: Physical separation of differentially expressed cDNAs


* Lynx Therapeutics, Inc., 25861 Industrial Boulevard, Hayward, CA 94545; and †BASF-LYNX Bioscience, Im Neuenheimer Feld 515, Heidelberg 69120, Germany

Contributed by Sydney Brenner, December 14, 1999

We describe a method for cloning nucleic acid molecules onto the surfaces of 5-μm microbeads rather than in biological hosts. A unique tag sequence is attached to each molecule, and the tagged library is amplified. Unique tagging of the molecules is achieved by sampling a small fraction (1%) of a very large repertoire of tag sequences. The resulting library is hybridized to microbeads that each carry ~10^5 strands complementary to one of the tags. About 10^3 copies of each molecule are collected on each microbead. Because such clones are segregated on microbeads, they can be operated on simultaneously and then assayed separately. To demonstrate the utility of this approach, we show how to label and extract microbeads bearing clones differentially expressed between two libraries by using a fluorescence-activated cell sorter (FACS). Because no prior information about the cloned molecules is required, this process is obviously useful where sequence databases are incomplete or nonexistent. More importantly, the process also permits the isolation of clones that are expressed only in given tissues or that are differentially expressed between normal and diseased states. Such clones may then be spotted on much more cost-effective, tissue- or disease-directed, low-density planar microarrays.

DNA analysis | gene expression | parallel cloning | fluid microarray

Analysis of complex genomes requires methods for generating and fractionating many tens of thousands of DNA fragments in quantities and formats amenable to biochemical analysis. The difficulty of such analysis is illustrated by human gene expression: the human genome is estimated to contain about 100,000 genes, of which 10–30%, or about 20–40 Mb, are actively expressed in any given tissue (1–2). Such large numbers of expressed genes make it difficult to track changes in expression patterns, particularly in view of the large fraction of genes that are expressed at very low levels: it has been estimated that as much as 30% of mammalian messenger RNA consists of many thousands of distinct species each making up far less than a few tenths of a percent of the total, and typically averaging less than a few tens of copies per cell (3–4). Yet relatively minor alterations in gene expression patterns are associated with profound changes in cell physiology and, more broadly, in the state of an organism’s health, its longevity, and its survival (5–7). A variety of techniques has been developed for analyzing gene expression that differ widely in convenience, expense, and sensitivity. Presently, techniques based on direct sequence analysis or specific hybridization of complex polynucleotide probes to microarrays of oligonucleotides or polynucleotides provide the most comprehensive and sensitive analysis of gene expression (8–9). However, in both approaches, the sequences to be analyzed must either be known or cloned and processed individually beforehand, usually with the aid of complex robotics systems. This makes it difficult to isolate and/or monitor many potentially important genes that are differentially expressed at low absolute levels against a background of more abundantly expressed genes.

To address some of these problems, we describe an approach that greatly simplifies the handling and analysis of complex mixtures of cDNA or genomic fragments. We show how millions of nucleic acid molecules, amplified with one set of common primers, can be cloned and specifically attached to 5-μm microbeads in a few single-tube reactions. Central to the method is the formation of a repertoire of oligonucleotide tags assembled combinatorially from a defined set of subunits, or “words,” and their attachment to individual polynucleotides of a complex mixture. By making the repertoire of tags large relative to the number of polynucleotides in the mixture, samples of the tag-polynucleotide conjugates may be selected with virtually every polynucleotide having a unique tag. Samples of conjugates then may be amplified and specifically hybridized to their complementary sequences (anti-tags) on separate microbeads in a single reaction to form a library of microbeads, each having attached a clonal population of one polynucleotide from the original mixture.

The key advantage of this approach over biological cloning is that the DNA on the surface of each microbead is readily accessible for biochemical analysis without further processing. In addition, all clones in a microbead library can be interrogated simultaneously by analytical probes designed to assay specific properties. This enables a broad range of applications in which clones are identified or physically fractionated by virtue of their sequence or their abundance or their ability to bind particular ligands. Below, we show how such “microbead clones” can be used to extract differentially regulated genes, independently of any sequence information, in the human acute monocytic leukemia cell line, THP-1, induced by phorbol 12-myristate 13-acetate (PMA) and lipopolysaccharide (LPS). After competitive hybridization of cDNA probes from the induced and noninduced cells, microbeads sorted on the basis of fluorescence intensity ratios were found to carry sequences known to be up-regulated and down-regulated in THP-1 cells after induction by PMA and LPS. Further validation is provided by plaque hybridization with probes constructed from several genes fractionated by our microbead procedure.

Materials and Methods

Construction of Oligonucleotide Tag and Anti-Tag Libraries. A library of 32-mer tags was synthesized based on eight 4-mer “words”

Abbreviations: PMA, phorbol 12-myristate 13-acetate; LPS, lipopolysaccharide; FACS, fluorescence-activated cell sorting; FAM, 6-carboxyl-fluorescein.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AW059512–AW059953).

* To whom reprint requests should be addressed. E-mail: galbrecht@lynxgen.com.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.
lacking dG described in Results by eight rounds of “mix and divide” combinatorial synthesis (Fig. 1) on 40 billion 5-μm glycidal methacrylate microbeads (Bangs Laboratories, Carmel, IN). The microbeads were prepared with ~10% of the initial nucleotides linked by a base-labile group (5'9-phosphate-ON, CLONTECH). A 28-mer 3'9 spacer having a primer-binding site and a PacI site was synthesized, followed by synthesis of the 32-mer tags and three cytidylate residues. Five billion of these microbeads were removed and by further synthesis a Bsp120I site and a 5'9 primer-binding site were added. Aliquots of the remaining 35 billion microbeads were prepared for capturing tagged cDNAs as follows: 2.5 × 108 microbeads suspended in 100 μl of H2O were combined with 100 μl of 10× NEB buffer 2 (New England Biolabs), 10 μl of 100 mM ATP, 1 μl of 10% Tween 20, 17 μl of T4 polynucleotide kinase (10 units/μl), and 772 of μl H2O for a final volume of 1,000 μl. After incubating for 2 hr at 37°C with vortexing, the temperature was increased to 65°C for 20 min to inactivate the kinase, with continued vortexing. After incubation, the microbeads were washed twice by spinning down the microbeads and resuspending them in 1 ml of Tris-EDTA containing 0.01% Tween 20. Postsynthesis analysis of these showed that each microbead carries about 10 million copies of a given anti-tag sequence (data not shown).

The aliquot of five billion microbeads was treated to cleave DNA attached by base-labile linkers to yield a total of 1 × 1016 sequences. After deprotection and purification by reverse-phase chromatography, sequences were isolated by ethanol precipitation, resuspended, and used as the template for a T7 polymerase fill-in reaction with a primer complementary to the 3'9 primer-binding site. The entire duplex then was PCR-amplified with primers specific for the 3' and 5' primer-binding sites, after which the amplicon was digested with PacI and Bsp120I and inserted into pLCV1 (Fig. 2). The product was electroporated into Escherichia coli to produce more than 200 million independent chloramphenicol-resistant clones, designated pLCV2.

cDNA Synthesis and Attachment of Tags. mRNA was converted to cDNA, tagged, and loaded onto microbeads as described (10). Briefly, poly(A)+ RNA was extracted by using a FastTrack 2.0 kit (Invitrogen), using the manufacturer’s protocol. Double-stranded cDNA was synthesized by using a cDNA Synthesis kit (Stratagene), using the manufacturer’s protocol with the following modifications: 2.5 μg of mRNA was used for each synthesis and the primer for first strand synthesis was 5'9-biotin-GACATGCTCGTCTCTGAC19V. After second-strand synthesis, the cDNA was size-fractionated on a SizeSep 400 column (Amersham Pharmacia) and ethanol-precipitated. The cDNA was resuspended and digested with 100 units of DpnII at 37°C for 2 hr, after which the biotinylated cDNA was purified with Dynal M-280 streptavidin beads. cDNA fragments were digested off the beads by using 40 units of BsmBI for 2 hr in an Eppendorf Thermomixer set at medium speed at 37°C. The supernatant was collected, and the beads were resuspended for a second BsmBI digestion. cDNA fragments from the first and second digestions were pooled, ethanol-precipitated, and resuspended in 10 μl of H2O. Tag vector pLCV2 was used for loading libraries on microbeads. pECV, which lacks tags, was used for probe libraries. The vectors were digested with BbsI and BamHI and dephosphorylated, and aliquots of each cDNA preparation were ligated with each vector. Sequences adjacent to BbsI and BamHI were chosen to ensure compatible ends. DNA was transformed...
into electro-competent *E. coli* TOP10 cells (Invitrogen). Aliquots were titered on LB agar plates containing 30 μg/ml chloramphenicol, and library pools were grown in liquid cultures. For the microbead libraries, six pools of 160,000 clones each were grown in 50-ml liquid cultures. For the probe libraries, 1 × 10⁷ clones were grown in 1-liter cultures. Plasmid DNA was prepared and used for subsequent manipulations.

**Microbead Loading.** Tagged cDNAs (160,000) were amplified in the presence of 5-methyl dCTP using flanking PCR primers (designated biotin-PCR-F and FAM (6-carboxyfluorescein)-PCR-R; labeling reagents from CLONTECH), and the products were digested to completion with *PacI* and affinity-purified. To expose the tags as single strands, amplified DNA was treated with T4 polymerase in the presence of 1 mM dGTP for 60 min, thereby digesting a single strand of the tag to the GC-rich *Bsp*I10I site. The reaction was stopped with EDTA, and the enzyme was denatured by heating (72°C for 15 min). Fifty micrograms of this mixture then was combined with an aliquot of 16.7 million microbeads, each having about 10⁵ copies of a single anti-tag, in a 100-μl reaction containing 500 mM NaCl, 10 mM sodium phosphate, 0.01% Tween 20, and 5% dextran sulfate. The sample was incubated for 3 days at 72°C. The microbeads then were washed twice, first in 50 mM Tris, 50 mM NaCl, 3 mM Mg-Cl, and then in 10 mM Tris-HCl (pH 8), 1 mM EDTA, 0.01% Tween 20, after which the 1% brightest beads were sorted on a Cytomation MoFlo cytometer. cDNAs (10⁴-10⁵ per microbead) were loaded, determined by comparing the amount of test probe hybridized to saturation to 2,000 identical micrograms of this mixture then was combined with an aliquot of 1,000,000 molecules can be converted into a library containing a family’s unique tag. This permits them to be collected on a microbead library of about as many microbeads, each carrying about 10⁶ copies of a unique tag. This permits them to be collected on a microbead library of about as many microbeads, each carrying about 10⁶ copies of a unique tag. This permits them to be collected on a microbead library of about as many microbeads, each carrying about 10⁶ copies of a unique tag.

**PCR, Cloning, and Sequencing of Isolated cDNA.** Sorted microbeads were used directly in a PCR, and the product was cloned by using the TA Cloning procedure (Invitrogen) and sequenced.

**Cell Culture.** THP-1 cells (ATCC accession no. TIB-202) were grown in DMEM/F-12 media supplemented with 10% heat-inactivated FBS (56°C, 30 min), 5 × 10⁻⁴ M 2-mercaptoethanol, 100 units/ml penicillin, and 100 μg/ml streptomycin. Cultures seeded with 1 × 10⁷ cells per ml were grown to a cell density of 1 × 10⁶ cells per ml (approximately 4 days). Cells were induced by adding PMA (stock concentration 1 mM in DMSO) to a final concentration of 100 nM and were grown for 48 hr with PMA. DMSO was added to noninduced cells. During this period, approximately 80% of the PMA-treated cells became adherent, whereas about 20% remained in suspension. After 48 hr, the medium and nonadherent cells were removed from PMA-induced cells and replaced with fresh medium containing 5 μg/ml LPS from *E. coli* serotype 0111:B4 (Sigma). Control cells were pelleted by centrifugation at 1,400 × g for 4 min and resuspended in fresh equilibrated medium, and incubation continued. The cells were harvested after 4 hr. Adherent PMA- and LPS-treated cells were dislodged by using a cell scraper and pelleted by centrifugation at 1,400 × g. Control cells, which remained in suspension, were simply pelleted.

**Results**

**Principle of the Method.** Each nucleic acid molecule in a complex mixture first is labeled with an oligonucleotide tag, after which a sample of the resulting library of tag–molecule conjugates is amplified. After amplification, a sample of *n* molecules will yield a library containing *n* families of clones, each identified by a unique tag. This permits them to be collected on a microbead library of about as many microbeads, each carrying about 100,000 copies of one of the templates. To achieve this, the nucleic acid sequences used as tags and anti-tags must be chosen to maximize the efficiency and discrimination of hybridization of a given clone to a given microbead.

Six criteria were used to design the repertoire of tags: (i) the repertoire must be diverse enough to enable the unique tagging of all (or nearly all) the molecules in large libraries; (ii) the tags must remain physically attached to the tagged molecules after operations such as cleavage by restriction enzymes; (iii) the melting temperatures (*T*m) of all tag/anti-tag duplexes must be isothermal; (iv) the difference in *T*m between any perfectly matched tag/anti-tag duplex and any duplex with a single mismatch must be both the same for all sequences and large enough to discriminate strongly in favor of the perfect match; (v) a practical and effective manner must exist to enable construction of such a repertoire; and (vi) there must be a simple way of applying the method, in parallel, and in one reaction, to any ensemble of nucleic acid molecules. A library of sequences with these properties can be constructed from a special DNA language with a vocabulary made up of eight four-base “words”: TTAC, AATC, TACT, ATCA, ACAT, TCTA, CTAA, and CAAA.

Each word uses only three (A, T, and C) of the DNA bases and differs from all of the other words in three of the four bases, and they all form, with their respective complements, three A:T and three C:G base pairs. Limiting the composition of the four-
nucleotide words to three bases eliminates self-complementarity within any sequence made up of the words and prevents their cleavage by any restriction enzyme with symmetrical recognition sites containing both G and C. The repertoire of tags is constructed by a “mix and divide” combinatorial synthesis of all possible eight-word combinations. There are 88 such combinations, forming a tag repertoire of 16,777,216 sequences 32 bases long. An example of one tag and its anti-tag complement is shown below:

\[
5'-\text{TACT.TTAC.ACAT.CTTT.CTTT.CAAA.AATC}-3' \\
3'-\text{ATGA.AATG.TGTA.TAGT.GAAA.GAAA.GTTT.TTAG}-5'
\]

Because of the words used to construct the tag repertoire, all 16,777,216 eight-word tag/anti-tag duplexes have the same base pair composition, i.e., 24 A:T and eight G:C base pairs. Consequently, all of the tags in the entire repertoire should have about the same \( T_m \). Because all tag sequences differ from their nearest neighbors by at least one word, or three base mismatches, there is good discrimination between perfect matches and one-word mismatches. Fig. 3A depicts the melting curves of eight-word duplexes that have zero-, one-, or two-word mismatches. It shows that at 68°C, a duplex with a single-word mismatch dissociates, whereas the duplex with an exact match does not. Fig. 3B shows that anti-tags immobilized on a solid surface do not lose their ability to discriminate in favor of the correct tags in hybridization experiments. To estimate the accuracy of loading, we deposited individual microbeads into 96-well microtiter plates by using fluorescence-activated cell sorting (FACS) and subjected them to multiple rounds of PCR amplification. Ten percent of the beads produced no detectable product, 70% yielded single bands, and the remaining 20% yielded more than one band (data not shown). Of this latter category, one-fourth were estimated to have contained two microbeads; we estimate, therefore, that at least 85% of the microbeads contain only one sequence.

Allocation of tags to individual molecules is outlined in Fig. 2. First, the complete repertoire of tags in a plasmid library is ligated to the entire population of cDNAs to give a population of tag–cDNA conjugates that contains a conjugate between every tag and every cDNA. Next, a sample of 160,000 tag–cDNA conjugates is taken that includes only 1% of the full repertoire and, thus, the probability that two different cDNAs will have the same tag sequence is very small, about 0.01%. The vast majority will be uniquely tagged. The small sample size also ensures greater specificity in hybridization, because most of the sampled tags are likely to differ from one another by two or more words. The 160,000 tag–cDNA conjugates are amplified by PCR to give fluorescently labeled biotinylated amplicons. After purification with a streptavidinated support, the DNA is treated with T4 DNA polymerase in the presence of dGTP to remove one strand from the tag/anti-tag duplex using the 3′ → 5′ exonuclease and exchange reaction activities of the enzyme. The resulting tag–cDNA conjugates are loaded onto microbeads by mixing them with the full 16.7 million repertoire of microbeads, each carrying a specific anti-tag sequence. The loading reaction is carried out under equilibrium conditions (i.e., 72 hr at 72°C) to ensure that hybridization discrimination is achieved through the higher off-rate of mismatched sequences. After separating loaded microbeads from unloaded microbeads by FACS, as shown in Fig. 4, the hybridized DNA is ligated to the anti-tag, covalently attaching one strand of the DNA to the microbead’s surface. This permits easy removal of the noncovalently attached strand.

**Analysis of Differential Gene Expression.** To illustrate the first of several applications of our cloning methodology, we describe an experiment for physically extracting clones that are differentially represented in two libraries. After melting off the noncovalently attached DNA strands from the microbeads, the remaining strands are available for competitive hybridization with two probes labeled with different fluorophores, to produce, in effect, a fluid microarray of millions of microbeads. The ratio of probes hybridizing to each bead can easily be measured at high throughput (20,000 beads per sec) by using available FACS equipment. Thus, microbeads labeled with more of one fluorophore than the other can be easily identified and physically separated from the rest of the library.

The sensitivity of this approach was tested in an experiment where two fluorescent probes in ratios of 1:8, 8:1, 1:4, 4:1, 1:2,
2:1, 1:1, 0:1, and 1:0 were hybridized to a complementary set of microbeads. The results (Fig. 5) show that the probe ratios can be clearly measured and that differences as low as 2-fold can be detected. The range of fluorescence intensities exhibited by the microbeads reflects the variation in the number of cDNAs loaded on each microbead, which determines the total number of probe molecules that can bind.

To see whether we could enrich for differentially expressed genes by using FACS, we made a microbead library with cDNA pooled from THP-1 before and after treatment with PMA and LPS. Four cDNA probes were constructed: two from the same library of induced cDNAs, which were labeled with either R110 or Cy5, and one each from libraries of induced and noninduced cDNAs, which were labeled, respectively, with Cy5 and R110. The former were used in a control experiment to compare the library with itself. Fig. 6A is a FACS plot of the microbead library hybridized with a 1:1 mixture of differently labeled probes from the same library, whereas Fig. 6B is a plot of the microbead library hybridized with a 1:1 mixture of cDNA probes from the induced and noninduced cells. The distribution of microbeads in Fig. 6A allowed us to set gates for collecting microbeads that were more heavily labeled with either Cy5 or R110 (Fig. 6B). The triangles in Fig. 6B represent the gates from which microbeads carrying up-regulated or down-regulated clones were collected. Each gate was set to collect about 1% of the total number of clones in the library; gate 1 should collect clones that are 10-fold or more up-regulated, whereas gate 2 should collect clones that are 2-fold or more down-regulated.

Of the 1,600,100 microbeads used in this experiment, 13,988 (0.87%) were collected in the up-regulated fraction, and 17,303 (1.01%) in the down-regulated fraction. The DNA in these fractions was recovered by PCR, and 956 of the up-regulated clones and 985 of the down-regulated clones were sequenced. Clones from the up-regulated fraction included 68 known genes, 35 expressed sequence tags, and nine novel sequences, whereas clones from the down-regulated fraction included 209 known genes, 111 expressed sequence tags, and five novel sequences. In the up-regulated fraction there were many known PMA-induced genes such as IL-8, tumor necrosis factor \( \alpha \), macrophage inflammatory protein 1, and superoxide dismutase 2 (11–14). There was very little overlap between the up- and down-regulated sets of genes, except for the presence of a small number of microbeads bearing up-regulated B94 (M92237) cDNAs in the down-regulated fraction and some microbeads bearing down-regulated 23 kDa basic protein transcript (X56932) in the up-regulated fraction. These were likely the result of sorting errors caused by clustering of microbeads in adjacent drops.

**Plaque Hybridizations.** To validate our procedure, we independently cloned cDNA from the induced and noninduced THP-1 cells in bacteriophage lambda and screened filters each containing 50,000 plaques with probes derived from 19 of the genes identified by sequencing the DNA from the fractionated beads. Table 1 summarizes these results. In every case, the sequences
Down-regulated genes
sequences differentially regulated at very low levels. All that is
analyses, such as hybridization or sequencing.

further biochemical processing is required for various types
the DNA on microbeads is accessible for direct interrogation, no

genes that are differentially expressed. Microbead clones can be
fragments on the surfaces of microbeads and have shown how

Discussion

contaminated by genes that were unaffected by PMA stimulation
addition, the pool of microbeads in gate 2 was not significantly

of 7.8 per gene. We estimate the enrichment to be 5
proteins with recurrences ranging from 1 to 42 and an average

differentiation. The first process should be associated with the
cycle followed by the induction of the genes required for

0.02–2%.

THP-1 cells respond to PMA by first coming out of the cell
cycle followed by the induction of the genes required for
differentiation. The first process should be associated with the
down-regulation of a large number of genes involved in the
growth of cells, such as ribosomal protein-coding genes. Plaque
hybridization with several genes in this class indicates that these
are down-regulated between 2- and 4-fold (Table 1). Our method
detects these relatively small changes; we found an enrichment
of these cDNAs in the pool of microbeads collected in gate 2. Of
the 989 clones sequenced, 375 could be assigned to 48 ribosomal
genes in this class indicates that these
are down-regulated between 2- and 4-fold (Table 1). Our method
detects these relatively small changes; we found an enrichment
of these cDNAs in the pool of microbeads collected in gate 2. Of
the 989 clones sequenced, 375 could be assigned to 48 ribosomal

Table 1. THP-1 expression levels by plaque hybridization assays

<table>
<thead>
<tr>
<th>GeneBank Number</th>
<th>Gene</th>
<th>Noninduced, replicate counts</th>
<th>PMA/LPS induced, replicate counts</th>
<th>Up- or down-regulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up-regulated genes (Gate 1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y00787</td>
<td>IL-8</td>
<td>1/0</td>
<td>663/808</td>
<td>&gt;500×</td>
</tr>
<tr>
<td>X65965</td>
<td>SOD-2</td>
<td>0/0</td>
<td>1636/1096</td>
<td>&gt;500×</td>
</tr>
<tr>
<td>M25315</td>
<td>MIP-1</td>
<td>0/0</td>
<td>1120</td>
<td>&gt;500×</td>
</tr>
<tr>
<td>J04130</td>
<td>Act-2</td>
<td>0/0</td>
<td>453/542/572/606</td>
<td>&gt;500×</td>
</tr>
<tr>
<td>X02910</td>
<td>TGF-α</td>
<td>0/0</td>
<td>81/82</td>
<td>&gt;500×</td>
</tr>
<tr>
<td>J03210</td>
<td>Collagenase type IV</td>
<td>0/0</td>
<td>92/93</td>
<td>&gt;500×</td>
</tr>
<tr>
<td>M21121</td>
<td>RANTES</td>
<td>0/7</td>
<td>113/92</td>
<td>29×</td>
</tr>
<tr>
<td>M92357</td>
<td>B94</td>
<td>7/3/0/2</td>
<td>155/150/124/174</td>
<td>50×</td>
</tr>
<tr>
<td>Down-regulated genes (Gate 2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X13546</td>
<td>HMG-17</td>
<td>75/83</td>
<td>6/4</td>
<td>15.8×</td>
</tr>
<tr>
<td>M17885</td>
<td>Acidic ribosomal protein P0</td>
<td>191/195</td>
<td>84/54</td>
<td>2.8×</td>
</tr>
<tr>
<td>X16869</td>
<td>Elongation factor 1α</td>
<td>577/607</td>
<td>268/270</td>
<td>2.2×</td>
</tr>
</tbody>
</table>

derived from beads collected in gates 1 or 2 were shown to be
either up-regulated 29- to 500-fold or down-regulated 2- to
16-fold, respectively. These changes spanned expression levels
necessary is that the microbead library be large enough to
contain nearly all of the sequences from the libraries being
compared. No previous knowledge of any of the sequences
is required. Nucleic acid samples therefore can be from any source.
The microbead libraries used for comparisons are constructed
from a combination of the libraries being analyzed. Probes are
constructed separately from the libraries and are labeled
differentially. The ratio with which probes hybridize to the microbeads
is equally accurate for microbeads carrying rare and abundant
species, because each microbead responds to the probes inde-
pendently of how many others carry the same cDNA.

Most importantly, our method allows libraries to be probed
very deeply, much more deeply than is practical by other
methods. To attempt identification of differentially expressed
clones by deep, random, expressed sequence tag sequencing of
the libraries being compared would be prohibitive. To do so with
high-density planar microarrays (chips or gridded clones on
filters) would require access to chips or filters containing all
sequences expressed by the system and, even then, might not
adequately detect low abundance genes. To do so with serial
analysis of gene expression experiments limited to 2,000 se-
queencing runs (the same number we used) would be equivalent
to probing only about 50,000 signatures deep whereas our
extracted, differentially expressed, clones were the result of
probing the libraries 2,000,000 deep. Because our technique
readily selects out a much-reduced set of clones relevant to a
biological system or an experiment, it can be deployed as a
discovery tool that enables easier and more effective uses of
other techniques.

We thank Laurie J. Goodman, Holger Hiemisch, Hans Christian Kornau,
An Lai, Ling Yau, Jason Zeng, and Eugene A. Zhukovsky for their
assistance in this project.