Predictive genomics

Microarrays: an important tool for predictive genomics

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
- How microarrays benefit population research
- How microarrays are used for pharmacogenomics

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, is used in Applied Biosystems™ microarray products from 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 fluorescent 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 of 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 implemented. 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 nonpolymorphic 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, fabrication, 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 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

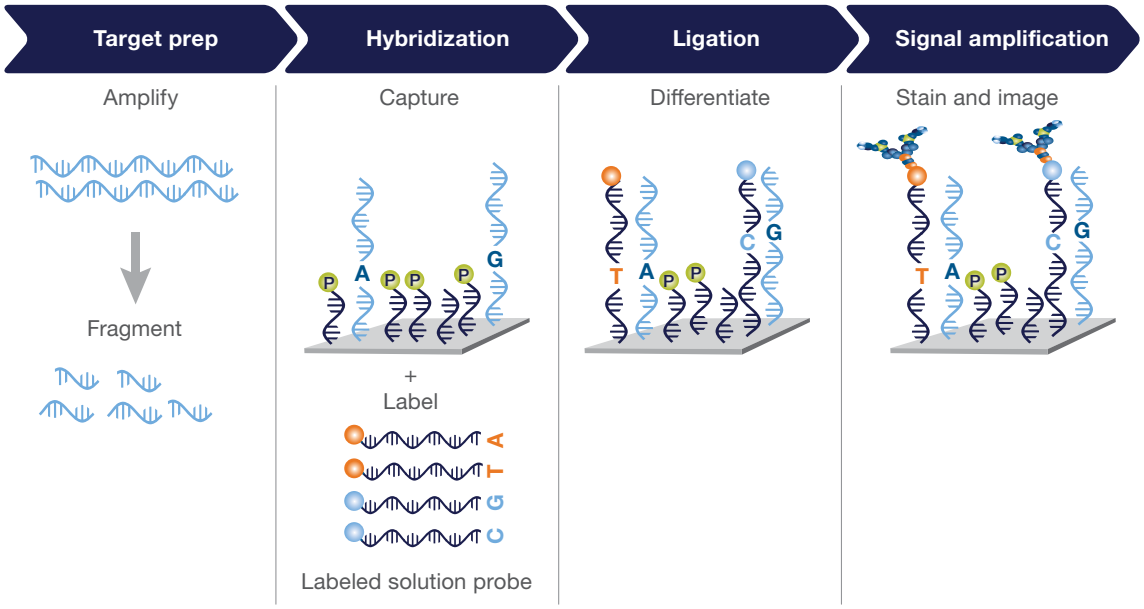

Figure 1. Overview of Axiom microarray chemistry.

for automation on Hamilton NIMBUS®, Tecan™, or Beckman Coulter™ instruments. Arrays can be custom designed for human populations or nonhuman organisms, providing tremendous genotyping flexibility and potential.

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 best-practice genotyping workflows and generates accurate results in a single step. It integrates single-nucleotide polymorphism (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 recognize 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 are used in genomic research to help advance the development of prenatal and postnatal genetic screening, carrier screening, and tumor analysis. 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]. Analyses 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, lncRNAs, and miRNAs, giving a complete picture of the transcriptome. Clariom assays are therefore ideal for comparative transcriptomic experiments [15].

Next-generation sequencing and predictive genomics

Next-generation sequencing (NGS) has emerged as a powerful technology that can be used for discovery-based research and answer many of the same questions as microarrays. However, NGS strategies might not be appropriate for all population-scale predictive genomics investigations. One strategy using NGS is whole-genome sequencing (WGS). WGS produces large amounts of data, but generating high-quality data can be costly on a per-sample basis, and can be low in throughput. A minimum ≥20x (recommended 30x–50x) average read depth is theoretically required for *de novo* variant discovery and direct calling of genotypes from sequence data. This read depth requirement affects scalability and cost. Moreover, data generation and analysis can be very complex, often requiring days or weeks of analysis before final answers are known. However, WGS is ideal for discovering putative new sequence variants.

Another strategy is low-pass sequencing (LPS). This is an emerging approach that is becoming more popular with researchers. Typically, this involves sequencing at a 0.4x to 4x average read depth. This prioritizes cost per sample, but at the expense of data quality and the ability to identify rare variants. Some *de novo* variant discovery is possible, but all genotypes must be imputed and there is no direct genotyping option for high-accuracy calling of rare variants or critical biomarkers. And data generation and analysis are still highly complex. While there are definite advantages for NGS-based approaches, microarrays provide a cost-effective and scalable (millions of samples) strategy for analyzing genotypic variation in populations. For more details, see reference 13.

Predictive genomics and disease risk

Using microarrays, researchers have made tremendous progress translating genetic variation into disease risk. Since 2003, more than 900 peer-reviewed publications have reported GWAS results (750 human) using Applied Biosystems microarrays. One of the ways researchers quantify the link between genetic load and disease risk is to calculate a polygenic risk score (PRS). The National Cancer Institute defines this as “an assessment of the risk of a specific condition based on the collective influence of many genetic variants. These can include variants associated with genes of known function and variants not known to be associated with genes relevant to the condition” [14]. Once a PRS has been associated with a disease, clinical researchers can use microarrays or other genotyping methods to better understand how susceptible an individual is to that disease.

In many cases, the efficiency and relatively low cost of microarrays provide the means for researchers to discover loci that are used to calculate risk scores. These are later refined with larger cohorts using microarrays or other technologies. Some recent examples of this approach include PRS calculations and refinements for adenomatous colon polyps [15], prostate cancer aggressiveness [16], and Alzheimer’s disease [17]. In some cases, PRS determination may be more accurate for a population if the background ancestry of a population is considered. For example, childhood myopia is a common disorder in East Asian populations.

Chen et al. identified polymorphisms that suggested increased risk in Han Chinese children [18], and in a different study Hsu et al. identified variants that are associated with Gilbert's syndrome in the Taiwanese population [19]. Thus, data generated on microarrays continue to have an important role for discovering variants that can be linked to disease risk.

Predictive genomics are facilitated by population genomics

The major goal of human genomic analysis is to relate phenotypic variation to genotypic variation. Phenotypes may manifest as differences in observable traits, in disease etiology or response, or even in behaviors. GWAS have been enormously successful for establishing some of these relationships. The power and accuracy of GWAS associations always increases when more sample health data are collected and analyzed. Therefore, there is great interest from national and government organizations to set up biobanks that can collect health-based phenotypic data as well as genotypic data from the same individuals.

One of the first organizations to undertake such a collection was the UK Biobank (UKBB). Its database includes information from over half a million individuals, collected between 2006 and 2010 [20]. The UKBB collected extensive phenotypic data related to lifestyle, health, and other biometric values. In addition, the UKBB collected genomic information on these individuals using custom Axiom genotyping microarrays [21]. These arrays were designed by and for the researchers for high-throughput, high-value genotyping of large sample cohorts with a single, comprehensive, low-cost solution. The arrays are on one 96-array plate that allows for genome-wide genotyping of large sample collections, and allowed the UKBB to collect data on the hundreds of thousands of samples efficiently. As of 2022, over 500 peer-reviewed GWAS studies have been published using this resource [20]. For example, Hyams et al. used the Applied Biosystems™ UK Biobank Axiom™ Array to characterize genetic contributors in children with ulcerative colitis who are refractory to current therapies [22]. Another study identified loci involved in heart development and failure [23]. In a study of severe mental illness, Kochunov et al. were able to distinguish between brain regional vulnerability index (RVI) measurements and genome-wide biomarkers using genetic data from the UKBB [24]. Finally, polygenic risk scores were developed and validated for congenital heart disease [25] and endometrial cancer [26] using UKBB data.

Although the UKBB is able to leverage the diversity of the UK population for their biobank, other insights can be gained by performing similar studies on more isolated populations. This is because deleterious alleles, resulting from bottlenecks, might be more common in such populations. The population of Finland is thought to have undergone such a population bottleneck about 120 generations in the past. The FinnGen project is a collaborative biobanking project that makes use of health demographic information and genotyping information of the Finnish population [27]. For these studies, a custom Axiom genotyping array was designed with alleles and variants more commonly seen in Finnish people. Some of the preliminary findings using this array and biobank have recently been described [27,28].

Similarly, the Million Veterans Program has collected samples and genetic information from US military veterans. For the genotyping study, they used a custom Axiom chip built using variants commonly seen in the US population, including people of European, Asian, African, and Hispanic ancestry [29]. Data from this array were recently used to identify genetic risk loci for suicidal thoughts among US military veterans [30].

As successful as these studies have been, it is becoming clear that although some variants can indeed be directly associated with traits, there can be tremendous variation in the influence of these variants in different individuals. Thus, not only the primary variant but also the background variants (genetic load) of an individual can be just as important. And these background variants can be significantly associated with ethnic background. Therefore, to determine the link most accurately between variation and phenotypic or pathological traits, the background variation, which might be population-derived, should be considered. Indeed, most of the GWAS studies and risk scoring that have been performed have been in populations of mostly European origin—and thus when results are applied to other ancestries, fail to accurately predict outcomes (for example, see [31]). This can exacerbate health disparities and may miss important insights into disease biology.

To begin to collect genomic data from divergent ancestries, some biobanks have developed Axiom microarrays specific for their populations. For example, Thermo Fisher Scientific and the Qatar Genome Program will develop an Axiom custom genotyping array for pan-Arab populations using whole-genome sequencing data from 19 Arab countries [32]. Similarly, the Korean Genome and Epidemiology Study (KoGES) developed the Korea Chip [33], to be used for studies focused on Korean, Japanese, and other East Asian populations. Using this array, Nam et al. identified 379 novel associations and demonstrated improved risk scores for East Asian populations [34]. And the Taiwan Precision Medicine Initiative (TPMI) has a goal of building a biobank of 1 million people in Taiwan that would include health and genetic information.

To generate the genetic information, they developed a custom Axiom array that is enriched for variants found in Han Chinese population [35]. Ko et al. used this array and sample collection to identify SNPs associated with anklyosing spondylitis that were specific to Han Chinese populations—and were not identified in studies focused on non-Chinese populations [36]. These studies demonstrate that fully understanding the genetic contribution to a trait in a specific population requires analysis of the underlying genetic background of that population. Custom Axiom microarrays are ideal for this kind of study.

Preemptive pharmacogenomics—predictive genomics to reduce ADEs

A major downside of pharmaceutical treatments, one that has the highest impact on individuals and health care costs, is when a drug produces an unwanted effect or injury—also known as an adverse drug event (ADE). According to the U.S. Department of Health and Human Services (HHS) Office of Disease Prevention and Health Promotion [37], annually, adverse drug effects can:

- Account for an estimated 1 in 3 of all hospital adverse events
- Affect about 2 million hospital stays each year
- Prolong hospital stays by 1.7 to 4.6 days
- Account for over 3.5 million physician office visits
- Result in an estimated 1 million emergency department visits
- Drive approximately 125,000 hospital admissions

ADEs cause discomfort and stress for those affected. But ADEs are also very expensive to manage, and therefore contribute to high health care costs. For more information, see [38].

In many cases, an ADE is influenced by the genotype of the individual taking the drug. For example, the variants present in drug metabolizing enzyme genes can affect the rate at which drugs are cleared from an individual's system—if they are not cleared quickly enough, then an overdose-like reaction might occur. Many other aspects of the human genome play a role in drug safety [39], and if these can be predicted, both the patient (less stress) and the health care provider (reduced costs) stand to benefit [40,41].

Pharmacogenomics aims to understand the relationship between a patient's genotype and their response to a drug, and use that information to predict how a patient might minimize side effects and better respond to the drug treatment. To facilitate these investigations, Thermo Fisher Scientific developed specialized microarrays that contains alleles known to affect adsorption, distribution, metabolism, and excretion (ADME) gene function. One of these, the Applied Biosystems™ Axiom™ Precision Medicine Research Array, contains probes for rare inherited diseases, genetic risk profiling, immune response, and pharmacogenomics research. Another array, the Applied Biosystems™ PharmacoScan™ array, is a hybrid array that focuses on ADME genes; both arrays come in the Axiom format for high-throughput analysis. The Association for Molecular Pathology and College of American Pathologists used the PharmacoScan array to provide recommendations for genotyping for warfarin responsiveness [42]. Finally, one study found that the combination of the Axiom Precision Medicine and PharmacoScan arrays had the best genome-wide and pharmacogene coverage of commercial arrays—and this is crucial for discovering new variants that may be involved in drug metabolism and drug adverse effects [43].

Conclusions

Harnessing the cornucopia of information coming from human genomic studies requires ways to 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 [44]. The 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 fulfilled, and will continue to fulfill, the needs of health researchers at the cutting edge.

Predictive genomics is the next step in leveraging the genotypic information. Axiom microarrays have been adopted as the genotyping platform of choice by several states and organizations due to their efficiency, customizability, and reproducibility. These features allow them to inform health care decisions for the populations served by these organizations.

Glossary

- **Predictive genomics:** Determining phenotypic outcomes in areas such as complex multifactorial diseases. Predictive genomics is at the intersection of multiple disciplines: predictive medicine, personal genomics, and translational bioinformatics.
- **Polygenic risk score (PRS):** An assessment of the risk of a specific condition based on the collective influence of many genetic variants. These can include variants associated with genes of known function and variants not known to be associated with genes relevant to the condition.
- **Genetic load:** The influence in fitness of the average individual in a population relative to the fittest genotype due to the presence of deleterious or other interactive genes in the gene pool.
- **Phenotype:** The observable characteristics or traits of an organism.
- **Genotype:** All or part of the genetic constitution of an individual or group.
- **Pharmacogenomics:** Field of research concerned with understanding how genetic differences among individuals cause varied responses to the same drug and developing drug therapies to compensate for these differences.
- **Adverse drug event (ADE):** An injury resulting from the use of a drug. Under this definition, the term ADE includes harm caused by the drug (adverse drug reactions and overdoses) and harm from the use of the drug (including dose reductions and discontinuations of drug therapy).
- **Drug metabolizing enzyme (DME):** Enzymes that are usually responsible for breaking down drugs and may include mixed-function oxidases or monooxygenases. May include cytochrome P450, cytochrome b5, and NADPH-cytochrome P450 reductase and similar enzymes.
- **Adsorption, distribution, metabolism, and excretion (ADME):** Four key properties that influence the pharmacokinetic behavior of a drug.

References

1. Green MR, Sambrook J (2012) *Molecular Cloning: A Laboratory Manual.* ISBN 978-1-936113-42-2. Cold Spring Harbor Laboratory Press.
2. Scitable. Microarray definition. **nature.com/scitable/definition/microarray-202/**
3. National Center for Biotechnology Information. Gene Expression Omnibus (GEO). **ncbi.nlm.nih.gov/geo/**
4. National Center for Biotechnology Information. Database of Genotypes and Phenotypes (dbGaP). **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. **doi.org/10.1093/nar/gky1120**
6. Thermo Fisher Scientific (2018) Axiom genotyping arrays for human studies. **thermofisher.com/TFS-Assets/BID/brochures/axiom-genotyping-arrays-human-studies-brochure.pdf**
7. Thermo Fisher Scientific (2021) Not all chromosomal microarrays for cytogenetics are created equal. **assets.thermofisher.com/TFS-Assets/LSG/Flyers/not-all-arrays-are-created-equal.pdf**
8. Thermo Fisher Scientific. CytoScan HD Array Kit and Reagent Bundle. **thermofisher.com/order/catalog/product/901835**
9. Thermo Fisher Scientific. Microarray analysis. **thermofisher.com/us/en/home/life-science/microarray-analysis/human-genotyping-pharmacogenomic-microbiome-solutions-microarrays.html**
10. Thermo Fisher Scientific (2019) OncoScan CNV and CNV Plus Assays: comprehensive solutions for whole-genome copy number analysis of FFPE tumor samples. **thermofisher.com/content/dam/LifeTech/Documents/PDFs/oncoscan-assays-whole-genome-copy-number-analysis-for-ffpe-tumor-samples-white-paper.pdf**
11. Thermo Fisher Scientific (2020) The power of high-resolution SNP arrays for detecting inherited chromosomal anomalies. **assets.thermofisher.com/TFS-Assets/GSD/Reference-Materials/the-power-of-snps-white-paper.pdf**
12. Thermo Fisher Scientific (2019) Clariom D microarrays provide a deep view of our transcriptome. **assets.thermofisher.com/TFS-Assets/GSD/Technical-Notes/transcriptome-microarray.pdf**
13. Thermo Fisher Scientific (2022) Navigating the technology landscape for population-scale predictive genomics. **assets.thermofisher.com/TFS-Assets/GSD/Reference-Materials/Whitepaper-Technology-Predictive-Genomics.pdf**
14. National Cancer Institute. Polygenic risk score. **cancer.gov/publications/dictionaries/genetics-dictionary/def/polygenic-risk-score**
15. Northcutt MJ et al. (2021) Polygenic risk score is a predictor of adenomatous polyps at screening colonoscopy. *BMC Gastroenterol* 21:65. **doi.org/10.1186/s12876-021-01645-4**
16. Brezina S et al. (2020) Genome-wide association study of germline copy number variations reveals an association with prostate cancer aggressiveness. *Mutagenesis* 35:283–290. **doi.org/10.1093/mutage/geaa010**
17. Escott-Price V et al. (2017) Polygenic risk score analysis of pathologically confirmed Alzheimer's disease. *Ann Neurol* 82(2):311–314. **doi.org/10.1002/ana.24999**
18. Chen LJ et al. (2021) Association of polymorphisms in *ZFHX1B*, *KCNQ5* and *GJD2* with myopia progression and polygenic risk prediction in children. *Br J Opthalmol* 105:1751–1757. **doi.org/10.1136/bjophthalmol-2020-318708**
19. Hsu PW-C et al. (2022) The mutation hotspots at *UGT1A* locus may be associated with Gilbert's syndrome affecting the Taiwanese population. *Int J Mol Sci* 23:12709. **doi.org/10.3390/ijms232012709**
20. Tan VY et al. (2022) The UK Biobank: A shining example of genome-wide association study science with the power to detect the murky complications of real-world epidemiology. *Ann Rev Genom Hum Genet* 23:569–589. **doi.org/10.1146/annurev-genom-121321-093606**
21. Bycroft C et al. (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562:203–209. **doi.org/10.1038/s41586-018-0579-z**
22. Hyams JS et al. (2021) Clinical and host biological factors predict colectomy risk in children newly diagnosed with ulcerative colitis. *Inflamm Bowel Dis* 28:151–160. **doi.org/10.1093/ibd/izab061**

23. Aung N et al. (2021) Genome-wide analysis of left ventricular image-derived phenotypes identifies fourteen loci associated with cardiac morphogenesis and heart failure development. *Circulation* 140:1318–1330. **doi.org/10.1161/circulationaha.119.041161**
24. Kochunov P et al. (2022) Brain-wide versus genome-wide vulnerability biomarkers for severe mental illness. *Hum Brain Mapp* 43:4970–4983. **doi.org/10.1002/hbm.26056**
25. Spendlove SJ et al. (2022) Polygenic risk scores of endo-phenotypes identify the effect of genetic background in congenital heart disease. *Hum Genet Genom Adv* 3:100112. **doi.org/10.1016/j.xhgg.2022.100112**
26. Bafligil C et al. (2022) Development and evaluation of polygenic risk scores for prediction of endometrial cancer risk in European women. *Genet Med* 24:1847–1846. **doi.org/10.1016/j.gim.2022.05.014**
27. Kurki MI et al. (2023) FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* 613:508–518. **doi.org/10.1038/s41586-022-05473-8**
28. Heyne HO et al. (2023) Mono- and biallelic variant effects on disease at biobank scale. *Nature* 613:519–525. **doi.org/10.1038/s41586-022-05420-7**
29. Hunter-Zinck H et al. (2020) Genotyping array design and data quality control in the Million Veteran Program. *Am J Hum Genet* 106(4):535–548. **doi.org/10.1016/j.ajhg.2020.03.004**
30. Kimbrel NA et al. (2022) Identification of novel, replicable genetic risk loci for suicidal thoughts and behaviors among US military veterans. *JAMA Psychiatry* 80(2):135–145. **doi.org/10.1001/jamapsychiatry.2022.3896**
31. Sun Q et al. (2022) Analyses of biomarker traits in diverse UK biobank participants identify associations missed by European-centric analysis strategies. *J Hum Genet* 67(2):87–93. **doi.org/10.1038/s10038-021-00968-0**
32. Thermo Fisher Scientific (2022) Thermo Fisher Scientific and Qatar Genome Program partner to advance precision medicine. Press release. **corporate.thermofisher.com/content/tfcorpsite/us/en/index/newsroom/press-releases/2022/May/23-Thermo-Fisher-Scientific-and-Qatar-Genome-Program-Partner-to-Advance-Precision-Medicine.html**
33. Moon S et al. (2019) The Korea Biobank Array: design and verification of coding variants associated with blood biochemical traits. *Sci Rep* 9:1382. **doi.org/10.1038/s41598-018-37832-9**
34. Nam K et al. (2022) Genome-wide study on 72,298 individuals in Korean biobank data for 76 traits. *Cell Genomics* 2:100189. **doi.org/10.1016/j.xgen.2022.100189**
35. Wei C-Y et al. (2021) Genetic profiles of 103,106 individuals in the Taiwan Biobank provide insights into the health and history of Han Chinese. *NPJ Genom Med* 6:10. **doi.org/10.1038/s41525-021-00178-9**
36. Ko C-L et al. (2022) Genome-wide association study reveals ethnicity-specific SNPs associated with ankylosing spondylitis in the Taiwanese population. *J Transl Med* 20:589. **doi.org/10.1186/s12967-022-03701-3**
37. U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion (2022) National action plan for ADE prevention. **health.gov/our-work/health-care-quality/adverse-drug-events#ref1**
38. Thermo Fisher Scientific (2022) Preemptive pharmacogenomics: investigating the relationship between genotype and drug response. **assets.thermofisher.com/TFS-Assets/GSD/Reference-Materials/Preemptive-pharmacogenomics-whitepaper.pdf**
39. Carrs KJ et al. (2023) Using human genetics to improve safety assessment of therapeutics. *Nat Rev Drug Discov* 22:145–162. **doi.org/10.1038/s41573-022-00561-w**
40. Arbitrio M et al. (2021) Pharmacogenomics biomarker discovery and validation for translation in clinical practice. *Clin Transl Sci* 14:113–119. **doi.org/10.1111/cts.12869**
41. Jørgensen JT (2021) Predictive genomics and clinical evidence. *Basic Clin Pharmacol Toxicol* 128:642–648. **doi.org/10.1111/bcpt.13578**
42. Pratt VM et al. (2020) Recommendations for clinical warfarin genotyping allele selection: A report of the Association for Molecular Pathology and the College of American Pathologists. *J Mol Diagn* 22(7):847–859. **doi.org/10.1016/j.jmoldx.2020.04.204**
43. Lemieux Perreault LP et al. (2018) Pharmacogenetic content of commercial genome-wide genotyping arrays. *Pharmacogenomics* 19(15):1159–1167. **doi.org/10.2217/pgs-2017-0129**
44. Uffelmann E et al. (2021) Genome-wide association studies. *Nat Rev Methods Primers* 1:59. **doi.org/10.1038/s43586-021-00056-9**

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

applied biosystems