There are various types of arrays utilized in gene expression studies, each with distinct features and benefits. The first type is spotted arrays, which are produced on poly-lysine coated glass microscope slides. These arrays allow for high-density DNA and fluorescent labeling of the sample. The second type is in-situ oligonucleotide arrays, also known as Affymetrix Gene Chips, which use photolithography technology for production and allow the binding of many 25-mers. The third type is ink-jet microarrays, which involve the direct printing of 25-60-mers onto the glass slide and offer greater flexibility, sensitivity, and specificity.

Array fabrication is a crucial step in gene expression analysis, and there are two primary methods utilized: spotting and photolithography. In the spotting method, oligonucleotides are chemically synthesized and deposited at specified locations using a robotic pen. The DNA is typically amplified through PCR, and approximately 1-10 ng is deposited per spot for experimentation against 40-70-mers (targets). Microarray spotters are high-precision robots with metal pins that can dip into the DNA solution for input and tap down on the slide with high precision, allowing for up to 48,000 spots on a single slide.

In contrast, photolithography involves the use of light to create a pattern. This method relies on UV masking and light-directed chemical synthesis on a solid support to synthesize probes directly onto the array's surface, one nucleotide at a time per spot, for many spots simultaneously. Each oligonucleotide has 107 copies in a 24 x 24 um square, and 20 pairs of different 25 (170)-mers of each gene are used (multiple probes). Perfect match and mismatch between the probe and the target allow for the discrimination of single mismatched base pairs. The masking allows for the choice of exposure sites and coordinates to be specified on the array where the nucleotide needs to be attached. Each spot requires four masks per round, one of which ensures the nucleotide addition, and three other masks to prevent light from deprotecting the same spot while adding other three nucleotides.

There are multiple copies of a single probe present in a single against a single gene that attach to multiple copies of the gene, providing the intensity level of the gene being expressed. CG islands tend to attach more to the probe due to high numbers of hydrogen bonding. Therefore, it is preferred to have CG island probes. If we have a gene set, all of which have CG islands, then we can use CG island probes. Otherwise, we have to use probes that are not taken from the CG islands of the genes to ensure we do not get nonspecific intensity. Two types of probes are utilized in the same well: PM (perfect match) and MM (mismatch) probes. PM attaches to the gene in perfect precision, whereas MM has a mutation at any position leading to mismatch against the gene. White color represents the best match, black represents the mismatch probe. This is to make sure that only the gene of interest is being attached to the probe, preventing false positives. The gene expression is counted as the intensity difference between PM – MM for noise removal. If there is a mutation in the gene in one patient, and all other patients do not have that mutation, and we see a white light at their expression intensity, that means the gene from the patient has a mutation, which is why it is black. We perform a repetitive experiment to reduce false positives. If there is a mutation in the sample, then black intensity is shown for the mutated cell. Features are the genes in the case of microarray/NGS.

Finally, the inkjet method utilizes a spotting process for the synthesis of in-situ oligonucleotides using a five-ink printing system of the four nucleotide precursors. This method does not require a masking step and is therefore more flexible than the photolithography-based method. The probes synthesized by this method are typically 60-70-mers in length and utilize phosphonamidite chemistry to deliver extremely small volumes (picoliters) which are then spotted on the slide.

In conclusion, microarray technology offers a powerful tool for analyzing gene expression levels and can be used to identify potential biomarkers for disease diagnosis and treatment. Understanding the different types of microarrays and their fabrication methods is essential for designing effective experiments and interpreting the resulting data accurately.

DNA Microarrays

Microarrays are powerful tools for the detection and quantification of gene expression, as well as the detection of single nucleotide polymorphisms (SNPs) and genotyping. There are various types of microarrays, including DNA microarrays or gene chips, which are matrices of microwells, each containing a short segment of known DNA sequence. The location of each well, along with the specific DNA sequence present in each well, is known. The DNA sequences present in the wells are called probes or reporters, and they are designed to be complementary against the cDNA sample that is obtained through a reverse transcriptase procedure.

The use of complementary bases between the probe and the target is advantageous in identifying genes present in unknown DNA samples. Probe selection can be challenging when some genes are short, or have 3' or orientation bias. Probes can cross-hybridize to wrong targets, and genes can be duplicated, or have whole genome duplications, transposable elements, spontaneous mutations, pseudogenes, or exon shuffling. However, multiple probes can be provided for a single gene from its various regions.

Targets are tagged with fluorescent dyes such as Cy3 and Cy5, which allow the detection and quantification of the probe-target hybridization. Once the hybridization has been performed, a laser is shone over each well, and a scanner detects and measures the light intensity of each well of the fluorescent dye. If multiple targets attach to a particular probe in a well, the fluorescent dye will be greater at that spot.

There are two types of microarray experiments: Agilent and Affymetrix. Agilent mixes both samples on the same slide, while Affymetrix uses individual slides for each sample. Dye is attached to the gene/sample rather than the probes, and probes are present within the cells. Colored slides that provide green/red/yellow format belong to Agilent and not to Affymetrix because it uses multiple slides for each sample.

There can be three issues with the quality of the data produced with microarray analysis: inaccurate fluorescent signal, low signal intensity, and fluorescence signal fidelity. Fold-change is the value obtained when the value of Cy5 is divided by Cy3, indicating how many times the

gene of the disease has been expressed more as compared to the other gene. We take log2 of the fold change value to put the values of the under-expression on the left side of 0 and overexpression on the right side of 0.

Column-wise information in microarray data is experiments/samples, whereas row-wise data are genes. DNA microarrays can contain thousands of probes, which can help in the determination of the presence of many genes in multiple samples in parallel. Microarrays can include an entire genome of a particular species.

DNA microarrays can be used for the detection of differences in gene expression under different conditions such as healthy vs. disease, before or after treatment, one environmental condition vs. another, and cancerous vs. non-cancerous tissue. They can also be used for the detection of genes expressed at a given time, detection of SNPs, detection of genotype or resequencing of mutant genomes, multiple samples in the same experiment, genetic footprinting, and species cross-hybridization (existence of a specific pathway in a related species).

If the sequences are not known, then microarray analysis cannot be performed because the sequences are required to create the probes. In this case, next-generation sequencing (NGS) should be used. If there is only one gene of interest, then reverse transcriptase-polymerase chain reaction (RT-PCR) should be used instead.