ANL/CMB/PP--83848

For submission to *Biotechniques - BioFeatures*

Coupling Sequencing by Hybridization (SBH) with Gel Sequencing for an Inexpensive Analysis of Genes and Genomes

Snezana Drmanac, Ivan Labat, Brian Hauser, and Radoje Drmanac

Integral Genetics Group

Center for Mechanistic Biology and Biotechnology

Argonne National Laboratory

9700 South Cass Avenue

Argonne, IL 60439

Address correspondence to: Radoje Drmanac

Key words: sequencing by hybridization, gel sequencing, genome project, partial sequencing, cDNA screening, clone ordering

DISCLAIMER

ABSTRACT

The speed and cost of DNA sequencing are bottlenecks in the analysis of genes end genomes. Sequencing by hybridization (SBH) is a versatile method with several applications which can accelerated DNA screening, mapping and sequencing. Requirements, achievements and problems in the development of the SBH format 1 (DNA samples arrayed) are presented and schemes for its synergetic coupling with gel sequencing techniques are discussed. It appears that by one hybridization machine with 24 boxes and four ABI gel sequencers 100-300 Mb of DNA sequence can be determined per year. Various genetic studies based on computer assisted analysis of large collections of partial or complete DNA sequences ("sequenetics") may be achieved in this century.

INTRODUCTION

Many biological problems can be solved by analyzing relevant DNA sequences. The identification of genes and other functional elements and definition of their functions are likely to be achieved in the majority of cases from partially determined sequences (1,16) or from sequences containing errors (28). Nevertheless, reading 100 billion base pairs contained in the genomes of 30-50 species and individuals by conventional methods is impractical. New high-speed sequencing methods are needed to bring about an advancement like that introduced by gel sequencing (25,27). The nature of the DNA molecule and the size and composition of genomic DNAs probably requires a combination of methods for efficient high-volume sequencing.

The chemical complementarity of bases underlies the processes of copying and retrieving genetic information. It also offers a possibility for defining the sequence of the four bases in genomic DNAs. Chemically synthesized oligonucleotides of known sequence can be hybridized with target DNA molecules to reveal the presence of complementary oligomers. Scoring a sufficient number of different oligonucleotides permits the assembly of continuous sequences. The sequencing by hybridization (SBH) method has been proposed by several groups (reviewed in 7 and 10) and has been proved by sequencing 340 bp without error in a blind test (13).

SBH has an enormous technological versatility and variety of applications. In format 1, DNA samples are arrayed on a support and labeled oligonucleotide probes are applied. In format 2, different DNA samples are labeled and sequentially applied to an array of oligonucleotides. Combinations of the two formats are also possible. In both formats, the arrays can be miniaturized from the present millimeter scale of dot diameters to the micrometer level referred to as "sequencing chips" (17, reviewed in 7 and 10). The technology can be used for DNA screening, fine resolution mapping and partial sequencing (16), diagnostics (reviewed in 23) and large-scale sequencing (14). Achievements in the development of SBH format 1 and examples of applications are presented in this article. In addition, we discuss schemes for the integration of SBH and gel sequencing to inexpensively produce the data necessary for computer-based genetic studies.

SBH FORMAT 1: DATA PRODUCTION FACILITIES

SBH format 1 is in essence a large-scale dot-blot analysis of 0.5- to 5-kb DNA fragments with thousands of short oligonucleotides (6-12 bases). This is a high-volume but relatively slow (months) process in which the consistency of procedures is of primary importance. In addition, the success of the method depends on automatic monitoring of the experiments (which clones, probes, filters, and experimental conditions have to be used and are actually used) and on an automatic input of

the annotation and data into a data base. Finally, user friendly software for data evaluation and selection is necessary to be able to recognize and repeat unsuccessful probings and prepare complete data sets for application programs.

To achieve high data throughput, samples are stored and manipulated in high-density multiwell plates using corresponding pin tools and are spotted in packed-dot arrays. Results, problems, and plans concerning automated production and acquisition of the hybridization data are described in the following sections.

Sample processing

Critical requirements in sample preparation are a high concentration of the target DNA and a low contamination with bacterial or vector DNA. Bacteriological and biochemical procedures for preparation of M13 and plasmid clones have been developed (8,9). The steps of the process are listed in Table 1. A careful manipulation of the transformation mixture (aliquoting and storing at -80 °C in the freezing medium (8); and sufficient mixing of the cells in the aliquoting and dilution steps) are crucial for consistent results. Clones are arrayed either by picking or by dispensing optimally diluted transformation mixture (8). The use of 864-well plates as master plates reduces storage space and manipulations and allows parallel processing of 864 samples by a corresponding pin array. If large colonies are

picked, enough bacterial cells are transferred in 15-µl cultures and subsequent growth is not necessary. Step 3 is optional and provides backup or optimization of the conditions for the preparative step. We found that PCRs are more successful from the diluted cultures (12). Furthermore, replica plates allow nonsterile inoculation of PCR reactions.

Table 1. Steps of the sample preparation process

1. Library transformation and titration
2. Clone arraying in multiwell plates
3. Replica plate preparation
4. PCR or preparative growth
5. Sample checking and preparation for spotting

Step 4 is critical. If it is highly successful, step 5 is unnecessary. For example, M13 cultures have an insufficient phage concentration, and evaporation of the medium is used as one way to solve the problem (8). Fortunately, high-yield PCRs (>30 ng/µl) allow direct spotting without removing oil (8). We developed high-throughput PCR directly from bacterial cells using a BioOven III with six (and recently ten) multiwell plates (8). An example of consistently successful PCR of cDNA clones is shown

on Fig. 1. In this streamlined procedure, the steps of DNA purification or chemical modification are eliminated.

Procedures for the automatic transfer of liquid between and into 96- and 864-well plates have been developed for a modified Biomek1000 workstation, using a MP200 pipetting tool and fine pipetting tips (Cat. No. 609044, Beckman Instruments, Fullerton, CA and Cat. No. P-3701-CP, Denville Scientific, Inc., Denville, NJ). Filling of one 864-well plates takes approximately 5 min. Plates are labeled with bar codes using TPS Electronics (TPS Electronics, Palo Alto, CA) bar code printing software. We are using a 12-character bar code, which is sufficient for defining the operator's initial, unique plate number, name of the library, and type of samples (master cultures, dilutions, or PCR samples).

Spotting

Spotting DNA using a metal pin array has been developed (8). We automated the procedure to generated extremely dense dot arrays by an 864-floating-pin tool using an adapted Biomek1000 (I.L., R.D. and B. Hauser). The original tablet is replaced by a 1-cm thick lucite plate with holders for a source plate and a box with denaturation solution for pin washing (Fig. 2a). A rail of bearings is added at front to keep the tablet horizontal. The new Biomek2000 has a fixed tablet and doesn't need this solution. The design of pins (Figure 2b) is also optimized to allow

reproducible spotting from 2-mm wells and to prevent pin bending. A sophisticated software is developed to execute specific movements of pins for picking samples through the oil cover (8), printing dots on the filter, and washing and drying the pins between consecutively spotted plates. Also, a user-friendly interface allows the selection of filter size, density of dots (Fig. 2c), plate pattern, file, and number of repeated spottings in the same position. Plate pattern files are predefined by scanning bar codes from the selected plates.

Filters are laid down on two 3MM Whatman papers soaked in the denaturation solution. Whole setup has to be dried for 15 min to avoid spreading DNA. We routinely prepare filters with 31,104 dots (4× [3 × 3] 864-well plates, Fig. 3). It takes one hour per filter including the filter-drying step. Each plate is put manually in the holder and checked by bar code reader. The replacement of plates by a side loader extends spotting time for about 30 min. From a 15-µl PCR, up to 100 replica filters can be spotted. Filters are internally encoded by a predefined pattern of empty wells.

Hybridization process and signal reading

The core SBH process is full match/mismatch discriminative hybridization reaction (10,17). This step may represent the bottleneck of the entire process. Currently we are labeling probes with γ-^{33}P-ATP using T4 kinase. The procedure is very

simple but quite expensive. Probe dilutions of 10 ng/µl are stored and aliquots are labeled in 96-well plates. One filter is hybridized per box in 80 ml of buffer for 1-16 h. After three washings, the excess buffer is drained by putting the membranes between Whatman papers, and then the naked membranes are laid down in exposure cassettes. Storage phosphor plates are exposed for 2-18 h and scanned on a PhosphorImager (Molecular Dynamics, Sunnyvale, CA). Two filters (62,000 dots) can be scanned in 4 min. High discrimination of full-match and end-mismatch targets is obtained (Fig. 4). Positive clones repeatedly spotted on four positions give similar hybridization signals. To reuse the filters, the hybridized probes are stripped off in hybridization buffer for 1-20 h at 37-65 °C, depending on the C+G content, the length of the probes, and convenience. Some filters are reused 50 or more times without significant quality deterioration.

We are now using three semi-automatic hybridization systems with four boxes each. In collaboration with Intelligent Automation Systems (Boston, Massachusetts), we are developing a fully automated, PC-operated hybridization machine with 24 boxes organized in 6 drawers. A flow-through 24-well cassette (8 × 3) is designed for automatic labeling and direct introducing of probes in the boxes to avoid incorrect annotation about which probe is used with which filter. The cassettes will be barcoded, and eventually they will be automatically inserted in the machine. Three to four hybridization cycles can be finished per day using ^{33}P-labeled probes.

The only step that is difficult to automate is the image acquisition using phosphor screen technology, which requires an extensive manipulation of filters, screens, and cassettes. Direct imaging of the filters inside the boxes using a CCD camera can be integrated with the hybridization machine if fluorescent label(s) are used (18). An XYZ table can be used to open the drawers and bring the CCD camera and a source of excitation light to each of the boxes. Imaging time of up to 30 s per filter (12 min for 24 boxes) is acceptable. With this improvement, a closed hybridization process can be established. After imaging, filters will be automatically stripped off and a new set of 24 probes (a new cassette) has to be inserted in the machine. By such a hybridization and detection system, 8-12 cycles (192-288 probes) can be processed per day. Fluorescent labels allow a fourfold increase of the dot density and the simultaneous screening of at least four probes labeled with specific dyes. By implementing these options, as many as 100 million clone-probe hybridization results can be scored per day per machine.

Image analysis and data evaluation and scaling

The precision of scoring intensities of the hybridization signals is as important as consistency of the hybridization itself. Our image analysis program (DOTS, 12) finds rows and columns of dots automatically and reports the average signal for several pixels around the center of mass for each dot. In the

density of 9 x 9 dots per 9 x 9 mm (one well of a 96-well plate), there are about 36 pixels (at low resolution of the Phosphorimager) per dot. The average difference between two readings for the same dot obtained from two exposures of the same filter is about 15%.

Even more versatile software is necessary for data manipulation, evaluation and scaling. A program (SCORES) with a user-friendly X windows interface has been developed for this purpose (N.A. Stavropoulos and R.D.). The program allows visual selection of filters, probes, multiwell plates, or clones. It includes several statistical functions for data evaluation. For example, the ratio of the average hybridization signal for the dots which do not have a full-match target and of the average background value (dots without DNA) is defined for each probe: the ratios are supposed to be greater than one but not much greater. If the values are less than one, it indicates unsuccessful hybridizations, errors in the image analysis, or the presence of nonspecific signals. If the value of this parameter is greater then four, usually a nondiscriminative hybridization is scored.

SCORES adjusts signals for the difference in the amount of DNA among the dots and scales the absolute hybridization values from independent hybridization experiments into the same range to allow comparisons. Scaled values (scores) are less than 2 for clones with no matches and are usually not greater then 20 (8 is the average) for clones having full-match targets. The scores are

more informative than a binary (0/1) derivative of the raw data. The final result of the SCORES program is tables of clone-probe scores. Scores for one clone represent a signature of that clone. The signatures are used to define clone similarities in screening and mapping applications (11).

IDENTIFICATION OF EXON-CONTAINING CLONES

One application of the oligomer screening procedure is to select short genomic clones bearing one or more exons. We defined a set of 46 seven-mers that recognize 70% of the 1-kb genomic fragments containing exon sequences (I.L. and R.D.). The probes are selected and the tests are performed on a 1.3-Mb human genomic sequence made by merging 70 longest GenBank entries. This sequence comprises 631 annotated exons (62,176 bp). Heptamers which occur five times more frequently in exons than in the total sequence are defined first. From this subset we selected probes with the highest ratio of frequency in exon sequences to frequency in total sequence and which had less-than-average occurrence in 501 Alu repeats present in the 1.3-Mb test sequence (Table 2). The rate of false positive calls obtained in these tests was 32%. The screening produces a set of clones enriched with coding sequences similar to those obtained by exon-trapping. The probes are also used in cDNA screening. For a discriminative screening of 2-kb fragments, probes with eight bases have to be used.

Table 2. Probes for recognition of coding sequences in human DNA

1. GGTGATG	11. GGAGAAG	21. GGATGAG	31. CCCGTGC	41. GCCGCCC
2. TGGAGGA	12. ACCTGGA	22. TGGAGCA	32. CCACGGG	42. CCCCCCC
3. GAAGAAG	13. AGCAGCT	23. CTCACCA	33. GCCCACG	43. CCACGGC
4. CAAGAAG	14. GGACAAG	24. TGGACAA	34. GTGCCGC	44. CCTGCGC
5. TGGTGGA	15. CAAGGAG	25. AGCTGAC	35. GACATCG	45. GGACATC
6. GGAGATG	16. GTGCTGA	26. CACCAAG	36. CCCACGG	46. CCGCCTG
7. TGATGGT	17. AGAAGGT	27. CACCAAC	37. TGCCGCC	
8. TGCTGGA	18. GGTGAAG	28. ACCTGCA	38. CTGCGCA	
9. TGGATGA	19. CTACAGC	29. CTGACCA	39. GCGGCTG	
10. AGCTGCA	20. TGGACCT	30. CGGGTGA	40. CGCCCAC	

cDNA SCREENING

Screening cDNA clones using 100-200 probes (16,24) can facilitate the effort in cataloging genes by expressed sequence tag (EST) approach (1,19,26). We are analyzing an infant brain library obtained from B. Soares (Columbia University, New York). Three filters, each in 20 replicas (over 70,000 scorable dots), are prepared and hybridized with 250-350 probes. Hybridization patterns of 4 probes are shown in Fig. 5. Computational procedures have been developed for pairwise comparisons of the clone signatures and for clustering of significantly similar

clones (11). Test comparisons of 18,600 clones from an ordinary library and 14,200 clones from a normalized library give 8,000 and 9,070 distinct clusters, respectively. The redundancy (number of clones/number of clusters) is 2.33 for the ordinary and 1.56 (one-third less) for normalized library. Because of the large number of genes expressed in the brain, 15,000 clones is not sufficient to determine the real difference in redundancy. Due to the influence of statistical factors and limitations of the normalization process (11), we expect that the maximal decrease of redundancy in the normalized library is not more than two-thirds.

By hybridization selection of unique clones from ordinary libraries, an ultimately normalized set of clones and relative expression levels of the recognized genes can be defined. Due to a low cost (50 cents per clone), 10 million necessary clones from various tissues (11) can be screened in two to three years. A few percent of the representative clones will be selected to define ESTs or to determine complete sequences.

SHOTGUN CLONE MAPPING

We have developed a heuristic algorithm to determine the order of clones from shotgun libraries on the basis of the signatures generated by 100-200 oligonucleotide probes (R.D, I.L., and J. Schmutz). The same type of pairwise comparisons as in the cDNA sorting is performed to identify highly overlapped

clones. The most similar pairs are used as seeds for further ordering. The complete probe set, or subsets of non-shared probes defined from overlapped clones (14), are used in consecutive linking steps. The clone ordering program (CORD) has been developed for UNIX platforms using GNU C compiler.

The maps obtained from a number of simulation experiments on 40,000 and 1,000,000 bases show small local swapping of neighboring clones (Fig. 6) even with the introduction of more errors than are obtained in real hybridization experiments. Using CORD, M13 clones from a cosmid shotgun library (provided by B. Koop) have been ordered on the bases of hybridization data of 250 probes. The overlapped clones are confirmed by restriction mapping. Simulation experiments show that maps of shotgun clones reduce sequencing redundancy to about two readings per base.

SCHEMES FOR COUPLING SBH WITH GEL SEQUENCING

The basic technical difference between DNA mapping and DNA sequencing by hybridization is in the number of probes needed. The original SBH proposal was based on 100,000 probes (14). Recently, a more efficient technique has been proposed (4,10). Single-pass gel sequences and hybridization data from about 3000 probes (mainly 7-mers) have to be prepared for the same clones (Fig. 7). These low-redundancy data complement each other and provide complete sequences. Gel sequences with up to 20% errors are used as a frame to put together positively hybridized probes

and solve branches (14). The probes read each base three to four times, providing local accuracy.

This is a high-volume and cost-effective sequencing scheme. A minimal set of clones (1000 1- to 2-kb clones per million bases) will be first selected from shotgun libraries in the mapping phase. Our filters with 31,000 dots contain at least 20 Mb of DNA assuming that 30% of the dots will represent control or unsuccessful clones. With the previously described machine using ^{33}P as the label, 3000 probes can be scored in two months (72 probes per day using 72 replica membranes). The cost projected from our cDNA screening experience should be about $8/clone, or much less if the ^{33}P is replaced by fluorescent label. Continuous single-pass gel sequences with as much as 20% error in some segments, can be produced by a similar cost. Each clone can be read through starting from one or both ends. Present cost per gel read is about $10. The total cost could be less than 10 cents per base pair including mapping. Four sequencers (48 lines, Applied Biosystems, Inc.) running two cycles per day are necessary to match one ^{33}P hybridization machine and to produce about 100 Mb of the complete sequence per year.

Further improvements are possible if the hybridization data from 70- to 95%-similar sequences are integrated (4,7,22). The data necessary to sequence four similar bacterial genome (20 Mb total) are listed in Table 4. In this case, about 5000 targeted segments

Table 3. Data required for sequencing of four similar bacterial genomes

Steps	Filters(a)	Probes	Days(b)
Mapping	8	200	25
Partial Sequencing	2	1000	30
Complete Sequencing	1	2000	30

a) Filters containing 31,000 dots.

b) Rounded numbers of working days of one hybridization machine with 24 boxes running three cycles per day.

have to be sequenced by gel methods. The mapping step represents 30% of the effort but maps of this precision will minimize the sequence closure phase. Partial sequencing is an optional step which can be sufficient for many purposes. Using benefits from similar sequences and from a growing volume of known sequences, one fluorescence-based hybridization machine with a throughput of 40 million clone-probe scores per day and four gel sequencers with an expected twofold increased capacity should be able to determine more than 1 Mb of finished sequence per day.

The proposed marriage of SBH and gel sequencing provides two striking opportunities: immediate access to the most important

genetic information through continuous raw sequence and a rational selection of the unique genomic segments for complete sequencing. A partial sequencing of the human genome using 1000 seven-mers should cost about $20 million. Assuming ordered clones, single-pass sequencing (5-10% error) of the human genome can be achieved by the end of the century by 150 machines at a modest cost of about $140 million (10% of the projected spending for the genome project). Computer-assisted identification of genes and their function by the analysis of partial sequences ("sequenetics"), especially if these data are available for a few genomes, can be extensively practiced in this century.

In the future, miniaturized arrays (6,15,20,21) may provide redundant hybridization data inexpensively and the use of single-pass gel sequences may not increase sequencing efficiency. For example, by ligation of the probes from miniaturized arrays (format 2) with labeled probes in solution (format 1), sequencing 1-10 Mb without subcloning is projected (format 3) (5). The ligation step has been shown to be very effective (2). For larger genomes, various ligation-based fractionating of genomic segments (3,6) can be used as an integral component of a genome sequencing machine. For example, an array of 1024 DNA anchors, each with one specific five-base-long protruding end, allows separation of the genomic fragments generated by *Hga* I (GACGC(N5')/CTGCG(N10') like restriction enzymes in 1024 fractions. Each fraction will contain about 3 Mb of DNA and can be directly sequenced by format 3. SBH combined with gel sequencing methods can be now used for

low-cost large-scale DNA screening, mapping, and sequencing, and it has the capacity to provide routine genome analysis and a comprehensive and affordable diagnostics in the future.

ACKNOWLEDGMENTS

We thank David E. Nadziejka for technical editing and Kay Bexson for technical assistance. Work supported by the U.S. Department of Energy, Office of Health and Environmental Research, under Contract No. W-31-109-ENG-38.

REFERENCES

1. **Adams, M.D., M. Dubnick, A.R. Kerlavage, R. Moreno, J.M. Kelley, R.R. Utterback, J.W. Nagle, C. Fields and J.C. Venter.** 1992. Sequence identification of 2,375 human brain genes. Nature *355*:632-634.

2. **Broude, N.E., T. Sano, C.L. Smith and C.R. Cantor.** 1994. Enhanced DNA sequencing by hybridization. Proc. Natl. Acad. Sci. USA *91*:3072-3076.

3. **Chetverin, A.B. and F.R. Kramer.** 1993. Sequencing of pools of nucleic acids on oligonucleotide arrays. BioSystems *30*:215-231.

4. **Drmanac. R.** 1992. How 1000 base pairs with 10% error become 10,000 base pairs of correct sequence: a felicitous marriage of the gel and hybridization sequencing methods, p. 318. Abstracts of Genome Mapping and Sequencing Meeting. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

5. **Drmanac, R.** 1994. Genome sequencing machine, p. 60. Abstracts of Genome Mapping and Sequencing Meeting. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

6. **Drmanac, R. and R. Crkvenjakov. 1990.** Prospects for a miniaturized, simplified and frugal human genome project. Scientia Yugoslavica 161(1-2), 97-107.

7. **Drmanac, R. and R. Crkvenjakov.** 1992. Sequencing by hybridization (SBH) with oligonucleotide probes as an integral approach for the analysis of complex genomes. Int. J. Genome Res. *1*:59-79.

8. **Drmanac, S. and R. Drmanac.** An efficient processing of cDNA and genomic kb-size clones for massive screening, mapping and sequencing by hybridization, *Biotechniques*, in press.

9. **Drmanac, R., S. Drmanac, I. Labat, R. Crkvenjakov, A. Vicentic and A. Gemmell.** 1992. Sequencing by hybridization: towards an automated sequencing of one million M13 clones arrayed on membranes. Electrophoresis *13*:566-573.

10. **Drmanac, R., S. Drmanac, J. Jarvis and I. Labat.** 1994. Sequencing by hybridization. *In* Automated DNA Sequencing and Analysis Techniques (M.D. Adams, C. Fields and J.C. Venter, eds.), pp 29-36, Academic Press Limited, London.

11. **Drmanac, R., S. Drmanac, I. Labat and N. Stavropolous.** Requirements in screening cDNA libraries for new genes and

solutions offered by SBH technology, Proceedings of 3rd International Workshop of Transcribed Sequences, in press.

12. **Drmanac, R., S. Drmanac, I. Labat, A. Vicentic, A. Gemmell, N. Stavropoulos and J. Jarvis.** 1993. SBH and the integration of complementary approaches in the mapping, sequencing, and understanding of complex genomes, p. 121-134. *In* H.A. Lim, J.W. Fickett, C.R. Cantor, and R.J. Robbins (Eds.), Proceedings of Second International2Pros Conference on Bioinformatics, Supercomputing and Complex Genome Analysis.

13. **Drmanac, R., S. Drmanac, Z. Strezoska, T. Paunesku, I. Labat, M. Zeremski, J. Snoddy, W. K. Funkhouser, B. Koop, L. Hood and R. Crkvenjakov.** 1993. DNA sequence determination by hybridization: a strategy for efficient large-scale sequencing. Science *260:*1649-1652.

14. **Drmanac, R., I. Labat, I. Brukner and R. Crkvenjakov.** 1989. Sequencing of megabase plus DNA by hybridization: theory of the method. Genomics *4*:114-128.

15. **Drmanac, R., I. Labat, Z. Strezoska, T. Paunesku, D. Radosavljevic, S. Drmanac and R. Crkvenjakov.** 1991. Sequencing by oligonucleotide hybridization: a promising framework in decoding of the genome program? *In* Proceedings of the First International Conference on Electrophoresis,

Supercomputing and the Human Genome, World Scientific Publishing Co., 47-59.

16. **Drmanac, R., G. Lennon, S. Drmanac, I. Labat, R. Crkvenjakov and H. Lehrach.** 1991. Partial sequencing by olig-hybridization: concept and applications in genome analysis, p. 60-74. *In* C.R. Cantor and H.A. Lim (Eds.), Electrophoreses, Supercomputing and the Human Genome, World Scientific, Singapore.

17. **Drmanac, R., Z. Strezoska, I. Labat, S. Drmanac and R. Crkvenjakov.** 1990. Reliable hybridization of oligonucleotides as short as six nucleotides. DNA Cell Biol. *9*:527-534.

18. **Graeser, D., V. Nowotny and D.J. States.** 1994. Efficient fluorescent detection of oligonucleotide hybridizzation for physical mapping, p. 77. Abstracts of Genome Mapping and Sequencing Meeting. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

19. **Khan, A.S., A.S. Wilcox, M.H. Polymeropoulos, J.A. Hopkins, T.J. Stevens, M. Robinson, A.K. Orpana and J.M. Sikela.** 1992. Single pass sequencing and physical and genetic mapping of human brain cDNAs. Nature Genetics *2*:180-185.

20. **Khrapko, K.R., Yu.P. Lysov, A.A. Khorlin, I.B. Ivanov, G.M. Yershow, S.K. Vasilenko, V.L. Florentiev and A.D. Mirzabekov.** 1991. A method for DNA sequencing by hybridization with oligonucleotide matrix. *In* DNA Sequence-J. DNA Sequencing and Mapping Vol. 1, pp. 375-388, Harwood Acad. Publishers, UK.

21. **Pease, A.C., D. Solas, E.J. Sullivan, M.T. Cronin, C.P. Holmes and S.P.A. Fodor.** 1994. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc. Natl. Acad. Sci. USA *91*:5022-5026.

22. **Labat, I. and R. Drmanac.** 1993. Simulations of ordering and sequence reconstruction of random DNA clones hybridized with a small number of oligomeric probes, p. 555-565. *In* H.A. Lim, J.W. Fickett, C.R. Cantor, and R.J. Robbins (Eds.), Proceedings of Second International Conference on Bioinformatics, Supercomputing and Complex Genome Analysis. World Scientific, Singapore.

23. **Landgren, U.** 1992. Detection of mutations in human DNA. GATA *9(1)*:3-8.

24. **Lennon, G.S. and H. Lehrach.** 1991. Hybridization analyses of arrayed cDNA libraries. TIG 7:314-317.

25. **Maxam, A.M. and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA *74*:560-564.

26. **Okubo, K., N. Hori, R. Matoba, T. Niiyama, A. Fukushima, Y. Kojima, and K. Matsubara.** 1992. Large scale cDNA sequencing for analysis of quantitative aspects of gene expression. Nature Genetics *2*:173-179.

27. **Sanger, F., S. Nicklen and A.R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*:5463-5467.

28. **States, D.J. and D. Botstein.** 1991. Molecular sequencing accuracy and the analysis of protein coding regions. Proc. Natl. Acad. Sci. USA *88*:5518-5522.

FIGURE LEGENDS

Figure 1. PCR reproducibility. Four PCR plates are prepared from a master plate using same condition. Three microliters of the samples from row A are loaded on the gel.

Figure 2a. The scheme of the Biomek 1000 modified tablet; 96- or 864-well plates with DNA samples can be used. A 96-well plate or an open box of the same size filled with denaturation buffer is used for pin washing. The tablet and plate holders are made of lucite.

Figure 2b. The design of spotting pins. The pins are made of 316 stainless steel. The dimensions are given in millimeters.

Figure 2c. Spotting patterns generated by 96- or 864-pin tool. Dots that fit in the area of one well (9 x 9 mm) are depicted. The numbers in the "No. of dots (plate)" column refer to 96-well plates. The same number of dots will be generated from one-ninth the number of 864-well plates.

Figure 3. High-density spotting by metal pin array. A filter containing 31,104 dots is generated by spotting PCR samples from 324 plates (96-well) by a corresponding array of metal pins having a 0.3-mm tip diameter. The image of a quarter of the filter (7776 dots) hybridized with a probe complementary to the co-amplified vector ends (CCATGATTAC) is shown. The superimposed grid is generated by the DOTS program.

Figure 4. Hybridization specificity. A filter containing 96 dots from M13 clones was hybridized with a GGGGAGCG octamer probe. The image with a grid generated by ImageQuant software (Molecular Dynamics, Sunnyvale, CA) and the hybridization intensities from column 6 (marked by arrow) are shown. Every second square in both directions contains a dot. Control clones 1, 7, and 8 (8) are marked. Clone 1 has full-match target and clone 8 an end-mismatch target for the probe. Peak 7 (clone 1) is over fivefold larger than peak 2 (clone 8). Peaks 6 and 8 are produced by an additional control clone, which has a full-match target and which is grown in the two corresponding wells.

Figure 5. cDNA patterns. Images of the top left corner of the same filter hybridized consecutively with 8-mer probes (NAAATGTGGN and NCCAGGCCCN, top) and two exon-specific 7-mer probes (NCCACGGCN and NNTGGAGGAN, bottom) are shown. The images represent a twofold enlargement of the actual size of the pictures. The images with the superimposed grids are generated by the DOTS program.

Figure 6. Sequence-ready map of shotgun clones. A random sequence of 40,000 bases was randomly subdivided into 334 clones, 1200 bases in length, and hybridization data with 200 random heptamer probes was simulated for each clone. Superimposed overviews of derived (red) and actual (blue) maps are shown (top). Visible blue bars indicate inconsistencies among the maps.

Green bars show the portion of the map magnified (bottom). Only local swaps of neighboring clones occur.

Figure 7. Integration of single-pass gel sequences and partial SBH data. The DNA segment in the open box is represented by a single pass sequence with two wrongly scored T's, and by four positive 6-mers. Overlapped probes identify errors in the single-pass sequence, and the single-pass sequence solves the branching problem depicted by the two probes in the boxes, which are positive on the other place in the target DNA. Without gel sequencing, it can not be decided whether sequence ACCGACA is followed by TC or by GT.

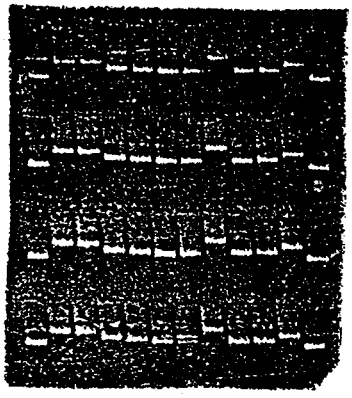


FIGURE 1

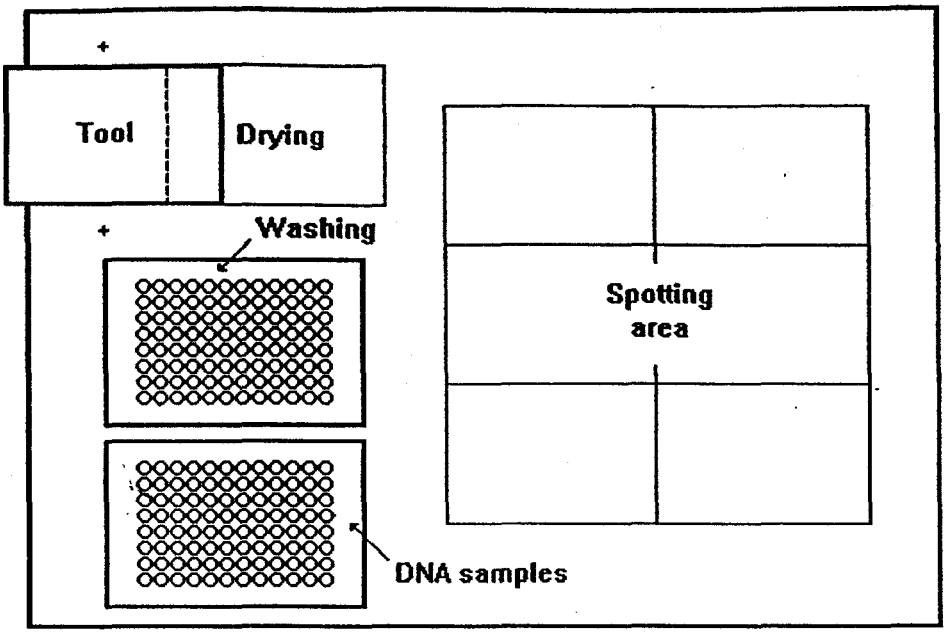


Figure 2a

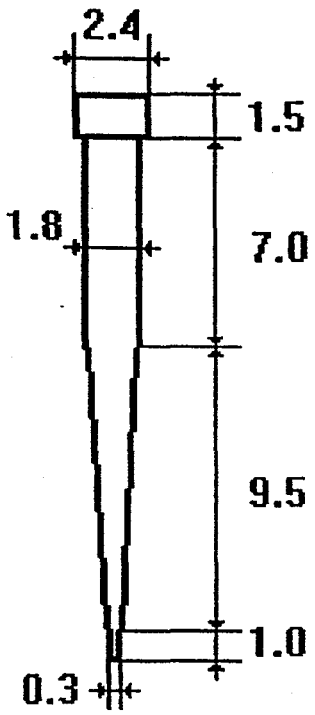
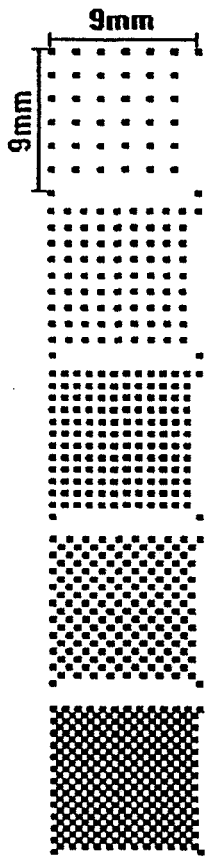


Figure 2L



No. of dots (plates)	dot diameter (mm)	distance (mm)	diagonal distance (mm)
6x6=36	0.375	1.125	1.745
9x9=81	0.375	0.625	1.038
12x12=144	0.375	0.375	0.685
2x9x9=162	0.375	0.625	0.331
2x12x12=288	0.375	0.375	0.155

Figure 2C

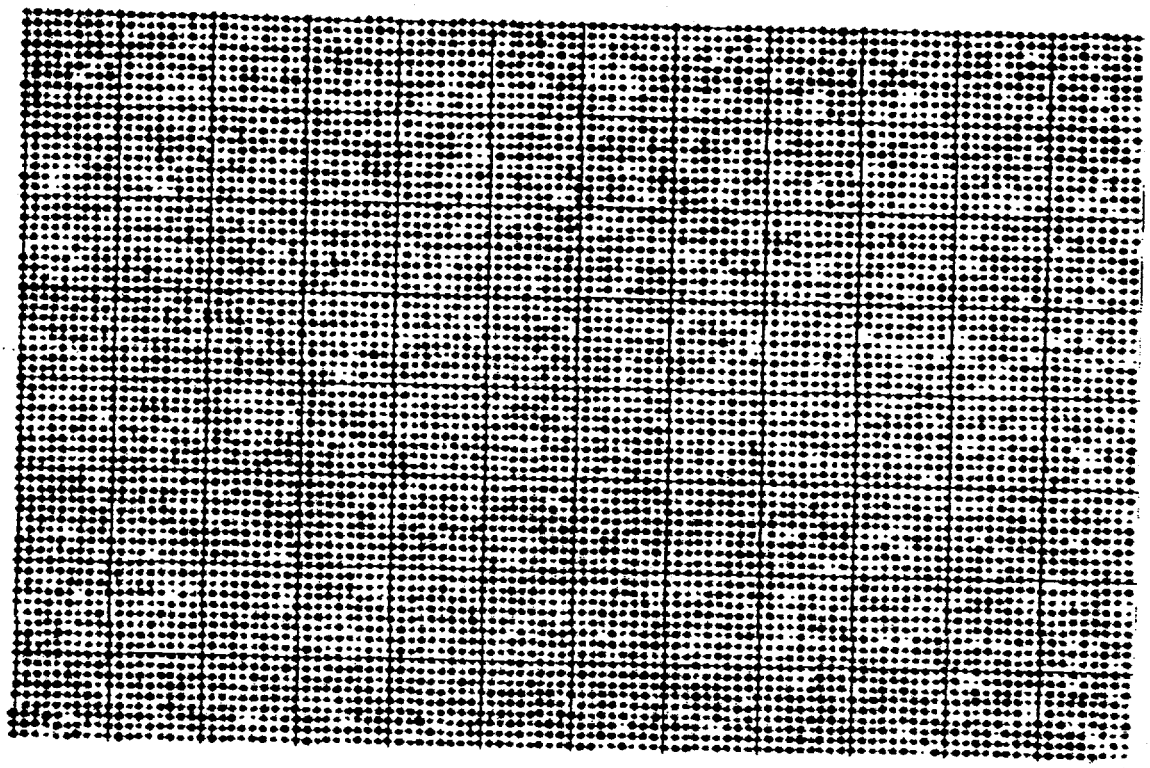
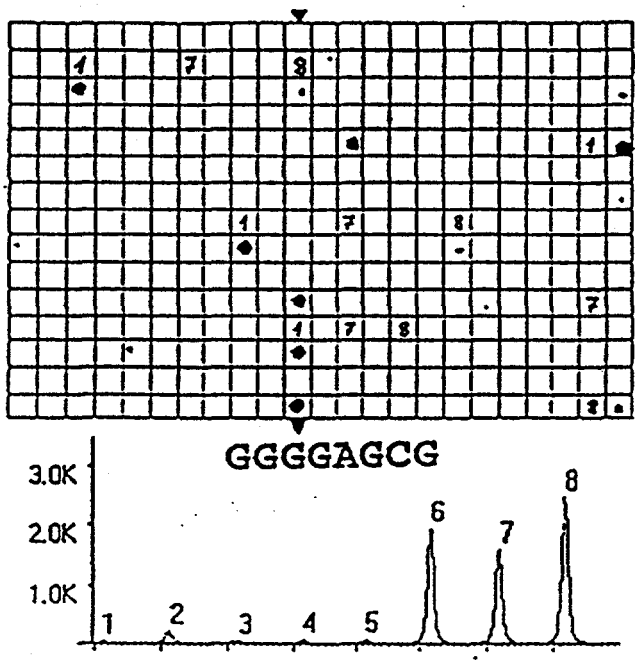
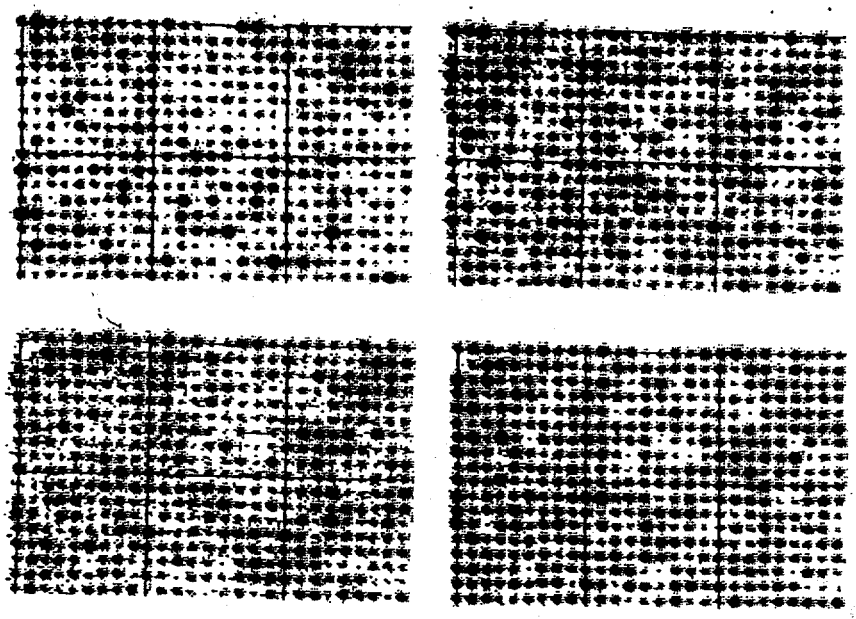
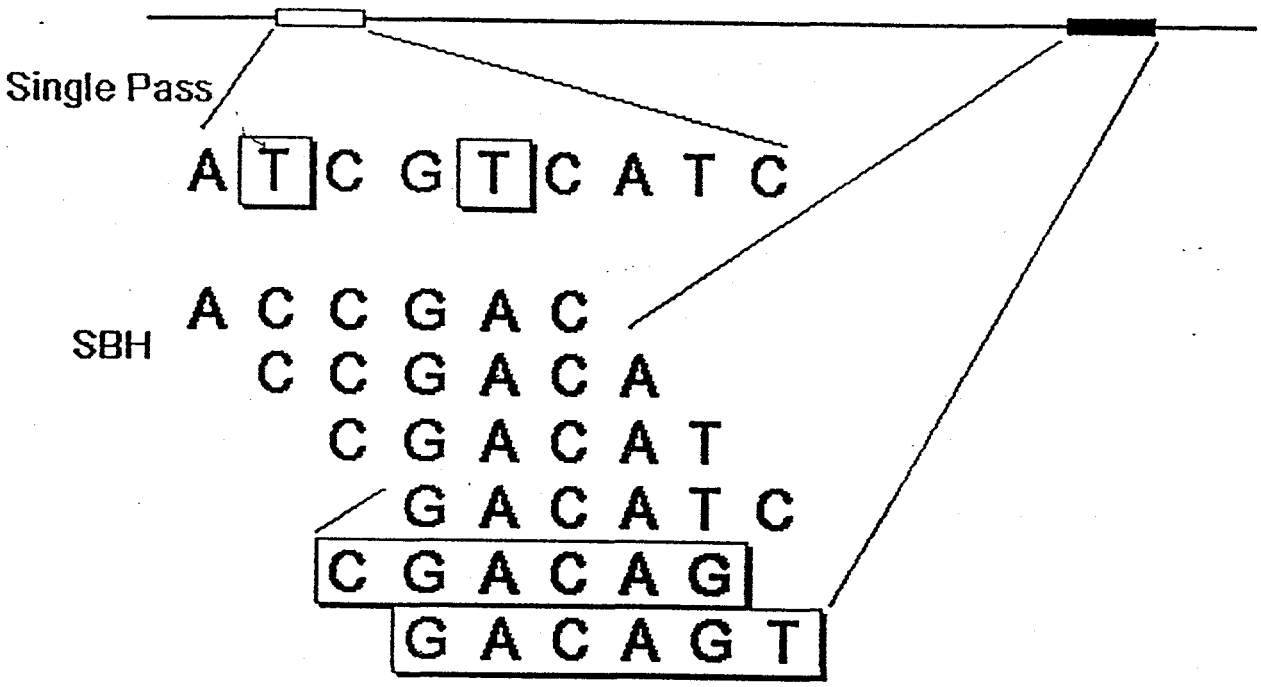


Figure 3


GGGGAGCG
3.0K
2.0K
1.0K
1
2
3
4
5
6
7
8




Single Pass
A T C G T C A T C
SBH
A C C G A C
C C G A C A
C G A C A T
G A C A T C
C G A C A G
G A C A G T