Human genotyping

Improvements in human genotyping research with SwiftArrayStudio Microarray Analyzer and Axiom PangenomePro Arrays

In this document, we:

- Introduce the PangenomePro Genotyping Microarray
- Describe the applications that can use the PangenomePro array
- Describe the markers and coverage that are new to the PangenomePro array

Introduction

The Human Genome Project provided the first full sequence of the human genome and laid the foundation for modern population genomics. Since then, advances in next-generation sequencing (NGS) have generated massive amounts of genomic data from populations around the world. These data have improved our understanding of genetic variation, disease risk, and ancestry. They also support precision medicine, which depends on knowing specific DNA sequences and how they affect health. This knowledge enables tools such as polygenic risk scores (PRS) to research disease susceptibility, pharmacogenomics to research drug selection and dosing, and population-specific health investigations tailored to genetic backgrounds.

Advances in next-generation sequencing have dramatically expanded our understanding of human genetic variation. However, whole genome sequencing often generates far more data than is required for many research and clinical applications, increasing costs and interpretive complexity. Many detected variants are classified as variants of uncertain significance (VUS), making interpretation difficult for clinicians and confusing for patients. Managing and analyzing these large datasets also requires significant investment in computational infrastructure, data storage, and expert review.

Targeted genotyping microarrays offer a practical alternative by focusing on well-characterized variants relevant to population genetics, disease research, pharmacogenomics, and polygenic risk prediction. They are less expensive to run than whole genome sequencing and generate far smaller datasets, which reduces storage and analysis costs. When the goal is to assess established variants rather than discover new ones, microarrays provide a more efficient and cost-effective approach while still supporting precision medicine initiatives.

The Applied Biosystems™ Axiom™ PangenomePro™ Array, used with the Applied Biosystems™ SwiftArrayStudio™ assay workflow, builds on the earlier Applied Biosystems™ Axiom™ PangenomiX™ design by introducing updated genomic content, expanded population coverage, and faster assay processing. Together, these improvements enable high-throughput genotyping with broad genome-wide coverage and enhanced pathologically relevant variant detection. In this document, we outline some of the key enhancements that are included in the PangenomePro array.

applied biosystems

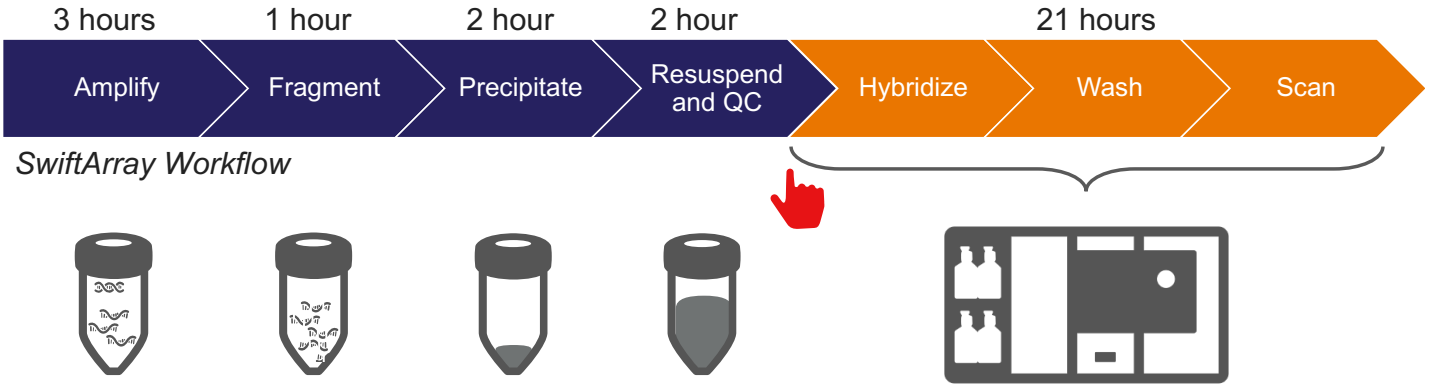

Figure 1. **Introducing the SwiftArray Microarray Analysis solution.** The Applied Biosystems™ SwiftArray™ target preparation steps can be completed in about eight hours in a single working day. On-instrument, autonomous array processing can generate data within 21 hours of loading. Overall, a full set of genotypes and copy number calls from an Applied Biosystems™ Axiom™ plate can be generated in about 30 hours. Different colors represent workflow steps on different days; finger icon represents the touch point on the instrument, where plates and/or reagents are loaded onto the instrument.

What's new in the PangenomePro Array

- Updated GWAS backbone with more than 940,000 imputation-optimized markers
- Expanded population-specific marker coverage (30k–100k additional variants across key populations)
- More than 14,000 curated variants across ACMG 73 genes
- More than 5,500 pharmacogenomic variants with expanded CYP coverage
- Markers supporting rare blood typing and lifestyle traits
- Optimized probe design for challenging GC-rich regions
- Compatibility with SwiftArrayStudio workflow, reducing assay time from 5 days to 2 days (Figure 1)

Thermo Fisher Scientific launched the Applied Biosystems Axiom PangenomiX Array as a human genotyping research tool designed for whole-genome imputation across diverse global populations. Building on this foundation, we introduced several new innovations. Workflow time was reduced from five days to two days with the SwiftArrayStudio Assay and Microarray Analyzer. We also updated the content to create the Axiom PangenomePro array, incorporating the latest genomic insights on populations, diseases, and risk factors and optimizing it for the new assay chemistry. Like the original version, the PangenomePro enables SNP genotyping, copy number analysis, HLA typing, and more in a single, cost-effective assay with streamlined data analysis. Expanded marker coverage and improved design further strengthen its performance for GWAS, population health studies, polygenic risk score research, and clinical research in drug development. This document outlines the key enhancements included in the PangenomePro array.

Imputation-aware design

One of the key distinguishing features of Thermo Fisher Scientific arrays is their imputation-aware design. This helps reduce costs by reducing the number of markers required for genotypic analysis. Genotype imputation is a statistical method used to infer unobserved genetic variants based on known haplotype reference panels and linkage disequilibrium patterns, thereby increasing genomic coverage beyond directly genotyped markers. Rather than examining every position in the genome, researchers genotype a carefully selected subset of variants and computationally predict additional variants with high accuracy. Axiom microarrays are specifically designed to support this approach through proprietary imputation-based SNP selection algorithms that optimize marker content for both common and low-frequency variation across targeted populations. These imputation-aware arrays combine highly accurate direct genotyping with robust downstream imputation, enabling cost-effective whole-genome interrogation. For more details, see our whitepaper highlighting our imputation-aware designs.

Population mapping

These imputation-aware strategies enable more detailed population genotyping studies, opening insights into ancestry and health by enabling large-scale analysis of genetic variation across diverse groups. By interrogating carefully selected markers, researchers can characterize ancestry-specific allele frequencies, identify population structure, and detect variants associated with disease risk. Studies performed with the UK Biobank microarray demonstrate the value of leveraging diverse populations [1], but additional insights often emerge by examining groups where historical bottlenecks or geographic isolation have increased the frequency of certain deleterious alleles [2]. These insights underscore the importance of identifying harmful variants in under-

represented populations, where lack of representation in global reference datasets can otherwise limit risk prediction accuracy, delay diagnosis, and widen health disparities.

The Axiom PangenomePro Array GWAS module integrates more than 940,000 markers selected using a proprietary, imputation-driven strategy designed to maximize genome-wide coverage across European, African, admixed American, East Asian, and South Asian populations (Table 1). This approach facilitates high-resolution imputation performance, accurately capturing both common variants (minor allele frequency [MAF] >5%) and low-frequency variants (MAF >1%) across diverse ancestries. The array also supports robust population genetics and ancestry analyses through inclusion of autosomal, mitochondrial, and Y-chromosome markers, enabling fine-scale ancestry inference and migration pattern studies. Its design emphasizes efficiency, optimizing marker content to minimize redundancy while preserving comprehensive genomic representation, thereby supporting statistically powered, large-scale association analyses. The PangenomePro array expands the range of populations covered, with 30,000-100,000 novel markers with wide population distribution. In addition, it has 1356 new markers that are unique for the populations listed in Table 1. Collectively, these features position the array as a scalable and analytically robust genotyping solution for GWAS, biobank initiatives, and population-scale genomics research.

Population	Number of new probes for populations	Number of probes unique for population
ASW=Americans of African Ancestry in SW USA	72081	100
CEU=Utah Residents (CEPH) with Northern and Western European ancestry	108282	62
CHB=Han Chinese in Beijing, China	36780	42
CHS=Southern Han Chinese	32156	48
CLM=Colombians from Medellin, Colombia	100663	40
FIN=Finnish in Finland	102225	29
GBR=British in England and Scotland	106292	55
JPT=Japanese in Tokyo, Japan	30864	93
LWK=Luhya in Webuye, Kenya	42503	606
MXL=Mexican Ancestry from Los Angeles USA	89852	27
PUR=Puerto Ricans from Puerto Rico	103219	48
TSI=Toscani in Italy	107101	95
YRI=Yoruba in Ibadan, Nigeria	34250	111

Table 1. Ancestry-derived markers present on the PangenomePro microarray. The first column indicates the number of new markers that are found in the indicated populations. Note that there may be overlap (i.e., one marker might be found in different populations, at different frequencies). The second column indicates the number of new markers that are found ONLY in that population.

Classification	Total number of PangenomePro Array markers	Markers novel for PangenomePro array
AIMs 10K	9,294	7,867
Blood	1,247	521
Cancer	5,940	183
ClinVar ACMG	530	1,396
Copy Number Aware regions	4,726	3,258
DTC genetic testing	1,356	125
eQTLs	31,308	1
FORCE alleles	4,840	3,445
GWAS grid	847,754	103,957
HLA	10,274	28
Nonsynonymous	79,543	55
Pharmacogenetics/ADME	5,912	4,353
Published GWAS hits	32,362	89
Sample tracking: fingerprint	344	0
Y and MT markers	3,389	1,444

Table 2. Functional classification of the markers on the PangeomePro array. In the first column are classifications for all the markers; in the second column are the markers that are new to the PangenomePro array. AIMs: Ancestry-informed markers [3]; GWAS Grid: markers chosen specifically for imputation; FORCE: markers from the FORensic Capture Enrichment (FORCE) panel for forensic applications [4]; DTC: markers commonly found in direct-to-consumer applications; GWAS: genome-wide association study; Y: Y chromosome markers; MT: mitochondrial markers; eQTLs: markers identified as important expression quantitative trait loci.

Inherited disease and somatic cancer screening research

Genotyping microarrays play a central role in inherited disease screening by enabling accurate, high-throughput detection of known pathogenic and likely pathogenic variants across large numbers of samples. Unlike whole genome sequencing, which generates extensive data with many variants of uncertain significance, targeted microarrays focus on clinically verified mutations in genes associated with Mendelian disorders, improving interpretability and turnaround time. In addition, microarrays are used in newborn screening research and reproductive health programs to identify at-risk couples before conception. By combining high analytical accuracy, scalability, and cost-efficiency, microarrays provide a practical and clinically actionable approach to inherited disease screening, particularly when screening focuses on well-characterized, high-impact variants.

The Axiom PangenomePro Array incorporates more than 14,000 expertly curated variants spanning all 73 ACMG-recommended genes (ref), directly aligning genomic interrogation with loci influencing morbidity and mortality. The ACMG 73 and ClinVar module delivers comprehensive coverage of high-impact genes—including *BRCA1/2, CFTR, DMD, RYR1/2, LDLR,* and *APOB*—while leveraging advanced probe design optimized for accurate genotyping in technically challenging regions, including sequences with GC content exceeding 78%. Its content incorporates ClinVar-classified pathogenic and likely pathogenic variants to maintain translational relevance for disease research. Functionally, these loci support targeted investigation across oncology, cardiovascular disease, neurology, and rare genetic disorders (Table 3). Integrated with the array's genome-wide backbone, these evidence-based modules collectively support a balanced architecture that supports both expansive genomic discovery and focused analysis of actionable variation for clinical research.

Pathology	Updated markers on Pangenome Pro array
{Malignant hyperthermia susceptibility 1}	259
Central core disease	259
King-Denborough syndrome	259
Minicore myopathy with external ophthalmoplegia	259
Neuromuscular disease, congenital, with uniform type 1 fiber	259
{Breast-ovarian cancer, familial, 1}	252
{Pancreatic cancer, susceptibility to, 4}	252
Fanconi anemia, complementation group S	252
{Breast cancer, male, susceptibility to}	223
{Breast-ovarian cancer, familial, 2}	223
{Glioblastoma 3}	223
{Medulloblastoma}	223
{Pancreatic cancer 2}	223
{Prostate cancer}	223
Fanconi anemia, complementation group D1	223
Wilms tumor	223
Mismatch repair cancer syndrome	191
5-fluorouracil toxicity	143
Dihydropyrimidine dehydrogenase deficiency	143
Muir-Torre syndrome	131
Colorectal cancer, hereditary nonpolyposis, type 2	78
Acromicric dysplasia	65
Ectopia lentis, familial	65
Geleophysic dysplasia 2	65
Marfan lipodystrophy syndrome	65
Marfan syndrome	65
MASS syndrome	65
Stiff skin syndrome	65
Weill-Marchesani syndrome 2, dominant	65

Table 3. **Updated markers on the PangenomePro array for somatic and inherited diseases.** Braces "{ }" indicate markers that contribute to susceptibility to multifactorial disorders (e.g., diabetes, asthma) or to susceptibility to infection (e.g., malaria), as defined by the OMIM and UCSC genome browser. Only those syndromes with more than 50 markers are shown; overall over 2000 syndromes are represented on the array.

Pharmacogenomic research

Pharmacogenomics links a person's genotype to their drug response in order to optimize efficacy and minimize adverse effects. Microarrays play a critical role in pharmacogenomic (PGx) screening by enabling accurate detection of genetic variants that influence drug absorption, distribution, metabolism, and excretion (ADME). To support this, Thermo Fisher Scientific has probes that target relevant pharmacogenes in the PangenomePro array.

The Axiom PangenomePro Array incorporates updated pharmacogenomic (PGx) content comprising more than 5,500 evidence-based variants, enabling researchers to interrogate genetic determinants of drug metabolism supported by high, moderate, and preliminary levels of scientific evidence. Of these, 4,353 markers are new to the PangenomePro array. The PGx module includes 2,000 markers in high-priority pharmacogenes identified by the ClinPGx database, 300 markers associated with ClinPGx level 1A–2B annotations, and more than 550 reportable alleles referenced in CPIC guidelines. The array provides robust genotype calling across this PGx content, and when paired with the Applied Biosystems™ Axiom™ PangenomePro™ Plus SwiftArray™ Assay, unlocks over 100 additional haplotype-defining markers in clinically significant genes such as *CYP2D6, CYP2A6, CYP2B6, CYP2C19, CYP1A2, CYP2C8,* and *SULT1A1*. Leveraging gene-specific amplification, this advanced assay enables accurate genotyping within highly homologous genomic regions, strengthening its utility for combined GWAS and PGx research applications. The PangenomePro Plus Array further extends this capability by providing pharmacogenomic translation reports that include star allele assignment and phenotype prediction, with star allele calling informed by copy number analysis for genes such as *CYP2D6*, which requires integration of copy number state and SNP genotypes for accurate diplotype and metabolizer classification.

Lifestyle phenotypes and traits	Number of markers
Alcohol dependence and sensitivity	more than 500
Asthma	more than 700
Allergies	more than 1,000
Smoking and addiction	more than 1,000
Vitamin absorption	more than 100
Weight and obesity	more than 900

Table 4. Markers associated with lifestyle and trait-associated variants found on the PangenomePro array.

Polygenic risk scores

Microarrays are widely used to generate the genotype data needed for polygenic risk score (PRS) determination by enabling large-scale research of genetic variants associated with complex diseases. Over the past two decades, Applied Biosystems microarrays have supported more than 900 peer-reviewed GWAS publications (including approximately 750 in humans), contributing to the discovery of loci that collectively inform disease susceptibility models. Importantly, PRS accuracy can improve when population ancestry is considered; for example, polymorphisms associated with childhood myopia have been identified in Han Chinese children, and variants linked to Gilbert's syndrome have been characterized in Taiwanese populations [5,6]. By enabling high-throughput, population-specific genotyping, microarrays continue to play a central role in translating genetic variation into clinically meaningful risk prediction.

The Axiom PangenomePro Array extends well beyond traditional GWAS applications, supporting predictive genomics and translational research through a diverse and strategically curated content portfolio. It includes ancestry-informative markers optimized across multiple global populations that enable risk stratification based on population diversity (Table 5). The array also contains extensive rare blood group coverage to support immunohematology and transfusion research, alongside lifestyle- and trait-associated variants linked to phenotypes such as diet, weight, metabolism, and taste (Table 4). In addition, it incorporates marker sets aligned with large-scale biobank findings, including UK Biobank data, to enable polygenic risk score (PRS) development and risk prediction for common complex diseases. Collectively, these capabilities position the array as a comprehensive and versatile platform for disease research, predictive genomics initiatives, and clinical trial research applications.

Summary

Following the Human Genome Project, advances in sequencing expanded precision medicine but introduced high costs and interpretive complexity, making targeted genotyping microarrays an efficient alternative for many applications. The Axiom PangenomePro Array leverages imputation-aware design that facilitates broad, multi-ancestry genome-wide coverage while focusing on clinically actionable content. It integrates more than 940,000 GWAS markers, over 14,000 curated variants across all 73 ACMG-recommended genes, and more than 5,500 pharmacogenomic variants—including enhanced coverage for complex loci such as *CYP2D6*. Supporting ancestry analysis, inherited disease research, pharmacogenomics, rare blood typing, lifestyle traits, and polygenic risk score development, the array provides a scalable and translationally relevant platform for population genomics and clinical research.

Categories and subcategories	Markers on PangenomePro array	Updated markers on PangenomePro array
Cancer risk variants	more than 13,000	3,126
Myeloma	more than 70	0
Lung cancer	more than 400	7
Breast cancer	more than 1,800	257
Ovarian cancer	more than 1,500	479
Gastric cancer	more than 900	46
Leukemia	more than 3,000	26
Lymphoma	more than 700	4
Colorectal cancer	more than 2,200	276
Mental behavioral, neurological, and neurodevelopmental risk variants	more than 4,300	72
Alzheimer's disease	more than 300	0
Parkinson's disease	more than 300	1
Schizophrenia	more than 700	0
Autism	more than 200	29
Inherited eye disease risk variants	more than 3,700	17
Macular degeneration	more than 500	15
Glaucoma	more than 150	4
Retinal dystrophy	more than 100	10
Retinitis pigmentosa	more than 400	70
Optic atrophy	more than 10	5
Autoimmune and inflammatory disease risk variants	more than 1,150	7
Celiac disease	more than 90	0
Crohn's disease	more than 400	0
Graves' disease	more than 35	0
Fanconi anemia	more than 60	485
Cystic fibrosis	more than 3	0
Thalassemia	more than 3	62
Familial hypercholesterolemia	more than 20	35
Mitochondrial diseases	more than 10	21
Cardiovascular disease risk variants	more than 8,500	288
Respiration disorder risk variants	more than 500	5

Table 5. Markers on the PangenomePro array that target risk and susceptibility. The second column indicates the number of all markers on the array; the third column indicates the number of updated markers on the array.

References

1. 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
2. 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
3. Resutik, P. et al. (2023) Comparative evaluation of the MAPlex, Precision ID Ancestry Panel, and VISAGE Basic Tool for biogeographical ancestry inference. Forensic Science International: Genetics 64, 102850 doi: 10.1016/j.fsigen.2023.102850
4. Tillmar A, et al. (2021). The FORCE Panel: An All-in-One SNP Marker Set for Confirming Investigative Genetic Genealogy Leads and for General Forensics. Applications. Genes (Basel) 12(12):1968. doi: 10.3390/genes12121968.
5. Chen LJ et al. (2021) Association of polymorphisms in ZFHX1B, KCNQ5 and GJD2 with myopia progression and polygenic risk prediction in children. Br J Ophthalmol 105:1751–1757. doi.org/10.1136/bjophthalmol-2020-318708
6. 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

Learn more at thermofisher.com/pangenomepro