DNA Chip

The What's, the Why's and the How's

Bhismadev Chakrabarti and Samir K Brahmachari

Bhismadev Chakrabarti is studying B.Sc(Hons.) Chemistry at St.Stephen's College, Delhi. He is passionately interested in genomics and Indian music. He plays the tabla and enjoys reading. Currently he is a Scholar Associate at the Centre for Biochemical Technology, Delhi, under S K Brahmachari.

Samir K Brahmachari is the Director, Centre for Biochemical Technology(CSIR), Delhi and Professor, Molecular Biophysics Unit, Indian Institute of Science, Bangalore. Primarily interested in functional genomics, his present fascination and area of work is functional polymorphism scanning and the genetics of neurodegenerative disorders.

This article presents a brief overview of DNA chips, or microarrays. Microarrays are a very significant technological development in molecular biology, and are perhaps the most efficient tool available for functional genomics today.

What is a DNA Chip?

Imagine a world without identity cards; no I-cards for the college or office or bank account or anything! All you are carrying is a small (say, 2cm $\times$ 2cm) 'DNA-chip', which has the whole of your genetic profile on it. Your identity cannot get more authentic than that. Imagine a world where marriages are not decided by matching horoscopes, but by testing the compatibility of the DNA chips of the prospective couple! Welcome to the whole new world of DNA chips or microarrays. As evident from the name, microarrays essentially consist of an array of either *oligonucleotides or cDNA* (see *Box* 1) fixed on a substrate, usually solid, of micro-dimensions.

Why DNA Chips?

There has been an explosion of information in the field of genomics in the last five years. Genomes of several organisms have been fully sequenced. The next step necessarily involves the analysis of comparative expression levels of various genes and to identify all the possible variations of sequence present in each of the genes, or in the non-coding regulatory regions, obtained from a particular population. Handling such large volumes of data requires techniques which necessitate miniaturization and massive-scale parallelism. Hence the DNA chip comes into the picture. So vast is the scope of its use that many

scientists predict its playing a more vital role in future than what is played by its semiconductor cousin now!

Gone are the days of a molecular biologist groping about a host of oligonucleotides in search of an elusive gene of interest. The microarray technology brings with it several advantages like those of large-scale parallelism (with the current pace of developments, the day when we will have the whole human genome on a microarray of dimensions say 1.5cm $\times$ 1.5cm is not far off), multiplexing, automation, and obviously, miniaturization. So, all the modern-day molecular biologist has to do is to set up a microarray of all possible *n*-mers (where *n* ranges from 6 to 10 nucleotides), and hybridize it with the *labeled* segment (see *Box* 2) of interest from the desired genome. The spots on the chip that 'light up' after the hybridization would correspond to the presence of complimentary sequences which then can be arranged computationally to predict the exact sequence of the gene.

Down the years, as molecular biology has given way to genomics [1] the scale of research has multiplied manifold, from identifying one gene at a time to identifying thousands of genes at a time with the help of a microarray. Such massive parallelism greatly increases experimental efficiency and allows meaningful comparisons to be made between the genes or gene products represented in the microarray. The principal uses of microarrays can be classified under two heads:

Resequencing: This broadly refers to the experiments performed to determine the exact sequence of a gene, and/or to identify the possible insertions/deletions/substitutions in a gene, or to identify the single nucleotide polymorphisms (*SNPs*; see *Box* 3) that maybe present in the sequence. For example, given a reference sequence of DNA, four *probes* (see *Box* 4) are designed to interrogate a single position. One of the probes is made exactly complementary to the reference sequence while the other three are identical to the first, except at the position of interrogation, where each of the other three bases are substituted. When

Box 1. Oligonucleotides & cDNA

We refer to synthetic oligonucleotides when we mention oligonucleotides. These have characteristic 5'-OH and 3'-OH groups. cDNAs have characteristic 5'-phosphoryl and 3'-hydroxyl groups. (see also *Box* 5)

Box 2. Principle of the DNA chip

School level biology is enough to understand the principle of the DNA chip. This whole technique is based on Watson–Crick base pairing (where A binds with T using 2 Hydrogen bonds and G binds with C using 3 Hydrogen bonds) between the probe (fixed on support) and a labeled target sequence (in solution).

Box 3. SNPs

SNPs are single nucleotide polymorphisms, i.e. single base variants at a particular locus in a given sequence of genomic DNA. These are the most frequent type of variations in the human genome (about 1 in every 500 nucleotides in the human genome has an SNP.) Hence they are used as 'markers' in human genetics.)

hybridized with the reference sequence, the probe exactly complimentary to the reference sequence 'lights up' most. When hybridized with another sample of the reference sequence (isolated from a different source), the probe exactly complementary to that shows up most prominently.

Following this train of logic, it is easy to predict that one thousand such sets of four probes will be required to identify all possible SNPs in a reference sequence of 1000 nucleotides. It can be estimated that for complete human genome the difference of DNA sequence (e.g. SNPs) will be in 6 million locations. Hence 24 million probes will be sufficient to find all the variations in the human genome sequence. However, all SNPs or other polymorphisms may not be functionally significant.

Gene Expression Studies: The expression pattern of a gene provides indirect information about its function. A gene expressed only in the kidney is not very likely to be directly involved in the pathology of schizophrenia. In the human genome, the ratio between coding and non-coding DNA is very low (less than 3% of the human genome is expressed as proteins). In order to obtain a holistic view of comparative expression levels of particular genes of the genome in the various organs of the body, or to compare the expression levels of genes between

Box 4. Hybridization: Labelling

The target sequence will bind with the complementary probe sequence only in the presence of certain hybridization buffers within a particular pre-determined temperature range. 'Earmarking' the target sequences becomes very essential in order to detect spots on the chip where hybridization has occurred successfully. To achieve this, target sequences are radioactively or fluorescently labeled (causing the famous FISH and CHIP combination, but FISH here stands for fluorescent in-situ hybridization!). A recent approach involves the use of 2-colour fluorescence, where the probes (cDNAs) from two different sources or the same source under two different conditions are labeled with two different fluorescent tags (say red and green), and are hybridized on the target. Quite obviously, those spots on the chip which exhibit yellow colour will indicate that those genes are expressed nearly equally where as under or over expressed genes would show up as red or green spots.

the diseased and normal states, we use cDNA microarrays (see *Boxes* 1 and 5). These have direct use in drug discovery where such comparative expression levels between normal (called wild type), diseased and drug-treated cells provide useful data. Expression analysis, thus, is the single most biologically informative application of microarrays.

As few as 100,000 (relatively speaking!) unique human cDNAs are enough to monitor the expression levels of the whole human genome on a single chip!

Box 4. cDNA

cDNA stands for complementary DNA. In order to identify those segments of genome which are functionally expressed (as proteins), mRNA is isolated from the body cells. This mRNA when treated with reverse transcriptase enzyme, under certain reaction conditions, gives rise to cDNA, which is a single strand complimentary copy of the mRNA.

How is a DNA Chip made?

Microarrays that have been made till date have mostly used one of the three following substrates:

1) Nylon/nitrocellulose membranes

2) Glass

3) Silicon

While nylon and other porous membranes have the disadvantage of low *signal to noise ratio* (see *Box* 6), the impermeable and rigid substrates have several advantages working in their favour. These are:

Box 6. Signal to Noise Ratio

The ratio between the fluorescence intensity of the spot where hybridization has occurred on the chip to that of the support itself is known as the signal-to-noise ratio. When S/N ratio is high, we can clearly detect the spots where hybridization has taken place while when this ratio is low, the whole chip appears uniformly bright.

i) As liquid cannot penetrate the surface of the support, target nucleic acids can find immediate access to the probes without diffusing into pores.

ii) Due to enhanced mixing, the rate of hybridization is stepped up.

Glass supports have an advantage due to their flatness, rigidity and transparency with reference to image acquisition and processing (image refers to the fluorescent spots, post-hybridization). Silicon offers an advantage of having very closely spaced functional groups on its surface to which the oligonucleotides/ cDNA may bind. This high density of functional groups possible on silicon surfaces helps in miniaturization.

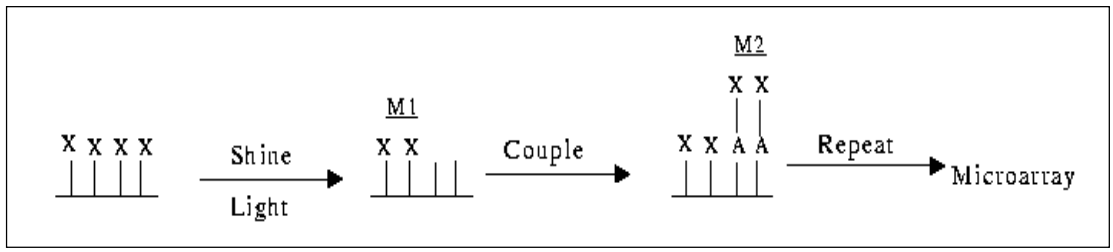

Figure 1. Photolithography (the growers) X=photolabile protecting group; M1, M2=photomasks.

There are three major approaches to the making of a microarray. They are:

Gridders: These are instruments that pick up nanolitre amounts of the sample (oligo/cDNA) from a microtitre plate and deposit it according to a particular pre-determined order on definite locations of the substrate (see *Figure* 1).

The support needs to be functionalized in order for the oligos/cDNA to be properly fixed. One of the standard protocols involves aminating the support and phosphorylating the cDNA/oligo at the 5'-end using a kinase enzyme. Following this, when the derivatised oligos/cDNAs are placed on the fictionalized support in the presence of certain coupling buffers and incubated for a few hours, the oligos/cDNAs get fixed firmly on the substrate.

However, ultraviolet light induced cross-linking is a preferred method for fixing cDNA on glass slides. This approach is technically referred to as the *immobilisation approach.*

Grabbers: These refer to supports where positively charged microelectrodes are fixed, to which the negatively charged oligos/cDNAs attach themselves and thus the array gets created. This approach is relatively less used (see [3]).

Growers: This approach, known as the in situ synthesis of oligonucleotides on the support (silicon, in this case) uses light-directed combinatorial chemistry to 'grow' the oligos on the support itself, nucleotide by nucleotide (see *Figure* 2). A lot of work has been done by Affymetrix, a company based in Califor-

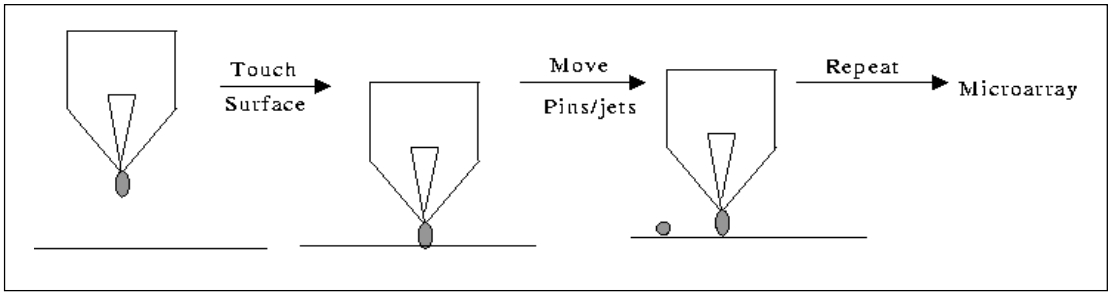

Figure 2. Spotting and immobilisation (using the gridders): Microspotting of functionalized presynthesised oligonucleotides/ cDNA is done robotically.

nia, USA. They have made Gene Chips™ having 40, 000 oligos on a 1.5cm $\times$ 1.5cm silicon chip.

And now, the Disadvantages…

Thankfully, they are few in number. Here goes:

1) *Generic microarrays* (those where the probes include all possible *n*-mers, where *n* ranges from 6 to 10 nucleotides) are relatively inefficient in testing for sequences with repeats, something which is very common in the non-coding regions of the human genome.

2) The hybridization between the target sequences and the probe sequences may be nonspecific if the difference of length between the two sequences is high. However, this problem can largely be eliminated by giving the chip some stringent post-hybridization washes.

3) The present signal detection system is based on fluorescence which is unsuitable for accurate quantification.

Box 7. Nonspecific Hybridization

Simply put, non-complementary binding of nucleotide bases, e.g. when A binds to C and not to T, hybridization is said to be nonspecific. Such nonspecific hybridization may occur due any one or more of the following reasons:

i) The difference of length between the target sequence and the probe sequence is high, i.e. the length of overhangs is more than that of specific hybridizing sequences.

ii) The temperature restrictions during the hybridization reaction were not strictly followed.

iii) There are 3'-terminal mismatches between the probe and the target sequences.

Box 8. T_m

T_m stands for the temperature of 'melting' or denaturation of a particular duplex DNA sequence. It is the temperature at which half of the duplexes separate to give single stranded DNAs. The hybridization temperature of a particular sequence is usually 8-10 ºC below the T_m. T_m of a particular sequence is determined mainly by its base composition, and the order in which the bases are arranged.

Challenge

One of the primary challenges facing the array technology is the construction of a T_m-specific chip, i.e. all the oligos/cDNA on the chip will hybridize at the same temperature. The techniques available are still not able to create a chip where all the oligos/ cDNAs will have similar T_ms. Such a T_m-specific chip will naturally eliminate the probability of non-specific hybridization to a large extent (see *Boxes* 7 and 8).

Final Remarks

The microarray technology has suddenly magnified the scale of genomic research manifold, so much so that a leading science journal refers to it as "biotechnology's discovery platform for functional genomics". Immense amount of data generated by these biochips will have to be properly managed by bioinformatics tools capable of relational data analysis. This will, in turn, open up unprecedented access to key areas of human health, including disease prognosis and diagnosis, drug discovery, toxicology, aging and mental illness.

Address for Correspondence
Bhismadev Chakrabarti
Room H1, Allnutt North Block
St. Stephen's College
Delhi 110 007, India.
Email: cbhisma@go.com

S K Brahmachari
Director
Centre for Biochemical Technology
Mall Road (Opp: Jubilee Hall)
Delhi 110 007, India.
Email: skb@cbt.res.in

Suggested Reading

[1] S K Brahmachari and others, The multitude of omics and omes: Evolution of scientific terms in Molecular Biology in the new millennium, *Current Science*, 77, 7, 10 October, 1999.

[2] *Nature Genetics*, 21, Jan 1999 (supplement); the whole issue is deveoted to this topic

[3] B Sinclair, *Scientist*, 13 (11), May 24, 1999.

[4] Schena M and others, Microarrays: Biotechnology's discovery platform for functional genomics, TIBTECH, 16, 301-306, July 1998.