illumina®

Infinium® Expanded Multi-Ethnic Genotyping Array (MEGA^{EX})

A consortium-built array with increased power for understanding complex disease in diverse human populations.

Introduction

The Expanded Multi-Ethnic Genotyping Array (MEGA^{EX}) leverages content from Phase 3 of the 1000 Genomes Project (1kGP)[1], Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA), Population Architecture using Genomics and Epidemiology (PAGE), T2D-Genes Consortium, OMIM, ClinVar, ACGM, carrier screening panels, and other resources to create a true multipurpose, multiethnic array. There has been significant investment in detecting common genetic variants associated with complex disease in European populations; however, there are advantages to studies focused on diverse human populations. These studies are essential for future enablement of precision medicine, adding value to biobank repositories, empowering the next generation of genetic studies, and understanding and measuring fine-scale population structure and its association with biomedical traits. With > 2 million expertly selected markers, MEGA^{EX} enables identification of genetic associations with common and rare traits, providing insight across diverse populations to epidemiologists, health care researchers, population geneticists, and genomic researchers (Table 1 and 2).

Maximized Imputation Accuracy

Consortium partners developed MEGA^{EX} using tagging strategies with the power to perform more effective association studies in diverse populations (Table 3). The novel algorithm selects population-specific and transethnic tag SNPs that maximize imputation accuracy, rather than pairwise coverage. Optimization focused on genotype imputation as it has become a standard practice in the interpretation of genotyping data and allows for more accurate statistical inference of genotypes not directly genotyped.

Table 1: MEGA^{EX} Product Information

Feature	Description	
Total No. of Markers	2,036,060	
Capacity for Custom Markers	300,000	
No. Samples per BeadChip	8	
DNA Input	200 ng	
Assay Chemistry	Infinium® LCG	
Instrument Support	iScan® or HiScan® System	
Sample Throughput[a]	~ 1067 samples/week	
Scan Time per Sample	iScan System 11.3 min	HiScan System 6.5 min

a. Estimated sample throughput based on use of 1 HiScan System, 1 AutoLoader 2.x, 1 Tecan robot, and a 5-day work week.

Table 2: MEGA^{EX} Marker Information

Marker Category	No. of Markers
Chromosome X	60,547
Chromosome Y	7216
PAR / Homologous	5342
Mitochondrial	890
Indels	28,316
Exonic	466,011
Missense SNPs	22,438
Nonsense SNPs	21,892
Synonymous SNPs	26,657
UTR	31,094
Intronic	623,950

Table 3: Imputation Accuracy[a] for 6 Populations from 1kGP at Different MAF Thresholds

	Minor Allele Frequency (MAF) Threshold		
Population[b]	0.5–1%	1–5%	≥ 5%
ACB + ASW	80.8%	91.0%	94.8%
AFR	79.7%	89.8%	94.2%
AMR	83.3%	89.9%	96.3%
EAS	60.5%	83.3%	95.0%
EUR	74.1%	88.5%	96.6%
SAS	67.6%	85.6%	95.8%

a. Imputation accuracy analysis performed by Bustamante Lab of Stanford University.
b. ACB: African Caribbeans in Barbados; ASW: Americans of African Ancestry in SW USA; AFR: African; AMR: Ad-mixed American; EAS: East Asian; EUR: European; SAS: South Asian.[1]

Expert-Selected Content

MEGA^{EX} combines expertly selected markers and content from the most popular Illumina commercial arrays and backwards compatibility with the most current genomic information. Researchers can detect both common and rare variants across the most commonly studied 6 major population groups and impute variants in a vast number of subpopulations (Table 3).

MEGA^{EX} contains the following content (Table 4):

- Infinium HumanCore BeadChip content with highly informative genome-wide tag SNPs
- African Diaspora Consortium Power Chip content identified through sequencing of 692 individuals by CAAPA
- High-information, genome-wide coverage for diverse populations selected by PAGE using a new cross-population tagging strategy from 1kGP Phase 3
- Total exonic content of > 400,000 markers
  - Infinium HumanExome BeadChip content with exonic variants selected from sequencing > 12,000 individuals
  - Multiethnic exome content designed by PAGE from sequencing > 36,000 individuals in diverse ethnic groups, emphasizing loss-of-function and splice variants
- Over 17,000 variants chosen to be relevant to clinical and pharmacogenetic studies
- Additional 23,000 hand-curated variants picked for functional, immunological, oncological, ancestry, forensic, and common and rare disease research applications
- Capacity to add up to 300,000 custom beadtypes to the array

Learn More

For more information, contact consortiamanager@illumina.com or your local genotyping specialist:
North America: 800.809.4566
Europe, Middle East, Africa: +44.1799.534000
Other regions: www.illumina.com/company/contact-us.html

References

1. 1000 Genomes (www.1000genomes.org). Accessed July 21, 2015.
2. COSMIC: Catalogue of somatic mutations in cancer (cancer.sanger.ac.uk/cosmic). Accessed July 21, 2015.
3. Gene Ontology Consortium (geneontology.org). Accessed July 21, 2015.

Table 4: MEGA^{EX} High-Value Content

Content	No. of Markers	Research Application / Note
ADME Core and Extended Genes[a]	20,563	Drug metabolism and excretion
ADME Core and Extended Genes[a] +/- 10 kb	25,022	Drug metabolism and excretion (+ regulatory regions)
APOE	144	Cardiovascular disease, Alzheimer's disease, immunoregulation, and cognition
Blood Phenotype Genes	3309	Blood phenotypes
COSMIC Genes[a]	1,018,089	Somatic mutations in cancer
GO CVS Genes[a]	273,024	Cardiovascular conditions
Database of Genomic Variants	1,596,466	Genomic structural variation
eQTLs	6689	Genomic loci regulating mRNA expression levels
Fingerprint SNPs	731	Human identification
HLA[a]	2050	Disease defense, transplant rejection, and autoimmune disorders
KIR[a]	473	Autoimmune disorders and disease defense
MHC[a]	23,946	Disease defense, transplant rejection, and autoimmune disorders
Neanderthal SNPs	2226	Neanderthal ancestry and human population migration
NHGRI GWAS Catalog[a]	11,631	Markers from published genome-wide association studies
RefSeq 3' UTRs	45,638	3' untranslated regions of known genes
RefSeq 5' UTRs	31,390	5' untranslated regions of known genes
RefSeq All UTRs	74,771	All untranslated regions of known genes
RefSeq	1,127,714	All known genes
RefSeq +/-10 kb	1,285,332	All known genes +/- 10 kb to include regulatory regions
RefSeq Promoters	49,597	2 kb 5' of all known genes to include promoter regions
RefSeq Splice Regions	10,855	Variants at splice sites in all known genes

a. ADME: absorption, distribution, metabolism, and excretion; COSMIC: catalog of somatic mutations in cancer[2]; GO CVS: Gene Ontology annotation of the cardiovascular system[3]; eQTL: expression quantitative trait loci; HLA: human leukocyte antigen; KIR: killer cell Ig-like receptor; MHC: major histocompatibility complex; NHGRI: national human genome research institute.

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

For Research Use Only. Not for use in diagnostic procedures.

illumina®