illumina®

Infinium™ H3Africa Consortium Array v2

A powerful array for genetic studies focused on African populations.

Overview

The Infinium H3Africa Consortium Array v2 (Figure 1, Table 1) is a powerful array for research studies that enables identification of genetic associations with common and rare traits among African populations (Table 2). Specifically designed by the H3Africa consortium, the array harnesses content from Illumina Omni2.5-8 and Omni5-4 BeadChips, in addition to custom content selected by the consortium based on whole-genome sequencing (WGS) data. The custom content was selected by including specific single nucleotide polymorphisms (SNPs) requested by H3Africa consortium projects, SNPs within the MHC region, X-chromosome and mitochondrion SNPs, and SNPs of clinical or pharmacogenomic interest. Remaining custom SNPs were selected to improve coverage, imputation accuracy, and enrichment in novel but common variants in African populations based on sequencing data.

Consortium-selected content

The H3Africa consortium selected approximately 10,000 variants relevant to specific diseases of interest including variants known to be associated with kidney disease, diabetes, sickle cell disease, cardiometabolic diseases and susceptibility to infectious diseases. Additional variants from PharmGKB, genome-wide association study (GWAS) catalog, ClinVar and the COSMIC database were also identified and selected by the consortium. For PharmGKB and the GWAS catalog, 4000 and 24,000 variants, respectively, that occur with a minor allele frequency (MAF) ≥ 0.01 in at least one of the African populations were selected for inclusion. For ClinVar the consortium restricted their selection to 27,000 SNPs with MAF ≥ 0.05 in at least one of the African populations. For variants in the COSMIC database the consortium retained 20,000 variants that were substitutions, characterized as pathogenic in the database and showed MAF ≥ 0.05 in at least one of the African populations. A part of this list was already present in the fixed content, and the remaining 60,000 SNPs were added as a component of the custom content. To optimize ancestry inference, the consortium identified a panel consisting of ~2000 variants of carefully chosen mitochondrial, Y-chromosome, and African-centric ancestry informative markers (AIMs) (Table 2, Table 3).

Table 1: Product information

Feature	Description
Species	Human
Total number of markers	2,271,503
Number of samples per BeadChip	8
DNA input requirement	200 ng
Assay chemistry	Infinium LCG
Instrument support	iScan™ System
Sample throughput	~1728 samples/week
Scan time per sample	35 minutes

Figure 1: Infinium H3Africa Consortium Array v2—The 8-sample Infinium H3Africa Consortium Array supports rapid cost-effective studies.

Table 2: Marker information

Marker categories[a]			No. of markers
Exonic markers[b]			88,785
Intronic markers[b]			1,074,881
Nonsense markers[c]			313
Missense markers[c]			23,548
Synonymous markers[c]			21,654
MItochondrial markers[c]			234
Indels[d]			318
Sex chromosomes[c]	X	Y	
	36,347	2528	

a. Number of markers are calculated from the consortium manifest.
b. RefSeq: NCBI Reference Sequence Database. Accessed August 30, 2020.[18]
c. Compared against the UCSC Genome Browser. Accessed August 30, 2020.[2]
d. NCBI Genome Reference Consortium, Version GRCh37. Accessed August 30, 2020.[19]
Abbreviations: indel: insertion/deletion, PAR: pseudoautosomal region.

Infinium H3Africa Array reference samples

WGS data obtained from ~3480 individuals from 17 African countries were used in the array design. The sequencing coverage for these samples varies from 4× to 30×. Included in this cohort, the H3Africa consortium contributed ~350 samples for sequencing at the Baylor College of Medicine to generate high coverage data and fill some of the gaps in missing populations or countries. The TrypanoGen project from H3Africa had additional medium coverage sequence data for 118 samples which were contributed for the design.

Table 3: High-value content

Content	No. of markers[a]	Research application/note
ACMG[1] 59 2016 gene coverage	5201	Variants with known clinical significance identified from clinical WGS and WES samples
ADME[2] CPIC genes	3258	Drug metabolism and excretion (includes regulatory regions)
ADME[2] core and extended + CPIC genes +/- 10 kb	24,168	
AIMs	2563	Ancestry-informative markers
ClinVar[3] variants	8590	Relationships among variation, phenotypes, and human health
ClinVar[3] pathogenic	85	
ClinVar[3] likely pathogenic	24	
ClinVar[3] benign	4679	
ClinVar[3] likely benign	4097	
COSMIC[4] genes	81,841	Somatic mutations in cancer
eQTLs	8219	Genomic loci regulating mRNA expression levels
gnomAD exome	65,982	WES and WGS results from unrelated individuals from various studies
HLA genes	1164	Disease defense, transplant rejection, and autoimmune disorders
Extended MHC[5b]	24,411	Disease defense, transplant rejection, and autoimmune disorders
NHGRI[6]-EBI GWAS catalog	44,344	Markers from published GWAS
PharmGKB[7] phenotype annotation	1846	Human genetic variation associated with drug responses, variants that affect a phenotype, with or without drug information
PharmGKB[8] drug annotation	1729	Variants that affect drug dose, response, metabolism, etc
PharmGKB[7] functional analysis annotation	149	*In vitro* and functional analysis-type associations

a. The number of markers for each category may be subject to change
b. Extended MHC is a 8 Mb region

Abbreviations: ACMG: American College of Medical Genetics; ADME: absorption, distribution, metabolism, and excretion; AIM: ancestry-informative marker; COSMIC: catalog of somatic mutations in cancer; eQTL: expression quantitative trait loci; gnomAD: Genome Aggregation Database; HLA: human leukocyte antigen; MHC: major histocompatibility complex; NHGRI: national human genome research institute; PharmGKB: Pharmacogenomics Knowledgebase

Ordering Information

Infinium H3Africa Kit	Catalog no.
48 samples	15056943
96 samples	15056944
384 samples	15056945

Learn more

To learn more about the Infinium H3Africa Consortium Array, contact your genotyping specialist or a local sales representative:

North America: 800.809.4566

Europe, Middle East, Africa: +44.1799.534000

Other regions: www.illumina.com/company/contact-us.html

To learn more about the H3Africa Consortium, visit h3africa.org/

References

1. ACMG Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing. Accessed September 20, 2020.
2. PharmaADME Gene List. Accessed September 20, 2020.
3. ClinVar Database. Accessed September 20, 2020.
4. COSMIC. Accessed September 20, 2020.
5. de Bakker PIW, McVean G, Sabeti PC, et al. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat Genet*. 2006;38:1166–1172.
6. National Human Genome Research Institute. Accessed September 20, 2020.
7. PharmGKB. Accessed September 20, 2020.
8. PharmGKB, Clinical Annotation Levels of Evidence. Accessed September 20, 2020.

Illumina • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com