Thermo Fisher SCIENTIFIC

The A to Z of microarrays—evolution of a revolutionary solution

Microarrays

In this white paper, we discuss:

- How microarrays have been instrumental in human genomic analyses
- Advancements in microarray technology that increase efficiency
- Types of research applications that utilize microarrays

Introduction

The Human Genome Project was a monumental effort over more than a decade to determine the human DNA sequence and analyze genetic variation among individuals. The hope was that, armed with this understanding, phenotypic differences would be explainable by genetic variation, and health benefits could follow. However, there was a need for efficient solutions that could collect and analyze all the sequence information being generated. Some of the initial attempts to determine presence, absence, or abundance of specific sequences involved hybridizing radioactive or other labeled probes to sequences immobilized on nitrocellulose or nylon membranes [1]. However, these methods did not enable analysis of large numbers of sequences at a time.

The rise of microarrays

To meet researchers' need to efficiently query large numbers of sequences in a single experiment, DNA microarrays were developed. Glass or silicon chips were manufactured with oligonucleotides of defined sequence immobilized on very small, specific regions on the substrate. Sample DNA was labeled with a fluorophore and hybridized to the chip. The presence or absence of sequences was determined by measuring the fluorescence of each immobilized oligonucleotide, thus revealing the presence or absence of a sequence in that sample [2]. DNA microarrays revolutionized genomic analysis by providing researchers with the ability to query hundreds to thousands of defined sequences simultaneously.

There are two basic strategies for using oligonucleotide-based arrays for analyzing human genomic sequences. One of these, array comparative genomic hybridization (aCGH), measures the relative abundance of sequences in test and normal samples. For this method, the genomic DNA of a test sample is labeled with one fluorophore, and the genomic DNA of the normal sample is labeled with a different fluorophore. The labeled DNAs are mixed in equal proportions and hybridized to the array. The test and control DNA competitively hybridize to their complementary sequences. The brightness measured is directly proportional to the abundance of a sequence in the sample. The fluorescence signal of both colors is measured at every position on the array. Various manipulations of the signal measurements are used to convert the fluorescence signals to a ratio of test-to-control intensity. A typical clinical CGH microarray contains a few hundred thousand probes while the number of probes on research CGH microarrays may be in the millions.

Another approach, known as a single-nucleotide polymorphism (SNP) array, was developed by Affymetrix, now part of the Applied Biosystems™ brand at Thermo Fisher Scientific. In these arrays, oligonucleotide DNA probes, based on regions in the genome that show single-nucleotide diversity among individuals, are immobilized on a glass or silicon wafer. The genomic DNA is labeled with a fluorescent dye, and after hybridization and washing, the absolute fluorescence at each position with an immobilized oligonucleotide is measured. Unlike aCGH analysis, only genomic DNA from a test sample is needed; no simultaneous analysis of a normal control sample is necessary. The presence or absence of sequences is determined by comparing fluorescence intensities with numerous normal control samples that were run independently and combined to create a reference data set. The resulting ratio between test and normal reference samples is calculated *in silico*. Although the first arrays were originally designed for identifying HIV sequences, they were rapidly recognized as being efficient tools for analyzing human genomic sequences.

The original SNP genotyping arrays encapsulated the oligonucleotide-containing microarray chip in a cartridge that was used for automated hybridization, washing, and visualization of the results. One of the most popular SNP arrays, the Applied Biosystems™ Genome-Wide Human SNP Array 6.0, was used in over 2,000 genome-wide association studies (GWAS). Moreover, public data repositories that contain data from hundreds of thousands of gene chips, such as the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) for gene expression experiments [3] and the Database for Genotypes and Phenotypes (dbGaP) for genotyping [4], have facilitated understanding of the relationship between DNA variation and disease.

Although a tremendous amount of research has been performed using these SNP arrays [5], numerous improvements to the basic SNP array have been developed. For example, the density of probes on a single chip has increased—current chips can query up to 6 million unique sequences in a single-chip hybridization. Some arrays have been developed that contain large numbers of both SNP probes and non-polymorphic probes, facilitating highly accurate analysis of copy number variation. In addition, the ease and relatively low cost of synthesizing oligos on chips means that novel sequences, such as new variants or long intergenic noncoding RNA (lincRNA) genes, can easily be incorporated as their importance becomes recognized by the research community.

Advances leading to Axiom genotyping arrays

The latest innovations in bioinformatics, biofabrication, and chemistries have been incorporated into the latest generation of SNP arrays, the Applied Biosystems™ Axiom™ family of genotyping microarrays [6]. Axiom predesigned whole-genome and custom arrays are synthesized *in situ* using a proprietary photolithographic template, which is designed for assay-to-assay consistency between and across manufacturing batches. The arrays allow analysis of 1,500–1,000,000 markers per sample. Millions of wet lab–verified markers in the Axiom database are available, covering diverse populations. Axiom genotyping arrays use unique imputation algorithms that leverage sequence data from private initiatives or phase 3 of the 1000 Genomes Project to design array content. The flexible, innovative, and simplified custom design pipeline and available bioinformatics support facilitate complex assay design. Arrays can be custom designed for any human population or nonhuman organism, providing tremendous genotyping flexibility and potential. Finally, the automated, hands-free microarray processing on the Applied Biosystems™ GeneTitan™ Multi-Channel (MC) Instrument means there is minimal manual intervention and it provides run-to-run consistency.

The basic workflow is straightforward and easy to follow (Figure 1). Genomic DNA is isothermally amplified and randomly fragmented into 25–125 base pair (bp) fragments. These fragments are purified, resuspended, and hybridized to customized Axiom array plates. Following hybridization, the bound target is washed under stringent conditions to remove nonspecific background. Each polymorphic nucleotide is queried via a ligation event carried out on the array surface. After ligation, the arrays are stained and imaged

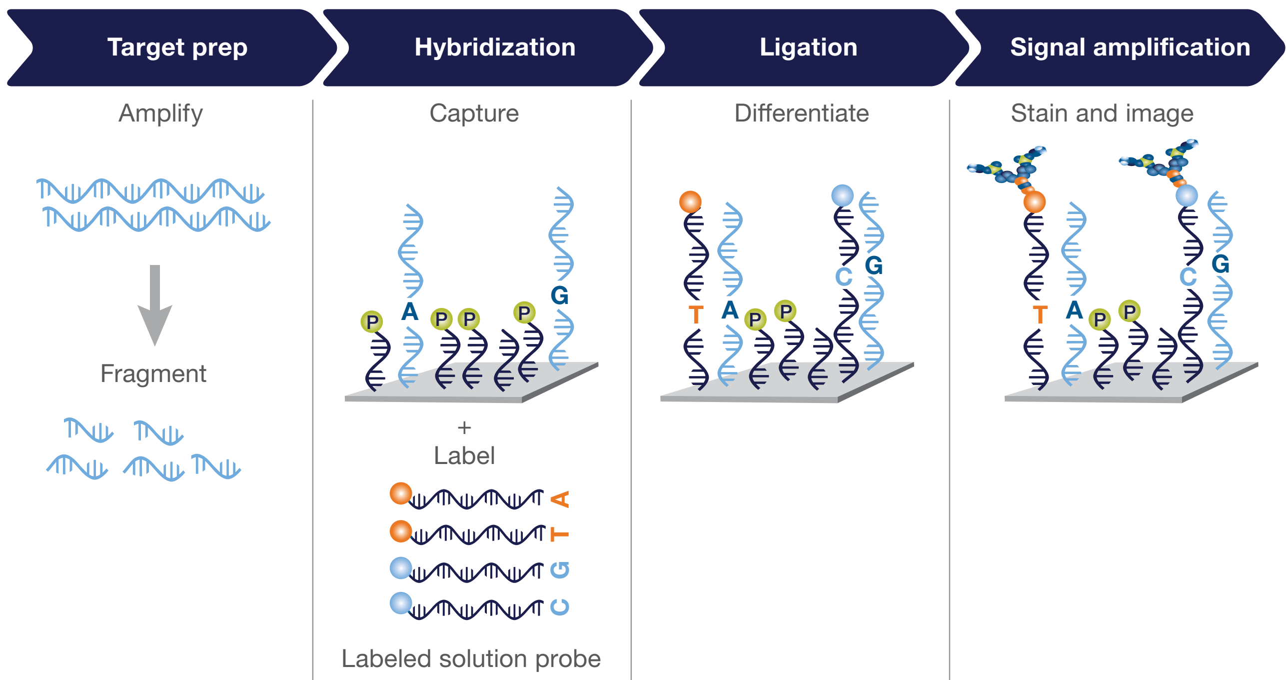

Figure 1. Applied Biosystems™ Axiom™ microarray biochemistry and workflow overview.

on the GeneTitan MC instrument. This basic workflow was adapted to work either manually or on high-throughput liquid handling robots. Optimized protocols have been developed for automation on Hamilton NIMBUS®, Tecan™, or Beckman Coulter™ instruments.

The revolutionary power of the Axiom genotyping solution comes from the ability to simultaneously analyze a large number of arrays. To do this, the silicon-based microarrays are themselves arrayed on pillars. Plates with 96 or 384 arrays, each of which is the size of a 96-well microtiter plate, allow researchers to process 96 or 384 samples in parallel (Figure 2).

Generating large amounts of data requires a method for analyzing and extracting meaningful insights from the data. The Applied Biosystems™ Axiom™ Analysis Suite is a simple-to-use software package that has been designed to assess data generated from Axiom arrays. The software features an easy-to-use graphical interface and data visualization that automates standard genotyping workflows and generates accurate results in a single step. It integrates SNP genotyping, insertion/deletion (indel) detection, multiallelic variant analysis, copy number analysis, and off-target variant (OTV) calling of simple and complex genomes into its research workflow. Data can be exported in multiple formats for further analysis if needed. This package enables researchers to discover and quickly extract genotyping information with newly added visualization capabilities.

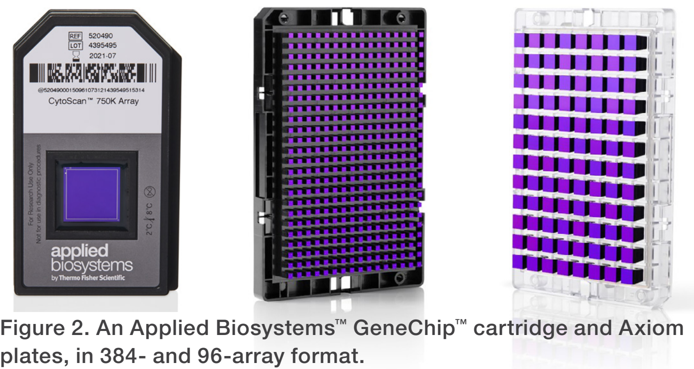

Figure 2. An Applied Biosystems™ GeneChip™ cartridge and Axiom plates, in 384- and 96-array format.

Advances leading to CytoScan arrays

Applied Biosystems™ CytoScan™ hybrid SNP arrays contain probes against polymorphic SNPs and against nonpolymorphic sequences [7]. This combination of features ensures confidence in identifying single-nucleotide variants as well as breakpoint determination for independent confirmation of copy number events throughout the entire genome [8].

The high-density Applied Biosystems™ CytoScan™ HD Array includes 2.67 million markers for copy number (CN) analysis, including 750,000 SNP probes and 1.9 million nonpolymorphic probes for comprehensive whole-genome coverage. The array was designed by selecting probes from a pool of over 20 million loci that were then analyzed with over 3,000 samples to choose those that performed best for whole-genome cytogenetic applications. CytoScan arrays cover all genes in the genome, thus helping to future-proof the technology investment and eliminate revalidation burden. They identify many types of chromosomal aberrations—including large deletions and duplications, copy number gains and losses, and copy-neutral events such as absence of heterozygosity (AOH)—at high resolution in a single test. CytoScan arrays provide sensitive mosaic detection that can elucidate patterns of clonal evolution, structural inconsistencies, and cellular contamination, for genomic research.

Research areas that benefit from microarray technologies

The analysis of nucleic acids on microarrays is useful for many different types of studies. In general, due to the large number of sequences targeted, microarrays are ideally suited for unbiased queries of large numbers of sequences, for large-scale database creation and data stratification amongst disease, ethnicity, and other phenotypes. For example, high-density SNP arrays are very useful for identifying single-nucleotide differences among known sequences in different genomes. These comparisons could be used in oncology research to compare genomes in tumor vs. normal samples; to find allelic variants that may be associated with inherited traits or diseases in GWAS; for phamacogenomic studies that help understand drug safety and efficacy; or for understanding genetic differences in populations and how those relate to health problems. Axiom genotyping solutions are ideal for research in these areas [9].

Hybrid SNP arrays such as the CytoScan family of arrays are used extensively for prenatal and postnatal genetic studies, carrier screening, and tumor analyses. These arrays can reveal a wide variety of genetic anomalies, such as chromosomal aberrations, loss of heterozygosity, regions of uniparental disomy, or chromothripsis (Applied Biosystems™ OncoScan™ and CytoScan assays) [10,11]. Analysis using such hybrid arrays are useful for providing a deeper understanding of certain types of cancer, and for reproductive health research to identify chromosomal aberrations and the genetic causes of developmental dysmorphologies. In fact, they are recommended by several professional health societies for these types of analyses [12-14].

Arrays can also be used to analyze gene expression. Here, oligonucleotide probes are synthesized on the array as for Axiom and CytoScan products, and the array is hybridized with labeled cDNA. Applied Biosystems™ Clariom™ assays are efficient tools for finding high-fidelity gene expression biomarkers. In addition to protein-coding regions, these arrays can also have oligonucleotide probes for splice variants, predicted genes, and lncRNAs and miRNAs, giving a complete picture of the transcriptome. Clariom assays are therefore ideal for comparative transcriptomic experiments [15].

Finally, arrays can also be used in nonhuman applications, such as agrigenomics. Research in this area is growing as climate change, population growth, and urbanization threaten farmers' ability to meet the world's food demands. To address these needs, breeders and farmers are employing new genomic strategies to use fewer environmental resources to develop higher-producing livestock, poultry, and crops, and to use fewer antibiotics and pesticides. Arrays are used for a wide variety of applications. Genomic selection aims to improve quantitative traits in large breeding populations using whole-genome molecular markers. Marker-assisted selection (MAS) and marker-assisted breeding (MAB) focus on selecting desirable traits during breeding, where a trait of interest is selected based on a marker linked to that trait. Parentage is another application area where breeders can confidently select animals, knowing their ancestry is correct. Lastly, arrays are used for the characterization of genetically modified organisms (GMOs) to increase yield, decrease susceptibility to disease, and minimize the use of antibiotics in production facilities.

Thermo Fisher SCIENTIFIC

Conclusions

Harnessing the cornucopia of information coming from genomic studies requires methods that can efficiently analyze large numbers of sequences. Genome-wide association studies (GWAS), often using microarrays to analyze genotypes, have linked variations in the genome with either human traits or pathologies [16]. Additionally, the ability to analyze millions of genomic loci has been an enormous boon in non-human applications, as exemplified by the benefits in breeding for food production. Axiom, OncoScan, CytoScan, and Clariom assays, leveraging years of advances in bioinformatics, biofabrication science, and chemistries, make them a far cry from the arrays that were based on the techniques and data in the early Human Genome Project era. These oligonucleotide-based microarrays have been and will continue to be fulfilling the needs of health researchers at the cutting edge.

References

1. Green MR and Sambrook J (2012) Molecular cloning: a laboratory manual. ISBN 978-1-936113-42-2. Cold Spring Harbor Laboratory Press.
2. **nature.com/scitable/definition/microarray-202/**
3. **ncbi.nlm.nih.gov/geo/**
4. **ncbi.nlm.nih.gov/gap/**
5. Buniello A et al. (2019) The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 47(D1):D1005-D1012 (2019). doi: 10.1093/nar/gky1120.
6. **thermofisher.com/document-connect/document-connect.html?url=https://assets.thermofisher.com/TFS-Assets/BID/brochures/axiom-genotyping-arrays-human-studies-brochurepdf&title=QnJvY2h1cmU6IEF4aW9tIGdlbm90eXBpbmcgYXJyYXlzIGZvciBodW1hbiBzdHVkaWVz**
7. **assets.thermofisher.com/TFS-Assets/LSG/Flyers/not-all-arrays-are-created-equal.pdf**
8. **thermofisher.com/us/en/home/life-science/microarray-analysis/human-genotyping-pharmacogenomic-microbiome-solutions-microarrays.html**
9. **assets.thermofisher.com/TFS-Assets/GSD/Reference-Materials/karyotyping-vs-cma-white-paper.pdf**
10. **thermofisher.com/content/dam/LifeTech/Documents/PDFs/oncoscanassays-whole-genome-copy-number-analysis-for-ffpe-tumor-samples-whitepaper.pdf**
11. **assets.thermofisher.com/TFS-Assets/GSD/Reference-Materials/the-power-of-snps-white-paper.pdf**
12. Kearney HM et al. (2011) American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genet Med* 13:680–685.
13. Shao L et al. (2021) Chromosomal microarray analysis, including constitutional and neoplastic disease applications, 2021 revision: a technical standard of the American College of Medical Genetics and Genomics (ACMG). *Genet Med* 23:1818–1829.
14. Silva M et al. (2019) European guidelines for constitutional cytogenomic analysis. *Eur J Hum Genet* 27:1–16.
15. **assets.thermofisher.com/TFS-Assets/GSD/Technical-Notes/transcriptome-microarray.pdf**
16. Uffelmann E et al. (2021) Genome-wide association studies. *Nature Reviews Methods Primers* 1:59. **doi.org/10.1038/s43586-021-0005**

Learn more at **thermofisher.com/microarrays**

applied biosystems

For Research Use Only. Not for use in diagnostic procedures. © 2023 Thermo Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified. NIMBUS is a registered trademark of Hamilton Co. Tecan is a trademark of Tecan Group Ltd. Beckman Coulter is a trademark of Beckman Coulter Inc. **COL27666 0623**