Axiom Asia Precision Medicine Research Array

Driving deeper insights into genetics, ancestry, and susceptibility to complex diseases in subpopulations in Asia to advance population genomics studies

Precision medicine research involves the analysis of a wide range of evidence-based markers implicated in disease, pharmacogenomics, and immunology. In precision medicine research studies, it is important to capture the variation in individual populations to detect disease and trait associations. Imputation-aware, population-optimized Applied Biosystems™ Axiom™ genotyping arrays maximize the discovery of novel associations in target populations through the inclusion of population-private genetic variants. These imputation-aware Axiom genotyping arrays are now used as custom arrays in several large biobank studies and by direct-to-consumer companies because of the high imputation performance and coverage of populations.

The Applied Biosystems™ Axiom™ Asia Precision Medicine Research Array (Asia PMRA) leverages a proprietary imputation algorithm to maximize genetic coverage in target multiethnic populations throughout Asia. The array was designed to include an efficient set of markers uniquely suited for imputation in each of the populations in China, Vietnam, Japan, and populations in the Indian subcontinent, enabling superior imputation accuracy. The modularly designed array has over 750,000 markers in total, including a genome-wide association studies (GWAS) imputation module with more than 540,000 markers.

This population-specific approach for variant selection was extended to loss-of-function (LOF) mutations and potentially pathogenic variants from ClinVar. Variants obtained from public databases such as ClinVar, the National Human Genome Research Institute (NHGRI) GWAS catalog, and Pharmacogenomics Knowledgebase (PharmGKB), as well as additional recent publications [1-4], were selected based on their alternate allele frequencies in East Asian (EAS) and South Asian (SAS) populations, enabling deep insight in genomic variation relevant in these target ethnic groups.

Leveraging the customizability of the Applied Biosystems™ Axiom™ Genotyping Solution, the Axiom Asia PMRA content can also serve as a valuable scaffold to power prospective epidemiological studies in any of the targeted populations by supplementing its content with further evidence-based, population-private, clinically actionable markers.

Highlights

- Available in convenient 24- and 96-array formats with Applied Biosystems™ Axiom™ 2.0 reagents and GeneTitan™ Multi-Channel (MC) Instrument consumables
- Powered by the Axiom Genotyping Solution used by leading biobanks worldwide, Axiom Asia PMRA offers a genome-wide imputation grid from phase III of the 1000 Genomes Project and NHGRI–European Bioinformatics Institute (EBI) GWAS catalog published as of May 2017, providing the most up-to-date content, broadest coverage, and highest accuracy for disease association studies across populations [5-7]
- Includes optimally selected, potentially clinically significant pathogenic variants from ClinVar (February 2017), including those implicated for actionable genetic risk in EAS and SAS ancestral populations, through a proprietary imputation algorithm
- Includes over 2,500 pharmacogenomics variants covering absorption, distribution, metabolism, and excretion (ADME) markers, of which 1,100 are evidence-based markers identified by PharmGKB (Table 1 and Figure 1)

Table 1. Axiom Asia PMRA key marker groups.

Variant category	Number of markers*
Genome-wide imputation grid**—focus on EAS and SAS populations	>540,000
NHGRI-EBI GWAS catalog	>23,400
Markers of clinical relevance	
ClinVar	>43,000
ACMG	>9,200
Pharmacogenomics and ADME	>2,600
Additional high-value markers (subset of ClinVar: *APOE, BRCA1/2, DMD, CFTR*)	> 2,000
Immune-related markers	
Human leukocyte antigen (HLA)	>9,000
Killer immunoglobulin-like receptor (KIR)	>1,400
Autoimmune and inflammatory	>250
Functional markers	
LOF	>43,000
Very rare nonsynonymous variants (minor allele frequency (MAF) > 0.01%)	>35,000
Expression quantitative trait loci (eQTL)	>15,000
Lung function phenotypes	>7,600
Disease-related markers	
Alzheimer's disease	>900
Cardio-metabolic	>360
Neurological disorders	~16,000
Diabetes	>500
Common variants in cancer	>300
Rare missense variants in cancer predisposition genes	>2,600
Rare variants in cardiac predisposition genes	>830
Rare polymorphic variants from Exome Aggregation Consortium (ExAC) data	>4,700
Miscellaneous	
Fingerprinting and sample tracking	>300
Y chromosome	~400
Mitochondrial	~500
Gender determination	~1,000
Chromosome X SNPs and indels	~25,000
Custom variants**	
Add 50,000 custom markers, or fully customize as required	
Total markers	**>750,000**

* Content in categories may overlap.
** 50,000 markers in the GWAS grid can be replaced with custom content without impacting coverage or accuracy.

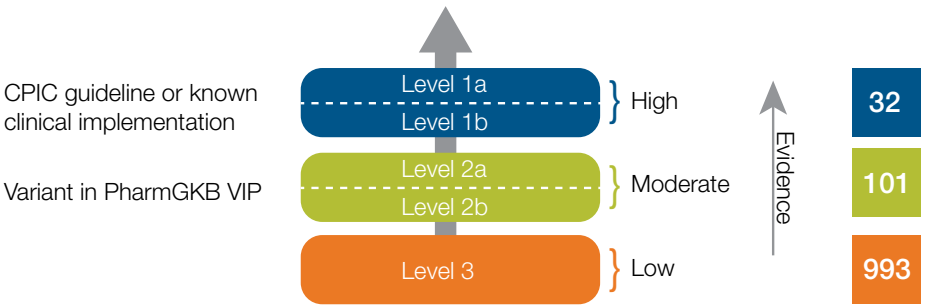

Figure 1. Pharmacogenomic markers classified into evidence-based categories based on variant description in PharmGKB. Over 1,100 of the 2,500 markers on the array cover 600 genes and are classified as high-, medium-, or low-evidence marker. Refer to **www.pharmgkb.org/page/clinAnnLevels** for a full description of the annotation levels of evidence. CPIC: Clinical Pharmacogenetics Implementation Consortium; VIP: Very Important Pharmacogene.

Variants covering diseases and human health conditions

The evidence-based clinical research content on this array enables better understanding of susceptibility to complex diseases. The array includes several thousand evidence-based risk variants that have been implicated in human diseases and conditions in EAS and SAS ancestral populations. The array includes pathogenic variants selected from ClinVar, including apolipoprotein E (ApoE) markers associated with Alzheimer's disease. Table 2 shows the number of markers that fall into the various categories per the Online Medelian Inheritance in Man (OMIM) and ClinVar databases.

Table 2. Markers classified into disease research categories and important subcategories according to OMIM and ClinVar databases.

Categories and subcategories	Number of markers
Cancer-associated variants	**>8,000**
Myeloma	>50
Lung cancer	>200
Breast cancer	>300
Ovarian cancer	>250
Gastric cancer	>40
Leukemia	>3,000
Colorectal cancer	>400
Mental, behavioral, and neurodevelopmental risk variants	**>19,000**
Alzheimer's disease	>550
Parkinson's disease	>400
Schizophrenia	>300
Autism	>140
Autoimmune and inflammatory disease–associated variants	**>200**
Celiac disease	88
Crohn's disease	34
Graves' disease	28
Clinical variants	**>44,000**
Pathogenic	>33,000
Likely pathogenic	>10,000
LOF variants	**>47,000**
Autosomal recessive	>150
Autosomal dominant	>1,500
Mitochondrial	>60
Pathogenic or likely pathogenic	>2,400
Congenital conditions	>230
Cardiovascular disease–associated variants	**>2,000**
Health status–associated variants	**>2,000**
Smoking and addiction	>40
Alcohol dependence and sensitivity	>150
Asthma	>200
Caffeine consumption	>45
Allergy variants (egg, milk, food, lactose)	121
Skin, hair, or eye pigmentation	>1,000
Respiratory disease–associated variants	**>700**
Glaucoma risk variants	**>130**
Diabetes risk variants	**>500**
Obesity risk variants	**>950**

Genome-wide imputation grid—This grid includes markers to maximize coverage in all EAS and SAS populations, especially in the 1–5% MAF range, enabling cross-platform and cross-cohort metadata analysis. The highest-ranked 540,000 markers offer imputation accuracy of 94% for common variants with MAF >5% in populations with EAS ancestry. Likewise, the array offers imputation accuracy of 95% for common variants with MAF >5% in populations with SAS ancestry. The imputation coverage and accuracy are provided in Tables 3 and 4, respectively.

NHGRI-GWAS catalog variants—This includes content covering the complete NHGRI catalog of published GWAS as of May 2017.

Markers of clinical relevance—These variants cover pathogenic or likely pathogenic associations from ClinVar (accessed February 2017) archives. The list also includes a set of markers from the list of genes published by the American College of Medical Genetics (ACMG) with intersection in the ClinVar archives. Additional markers known to be of high clinical importance, such as those in *BRCA1, BRCA2, CFTR, DMD*, and *APOE* genes (with >78% GC content in flanking sequences), are included in this module.

Pharmacogenomic variants—These variants include ADME markers from the list of variants in the Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines. The array includes over 2,600 markers covering 600 genes that are associated with pharmacogenomic information.

Immune-related markers—This list includes markers from *HLA*, *KIR* genes, and variants associated with specific autoimmune and inflammatory disorders, including ulcerative colitis, Crohn's disease, Graves' disease, Hashimoto's thyroiditis, and celiac disease. Analysis of HLA markers from the extended major histocompatibility complex (MHC) region is accomplished with Applied Biosystems™ Axiom™ HLA Analysis Software for improved imputation of HLA alleles in multiethnic populations.

Functional variants—This module covers (i) LOF variants, including rare and likely deleterious alleles, to detect genetic changes predicted to completely disrupt the function of protein-coding genes; (ii) lung function variants having an established or putative association with lung function, lung diseases (asthma, cystic fibrosis, chronic obstructive pulmonary disease, idiopathic pulmonary fibrosis, and lung cancer), and smoking behavior; (iii) rare and nonsynonymous variants with very low MAF (>0.01%); and (iv) eQTLs with MAF >0.01% to support mapping functional noncoding variations to identify associations with gene transcription variability and differential gene expression.

Table 3. Number of imputed markers with r^2 >0.8 and MAF >1%.

Population	MAF >1%
East Asian ancestry	8.99 million
South Asian ancestry	10.1 million

Table 4. Imputation accuracy (mean r^2) calculated across autosomal SNPs from highest-ranked 540,000 markers.

		Accuracy (mean r^2 ≥0.8)	
		MAF >1%	MAF >5%
East Asian ancestry	All populations	0.90	0.94
	Chinese Dai	0.89	0.95
	Chinese Han	0.90	0.94
	Chinese Southern Han	0.90	0.94
	Japanese	0.90	0.94
	Vietnamese Kinh in Ho	0.88	0.94
South Asian ancestry from Indian subcontinent	All populations	0.90	0.95
	Bengali	0.88	0.94
	Gujrati	0.92	0.95
	Telugu	0.90	0.94
	Punjabi	0.89	0.95
	Sri Lankan	0.90	0.94

Disease-related markers—The disease-associated markers are in various disease-specific modules. These include:

- **Alzheimer's disease**—Variants associated with Alzheimer's disease were selected from a meta-analysis of Alzheimer's disease association studies and a set of mitochondrial markers suspected to be associated with the disease from an Alzheimer's disease research group.
- **Neurological conditions**—Markers include a large number of rare mutations possibly associated with a variety of neurological diseases, recently identified association hits, and a number of markers chosen from exome-sequencing studies. Cohorts of diseases considered include Alzheimer's disease, frontotemporal dementia, progressive supranuclear palsy, amyotrophic lateral sclerosis, and Parkinson's disease.
- **Diabetes**—Variants associated with diabetes, identified via GWAS according to the NHGRI-EBI GWAS catalog (accessed May 2017) and from recent publications, that have been shown to be associated with diabetes in EAS and SAS populations.
- **Rare cardiac variants**—Markers for cardiac disease predisposition including rare variants in the genes *MYBPC3* and *MYH7*, markers chosen from the ARVD/C locus-specific database to look at variants in the genes *DSC2, DSG2, DSP, JUP,* and *PKP2*, and a set of markers chosen from human disease mutation databases for genes related to cardiac disease and hemochromatosis.
- **Cardio-metabolic variants**—Known variants associated with various cardiometabolic traits (i.e., coronary disease, lipids, anthropometry, glycemic markers, and blood pressure).
- **Rare cancer variants**—Markers for rare missense variants in proven cancer predisposition genes, selected from sources including locus-specific and human disease mutation databases.
- **Common cancer variants**—Variants from the list of published common variants associated with cancer phenotypes, identified via GWAS according to the NHGRI-EBI GWAS catalog, as well as some recently published and unpublished cancer-associated SNPs as of May 2017.

High–imputation coverage accuracy

The Axiom Asia PMRA GWAS markers are common variants that are intelligently selected via a proprietary imputation-based marker selection strategy for genome-wide coverage in EAS and SAS populations, as defined by the 1000 Genomes Project. Table 3 provides the number of imputed markers available on Axiom Asia PMRA for the two ancestral populations; Table 4 provides the imputation accuracy for GWAS markers with MAF >1% and >5%.

Precision medicine initiatives and continuity of studies

The manufacturing technology of Axiom genotyping arrays produces 100% fidelity. All markers are present on every manufacturing batch for as long as you need. Precision medicine research studies involving large cohorts are typically conducted over several years and therefore require a platform that can offer multiyear availability of 100% of the specific array content. Axiom genotyping arrays are able to deliver all of these product requirements. In contrast, bead-based technologies and the corresponding catalog products require synthesis of a new bead pool with each new batch, and will potentially have batch-to-batch variability and SNP dropouts.

Array processing

The Axiom Genotyping Solution provides superior lot-to-lot reproducibility and utilizes a powerful 24- and 96-array format. The GeneTitan MC Instrument automates array processing from target hybridization to data generation and can process up to 8-array plates per week in a standard workflow. Applied Biosystems™ Axiom™ Analysis Suite software includes allele-calling algorithms and user-friendly visualization tools. The assay and workflow are available as a manual or fully automated process and offer reduced hands-on time compared to other genotyping platforms, allowing the processing of more samples for a more comprehensive study.

applied biosystems

Specifications

Genotyping performance has been evaluated on 1,176 samples from the International HapMap Project, using stringent quality-control metrics that cover average sample call rate, sample concordance, and reproducibility (Table 5). Table 6 provides sample call rate and concordance for each of the 10 individual populations on the array.

Table 5. Performance of Axiom PMRA across 1,176 samples.

Metric	Specification (%)	Performance (%)
Sample pass rate	≥95	98.64
Average call rate	≥95	99.81
Reproducibility	≥99.80	99.92
Average concordance with 1000 Genomes Project	≥99.50%	99.70%

Analysis workflow for Axiom Asia PMRA

The analysis workflow as described in the Axiom Genotyping Solution Data Analysis Guide (P/N 702961) is an advanced analysis technique that provides the greatest flexibility in finding the most informative content for each study.

Table 6. Call rate across each of the individual populations covered in Axiom Asia PMRA.

	1000 Genomes Project populations	Call rate (%)	Overall concordance (%)
East Asian ancestry	Chinese Dai	99.84	99.70
	Chinese Han	99.84	99.70
	Chinese Southern Han	99.78	99.70
	Japanese	99.93	99.70
	Vietnamese Kinh in Ho	99.92	99.80
South Asian ancestry	Bengali	99.88	99.70
	Gujrati	99.72	99.80
	Telugu	99.64	99.60
	Punjabi	99.91	99.70
	Tamil	99.80	99.70

Ordering information

Product	Details	Cat. No.
Axiom Asia Precision Medicine Research Array Kit (96-array format)	Contains one 96-array plate and all GeneTitan Multi-Channel Instrument consumables and Axiom reagents (except ispropanol) for processing one 96-array plate	905423
Axiom Asia Precision Medicine Research Array Kit (24-array format)	Contains four 24-array plates and all GeneTitan Multi-Channel Instrument consumables and Axiom reagents (except ispropanol) for processing four 24-array plates	905421

References

1. Mittal A et al. (2017) Design of custom genotyping array optimized for East and South Asian populations using improved SNP selection algorithms. Poster 1510. Presented at the 67th Annual Meeting of *The American Society of Human Genetics*, October 18, 2017, in Orlando, Florida.
2. Giri AK et al. (2014) Pharmacogenetic landscape of clopidogrel in North Indians suggest distinct interpopulation differences in allele frequencies. *Pharmacogenomics* 15(5):643-653.
3. Chauhan G et al. (2010) Impact of common variants of *PPARG*, *KCNJ11*, *TCF7L2*, *SLC30A8*, *HHEX*, *CDKN2A*, *IGF2BP2*, and *CDKAL1* on the risk of type 2 diabetes in 5,164 Indians. *Diabetes* 59:2068-2074.
4. Giri AK et al. (2016) Genome-wide association study of uric acid in Indian population and interaction of identified variants with type 2 diabetes. *Sci Rep* 6:21440.
5. Ma J et al. (2016) Association analysis of the *cubilin (CUBN)* and *megalin (LRP2)* genes with ESRD in African Americans. *Clin J Am Soc Nephrol* 11(6):1034-1043.
6. Davies G et al. (2016) Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112 151). *Mol Psychiatry* 21:758-767.
7. Day FR et al. (2016) Physical and neurobehavioral determinants of reproductive onset and success. *Nat Genet* 48:617-623.

Find out more at **thermofisher.com/microarrays**

ThermoFisher SCIENTIFIC