

# Production of complex nucleic acid libraries using highly parallel *in situ* oligonucleotide synthesis

Michele A Cleary<sup>1</sup>, Kristopher Kilian<sup>1</sup>, Yanqun Wang<sup>1</sup>, Jeff Bradshaw<sup>1</sup>, Guy Cavet<sup>1</sup>, Wei Ge<sup>1</sup>, Amit Kulkarni<sup>1</sup>, Patrick J Paddison<sup>2</sup>, Kenneth Chang<sup>2</sup>, Nihar Sheth<sup>2</sup>, Eric Leproust<sup>3</sup>, Ernest M Coffey<sup>1</sup>, Julja Burchard<sup>1</sup>, W Richard McCombie<sup>2</sup>, Peter Linsley<sup>1</sup> & Gregory J Hannon<sup>2</sup>

**Generation of complex libraries of defined nucleic acid sequences can greatly aid the functional analysis of protein and gene function. Previously, such studies relied either on individually synthesized oligonucleotides or on cellular nucleic acids as the starting material. As each method has disadvantages, we have developed a rapid and cost-effective alternative for construction of small-fragment DNA libraries of defined sequences. This approach uses *in situ* microarray DNA synthesis for generation of complex oligonucleotide populations. These populations can be recovered and either used directly or immortalized by cloning. From a single microarray, a library containing thousands of unique sequences can be generated. As an example of the potential applications of this technology, we have tested the approach for the production of plasmids encoding short hairpin RNAs (shRNAs) targeting numerous human and mouse genes. We achieved high-fidelity clone retrieval with a uniform representation of intended library sequences.**

**np** Nucleic acid libraries provide some of the most versatile tools for functional analysis of genomes, individual proteins or complexes<sup>1–14</sup>. These libraries can be constructed using either biologically derived or chemically synthesized nucleic acids as substrates. Libraries generated from natural sources generally do not cover all expressed sequences in the genome, largely owing to tissue-specific mRNA expression and variations in mRNA abundance, limiting the complexity and uniform representation of the cDNA source material. Chemically synthesized oligonucleotides have also been used to construct libraries for biological analysis<sup>14</sup>. Although these allow defined and uniform representation, the cost of source material for library construction is quite high.

To reduce the cost inherent in the use of conventional methods for the generation of complex libraries of defined nucleic acids, we have developed an approach that uses printed microarrays as a source material for complex oligonucleotide populations<sup>15–18</sup> (Fig. 1). Although such an approach can be applied in many different ways, we have tested the methodology for one specific

application, namely for the construction of libraries of shRNA expression constructs.

## RESULTS

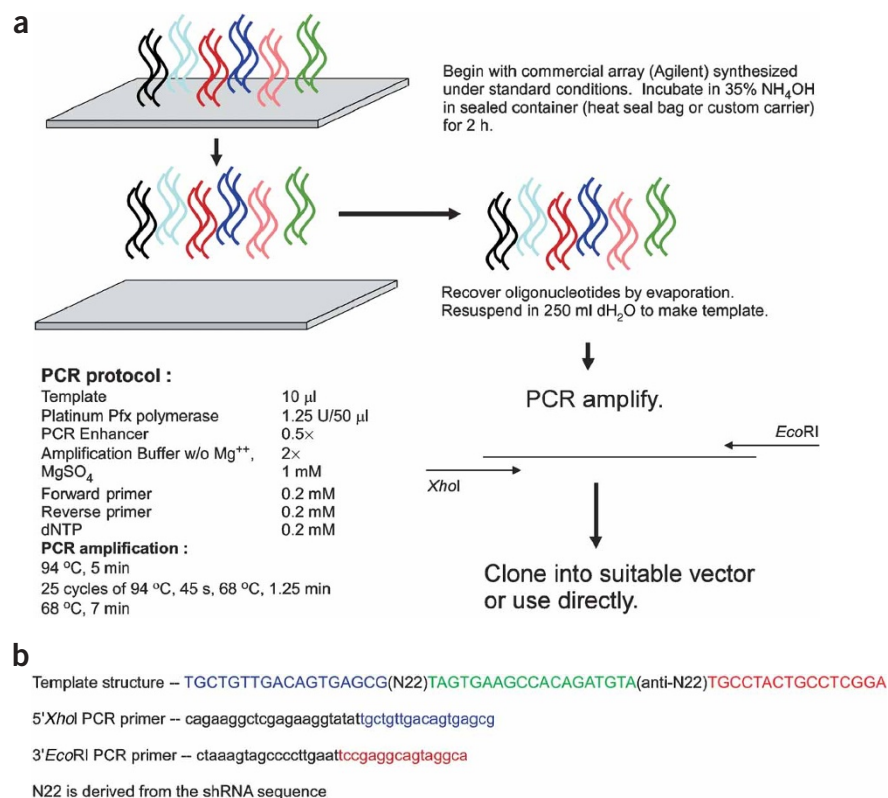
### Ink-jet synthesis of oligonucleotides

Ink-jet technology has been optimized for hybridization microarrays using oligonucleotides of 60 bases or less on slides that contain ~25,000 individual spots<sup>15–18</sup>. But no tests have suggested whether this method produced DNA of sufficient quality or quantity for use as source material for library construction. To address this question, we designed and printed arrays containing 110 unique 59-nucleotide (nt) DNA sequences, each containing identical flanking PCR primer binding sites. Initially, each oligonucleotide was synthesized redundantly in ~220 different locations on an array containing 24,200 probes to give an overall complexity of 1,000 different sequences. Oligonucleotide populations were recovered from the microarray surface using one of two approaches. The first, simpler approach involved treatment of standard arrays with ammonium hydroxide<sup>19</sup> (Fig. 1a). The second approach required derivatizing slides with a photocleavable linker before synthesis, and the oligonucleotides were ultimately recovered after a brief treatment with UV light. After harvesting the oligonucleotides, we amplified the pooled material by PCR and cloned the products. Of the clones obtained from ammonium hydroxide cleaved material, five of five readable sequences were of the correct length, exactly matched one of the sequences in the array pattern design, and were unique. Of the clones obtained from photocleaved material, four of five readable sequences had the correct length and each perfectly matched a unique sequence in the array pattern design. These results suggested that the use of this highly parallel synthesis approach was feasible for producing clones with ~60-base-pair (bp) inserts.

### Accurate synthesis of long oligonucleotides

The ability to produce complex libraries comprised of defined 60-nt fragments is sufficient for many purposes, and such arrays can be purchased as a standard product from Agilent Technologies, the

<sup>1</sup>Rosetta Inpharmatics LLC, a wholly owned subsidiary of Merck & Co., Inc., 401 Terry Ave. North, Seattle, Washington 98109, USA. <sup>2</sup>Cold Spring Harbor Laboratory, Watson School of Biological Sciences, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. <sup>3</sup>Agilent Technologies, 3500 Deer Creek Road, Palo Alto, California 94304, USA. Correspondence should be addressed to M.A.C. (Michele\_Cleary@merck.com) and G.J.H. (hannon@cshl.edu).



**Figure 1** | Cloning strategy using *in situ* oligonucleotide synthesis. To create a pool of sequences for library cloning, oligonucleotides were printed on a microarray substrate, cleaved by treatment with a strong base or ultraviolet light and amplified by PCR. The amplified products were treated with restriction enzymes or used directly for ligation as a pool into a vector of choice.

commercial source of arrays used in this report. However, some specialized applications may require longer oligonucleotides. For our purposes, the design of optimized shRNA libraries requires synthesis of oligonucleotides ~100 bases in length. Although these are not a standard product, the use of 100-mers provided a very stringent test of the array methodology for library production. We designed arrays containing 96-nt sequences deposited either once per array or at variable representation, ranging from 1 to 1,024 times. With PCR products derived from ammonium hydroxide-cleaved material, we found an average of ~63% of clones (total of 30 in three separate cloning trials) with the correct sequence and length (Fig. 1b, primer and template structure). From arrays printed at variable oligonucleotide representation, we recovered an overwhelming majority of accurate clones corresponding to the sequence spotted 1,024 times. We could not clone from photo-cleaved 96-nt material.

The use of RNA interference (RNAi) has opened the door for loss-of-function genetic approaches in numerous organisms, including mammals<sup>20</sup>. One method to achieve RNAi is the expression of shRNAs from DNA vectors<sup>21–27</sup>. We therefore set out to use *in situ*-synthesized sequences to build shRNA expression libraries targeting nearly every identified and predicted gene in the genomes of several species, including human, mouse and rat. Similar libraries have previously been constructed using conventional oligonucleotides or natural nucleic acids as starting material<sup>28–33</sup>. To maximize recovery of accurate clones from our highly structured

templates, we used thermostable polymerases that have proofreading capability and are able to effect strand displacement. We also added PCR enhancing agents such as DMSO or betaine. Through a combination of these strategies (Fig. 1 and Methods), we were able to achieve success rates consistently ranging from ~25% to >60% for cloning of perfect shRNAs.

### Construction of shRNA libraries

In our effort to create large-scale human and mouse shRNA libraries, we designed oligonucleotides corresponding to more than 32,000 known and predicted genes each in human and mouse. These yield 195,077 oligonucleotides homologous to murine genes and 187,905 oligonucleotides homologous to human genes. Each oligonucleotide was synthesized once on each array, necessitating the use of a minimum of 21 arrays to completely cover genes in both organisms with up to six shRNAs each. Iterative cycles of sequencing and synthesis were used to maximize the efficiency of obtaining correct clones. Recovery of unique, perfect shRNA vectors from the population can be hampered by two types of errors. The first is inaccuracies in the synthesis, amplification or sequencing leading to inserts that are or appear inaccurate. The second is biases in the synthesis, amplification and cloning procedures leading to

imperfect representation of the desired oligonucleotide population in the cloned pools. We have examined each problem separately. Table 1 shows data relevant to the first type of error, comparing the accuracy of array synthesis and chemical synthesis. Thus far, we have sample-sequenced clones from 23 separate arrays covering a total of 447,410 printed sequences with 216,945 informative sequencing reads. An informative read is defined as a sequencing run that gives high-quality sequence (PHRED score >20 over the length of the insert). The rates of successfully obtaining perfect clones varied from 21% to 58%, depending on the synthesis run. The ~220,000 reads yielded 76,960 perfect clones overall, of which 53,478 represented unique sequences. We noted no bias for correct versus incorrect clones based on the oligonucleotide position on the arrays. For comparison, we obtained 7,360 oligonucleotides in six independent batches that were produced using conventional synthesis methods by a commercial manufacturer. From 18,554 informative sequencing reads, 3,526 perfect clones were obtained, with success rates from individual pools ranging from 9.9% to 26.5%. These pools represent the upper range of success rates with conventional oligonucleotides obtained from a number of different suppliers. As it is difficult to directly compare the quality of array synthesized material to conventionally synthesized material given differences in pool complexities, it seems reasonable to conclude that the array synthesized material, treated and cloned in the manner described herein, is of a quality that is at least equivalent to that of conventionally synthesized material.

**Table 1** | Characterization of cloned shRNAs

Source	Complexity	Reads	Correct	1 Mm	2 Mm	> 2 Mm	Success rate
<i>In situ</i> -synthesized oligonucleotides							
Chip 1	19,253	16,103	3,671	3,722	2,050	6,660	0.227
Chip 2	19,230	12,531	3,749	3,962	1,782	3,038	0.299
Chip 3	19,244	6,877	2,533	1,859	728	1,757	0.368
Chip 4	19,245	12,749	3,312	3,321	1,878	4,238	0.259
Chip 5	19,251	10,852	5,232	2,004	527	3,089	0.482
Chip 6	19,234	13,079	3,847	3,703	1,594	3,935	0.294
Chip 7	19,216	12,616	3,260	3,861	1,934	3,561	0.258
Chip 8	19,226	12,584	4,236	3,805	1,630	2,913	0.336
Chip 9	19,228	12,542	2,635	4,052	2,312	3,543	0.21
Chip 10	22,089	7,550	3,880	1,436	412	1,822	0.513
Chip 11	19,236	7,402	4,342	1,533	386	1,141	0.586
Chip 12	22,077	7,351	3,596	1,447	449	1,859	0.489
Chip 13	21,524	7,502	3,775	1,542	415	1,770	0.503
Chip 14	22,000	5,826	1,439	1,791	847	1,749	0.246
Chip 15	17,621	11,911	6,579	2,598	582	2,152	0.552
Chip 16	20,545	6,030	2,626	1,611	416	1,377	0.435
Chip 17	20,550	5,534	1,645	1,769	633	1,487	0.297
Chip 18	20,546	14,229	3,279	4,264	2,268	4,418	0.23
Chip 19	17,620	5,425	1,783	1,785	573	1,284	0.328
Chip 20	17,620	5,756	2,007	1,828	585	1,336	0.348
Chip 21	17,616	5,503	2,013	1,758	517	1,215	0.365
Chip 22	17,621	11,600	5,087	2,678	840	2,995	0.438
Chip 23	17,618	5,393	2,434	1,399	356	1,204	0.451
Conventionally synthesized oligonucleotides							
Pool A	576	1,827	396	240	120	1,071	0.216
Pool B	2,000	6,751	1,792	1,189	645	3,125	0.265
Pool C	2,000	2,726	180	302	173	2,071	0.066
Pool D	1,440	3,851	695	529	305	2,322	0.18
Pool E	768	1,510	150	183	159	1,018	0.099
Pool F	576	1,889	313	426	224	926	0.165

Oligonucleotides from the indicated sources were amplified by PCR and cloned into pSM2 (G.J.H., J. Silva, P.J.P., M.A.C., S. Elledge, D. Siolas *et al.*, data not shown) for sequencing. The complexity of each population is indicated. Informative reads (Reads) were scored for perfect clones (Correct) or for clones with a single mismatch (1 Mm), two mismatches (2 Mm) or more than two mismatches (> 2 Mm). An informative read is defined as a trace giving a sequence of sufficient quality for analysis (PHRED score > 20 over the length of the insert). The success rate is calculated as the fraction of perfect clones from the total informative reads.

**Table 2** presents data that tracks the second type of error, measuring the frequency with which we recovered individual oligonucleotide sequences as clones. To examine the data in the most consistent fashion, we examined the performance of each pool when the sampling by sequencing had reached  $0.5\times$ . We scored all identifiable clones, defined as those with a sequence with fewer than three mismatches to the target. Overall, pools made by both synthesis methods behaved similarly. Both sources of material yielded clone populations that matched slightly fewer oligonucleotides than was expected from a Poisson distribution, indicating that there were inherent biases in either the synthesis or the amplification of each oligonucleotide population. With conventionally synthesized material the rate at which cloned oligonucleotides were recovered in a nonredundant fashion varied from 34% to 68%, whereas with array-synthesized material this varied from 51% to 70%. At  $0.5\times$  sampling,  $\sim 78\%$  of reads were expected to represent unique oligonucleotides.

To examine a single population as an example, consider chip 15 (Tables 1 and 2). Of 11,911 informative reads, 6,579 perfectly matched printed oligonucleotide sequences, giving an accuracy rate

of 55.2%. To measure sampling error, we considered only the first 8,810 reads that unambiguously matched oligonucleotides printed on the array so that our sampling rate was normalized with other populations at  $\sim 0.5\times$ . Within those 8,810 reads, we expected that 6,933 printed oligonucleotide sequences would be represented (78% of 8,810). Instead, we found 4,780 printed sequences.

An examination of the melting temperature ( $T_m$ ) profile of the recovered, perfect shRNAs showed that it largely reflected the  $T_m$  profile of the total library oligonucleotide population, although there was a shift toward lower  $T_m$  for perfect clones (Fig. 2a). Similar results were obtained for conventional oligonucleotides (Fig. 2b). These results suggested that the PCR and cloning procedures used had a small preference for amplification of hairpins with lower thermal stability. The difference in  $T_m$  between the perfect and expected clones represents a shift corresponding to approximately two additional G-C base pairs. Furthermore, an examination of the error profile of the sequences suggested that there exists a bias for errors within the stem regions (Fig. 2c,d). This same bias was seen irrespective of the source of the oligonucleotides, with conventional and ink-jet samples giving similar results.

**Table 2** | Sampling of shRNA populations from Chip and conventional oligonucleotides

Source	Complexity	Reads	Sampling rate	Expected unique	Actual unique	Unique rate
<i>In situ</i> -synthesized oligonucleotides						
Chip 1	19,253	9,626	0.499	7,575	6,821	0.708
Chip 2	19,230	9,615	0.5	7,566	5,617	0.584
Chip 4	19,245	9,622	0.499	7,572	6,439	0.669
Chip 5	19,251	9,625	0.499	7,574	5,756	0.598
Chip 6	19,234	9,617	0.5	7,567	5,694	0.592
Chip 7	19,216	9,608	0.5	7,560	6,116	0.636
Chip 8	19,226	9,613	0.5	7,564	5,521	0.574
Chip 9	19,228	9,614	0.5	7,565	5,745	0.597
Chip 15	17,621	8,810	0.499	6,933	4,780	0.542
Chip 18	20,546	10,273	0.5	8,084	7,330	0.713
Chip 22	17,621	8,810	0.499	6,933	4,529	0.514
Conventionally synthesized oligonucleotides						
Pool A	576	288	0.5	226	197	0.684
Pool B	2,000	1,000	0.5	786	575	0.575
Pool C	2,000	1,000	0.5	786	489	0.489
Pool D	1,440	720	0.5	566	480	0.666
Pool E	768	384	0.5	302	235	0.611
Pool F	576	288	0.5	226	99	0.343

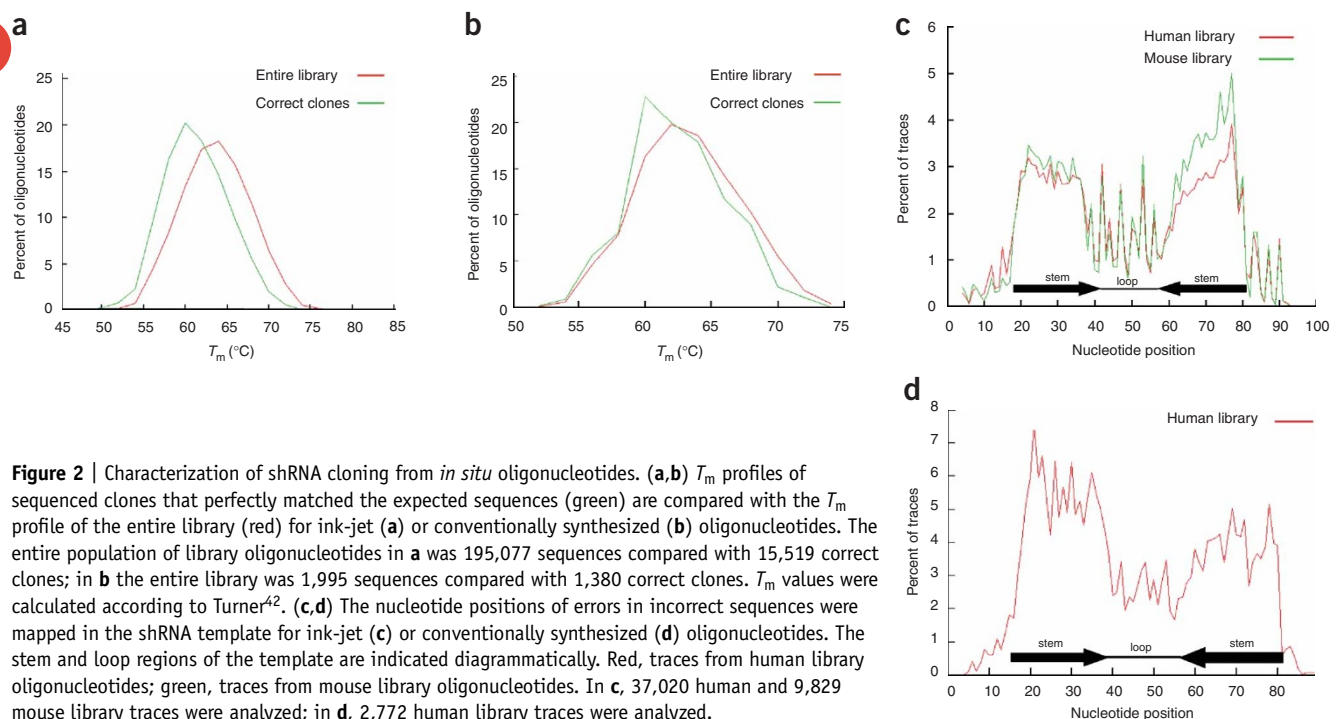
For all populations sampled at more than  $0.5\times$  coverage, we determined the redundancy rate in the informative reads. So that each population could be compared, we restricted our analysis to sequencing reads that comprised only  $0.5\times$  coverage. The actual and expected unique matches to printed oligonucleotides for each pool, according to a Poisson distribution, are given along with the rate at which unique matches are obtained.

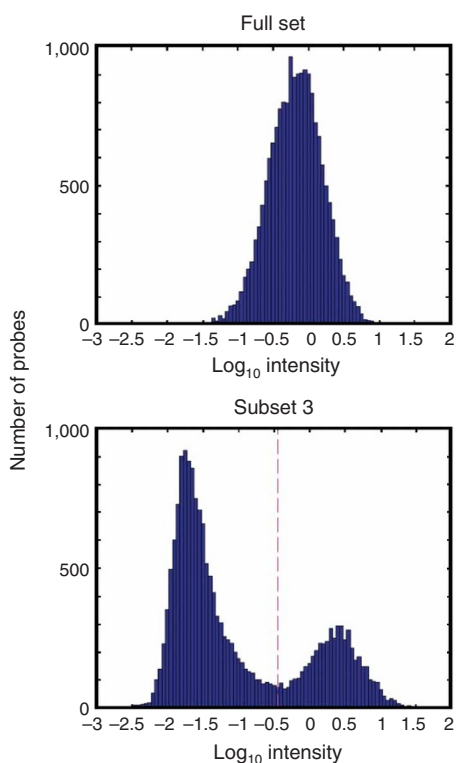
The peaks of errors that are observed within the loop region do not correspond to any regions of known structure. All represent adenine residues, however, potentially indicating some bias in the chemical synthesis procedure.

### Array-based assessments of synthesis bias

To assess the representation of the printed sequences in the amplified oligonucleotide pools, we used standard microarray

hybridization. We printed and cleaved a set of 18,723 unique 97-base oligonucleotides encoding shRNAs each spotted once on the array. We also designed four subset arrays, each containing 5,152 of the 18,723 sequences, with each subset overlapping the subsequent subset by  $\sim 600$  sequences. We used a T7 promoter-adapted PCR primer to amplify double-stranded templates for *in vitro* transcription (IVT), transcribed these templates in the presence of amino allyl UTP and coupled the resulting IVT

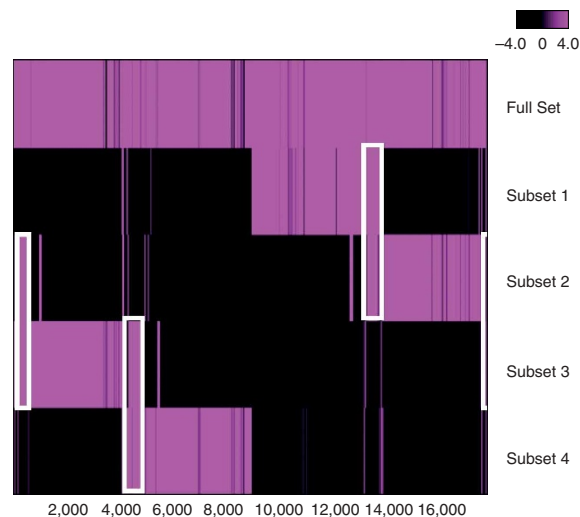




**Figure 3** | Histograms of the average intensity of the 18,723 probes when hybridized to IVT products derived from the pool of the full-set of sequences (top) and one representative subset of 5,152 sequences (bottom). Subset arrays 1, 2 and 4 showed similar bimodal distributions.

products to Cy3 and Cy5 dyes. After coupling, we hybridized dye-labeled material to a ‘diagnostic’ microarray containing 60-mer probes of all 18,723 sequences, along with controls. To minimize cross-hybridization, we eliminated the common primer binding sites from the oligonucleotides on the diagnostic array. In these shRNAs, up to three G-C base pairs in the stems were converted to encode G-U base pairs in the expressed shRNAs<sup>28</sup>. This approach alleviates secondary structure at the DNA level and increases stability during amplification, cloning and propagation in bacteria. Newer shRNA designs, such as those used for the sequence analysis of shRNA populations described above, do not incorporate this strategy, but the inclusion of G-U mismatches in the stem region should have no impact on the relative degree to which cleaved populations represent the total pool of synthesized material.

We observed a single-mode distribution of hybridizing probes (high and low intensity) on the diagnostic microarray for the full-set pool and, as expected, bimodal distributions for the subset pools (Fig. 3). After subtraction of background hybridization using negative controls on the microarray, the distributions were segmented to estimate the probes with intensity above background as follows. For hybridization to the subset pools, we used the data from the subset detection arrays to calculate false positive and false negative rates. A false positive for a subset array is a sequence determined to be represented in the hybridization but not included in the 5,152 sequences actually printed on the array from which the pool was derived. A false negative is a sequence that was not represented in the hybridization, despite being an intended



**Figure 4** | The subset sequences gave unique signatures of bright-intensity probes and showed the expected overlap. The heat map shows the results of two-dimensional clustering of logarithmic intensities of 17,552 good probes, representing >98% of the 17,898 valid probes (excluding 825 total GGGTTGGCTC-containing sequences) on the full set and subset cloning array samples. Pink, bright-intensity probes; black, dim-intensity probes; white boxes, probes with expected overlap among the subset arrays. Note that the probe intensity from each array is normalized by its computed threshold for representation so that a sequence is considered represented when its logarithmic intensity is >0.

sequence of the set. For each subset array, the threshold for representation was set such that the sum of the false positive rate and the false negative rate was minimized. The computed threshold essentially segments the bimodal probe intensity distribution into two groups, represented sequences and background (Fig. 3). The same approach can be extended to the full-set array to estimate the number of sequences deemed represented, in which case the representation threshold segments the full-set probes (represented) from the negative control probes (background).

By this approach, labeled IVT products from the full-set of sequences hybridized to 18,686 (99.8%) of the 18,723 unique sequence probes. The collective data for the four subset oligonucleotide pools revealed 390 sequences that overlapped in all four hybridization experiments. This overlap was not intended in the array design. On further inspection, it became apparent that members of this set of sequences shared a highly conserved internal core of approximately ten consecutive bases (GGGTTGGCTC) that included the conserved shRNA loop structure (Supplementary Fig. 1 online). These fortuitous stretches of sequence conservation likely explain the cross-hybridization observed. Of the probes on the microarray, 825 sequences contain the sequence GGGTTGGCTC from positions 27–36.

As a visual illustration of the coverage afforded by our library pools, we eliminated the 825 probes with the common core sequence GGGTTGGCTC and studied only the 17,898 remaining valid probes. Using the segmentation method described earlier, we obtained 17,552 probes with hybridization intensity substantially above background in at least one subset detection array (representing more than 98% of the 17,898 valid probes) and carried out a two-dimensional intensity clustering analysis of these probes. Each

cleaved subset array gave a unique signature (Fig. 4). As expected, we observed small clusters of bright probes for each array that were also bright for intended overlapping arrays (white boxes). With this approach, we obtained an average false positive rate of 6.15% and an average false negative rate of 1.99%. The higher, but still quite low, false positive rate likely reflects a much smaller set of redundant sequences that remains after removal of the 825 GGGTTGGCTC-containing sequences (data not shown). Thus, the true false positive rate probably approaches that of the false negative rate. Considered together with the sample sequencing, these data suggest that pools of oligonucleotides cleaved from microarrays are well represented.

## DISCUSSION

Cost-effective approaches for cloning complex libraries of predefined nucleic acid sequences are very limited. Typically, if there is no natural source of the nucleic acid, oligonucleotides must be synthesized individually for engineering into the larger library. This traditional approach is disadvantageous in several respects. First, it is costly, which limits the number of sequences that can be included in the library. Second, the approach is labor intensive, as each individual oligonucleotide must be manipulated for engineering into the library. Even in cases where natural sources of nucleic acid are available, cloning and manipulation of these might not produce ideally structured populations. Our data show that microarray-based library cloning provides a rapid, cost-effective and flexible approach for the generation of complex, uniformly distributed libraries of defined oligonucleotides.

Ink-jet microarray synthesis has been optimized for production of oligonucleotides of 60 bases or fewer, and such standard arrays will be suitable for many purposes. We have shown that we can use ink-jet synthesis to produce very high-fidelity cloned populations with oligonucleotides of up to 96 bases. Although arrays carrying oligonucleotides of this length are not standard reagents, we used these materials to provide a very stringent test for the performance of array-synthesized oligonucleotides. We noted high fidelity and only modest biases in the amplification of complex populations of highly structured templates. Overall, considering only the accuracy of cloned populations, we consistently recovered 45–55% of clones with perfect sequences. Also considering biases in amplified populations, 25–30% of all clones represent unique and perfect shRNAs. Both the rates themselves and the importance of each metric will vary with individual applications of the approach. For our specific purpose, success rates in generating viable shRNA clones using ink-jet-synthesized oligonucleotides are sufficient to allow this method to be used for the large-scale construction of both mixed and sequence verified libraries, as it does not substantially differ from success rates observed in our previous efforts at library construction using conventionally synthesized material<sup>28</sup>.

The creation of complex libraries by ink-jet DNA synthesis can be applied to address numerous biological problems. For example, this method would be ideal for generating libraries for antibody diversity studies, phage display, combinatorial peptide sequence generation, DNA binding site selection, promoter region analysis and restriction enzyme site analysis. In each case, the necessary oligonucleotide length, the requirement for sequence verification and the arraying clones will vary. Accordingly, the cost savings afforded by this technology will also vary. Overall, our data suggest a parity between the quality of ink-jet synthesized material

and material obtained by mixing populations of conventional oligonucleotides. Given the accuracy and flexibility of ink-jet oligonucleotide synthesis, it is likely that the approach described here will become an important method for constructing diverse library-based tools for functional genomic studies.

## METHODS

**Oligonucleotide design and microarray synthesis.** For this cloning method, any microarray technology capable of *in situ* synthesis of oligonucleotides of the desired length for the application may be appropriate. For our studies, however, we primarily used and validated oligonucleotide microarrays printed at Agilent Technologies using ink-jet technology as described previously<sup>15</sup> with essentially no modifications to standard manufacturer's protocols. Detailed methods for generating ink-jet microarrays can be found in U.S. Patents numbered 6,419,883 and 6,028,189.

Sequences to be included in a library were designed such that each was flanked by 5' and 3' common 14- to 18-base PCR primer recognition sites (Fig. 1b). Before the oligonucleotides were harvested, quality control testing was performed using a functional hybridization of representative arrays that were produced on the same manufactured glass substrates.

**Oligonucleotide cleavage with a photocleavable spacer.** Photocleavable spacer phosphoramidite (Glen Research) monomers were synthesized on a silanized 3 inch × 3 inch × 0.004 inch glass wafer with hydroxyl functionality. Silanization of glass surfaces for oligonucleotide applications has been described<sup>30,32</sup> and silanes with various functionalities are commercially available (Gelest). For these studies a 50:1 mixture of decyl trichlorosilane and 11-trichlorosilyl-1-undecene was used. All reaction steps and reagent preparations were performed under nitrogen in a PLAS-LABS 830-ABC glove box (PLAS-LABS). One microliter of anhydrous acetonitrile (Fisher Scientific) was added by syringe injection to 100 μmol of freeze-dried photocleavable spacer phosphoramidite to yield a 0.1 M solution. Next, 62 ml of anhydrous acetonitrile was added to 2 g of freeze-dried 5-ethylthiol-<sup>1</sup>H-tetrazole (Glen Research) to yield a 0.25 M solution for phosphoramidite activation. The solutions were vortexed briefly and allowed to equilibrate at room temperature for 30 min. One milliliter of tetrazole solution was transferred by syringe to the photocleavable spacer solution, and the mixture was vortexed for 10 s. Two silanized wafers were placed 'reactive side' up and 2 ml of the active photocleavable spacer-tetrazole solution was added to the surface of the first wafer. The second wafer was placed sandwich-like on top of the first, allowing the fluid to distribute uniformly between the surfaces. The wafers were incubated at room temperature for 2 min, separated, placed in a Teflon rack and immersed in a bath of acetonitrile. The rack was agitated in the bath for 2 min to ensure complete rinsing of excess photocleavable spacer and dried by centrifugation. Formation of the stable pentavalent phosphodiester and removal of the dimethoxytrityl protecting group were carried out according to standard oligonucleotide synthesis procedures<sup>15,17</sup>. Synthesis of oligonucleotides on photocleavable spacer-functionalized substrate was performed as described above.

For arrays synthesized with a photocleavable spacer, the oligonucleotides were cleaved in 1 ml of 25 mM Tris-buffer solution (pH 7.4) by placing the array in almost direct contact with a UV

irradiation source (UVM-57, UVP, Inc.; 302 nm wavelength) for 20 min. The solution was transferred to a 1.5-ml microcentrifuge tube and speed vacuumed at 45 °C overnight.

**Oligonucleotide cleavage using ammonium hydroxide.** To cleave oligonucleotides synthesized without a photocleavable spacer, the microarrays were treated for 2 h with 2–3 ml of 35% NH<sub>4</sub>OH solution (Fisher Scientific) at room temperature. The solution was transferred to 1.5-ml microcentrifuge tubes and speed vacuum dried at 45 °C overnight.

**PCR amplification of cleaved oligonucleotides.** Dried material containing oligonucleotides cleaved from each microarray was resuspended in 250 µl of RNase- and DNase-free water. For the PCR template, a range of volumes (0.1–5.0 µl) was tested to determine the amount that gave the best yield with the lowest incidence of nonspecific product. We carried out PCR amplification of the initial 59- and 96-nt test sequences in 50 µl reactions containing 1x PCR buffer without Mg (Invitrogen), 9% sucrose, 1.5 mM MgCl<sub>2</sub>, 1 ng/µl forward and reverse primers, 125 µM dNTPs and 0.05 U/µl *Taq* polymerase. Thermal cycler conditions depended on the length of the oligonucleotides and the melting temperatures of the forward and reverse primers. In general, 30 cycles of 94 °C denaturing for 30 sec, annealing at the appropriate temperature for 30 s, and extension at 72 °C for 90 s worked well. If the PCR products were to be cloned using a TA cloning system such as the Topo TA cloning system (Invitrogen), we used *Taq* polymerase and followed the 30-cycle PCR with a 10 min extension at 72 °C. For the cloning of shRNA libraries, Vent polymerase (New England Biolabs) or Pfx polymerase (Invitrogen) in the presence of DMSO and/or betaine was used to reduce the incidence of nucleotide misincorporation during the PCR. We optimized conditions independently for each primer set used. For the cloning of mouse and human shRNA libraries discussed here, we used Platinum Pfx (Invitrogen) with a 2× final concentration of the manufacturer's provided amplification buffer, a 0.5× final concentration of the provided PCR enhancer, 10 µl of template (1/50 of the oligonucleotides cleaved from each array), 1 mM MgSO<sub>4</sub>, 0.2 µM each of the forward and reverse primers, and 0.2 mM final concentration of all four dNTPs. Thermal cycler conditions were 94 °C for 5 min followed by 25 cycles of 94 °C for 45 s, 68 °C for 1 min 15 s, and finally extension at 68 °C for 7 min. In some cases, PCR products were cleaned up by gel purification using the QIAquick Gel Extraction protocol (QIAGEN). In other cases, the PCR products were cleaned up using the QIAquick PCR purification protocol (QIAGEN).

**Reverse transcription–*in vitro* transcription (RT-IVT) and microarray hybridization.** To prepare templates for T7 IVT, we pooled PCR material from two individual reactions. Unincorporated nucleotides and polymerase were removed from the pooled PCR products by QIAquick PCR purification (QIAGEN) and eluted in 50 µl of RNase and DNase-free water. Eluates were speed-vacuum dried to concentrate two-fold and 7.25 µl was used as template in a T7 RNA polymerization reaction, using a modified MEGAscript protocol (Ambion). In lieu of 2 µl of 75 mM UTP, we used 2.25 µl of 50 mM amino allyl UTP (aa-UTP; Ambion) plus 0.5 µl of the 75 mM UTP provided with the kit. The reactions were carried out at 37 °C overnight. Then, 1 µl of DNase was added for

15 min at room temperature. Next, the samples were phenol/chloroform/isoamyl alcohol extracted and ethanol precipitated. The final product was resuspended in 40 µl of water.

Amino allyl UTP-incorporated cRNA was divided into aliquots in two 96-well plates (5 µg per reaction well). One plate for Cy3 NHS-ester coupling and one for Cy5 NHS-ester coupling were prepared (dyes were obtained from Amersham Biosciences). Samples were reacted with the dyes, mixed for performance of two-color ratio experiments and subsequently purified using Micro Bio-Spin columns P-30 Tris (Bio-Rad Laboratories). Purified dye-labeled samples were then hybridized to the detection microarray for 24 h, washed, scanned on an Agilent Scanner and analyzed. Rosetta standard coupling and hybridization processes were employed as previously described<sup>15</sup>.

*Note: Supplementary information is available on the Nature Methods website.*

#### ACKNOWLEDGMENTS

G.J.H. is supported by an Innovator Award from the US Army Breast Cancer Research Program. This work was also supported by a grant from the US National Institutes of Health (G.J.H.). We thank H. Dai for suggestions regarding microarray analysis of the library population, T. Fare for helpful comments on the manuscript, and the Rosetta Gene Expression Laboratory for microarray RNA processing and hybridizations.

#### COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Methods* website for details).

Received 26 July; accepted 20 October 2004

Published online at <http://www.nature.com/naturemethods/>

- Hudson, J.D. *et al.* A proinflammatory cytokine inhibits p53 tumor suppressor activity. *J. Exp. Med.* **190**, 1375–1382 (1999).
- Brummelkamp, T.R. *et al.* TBX-3, the gene mutated in Ulnar-Mammary Syndrome, is a negative regulator of p19ARF and inhibits senescence. *J. Biol. Chem.* **277**, 6567–6572 (2002).
- Maestro, R. *et al.* Twist is a potential oncogene that inhibits apoptosis. *Genes Dev.* **13**, 2207–2217 (1999).
- Raveh, T., Berissi, H., Eisenstein, M., Spivak, T. & Kimchi, A. A functional genetic screen identifies regions at the C-terminal tail and death-domain of death-associated protein kinase that are critical for its proapoptotic activity. *Proc. Natl. Acad. Sci. USA* **97**, 1572–1577 (2000).
- Fletcher, B.S., Dragstedt, C., Nottenpek, L. & Nolan, G.P. Functional cloning of SPIN-2, a nuclear anti-apoptotic protein with roles in cell cycle progression. *Leukemia* **16**, 1507–1518 (2002).
- Rayner, J.R. & Gonda, T.J. A simple and efficient procedure for generating stable expression libraries by cDNA cloning in a retroviral vector. *Mol. Cell. Biol.* **14**, 880–887 (1994).
- Kitamura, T. *et al.* Efficient screening of retroviral cDNA expression libraries. *Proc. Natl. Acad. Sci. USA* **92**, 9146–9150 (1995).
- Kojima, T. & Kitamura, T. A signal sequence trap based on a constitutively active cytokine receptor. *Nat. Biotechnol.* **17**, 487–490 (1999).
- Golovkina, T.V. *et al.* A novel membrane protein is a mouse mammary tumor virus receptor. *J. Virol.* **72**, 3066–3071 (1998).
- Battini, J.L., Rasko, J.E. & Miller, A.D. A human cell-surface receptor for xenotropic and polytropic murine leukemia viruses: possible role in G protein-coupled signal transduction. *Proc. Natl. Acad. Sci. USA* **96**, 1385–1390 (1999).
- Gallagher, W.M., Cairney, M., Schott, B., Roninson, I.B. & Brown, R. Identification of p53 genetic suppressor elements which confer resistance to cisplatin. *Oncogene* **14**, 185–193 (1997).
- Garkavtsev, I., Kazarov, A., Gudkov, A. & Riabowol, K. Suppression of the novel growth inhibitor p33ING1 promotes neoplastic transformation. *Nat. Genet.* **14**, 415–420 (1996).
- Bock, L.C., Griffin, L.C., Latham, J.A., Vermaas, E.H. & Toole, J.J. Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* **355**, 564–566 (1992).
- Tuerk, C., MacDougall, S. & Gold, L. RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc. Natl. Acad. Sci. USA* **89**, 6988–6992 (1992).

15. Hughes, T.R. *et al.* Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat. Biotechnol.* **19**, 342–347 (2001).
16. Halliwell, C. & Cass, A.E. A factorial analysis of silanization conditions for the immobilization of oligonucleotides on glass surfaces. *Anal. Chem.* **73**, 2476–2483 (2001).
17. Brown, D.M. A brief history of oligonucleotide synthesis. *Methods Mol. Biol.* **20**, 1–17 (1993).
18. Bourdeiu, L., Silberzan, P. & Chatenay, D. Langmuir-Blodgett films: From micron to angstrom. *Physical Review Letters* **7**, 2029–2032 (1991).
19. LeProust, E., Zhang, H., Yu, P., Zhou, X. & Gao, X. Characterization of oligodeoxyribonucleotide synthesis on glass plates. *Nucleic Acids Res.* **29**, 2171–2180 (2001).
20. Hannon, G.J. & Rossi, J.J. Unlocking the potential of the human genome with RNA interference. *Nature* **431**, 371–378 (2004).
21. Brummelkamp, T.R., Bernards, R. & Agami, R. A System for Stable Expression of Short Interfering RNAs in Mammalian Cells. *Science* **296**, 550–553 (2002).
22. Paddison, P.J., Caudy, A.A., Bernstein, E., Hannon, G.J. & Conklin, D.S. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev.* **16**, 948–958 (2002).
23. Paul, C.P., Good, P.D., Winer, I. & Engelke, D.R. Effective expression of small interfering RNA in human cells. *Nat. Biotechnol.* **20**, 505–508 (2002).
24. Lee, N.S. *et al.* Expression of small interfering RNAs targeted against HIV-1 rev transcripts in human cells. *Nat. Biotechnol.* **20**, 500–505 (2002).
25. Sui, G. *et al.* A DNA vector-based RNAi technology to suppress gene expression in mammalian cells. *Proc. Natl. Acad. Sci. USA* **99**, 5515–5520 (2002).
26. Kawasaki, H. & Taira, K. Short hairpin type of dsRNAs that are controlled by tRNA(Val) promoter significantly induce RNAi-mediated gene silencing in the cytoplasm of human cells. *Nucleic Acids Res.* **31**, 700–707 (2003).
27. Miyagishi, M. & Taira, K. U6 promoter-driven siRNAs with four uridine 3' overhangs efficiently suppress targeted gene expression in mammalian cells. *Nat. Biotechnol.* **20**, 497–500 (2002).
28. Paddison, P.J. A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428**, 427–431 (2004).
29. Berns, K. *et al.* A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431–437 (2004).
30. Sen, G., Wehrman, T.S., Myers, J.W. & Blau, H.M. Restriction enzyme-generated siRNA (REGS) vectors and libraries. *Nat. Genet.* **36**, 183–189 (2004).
31. Shirane, D. *et al.* Enzymatic production of RNAi libraries from cDNAs. *Nat. Genet.* **36**, 190–196 (2004).
32. Luo, B., Heard, A.D. & Lodish, H.F. Small interfering RNA production by enzymatic engineering of DNA (SPEED). *Proc. Natl. Acad. Sci. USA* **101**, 5494–5499 (2004).
33. Hsieh, A.C. *et al.* A library of siRNA duplexes targeting the phosphoinositide 3-kinase pathway: determinants of gene silencing for use in cell-based screens. *Nucleic Acids Res.* **32**, 893–901 (2004).