Infinium™ Global Screening Array-24 v2.0

A powerful, high-quality, cost-effective array for population-scale genetic studies.

Highlights

- **Global content**
  Includes a multiethnic genome-wide backbone, expertly designed clinical research variants, quality control (QC) markers, and the flexibility to add content
- **Broad clinical research applications**
  Enables genotyping for a broad range of applications, including complex disease studies, pharmacogenomics research, lifestyle and wellness characterization, and more
- **High-throughput workflow**
  Supports high-throughput processing of thousands of samples per week for population-scale studies
- **Robust, high-quality assay**
  Uses trusted Infinium technology to deliver call rates > 99% and reproducibility >99.9

Introduction

The Infinium Global Screening Array-24 v2.0 (GSA) BeadChip is an advanced genotyping array that provides a cost-effective solution for population-scale genetic studies, variant screening, and precision medicine research. Using the proven iScan™ System, integrated analysis software, and Infinium high-throughput screening (HTS) Assay, this high-density, 24-sample BeadChip (Figure 1) provides optimized content for a broad range of applications, delivered with the same high-quality, reproducible data that Illumina genotyping arrays have provided for over a decade (Table 1). The GSA Kit includes convenient packaging containing BeadChips and reagents for amplifying, fragmenting, hybridizing, labeling, and detecting genetic variants using the high-throughput, streamlined Infinium workflow.

Table 1: Product information

Feature	Description	
Species	Human	
Total number of markers	665,608	
Capacity for custom bead types	50,000	
Number of samples per BeadChip	24	
DNA input requirement	200 ng	
Assay chemistry	Infinium HTS	
Instrument support	iScan or HiScan™ System	
Sample throughput[a]	~ 2304 samples/week	
Scan time per sample	iScan System 2.5 min	HiScan System 2.0 min

a. Estimate assumes 1 iScan System, 1 AutoLoader, 2 Tecan robots, and a 5-day work week.

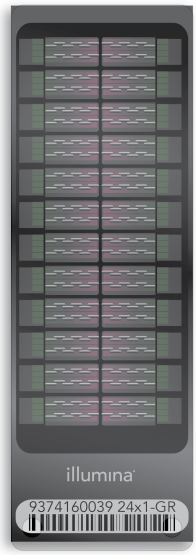

Figure 1: The Infinium Global Screening Array-24 v2.0 BeadChip—The Infinium GSA-24 v2.0 BeadChip is built on the trusted 24-sample Infinium HTS platform.

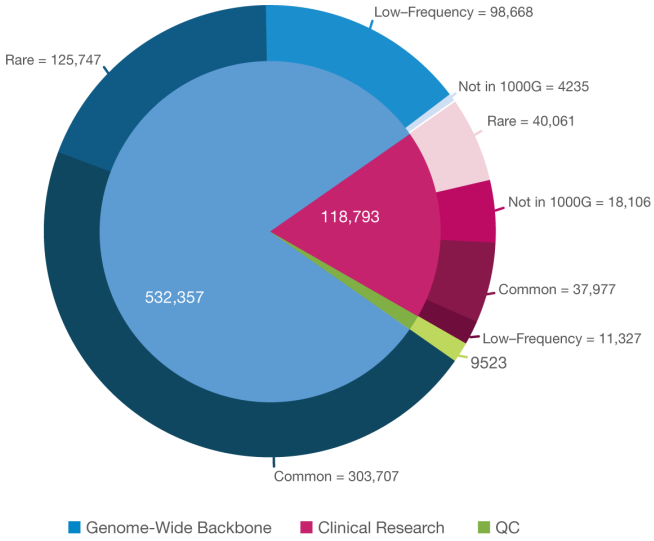

Figure 2: Summary of content–Genome-wide content enables a broad range of clinical research and genetic variant screening applications. Plotted in the inner pie is the proportion of the array that was selected for genome-wide coverage (blue), clinical research (pink), and quality control (green). The outer ring summarizes the weighted reference global allele frequency for unique variants present in the 1000 Genomes Project (1000G).[1] Variants not in 1000G are labeled.

Table 2: High-value content

Content	No. of markers	Research application/note	Content	No. of markers	Research application/note
ACMG[2] 2016 gene list	17,075	List of genes recommended for reporting of secondary findings from genome and exome sequencing by the American College of Medical Genetics	gnomAD exome	67,371	The Genome Aggregation Database (gnomAD) consists of exome and whole-genome sequences from unrelated individuals sequenced as part of various studies
ADME core and extended + CPIC genes[3]	15,023	Drug metabolism and excretion	HLA genes[4]	466	Disease defense, transplant rejection, and autoimmune disorders
ADME core and extended + CPIC genes +/- 10 kb	17,750	Drug metabolism and excretion (plus regulatory regions)	Extended MHC[12a]	8546	Disease defense, transplant rejection, and autoimmune disorders
AIMs	2965	Ancestry-informative markers	KIR genes[4]	25	Autoimmune disorders and disease defense
APOE[4]	18	Cardiovascular disease, Alzheimer's disease, immunoregulation, and cognition	Neanderthal SNPs[13]	1553	Neanderthal ancestry and human population migration
Blood phenotype genes[5]	2002	Blood phenotypes	Newborn/carrier screening gene coverage	26,567	genes associated with severe, recessive childhood diseases included in the TruSight™ Inherited Disease Sequencing Panel
ClinVar[6] variants	43,315	Relationships among variation, phenotypes, and human health	NHGRI-EBI GWAS catalog[14]	10,663	Markers from published genome-wide association studies
ClinVar pathogenic	14,387	Cause Mendelian disorders based on criteria using typical variant evidence (eg, population, computational, functional, or segregation data)	PharmGKB[15]	4642	Human genetic variation associated with drug responses
ClinVar likely pathogenic	5625	Likely cause Mendelian disorders based on criteria using typical variant evidence (eg, population, computational, functional, or segregation data)	RefSeq[17] 3' UTRs	13,481	3' untranslated regions of known genes
ClinVar benign	4816	Do not cause Mendelian disorders based on criteria using typical variant evidence (eg, population, computational, functional, or segregation data)	RefSeq 5' UTRs	6485	5' untranslated regions of known genes
ClinVar likely benign	3864	Likely do not cause Mendelian disorders based on criteria using typical variant evidence (eg, population, computational, functional, or segregation data)	RefSeq All UTRs	19,401	All untranslated regions of known genes
COSMIC[7] genes	305,130	Somatic mutations in cancer	RefSeq	340,002	All known genes
GO[8] CVS genes	104,357	Cardiovascular conditions	RefSeq +/- 10 kb	396,916	All known genes plus regulatory regions
Database of Genomic Variants[9]	516,749	Genomic structural variation	RefSeq Promoters	15,046	2 kb upstream of all known genes to include promoter regions
eQTLs[10]	2734	Genomic loci regulating mRNA expression levels	RefSeq Splice Regions	3319	Variants at splice sites in all known genes
Fingerprint SNPs[11]	577	Human identification			

a. Extended MHC is a ~ 8 Mb region.

Abbreviations: ACMG: American College of Medical Genetics; ADME: absorption, distribution, metabolism, and excretion; AIM: ancestry-informative marker; APOE: apolipoprotein E; COSMIC: catalog of somatic mutations in cancer; CPIC: Clinical Pharmacogenetics Implementation Consortium; EBI: European Bioinformatics Institute; eQTL: expression quantitative trait loci; gnomAD: Genome Aggregation Database; GO CVS: gene ontology annotation of the cardiovascular system; GWAS: genome-wide association study; HLA: human leukocyte antigen; KIR: killer cell immunoglobulin-like receptor; MHC: major histocompatibility complex; NHGRI: national human genome research institute; PharmGKB: Pharmacogenomics Knowledgebase; RefSeq: reference sequence; UTR: untranslated region.

Widespread adoption

The Infinium GSA-24 v2.0 BeadChip builds on the success of the consortium version of the Infinium Global Screening Array developed by a community of human disease researchers, health care networks, consumer genomics companies, and genomic service providers. The Infinium GSA-24 v2.0 BeadChip has been widely adopted with over 6 million BeadChips ordered by a global community that provides a network of users powering discovery through collaboration and data sharing.

Optimized and relevant global content

The Infinium GSA-24 v2.0 BeadChip combines highly optimized multiethnic genome-wide content, curated clinical research variants, and QC markers for a broad range of clinical research and variant screening applications. These applications include disease association and risk profiling studies, pharmacogenomics research, disease characterization, lifestyle and wellness characterization, and marker discovery in complex disease research (Figure 2).

The launch of the Infinium GSA-24 v2.0 BeadChip includes the addition of more than 60,000 markers, enhancing coverage in key content categories of pharmacogenomics, ClinVar, HLA, and ACMG variants.

Broad clinical research applications

The clinical research content of the Infinium GSA-24 v2.0 BeadChip was designed through collaboration with medical genomics experts using multiple annotation databases[2-9] to create an informative, cost-effective panel for clinical research applications (Table 2 and Table 3).

Table 3: Marker information

Marker categories			No. of markers
Exonic markers[a]			88,377
Intronic markers[a]			268,023
Nonsense markers[b]			5639
Missense markers[b]			53,093
Synonymous markers[b]			9071
Mitochondrial markers[c]			1180
Indels[c]			8626
Sex chromosomes[c]	X 28,594	Y 4964	PAR/homologous 1229

a. RefSeq - NCBI Reference Sequence Database[6]
b. Compared against the University of California, Santa Cruz (UCSC) Genome Browser[4]
c. NCBI Genome Reference Consortium, Version GRCh37[20]

Abbreviations: PAR: pseudoautosomal region

Expertly selected content

Variants included on the array consist of markers with known disease association based on ClinVar,[6] the Pharmacogenomics Knowledgebase (PharmGKB),[15,16] and the National Human Genome Research Institute (NHGRI)-EBI database[14] (Figure 3). In addition to disease-associated markers, the GSA contains imputation-based tagSNPs for HLA alleles, extended MHC region, the KIR gene, and exonic content from the gnomAD database.[4]

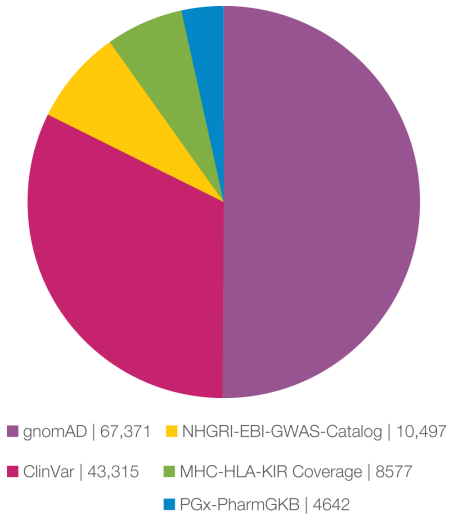

Figure 3: Clinical research content on the Infinium GSA-24 v2.0 BeadChip–Clinical research content was expertly selected from scientifically recognized databases to create a highly informative array for clinical research applications.

Broad spectrum of pharmacogenomics markers and exonic content

The Infinium GSA-24 v2.0 BeadChip features pharmacogenomics variants associated with absorption, distribution, metabolism, and excretion (ADME) phenotypes based on PharmGKB[16] and Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines[18] (Figure 2). It also features diverse exonic content from the ExAC database,[19] including both cross population and population specific markers (Figure 5) with either functionality or strong evidence for association.

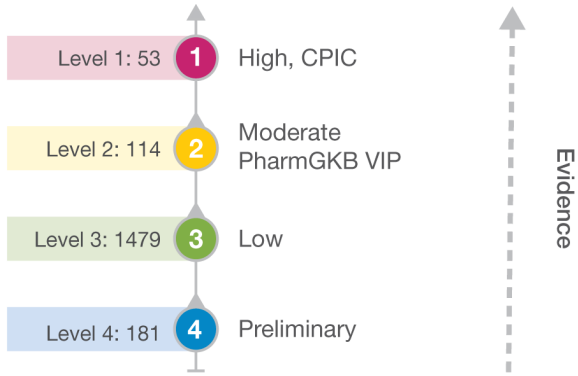

Figure 4: Broad spectrum of pharmacogenomics markers–Clinical research content features an extensive list of pharmacogenomics markers selected based on CPIC guidelines and the PharmGKB database.[10] Markers are arranged according to level of evidence as defined by the PharmGKB database. VIP: very important pharmacogene.

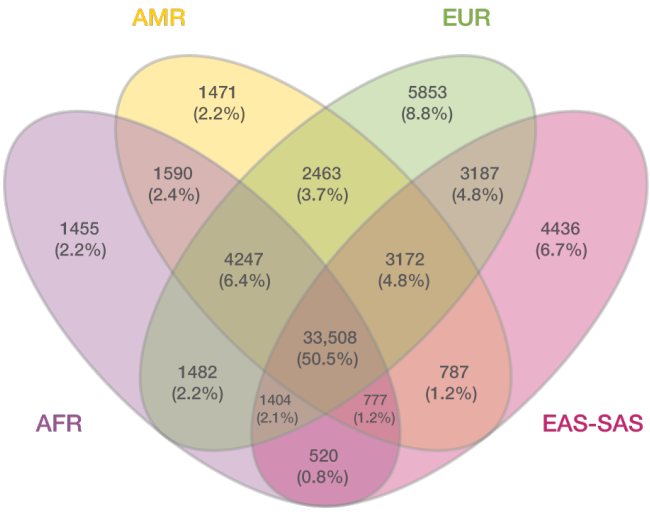

Figure 5: Global exonic content is cross-population and population-specific– Exonic content included on the Infinium GSA-24 v2.0 BeadChip contains content that is present across several populations as well as population-specific content. The Venn diagram displays proportion of total content that either overlaps or is specific to certain populations. Abbreviations: EAS: East Asian; SAS: South Asian; AMR: Ad Mixed American; AFR: African; EUR: European.

Extensive range of disease categories covered

Including over 28,000 variants with established clinical associations based on the ClinVar database,[6] clinical research content on the Infinium GSA-24 v2.0 BeadChip enables validation of disease associations, risk profiling, preemptive screening research, and pharmacogenomics studies. Variant selection includes a range of pathology classifications based on the ClinVar American College of Medical Genetics and Genomics (ACMG) annotations (Figure 6A).[2] There are over 10,000 disease and trait associations from the ClinVar database (Figure 6B) and over 7000 variants selected from the NHGRI-GWAS catalog[14] (Figure 7), representing a broad range of phenotypes and disease classifications.

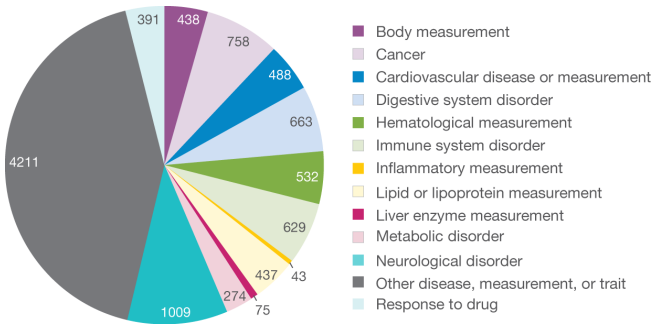

Figure 7: NHGRI disease categories– Infinium GSA-24 v2.0 BeadChip clinical research content features over 9000 markers across 20 disease categories based on the NHGRI database.

QC markers for sample identification, tracking, and stratification

The Infinium GSA-24 v2.0 BeadChip includes QC and high-value markers for large-scale studies, enabling sample identification, tracking, ancestry determination, and stratification (Figure 8).

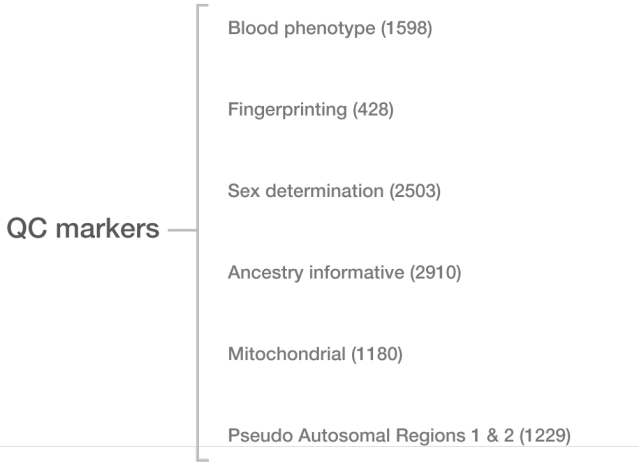

Figure 8: QC markers–QC variants on the Infinium GSA-24 v2.0 BeadChip enable a variety of capabilities for sample tracking such as sex determination, continental ancestry, and forensics.

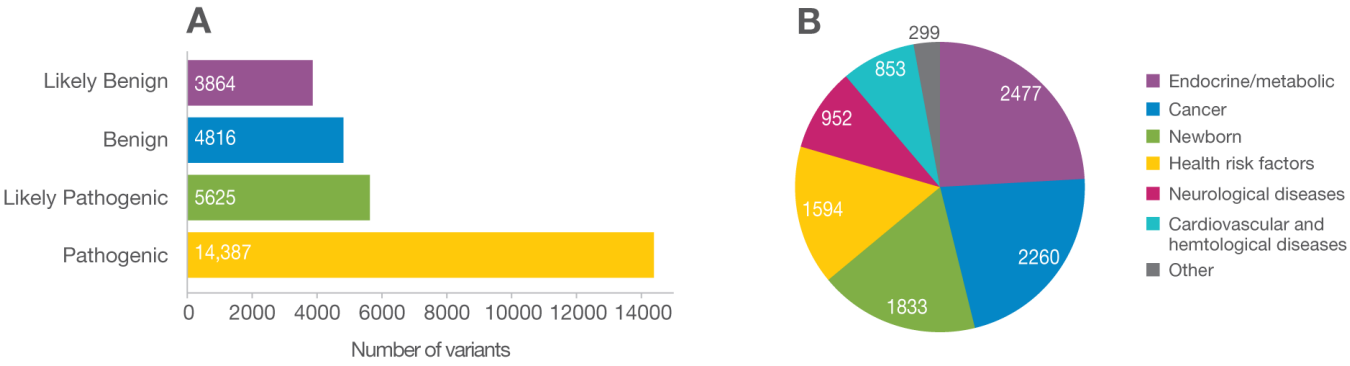

Figure 6: Broad coverage of disease categories– (A) Variants sorted by range of pathology classifications according to ClinVar American College of Medical Genetics (ACMG) annotations. (B) Infinium GSA-24 v2.0 BeadChip clinical research content features over 10,000 markers based on the ClinVar database.

Flexible content options

The Infinium GSA-24 v2.0 BeadChip can be customized to incorporate up to 50,000 custom bead types or predesigned content panels (Table 4).

Table 4: Flexible content options

Optional compatible content	No. of markers	Description
Custom content	≤ 50,000 bead types	Custom design virtually any target (eg, SNP, CNV, indel) using the DesignStudio™ Microarray Assay Designer[a]
Multi-disease drop-In panel	~ 50,000 markers	Fine-mapping content derived from exome sequencing and meta analysis of phenotype-specific consortia focused on the following traits: psychiatric, neurological, cancer, cardiometabolic, autoimmune, anthropometric
Infinium PsychArray-24 focused content panel	~ 30,000 m arkers	Markers from the Infinium PsychArray-24 v1.1 BeadChip[b] associated with common psychiatric disorders including, schizophrenia, bipolar disorder, autism spectrum disorders, attention deficit hyperactivity disorder, major depressive disorders, obsessive compulsive disorder, anorexia, Tourette's syndrome

a. www.illumina.com/designstudio.html.
b. www.illumina.com/products/by-type/microarray-kits/infinium-psycharray.html.

Abbreviations: SNP: single nucleotide polymorphism; CNV: copy number variation; indel: insertion/deletion.

High-throughput workflow

The Infinium GSA-24 v2.0 BeadChip uses the highly scalable 24-sample Infinium HTS format which enables laboratories to efficiently increase throughput as needed to support population-scale research and variant screening applications. The HTS format includes two different assay options, optimized based on processing throughput. For flexible throughput processing, the Infinium HTS assay provides the capability to run hundreds to thousands of samples per week. The Infinium HTS assay provides a rapid, three-day workflow that allows genotyping service providers and clinical researchers to gather data and advance studies quickly (Figure 9).

Optional integration of the Illumina Laboratory Information Management System (LIMS) into the workflow provides high laboratory efficiency with automation functionality, process tracking, and QC data tracking. The Illumina ArrayLab Consulting Service offers customized solutions to high-throughput genotyping labs that desire increased efficiency and overall operational excellence.

Robust, high-quality assay

The Infinium GSA-24 v2.0 BeadChip uses proven Infinium assay chemistry to deliver the same high-quality, reproducible data (Table 5) that Illumina genotyping arrays have provided for over a decade. The Infinium product line provides high call rates and high reproducibility for numerous sample types including, saliva, blood, solid tumors, fresh frozen, and buccal swabs. It is compatible with the Infinium FFPE QC and DNA Restoration Kits,[21] enabling genotyping of formalin-fixed, paraffin-embedded (FFPE) samples. In addition, the high signal-to-noise ratio of the individual genotyping calls from the Infinium assay provides researchers with access to genome-wide copy number variant (CNV) calling with a mean probe spacing of ~ 4.4 kb.

Table 5: Data performance and spacing

Data performance	Value[a]	Product specification[b]	
Call rate	99.5%	> 99% avg	
Reproducibility	99.99%	> 99.9%	
Log R deviation	0.109	< 0.30[c]	
Spacing			
Spacing (kb)	Mean	Median	90th%[c]
	4.4	2.3	10.5

a. Values are derived from genotyping 1316 HapMap reference samples.
b. Excludes Y chromosome markers for female samples.
c. Value expected for typical projects using standard Illumina protocols. Tumor samples and samples prepared by methods other than standard Illumina protocols are excluded.

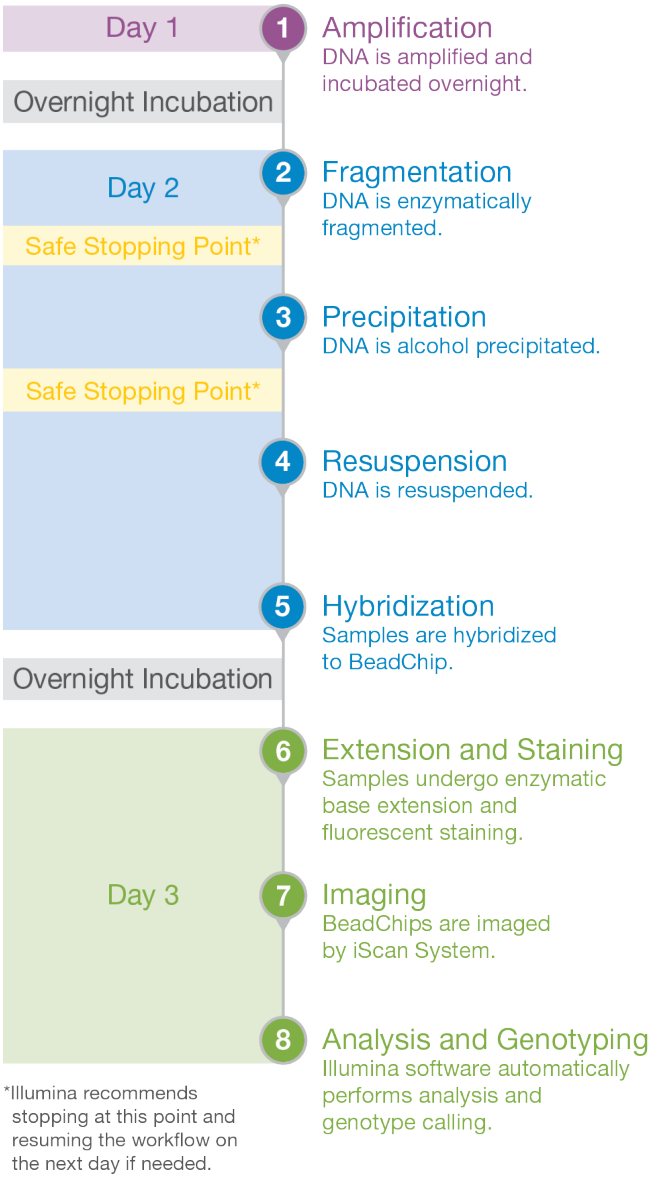

Figure 9: The Infinium HTS workflow–The Infinium HTS format provides a rapid 3-day workflow with minimal hands-on time.

High imputation accuracy for global populations

Leveraging available whole-genome reference data from over 26 global populations in Phase 3 of the 1000 Genomes Project,[1] the genome-wide content on the Infinium GSA-24 v2.0 BeadChip has been selected to generate high imputation accuracy for low-frequency and common variants (minor allele frequencies (MAF) of > 1%) (Tables 6–9). High imputation accuracy provides increased power to support population-scale disease research and population-specific causal variant detection.

Table 6: Imputation accuracy from 1000G[a] at various MAF thresholds

Population[b]	Imputation accuracy		
	MAF ≥ 5%	MAF ≥ 1%	MAF 1–5%
AFR	0.91	0.86	0.79
AMR	0.95	0.92	0.85
EAS	0.94	0.89	0.77
EUR	0.95	0.93	0.87
SAS	0.94	0.89	0.78

a. Compared against Phase 3, version 5 of the 1000 Genomes Project (1000G). www.1000genomes.org. Accessed July 2016.
b. www.1000genomes.org/category/frequently-asked-questions/population.

Abbreviations: MAF: minor allele frequency.

Table 7: Number of markers imputed at $r^2 \geq 0.80$ from 1000G[a]

Population[b]	Number of markers imputed at $r^2 \geq 0.80$ (% of total markers)		
	MAF ≥ 5%	MAF ≥ 1%	MAF 1–5%
AFR	6.5 M (76%)	11.1 M (70%)	4.6 M (63%)
AMR	5.6 M (89%)	12.0 M (90%)	6.4 M (91%)
EAS	4.8 M (86%)	8.6 M (86%)	3.8 M (85%)
EUR	5.5 M (90%)	9.7 M (89%)	4.2 M (87%)
SAS	5.4 M (87%)	9.6 M (85%)	4.3 M (82%)

a. Compared against Phase 3, version 5 of the 1000 Genomes Project (1000G). www.1000genomes.org. Accessed July 2016.
b. www.1000genomes.org/category/frequently-asked-questions/population.

Table 8: No. of markers LD $r^2 \geq 0.80$ from 1000G[a] at various MAF thresholds

1000G population[b]	LD coverage ($r^2 \geq 0.80$)		
	MAF ≥ 5%	MAF ≥ 1%	MAF 1–5%
AFR	1.8 M (22%)	2.2 M (14%)	279 K (4%)
AMR	2.9 M (47%)	3.7 M (38%)	750 K (21%)
EAS	3.2 M (59%)	4.0 M (53%)	818 K (39%)
EUR	3.1 M (52%)	4.3 M (50%)	1.3 M (47%)
SAS	3.1 M (51%)	3.8 M (43%)	660 K (24%)

a. Compared against Phase 3, version 5 of the 1000 Genomes Project (1000G). www.1000genomes.org. Accessed July 2016.
b. www.1000genomes.org/category/frequently-asked-questions/population.

Abbreviations: LD: linkage disequilibrium.

Table 9: LD mean r^2 from 1000G[a] at various MAF thresholds

Population[b]	LD coverage (mean r^2)		
	MAF ≥ 5%	MAF ≥ 1%	MAF 1–5%
AFR	0.44	0.30	0.11
AMR	0.69	0.58	0.35
EAS	0.75	0.68	0.49
EUR	0.71	0.68	0.58
SAS	0.71	0.61	0.35

a. Compared against Phase 3, version 5 of the 1000 Genomes Project (1000G). www.1000genomes.org. Accessed July 2016.
b. See www.1000genomes.org/category/frequently-asked-questions/population.

Summary

The Infinium Global Screening Array-24 v2.0 BeadChip provides a cost-effective solution for population-scale genetic studies, variant screening, and precision medicine research. The Infinium GSA-24 v2.0 BeadChip builds on the success of the consortium version of the Infinium Global Screening Array, which has been widely adopted with over 6 million BeadChips ordered worldwide. Using the proven iScan System, Infinium HTS Assay, and integrated analysis software, this high-density, 24-sample BeadChip provides optimized content for a broad range of clinical research applications.

Ordering information

Infinium Global Screening Array-24 v2.0 BeadChip Kit	Catalog no.
48 Samples	20024444
288 Samples	20024445
1152 Samples	20024446
Infinium Global Screening Array-24+ v2.0 BeadChip Kit*	**Catalog no.**
48 Samples	20024447
288 Samples	20024448
1152 Samples	20024449

*Enabled for custom content.

Learn more

To learn more about the Infinium Global Screening Array-24 v2.0 BeadChip and other Illumina genotyping products and services, visit www.illumina.com/genotyping.html.

References

1. The 1000 Genomes Project. www.1000genomes.org. Accessed July 16, 2016.
2. ACMG Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing. www.ncbi.nlm.nih.gov/clinvar/docs/acmg/. Accessed January 2017.
3. PharmaADME Gene List. www.pharmaadme.org. Accessed August 2014.
4. University of California, Santa Cruz (UCSC) Genome Browser. genome.ucsc.edu. Accessed July 2016.
5. NCBI Reference Sequence Blood Group Antigen Gene Mutation Database. www.ncbi.nlm.nih.gov/projects/gv/rbc/xslcgi.fcgi?cmd=bgmut/systems. Accessed July 2016.
6. ClinVar Database. www.ncbi.nlm.nih.gov/clinvar. Accessed October 2016.
7. Catalog of somatic mutations in cancer. cancer.sanger.uk/cosmic. Accessed July 2016.
8. Gene Ontology Consortium. www.geneontology.org. Accessed July 2016.
9. Database of Genomic Variants. dgv.tcag.ca/dgv/app/home. Accessed July 2016.
10. NCBI eQTL Database. www.ncbi.nlm.nih.gov/projects/gap/eqtl/index.cgi. Accessed July 2016.
11. The Allele Frequency Database. alfred.med.yale.edu/alfred/snpSets.asp. Accessed July 2016.
12. 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.
13. Neanderthal Genome Browser. neandertal.ensemblgenomes.org/index.html. Accessed July 2016.
14. National Human Genome Research Institute. www.genome.gov/. Accessed January 2017.
15. PharmGKB, The Pharmacogenomics Knowledgebase. www.pharmgkb.org. Accessed January 2017.
16. PharmGKB, Clinical Annotation Levels of Evidence. www.pharmgkb.org/page/clinAnnLevels. Accessed January 2017.
17. RefSeq - NCBI Reference Sequence Database. www.ncbi.nlm.nih.gov/refseq. Accessed September 2016.
18. Clinical Pharmacogenetics Implementation Consortium (CPIC). cpicpgx.org. Accessed October 2016
19. Exome Aggregation Consortium (ExAC) Browser. exac.broadinstitute.org. Accessed October 2016.
20. NCBI Genome Reference Consortium. Version GRCh37. www.ncbi.nlm.nih.gov/grc/human. Accessed July 2016.
21. Infinium FFPE QC and DNA Restoration Kit. www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_FFPE_DNA_restoration.pdf.

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