DATA SHEET

Infinium™ Global Diversity Array with Polygenic Risk Score Content-8 v1.0

A versatile, scalable solution for multi-ethnic PRS research

- Updated content includes 160K PRS markers for broad disorder coverage
- Comprehensive genome-wide coverage enables highly accurate, pan-ethnic PRS determination
- Compatibility with the Predict software module simplifies PRS computation and reporting

illumina®

Introduction

Accurate assessment of disease risk is a key aspect of preventive and personalized medicine. However, for diseases with polygenic inheritance, which are influenced by hundreds to thousands of genetic variants that act in conjunction with environmental factors, risk prediction remains challenging. Polygenic risk scores (PRS), also known as polygenic scores or genetic risk scores, are numerical indicators that represent the genetic propensity for an individual to develop a trait or disease.[1] PRS is a useful tool for researchers to study risk-stratification and better understand who would be most likely to benefit from additional monitoring or early preventive interventions.[1-3]

The Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 (Table 1) is a high-performance microarray solution that provides a flexible and scalable workflow to support PRS studies. The eight-sample BeadChip, built on the backbone of the proven Infinium Global Diversity Array, features updated genotyping content including 160K markers for accurate PRS determination (Table 2). Data from the BeadChip is analyzed optimally using the Predict software module, offering a complete genotype-to-risk solution for polygenic disease research.

Content optimized for PRS research

The genome-wide backbone of the Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 contains ~1.9M scientifically relevant markers. This comprehensive content from the Infinium Global Diversity Array-8 v1.0, the genotyping array used in the All of Us Research Program, is supplemented with 160K PRS markers from the expansive Polygenic Score (PGS) Catalog[4] to provide expanded coverage across high-value genomic regions (Table 2, Table 3). These markers, selected in collaboration with Allelica, provide exceptional coverage for six key PRS disorders including type 1 and type 2 diabetes, breast cancer, colon cancer, prostate cancer, coronary artery disease, and Alzheimer's disease.

To generate the PRS booster content, array samples were benchmarked against scores from the gold standard 1000 Genomes Project and common PGS Catalog scores (eg, MAV313) for all Allelica PRS. The purpose of this benchmark was to identify the variants with the most common errors in imputation by comparing imputation results with 30× whole-genome sequencing (WGS) data.

To develop the array, variants with the highest effect size on the PRS and the lowest imputation quality were selected. These variants were included directly in the array to be genotyped instead of being imputed. The PRS booster content includes variants with the highest effect size specific to each ancestry in the Allelica multiancestry scores to optimize the value of the array for individuals from all ancestries and increase concordance for the PRS.

Table 1: Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 overview

Feature	Description
Species	Human
Total number of markers[a]	2,028,571
Number of samples per BeadChip	8
DNA input requirement	200 ng
SNP replicates	15
Number of SNPs needed to call CNV	10
Assay chemistry	Infinium LCG
Instrument support	iScan System
Maximum iScan System sample throughput[b]	~1728 samples/week
Scan time per sample[b]	3–5 minutes

a. Content includes ~1.9M markers from the genome-wide backbone from Infinium Global Diversity Array-8 v1.0 plus 160K PRS-specific markers.
b. Approximate values. Scan times and maximum throughput will vary depending on laboratory and system configurations.

Table 2: Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 high-value content

Content	No. of markers[a]	Research application/notes
RefSeq hg19 genes	1,123,407	All known genes
RefSeq hg19 ± 10 kb	1,266,608	Regulatory regions[b]
RefSeq promoters	48,249	2 kb upstream to include promoter regions
ADME exonic	16,528	Drug absorption, distribution, metabolism, and excretion
ADME hg19 genes	32,117	
ADME hg19 ± 10 kb	37,468	Includes regulatory regions
HLA markers	17,595	Disease defense, transplant rejection, autoimmune disorders
HLA hg19 genes	1297	
MHC markers	22783	
COSMIC hg19 genes	1,079,088	Somatic mutations in cancer
GO hg19 genes	324,039	Gene ontology annotation

a. The number of markers for each category is subject to change.
b. Of all known genes.

Abbreviations: ADME, absorption, distribution, metabolism, and excretion; COSMIC, catalog of somatic mutations in cancer; GO, gene ontology database; hg19, human genome version 19; HLA, human leukocyte antigen; MHC, major histocompatibility complex; RefSeq, National Center for Biotechnology Information Reference Sequence Database.

Table 3: Marker information

Marker category	No. of markers[a]
Exonic markers[a]	530,644
Nonsense markers[b]	28,287
Missense markers[b]	396,783
Synonymous markers[b]	33,442
Silent SNPs[b]	40,964
Mitochondrial DNA markers[b]	1346
Indels[c]	37,289
Chr X markers[c]	62,103
Chr Y markers[c]	6449
PAR/homologous markers[c]	5485

a. RefSeq, NCBI Reference Sequence Database, ncbi.nlm.nih.gov/refseq.
b. Compared against the UCSC Genome Browser, genome.ucsc.edu.
c. NCBI Genome Reference Consortium, Version GRCh37 (hg19), ncbi.nlm.nih.gov/assembly/GCF_000001405.13.

Abbreviations: Indel, insertion/deletion; NCBI, National Center for Biotechnology Information; PAR, pseudoautosomal region; SNP, single nucleotide polymorphisms; UCSC, University of California Santa Cruz.

Broad coverage of disease-associated variants

The Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 provides coverage of variants selected from the National Human Genome Research Institute and European Bioinformatics Institute (NHGRI-EBI) GWAS catalog[5] representing a broad range of phenotypes and disease classifications (Figure 1). Variant selection also includes a range of pathology classifications based on ClinVar[6] and American College of Medical Genetics (ACMG)[7] annotations.

Clinical databases, such as ClinVar, are constantly evolving as new variants are added and variant designations change to 'Pathogenic' or 'Likely pathogenic'. The BeadChip provides updated coverage of many of these high-value variants contained within annotated databases. This comprehensive content provides powerful opportunities for researchers interested in studying diverse populations to test and validate associations previously found in European populations.

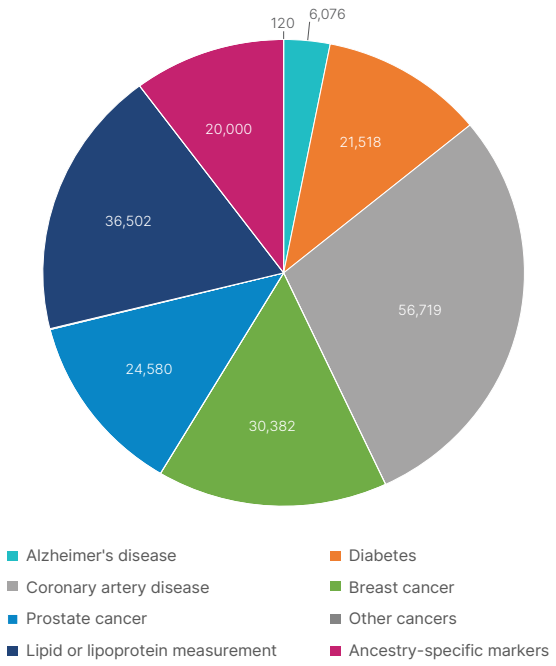

Figure 1: PRS booster content covers multiple phenotypes—The additional PRS content included on the Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 features SNP markers across a broad range of disease categories. A subset of markers ensure optimal performance for specific ancestral populations. Note that because many markers are shared across phenotypes, the sum of these values does not equal 160K markers.

Exceptional pan-ethnic exonic content

The Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 includes enhanced tagging in exonic regions and enriched coverage to map loci from genome-wide association studies (GWAS) with previously identified disease or trait associations. More than 400,000 exome markers were gathered from 36,000 individuals from diverse ethnic groups, including African Americans, Hispanics, Pacific Islanders, East Asians, Europeans, and individuals of mixed ancestry. The array also features diverse exonic content from the ExAC/gnomAD database,[8] including cross-population and population-specific markers with either functionality or strong evidence for association (Table 4).

Table 4: Exonic coverage across populations

Population(s)[a,b]	No. of markers
NFE	346,340
EAS	146,281
AMR	272,178
AFR	257,690
SAS	224,431
NFE/EAS/AMR/AFR/SAS	69,432

a. internationalgenome.org/category/population.
b. Based on gnomAD, gnomad.broadinstitute.org.

Abbreviations: NFE, non-Finnish European; EAS, East Asian; AMR, admixed American; AFR, African; SAS, South Asian.

QC markers for sample tracking

The Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 includes ~10K quality control (QC) markers. This QC marker content enables important sample tracking functions, ancestry determination, and stratification to facilitate high-throughput studies (Figure 2).

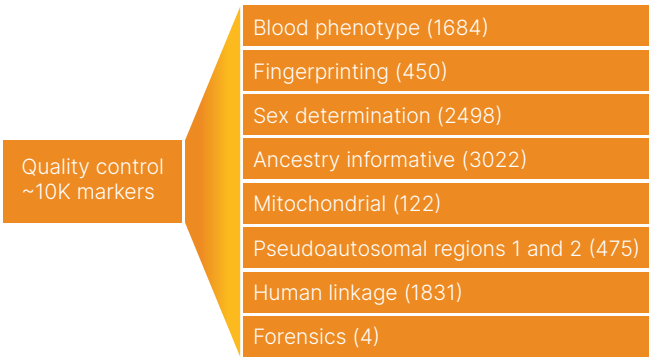

Figure 2: QC markers by category— QC variants on the array enable various capabilities for sample tracking such as sex determination, continental ancestry, human identification, and more.

High-throughput workflow

The Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 uses the Infinium eight-sample BeadChip format to enable laboratories to scale efficiently. For flexible throughput processing, the Infinium LCG assay provides the capability to run up to 1728 samples per week using a single iScan™ System. The Infinium assay provides a three-day workflow that allows researchers to gather and report data quickly with minimal hands-on time (Figure 3).

Trusted high-quality assay

The Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 uses trusted Infinium assay chemistry to deliver the same high-quality, reproducible data (Table 5) that Illumina genotyping arrays have provided for over a decade. In addition, the high signal-to-noise ratio of the individual genotyping calls from the Infinium assay provides access to genome-wide copy number variation (CNV) calling.

Table 5: Data performance and spacing

Data performance	Value[a,b]	Product specification[c]
Call rate	99.78%	> 99% avg
Reproducibility	99.99%	> 99.90% avg
Spacing		

	Mean	Median	90th%[d]
Spacing (kb)	1.439	0.619	3.680

a. Values are derived from genotyping 2228 HapMap reference samples.
b. Values expected for typical projects using standard Illumina protocols. Tumor samples and samples prepared by nonstandard protocols are excluded.
c. Excludes Y chromosome markers for female samples.
d. Based on results from GenTrain sample set.

Figure 3: Infinium LCG workflow— The Infinium assay uses a rapid three-day workflow requiring minimal hands-on time.

Flexible content options

The Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 can be customized to incorporate bead types that meet specific PRS research needs. Add-on content can be selected and exported from the PGS catalog and incorporated into the BeadChip using the DesignStudio™ Microarray Assay Designer, to increase variant coverage.

Powerful analysis pipeline for PRS research

Data generated by the Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 is compatible with analysis using BaseSpace™ Sequence Hub built on the trusted Illumina Connected Analytics (ICA) cloud-based infrastructure. For laboratories looking for a complete genotype-to-risk prediction workflow, the Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 is analyzed optimally with the Predict software module. This easy-to-use analysis software is designed to output individual PRS values in relation to those computed across a reference population with known phenotypes to generate automated PRS reports with risk prediction. The Predict software module can compute up to 24 PRS on up to 1152 samples in 15 minutes, for fast and accurate ancestry-informed PRS determination.

Summary

The Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 is part of a comprehensive toolkit that provides a streamlined genotype-to-risk prediction solution for PRS research. The Illumina PRS toolkit simplifies the PRS development process, saving months of data analysis typically performed by dedicated bioinformaticians. This flexible solution is compatible with low- and high-throughput applications, and can be customized for broad, ancestry-informed PRS reporting.

Learn more

Polygenic risk scores, illumina.com/areas-of-interest/complex-disease-genomics/polygenic-risk-scores

Infinium Global Diversity Array, illumina.com/products/by-type/microarray-kits/infinium-global-diversity

Predict software module, illumina.com/products/by-type/informatics-products/polygenic-risk-score-software

Ordering information

Product	Catalog no.
Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 (16 samples)	20090683
Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 (48 samples)	20090684
Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 (96 samples)	20090685
Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 (384 samples)	20090686

References

1. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 2018;50(9):1219-1224. doi:10.1038/s41588-018-0183-z
2. Lambert SA, Abraham G, Inouye M. Towards clinical utility of polygenic risk scores. Hum Mol Genet. 2019;28(R2):R133-R142. doi:10.1093/hmg/ddz187
3. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. Genome Med. 2020;12(1):44. doi:10.1186/s13073-020-00742-5
4. The Polygenic Score (PGS) Catalog website. pgscatalog.org. Accessed August 18, 2022.
5. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 2019;47(D1):D1005-D1012. doi:10.1093/nar/gky1120
6. NCBI. ClinVar Database website. ncbi.nlm.nih.gov/clinvar. Accessed August 18, 2022.
7. Green RC, Berg JS, Grody WW, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 2013;15(7):565-574. doi:10.1038/gim.2013.73
8. Broad Institute. Genome Aggregation Database (gnomAD) Browser website. gnomad.broadinstitute.org. Accessed August 18, 2022.

illumina®

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

M-GL-01187 V1.0