AD_____

GRANT NUMBER: DAMD17-94-J-4414

TITLE: cDNA Fingerprinting of Breast Cancer Tumor Cells

PRINCIPAL INVESTIGATOR: Cassandra L. Smith

CONTRACTING ORGANIZATION: Boston University
Boston, MA 02215

REPORT DATE: October 1996

TYPE OF REPORT: Annual

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

DISTRIBUTION STATEMENT: Approved for public release;
distribution unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave Blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	30 October 1996	Annual Report 30 Sep 95 - 29 Sep 96

4. TITLE AND SUBTITLE

cDNA Fingerprinting of Breast Cancer Tumor Cells

5. FUNDING NUMBERS

Contract DAMD17-94-J-4414

6. AUTHOR(S)

Dr. Cassandra L. Smith

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Dr. Cassandra L. Smith
Center for Advanced Biotechnology
Boston University
36 Cummington St.
Boston, MA 02215

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Material Command
Fort Detrick, MD 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b.

13. ABSTRACT (*Maximum 200 words*)

Recently, several methods have been described that highlight genetic differences between samples. However, none are robust enough to be applied routinely to a large number of samples. The long term goal of this proposal is to develop and to apply such methods to breast cancer. One set of experiments uses genomic DNA as an alternative to clone libraries for gene hunts. This is important, because clone libraries are expensive and time consuming to make and maintain, and cloned DNA may not maintain the true genomic organization. These experiments have identified over 60 candidate genes in the q13 region of chromosome 20 amplified in breast cancer tumor cells. A second set of experiments is designed to search for genomic rearrangements relevant to breast cancer in a parallel and efficient manner.

Other work focuses on using Matrix assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF MS) to monitor gene expression. Here, an indexing scheme (an index is a very short DNA sequence) is used to array and analyze pools of cDNA fragments. Each index is detected by MALDI-TOF MS which yields molecular masses with up to 1 Dalton accuracy and 1 part in 1000 resolution. Our goal is to bring the power of MS to cDNA analysis in breast cancer so that the quantum leap in efficiency it provides will allow the overall pattern of gene expression to be studied routinely as a new molecular monitor of cell physiology.

14. SUBJECT TERMS	15. NUMBER OF PAGES
breast cancer, cDNA, PCR, chromosome 20, pfg	12
	16. PRICE CODE

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

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

__X__ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

__X__ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

__X__ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

__X__ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

__X__ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

CL Smith 10/15/96

PI - Signature Date

Abstract:

Recently, several methods have been described that highlight genetic differences between samples. However, none are robust enough to be applied routinely to a large number of samples. The long term goal of this proposal is to develop and to apply such methods to breast cancer. One set of experiments uses genomic DNA as an alternative to clone libraries for gene hunts. This is important, because clone libraries are expensive and time consuming to make and maintain, and cloned DNA may not maintain the true genomic organization. These experiments have identified over 60 candidate genes in the q13 region of chromosome 20 amplified in breast cancer tumor cells. A second set of experiments is designed to search for genomic rearrangements relevant to breast cancer in a parallel and efficient manner.

Other work focuses on using Matrix assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF MS) to monitor gene expression. Here, an indexing scheme (an index is a very short DNA sequence) is used to array and analyze pools of cDNA fragments. Each index is detected by MALDI-TOF MS which yields molecular masses with up to 1 Dalton accuracy and 1 part in 1000 resolution. Our goal is to bring the power of MS to cDNA analysis in breast cancer so that the quantum leap in efficiency it provides will allow the overall pattern of gene expression to be studied routinely as a new molecular monitor of cell physiology.

Introduction

The objectives of this proposal are to identify important changes at the DNA or RNA levels associated with specific breast cancer characteristics. These changes may occur through point mutations, or larger DNA rearrangments on amplifications. Towards this end, we have focused on three approaches. In each case our plan is to develop and validate new methods, and then to apply these to a comparative study of five available breast cancer cell lines and tumors derived from these cells. The technique development is a broad program, and this grant pays only a portion of the total cost. The specific application of the new methods to breast cancer is funded only by this grant. The three methods are:
(1) the use of genomic DNA directly for positional cloning experiments for a region of chromosome 20 amplified in breast cancer tumor cells; applications of this method to breast cancer cells are well underway
(2) the differential display of genomic DNA restriction fragments *en masse* as a means of displaying genomic differences between normal and breast cancer tumor cells; these methods have been further refined during the past year, and they are now ready for application to breast cancer cell lines
(3) the analysis of cDNA arrays as a means of quantitating differential gene expression in normal and tumor cells. This part of the project has been redefined, because of considerable progress in the fields of DNA mass spectrometry and high throughput cDNA analysis

In positional cloning experiments, conventional genetic methods are used to narrow the search region. Then physical (molecular) methods are applied to further narrow the search region and to identify genes within the search region. Recently, these methods were used to isolate BCRA1 (Harshman *et al.*, 1995) and BCRA2 (Tavtigian *et al.*, 1996). The Human Genome Project has provided an increasing number of resources for finding diseases genes using positional cloning methods. Still, the task of finding genes involved in particular diseases is arduous. Multiple physical methods for identifying genes must be used in each gene search, because no single approach would guarantee the identification of all genes in a particular region.

Here, we have developed physical methods that allow us to use genomic DNA directly in place of large clone libraries during positional cloning experiments. This is important because the large genomic clones now in use have rearrangements and deletions. Many of the clones are chimeric; they contain DNAs from different genomic regions. Large insert clone libraries are time consuming and expensive to make and maintain. In contrast, our genomic DNA method can be rapidly applied to any DNA sample. Thus, our approach is quite useful in positional cloning searches to access a specific genomic region in a particular DNA sample. Our particular focus has been on the q13 region of chromosome 20 known to be amplified in many breast cancer tumor cells.

Thus far the positional genetic approaches have only identified major gene causes. However, the onset or progress of many diseases is governed by multigenic effects and interactions. Even major disease genes are not expressed alone but in a chorus of over 80,000 other genes. Given the spectrum of genomic changes thus far indentified in breast and other cancers, it is quite clear that efficient and reliable methods are needed to analyze the increasing number of genomic sequences important in tumor development, progression and response to

therapeutic regimes. Thus, a number of groups, including us, are focused on developing comparative methods for identifying multi-gene differences between samples that can be applied in a cost effective method to a large number of samples.

Although the published methods for multigene analysis are useful as research tools, none have proven to be robust enough to be routinely applied to samples that have the complexity of the human genome. The approaches include comparative genome hybridization (CGH; Kallioniemi *et al.*, 1994), differential display (Liang and Pardee, 1992; Liang *et al.*, 1994) and subtractive hybridization (Lisitzyn *et al.*, 1993a; Lisitzyn *et al.*, 1993b). In CGH, a mixture of differentially labeled cDNAs from two samples is hybridized to metaphase chromosomes. Genomic regions that are amplified or deleted in one of the test samples will be differentially labeled. Hence, this method can identify genomic regions important in disease states. In differential cDNA display experiments, mRNA levels of appropriate samples are analyzed. Here, total mRNA is amplified randomly and displayed by size, electrophoretically, from different appropriate samples. The differentially expressed cDNAs are then isolated and chacterized. In subtractive hybridization, sequences present in one cDNA library but missing in a second cDNA are isolated.

An alternative method of measuring mRNA level is the random sequencing of cDNA libraries made from particular cells. Although several pharmaceutical groups with a large number of resources are taking this approach for some diseases, it is quite clear that DNA sequencing costs at this time preclude the use of this method for routine application. Our original proposal intended to extend the principles of CGH to arrays of cDNAs. Since this proposal was written two methods for differential display of cDNA were described. One method (Schena *et al.*, 1995) is very similar to that described in the original proposal. The method involves hybridization of differentially labeled cDNA simultaneously to the same array of cDNA probe samples. Schena *et al.* (1995) reported on the application of CGH principles to arrays of yeast cDNAs. We also carried out a number of pilot studies on several arrays of cDNA. The other method (Velculescu *et al.*, 1995) to quantitate gene expression uses direct DNA sequencing of chimeric small clones that are composed of ligated pieces of cDNAs. Each of the ligated pieces is an index for a particular cDNA. Thus, one sequencing reaction gives information about many cDNAs. The chimeric clones are created in a manner that should preserve quantitative information on the occurrence of each cDNA.

These publications prompted us to rethink our cDNA profiling method. In particular, we are developing a hybrid system using matrix assisted laser desorption/ionization (MALDI) mass spectrometry (MS) as a tool for rapid, cost-effective, comparative studies of cDNAs fragments after specific hybridization capture steps to simplify the mixture of fragments. Recently we and others (Pieles *et al.*, 1993; Roskey *et al.*, 1996) showed that MALDI-MS is an effective tool for the rapid measurement of short (<35 nucleotide) DNA sequences. We have begun to develop the necessary simulation and (data) analytical software tools to adapt MALDI-MS as a method for measuring and characterizing genetic expression. An indexing scheme will be used to identify the cDNA strands corresponding to any given mRNA, or known sequence. Short (n = 8 - 15 nucleotides) sequences are used as identifiers. Such an identifier is capable of identifying 4^n different species. This should provide sufficient indices such that the majority of cDNA species are uniquely represented. Relative percentage abundances of cDNA species must retain those of the corresponding mRNA. The experimental program is investigating methods for maintaining

this quantitative information.

Body

(1) The use of genomic DNA directly for positional cloning experiments on a region of human chromosome 20 amplified in breast cancer tumor cells.

We have developed methods that allow us to use pulsed field gel- (PFG: Schwartz *et al.*, 1983; Schwartz and Cantor, 1994) fractionated genomic restriction fragments as a direct source of DNA (Bukanov, manuscript in preparation). Genomic DNA that has been cut with a restriction enzyme is fractionated by PFG under appropriate conditions. The gel lane containing DNA is cut into 2 mm slices. Each slice is melted in a solution containing 20 mM of ethanolamine by heating to 95° C 15 min. These samples can be stored indefinitely. The DNA in agarose can be used as a template in a number of reactions including PCR. For instance, PCR reaction can be used to test for the presence of particular STS's in slices.

We have used the DNA contained in slices to analyze a region of chromosome 20 amplified in breast cancer tumor cells. The experiments used genomic DNA from a monosomic hybrid cell line containing human chromosome 20. STS analysis of 22 sequences identified slices containing DNA from the amplified region. Then, long inter-*Alu* PCR was used to amplify and ^{32}P-label human DNA from the amplified region. The labeled DNA was used as a hybridization probe to screen a heterogeneous nuclear (hn)cDNA library. About ninety clones were identified that hybridized to this region. Other available genomic resources (e. g. cloned sequences) were also used as hybridization probes. Eight clones with high intensity hybridization signals were sequenced. Then, STS PCR primers were designed, and gel slices and available large insert clones in the amplified region were tested for the occurrence of the selected sequences. The results of these experiments indicate that the majority of these test clones come from the selected chromosomal region. This confirms other experiments done in collaboration with Joe Gray using FISH (flourescent *in situ* hybridization) that demonstrated that our gel slices provided region - specific DNA. Ongoing experiments are sequencing the remainder of the isolated clones. Then PCR primers will be designed and the location of these clones on the q13 region of chromosome 20 will be determined.

With this resource of DNA fragments from the region completed, we will be able to make detailed physical maps of the region in each of the available cell lines (SKBR-3, BT-474, UACC812, MCF7, and MDA157) and define the nature and extent of genome amplifications or other rearrangements that are present. The expression of genes encoding these seqences will also be assessed by Northern hybridization experiments using RNA isolated from the five breast cancer tumor lines. Of course, if any of the sequences looks promising as a possible gene for direct involvement in breast cancer, we will attempt to complete a full length cDNA sequence either by piecing together fragments already existing in publically accessible databases, or through collaboration with others possessing access to suitable materials or additional sequence information.

(2) Differential display of genomic DNA restriction fragments *en masse* as a means of displaying genomic differences between normal and breast cancer tumor cells.

We have developed a genomic differential display method that allows us to compare genomic DNA directly (Broude *et al.*, submitted). The method reduces genome complexity by capturing genome subsets (i. e. restriction fragments) that contain a targeted interspersed repeat. The captured fragments are labeled with fluorescein and fractionated by size on an automated DNA sequencing instrument. For this method to be generally useful, it must be quantitatively reproducible. Thus the sample preparation, the PCR amplification, and the final display must be performed very accurately and carefully. We have investigated the effect of different capture and labeling methods on the complexity and reproduciblity of the fragment display. For instance, the method used now captures DNA fragments that contain $(CAG)_n$ repeats by hybridization to an immobilized complementary probe. The captured fragments are released and then labeled in a PCR reaction using primers complementary to a known sequence that has been ligated onto the ends of the fragments and in some cases a primer that is complementary to the repeat sequence. Substantial differences in the pattern of fragments displayed are observed, depending on which primer contains the flourescent primer.

Ongoing experiments are focused on understanding these differences.
We are exploring further the variables that affect the reproducibility of our genomic differential display method. We are also developing methods for automatically analyzing the similarities and differences in our display methods. This will soon allow us to evaluate different experimental approaches and to determine the level of differences between samples. The approach that is used now captures genomic restriction fragments containing a targeted interspersed repeated before PCR amplication. We will explore simplifications of this procedure. For instance, we will test whether such fragments can be selectively amplified from genomic DNA directly without the capture step. During the next year genomic display will be used to compare the five available cell lines and then tumors derived from these lines. If time permits, we will also test these methods on cDNAs. Differential genomic display is especially powerful when two very closely related samples are available for analysis. This method shoudld find a natural application when normal and tumor tissue, or different stages of tumor tissue from the same individual are compared.

(3) The analysis of cDNA arrays as a means of quantitating differential gene expression in normal and tumor cells using MALDI-TOF MS.

Indexing techniques are being widely investigated for use in quantification of gene expression. Each method for preparation and selection has its own idiosyncrasies. However, the underlying steps are the same. Generation of an expression profile involves the following:
(1) the creation of cDNA samples using reverse transcriptase,
(2) an index of 10-15 nucleotides within each cDNA is isolated,
(3) PCR amplification of all of the indices is carried out in parallel, and
(4) the relative abundances of the cDNA indices chosen for each cDNA are measured. The key advantage of indexing is that the PCR amplification is carried out after all of the cDNAs have been reduced to short, same sized DNA fragments. This ought to improve the accuracy of the relative abundance information markedly. The challenge is finding a way to simplify the analysis of the enormous amount of data contained in a full set of indices.

The human genome is estimated to have more than 80,000 genes. The total number of expressed genes in a given cell is unknown but estimated to be several thousand. It is also

possible that most if not all genes are expressed in all cells, albeit at very low levels. Thus, it is quite clear that the number of possible indices and ways of generating them are quite numerous. Thus, we have begun to develop the necessary software tools, simulational and (data) analytical, that are need for developing and testing the various experimental approaches.

Some method of data reduction must be developed in order to reduce the complexity of measuring, potentially, thousands of signals. The data reduction step will be be an integral component of the indexing scheme. We will use array hybridization to simplify the mixture of index fragments. Thus, our method in esence combines some features of both indexing as originally suggested by Velculescu *et al.* (1995) with procedures used after more traditional rtPCR as described by Kato (1995, 1996) and Unrau and Deugau (1994). Each index fragment will be generated such that one (single indexing: SI) or both (double indexing:DI) ends have a single-stranded overhang. In each case, one end of the fragment will be hybridized to a spatially separated array of fixed hybridization probes; each probe has a unique single-stranded overhang, and each is analyzed separately by MS. The fixed probe array contains 4^m elements, where m is the number of nucleotides in single-stranded overhang. Our experiments (Broude *et al.*, 1994; Fu *et al.*, 1995) have shown that this greatly reduces the probability of mismatches between the anchored probes and their targets.

Further differentiation of cDNA species is dependent upon whether SI or DI indexing is used (See Appendix - Figure 1). In SI, further differentiation is obtained through mass measurement. In this protocol, only one strand (length N) of the cDNA is analyzed in the MS. Since, m nucleotides are known from the position in the array, this leaves N-m = k nucleotides be determined by MALDI MS. In a DI approach, a mixture of specifically designed floating probes is hybridized to the second single strand overhang after the cDN fragment has been hybridized into place in the array. For quantitative analysis, competitive hybridization can be used with a mass-labeled set of standards for each array elemeht.

Simulation experiments will guide and optimize the accompanying experimental program which will be focused on examining the most serious error sources
(1) accuracy of mass measurement by MALDI MS,
(2) hybridization of slightly mismatched probes,
(3) the quantitative represention of mRNAs by the RT-PCR generated cDNAs, and
(4) the coincident occurrence of identical or nearly identical mass labels on different mRNA species.

The experimental program will use test sequences from the ever increasing number of genes that have been identified in regions that are known to amplified or deleted in breast cancer. Thus far, we have compiled a list of over 50 such genes located all over the genome. In addition, there are a number of cDNAs that are known to be differentially expressed in breast tumor cells (Sager *et al.*, 1994). These genes will make up our test system since differential display has already been used to assess the level of these genes in about 20 different breast cancer cell lines and primary tumor cells.

Conclusions

Great progress has been made in isolating expressed genes from a region on human

chromosome 20 amplified in a large number of breast cancer tumor cells. These clones are now being sequenced and individually characterized in a large number of breast cancer tumor cells.

The major progress on genomic profiling entails the realization that the originally proposed method is not as powerful as newly developing MS methods. Thus, we decided to take a very forward looking approach to cDNA profiling, rather than use the current inefficient methods. Fortunately, the basic methods of DNA handling are the almost the same as those proposed in the original grant. Specific adaption to MS is now being done.

References

Brenner, S. and Livak, K. J. 1989. Proc. Nat. Acad. Sci., 86: 8902-8906.

Broude, N., Sano, T., Smith, C. L., and Cantor, C. R. 1994. Proc. Nat. Acad. Sci. USA 91: 3072-3076.

Broude, N., Lykyanov, S. A., Graber, J. H., Sverdlov, E. D., and Smith, C.L. 1996. Submitted.

Ermolaeva, O. D. and Sverdlov, E. D. 1995. Genetic Analysis (Biomolecular Engineering) 13: 49-54.

Fu, D. J., Broude, N. E., Koester, H, Smith, C. L., and Cantor, C. R. 1995. Genetic Analysis (Biomolecular Engineering) 12: 137-142.

Harshman, K., Bell, R., Rosenthal., J., *et al.*, 1995. Hum. Mol. Genet. 4: 1259-1266.

Kallioniemi, A., Kallioniemi, O.-P., Piper, J., Tanner, M., Stokke, T., Chen, L., Smith, H. S., Pinkel, D., Gray, J. W., and Waldman, F. M. 1994. Proc. Natl. Acad. Sci. USA 91: 2156-2160.

Kato, K. 1995. Nucleic Acids Res. 23: 3685-3690.

Kato, K. 1996. Nucleic Acids Res. 24: 394-395

Liang, P. and A. B. Pardee 1992. Science 257: 267

Liang, P., Zhu, W., Zhang, X., Guo, Z., O'Connell, Averbukh, L., Wang, F. and Pardee, A. 1994. Nucleic Acids REs. 22, 5763-5764.

Lisitzyn, N., Lisityn, N., and Wigler M. 1993a. Science 259, 946-951.

Lisitzyn, N., Rosenberg, M., Launer, G., Wagner, L., Potapov, V., Kolesnik, T., and Sverdlov, E. 1993b. Mol. Genet. Microbiol. and Virusol. 3, 26-29.

Sager, R, Sheng, S., Anisowixz, A., Sotiropouloou, G., Zou, A., Stenman, G., Swisshelm, Z., Chen, M., Hendrix, J.C., Pemberton, P., Rafidi, K., and Ryan, K. 1994. Cold spring Harbor Symposia on Quant. Bio. LIX: 537-545.

Schena, M., Shalon, D., Davis, R. W., and Brown, P.O. 1995. 270: 467-70

Smith, D. R. 1992. PCR Meth. Appl., 2: 21-27

Tanner, M., M. Tikkonen, A Kallioniemi, C. Collins, T. Stokke, R. Karhu, D. Kowbel, F. Shadravan, M. Hlntz, W.L. Kuo, f. M. Waldman, J. Isola and J. W. Gray. 1994. Cancer Research 54: 4257-4260.

Tavtigian, S. V., Simard, J., Rommens, J., *et al.* 1996. Nat. Genet. 12: 333-337.

Unrau, P. and Deugau, K. V. 1994. Gene 145: 163-169.

Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. 1995. Science 270: 484-487.

Appendices

Figure 1. SI and DI Approaches to MALDI MS cDNA profiling.

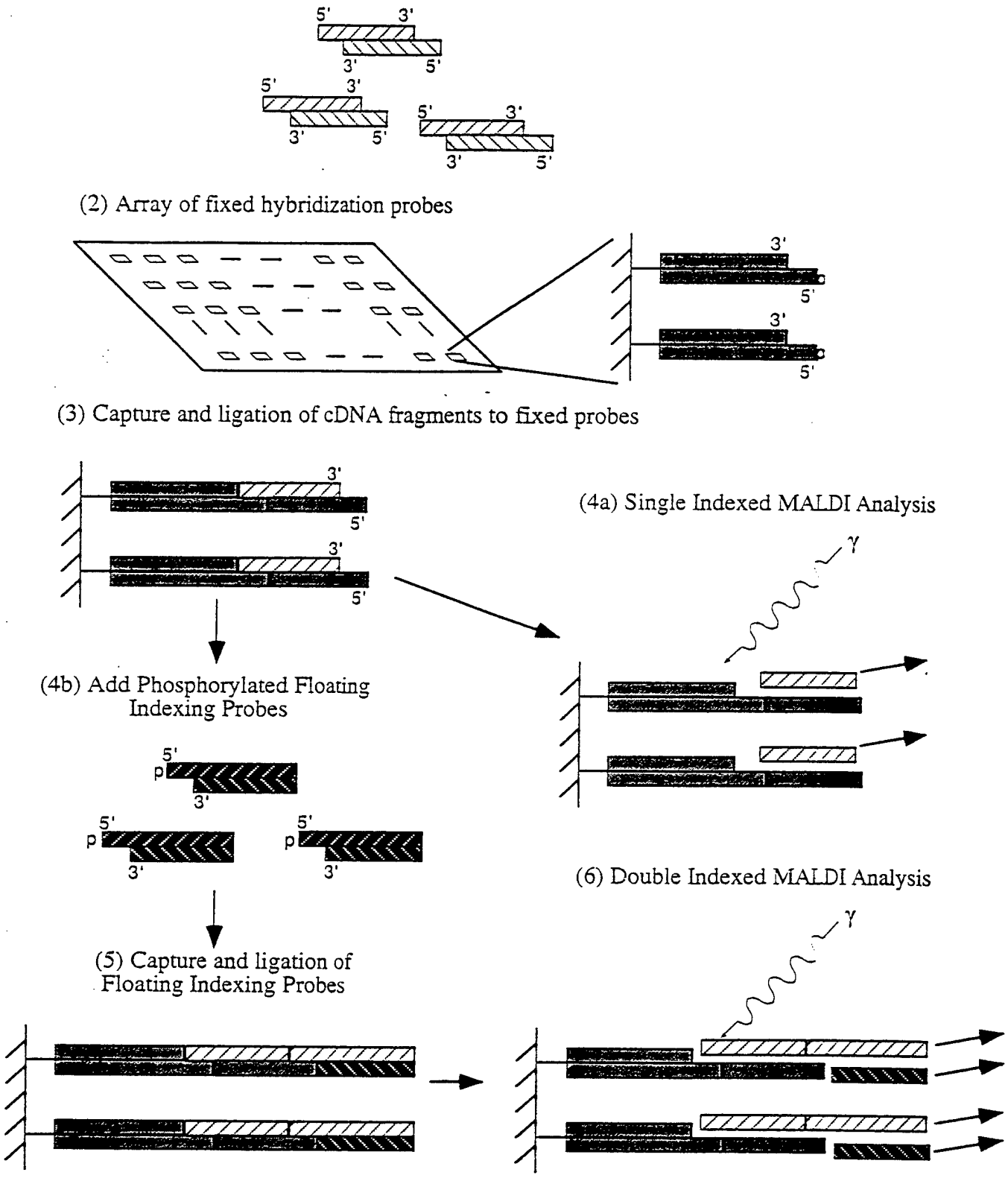

Figure 1. Graphical representation of MALDI measurement of genetic expression, in both the single indexing and double indexing schemes. Pattern changes in fragments are used to show the joining together of fragments through ligation.