A Decade of Differential Display


INTRODUCTION

One of the most elusive questions in biology has been how a fertilized egg that contains a complete set of genes unique to a species can give rise to so many different cell types that organize into different tissues and organs, which, in turn, make up a living organism. Complete sequencing of many genomes, from a few thousand kilobases for bacteria, to over 3 billion bp for human, has yet to provide the clue to the mystery of life. Of the estimated 40,000–50,000 genes embedded in our genome, only 10%–15% of them are “turned on” (expressed as mRNAs for protein synthesis) at any given time in each of our cells. Thus, interpretation of the genomic instruction will have to rely, at least in large part, on tools that can allow us to determine when and where a gene is to be turned on or off in a cell as it divides, differentiates, and ages. Obviously, such tools are also important for the detection of when and where a seemingly precise interpretation of genomic instruction goes awry, which underlies many disease states such as cancer. Differential display (DD) technology is one of the major tools that has already helped thousands of researchers worldwide interpret the genomic information (36).

The first comprehensive review on DD (37) dates back to 1995 when there were fewer than 100 publications describing the use and improvement of the method. Now, nearly 10 years after the initial description of DD methodology, the number of publications using DD has exploded to more than 3800, which outnumbers the combined total of those using all other competitive methodologies, including DNA microarrays, serial analysis of gene expression (SAGE) and representational difference analysis (RDA) (Table 1). On one hand, the vast number of successful DD applications is obviously a testimony to the power of this simple method; on the other hand, it makes a comprehensive review on DD technology a daunting task. More recent reviews on DD include those dealing with either methodology such as how to reduce false positives (40,43,59) and on various restriction fragment-based DD strategies (20,53), or DD applications in diverse biological systems (3,42).

Given the scope and time constrains, this review will focus on publications within the past five years to provide a glimpse of the current status of DD technology and highlight some of the outstanding examples of its successful applications in diverse biological systems. Some of the major fundamental differences between DD and DNA microarrays will also be covered.

DD Methodology

To speed up the identification of differentially expressed genes, DD was developed with the aim of overcoming limitations of previous methodologies that were error-prone, insensitive, nonsystematic, and laborious. To achieve this goal, the method had to be simple, sensitive, systematic, and reliable. The revolution in molecular biology, in fact, has been powered mostly by simple methodological breakthroughs such as recombinant DNA technology, DNA sequencing, and PCR. Simplicity also ensures the reliability and accuracy of a method. The method had to be sensitive so it could be applied to biological systems where...
scarce biological samples are available. The method had to be systematic so a complete search of all the expressed genes in a cell was possible. Based on these crucial criteria, DD was developed by integrating three of the most simple, powerful, and commonly used molecular biological methods, namely RT-PCR, DNA sequencing gel electrophoresis, and cDNA cloning (36). Figure 1 depicts the principle of DD. First, mRNAs from cells are converted to first-strand cDNAs using three individual anchored oligo-dT primers that differ from each other at the last 3'-non-T base. The use of these anchored primers enables the homogeneous initiation of cDNA synthesis at the beginning of the poly(A) tail for any given mRNA. The resulting three sub-populations of cDNAs are further amplified and labeled with either isotopes or fluorescent dyes by PCR in the presence of a set of second primers that are short and arbitrary in sequence. The length of an arbitrary primer is so designed that by probability each will recognize 50–100 mRNAs under a given PCR condition (36). As a result, mRNA 3' termini defined by any given pair of anchored primer and arbitrary primer are amplified and displayed by denaturing PAGE. Side-by-side comparisons of such cDNA patterns between or among relevant RNA samples would reveal differences in gene expression. Differentially expressed cDNA bands can be retrieved, cloned, and sequenced for further molecular characterization. Based on the finding that each arbitrary primer would recognize its corresponding mRNA targets with a minimum of seven bases match, mathematical models have been proposed to predict the relationship between the number of arbitrary primers and the coverage of expressed genes in any given eukaryotic cells (38,39). These models have yet to be validated by saturation DD screening using a large number of primer combinations. One of the major advantages of DD is that it does not require any prior knowledge in mRNA sequences, making the gene screening systematic and nonbiased.

### Factors Affecting the False-Positive Rate in DD

Some of the earlier efforts to improve the accuracy of DD had been focused on the identification of factors, both extrinsic and intrinsic of the method, that are the causes of false positives. These efforts have been quite fruitful in identifying many such factors, including the systems being compared, experimental designs, appropriate internal or positive controls, qualities of reagents and enzymes, reaction setup, criteria for picking bands, types of tubes and thermal cyclers, pipetting errors, and especially primer designs, etc. (13,43). It should be noted that the original primer designs for DD with two-base anchored primers and arbitrary decamers were found to be suboptimal and might have contributed to the high rate of false positives (43). Later, the development of one-base an-

<table>
<thead>
<tr>
<th>Methods</th>
<th>Year Invented</th>
<th>Medline Hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>DD</td>
<td>1992</td>
<td>3862</td>
</tr>
<tr>
<td>Microarrays</td>
<td>1995</td>
<td>1196</td>
</tr>
<tr>
<td>Subtractive Hybridization</td>
<td>1985</td>
<td>886</td>
</tr>
<tr>
<td>SAGE</td>
<td>1995</td>
<td>644</td>
</tr>
<tr>
<td>RDA</td>
<td>1993</td>
<td>288</td>
</tr>
</tbody>
</table>


Table 1. Impact of Major Technologies in Gene Expression Analysis

Figure 1. Schematic illustration of fluorescent DD.
chored primers and rationally designed 13-mers (43) have greatly improved the accuracy of DD and were represented by a larger number of successful applications published.

**RFLP-Based DD Strategies**

One of the assumptions for the false positives in DD has been that the short arbitrary primers (10–13 bases) hybridize promiscuously to the cDNA templates, in contrast to experimental data showing that these primers can hybridize predictably to their target templates with 7–10 bases matches at their 3’-ends (38,39). To this end, many attempts have been made to replace the selection of the mRNAs by arbitrary primers with restriction enzyme digestions (53). There have been close to a dozen such RFLP-based DD methods described so far, each with its own name, such as RFLP-coupled domain-directed DD (RC4D) (16), ordered differential display (ODD) (34), Gene-Calling (20), amplified differential gene expression (ADGE) (11), total gene expression analysis (TOGA) (48,66), and amplification of double-stranded cDNA ends restriction fragments (ADDER) (34). These methods in principle are essentially the same with a variation on a theme. The strategy is to reverse-transcribe the mRNA with oligo-dT primers with or without anchor followed by the second-strand cDNA synthesis. The dsDNAs are then digested with four base cutter restriction enzymes, which create a population of cDNA ends to which adaptor primer binding sites can be ligated. The cDNAs are then amplified with either the anchored oligo-dT primers in combinations with upstream ligated primers, or ligated primers alone. In the case of ODD, the amplifications of cDNA fragments other than the 3’ termini is suppressed by self-complementary primers ligated to the cDNAs, whereas for ADDER and TOGA, the cDNA fragments corresponding to the 3’ termini of mRNAs are purified and released by the use of a specially designed anchored primer attached to a solid phase, such as magnetic beads (34,48). Unlike the traditional DD method that requires only two steps, namely RT and PCR, the RFLP-based DD strategies all require many more additional experimental steps, such as second-strand cDNA synthesis, restriction digestion, and ligation of adaptor primers before mRNA samples are compared. Since each of the additional steps would inevitably expose a method to variability between or among RNA samples before they are compared, the accuracy of these strategies over that of traditional DD has yet to be substantiated by more successful applications. Importantly, it can never been emphasized enough that if a method is to be developed, or an existing method is to be improved upon, then one must first incorporate a positive control during a method development and then show a side-by-side comparison of the old and new method. Without these two important considerations, it will be difficult to judge the reliability and significance of a method or method improvement.

**Targeted DD**

In the past, several attempts have been made to use DD to display gene family-specific mRNAs selectively, but with little success. This could be explained by the limited specificity provided by the short arbitrary primers. However, strategies designed to target either N-termini of protein coding regions based on the KOZAK sequence or secreted proteins based on the signal peptide sequences appeared to be feasible (38,68). Both approaches used arbitrary primers instead of anchored oligo-dT primer to prime the first-strand cDNA synthesis, followed by PCR in the presence of a second primer targeting desired mRNAs. In theory, the successful targeting rate is likely to be very low with these approaches, and a comprehensive targeted DD screening could be as tedious as regular DD.

**Integration of Subtractive Hybridization with DD**

The other strategy attempted to improve DD is by combining subtractive hybridization with DD. Such approaches, one dubbed subtractive display (5), the other mutually subtracted RNA fingerprinting (SuRF), were designed first to remove most of the commonly expressed mRNA by subtractive hybridization and then to display the remaining differentially expressed mRNAs by DD (17). Conceptually, these approaches appeared to be straightforward, though without a positive control (such as a known differentially expressed gene) during the method development, their reliability and advantages over that of traditional DD have yet to be substantiated by future applications.

**DD Analysis of Prokaryotic mRNAs**

Although conventional DD was designed to analyze the mRNAs from eukaryotes, which have poly(A) tails, DD also has been adapted successfully to analyze gene expression in prokaryotes. The general strategy is based on the RNA arbitrarily primed PCR method in which one or two arbitrary primers, instead of one arbitrary and one anchored oligo-dT primer, are used to reverse-transcribe and amplify the bacterial mRNAs that lack poly(A) tails. Several more systematic approaches are of particular interest since they are likely more comprehensive in gene coverage. One such method uses primers representing highly iterated palindromic (HIP) elements found in certain bacterial genomes (6). Such HIPs are commonly found in half of the genes in a genome, providing a convenient global screening strategy of bacterial gene expression using a limited number of primers. Another screening strategy for bacterial gene expression appears to be more systematic and comprehensive in gene coverage. This strategy employs a set of 81 rationally designed arbitrary 17-mers, each of which only differs in sequence in one of the last four 3’ bases (9). When these primers were used individually for both RT and PCR steps, metabolic genes induced by a certain chemical compound were successfully identified. Some of these genes were found to encode metabolic enzymes that break down the chemical compound of interest (9).

**Integration of DD with DNA Microarrays**

DNA microarrays offer the potential of providing a snapshot of gene expression in a cell. However, in practice, this method has many pitfalls (30). One of
the most serious limitations of microarrays appears to be the quality of the cDNA probes, with regard to their sensitivity, complexity, and uniformity in labeling. To this end, it has been demonstrated that less complex cDNA probes generated by DD could greatly increase the sensitivity in gene detection for cDNA microarrays (69). Furthermore, the concept of dual-color fluorescent detection in cDNA microarray has been integrated into multicolored fluorescent DD (12).

**DD Automation**

Although DD is the most widely used method for isolating differentially expressed genes, most DD screenings take a shotgun approach by using only a limited number of primer combinations, with which a few genes are identified and characterized. Theoretically, a genome-wide comprehensive DD screening is possible (38,39) but would require hundreds of PCRs to be set up for each RNA sample being compared. Clearly, an automated DD platform that incorporates robotic liquid handling, digital data acquisition, and analysis will greatly increase not only the throughput but also the accuracy of DD screenings. To this end, fluorescent DD was developed (25) and optimized to have a similar sensitivity (12). Combined with robotics and digital data analysis, the automated fluorescent DD was shown to be accurate and with high throughput (12,44). Further streamlining the fluorescent DD process, any differentially expressed cDNAs of interest could be re-amplified and sequenced directly with a given arbitrary primer without subcloning (10). Computer programs were developed to allow positive band identification automatically from a fluorescent DD image (1,2,60). The most sophisticated attempt in fluorescent DD downstream automation could have been the development of a prototype computer-controlled system for positive band identification and retrieval by Hitachi (Alameda, CA, USA) (23). This approach employed capillary array gel electrophoresis coupled with fraction collection using sheath flow technology. Automation in PCR setup and fluorescent data analysis for TOGO and GeneCalling has also been described (20,48).

**Highlight of Successful DD Applications in Diverse Biological Systems**

Given more than 3800 publications describing the applications of DD, it will be a formidable task for anyone trying to provide a thorough overview of its utility. Clearly, the large number of published applications is a manifestation of both the simplicity and non-species-specific nature of DD methodology. A more comprehensive compilation of DD publications broken down into fields of applications can be found at the Web site of GenHunter (www.differentialdisplay.com). Here I will only highlight some of the excellent examples of more recent DD applications in diverse research fields and across many biological systems, to outline a tried-and-true roadmap for those who want to embark on a successful, rather than a superficial, gene hunting expedition.

**Disease Marker Identification**

Based on the principle of the method, DD offers a great potential for the rapid identification of markers for disease diagnosis or prognosis (41). Using a comprehensive DD screening, expression markers associated with chronic hepatitis B viral infection (29), fatal neurodegenerative diseases such as scrapie (54), and melanoma (15) have been documented. Melastatin (15), the first commercial marker for prognosis of melanoma metastasis, now has been marketed by Millennium Pharmaceuticals (Cambridge, MA, USA).

**Cell Cycle**

DD, coupled with subsequent genetic analysis of the genes identified, has time and again proven to be an extremely powerful approach that is readily applicable to biological systems such as yeast, zebrafish, *Arabidopsis*, and *Drosophila*, where genetic manipulation of genes can be easily and quickly accomplished. Using such an approach, Mai and Breeden (50) identified a number of G1 cyclins whose expression was coordinately repressed by Xbp1, a transcriptional repressor in yeast. Gene deletion analysis confirmed functionally that one of the cyclin genes crucial for sporulation in
budding yeast, CLN1, was the primary target of Xbp1 (50). Using Xenopus oocytes with and without stimulation of progesterone, which induces meiotic maturation, Antonio et al. (4) discovered a new kinesin-like gene, Xkid. Through a combination of molecular biological, cell biological, and biochemical characterizations, Xkid was beautifully shown as a key kinesin required for chromosome alignment on the metaphase plate (4).

Cancer Research

One of the most intensive research efforts in cancer research has been the search for p53 tumor-suppressor target genes. Since p53 functions as a transcription factor, identification of its target genes may shed light on the understanding of cancer. Among a handful of such p53 target genes identified so far, a significant number of them were discovered by DD. These include PAG608, which encodes a nuclear zinc finger protein (24), DDA3 (49), ei24 (21,35), a ribonucleotide reductase gene (67), Pidd, which is a new death-domain-containing protein (46), Reprimo (57), Noxa, which belongs to Bcl-2 protein family (56), and p53DINP1, which works as a cofactor for p53 phosphorylation (58).

Among oncogene targets identified by DD, many ras target genes have been identified (27,28), including a novel cytokine loop constitutively activated by ras oncogenes (75). This cytokine system now has been defined biochemically and designated as IL-24 receptors (71).

Neuroscience

Among the large number of DD applications in this field, studies on the circadian clock are of particular interest since transcriptional regulation of genes is known to play a key role. One of the circadian clock-regulated genes, nocturnin, was identified by DD and shown to be expressed beautifully at the day-night transition in the eye (19). Subsequent biochemical and cell biological studies demonstrated that nocturnin expression was restricted to the retina, and the promoter elements mediating the specificity in nocturnin gene expression have been identified (47). A direct transcriptional target of Drosophila CLOCK and CYCLE, vrille, was discovered using DD by comparing gene expression profile between brains of wild-type and per mutant flies (7). Functional characterizations of vrille knockout and transgenic flies confirmed that vrille was a key component of the circadian clock. While Drosophila exhibits a circadian rest-active cycle, it is unclear whether fly rest constitutes sleep. A comprehensive DD screening to compare gene expression differences in fly heads between “wake” and “rest” states yielded strong evidence that Drosophila rest has many similarities to mammalian sleep (65). Therefore, the fly may be used as a model system for the genetic analysis of sleep. Most interestingly, genes regulated by music in the brain of songbird have also been identified DD and characterized (14).

Plant Science

DD has been widely used in plant science. Several genes involved in floral development were identified by DD. These included a family of MADS-box genes (74) and NAP, a target gene of floral homeotic genes (63). Mis-expression of NAP in transgenic plants caused abnormality in floral, but not vegetative, development of the plant. One of the most beautiful examples of the successful application of DD, perhaps, was the identification of genes responsible for the modification of anthocyanin, which is a precursor for floral color determination (73). Three of the genes identified by DD to be expressed in purple-colored but not green-colored plants encoded enzymes for anthocyanin biosynthesis and modifications (73). One of these enzymes, 5-O-glucosyltransferase, was shown to be able to modify anthocyanin to form a purple-colored product (73).

Detection of mRNA Polymorphisms and Gene Silencing by DD

While most of the DD applications have been used to identify differentially expressed genes, in principle DD can also reveal sequence variations in mRNAs. Indeed, it was reported recently that small deletions were detected in certain mRNAs, which caused a size shift, instead of an absence or presence, of a cDNA band when RNA samples from different individuals were compared (45). In addition, it is worth noting that DD also has been used successfully to identify gene imprinting, a phenomenon referring to allele (maternal or paternal)-specific gene expression (22,26,64).

Identification of mRNAs Bound to RNA-Binding Proteins by DD

Another interesting DD application was designed to identify mRNA substrates for a particular RNA binding protein. In this approach, the RNA-protein complex was purified by pull-down assays using GST-tag or immuno-purification. The purified RNA-protein complex then was subjected to DD to resolve and identify the mRNA species bound to an RNA binding protein (70).

Looking into the Future: DD versus Other Gene Discovery Technologies

Given the other gene expression analysis tools available, especially DNA microarrays, many researchers are curious about which method is best
suited for their own studies. Some of
the fundamental differences between
DD and microarrays are discussed be-
low, which hopefully may help re-
searchers not only to choose between
the two methods but also to perform ei-
ther method more successfully.

Both DD and microarrays are con-
ceptually simple; however, the two
methods are principally different (Table
2). The fundamental difference is that
DD visualizes the mRNAs in subsets di-
rectly after their amplification and la-
beling by either isotopes or floures-
cence. In contrast, DNA microarrays
visualize the mRNAs indirectly, after
hybridization of complex mRNA
probes as first-strand cDNAs labeled
with fluorescent dyes, to a mostly in-
complete set of cDNA templates spot-
ted on a glass surface. For DD, hun-
dreds of PCRs must be prepared and
separated on a gel matrix for each RNA
sample to ensure that most of the
mRNAs are detected, whereas for mi-
croarrays, similar gene coverage may be
obtained with only one labeling reaction
per RNA sample, provided that most of
genes from a genome are spotted on a
"chip" in the first place. One disadvan-
tage of DD may be that once a differ-
ce is detected, one may not know the
nature of the gene at hand until it is re-
amplified and sequenced, which could
take additional days to accomplish. In
contrast, each spot on a microarray con-
tains a known sequence. Once a signal
is detected, so may be the nature of the
gene, provided that the signal is specif-
ic. However, one of the major chal-
lenges for microarrays has been how
one can be sure that a hybridization sig-
nal is specific to a known sequence laid
on a chip when a complex cDNA probe
is used. In fact, a cDNA probe used for
microarrays is so complex that it con-
ists of as many as 10,000 different
species ranging from a few to thousands
of copies per cell. Further compounding
the problem in signal specificity has
been the fact that eukaryotic genes often
come in families with many conserved
sequences among the family members.
Also for microarrays, one is going to be
limited to the detection of whatever
genes that are spotted on a slide, making
it a "closed" system for gene discovery.
Moreover, for each living organism, one
must first know and have all gene se-
quencies before a chip can be made. In
contrast, DD is sequence-dependent and
capable of detecting both known and
novel genes for any living organism.

It should be pointed out that, like DD
when it was first developed, DNA mi-
croarray is an emerging technique that
has many technical problems (30).
Chief among these are problems for re-
producibility, probe sensitivity, nonlin-
erarity in signal detection (61), probe
cross-hybridization due to homologous
cDNA sequences (62), and data man-
agement (18). There certainly has been
a misperception that microarray data
(which is essentially based on reverse
Northern blot) is as accurate and sensi-
tive as Northern blot itself. Therefore,
computer programs were developed to
make sense of a large number of uncon-
firmed gene expression differences ob-
tained often from a single run of an
experiment. Concepts such as "self-or-
ganizing data" emerged, as if nature’s
secret can be revealed simply by the
compilation of expression patterns of a
limited number of unconfirmed genes.
"Genome-wide gene expression profil-
ing" became a standard banner for
many microarray papers, which in fact
only evaluated a few thousand out of an
estimated 30,000-50,000 genes in a
genome. Ironically, such collections of
gene expression differences without fur-
ther confirmation and functional studies
had been recognized as a major flaw for
earlier DD studies, but it appeared to be
forgotten or ignored for microarrays.
Clearly, like any other methods includ-
ing DD, microarray technology is not
foolproof. For example, before a mi-
croarray experiment was carried out,
cDNA sequences deposited on an array
might have an error rate as high as 30%
(31); one major oligonucleotide array
company had to recall several of its
mouse chips because of as many as 65%
erroneous sequences synthesized on
them (51). The other major player in
cDNA arrays has recently dropped its
microarray business altogether, citing
competition instead of technical diffi-
culties as the major reason (32). Other
problems and challenges for microar-
rays have been the subject of several
critics (8,18,44,55,72). Knowing these
potential pitfalls will not only help to
improve the method further but also to
apply it more successfully.
It should be pointed out that other gene discovery methods, besides DD and microarrays, have also contributed greatly to functional genomic studies. For example, the SAGE technology has greatly simplified and sped up the Y' expressed sequence tag collection and sequencing. Like DD, SAGE is an “open” system. However, like RFLP-based DD methodologies, SAGE requires many more additional steps in cDNA generation and manipulations before the samples are compared. This makes the method also susceptible to artifactual differences that could not be easily determined until after thousands of runs of DNA sequencing reactions are carried out and analyzed. Because SAGE is a DNA sequencing-dependent gene dosage counting method, the gene expression measurement may be more quantitative than DD. However, given the large number of sequencing reactions needed for SAGE, the cost factor may be one of the major rate-limiting steps for its widespread application in most of the cost-conscious “cottage-industry” academic laboratories.

Obviously, no matter which gene discovery methodologies one chooses to use, it will ultimately be the functional characterizations of each isolated genes, by genetic, cell biological, and biochemical methods, that will likely provide the real proof (or disproof) of the relevance of the genes to a biological system under investigation. In a preface to a method book on protein purification (33), Dr. Arthur Kornberg had once quoted an admonition of Efraim Racker, “Don’t waste clean thinking on dirty enzymes”, to illustrate the importance of good biochemical practice, which is at the core of enzymology. A similar doctrine, “Don’t waste clear thinking on dirty data”, will certainly continue to help to produce a better quality of science in the field of gene expression analysis in the next 10 years.

ACKNOWLEDGMENTS

This special review is dedicated to Arthur Pardee on the occasion of his 80th birthday, because without his vision and wisdom, there would not have been DD. Given the scope and time constraints in preparing for this review, it is impossible for the author to cite most of the DD publications. As such, any omissions in citation in this review are purely unintentional. Given the potential conflict of interest of the author when commenting on other competing technologies, readers’ discretion is recommended. The author is grateful to Emily Thomas for critically proofreading the manuscript before its publication. This work was supported in part by National Institutes of Health grant no. CA76960.

REFERENCES

