Developmental Biology **228,** 270–286 (2000)
doi:10.1006/dbio.2000.9941, available online at http://www.idealibrary.com on IDEAL®

metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

Recovery of Developmentally Defined Gene Sets from High-Density cDNA Macroarrays

Jonathan P. Rast, Gabriele Amore, Cristina Calestani, Carolina B. Livi, Andrew Ransick, and Eric H. Davidson[1]
Division of Biology 156-29, California Institute of Technology, Pasadena, California 91125

New technologies for isolating differentially expressed genes from large arrayed cDNA libraries are reported. These methods can be used to identify genes that lie downstream of developmentally important transcription factors and genes that are expressed in specific tissues, processes, or stages of embryonic development. Though developed for the study of gene expression during the early embryogenesis of the sea urchin *Strongylocentrotus purpuratus,* these technologies can be applied generally. Hybridization parameters were determined for the reaction of complex cDNA probes to cDNA libraries carried on six nylon filters, each containing duplicate spots from 18,432 bacterial clones (macroarrays). These libraries are of sufficient size to include nearly all genes expressed in the embryo. The screening strategy we have devised is designed to overcome inherent sensitivity limitations of macroarray hybridization and thus to isolate differentially expressed genes that are represented only by low-prevalence mRNAs. To this end, we have developed improved methods for the amplification of cDNA from small amounts of tissue (as little as ~300 sea urchin embryos, or 2 × 10^5 cells, or about 10 ng of mRNA) and for the differential enhancement of probe sequence concentration by subtractive hybridization. Quantitative analysis of macroarray hybridization shows that these probes now suffice for detection of differentially expressed mRNAs down to a level below five molecules per average embryo cell. © 2000 Academic Press

Key Words: **macroarray; arrayed cDNA library; sea urchin; *Strongylocentrotus purpuratus;* subtractive hybridization.**

INTRODUCTION

The recovery of genes expressed specifically under given developmental conditions is a key initial step in unraveling genetic networks. For the few species for which extensive microarrays and large amounts of sequence information are already available, chip hybridization technology (Lipshutz *et al.,* 1999; Brown and Botstein, 1999) may suffice. For the majority of organisms, however, such resources are currently inaccessible. Here we describe technologies that will allow the detection of differentially expressed genes from any biological material from which cDNA libraries and probes can be obtained. These methods rely on arrayed cDNA libraries that have been gridded in the form of bacterial colonies onto nylon filters ("macroarrays"). The libraries used to generate macroarray filters can easily be made large enough to provide essentially complete coverage of the transcripts expressed in many developmental circumstances. Given the necessary robotic arraying device, macroarray filters can be produced with relatively little labor and at a relatively low cost. Although bacterial colony growth is inherently more variable than is direct spotting of DNA, we have found that in fact only moderate variation usually occurs among spots on a given filter, while the problem of variability between filters can be circumvented entirely by rescreening the same set of filters in any given comparative experiment.

The approach described in this paper was developed for the purpose of identifying genes that are linked within transcriptional networks that function in the embryogenesis of the purple sea urchin, *Strongylocentrotus purpuratus.* The major difficulty to be overcome is that most of the sequences of interest are of low prevalence in the embryo mRNA populations. On the other hand, the large size of macroarray filters requires that hybridization reactions be carried out in correspondingly large volumes. In consequence, after hybridization with complex cDNA probes, only message species of moderate to high prevalence achieve enough C_0t to drive hybridization to the corresponding macroarray spots sufficiently so that the signal intensity is significantly above background. Macroarray

[1] To whom correspondence should be addressed. Fax: (626) 793-3047. E-mail: davidson@mirsky.caltech.edu.

0012-1606/00 $35.00
Copyright © 2000 by Academic Press
All rights of reproduction in any form reserved.

filter screens are thus relatively insensitive to low-prevalence messages. Here we characterize in detail their sensitivity limits using complex probes derived from complete embryo or cellular mRNA populations. The main objective of this work has been the invention of means by which these sensitivity barriers can be bypassed so as to allow reliable detection of very low prevalence, differentially expressed sequences. We have revised procedures for global amplification of cDNA populations from small starting samples and for the generation of subtractive probes. Together these technologies transform macroarray screening into a powerful tool for gene discovery with potentially wide phylogenetic application.

MATERIALS AND METHODS

Hybridization Kinetics

Characterization of probe hybridization to macroarray filter spots. Small filters for probe excess and filter-driven hybridization experiments consisted of arrays of four blocks; each block included spot pairs of four probe target cDNA clones, negative control clone spot pairs, and empty background control positions. These were gridded by the same means as were actual library filters. Probes consisted of four 40-mer oligonucleotides representing regions of the four cDNA sequences. Probes were end-labeled by standard methods with ^{32}P using T4 polynucleotide kinase. Filters were hybridized to a specific C_0t in 0.41 M phosphate buffer (PB), 1.5% SDS, pH 7.4, at 59°C. Filters were then washed three times for 30 min each in 1× SSC (1× SSC is 0.15 M NaCl, 0.015 M sodium citrate), 0.1% SDS, once at room temperature and twice at 59°C. Bound probe was quantified by dry scintillation counting, and readings were converted to counts per minute. After being counted filters were exposed to film to confirm specific localization of hybridization.

Complex probe hybridization measurements. Total message from transgenic cells was amplified by the method described below. The precise prevalence of green fluorescent protein (GFP) sequence within the complex probe population that hybridized in solution to a 507-bp GFP fragment was determined by probe excess hybridization as outlined in Lee *et al.* (1986). Radiolabeled RNA probe was made from this same complex cDNA template. The GFP hybridization target on the macroarray filters consisted of bacterial clones containing a plasmid with an insert of the same 507-bp GFP fragment as was used for the probe characterization.

Arrayed cDNA Library Construction

A random-primed directional cDNA library was constructed from gastrula-stage sea urchin embryos (40 h after fertilization) as described previously (Lee *et al.*, 1999a). The library contains inserts with an average length of 1500 bp. Approximately 110,000 clones were spotted onto six 22 × 22-cm Hybond N$^+$ nylon filters (Amersham Pharmacia Biotech, Piscataway, NJ) with a "Q-Bot" robot (Genetix Ltd., New Milton, Hampshire, UK). In order to maintain consistent spot size and reduce streaking, bacterial filters were grown on 3% agar plates and colony size was rigorously monitored. Colonies were grown for approximately 12 h at 37°C, then processed for screening according to standard protocols (Clark *et al.*, 1999; Nizetic *et al.*, 1991).

cDNA Amplification and Subtractive Probe Synthesis

In order to maintain reproducibility and minimize prevalence skewing, careful consideration must be given to each step of the cDNA synthesis and amplification procedures. Here we present a description of our strategy. Space does not permit a detailed description of the procedures, but a full set of protocols is available on our Sea Urchin Genome Project Web site (http://sea-urchin.caltech.edu:8000/genome/).

RNA isolation. Total RNA was isolated from batches of staged sea urchin embryos or, for some experiments reported here, from 2 × 10^5 disaggregated embryonic cells that were expressing transgenes encoding the SpBrachyury transcription factor (Peterson *et al.*, 1999) and GFP, both under the control of ubiquitous promoters. Mosaic transgenic embryos were disaggregated (McClay, 1986) and GFP$^+$ cells, which carry the *brachyury* transgenes, were separated from nontransgenic cells by fluorescence-activated cell sorting (FACS). Thus pure populations of cells expressing a transgene were isolated. RNA was extracted with RNAzol (Leedo Medical Laboratories, Houston, TX). Polyadenylated RNA was isolated on Poly-$T_{(25)}$ magnetic beads (Dynal, Inc., Lake Success, NY).

cDNA synthesis. First-strand cDNA synthesis was primed with a biotinylated primer containing a T7 RNA polymerase binding site and a random hexamer sequence: LT7RND-BT, 5′-[biotin]-CGGAGGTAATACGACTCACTATAGGGAGNNN-NNN-3′. Second-strand cDNA was synthesized by RNase H-mediated nick translation (Gubler and Hoffman, 1983). cDNA synthesis enzymes and buffers were taken from a Marathon cDNA synthesis kit (Clontech, Palo Alto, CA).

A double-stranded linker (SLINKER) was ligated to the cDNA that was to be used as a target of subtraction (here termed "selectate"). The dideoxycytidine of the linker (ddC, shown below) blocks extension into the primer-complementary region of the linker, thereby minimizing the linker–dimer contribution to the cDNA PCR amplification product (Siebert *et al.*, 1995). The portion of the long strand that is underlined corresponds to the PCR primer used later for amplification. The short strand of the linker is 5′-phosphorylated (P).

S-linker: 5′-<u>GGGTGCTGTATTGTGTACTTGAACG</u>GGCGGCCGCA-3′
3′-**ddC**CCGCCGGCGT-**P**-5′

For cDNA that was to be used as driver, a similar linker but of different sequence (DLINKER; shown below with corresponding PCR primer underlined) was used in order to avoid linker-driven annealing during subtractive hybridization.

D-linker: 5′-<u>GCCAACGTATGTAAGGTTGAGTTCC</u>GGGCAGGT-3′
3′-**ddC**CCGTCCA-**P**-5′

cDNA amplification. Following linker ligation, a 300- to 500-bp fraction of selectate cDNA was size selected by agarose gel electrophoresis. When the initial cDNA mass was insufficient for direct size selection, a limited PCR amplification step was introduced prior to gel electrophoresis. Amplification reactions were carried out using a thermostable polymerase mix (Clontech Advantage-2; Clontech). Thermal cycling parameters were as follows: 95°C, 2-min initial denaturation followed by 3 to 20 cycles of 95°C, 1-min denaturation, 60°C 1-min primer annealing, 72°C 3-min extension.

The size-selected, double-stranded cDNA (whether preamplified or not) was used as template in an amplification reaction using

Copyright © 2000 by Academic Press. All rights of reproduction in any form reserved.

parameters identical to those described above, this time with a 5′-biotinylated version of the T7 primer (BT-LT7PRIMER). A pilot reaction was done and the optimized parameters were used in a scaled-up, minimal-cycle-number amplification reaction to produce approximately 3 μg of product that was biotinylated on the minus strand.

Preparation of single-stranded selectate. Plus-strand selectate was isolated from size-selected biotinylated cDNA by capture on streptavidin-coated magnetic beads (M-280; Dynal, Inc.) and elution under alkaline conditions according to the manufacturer's protocol for preparation of strand-specific probe template.

Preparation of Driver RNA. cDNA template for minus-strand driver RNA synthesis was produced from normal 15-h postfertilization embryos as described above except (1) it was ligated with a different linker (DLINKER), (2) only three to five cycles of PCR were used for amplification, and (3) no size selection was performed. Biotinylated RNA driver was synthesized using a Megascript T7 RNA polymerase kit (Ambion, Austin, TX). A 2:1 UTP:biotin-16–UTP solution (Roche Molecular Biochemicals, Indianapolis, IN) was used at the suggested UTP molarity. Driver RNA was prepared similarly for hydroxylapatite column separation, with the exception that only nonbiotinylated UTP was included in the T7 synthesis. RNA size was determined by denaturing gel electrophoresis and found to range in size from 500 to 2000 nt.

Subtractive Hybridization and Phenol:Chloroform Driver Extraction

Subtractive hybridization mixtures included 200 ng of plus-strand selectate and 5 μg of minus-strand biotinylated driver in a 10-μl volume of hybridization buffer (25 mM Hepes, pH 8.0, 5 mM EDTA, 0.1% SDS, 0.6 M NaCl, 0.01% sodium pyrophosphate). After complete denaturation, the sample was incubated at 65°C for 20 h. A streptavidin:phenol separation method was used for some experiments to separate nonhybridized from hybridized selectate, modified from a previously described method (Sive and St John, 1988). Phenol extraction was performed on ice in a cold room to minimize the effects of phenol:water emulsion on hybridization rate (Kohne *et al.,* 1977). After extraction, fresh biotinylated driver RNA was added and the sample was precipitated. The hybridization and extraction procedures were then repeated. Remaining single-stranded selectate was amplified with eight cycles of PCR using the linker and T7 primers to generate double-stranded template for a T7 RNA probe labeling reaction.

Hydroxylapatite Column Chromatography

Hydroxylapatite (HAP) column chromatography was used to improve isolation of nonhybridized single-stranded selectate (for general review of HAP chromatography methods, see Britten *et al.,* 1974). For these separations only one round of subtractive hybridization was performed, but the time of hybridization was extended to 40 h. After hybridization, HAP chromatography was performed on a water-jacketed column in 0.12 M PB at 60°C. At this temperature and salt criterion double-stranded nucleic acid remains bound to the column. Single-stranded selectate is eluted with a 0.12 M PB column wash. This eluate was desalted by concentration and dilution in a Centricon YM-10 spin filter (Millipore, Bedford, MA). Radiolabeled RNA probe was prepared as described below.

To assess the sequence enrichment after HAP chromatography, two 500-nt, single-stranded fragments of bacteriophage λ (λ1 and λ2) were employed as a controls. The complementary strand of λ2 was present as RNA in the driver, while λ1 sequence was unique to the selectate. Each sequence was present at a concentration corresponding to 10 copies per average embryo cell. Enrichment of the λ1 sequence relative to λ2 and endogenous ubiquitin was assessed by quantitative PCR.

Hybridization of Macroarray Filters

Probe labeling. Complex radiolabeled RNA probes were synthesized by incorporating [α^{32}P]UTP into T7 RNA polymerase transcripts using a T7 transcription kit (Maxiscript; Ambion) and 10 to 100 ng of complex cDNA template. Reactions were set up to yield 2 to 3 μg of RNA probe with a specific activity of approximately 1×10^8 cpm/μg.

Filter hybridization. Library filters were screened by conventional methods (described in detail on our Web site). The filters were hybridized with probe for 48 h. After hybridization filters were washed to high stringency (65°C; 0.1× SSPE, 0.1% SDS, 0.05% sodium pyrophosphate). The filters were exposed to a phosphor screen to allow near saturation for the most intense colony spots. The screen was scanned at 100 or 200 μm resolution on a phosphorimager (Storm 820, Molecular Dynamics, Sunnyvale, CA). Following data acquisition, filters were stripped by alkaline treatment (Clark *et al.,* 1999).

Filter analysis. Filters were analyzed using the VisualGrid program (www.gpc-biotech.com; Genome Pharmaceuticals Corp., Munich, Germany). Intensity of spot hybridization was calculated as the 80% quantile within a circle placed over the spot (i.e., the pixel value for which 80% of pixels were lower and 20% higher). Local background values were calculated for each 4 × 4 block of spots. Hybridization intensity values for each clone were calculated by averaging duplicate spot pixel intensities and then subtracting the background values taken from the corresponding blocks. To estimate values in counts per minute, a conversion factor was determined empirically by scintillation counting after timed phosphor screen exposures.

Quantitative PCR

cDNA sequences were used to design primer sets to amplify products of 125 to 150 bp length. RNA was isolated from groups of 500 transgenic or normal embryos by extraction with RNAzol (Leedo Medical Laboratories). First-strand cDNA was primed with random hexamers using MultiScribe reverse transcriptase according to manufacturer's instructions (PE Biosystems, Foster City, CA). Amplification reactions were analyzed on an ABI 5700 sequence detection system using SYBR Green chemistry (PE Biosystems). Reactions were run in triplicate with cDNA from 5 or 10 embryos. For analysis of the HAP column separation, quantitative PCR was performed on aliquots of pre- or postcolumn fractions such that approximately 10,000 target molecules of the λ controls were predicted to be present. Thermal cycling parameters were 95°C for 30 s, 60°C for 1 min, 40 cycles. Gel electrophoresis was used to confirm product size. Primer efficiencies (i.e., the amplification factors for each cycle) were found to exceed 1.9.

Copyright © 2000 by Academic Press. All rights of reproduction in any form reserved.

RESULTS

Macroarray Filter Screens: Basic Parameters

For discovery of genes expressed in a specific regulatory context there are two basic screening requirements. These are, first, the use of a probe in which the mRNA species that define that context are accentuated and, second, a detection system for these heightened message concentrations that operates on a cDNA library large enough to represent adequately the total mRNA population. We take up the first of these requirements in later sections and here consider the second. The essential issue is that the larger the format of a hybridization array, the less sensitive the screen, because of the large hybridization volume required. In the following we examine the factors that define hybridization sensitivity on cDNA library macroarrays that are sufficiently extensive to represent almost completely the mRNA populations of sea urchin embryos. The sensitivity limits turn out to depend entirely on the kinetics of hybridization. This result directly determines our approach to the problem of using macroarrays for the isolation of differentially expressed, but relatively rare, mRNA species. That is, our strategy must specifically increase the concentration of probe corresponding to these mRNAs beyond the threshold of macroarray detection.

Macroarray variability. The macroarray filters used for screening in our work each contain 18,432 pairs of spots, arranged in a grid pattern which uniquely specifies the plate and well coordinates of each clone in the parental cDNA libraries (4 × 4 array pattern; Maier *et al.,* 1994). The filters are spotted with the Q-Bot arraying robot (Genetix, Ltd.), which uses pins of 400-μm tip diameter to deposit paired samples of each clone. Following growth, the clones are converted to bound DNA spots on the charged nylon membrane filter, by the usual procedures of alkali denaturation and protease treatment (Nizetic *et al.,* 1991). The amount of DNA per spot of course depends on the amount of clone growth, which may vary from run to run, and also from clone to clone. But in our experience, more than 90% of spot pair members differ from one another by less than twofold and most by less than 20%. Although the difference between replicate filters is often greater than this, the filters can be utilized repeatedly, e.g., with different complex probe populations. Therefore, even small differences in the representation of gene sequences within these probes can be reliably detected, so long as the representation level is above the sensitivity limit.

Transcript representation on macroarray filters. The following considerations show that cDNA libraries of the size we used suffice to include most expressed mRNAs. Each of the sea urchin embryo cDNA libraries contains approximately 110,000 clones, arrayed on six 22 × 22-cm filters. Sea urchin embryos transcribe about 8–10 × 10^3 genes, depending on the developmental stage (for review see Davidson, 1986), and they lack any highly prevalent sequences. The majority of the message sequences are rare in the embryo mRNA population. For example, at the 500-cell stage there are about 10^5 mRNAs/cell, and of these about 40% are of a low-prevalence class (<10 copies/cell; average ~5 copies/cell; mode ~2 copies per cell). The prevalence of the remaining 60% averages about 80 copies/cell, ranging up to a few hundred copies per cell. Only about 25% of the total mass of mRNA consists of molecules present at ≥50 copies per cell. These values are, of course, given in terms of prevalence per average embryo cell, and for a transcript expressed in, say, 20% of cells, the prevalence in these cells will be five times higher (Lasky *et al.,* 1980; Flytzanis *et al.,* 1982; reviewed by Davidson, 1986; for recent EST data, which entirely support earlier conclusions, see Poustka *et al.,* 1999; Lee *et al.,* 1999a). It follows from the complexity and prevalence distribution that a 1.1 × 10^5 clone library is likely to contain representatives of most of the sequences being expressed.

Sensitivity limits under normal hybridization conditions. In an initial series of experiments we determined that even under the maximum practical extent of hybridization only about a quarter of the clones is represented in the cDNA probe mixture at sufficiently high concentration to produce quantifiable reactions on the macroarray filters. In these experiments a mass of 1–2 μg of cDNA probe (sp act ~10^8 cpm/μg) was prepared from mRNA of the same stage as was used to generate the arrayed library and reacted with the filters for 48 h. The criterion of detectability in this experiment was that radioactivity bound to hybridized clones should exceed 1.6× background. In terms of transcript prevalence, this hybridization level corresponds to species present at ≥40–50 copies per average embryo cell. This limit, which we verify for specific cases below, sets the practical baseline for screening sensitivity in the absence of special procedures that increase the concentration of given sequences in the probe.

Hybridization to macroarray spot pairs: Rate constants, target mass per spot, and kinetics. To attain a quantitative understanding of hybridization reactions on macroarray filters estimates of two different parameters are required. These are the rate constant for hybridization of probe molecules to filter-bound DNA spots and the amount of hybridizable target DNA present in each colony spot. Measurements to determine these parameters were carried out on a miniaturized scale. The 22 × 22-cm macroarrays consist of 2304 small squares (blocks), each containing eight spot pairs. Miniature filters (1.1 × 1.1 cm) were produced that consisted of four of these blocks onto which known cDNA clones had been spotted, as described under Materials and Methods. A series of replicate reactions was carried out, using 100- and 200-ng probe concentrations per well to ensure kinetic termination. From the termination values, given the known probe specific activity, the hybridizable DNA per spot could be obtained directly. The kinetics of the solution-to-filter hybridization reactions provided estimates of a rate constant for the solution-driven reaction that is applicable to any specific sequence in a macroarray screen. Results from three different sets of reactions are

Copyright © 2000 by Academic Press. All rights of reproduction in any form reserved.

TABLE 1
Reaction Kinetics for Probe Excess Hybridization to Macroarray DNA Spots

Expt[a]	Probe[b]	k (M^{-1} s^{-1})[c]	Error[d]
1	100 ng	1920	8.9%
2	200 ng	3270	12%
3	100 ng	1450	18%

[a] Reactions were as described in the legend to Fig. 1. A mixed probe consisting of four 40-base oligonucleotides was reacted with a 1.1-cm filter square containing 32 DNA spots. The oligonucleotide sequences were derived from four different cDNA clones, each present in four spot pairs on the sequence (see Materials and Methods and text for setup). Parameter values shown were obtained by nonlinear least squares (MATLAB5.3), assuming pseudo-first-order reaction kinetics (Davidson, 1986, Eq. (17), p. 538).

[b] Per 300-μl volume.

[c] As observed, i.e., not corrected for salt or length effects.

[d] Error is calculated as $[\Sigma \text{ (deviations at each point)}^2/n - 1]^{1/2}$. The amount of scatter associated with a given error is illustrated for Expt 3 in Fig. 1, which displays the largest scatter of the three experiments. Error is expressed here as percentage of maximum value.

shown in Table 1, and an example of the kinetics observed is illustrated in Fig. 1. The conclusions can be summarized as follows: there are about 2–4 ng of total hybridizable DNA (vector + insert) per spot (see legend to Fig. 1). For a 1-kb insert, each spot contains 250–500 pg of the transcribed strand of insert DNA. The hybridization rate constant measured was about 2×10^3 M^{-1} s^{-1} for this probe mixture, which is of 160-nucleotide aggregate complexity. This is close to the value expected from the solution kinetic standard (i.e., only about 1.5× slower; Galau *et al.,* 1977). The rate constant is particularly useful, for it dictates the amount of hybridization that will be obtained given the sequence concentration of any specific probe species.

For solution-driven reactions such as those of Table 1, as probe sequence concentration is decreased, the rate of the reaction will decrease. Eventually, there is more DNA on the filters than probe in the available and accessible solution. Filter DNA-driven reactions rather than probe-driven reactions should then occur. However, because the DNA is filter-bound, the rate of these reactions is not necessarily predictable from solution reaction kinetics, though these should certainly provide a useful guide. The uncertainties introduced by this combination of probe- and filter-driven reactions could complicate prevalence determinations by hybridization to macroarray filters. We therefore set out to characterize macroarray filter-driven reactions. In the following experiment, bound-DNA-excess kinetics were established for the 32-target spot filters described above in order to determine from the rate observed whether bound-DNA-driven kinetics could ever in practice affect a macroarray screening reaction. Initial considerations indicated that at the DNA per spot values derived from the experiments of Table 1, the bound DNA could not obtain sufficient C_0t to drive reactions to completion in the 0.3-ml volume used for previous experiments, and so a smaller chamber in which the reaction could be carried out in a 30-μl volume was devised. The total amount of hybridizable DNA per chamber is nominally 640 pg (i.e., 32 cDNA spots represented in the probe spots and 20 pg of DNA complementary to the 40-nucleotide probe per spot, from Fig. 1). Figure 2 shows the reactions of 20 and 100 pg of probe DNA with these filters in 30 μl of medium.

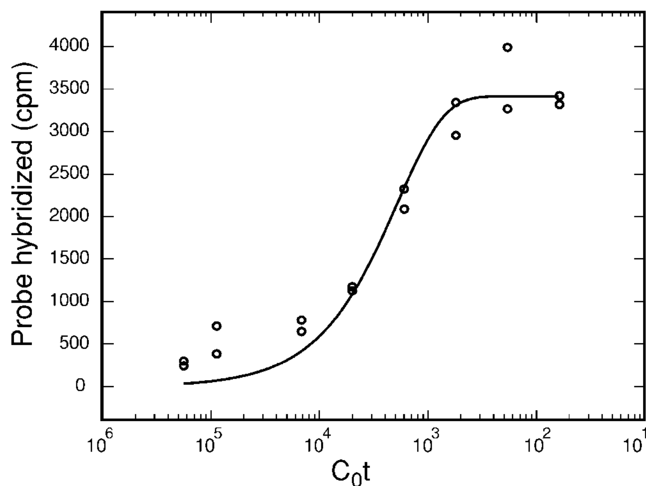

FIG. 1. Measurement of DNA per spot and kinetics of probe excess hybridization reactions. Filters (Hybond N^+) bearing DNA from four different cDNA clones, each represented in four spot pairs (32 total spots), were hybridized in 300 μl of a phosphate buffer (see Materials and Methods). Total probe complexity was 160 bases (four 40-base oligonucleotides; see text). In this experiment (No. 3 of Table 1) the termination value, 3450 cpm, represents 5.2 ng DNA per spot, given that the probe specific activity was 4.15×10^5 cpm/100 ng, and the average clone length is 4 kb including vector [(3450 cpm/32 spots) × (100 ng/4.15×10^5 cpm) × (2 × 4000 bases/40 bases) = 5.2 ng]. The rate constant expected would be about 6700 $M^{-1}s^{-1}$ had the reaction been carried out at a 400-base fragment length in solution, by reference to ϕX174 pseudo-first-order standard kinetics measured by Galau *et al.* (1977); the complexity of the ϕX174 genome is 5374 bases and the pseudo-first-order rate measured for that genome was about 200 M^{-1} s^{-1}. Here a linear fragment length effect is assumed, i.e., 6700 × 40/400 = 670 (the $L^{1/2}$ rate applied for solution reactions reflects retardation due to interpenetration of the excluded volumes of the reactants, here irrelevant because the target DNA is fixed in a more or less two-dimensional format and the probe DNA is very short). The reaction was run in 0.6 M [Na^+], which in pure solution reactions affords a 5× rate acceleration compared to the standard conditions of Galau *et al.* (1977) and Britten *et al.* (1974). Hence ideal expectation would be about 3300 M^{-1} s^{-1} (i.e., 670 M^{-1} s^{-1} × 5), but the measured rate was 1920 M^{-1} s^{-1}. Possible reasons for the slightly slower rate observed include a less steep salt-dependent rate factor for filter reactions, for which there is some evidence in others of these experiments (not shown).

Copyright © 2000 by Academic Press. All rights of reproduction in any form reserved.

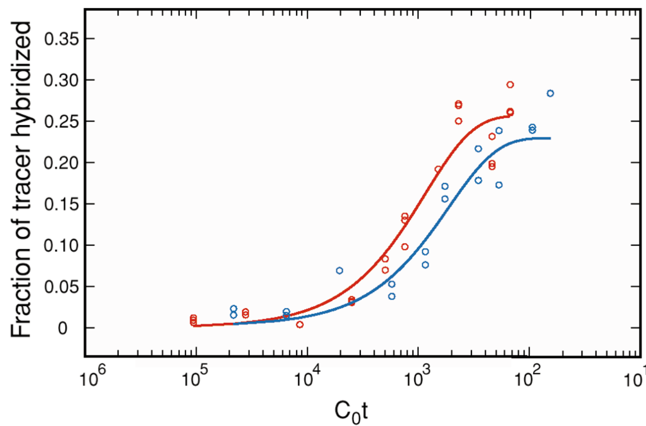

FIG. 2. Kinetics of bound-DNA-driven reactions. Filter squares containing 32 DNA spots as in the experiments of Fig. 1 and Table 1 were reacted with 20 and 100 pg of probe in a 30-μl volume (see Materials and Methods). The abscissa shows C_0t calculated on the basis of total bound-DNA content (i.e., about 640 pg in the 32 spots) and the total reaction volume. Blue symbols and curve represent the 100-pg reaction and red circles and curve the 20-pg reaction.

Under these conditions, the amount of reaction no longer depends on the probe C_0t, but only on the bound DNA C_0t calculated simply from total bound DNA per chamber volume, consistent with bound-DNA-driven kinetics. As expected the fraction of probe reacted at each point is almost the same for the two different probe concentrations though they differ by fivefold. Thus the rate constants obtained were 860 and 730 $M^{-1}\ s^{-1}$, for the 100 and 20 pg probe reactions. These values are in fact two- to threefold lower than the probe-driven rates of Table 1, perhaps because of some filter occlusion effect.

In Fig. 2, the total amount of probe hybridized was only about 30% as kinetic termination is approached. This implies that some fraction of the probe was inaccessible, perhaps nonproductively bound within the filter itself (i.e., until washing after the reaction) or on the chamber walls. Images of the filters following hybridization (not shown) confirm that after washing (see Materials and Methods) all of the detectable radioactivity is in fact found at the expected spot locations. In other experiments with different cDNA clones, however, the terminal values ranged from 60 to 70% hybridized. These numbers are not important in themselves; they simply indicate that when the reaction terminates a large fraction of the probe, though not all, has been driven into filter-bound hybrid at the 32 spots in bound-DNA-driven reactions, essentially as expected. The main significance of these experiments is to prove that the rate constants for bound-DNA-driven reactions would be much too low to affect macroarray screens in the large volumes that these require. Therefore, the amount of reaction obtained by such screens depends exclusively on the probe-driven kinetics, that is, on the sequence concentration for each species in the probe mixture.

Hybridization Kinetics for Components of Complex Probes Hybridized to Multiple Macroarray Filters

To demonstrate the predictive power of the kinetic determinations outlined in the last section for actual macroarray screening experiments, rate constants were estimated for individual components within complex cDNA probes. The complex probe used for this experiment was synthesized from the mRNA of embryos that had been injected with *brachyury* and *GFP* expression constructs, driven by ubiquitously active *cis*-regulatory elements. After injection

TABLE 2

Determination of Hybridization Rate to GFP Control Spots

Probe	GFP in probe[a]	C_0t for GFP[b]	Probe on spot[c]	Rate constant[d]
Unsubtracted	2.49 ng	2.13×10^{-5} M s	1.49 pg	563 $M^{-1}\ s^{-1}$
Subtracted	8.41 ng	7.08×10^{-5} M s	4.84 pg	558 $M^{-1}\ s^{-1}$

[a] The mass of GFP in the probe was determined by probe excess hybridization to unlabeled RNA that was synthesized from the same template as was the radiolabeled probe.

[b] The specific C_0t for GFP within the probe was calculated from the GFP mass, reaction volume, and hybridization time. The reaction volumes for the unsubtracted and subtracted probes were 62 and 63 ml, respectively. This was determined by adding 100 ml of hybridization buffer to the four dry filters (with separator meshes) and removing the measured excess of liquid. Both hybridization reactions were carried out for 44 h.

[c] The mass of probe bound to the spots was calculated by measuring the intensity of hybridization as determined by exposure to a phosphor screen followed by scanning. The value reported is an average for 10 GFP spot pairs. These are replicate colonies of a clone containing a 507-bp fragment of the GFP coding region. Background intensity was determined by measuring nearby spots for which no signal was apparent. Data from timed phosphor screen exposures were converted to cpm using a conversion factor determined in similar measurements for which corresponding scintillation values were obtained.

[d] Predicted rate constants were determined by assuming pseudo-first-order reaction kinetics. The strand-specific target DNA per spot is assumed to be 125 pg. The apparent rate constant (k) was thus calculated as $k = -\ln[(125\text{ pg} - \text{GFP on spot})/125\text{ pg}]/C_0t$.

Copyright © 2000 by Academic Press. All rights of reproduction in any form reserved.

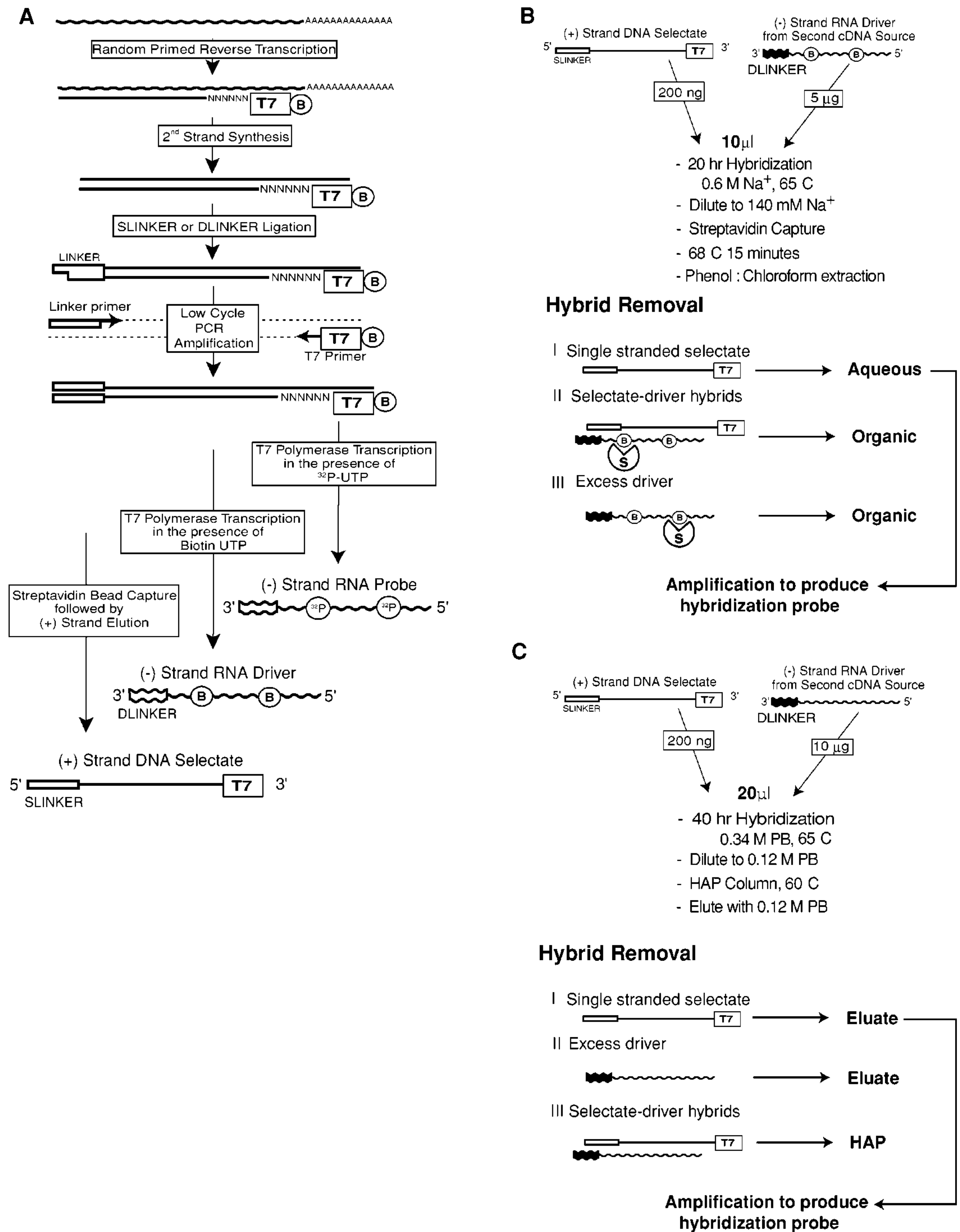

FIG. 3. A schematic illustration of cDNA synthesis and subtractive hybridization methods. (A) cDNA, probe, and selectate synthesis. RNA is shown as wavy lines; straight lines represent DNA. To begin isolated mRNA is reverse transcribed from a random primer. This primer is biotinylated ("B") and carries a T7 RNA polymerase binding site at its 5′ end (T7). Second-strand cDNA is synthesized by an RNase H-nick translation reaction using *Escherichia coli* DNA polymerase. Either of two linkers is then ligated, depending on whether the material is to be used as driver or selectate, and the cDNA is PCR amplified for a minimal number of cycles. Radiolabeled complex probe may then be synthesized by incorporating [^{32}P]UTP during RNA synthesis by T7 RNA polymerase. Alternatively, the cDNA can be enriched for differential transcripts in a subtractive hybridization reaction. For subtraction, the T7 primer used for PCR amplification carries a biotin molecule ("B") at its 5′ end. The PCR products are captured on streptavidin beads and the nonbiotinylated plus strand is eluted by alkaline denaturation. This strand is then used as selectate. Driver is produced in a T7 RNA polymerase reaction that incorporates

Copyright © 2000 by Academic Press. All rights of reproduction in any form reserved.

into fertilized sea urchin eggs, linearized exogenous DNA molecules are concatenated together (McMahon *et al.,* 1985) and co-incorporated into embryo cells in a mosaic fashion, usually at second, third, or fourth cleavage (Hough-Evans *et al.,* 1988; Livant *et al.,* 1991). Those cells expressing GFP in such an experiment also express whatever other construct is co-injected with the GFP construct (Arnone *et al.,* 1997), here the *brachyury* expression construct. The embryos were dissociated in calcium-free seawater, and cells displaying GFP fluorescence were separated out by FACS. The sorting step afforded approximately 10-fold enrichment of message specific to the transgenic cells, with respect to whole embryo mRNA.

We estimated rate constants for hybridization to GFP clones included in the macroarray library filters, on the basis of the mass of bound target sequence in a spot, as determined in the previous set of experiments, and the sequence concentration of GFP mRNA within the complex probe. The length of the GFP clone insert used for the experiments was 507 bp, and based on the above measurements, each spot should contain ~125 pg of hybridizable GFP sequence. Ten sets of duplicate colonies of this clone were gridded among the cDNA clones on four 22 × 22-cm high-density filters of the 40-h embryo cDNA library. Two complex probe mixtures in which the GFP concentrations differed by 3.4-fold were analyzed. The lower GFP concentration is equivalent to a transcript present in about 125 copies per average cell. Pseudo-first-order rate constants of 560 M^{-1} s^{-1} were estimated for both probe concentrations (see Table 2 for details). Considering that the probe sequence is about three times longer than the aggregate complexity used for the probe excess rate determination above (160 bp vs 507 bp), the rate constant estimated for the GFP sequence is almost exactly as expected. The results are also consistent with the hybridization intensities obtained for clones for which the corresponding mRNA prevalence is known. For example, ubiquitin, at approximately 100 copies per average cell (Nemer *et al.,* 1991), produces a signal of around twice that of GFP, close to what would be expected for a ubiquitin clone with a 1.5-kb insert. In summary, the individual sequence components in the complex probe mixtures behave just as do the experimental oligonucleotide probes of the last set of experiments, and the radioactivity detected in a spot pair after hybridization with a complex probe is an interpretable reflection of the prevalence of that species within the probe.

Amplification Strategies

cDNA synthesis and amplification for subtractive hybridization. Given the rate limitations that we describe above, it is clear that for macroarray screening to be useful, the concentration of sequences that are differentially represented in the probe must be increased. The absolute concentration of a particular cDNA species within a complex probe is of course dependent on both the relative prevalence of that species and the total probe mass. It can often be difficult to obtain sufficient probe mass in an embryological investigation for hybridization to macroarray filters. Although any form of cDNA amplification risks skewing the cDNA prevalence distribution (see below), the use of some magnification procedure is unavoidable when working with limiting quantities of embryos, tissues, or cells. To minimize skewing where this is important, while maximizing amplification, we employed elements from several previously described methods (Hampson *et al.,* 1996; Sagerstrom *et al.,* 1997; Siebert *et al.,* 1995). The resulting procedure is shown in outline form in Fig. 3A (see Materials and Methods for details). As much as possible of the burden of amplification is placed on a T7 RNA polymerase reaction, which generates copies of the initial template in a linear fashion rather than exponentially, as in PCR amplification (Philips and Eberwine, 1996; Kacharmina *et al.,* 1999). First-strand cDNA is synthesized from a random primer in order both to eliminate bias toward the mRNA 3′-end and to maximize sequence overlap with the random-primed arrayed cDNA libraries on which our analyses take place. In addition to the randomized primer sequence, the first-strand anchor contains a T7 RNA polymerase initiation site at its 5′-end that is used both as a PCR amplification primer site and later to generate RNA copies of the cDNA. After second-strand synthesis a directionally blocked linker (Siebert *et al.,* 1995) is ligated to the cDNA end corresponding to the 5′-mRNA end. Ligation of this linker to the cDNA end corresponding to the 3′-mRNA end is suppressed by the presence of a biotin moiety on the original T7 random primer (see Fig. 3A).

Selectate amplification. The cDNA that was to be used as the target of subtractive hybridization (here termed "selectate") was size separated by agarose gel electrophoresis, and a 300- to 500-bp fraction was selected and electroeluted from the gel. Following electroelution after size selection, the cDNA was used as template for PCR ampli-

biotin–UTP into the synthesis of minus-strand RNA. Different linkers are used for selectate (SLINKER) and driver (DLINKER) populations to eliminate linker-driven annealing during subtractive hybridization. (B) Subtractive hybridization. Reactions are performed in a 10-μl volume using 200 ng of selectate DNA and 5 μg of biotinylated driver RNA. After 20 h of hybridization the reactions are diluted and soluble streptavidin is added. The hybrids and unreacted driver are then removed by phenol:chloroform extraction. New driver is then added and the sample is subjected to a second round of subtraction. The remaining unhybridized selectate is PCR amplified for three to eight cycles to produce double-stranded product, and a radiolabeled probe is synthesized using T7 RNA polymerase. (C) Use of hydroxylapatite (HAP) instead of phenol:chloroform to separate unhybridized from hybridized selectate. This process significantly boosts separation efficiency, as described in the text (see Materials and Methods). A single cycle suffices for most purposes.

Copyright © 2000 by Academic Press. All rights of reproduction in any form reserved.

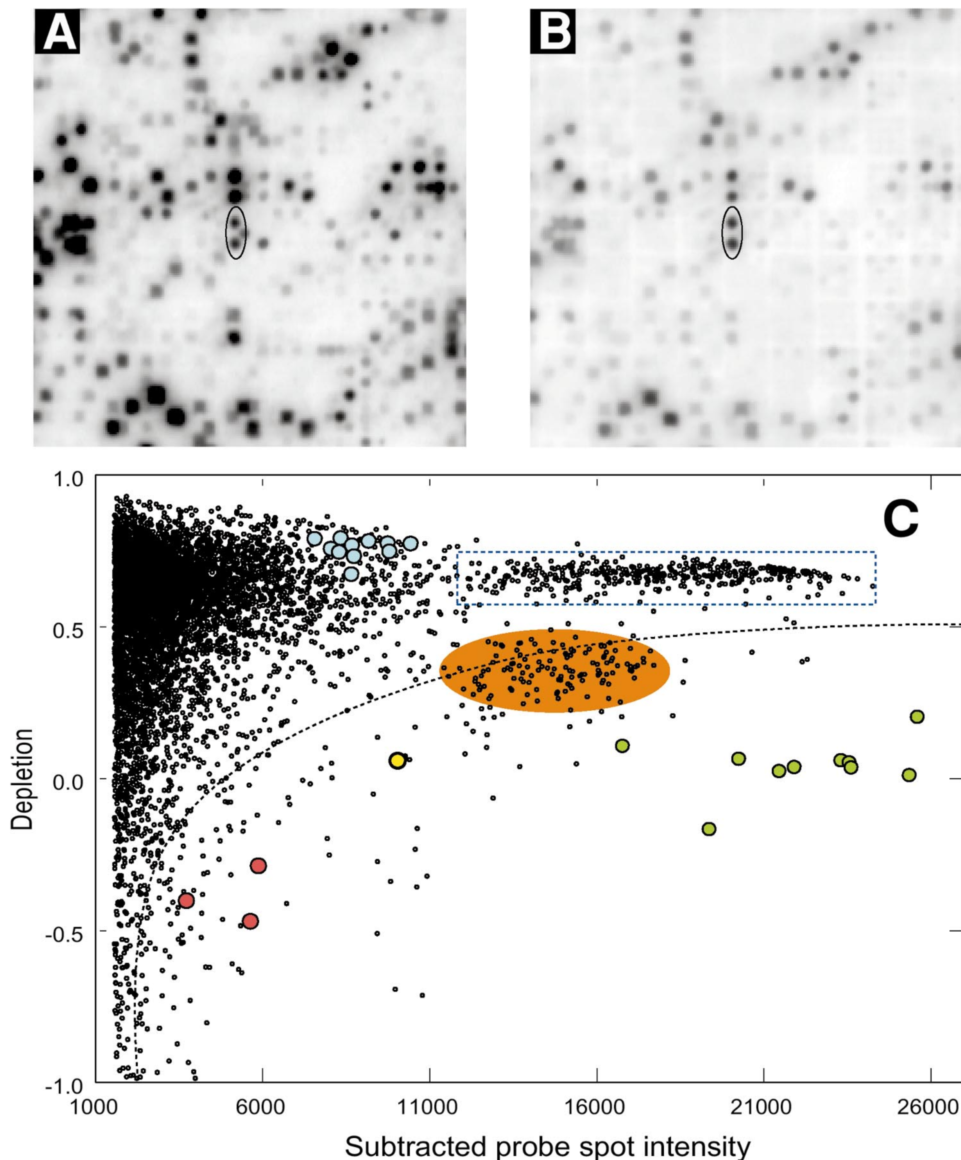

FIG. 4. Signal depletion for common sequences after subtractive probe hybridizations and concentration enhancement for differentially expressed sequences. (A and B) Images of a section of library filter, shown after successive hybridization with complex radiolabeled cDNA probes. (A) Unsubtracted probe; (B) subtracted probe. The section shown encompasses 648 clones, each spotted in duplicate. The selectate probes are made from transgenic sea urchin embryo cells that express GFP and also express the gene encoding the *brachyury* transcription factor. In the subtracted case (B), the probe is depleted by reaction with driver derived from normal 15-h embryos, which do not express GFP, and removal of driver–selectate duplex. Gray scale for the GFP spot pair at the center of each image (within ellipse) is held constant in (A) and (B) to illustrate the decrease in (B) of label bound to other clones, i.e., those common to both the driver and the selectate cell populations. (C) Analysis of probe hybridization to 6477 pairs of clone spots from the same experiment. These were all spots that exceeded 3 standard deviations above mean background upon hybridization with the subtracted probe. Signal depletion relative to GFP control spots after subtractive hybridization (ordinate) is plotted against spot intensity upon hybridization to the subtracted probe (abscissa). For the depletion calculation on the ordinate, all spot hybridization values are normalized to GFP, which is expressed exclusively in the transgenic (selectate) cell population and is not expected to be depleted. For this normalization, after clone spot pair readings are averaged and each measurement is corrected for background, the hybridization from the subtracted (S) and unsubtracted (U) probe was calculated as the ratio of counts detected in the spot pair of interest to the average counts hybridized to GFP spots (GFP_s and GFP_u, respectively). This average was calculated from 10 control spot pairs of a bacterial clone containing a 507-bp fragment of the GFP coding region. These were grown on the same library filter (green circles). The fractional depletion value plotted on the ordinate is calculated as $[1 - (S/GFP_s \div U/GFP_u)]$. Complete lack of depletion relative to GFP produces a fractional depletion value of 0, while complete depletion results in a value of 1. Enrichment relative to GFP produces a negative depletion fraction. The scale on the abscissa is taken from the phosphor screen scan and can be

Copyright © 2000 by Academic Press. All rights of reproduction in any form reserved.

fication. Only 5–10% of the original cDNA is retained, but the procedure eliminates primer–dimer contamination and homogenizes the product with respect to size so that length does not bias amplification efficiency. After amplification of cDNA from this template with a biotinylated minus-strand primer, plus-strand selectate is purified by elution from streptavidin beads. In all PCR steps, care is taken to limit cycle number to the minimum necessary and to allow for complete extension at each cycle. As we discuss below, skewing of prevalence in PCR can be tolerated (within limits) in the selectate, but must at all costs be avoided in a subtractive hybridization driver.

Driver amplification. The material used as driver to selectively capture sequence in common with selectate during subtractive hybridization consists of biotinylated minus-strand RNA. This is generated by a T7 RNA polymerase amplification reaction in the presence of biotin–UTP. Template for this transcription reaction is synthesized in a manner similar to that of selectate, but with a different linker. No size selection was performed for driver synthesis, and only three or four cycles of PCR (corresponding to two- to eightfold amplification because the first cycle merely replaces the minus strand) are required to prepare the amplicon ends for T7 polymerase. The T7 polymerase reaction is then used in place of any further PCR for amplification of driver mass.

Skewing of transcript prevalence by extensive PCR amplification. To characterize the fidelity and reproducibility of cDNA amplification by PCR alone, without the use of linear T7 polymerase amplification, and to assess skewing directly, comparisons were made among reverse-transcribed and amplified probes derived from the same sample of 20-h sea urchin embryo mRNA. Two amplification reactions were carried out independently to examine reproducibility, beginning with only 10 ng of mRNA as initial template. This is approximately the amount of mRNA obtained from 300 embryos. The amplifications were performed a little differently compared to that described above in that both the first- and the second-strand syntheses were initiated with random primers. The addition of a second random priming step could increase skewing, as a result of biased selection of initiation sites by the defined (5′) portion of the primer. Thus the comparison will provide a test of the upper limit of skewing induced by PCR. The amplification was carried out in 15 PCR cycles, using a primer sequence associated with the random hexamers (see Materials and Methods). Approximately 1 μg of antisense radiolabeled probe (sp act 1×10^8 cpm/μg) was used to screen a replicate of a 20-h embryo cDNA library macroarray filter. The two amplified probes were compared with one made without PCR amplification, i.e., by reverse transcription of a large amount of cDNA. Significant skewing of apparent prevalence is evident in comparing the reverse-transcribed to the amplified probes. Of the spots that were $>1.5\times$ background in the unamplified probe hybridization, 76% were also represented above $1.5\times$ background on filters hybridized with amplified probe. However, the remaining fourth of the total had fallen in prevalence below the $1.5\times$ background criterion. Upon visual inspection of autoradiography films, nearly all spots that were detectable above the surrounding background on hybridization with the unamplified probe were also detectable in the amplified probe hybridization. Qualitatively, the PCR-amplified probe was thus equivalent to the unamplified probe, and the total complexity of spots represented above background was similar for the amplified and unamplified probe hybridizations. Respectively, 26 and 20% of spot intensities were above $1.5\times$ background. These values represent the sensitivity limits for standard hybridization of complex probes to macroarray filters using about as much mass and radioactivity as is normally practical.

Reproducibility between the two independently amplified probes was very good, yielding a correlation coefficient of 0.96 when intensities for spots ($>1.5\times$ background) obtained with the two probes are compared. A separate experiment in which the same probe was hybridized to replicate copies of high-density library filters gave almost as good results (correlation coefficient 0.91). In conclusion, the main result is that skewing of apparent prevalence by PCR amplification is indeed a significant problem. But the behavior of each sequence appears to be highly reproducible in independent amplification reactions. Furthermore, even in PCR-amplified cDNA populations, the set of sequences recovered at the end remains qualitatively representative of the starting population.

Use of Subtractive Hybridization to Increase Sequence Concentration of Differentially Expressed Transcripts

Principles. The sensitivity limit observed in the previous experiment shows that low-prevalence sequences (i.e., below the top one-fourth or so of the population) remain

approximately related to accumulated radioactive counts per 200 × 200-μm pixel by multiplying by a factor of 1.3. Most spots are clearly depleted because they represent transcripts common to both driver and selectate. Spots that are candidates for differential expression are shown below an arbitrary demarcation line (black dotted line). A spot cluster that is likely to be made up almost entirely of cytochrome *c* oxidase clones is shown within a blue dotted rectangle (see text). A cluster composed mainly of ribosomal RNA clones is encompassed by an orange oval. Spots known to represent ubiquitin clones are shown as blue circles. Values for a clone representing *brachyury,* which is not expressed at significant levels in the 15-h embryo, are indicated by the yellow circle. As expected the *brachyury* message is not significantly depleted. Spots representing cDNAs of the *endo16* gene, which is also enriched after subtraction, are shown in red.

Copyright © 2000 by Academic Press. All rights of reproduction in any form reserved.

invisible when the total embryo mRNA is made into cDNA, amplified, and hybridized to a macroarray filter. We turned to subtractive hybridization in order to enrich the prevalence of sequences representing differentially expressed genes, which indeed for most developmental purposes are those that are desirable to recover. In subtractive hybridization, the sequence concentrations of differentially expressed cDNAs are enhanced at the expense of commonly expressed sequences, given that the overall probe mass is kept approximately constant. After subtractive hybridization the sequence concentration or prevalence in a hybridization probe of a particular differentially represented sequence (P_i) can be thought of as

$$P_{\mathrm{i}} = \frac{[\mathrm{D_i}]}{[\mathrm{D}] + [\mathrm{C}]}, \quad [1]$$

where [$\mathrm{D_i}$] is the concentration of the particular message species under consideration, [D] is the total concentration of all differentially expressed sequences (i.e., those present in the selectate and absent from the driver), and [C] is the total concentration of all sequences expressed commonly in the two parental RNA populations. By subtraction the value of [C] is reduced, thereby increasing P_{i} and the corresponding intensity of hybridization to cDNA spots representing that differentially expressed gene. In a typical subtractive hybridization experiment, [C] must be much larger than [D] or the procedure will be ineffective; that is the two mRNA populations must be in general closely similar. In this case [C] will provide a reasonable approximation of the value of the denominator of Eq. [1], and the factor by which [C] is reduced corresponds to the fold increase in the hybridization intensity of spots representing differential message species. The experimental scheme to accomplish this is outlined in Fig. 3B.

In our procedure selectate and driver are represented in the hybridization reaction in direction-specific, single-stranded forms. This eliminates the complicated mixtures of competing pseudo-first-order and second-order reactions that occur when one or both of the reaction components are present in double-stranded form. The rate of hybrid formation under our conditions is dependent entirely and in a direct way on the driver concentration. The fraction of selectate representing any particular species of sequence that is hybridized (C_{ih}) is

$$C_{\mathrm{ih}} = 1 - e^{-kC_{\mathrm{i0}}t}, \quad [2]$$

where k is the pseudo-first-order reaction rate constant for this species (about 600 $\mathrm{M}^{-1}\ \mathrm{s}^{-1}$ for the fragment lengths we used under standard reaction conditions and assuming a 1.5-kb complexity for typical mRNA; Galau *et al.,* 1977), C_{ih} is the concentration of hybridized molecules of the species i at time t, and C_{i0} is the initial concentration of that specific transcript species in the driver. As an example, the value of $C_{\mathrm{i0}}t$ at 30 h for probe RNA fragments derived from an mRNA of a prevalence of five copies per average embryo cell in our reactions (i.e., five copies per 10^5 transcripts or 250 pg of the 5 μg of driver in a 10-μl reaction volume) is 9.0×10^{-3} M s. For such an RNA, duplex formation is expected to proceed to very near completion (99.5%) over the 30-h period of hybridization, according to Eq. [2].

cDNA that had hybridized to biotinylated driver RNA was removed by streptavidin capture and phenol:chloroform extraction (Sagerstrom *et al.,* 1997; Sive and St John, 1988). Removal of biotinylated driver RNA (whether hybridized or not) by this method was generally determined to be around 95% in experiments in which its distribution was monitored by colabeling with ^{32}P. The remaining selectate was then precipitated in the presence of 5 μg of new driver and the hybridization process was repeated. Finally the surviving single-stranded selectate was converted to a double-stranded form by minimal PCR amplification (five to eight cycles) and as above used as template to prepare an amplified probe by T7 RNA polymerase transcription, in the presence of [^{32}P]UTP.

Additional studies demonstrated that it is more efficient to separate hybridized and unhybridized selectate by HAP chromatography than by phenol:chloroform extractions (this method is described in the following). HAP separation (illustrated in Fig. 3C) increases the sensitivity of macroarray screening, but does not qualitatively change the subtractive method, so the analysis below will apply similarly to screenings based on HAP subtraction.

Examples: enhancement of specific sequences by subtractive hybridization. For the example of subtractive hybridization analyzed in greatest detail here, the selectate was derived from dissociated embryo cells that express GFP and which also express the *brachyury* gene, as described above. Driver was synthesized from normal embryo cells of the same stage. Approximately 2 μg of radiolabeled probe was prepared from both unsubtracted and subtracted cDNA populations and hybridized in series to the same set of library filters. In this particular experiment, most mRNA species are expected to be present at similar levels in both selectate and driver populations. Because the filters are hybridized with the same total mass of pre- or postsubtraction probe, the majority of spot intensities should remain similar after subtractive hybridization, i.e., their individual sequence concentrations will have been boosted back to near their original levels after subtraction, while a minority of clones, which represent mRNAs that were not represented in the normal embryo cells, are expected to exhibit increased intensities when hybridized with the subtracted probe (see Eq. [1]). The combination of subtraction with reamplification for probe generation, and the use of macroarray screening, makes it possible to identify differences in minor components of the cDNA, i.e., in relatively rare mRNA species. Furthermore, because there is no need to extensively purify differential cDNA, this can be done without removing all of the common sequence and without

Copyright © 2000 by Academic Press. All rights of reproduction in any form reserved.

the drastic skewing of sequence prevalence that such purification often entails.

To estimate the subtraction efficiency for the common mRNAs and, conversely, the relative enrichment of species present only in selectate, all spot intensity readings were normalized to GFP control spots. Results of this experiment are shown in Fig. 4 (see legend for details). Messenger RNA derived from the transgenic GFP expression construct is present only in the selectate population and will not be specifically depleted by the subtraction procedure, since there are no complements in the driver RNA with which it can hybridize. The depletion of each common probe cDNA species identified by hybridization to a colony spot pair can thereby be calculated by comparison to GFP. Furthermore, any small amount of skewing that occurred during the few PCR cycles needed to initiate probe synthesis after subtraction (cf. Fig. 3) will be normalized out. That is, the pre- and postsubtraction probes being compared derive from the same selectate.

Depletion of the majority of message species in the *brachyury* overexpression subtraction is illustrated in Fig. 4. In Figs. 4A and 4B, a portion of a hybridized macroarray filter is shown in which the intensity of a central GFP spot pair has been held at constant gray scale to facilitate comparison of presubtraction (Fig. 4A) and postsubtraction (Fig. 4B) hybridization images. The extent to which most of the surrounding spots exhibit reduced hybridization intensity in Fig. 4B represents the subtractive depletion of their sequence concentrations, relative to that of GFP. As expected, the vast majority of transcript species are depleted during subtractive hybridization.

Analysis of total transcript population after subtraction. A quantitative analysis of all clone spots on a single macroarray filter for which hybridization intensity exceeded 3 standard deviations above mean background is shown in Fig. 4C. This is a limit criterion for detectability, equivalent to 1.25× background, and it resulted in detection of a total of 6477/18,432, or 35% of all clones. In Fig. 4C fractional depletion (ordinate) is plotted against the amount of signal from the subtracted probe (abscissa). The intensity of hybridization along the abscissa increases with sequence prevalence after subtraction. Fractional depletion is here defined as 1 minus the ratio of spot pair intensities after to before subtraction, for any given spot pair (once these values have been normalized to the average GFP spot pair intensities after and before subtraction; see legend to Fig. 4). A fractional depletion of 1 would correspond to complete removal of the probe sequence species upon subtraction, and a fractional depletion of 0 indicates that the spot analyzed remains at the same intensity relative to GFP after subtraction. Intensity and depletion readings for the 10 control GFP spot pairs on the filter are plotted as green circles. The average depletion of these 10 clones was placed at 0, by definition. The 10 pairs of GFP spots were in fact driven from a moderate intensity in the unsubtracted probe hybridization to a hybridization level that is among the most intense on the filter after probe subtraction. Most of the clones analyzed in Fig. 4 are depleted relative to GFP by a fraction of about 0.75 or, conversely, the GFP control is enriched fourfold. Presumably, sequence representation from other differentially expressed genes is enriched similarly and will be shifted down and to the right from the main body of clone spots. One such gene for which we have *a priori* information is *brachyury* itself. This gene is expressed at relatively low levels at 15 h and is undetectable on filters hybridized with driver probe, but is readily detectable with probe made from the transgenic cells. A single *brachyury* clone is present on the filter analyzed in Fig. 4C (shown as a yellow spot) and, as expected, it behaves similarly to the GFP control spots. Three clones derived from the *endo16* gene, expressed in the endomesodermal region of the embryo at this stage, are similarly undepleted (shown as red circles in Fig. 4C). This gene is also apparently upregulated in *brachyury*-expressing cells.

Clones from the depleted region of the graph were also analyzed. Eleven polyubiquitin clones whose identities were confirmed by sequencing are plotted as blue circles. A consistent level of depletion is exhibited by each of these clones and the range of presubtraction intensity is well within what is expected to result from insert size variation among clones. For example, the insert sizes of 15 ubiquitin clones taken from a second filter (hybridized simultaneously with the filter analyzed in Fig. 4) ranged from 600 to 2200 bp. Hybridization intensities correspond roughly to insert size for these clones ($R^2 = 0.67$). When 8 clones were arbitrarily picked from within the cluster encompassed by the blue box (Fig. 4C) and sequenced, all were found to represent the product of a mitochondrial gene (cytochrome *c* oxidase subunit I). All of these clones are depleted similarly. A second cluster (highlighted by an orange oval in Fig. 4C) consists mainly of ribosomal RNA clones. These were not as efficiently depleted in the hybridization as the majority, but they can easily be eliminated from consideration by avoiding clones within this cluster or by previously identifying ribosomal RNA clones. Clones bearing ribosomal RNA inserts generally make up less than 1% of these sea urchin embryo libraries, which are generated by random priming of purified poly(A) RNA.

cDNA clones that exhibit interesting hybridization ratios can be separated on the basis of their depletion values. The dotted demarcation line in Fig. 4C was arbitrarily set to delineate a manageable number of clones. Clones for which depletion values were below this line were chosen for further analysis. In this experiment, 1 to 2% of the clones represented above background fall below this line for each macroarray filter analyzed. Figure 5 shows the spot pattern for hybridization of three clones chosen in this way from two separate differential screening experiments. In each case the highlighted clone is undetectable on the filter hybridized with driver probe (left column), is marginally detectable with the unsubtracted selectate probe (right column), and is clearly enhanced relative to surrounding clones in the subtracted probe (middle column). Images in the top two rows are taken from the ectopic *brachyury*

Copyright © 2000 by Academic Press. All rights of reproduction in any form reserved.

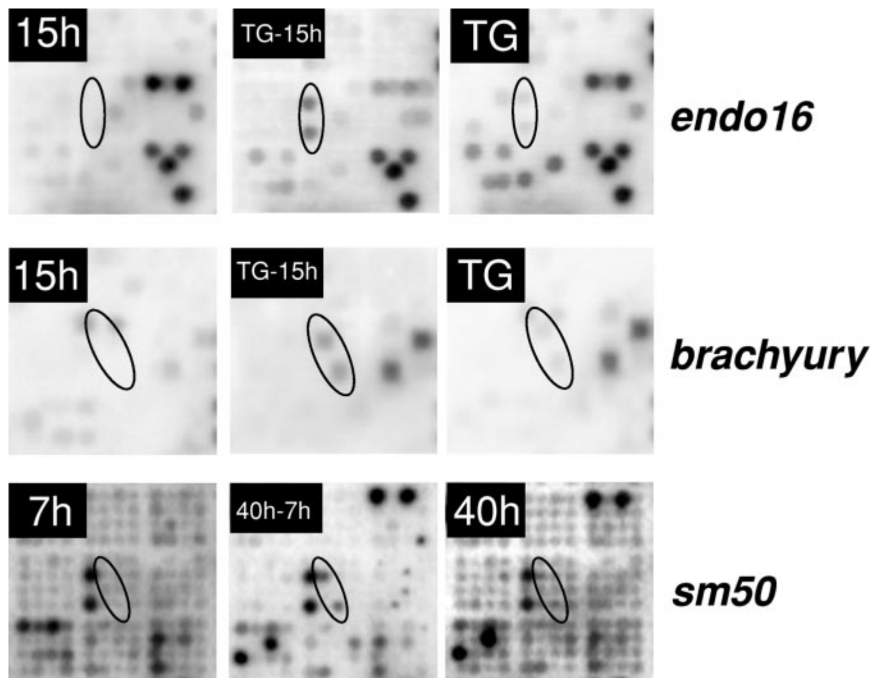

FIG. 5. Enhanced hybridization to colony spots representing genes that are differentially expressed. All images are from a 40-h embryo cDNA library. Spot pairs representing three differentially expressed genes (*endo16, brachyury,* and *sm50*) are shown within ellipses in each image. Each filter region has been hybridized with three complex probes. Driver probe hybridization is shown in the left column, selectate probe hybridization is shown in the right column, and subtracted probe hybridization is shown at the center. Top row: The *endo16* gene encodes a secreted protein that is first expressed in the endomesodermal region of the blastula. Signal from this spot cannot be seen in the 15-h driver hybridization (15h), but is clearly evident in the hybridization with the selectate probe (TG) made from the transgenic embryo cells which express the *brachyury* gene. The signal is enhanced relative to surrounding clones after subtraction (TG-15h). Middle row: the *brachyury* gene is expressed at levels that are undetectable by filter hybridization at 15 h and is overexpressed in the transgenic cells (TG). Hybridization is enhanced approximately sixfold after probe subtraction (TG-15h). Bottom row: The *sm50* gene encodes a structural protein involved in skeletogenesis. No hybridization is evident with a probe derived from 7-h embryo mRNA (7h). At this time the *sm50* gene is not yet transcribed. When hybridized with a probe derived from 40-h mRNA (40h), the *sm50* spots are just visible above background. At 40 h *sm50* is expressed at 30–40 copies per average cell (Killian and Wilt, 1987; Lee *et al.,* 1999b) and this prevalence level is just at the detection limit that we have determined for complex hybridization without differential sequence enhancement. *sm50* spots are clearly visible above background when hybridized with the subtracted probe (40h-7h).

expression experiment. The *endo16* transcript provides the first example and *brachyury* transcript the next in Fig. 5. In the hybridization with the subtracted probe, *brachyury* spot hybridization intensity is enhanced relative to surrounding clones and is not depleted relative to the GFP control.

Enhancement of a clone initially borderline in signal strength to a level well past the threshold of detectability is illustrated in the bottom row of Fig. 5. These images derive from a subtractive hybridization experiment comparing the normal patterns of gene expression in embryos at 7 and 40 h after fertilization. The spot pair shown represents *sm50,* a skeletal matrix protein gene that is not expressed in the 7-h embryo, but is expressed at approximately 40 copies per average embryo cell at 40 h of development (Lee *et al.,* 1999b; Killian and Wilt, 1989). This is near the detection limit of this type of screen without subtraction, and the *sm50* spot pair is just barely visible above background before subtraction. As expected, hybridization to the *sm50* clone spots is readily visible and is now enhanced relative to surrounding clones by about sixfold after reaction with the subtractive probe.

As an independent assessment of transcript prevalence from genes that were identified in this differential screen, we employed real-time quantitative PCR. This was used to determine the relative prevalence of transcripts isolated from 18-h embryos that ectopically express *brachyury* and from 15-h normal embryos, the sources of mRNA that were used to synthesize selectate and driver, respectively. These two embryo populations are essentially at the same stage (as a result of the slight developmental delay induced by DNA injection). Here we are simply interested in whether the genes that we have identified by differential screening are indeed expressed at higher levels in the selectate source, i.e., after forced *brachyury* expression. Table 3 shows the

Copyright © 2000 by Academic Press. All rights of reproduction in any form reserved.

TABLE 3

Quantitative PCR Analysis of a Subset of Genes Identified in the *brachyury* Expression Screen as Candidates for Upregulation

Gene	15-h control C_T[a]	18-h transgenic C_T	Ubiq. corr. ΔC_T[b]	Fold increase $(1.9^{\Delta CT})$[c]
Total *brachyury*[d]	27.57	18.46	9.46	433
Fascin	23.58	22.86	1.07	2.00
D-dopachrome tautomerase/MIF-like	28.80	26.24	2.91	7.51
Apolipophorin-like	27.30	23.37	4.28	15.6
Polyketide synthase-like	31.96	27.10	5.21	28.3
Sulfotransferase-like	31.34	30.08	1.61	2.81
endo16	30.44	26.38	4.41	16.96
PolyA binding-like	29.60	28.46	1.49	2.60
Ubiquitin	15.28	15.63	0	—

[a] C_T is the cycle number at which normalized product fluorescence exceeds a threshold set within the exponential range of the amplification reaction, as determined for each primer set.

[b] ΔC_T is the difference in C_T values between the control and the transgenic embryos after correction for the difference in C_T in ubiquitin amplifications. Ubiq. corr. ΔC_T is C_T after correction for differences in ubiquitin cycle number in the two samples.

[c] Fold increase is calculated as the per-cycle amplification efficiency (1.9) raised to the power ΔC_T. 1.9 is used as a conservative amplification efficiency estimate. Actual values lie between 1.9 and 2.0.

[d] Total *brachyury* is endogenous + transgenic expression.

results of a PCR analysis for a number of candidate genes and for *brachyury* itself. Measurements of ubiquitin expression levels were used to normalize the two mRNA sources. It is clear that these transcripts are present at higher prevalence in the transgenic embryo mRNA relative to that of the driver source. It is important to keep in mind that unlike the sample in the initial subtraction experiment, in which a pure population of sorted transgenic cells was used to synthesize the selectate, the embryos analyzed in this experiment were a mosaic of transgenic and wild-type cells, and only about 10% of the cells express the transgene. To effect a doubling of endogenous gene expression, the transgenic cells must upregulate expression of the genes being analyzed by up to approximately 10-fold (depending on the size of the endogenous population of cells that expresses the gene). Three of the clones in Table 3, representing ubiquitous "housekeeping" proteins (fascin, poly(A) binding protein, and a sulfotransferase) may not be significantly more highly represented in this experiment after forced *brachyury* expression. The remaining five clones were represented by much more prevalent transcripts after *brachyury* overexpression. In additional studies to be reported elsewhere we have isolated a large number of putative *brachyury* targets by this method and verified their enrichment in RNA samples from embryos ectopically expressing *brachyury.*

Enhancement of differential depletion by hydroxylapatite chromatography. While the previous experiments clearly demonstrate the utility of the phenol:chloroform separation method, the potential of subtractive nucleic acid hybridization for isolating sequence that is unique to a specific transcriptional state obviously exceeds the enrichment that we achieved. A problematic aspect of the phenol: chloroform procedure used for the quantitative experiment of Fig. 4 is the nonspecific loss of single-stranded selectate during the extraction. We therefore explored hydroxylapatite chromatography in combination with high C_0t hybridizations, as an alternative strategy for isolation of unhybridized selectate (outlined in Fig. 3C). Hybridization reactions were set up essentially as described above and incubated for 40 h. Enrichment of sequence that was unique to the selectate population was assessed by measuring the change in relative concentration of specific DNA fragments before and after hydroxylapatite chromatography. The driver and selectate were derived from similar but not identical sea urchin embryo RNA populations. The behavior of a single-stranded fragment present only in the selectate (λ1) was compared to that of a similar fragment (λ2) whose complementary sequence was present also as RNA in the driver. Both fragments contained the same primer sequences as did the actual selectate and driver and were added at very low concentrations, equivalent to approximately 10 copies per average cell. The subtractive hybridization mixture included embryonic selectate (200 ng) and driver (10 μg) in a 20-μl volume. We assessed the enrichment of the λ1 control relative to both the λ2 control fragment and the endogenous ubiquitin sequence. Ubiquitin and λ2 were present in similar concentrations in both selectate and driver. They are therefore expected to be depleted upon subtraction. Quantitative PCR was used to obtain these measurements. The experiment showed that after HAP column removal of hybridized selectate, λ1 sequence concentration increases 24.8-fold, relative to that of λ2, and 23.5-fold relative to that of ubiquitin. This is a vast improvement over the 6- to 10-fold enrichments that we have always obtained in similar experiments using the phenol:streptavidin method. Furthermore, this high level of enrichment is achieved in a single round of amplification.

Copyright © 2000 by Academic Press. All rights of reproduction in any form reserved.

DISCUSSION

Macroarray filters provide potentially complete representation of gene expression patterns without the need for prior sequence information, but screening them is inherently insensitive, because of the large volumes required. Using the largest easily practical amount of highly radioactive probe, we found that only about the top 25–30% of the clones on a typical sea urchin embryo macroarray can be reliably detected under optimal hybridization conditions. These are the clones representing mRNAs at the highest prevalence in the probe, i.e., sequences present at ≥40 copies/average embryo cell (in other words a prevalence of $\sim 4 \times 10^{-4}$ in the overall mRNA population at late blastula/gastrula stage). We showed that this limitation of sensitivity is entirely kinetic; that is, the amount of hybridization, considered as the fraction of the clone DNA per spot hybridized, depends completely on the probe C_0t for that sequence. Therefore the key to increased sensitivity is increased sequence concentration. For differentially expressed genes this can be attained by subtractive hybridization, combined with amplification of what initially may be very small complements of starting RNA from experimentally treated embryos. Amplification presents a further challenge in that skewing of the RNA frequency distribution in the driver used for subtractive hybridization could render the whole procedure useless. Here we show that a combination of total message amplification and subtractive enrichment of differential sequence can be used to isolate relatively rare message species by macroarray screening.

Amplification

We have designed these experiments to be able to process RNAs extracted from the small number of embryos that can be manually injected. In practice this means we would like to be able to prepare driver nucleic acids from as little as 10 ng of starting poly(A) RNA. To obtain high driver C_0t in the subtraction reactions, or even to make a direct probe for macroarray screening, 5- to 10-μg quantities of the amplified RNA are desirable; i.e., a 300–1000× amplification without significant skewing is required. We showed that concern about the use of PCR alone for this is indeed warranted; in the exponential PCR process small differences in efficiency of amplification are magnified and the emergent population looks quite different from the starting population. There are limits to PCR skewing in that few sequences are either completely lost or artificially introduced, but amplification in the preparation of a kinetically faithful driver by PCR alone is clearly inadmissible. Nonetheless, because the qualitative nature of the cDNA population survives, PCR amplification can be used for the preparation of a subtractive hybridization selectate. In the procedure summarized in Fig. 3, we have applied the linear amplification afforded by T7 polymerase (Phillips and Eberwine, 1996; Kacharmina *et al.,* 1999) by incorporating T7 polymerase primer sites in the initial stages of cDNA preparation. Only a few cycles of PCR are used. Essentially these act to repair the ragged products of first- and second-strand synthesis and provide just enough template for the T7 amplification to efficiently go to work on. We consider this aspect of the technology now sufficient for the task, and in fact the same procedure could probably be modified to deal with far smaller quantities of starting material if need be.

Generation of the Subtracted Probe

Using two rounds of driver challenge, without amplification between subtractive hybridization reactions, we show that the phenol:chloroform extraction method can achieve a significant enrichment in the probe of differentially expressed sequences. A meaningful prevalence distribution is maintained by the amplification and hybridization procedure described. In a number of experiments that we have performed using this technology, enrichment of unique sequence ranged from four- to sevenfold, based on measurements similar to those in Figs. 4 and 5 and Table 3 of this paper. This level of enrichment suffices to identify many differentially expressed genes. However, substitution of the hydroxylapatite separation method for the phenol:chloroform method affords a striking further increase in enrichment of differentially expressed sequences, 3–4× better in a single cycle than the phenol method in two. A strength of the procedure, irrespective of the hybrid separation method, is that only strand-specific components are used for subtraction, so that the reactions conform to pseudo-first-order kinetics, which are sharper than second-order kinetics. Thereby competing second-order reactions are avoided. The resulting prevalence depletion for any particular cDNA species depends directly on driver sequence concentration. Following the subtraction, the surviving selectate is amplified, again using T7 polymerase, to generate the large mass (1–2 μg) of highly radioactive probe needed for macroarray screening. The power of the combination of subtractive hybridization and macroarray screening is that nothing can hide from it: minor transcripts represented in the library, that would be very difficult to recover by procedures that require differential PCR product cloning or sequence analysis following subtraction, will be identified on the macroarrays if they are differentially expressed. Conversely, using macroarrays, it is unnecessary to deplete common sequences extensively before they can be distinguished from specific ones. This greatly reduces the risk of artificial loss of differentially expressed sequences through PCR skewing or other artifacts. Our present procedures extend macroarray screening sensitivity (for differentially utilized genes) from transcripts expressed at approximately ≥40 copies down to fewer than 5 copies per average sea urchin embryo cell.

Improved computational methods for interpreting macroarray hybridization results will further improve the accuracy and sensitivity with which we can identify differentially expressed genes. This is clear from the scatter plot in

Copyright © 2000 by Academic Press. All rights of reproduction in any form reserved.

Fig. 4. Much of the variance in depletion values for the lower intensity clones shown in Fig. 4C probably results from the large error in the determination of spot hybridization intensity as these values approach background levels in the depleted screen. More accurate estimates of bound probe will reduce the number of false-positive clones that need to be analyzed. The incidence of false positives may be further reduced by comparing multiple replicate filters, a process that will be facilitated by automated computational analysis. Finally, accuracy can be improved by screening procedures that lower the nonspecific background hybridization and thus the variance of signal determination.

CONCLUSIONS

Here we show that careful application of cDNA amplification in combination with subtractive hybridization can be used to bypass the inherent sensitivity limitations of macroarray filter hybridization. The differential enhancement achieved in the present study is sufficient to drive macroarray screening sensitivities to below five copies per average cell in the developing sea urchin embryo. The technology continues to improve but even as it now stands, the approach described here has shown the power of macroarray screening as a tool for gene discovery in development. We demonstrate this with a variety of standards and examples in this report, and forthcoming studies will illustrate the application of these methods in the recovery of components of several different gene networks operating in the *S. purpuratus* embryo. In principle these methods can be used to isolate any set of differentially expressed genes in any developing system from which RNA can be isolated.

ACKNOWLEDGMENTS

We are grateful to our colleagues Roy J. Britten at Caltech's Kerckhoff Marine Lab, Professor Ellen Rothenberg, and Dr. Michele Anderson for critical and perspicacious reviews of the manuscript. We thank Rochelle A. Diamond for guidance and help with the purification of transgenic cells by FACS and Xiaohong She and Miki Yun for technical assistance. This research was supported by the National Institute of Child Health and Human Development (HD-37105), the California Sea Grant College Program (NA65RG0477), the Beckman Institute of Caltech, and the Lucille P. Markey Charitable Trust. J.P.R. was supported by an NIH Individual NRSA (GM 18478). C.B.L. was supported by a Walter & Sylvia Treadway Fellowship.

REFERENCES

Arnone, M. I., Bogarad, L. D., Collazo, A., Kirchhamer, C. V., Cameron, R. A., Rast, J. P., Gregorians, A., and Davidson, E. H. (1997). Green fluorescent protein in the sea urchin: New experimental approaches to transcriptional regulatory analysis in embryos and larvae. *Development* **124,** 4649–4659.

Britten, R. J., Graham, D. E., and Neufeld, B. R. (1974). Analysis of repeating DNA sequences by reassociation. *Methods Enzymol.* **29,** 363–418.

Brown, P. O., and Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. *Nat. Genet.* **21,** 33–37.

Clark, M. D., Panopoulou, G. D., Cahill, D. J., Bussow, K., and Lehrach, H. (1999). Construction and analysis of arrayed cDNA libraries. *Methods Enzymol.* **303,** 205–233.

Davidson, E. H. (1986). "Gene Activity in Early Development." Academic Press, Orlando, FL.

Flytzanis, C. N., Brandhorst, B. P., Britten, R. J., and Davidson, E. H. (1982). Developmental patterns of cytoplasmic transcript prevalence in sea urchin embryos. *Dev. Biol.* **91,** 27–35.

Galau, G. A., Britten, R. J., and Davidson, E. H. (1977). Studies on nucleic acid reassociation kinetics: Rate of hybridization of excess RNA with DNA, compared to the rate of DNA renaturation. *Proc. Natl. Acad. Sci. USA* **74,** 1020–1023.

Gubler, U., and Hoffman, B. J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene* **25,** 263–269.

Hampson, I. N., Hampson, L., and Dexter, T. M. (1996). Directional random oligonucleotide primed (DROP) global amplification of cDNA: Its application to subtractive cDNA cloning. *Nucleic Acids Res.* **24,** 4832–4835.

Hough-Evans, B. R., Britten, R. J., and Davidson, E. H. (1988). Mosaic incorporation and regulated expression of an exogenous gene in the sea urchin embryo. *Dev. Biol.* **129,** 198–208.

Kacharmina, J. E., Crino, P. B., and Eberwine, J. (1999). Preparation of cDNA from single cells and subcellular regions. *Methods Enzymol.* **303,** 3–18.

Killian, C. E., and Wilt, F. H. (1989). The accumulation and translation of a spicule matrix protein mRNA during sea urchin embryo development. *Dev. Biol.* **133,** 148–156.

Kohne, D. E., Levison, S. A., and Byers, M. J. (1977). Room temperature method for increasing the rate of DNA reassociation by many thousandfold: The phenol emulsion reassociation technique. *Biochemistry* **16,** 5329–5341.

Lasky, L. A., Lev, Z., Xin, J. H., Britten, R. J., and Davidson, E. H. (1980). Messenger RNA prevalence in sea urchin embryos measured with cloned cDNAs. *Proc. Natl. Acad. Sci. USA* **77,** 5317–5321.

Lee, J. J., Calzone, F. J., Britten, R. J., Angerer, R. C., and Davidson, E. H. (1986). Activation of sea urchin actin genes during embryogenesis. Measurement of transcript accumulation from five different genes in *Strongylocentrotus purpuratus. J. Mol. Biol.* **188,** 173–183.

Lee, Y. H., Huang, G. M., Cameron, R. A., Graham, G., Davidson, E. H., Hood, L., and Britten, R. J. (1999a). EST analysis of gene expression in early cleavage-stage sea urchin embryos. *Development* **126,** 3857–3867.

Lee, Y. H., Britten, R. J., and Davidson, E. H. (1999b). *SM37,* a skeletogenic gene of the sea urchin embryo linked to the *SM50* gene. *Dev. Growth Differ.* **41,** 303–312.

Lipshutz, R. J., Fodor, S. P., Gingeras, T. R., and Lockhart, D. J. (1999). High density synthetic oligonucleotide arrays. *Nat. Genet.* **21,** 20–24.

Livant, D. L., Hough-Evans, B. R., Moore, J. G., Britten, R. J., and Davidson, E. H. (1991). Differential stability of expression of similarly specified endogenous and exogenous genes in the sea urchin embryo. *Development* **113,** 385–398.

Maier, E., Meier-Ewert, S., Ahmadi, A. R., Curtis, J., and Lehrach, H. (1994). Application of robotic technology to automated se-

Copyright © 2000 by Academic Press. All rights of reproduction in any form reserved.

quence fingerprint analysis by oligonucleotide hybridisation. *J. Biotechnol.* **35,** 191–203.
McClay, D. R. (1986). Embryo dissociation, cell isolation and cell reassociation. *Methods Cell Biol.* **27,** 309–323.
McMahon, A. P., Flytzanis, C. N., Hough-Evans, B. R., Katula, K. S., Britten, R. J., and Davidson, E. H. (1985). Introduction of cloned DNA into sea urchin egg cytoplasm: Replication and persistence during embryogenesis. *Dev. Biol.* **108,** 420–430.
Nemer, M., Rondinelli, E., Infante, D., and Infante, A. A. (1991). Polyubiquitin RNA characteristics and conditional induction in sea urchin embryos. *Dev. Biol.* **145,** 255–265.
Nizetic, D., Drmanac, R., and Lehrach, H. (1991). An improved bacterial colony lysis procedure enables direct DNA hybridisation using short (10, 11 bases) oligonucleotides to cosmids. *Nucleic Acids Res.* **19,** 182.
Peterson, K. J., Harada, Y., Cameron, R. A., and Davidson, E. H. (1999). Expression pattern of *brachyury* and *Not* in the sea urchin: Comparative implications for the origins of mesoderm in the basal deuterostomes. *Dev. Biol.* **207,** 419–431.
Phillips, J., and Eberwine, J. H. (1996). Antisense RNA amplification: A linear amplification method for analyzing the mRNA population from single living cells. *Methods: Companion Methods Enzymol.* **10,** 283–288.
Poustka, A. J., Herwig, R., Krause, A., Hennig, S., Meier-Ewert, S., and Lehrach, H. (1999). Toward the gene catalogue of sea urchin development: The construction and analysis of an unfertilized egg cDNA library highly normalized by oligonucleotide fingerprinting. *Genomics* **59,** 122–133.
Sagerstrom, C. G., Sun, B. I., and Sive, H. L. (1997). Subtractive cloning: Past, present, and future. *Annu. Rev. Biochem.* **66,** 751–783.
Siebert, P. D., Chenchik, A., Kellogg, D. E., Lukyanov, K. A., and Lukyanov, S. A. (1995). An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Res.* **23,** 1087–1088.
Sive, H. L., and St John, T. (1988). A simple subtractive hybridization technique employing photoactivatable biotin and phenol extraction. *Nucleic Acids Res.* **16,** 10937.

Received for publication August 24, 2000
Revised September 14, 2000
Accepted September 14, 2000
Published online November 17, 2000

Copyright © 2000 by Academic Press. All rights of reproduction in any form reserved.