Comparative Genomics: Differential Display and Subtractive Hybridization

Cassandra L. Smith, Joseph Bouchard, Gregg Surdi, Giang Nguyen, K. Pimpis, and Linda G. Tolstoi
Boston University, Boston, USA

An understanding of genome function within the context of human cells requires new methodology. This article describes comparative approaches that are used for genomes and genes in situ. However, the major focus of the article is on differential display (DD) and subtractive hybridization (SH) because these two methods are most accessible to a large number of researchers. The basic principles of these methods are described along with some fundamental reminders about the important factors that must be kept in mind for successful experiments.

1 INTRODUCTION

There are an increasing number of genomic sequences becoming available. Within the next 5 years even the human genome sequence should be completed. Understanding the meaning of these DNA sequences will take much more work and involve many comparative experiments.

In the past, gene function was defined by comparing a cell containing a specific mutation with a cell that did not have the mutation. Mutations are also used quite effectively for dissecting biochemical mechanisms and pathways. The power of using mutations lies in the fact that a very well defined change is being studied. In the past, most research was confined to model organisms with well-developed genetic systems that allowed gene manipulations. This type of approach is quite useful for experimental organisms that have well-developed genetic systems which allow gene manipulation. It is quite clear that these types of experiments will continue in the future. However, it is also clear that we need to understand gene function in complex organisms where gene manipulation is not possible or easily done.

Comparative studies in the complex genomes are difficult. For instance, it is estimated that there may be over 100 000 genes in the human genome. Of course, there are many naturally occurring mutations that aid geneticists and other researchers to understand disease and other natural processes. However, even putting aside the ethics of manipulating human genomes for experimental purposes, humans are not good experimental organisms because their life span is too long. Hence, it is quite clear that comparative methods focused on complex organisms with long life spans must be independent of gene manipulations. Furthermore, even with ''good'' experimental organisms, the number of known genes and interactions are too great to be dissected by past methods using conventional recombinant DNA methods. In such experiments, genes are cloned, sequenced, modified and studied in detail out of their native genomic environments.

A large number of methods and variations have been developed for comparing whole genomes and entire repertoires of genes. Most methods can be applied to large and small genomes. The technical difficulty of a particular approach is usually a function of genome size. This article will review genomic comparative techniques, focusing on DD and SH because these methods are most accessible to a large number of researchers.

For simplicity, all calculations done will assume a random DNA sequence composition. However, for many organisms the amount of sequence information available in the databases allows for more accurate calculations of experimental results and for some completely sequenced genomes, a total modeling of the experimental results.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.)

2 OPTIONS FOR COMPARING GENOMES

There are several approaches for comparing, globally, genome structure and function. These studies can be divided into those that compare the primary structure of genomic DNA and those that compare gene expression profiles. It should be noted that gene expression profiling requires quantitative analysis of levels that vary by 10^5-fold. This means that quantitative studies on gene expression are more complicated than direct genome comparisons.

Gene expression profiling is currently most often done at the mRNA level. Such experiments do not provide information about post-transcriptional regulation of gene expression. Two-dimensional gel electrophoresis has been used for some time to establish and compare protein profiles from different samples. Some recent experiments compare mRNA and protein profiles from the same sample. In yeast, Gygi et al.[1] found little correlation between protein and mRNA expression levels. It will be interesting to see what post-transcriptional regulatory mechanisms are uncovered and how these initial observations are extended to other organisms and systems. These types of comparative experiments and protein profiling in general will not be discussed here.

2.1 DNA Fingerprinting Methods

Genomic sequence comparisons between different samples is the only reliable method to detect all differences. Gene expression profiling can be done using DNA sequencing methodology. In these approaches, random cDNAs are sequenced and the frequency of occurrence of specific sequences is recorded in different cells or under different conditions such as disease states. Some have termed this electronic profiling and several successful biotechnology companies specialize in this approach. These types of experiments form the new field of pharmacogenetics. (Pharmacogenetics promises to tailor treatment and medication to individuals.)

A variation on random cDNA sequencing is called serial analysis of gene expression (SAGE).[2] SAGE links short bits of cDNA sequence together in a single sequencing template to increase efficiency of the analysis. Even so, the financial cost of current technology prevents widespread use of large-scale sequencing experiments for global comparative studies.

DNA fingerprint methods have been used to compare both genomic DNA and cDNA sequences. The first method that was developed hybridized a multilocus simple, tandem repeat probe to restriction enzyme cleaved genomic DNA fractionated electrophoretically.[3] Since then a number of easier polymerase chain reaction (PCR) methods have been developed for DNA fingerprinting. In some cases, PCR is used to amplify unique sequences located between primer sequences composed of arbitrary primers (randomly amplified polymerase chain reaction, RAPD).[4,5] In other cases, the primer sequences consist of interspersed repeat sequences, for example Alu sequences in the human genome.[6] The pool of PCR amplified fragments are size fractionated electrophoretically to create a complex DNA fingerprint. The DNA fingerprints of two or more samples are compared to assess the similarity and differences between samples.

DD[7] compares the size distribution pattern of DNA fragments generated from random amplification of two mRNA samples. Here, we will distinguish cDNA differential display (cDD) from genomic differential display (GDD). A more detailed discussion of these methods is given below.

Methods such as sequencing by hybridization (SBH, also known as DNA chip arrays)[8,9] replace time-consuming electrophoretic size fractionations with a single-stranded DNA array composed of different sequences. Hybridization of a single-stranded sample to a DNA array produces a DNA fingerprint revealing which array sequences are complementary to a position of the sample DNA. In conventional SBH, the presence but not the location of the sequence is obtained. The sample sequence may be reconstructed from the pattern of short oligonucleotides that hybridize to the template. An alternative method, called positional sequencing by hybridization (PSBH),[10] revealed both the presence and location of the complementary sequence. In this protocol, the array targets are partially double stranded and have a single-stranded end. In this format, the single-stranded end of the target sequence hybridizes to the probe element with stacking interactions from the duplex region insuring a high level of matching bases. The specificity of this hybridization may be increased by adding enzymatic steps, such as a DNA ligation and/or polymerase extension reaction.

In SBH, oligonucleotides are typically immobilized in an array format on a silicon chip to which a labeled template is hybridized. The presence of the template on each array element is scored as positive or negative when conventional radioactive, fluorescent or chemiluminescent labels are used.

Conventional SBH experiments have been severely handicapped by the hybridization of mismatched sequences, especially end mismatches. PSBH reduces the level of mismatches at least 10-fold. Further enhancement may be obtained when SBH experiments are analyzed by mass spectrometry (MS).[11]

MS analysis does not require a label. MS not only detects the presence of a species as do the other detection methods but also provides the mass of the molecule. Since the masses of the bases are known, different

sequence compositions with distinguishable masses may be distinguished when they are hybridized to the same array element. Potentially, MS allows the experimental results to be precisely correlated to DNA sequence(s).

DNA arrays may also be composed of other molecules. For instance, cloned cDNAs,[12,13] or genomic sequences[14] or even uncloned single or pools of genomic restriction fragments[15] may be used as targets.

Comparative genome hybridization (CGH)[16] experiments hybridize pairs of differentially labeled genomic DNA or cDNAs to normal metaphase chromosomes. Usually, one sample is labeled with fluorescein and the second with 4′,6-diamidino-2-phenylindole (DAPI). The samples are mixed together at equimolar ratios. Hybridization to metaphase chromosome means that the DNA probes are sorted to chromosomal regions. The fluorescein/DAPI signal ratio will remain contant unless there is a region that is amplified or deleted in one of the genomic DNA samples. Likewise, when cDNA probes are analyzed, the fluorescein/DAPI signal ratio will remain constant except for those regions where there is a difference in gene expression. CGH provides a low resolution (~10 Mb) genomic localization of differences and does not provide directly the sequence causing the difference. The advantage is that a controlled pairwise comparison is done, providing a much more robust quantitative analysis. The disadvantage is that the complexities of this method do not allow it to be adapted by a large number of laboratories. A similar type of ''two-color'' approach has also been applied to analyzing DNA arrays.[17]

2.1.1 Isolation of DNA Differences Between Samples

All of the experiments discussed above may be classified as DNA fingerprinting methods. These methods present similarities and differences between samples. An alternative to these approaches are methods that provide information only about the differences between samples.

In SH, only ''candidate'' differences are provided to the experimenter. An advantage of the DNA fingerprinting methods are that they can be quantitative and provide ongoing information to the experimenter. An advantage of the SH experiments is that they provide directly the candidate differences between samples. The DNA fingerprinting experiments provide information about differences that then must be isolated for further testing. In both approaches, the first chore of the experimenter is to prove that the indicated differences are real. This will allow the level of false differences between samples (i.e. false positives) to be assessed. The level of false negatives (false similarities) between samples is more difficult to assess. Our own approach to this problem has been to verify our method using a model system where the experimental results can be compared with the theoretical results (see below) and where few, if any, differences should exist between samples.

DD and SH are technically demanding. The key to success lies in understanding the underlying biochemistry and biophysics. The goal of this article is to provide the user with a basic understanding of the principles of these methods and factors that affect reliability. Also discussed are some of the many options and applications.

3 DIFFERENTIAL DISPLAY

3.1 Principles

DD was originally developed by Liang and Pardee[18] as a PCR method for amplifying and labeling subsets of mRNA to identify differences in gene expression. A recent useful collation of a variety of DD articles can be found in Liang and Pardee.[7] The pool of amplified products <500 bp (base pairs) is size-fractionated on a high-resolution DNA sequencing gel. This produces a display consisting of the size-dependent banding pattern of the pool of PCR products. The display produced from two or more samples is compared to identify differences in gene expression patterns.

In the originally described method,[18] the starting material was purified mRNA which is converted to cDNA using conventional methods. Then PCR was used to amplify subsets of the cDNA library. One of the PCR primer pairs was complementary to the polyA tail of eukaryotic mRNA with additional 3′-anchor bases (e.g. $5'T_{11}CA3'$). The use of a homopolymer T primer with a unique dinucleotide 3′-end serves two purposes. First, the unique dinucleotides anchored the primer to the unique genomic sequences adjacent to the 5′-end of the polyA mRNA tail. This insured the polymerase extension reaction initiated at the sequence adjacent to the polyA tail and not randomly within the mRNA polyA tail. The anchor bases also control the complexity of the PCR amplification products. Twelve different dinucleotide (e.g. CA, CT, CG, CC, GA, GT, GG, GC, AA, AT, AG, AC) anchors are needed to amplify all mRNA specifically. In a random sequence, a dinucleotide anchored primer sequence would reduce the number of PCR amplification with a specific dinucleotide anchor products by 1/12.

The second PCR primer was a 10-mer composed of an arbitrary sequence (e.g. Ltk3, CTTGATTGCC). There are 1049 different possible 10-mer combinations and any one 10-mer sequence would be expected to occur once in 10^6 bp. Since an average mRNA is about 10^3 bp in length, a single 10-mer should hybridize to 1 in 10^3 cDNAs.

DD analysis is usually restricted to fragments less than 10^3, hence the number of amplified products should be even less. The arbitrary primers used in these experiments

were chosen after testing to give reproducible and specific amplification products with the polyT anchored primers. Subsequent studies have taken a more critical look at the theoretical distribution of primers so that the speed and efficiency of DD could be maximized.[19]

3.2 Options

3.2.1 Genomic Analysis

Genomic comparisons of simple genomes may be done directly. For instance, it is quite easy using pulsed field gradient (PFG) techniques to compare chromosomal fingerprints of small genomes like bacteria, yeast and protozoan.[20] The resolution of these methods can be improved by establishing fingerprints using restriction enzymes that cleave infrequently and in some cases restriction enzymes that cleave frequently. These types of approaches can be applied to organisms whose genomes range in size up to about 100 million bp. In contrast, the human genome is 1000-fold greater than this. With today's technology it is still impossible to compare directly total human genomes. Instead genome complexity must be reduced.

We[21–25] and others[26–28] have reported on the use of targeting in DD. In this approach, a class of fragments containing a specific sequence (a target sequence) is selected for analysis (see below). In contrast to others, our methods have focused on analyzing selected genomic fragments rather than cDNAs (targeted genomic differential display (TGDD) vs targeted cDNA differential display (TcDD) respectively). The focus on genomic DNA allows for an exact modeling of the expected results using DNA from genomes whose entire sequence is known and was easier to develop as a quantitative method because of the smaller dynamic range requirements (e.g. ~2-fold versus 10^5-fold for genomic vs cDNA studies, respectively).

Two methods for TGDD were established by us.[29] In both methods, oligonucleotides of known sequence are ligated to the ends of genomic restriction fragments. In Method I, there is an initial hybridization – capture step which is used to purify the restriction fragments that contain the target sequence away from the remaining restriction fragments. An immobilized oligonucleotide complementary to the target sequence is used for the hybridization – capture step. Subsequently, the captured fragments are amplified by PCR. A single PCR primer complementary to the adapter sequences ligated onto the ends of the genomic fragments may be used. Alternatively, the adapter primer may be used with a primer complementary to the target sequence. In the former case, sequences from each side of the target are amplified; in the latter case the sequence from only one side of the target is amplified.

Method I eliminates the capture step and uses long (40-mer) adapter primers to enhance self-annealing of the adapter primers, an effect called PCR suppression.[30] In this method, a primer to the target sequence and a short 20-mer primer to the adapter sequence are used. Only fragments that contain a target sequence can be amplified. This version of the method has the advantage that it is easier to perform. Although, the adaption of interspersed repetitive element-bubble PCR[31] with Method I should prevent self-annealing of the adapter sequences,

Targeting has been adapted to cDNA studies (Nguyen et al., unpublished results). In TcDD, the adapter primer was replaced by a primer complementary to the polyA tail.

3.2.2 Targeting

A detailed discussion of targeting issues can be found in Bouchard et al.[29] In brief, it is important to realize that targeting can be done with a simple repeating sequence (the polyA tail of eukaryotic cDNA is one example of this) or more complicated sequence coding for a protein motif, a *cis*-acting sequence such as transcriptional activation signal. The method can analyze the target sequence itself and/or the surrounding sequence depending on the PCR primer that is used. Usually, the target is interspersed throughout the genome. This means that the method samples sequences spread throughout the genome. In some cases the target may be more confined to a specific chromosomal region, for example telomeric sequences. In the case of simple repeating sequences, unique bases added to the 3′- or 5′-end of the primer serves to anchor the primer to the target end and to reduce complexity (see above).

The question of how to design more complex motif primers that target gene families is not yet answered. Most, if not all, gene families consensus sequences were developed comparing amino acid compositions and not DNA sequences. In fact, very little work has been done on characterizing gene families at the DNA sequence level because it is clear that there can be much more variability at the DNA sequence level because of the redundancy of the genetic code. Additionally, different preferential codon usage is observed in different organisms and an emerging complication is that codon usage in a particular species may reflect the expression level of the protein sequence and/or be protein class specific.

3.3 Trouble Shooting

The major problem with DD experiments is the high number of false positives. A suggested way of dealing with this issue is to reconfirm each alleged difference individually by hybridization. This requires that the

DNA fragment be isolated and/or sequenced so that a hybridization probe can be generated.

Our own experience with TGDD has suggested several ways of improving the experimental results. One of the most important factors is to analyze high-resolution Gaussian-like distributions of the fragment size fractionations. In most DD experiments, the sample is radioactively labeled and the fractionation results recorded by exposure of the polyacrylamide gel to film to produce a banding pattern. Although this type of analysis is useful, especially if the focus is kept on the differential presence of very dark bands, a much more realistic pattern of intensity changes is obtained when higher resolution data are collected in real time on an automated DNA sequencing instrument.

Our extensive characterization of the factors that influence the reliability of DD has led us to the conclusion that the most important aspect is to match the starting DNA concentrations. There are several reasons for doing this; in particular, the occurrence of unequal amplification of the PCR product during the late cycles. This problem is likely related to the specific sequence that needs to be amplified. For example, although the human genome is 60% A + T, usually equal amounts of the trinucleotide precursors are added to the PCR. This may mean that in each cycle of PCR there is a 20% difference in the use of trinucleotide precursors dATP and TTP vs dCTP and dGTP. Furthermore, the Michaels constant (K_m) of each trinucleotide is different. Hence, the cycle at which precursors become limiting will differ. Of course, this is only a theoretical consideration. The real inequities must be based on the base composition of each fragment in the amplified pools. Another cause of unequal amplification may be the rapid annealing of high concentration PCR products, which would then prevent the subsequent annealing or extension of primers. These considerations emphasize the importance of using samples with matched DNA concentrations and a low number of PCR cycles.

Another consideration in any experiment involving nucleic acids, and especially those involving DNA, is the accurate determination of the sample concentration. It is difficult to accurately determine DNA concentrations. In the cell, 90% of the nucleic acid is RNA. This means that the greatest detriment to accurately determining DNA concentration from a cell extract is contaminating RNA sequences. Although it is possible to purify DNA from RNA via density gradient centrifugations, this is not usually done because of the limiting amount of material that is available. RNA or DNA specific dyes can be used but they will bind to the alternative species with less affinity as pointed out by many of the manufacturers. The problem is that the amount of contaminating material is unknown. Hence, it is more useful to be consistent in the manner in which DNA concentrations are determined and correlating this with useful results. In fact, since the entire scientific community has this problem it might be best to try several initial (two-fold variations below and above the target) DNA concentrations. Additionally, one might consider varying the number of PCR cycles by two cycles around the target to minimize the number of experiments needed, especially when starting out.

An important issue that cannot be ignored is the ratio and concentration of the primers in the PCR. For small oligonucleotide primers, the annealing temperature will be concentration dependent in addition to being sequence dependent.[32] Other considerations in these types of multilocus are the frequency of occurrence of particular sequences in the template DNA as well as the sequence conservation. The best way to handle these variables is to determine the best ratio and concentration of primer pairs to use empirically. Typically we titrate multilocus primers against each other in the 0.3 to 2 mM range.

3.4 Applications

There are many applications for DD and an increasing number of publications are appearing using this technology. Most publications focus on cDNA applications. Our own studies using TGDD on closely matched genomes compare the genomes of twins, families and different tissues from the same individual. The goal of these experiments is to determine the effect of genome stability and heritability on phenotype.

4 SUBTRACTIVE HYBRIDIZATION

4.1 Principles of Subtractive Hybridization

The goal of SH experiments is to isolate a target sequence(s) (T) that is present in one sample, called the test DNA (S), and absent in a matched sample, called the driver DNA (D). There are a large number of published protocols. In all cases, a mixture of excess D mixed with S, is denatured and allowed to reanneal. In this reaction there are three single-stranded species which can be distinguished as shown in Figure 1.

The D and S DNAs can form parental homoduplexes or heteroduplexes with complementary strands. All

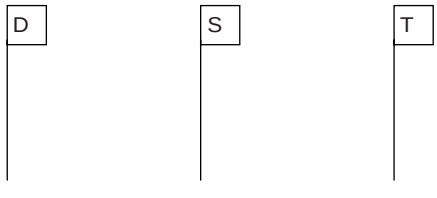

Figure 1 Single-stranded DNA.

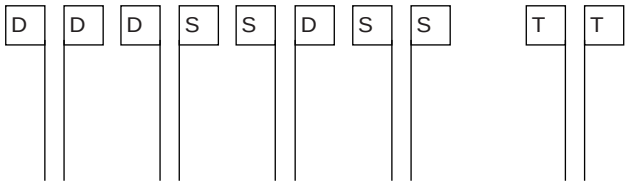

Figure 2 T DNA homoduplexes.

sequences are in common except for the T DNA. The T DNA can only form homoduplexes (Figure 2).

Equation (1) describes the kinetics of double-stranded DNA formation.[32]

$$f_{ds} = \frac{k_2 C_0 t}{1 + k_2 C_0 t} \tag{1}$$

where k_2 is the second-order rate constant, C_0 is the initial concentration of single-stranded DNA segments, and t is time. The fraction of DNA that has formed double strands, f_{ds}, can be calculated from the initial concentration as shown in Equation (2):

$$C_{ds} = f_{ds} C_0 = k_2 C_0^2 t \tag{2}$$

When this equation is applied to SH, two DNA samples must be considered. The first sample contains S- and T-type DNA; the second sample, D, contains only S-type DNA. Since T strands will only form double strands with other T strands the concentration of T : T, denoted C_{Tds}, can be determined from Equation (2). Equation (3) shows:

$$C_{Tds} = \frac{k_2 C_T^2 t}{1 + k_2 C_T t} \tag{3}$$

The formation of S : S, S : D, D : D can be determined by considering first the kinetics of D : D formation and then extracting the amount of S : S formed from the mole fraction of S : S, $X_{S:S}$ (see Equation 4). This can be done under the conditions where the concentration of S is negligible compared to the concentration of D.

$$X_{S:S} = \frac{C_{S1}^2}{C_{D1}^2} \tag{4}$$

where C_{S1} and C_{D1} denote the initial concentrations of S- and D-type strands respectively.

Equation (5) shows

$$(C_{Sds}) = (C_{D1}) \times X_{S:S} = \frac{k_2 C_{D1}^2 t \times C_{S1}^2}{(1 + k_2 C_{D1} t) C_{D1}^2}$$
$$= \frac{k_2 C_{S1}^2 t}{1 + k_2 C_{D1} t} \tag{5}$$

After the hybridization the next step is enrichment by PCR. Only the T : T and S : S strands have matching primers so only those strands will be exponentially amplified. This effectively removes S : D and D : D strands reducing the amount of S contaminants and enriching the amount of T strands. The ratio of enrichment resulting from the first subtraction round is given by Equation (6):

$$E = \frac{(C_{Tds})}{(C_{Sds})} = \frac{(k_2 C_T^2 t)}{(1 + k_2 C_T t)} \times \frac{(1 + k_2 C_{D1} t)}{k_2 C_{S1}^2 t}$$
$$= \frac{(C_T^2) \times (1 + k_2 C_{D1} t)}{(C_{S1}^2) \times (1 + k_2 C_T t)} \tag{6}$$

Since the T and the S strands come from the same genome they have the same concentration. Under this assumption the enrichment effectively simplifies to Equation (7):

$$E = \frac{(C_{Tds})}{(C_{Sds})} = \frac{C_{D1}}{C_T} \tag{7}$$

Therefore the enrichment ratio is increased or the amount of contaminating S strands is decreased simply by using a much higher concentration of D strands relative to T.

4.2 Options

Many variations and ''improvements'' on the SH concept have been published. Despite this, routine success has been elusive, although it is becoming more frequent. The major improvements in the technology included a minimization and simplifying the number of manipulations, use of PCR amplification, and developing more effective protocols through an improved understanding and modeling of the hybridization kinetics of a variety of approaches.

4.2.1 Polymerase Chain Reaction Subtraction

A commercial SH kit sold by Clontech is called ''PCR-Subtraction''. The principles but not the details of this procedure will not be presented here. (Details of this protocol can be found in material available from Clontech.)

This protocol was developed from modeling experiments of a variety of SBH protocols.[33–35] The authors considered several protocols: double-stranded DD and double-stranded SS (+TT); double-stranded DD and complementary single-stranded S (+T); single-stranded D and double-stranded SS (+TT); single-stranded D and complementary single-stranded SS (+TT). The initial modeling experiments focused on cDNAs and a single cycle protocol was developed. This method, called PCR subtraction, allows differential expressed genes to be isolated in a single round of subtraction for complex genomes and genomic differences to be isolated for microorganisms of genomes less than 10 Mb (Figure 3).

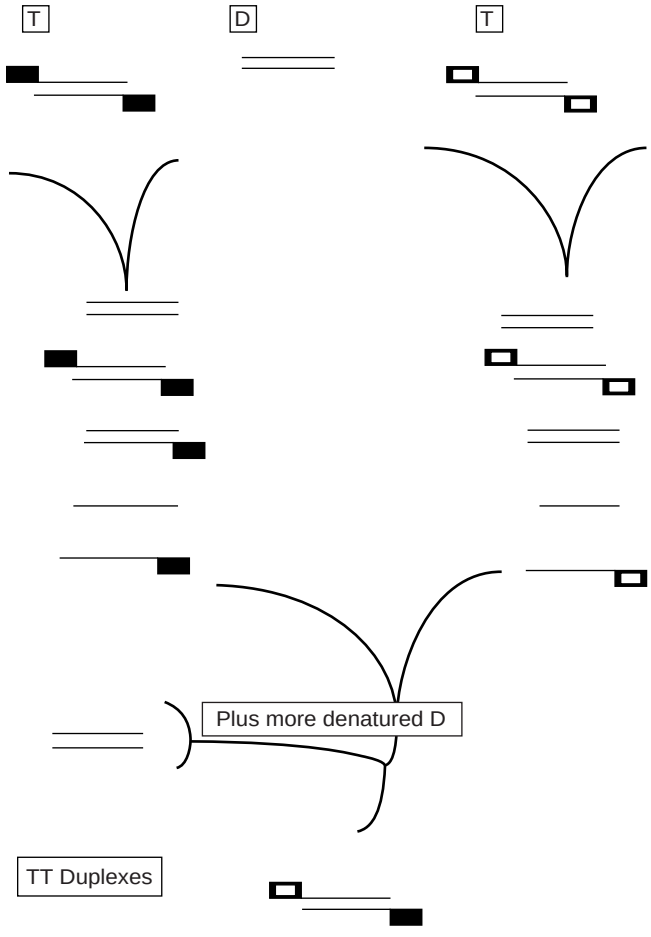

Figure 3 PCR subtraction.

In this protocol, first denatured 3 g D and 0.1 g S (+T) DNAs are annealed together in 1.5 L for 20 h at 68 °C. This allows an approximate equalization of the differentially expressed genes in the single-stranded fraction because of the second-order hybridization kinetics of the different cDNA abundance classes. In practice, two simultaneous initial SH reactions are carried out. The only difference in the reactions is that the S DNAs have different adapter sequences ligated onto their 5′-ends. The 5′-ends do not have a ligated adapter sequence because double-stranded unphosphorylated adapters were used. No adapters are ligated onto the D samples. The two reactions are mixed together and added, undenatured, to a new fraction of denatured D. During the second hybridization, the fragments shared between D and S are reduced further. The single-stranded T fragments in the two reactions will now form a TT with different 5′-end sequences. In a subsequent PCR the ends are filled in. Only duplexes with unmatched ends are amplified because the homoduplexes with complementary adapters self-anneal and preventing annealing of the shorter PCR primers, a phenomenon called suppression PCR.[30] This method takes advantage of rapid annealing of T DNA relative to the remaining S DNA in the single-stranded fraction.

4.3 Trouble Shooting

The major nonobvious difficulties in SH experiments, are similar to those encountered in DD experiments. Many of the problems can be alleviated by careful attention to DNA concentrations. However, a major difficulty is the assessment of DNA concentrations that are extremely critical and may best be handled empirically, especially when getting started.

4.4 Applications

There are many applications of SH experiments and progress is being made in identifying cDNAs differentially expressed. The major problems that need to be resolved are how to detect single base pair differences between samples and how to carry out total genomic subtraction experiments. In fact, more complex protocols have been developed using a complex reduction protocol with a gel electrophoresis step called in-gel competitive hybridization (IGCH)[36–38] and/or a preferential PCR amplification step called representational difference analysis (RDA).[39] In IGCH, excess D and T are mixed together, denatured and then fractionated using high-resolution electrophoresis. Individual size fractions are isolated from the gel and allowed to anneal.

5 OTHER CONSIDERATIONS

Most genome comparative experiments are not done on single cells because of the added difficulty of working with small amounts of DNA and in the case of genomic DNA, two molecules. Hence, the results present an average of several and usually, many cells. It is clear that it would be valuable to develop comparative methodology that allows the analysis of single cells. For instance, the analysis of single cells making up a tissue would provide information on the range of changes occurring in a tissue rather than the average of the changes.

There are several methods that can use the PCR to amplify DNA from low copy, including single molecules. One of the first methods was primer extension pre-amplification (PEP).[40] Reasonable success was reported with a primer pool of length $n = 15$, containing every possible 4^{15} primer. The concentration of any unique species is extremely low, but apparently high enough to insure amplification (e.g. 30 cycles) of most (estimated to be ~80%) of the genome after 50 cycles of PCR. A major

problem with this method is the potential for the primers to interact with each other.

An improvement to this method used a chimeric primer, with a variable portion and a conserved portion. This method was called tagged polymerase chain reaction (T-PCR)[41] because genomic fragments were tagged with a common adapter sequence during the PCR. In T-PCR, each primer had a 5′-end with a constant 17 base sequence and a 3′-end with a variable 9 base sequence. The first few rounds of PCR were done with the chimeric primer, then this primer was removed and the remaining PCR cycles were done with a primer solely composed of the constant region located at the ends of the tagged amplified fragments. A number of variations were developed, such as degenerative oligonucleotide-primed polymerase chain reaction (DOP-PCR).[42] Here the primer contains an internal variable portion (9 bases in length), a long 5′ constant region of 13 bases in length and a short 3′ constant region (3 bases in length). Cheung and Nelson[43] demonstrated that, with a slight modification, DOP-PCR could be used to amplify random genetic markers from small amounts of DNA.

6 PERSPECTIVE

Comparative genomics is a rapidly changing field. Perhaps the most important aspect in the application of a specific technology is adaptability. Not only is it important to stay advised of critical improvements in methods and conceptualization that create quantum improvement in the technology, but it is also important to be able to adapt new technologies as they become available.

The question of how one decides on a specific approach is more difficult. DD does not, in general, provide sequences of interest but rather shows you where they are so they may be isolated. DD provides quantitative information on the number and the extent of differences. The end results of an SH experiment are the sequences of interest. However, SH does not provide information about the number or the level of differences.

ACKNOWLEDGMENTS

This work was supported by grants NIH (1P50 HL55001) and DOA DAMD (17-94-J-414) to CLS.

ABBREVIATIONS AND ACRONYMS

bp	Base Pairs
cDD	cDNA Differential Display
CGH	Comparative Genome Hybridization
DAPI	4′,6-Diamidino-2-phenylindole
DD	Differential Display
DOP-PCR	Degenerative Oligonucleotide-primed Polymerase Chain Reaction
GDD	Genomic Differential Display
IGCH	In-gel Competitive Hybridization
MS	Mass Spectrometry
PCR	Polymerase Chain Reaction
PEP	Primer Extension Pre-amplification
PFG	Pulsed Field Gradient
PSBH	Positional Sequencing by Hybridization
RAPD	Randomly Amplified Polymerase Chain Reaction
RDA	Representational Difference Analysis
SAGE	Serial Analysis of Gene Expression
SBH	Sequencing by Hybridization
SH	Subtractive Hybridization
TcDD	Targeted cDNA Differential Display
TGDD	Targeted Genomic Differential Display
T-PCR	Tagged Polymerase Chain Reaction

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
DNA Arrays: Preparation and Application • Nucleic Acid Analysis in Clinical Chemistry

Liquid Chromatography **(Volume 13)**
Capillary Electrophoresis

Nucleic Acids Structure and Mapping **(Volume 6)**
Nucleic Acids Structure and Mapping: Introduction • Capillary Electrophoresis of Nucleic Acids • DNA Molecules, Properties and Detection of Single • DNA Probes • Fluorescence In Situ Hybridization • Polymerase Chain Reaction and Other Amplification Systems • Sequencing Strategies and Tactics in DNA and RNA Analysis

REFERENCES

1. S.P. Gygi, Y. Rochon, B.R. Franza, R. Aebersold, 'Correlation Between Protein and mRNA Abundance in Yeast', *Mol.Cell. Biol.*, **19**(3), 1720–1730 (1999).
2. V.E. Velculescu, L. Zhang, B. Vogelstein, K.W. Kinzler, 'Serial Analysis of Gene Expression', *Science*, **270**, 484–487 (1995).
3. A.J. Jeffreys, V. Wilson, S.L. Thein, 'Individual-specific 'Fingerprints' of Human DNA', *Nature*, **316**, 76–79 (1985).

4. J.G. Williams, A.R. Kubelik, K.J. Livak, J.A. Rafalski, S.V. Tingey, 'DNA Polymorphisms Amplified by Arbitrary Primers are Useful as Genetic Markers', *Nucleic Acids Res.*, **18**(22), 6531–6535 (1990).
5. J. Welsh, M. McClelland, 'Fingerprinting Genomes Using PCR with Arbitrary Primers', *Nucleic Acids Res.*, **18**(24), 7113–7218 (1990).
6. D.L. Nelson, S.A. Ledbetter, L. Cordo, M.F. Victoria, R. Ramirez-Solis, T.D. Webster, D.H. Ledbedder, C.T. Caskey, 'Alu Polymerase Chain Reaction: A Method for Rapid Isolation of Human-specific Sequences from Complex DNA Sources', *Proc. Natl. Acad. Sci. USA*, **86**(17), 6686–6690 (1989).
7. P. Liang, A. Pardee (eds.), *Methods in Molecular Biology, Differential Display Methods and Protocols*, Humana Press Inc., Totowa, Vol. 85, 1997.
8. S. Strezoska, T. Pauneska, D. Radosavljevic, I. Labat, R. Drmanac, R. Crkvenjakov, 'DNA Sequencing by Hybridization: 100 Bases Read by a Non-gel-based Method', *Proc. Natl. Acad. Sci. USA*, **88**(22), 10089–10093 (1991).
9. R. Drmanac, S. Drmanac, Z. Strezoska, T. Paunesku, L. Labat, M. Zeremski, J. Snoddy, W.K. Funhouser, B. Koop, L. Hood, R. Crkvenjakov, 'DNA Sequence Determine by Hybridization: A Strategy for Efficient Large-scale Sequencing', *Science*, **260**, 1649–1652 (1993).
10. N.E. Broude, T. Sano, C.L. Smith, C.R. Cantor, 'Enhanced DNA Sequencing by Hybridization', *Proc. Natl. Acad. Sci. USA*, **91**(8), 3072–3076 (1994).
11. H. Koster, K. Tang, D.J. Fu, A. Braun, D. Van den Boom, C.L. Smith, R.J. Cotter, C.R. Cantor, 'A Strategy for Rapid and Efficient DNA Sequencing by Mass Spectrometry', *Natl. Biotechnol.*, **14**(9), 1123–1128 (1996).
12. L. Wodicka, H. Dong, M. Mittmann, M.H. Ho, D.J. Lockhart, 'Genome-wide Expression Monitoring in *Saccharomyces cerevisiae*', *Natl. Biotechnol.*, **15**(13), 1359–1367 (1997).
13. D.J. Duggan, M. Bittner, Y. Chen, P. Meltzer, J.M. Trent, 'Expression Profiling Using cDNA Microassays', *Natl. Genet.*, **21**(1Suppl.), 10–14 (1999).
14. D. Grotheus, C.R. Cantor, C.L. Smith, 'Top-down Construction of an Ordered Schizosaccharomyces *pombe* Cosmid Library', *Proc. Natl. Acad. Sci. USA*, **91**(10), 4461–4465 (1994).
15. C.L. Smith et al., 'Cloneless Libraries: Long Range Mapping of Chromosome 20 and cDNA Selection for the 20q13 Amplified in Cancer Cells', in preparation.
16. A. Kallioniemi, O.P. Kallioniemi, D. Sudar, D. Rutovita, J.W. Grey, F. Waldman, D. Pinkel, 'Comparative Genomic Hybridization for Molecular Cytogenetic Analysis of Solid Tumors', *Science*, **258**, 818–821 (1992).
17. D. Shalon, S.J. Smith, P.O. Brown, 'A DNA Microarray System for Analyzing Complex DNA Samples Using Two-color Fluorescent Probe Hybridization', *Genome Res.*, **6**(7), 639–645 (1996).
18. P. Liang, A.B. Pardee, 'Differential Display of Eukaryotic Messenger RNA by Means of the Polymerase Chain Reaction', *Science*, **257**, 967–971 (1992).
19. D. Bauer, H. Muller, J. Reich, H. Riedl, V. Ahrenkiel, P. Warthoe, M. Strauss, 'Identification of Differentially Expressed mRNA Species by an Improved Display Technique (DDRT-PCR)', *Nucleic Acids Res.*, **21**(18), 4272–4280 (1993).
20. C.L. Smith, S. Klco, T.Y. Zhang, H. Fang, R. Oliva, J.B. Fan, M. Bremer, S. Lawrance, 'Analysis of Megabase DNA Using Pulsed Field Gel Electrophoresis', in *Methods in Molecular Genetics*, ed. K.W. Adolph, Academic Press, San Diego, 155–196, 1993.
21. N. Broude, A. Chandra, C.L. Smith, 'Differential Display of Genome Subsets Containing Specific Interspersed Repeats', *Proc. Natl. Acad. Sci. USA*, **94**(9), 4548–4553 (1997).
22. N. Broude, N. Storm, S. Malpel, J.H. Graber, S. Lykyanov, E. Sverdlov, C.L. Smith, 'PCR Based Targeted Genomic and cDNA Differential Display', *Genet. Anal.*, **15**(2), 51–63 (1999).
23. R.P. Oliveria, N.E. Broude, A.M. Macedo, C.R. Cantor, C.L. Smith, S.D.J. Pena, 'Probing the Genetic Population Structure of *Trypanosoma cruzi* with Polymorphic Microsatellites', *Proc. Natl. Acad. Sci. USA*, **95**(7), 3776–3780 (1998).
24. I. Lavrentieva, N. Broude, Y. Lebedev, I.I. Gottesman, S.A. Lukyanov, L. Smith, 'High Polymorphism Level of Genomic Sequences Flanking Insertion Sites of Human Endogenous Retroviral Long Terminal Repeats', *FEBS Lett.*, **443**(3), 341–347 (1999).
25. J. Bouchard, C. Foulon, G. Nguyen, N. Storm, C.L. Smith, 'Genomic Discordance in Mongozygotic Twins', in *Techniques in the Behavioral and Neural Sciences*, eds. W. Crusio, R. Gerlai, Elsevier, Amsterdam, 237–256, 1999.
26. R.P. Kandpal, G. Kandpal, S.M. Weissman, 'Construction of Libraries Enriched for Sequence Repeats and Jumping Clones, and Hybridization Selection for Region-specific Markers', *Proc. Natl. Acad. Sci. USA*, **91**(1), 88–92 (1994).
27. B. Stone, W. Wharton, 'Targeted RNA Fingerprinting: the Cloning of Differentially-expressed cDNA Fragments Enriched for Members of the Zinc Finger Gene Family', *Nucleic Acids Res.*, **22**(3), 2612–2618 (1994).
28. D. Graf, A.G. Fisher, M. Merkenschlager, 'Rational Primer Design Greatly Improves Differential Display-PCR (DD-PCR)', *Nucleic Acids Res.*, **25**(11), 2239–2240 (1997).
29. J. Bouchard, N. Storm, C.L. Smith, 'Enhancing Targeting Genomic and cDNA Differential Display Using a *Saccharomyces cerevisiae* Model System', in preparation.
30. P.D. Siebert, A. Chenchik, D.E. Kellog, K.A. Lukyanov, S.A. Lukyanov, 'An Improved PCR Method for Walking in Uncloned Genomic DNA', *Nucleic Acids Res.*, **23**(6), 1087–1088 (1995).

31. D.J. Munroe, M. Haas, E. Bric, T. Whitton, H. Aburatani, K. Hunter, D. Ward, D.E. Housman, 'IRE-Bubble PCR: A Rapid Method for Efficient and Representative Amplification of Human Genomic DNA Sequences from Complex Sources', *Genomics*, **19**(3), 506–514 (1994).
32. C.R. Cantor, C.L. Smith, *Genomics: The Science and Technology Behind the Human Genome Project*, Wiley & Sons, New York, 1999.
33. O.D. Ermolaeva, E.D. Sverdlov, 'Subtractive Hybridization, a Technique for Extraction of DNA Sequences Distinguishing Two Closely Related Genomes: Critical Analysis', *Genet. Anal.*, **13**(2), 49–58 (1996).
34. O.D. Ermolaeva, S.A. Lukyanov, E.D. Sverdlov, 'The Mathematical Model of Subtractive Hybridization and its Practical Application', *Ismb.*, **4**, 52–58 (1996).
35. N.G. Gurskaya, L. Diatechenko, A. Chencik, P.D. Siebert, G.L. Khaspekov, L.A. Lukyanov, L.L. Vagner, O.D. Ermolaeva, S.A. Lukyanov, E.D. Sverdlov, 'Equalizing cDNA Subtraction Based on Selective Suppression of Polymerase Chain Reaction: Cloning of Jurkat Cell Transcripts Induced by Phytoghemaglutinin and Phorbol 12-Myristate 13-Acetate', *Anal. Chem.*, **240**(1), 90–97 (1996).
36. H. Yokota, T. Iwasaki, M. Takahasi, M. Oishi, 'A Tissue-specific Change in Repetitive DNA in Rats', *Proc. Natl. Acad. Sci. USA*, **86**(23), 9233–9237 (1989).
37. H. Yokota, M. Oishi, 'Differential Cloning of Genomic DNA: Cloning of DNA with an Altered Primary Structure In-gel Competitive Reassociation', *Proc. Natl. Acad. Sci. USA*, **87**(16), 6398–6402 (1990).
38. H. Yokota, S. Amano, T. Yamane, K. Ataka, E. Kikyua, M. Oishi, 'A Differential Cloning Procedure of Complex Genomic DNA Fragments', *Anal. Biochem.*, **219**(1), 131–138 (1994).
39. N.A. Lisitsyn, N. Lisitsyn, M. Wigler, 'Cloning the Difference Between Two Complex Genomes', *Science*, **259**(5097), 946–951 (1993).
40. L. Zhang, X. Cui, K. Schmitt, R. Hubert, W. Navidi, N. Arnheim, 'Whole Genome Amplification from a Single Cell: Implications for Genetic Analysis', *Proc. Natl. Acad. Sci. USA*, **89**(13), 5847–5851 (1992).
41. D. Grothues, C.R. Cantor, C.L. Smith, 'PCR Amplification of Megabase DNA with Tagged Random Primers (T-PCR)', *Nucleic Acids Res.*, **21**(5), 1321–1322 (1993).
42. H. Telenius, N.P. Carter, C.E. Bebb, M. Nordenskjold, B.A. Ponder, A. Tunnacliffe, 'Degenerate Oligonucleotide Primed PCR: General Amplification of Target DNA by a Single Degenerate Primer', *Genomics*, **13**(3), 718–725 (1992).
43. V.G. Cheung, S.F. Nelson, 'Whole Genome Amplification Using a Degenerate Oligonucleotide Primer Allows Hundreds of Genotypes to be Performed on Less than One Nanogram of Genomic DNA', *Proc. Natl. Acad. Sci. USA*, **93**, 14 676–14 679 (1966).