DATA SHEET

Infinium™ Global Diversity Array with Enhanced PGx-8 v1.0

Versatile genotyping solution for pharmacogenomics, disease, and diversity studies

- Comprehensive coverage of over 6000 annotated variants from public PGx databases, including key genes like *CYP2D6*
- Genome-wide scaffold to detect common and low-frequency variants across a range of phenotypes
- Robust CNV detection and targeted amplification to allow PGx pseudogene disambiguation

illumina®

Introduction

The Infinium Global Diversity Array with Enhanced PGx-8 v1.0 BeadChip is the most comprehensive genotyping microarray on the market for supporting pharmacogenomic (PGx) research along with polygenic risk score development, ancestry determination, and genetic disease research. The array is built on an eight-sample BeadChip that contains more than 1.9 million markers on a genome-wide backbone (Figure 1, Table 1).

Other noteworthy attributes are the inclusion of more than 44,000 absorption, distribution, metabolism, and excretion (ADME) markers spanning more than 2000 genes and exceptional coverage of priority level A and B Clinical Pharmacogenetics Implementation Consortium (CPIC) variants (Figure 2, Table 2).[1,2] High-impact PGx genes like *CYP2D6*, *CYP2B6*, and *TPMT* that have historically been challenging to discern are now accessible due to significant workflow improvements that enable disambiguation of pseudogenes. In addition, optional DRAGEN™ (Dynamic Read Analysis for GENomics) Array secondary analysis software allows generation of a report containing metabolizer status and star allele calls.

Table 1: Product information[a]

Feature	Description
Species	Human
Total number of markers[b]	1,933,117
Capacity for custom bead types	125,000
Number of samples per BeadChip	8
DNA input requirement	200 ng
Assay chemistry	Infinium LCG
Instrument support	iScan System
Maximum iScan System sample throughput	~1728 samples/week
Scan time per sample	4.4 minutes

a. Approximate values, scan times, and maximum throughput will vary depending on laboratory and system configurations.
b. Variants found on commercial manifest.

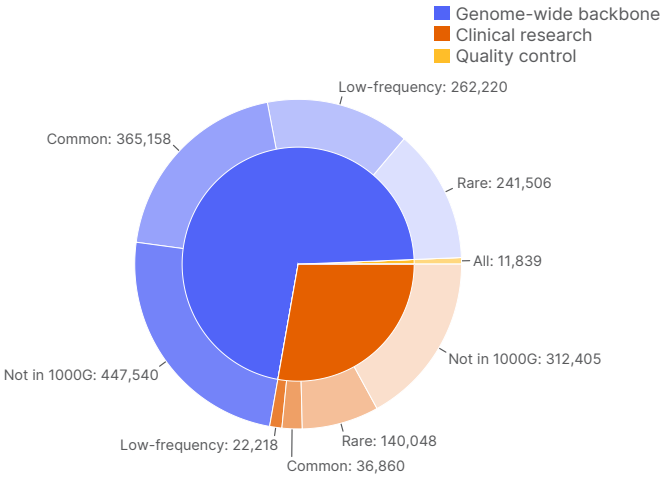

Figure 1: Summary of content—Plotted in the inner pie is the proportion of the array selected for genome-wide coverage, clinical research, and quality control (QC). The outer ring summarizes the weighted reference global allele frequency for unique variants present in the 1000 Genomes Project (1000G).[3] Variants not in 1000G are labeled. Counts represent unique variants.

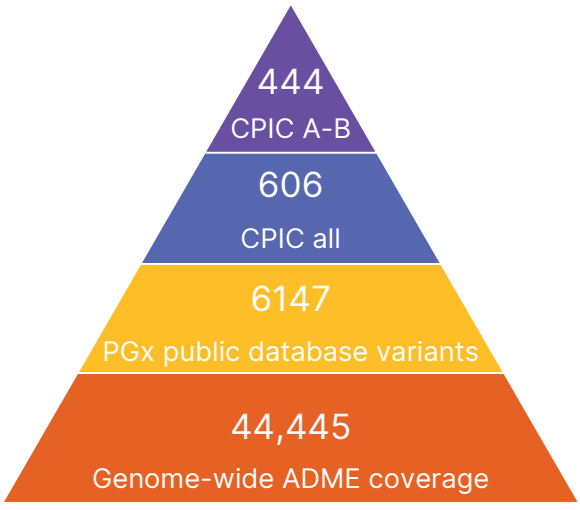

Figure 2: Broad spectrum of pharmacogenomics markers—Clinical research content developed from an extensive list of pharmacogenomics markers selected based on CPIC guidelines and the PharmGKB database.[1,2] Content includes PGx public database variants, variants annotated in PharmGKB, PharmVar,[4] CPIC, genome-wide PGx coverage, extended ADME genes, CPIC level A genes, including targeted imputation tag SNPs, and CPIC level A CNV tags.

Table 2: Infinium Global Diversity Array with Enhanced PGx-8 v1.0 high-value content

Content	No. of markers[a]	Research application/note	Content	No. of markers	Research application/note
ACMG[5] 59 2016 gene coverage	30,878	Variants with known clinical significance identified from clinical WGS and WES samples	GO[9] CVS genes	331,520	Cardiovascular conditions
ACMG 59 all annotations	22,812		Database of Genomic Variants[10]	1,440,531	Genomic structural variation
ACMG 59 pathogenic	6753		eQTLs[11]	6394	Genomic loci regulating mRNA expression levels
ACMG 59 likely pathogenic	2946		Fingerprint SNPs[12]	481	Human identification
ACMG 59 benign	2051		gnomAD[13] exome	502,547	WES and WGS results from unrelated individuals from various studies
ACMG 59 likely benign	3701		HLA genes[14]	18,893	Disease defense, transplant rejection, and autoimmune disorders
ACMG 59 VUS	5330		Extended MHC[14,c]	23,965	Disease defense, transplant rejection, and autoimmune disorders
AIMs[b]	3047	Ancestry-informative markers	KIR genes[6]	154	Autoimmune disorders and disease defense
			Neanderthal SNPs[16]	2095	Neanderthal ancestry and human population migration
			Newborn/carrier screening gene coverage	61,902	Genes associated with childhood diseases included in the TruSight™ Inherited Disease Sequencing Panel[16]
APOE[6]	86	Cardiovascular disease, Alzheimer's disease, and cognition	NHGRI-EBI GWAS catalog[17]	32,585	Markers from published GWAS
			PharmGKB[1,4] all	5116	Human genetic variation associated with drug responses
ClinVar[7] variants	110,608	Relationships among variation, phenotypes, and human health	PharmGKB level 1A	297	
ClinVar pathogenic	20,719		PharmGKB level 1B	8	
ClinVar likely pathogenic	8241		PharmGKB level 2A	56	
ClinVar benign	29,366		PharmGKB level 2B	49	
ClinVar likely benign	19,298		PharmGKB level 3	1911	
ClinVar VUS	24,342		PharmGKB level 4	446	
COSMIC[8] genes	1,043,886	Somatic mutations in cancer	RefSeq[18] 3′ UTRs	46,399	3′ untranslated regions[d]
CPIC[2] all	606	Variants with potential guidelines to optimize drug therapy	RefSeq 5′ UTRs	30,386	5′ untranslated regions[d]
CPIC-A	413		RefSeq All UTRs	74,608	Untranslated regions[d]
CPIC-A/B	3		RefSeq	1,121,140	All known genes
CPIC-B	28		RefSeq +/- 10 kb	1,262,045	Regulatory regions[d]
CPIC-C	43		RefSeq Promoters	45,221	2 kb upstream to include promoter regions[d]
CPIC-C/D	2		RefSeq splice regions	12,106	Variants at splice sites[d]
CPIC-D	60				

a. The number of markers for each category are subject to change.
b. Based on internal calculations.
c. Extended MHC is a 8 Mb region.
d. Of all known genes.

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, NCBI Reference Sequence Database; NCBI, National Center for Biotechnology Information; UTR, untranslated region; VUS, variant of unknown significance; WES, whole-exome sequencing; WGS, whole-genome sequencing.

The Global Diversity Array with Enhanced PGx-8 BeadChip is built on a high-density single nucleotide polymorphism (SNP) global backbone optimized for cross-population imputation genome coverage (Figure 1, Table 3). The combination of a high-density SNP backbone and clinical research variant coverage helps make the Infinium Global Diversity Array-8 v1.0 BeadChip the most cost-effective array within the Illumina portfolio. It is the array chosen by the All of Us Research Program that aims to sequence and genotype over 1 million individuals. The array is ideal for precision medicine programs interested in maximizing their return on genotyping investments.

Each Global Diversity Array with Enhanced PGx-8 v1.0 Kit includes BeadChips and reagents for amplifying, fragmenting, hybridizing, labeling, and detecting genetic variants using the high-throughput, streamlined Infinium workflow.

Table 3: Marker information

Marker categories			No. of markers
Exonic markers[a]			531,191
Intronic markers[a]			664,016
Promoters[a]			53,311
Nonsense markers[b]			28,224
Missense markers[b]			398,598
Synonymous markers[b]			34,000
Mitochondrial markers[b]			1318
Indels[c]			39,257
Sex chromosomes[c]	X	Y	PAR/homologous
	63,810	6215	5477

a. RefSeq, NCBI Reference Sequence Database.[19]
b. Compared against the UCSC Genome Browser.[6]
c. NCBI Genome Reference Consortium, Version GRCh37.[19]

indel, insertion/deletion; PAR, pseudoautosomal region; UCSC, University of California Santa Cruz; NCBI, National Center for Biotechnology Information.

Exceptional exonic content

The Infinium Global Diversity Array with Enhanced PGx-8 v1.0 BeadChip 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 database,[20] 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
NEF	346,340
EAS	146,281
AMR	272,178
AFR	257,690
SAS	224,431
NEF/EAS/AMR/AFR/SAS	69,432

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

NEF, non-Finish European; EAS, East Asian; AMR, ad mixed American; AFR, African; SAS, South Asian.

Broad coverage of variants with known disease associations

The Infinium Global Diversity Array with Enhanced PGx-8 v1.0 BeadChip provides coverage of variants selected from the National Human Genome Research Institute genome-wide association studies (NHGRI-GWAS) catalog[17] representing a broad range of phenotypes and disease classifications. This content provides powerful opportunities for researchers interested in studying diverse populations to test and validate associations previously found in European populations.

Clinical research content on the BeadChip enables validation of disease associations, risk profiling, preemptive screening research, and PGx studies. Variant selection includes a range of pathology classifications based on ClinVar and American College of Medical Genetics (ACMG) annotations.[5] The BeadChip contains extensive coverage of phenotypes and disease classifications based on ClinVar and the NHGRI-GWAS catalog (Figure 3). Markers cover ACMG and ClinVar database variants with a range of phenotypes pathogenic, likely pathogenic, and variants of unknown significance (VUS), as well as benign variants (Figure 4).

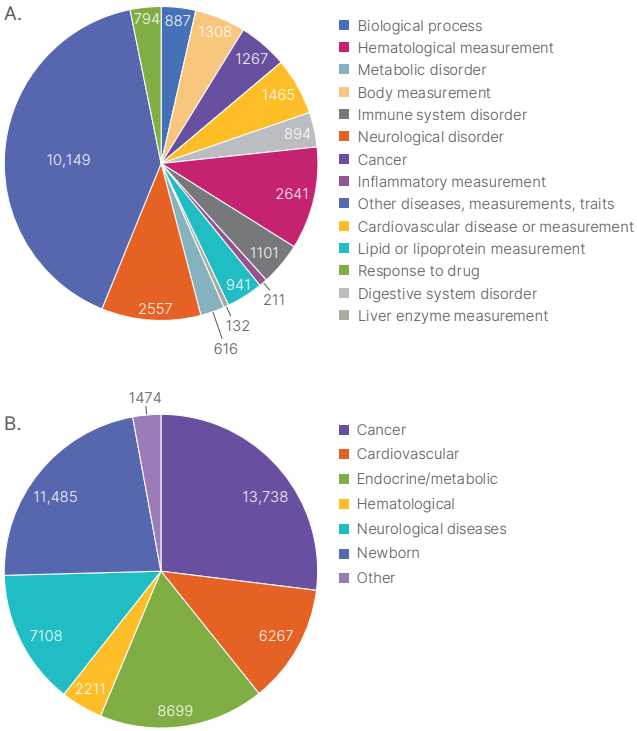

Figure 3: Disease research content covering diverse populations—The Global Diversity Array with Enhanced PGx-8 includes extensive coverage of numerous phenotypes and disease classifications based on (A) ClinVar categories and (B) NHGRI-GWAS categories.

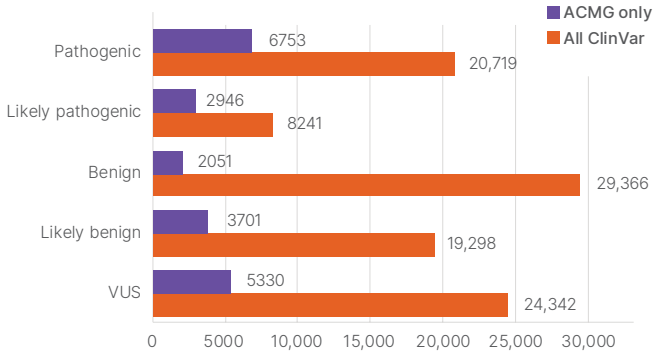

Figure 4: Distribution of variant pathology classifications according to ClinVar and ACMG annotations—Variants cover a range of pathogenic and nonpathogenic evidence.

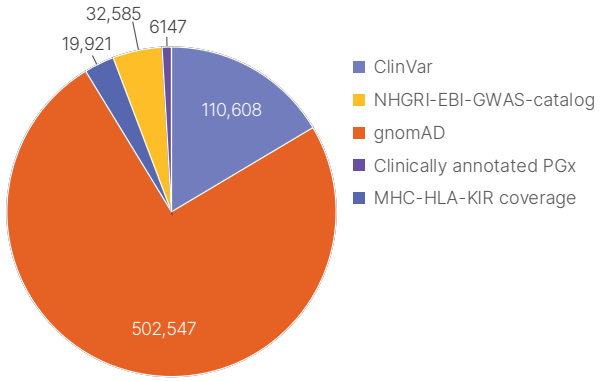

Figure 5: Clinical research content—The Infinium Global Diversity Array with Enhanced PGx-8 v1.0 BeadChip incorporates expertly selected clinical research content from key databases, supporting a broad range of applications.

Updated and relevant clinical research content

Clinical databases, such as ClinVar, are constantly evolving as new variants are added and variants change designation to "pathogenic" or "likely pathogenic." The Infinium Global Diversity Array with Enhanced PGx-8 v1.0 BeadChip provides updated coverage of many of the high-value variants contained within these annotated databases. Variants included on the array consist of markers with known disease association based on ClinVar, the PharmGKB, and the National Human Genome Research Institute (NHGRI)-EBI database.[17] The BeadChip also provides imputation-based tagSNPs for HLA alleles, extended MHC region, the KIR gene, and exonic content from the gnomAD[13] database (Table 2, Figure 5).

QC markers for sample identification

The Infinium Global Diversity Array with Enhanced PGx-8 v1.0 BeadChip includes quality control (QC) markers for large-scale studies, enabling sample identification, tracking, ancestry determination, stratification, and more (Figure 6).

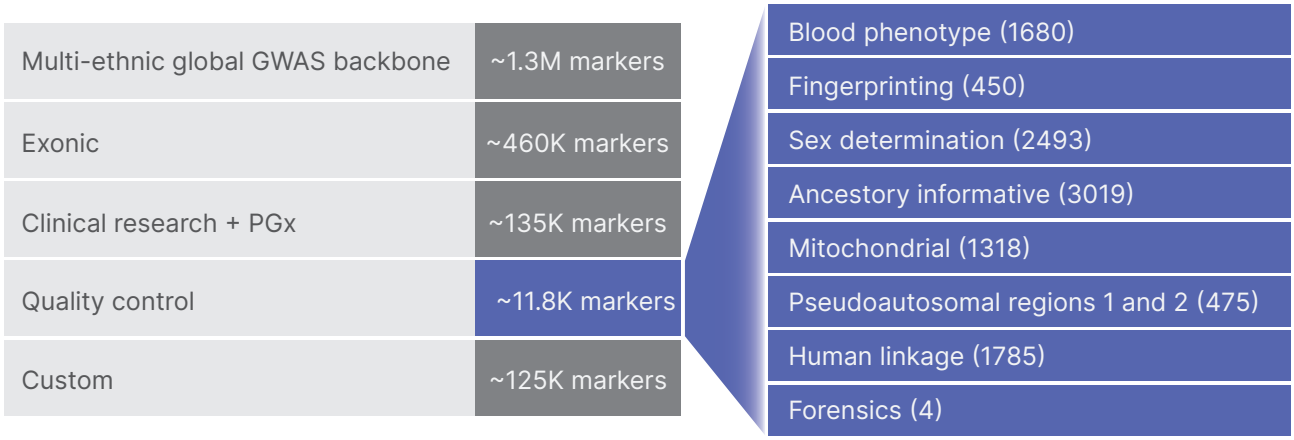

Figure 6: QC markers—QC variants on the array enable various capabilities for sample tracking such as sex determination, continental ancestry, human identification, and more.

Flexible content options

The Infinium Global Diversity Array with Enhanced PGx-8 v1.0 BeadChip can be customized to incorporate up to 125K custom bead types. The DesignStudio™ Microarray Assay Designer can be used to design targets such as SNPs, copy number variants (CNVs), and indels.

High-throughput workflow

The Infinium Global Diversity Array with Enhanced PGx-8 v1.0 BeadChip uses the proven Infinium 8-sample BeadChip format to enable laboratories to scale efficiently. For flexible throughput processing, the Infinium 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 users to gather and report data quickly (Figure 7).

Trusted high-quality assay

The Infinium Global Diversity Array with Enhanced PGx-8 v1.0 BeadChip uses trusted Infinium assay chemistry to deliver the 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 CNV calling.

Table 5: Data performance and spacing

Data performance	Value[a]	Product specification[b]
Call rate	99.7%	> 99.0% Avg
Reproducibility	99.99%	> 99.90%
Spacing		

Spacing (kb)	Mean	Median	90th%[c]
	1.53	0.61	3.91

a. Values are derived from genotyping 2228 HapMap reference samples
b. Excludes Y chromosome markers for female samples
c. Based on results from GenTrain sample set

Accurate and efficient secondary analysis

DRAGEN Array secondary analysis software is recommended for analysis of the Infinium Global Diversity Array with Enhanced PGx-8 v1.0 BeadChip. DRAGEN Array is a powerful bioinformatics software that uses cutting-edge data analysis tools to maximize genomic insights. The software includes SNP analysis, PGx star allele and variant coverage across 1700+ targets for over 50 genes, hybrid and allele-specific copy number detection, PGx CNV coverage on six target genes across nine target regions, and more.

DRAGEN Array secondary analysis is able to generate accurate results in multiple output file formats for easy downstream analysis, including the capability to generate VCF files from Infinium array-based assays in as little as 35 seconds per sample and full PGx analysis results in approximately one minute per sample.[21]

Figure 7: The Infinium eight-sample format workflow—The Infinium workflow provides a rapid three-day workflow with minimal hands-on time.

DRAGEN Array has two deployment options. A local analysis option provides a command-line interface for granular control. No specialized DRAGEN server or FPGA hardware is required for the local installation solution. A cloud-based package with an intuitive graphical user interface is also available with the user-friendly BaseSpace™ Sequence Hub. The cloud-based option offers easy access and additional functionality, such as polygenic risk scoring for arrays.

Summary

The high-density Infinium Global Diversity Array with Enhanced PGx-8 v1.0 BeadChip (Figure 8) provides a cost-effective solution for population-scale genetic studies, variant screening, and precision medicine research. The iScan System, Infinium assay technology, and DRAGEN Array secondary analysis software work together to create a versatile and comprehensive genotyping solution.

Figure 8: Infinium Global Diversity Array with Enhanced PGx-8 v1.0 BeadChip—Built on the trusted eight-sample Infinium platform.

Learn more

The Infinium Global Diversity Array with Enhanced PGx-8 v1.0 BeadChip

Pharmacogenomics

DRAGEN Array secondary analysis

Ordering information

Product	Catalog no.
Infinium Global Diversity Array with Enhanced PGx-8 v1.0 kit (48 samples)	20044822
Infinium Global Diversity Array with Enhanced PGx-8 v1.0 kit (384 samples)	20044823
Infinium Global Diversity Array with Enhanced PGx-8+ v1.0 kit (48 samples, with add-on content)	20048347
Infinium Global Diversity Array with Enhanced PGx-8+ v1.0 kit (384 samples, with add-on content)	20048348
Product	**Catalog no.**
DRAGEN Array local - star allele annotation (1 sample)	20109885
DRAGEN Array cloud - star allele annotation (1 sample)[a]	20109886

a. An Illumina Connected Analytics annual subscription is required for cloud analysis along with iCredits for data storage and analysis

References

1. Whirl-Carrillo M, Huddart R, Gong L, et al. An Evidence-Based Framework for Evaluating Pharmacogenomics Knowledge for Personalized Medicine. *Clin Pharmacol Ther.* 2021;110(3):563-572. doi:10.1002/cpt.2350
2. Relling MV, Klein TE. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. *Clin Pharmacol Ther.* 2011;89(3):464-467. doi:10.1038/clpt.2010.279
3. Fairley S, Lowy-Gallego E, Perry E, Flicek P. The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Res.* 2020;48(D1):D941-D947. doi:10.1093/nar/gkz836
4. Gaedigk A, Casey ST, Whirl-Carrillo M, Miller NA, Klein TE. Pharmacogene Variation Consortium: A Global Resource and Repository for Pharmacogene Variation. *Clin Pharmacol Ther.* 2021;110(3):542-545. doi:10.1002/cpt.2321
5. Green RC, Berg JS, Grody WW, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing [published correction appears in *Genet Med.* 2017;19(5):606]. *Genet Med.* 2013;15(7):565-574. doi:10.1038/gim.2013.73
6. Navarro Gonzalez J, Zweig AS, Speir ML, et al. The UCSC Genome Browser database: 2021 update. *Nucleic Acids Res.* 2021;49(D1):D1046-D1057. doi:10.1093/nar/gkaa1070
7. NCBI. ClinVar Database website. ncbi.nlm.nih.gov/clinvar. Accessed November 7, 2023.
8. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* 2019;47(D1):D941-D947. doi:10.1093/nar/gky1015
9. Gene Ontology Consortium, Aleksander SA, Balhoff J, et al. The Gene Ontology knowledgebase in 2023. *Genetics.* 2023;224(1):iyad031. doi:10.1093/genetics/iyad031
10. MacDonald JR, Ziman R, Yuen RK, Feuk L, Scherer SW. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res.* 2014;42(Database issue):D986-D992. doi:10.1093/nar/gkt958
11. Wong KM, Langlais K, Tobias GS, et al. The dbGaP data browser: a new tool for browsing dbGaP controlled-access genomic data. *Nucleic Acids Res.* 2017;45(D1):D819-D826. doi:10.1093/nar/gkw1139
12. Rajeevan H, Osier MV, Cheung KH, et al. ALFRED: the ALelle FREquency Database. Update. *Nucleic Acids Res.* 2003;31(1):270-271. doi:10.1093/nar/gkg043
13. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans [published correction appears in Nature. 2021 Feb;590(7846):E53] [published correction appears in Nature. 2021 Sep;597(7874):E3-E4]. *Nature.* 2020;581(7809):434-443. doi:10.1038/s41586-020-2308-7

14. de Bakker PI, 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(10):1166-1172. doi:10.1038/ng1885
15. Prüfer K, Racimo F, Patterson N, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature.* 2014;505(7481):43-49. doi:10.1038/nature12886
16. Illumina. TruSight Inherited Disease Sequencing Panel Data Sheet. http://illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_trusight_inherited_disease.pdf. Accessed July 2016.
17. NHGRI. National Human Genome Research Institute website. http://genome.gov. Accessed January 23, 2021.
18. NCBI. Reference Sequence Database website. http://ncbi.nlm.nih.gov/refseq. Accessed November 9, 2023.
19. Genome Reference Consortium. Human Genome Overview Version GRCh37 website. http://ncbi.nlm.nih.gov/grc/human. Accessed November 9, 2023.
20. Karczewski KJ, Weisburd B, Thomas B, et al. The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Res.* 2017;45(D1):D840-D845. doi:10.1093/nar/gkw971
21. Illumina data on file, November 7 2023.

illumina®

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

Pub. no. M-GL-00031 v3.0