DATA SHEET

Infinium™ Global Diversity Array-8 v1.0

A powerful, cost-effective array with optimized multi-ethnic, genome-wide content

- Chosen by the *All of Us* precision medicine genomic initiative to genotype 1 million+ people
- Updated coverage of clinical research variants for a broad range of applications
- Optimized, multi-ethnic content meets the need for diversity in genomics studies
- High-quality, reproducible data using trusted Infinium chemistry with a scalable workflow

illumina®

Introduction

The 8-sample Infinium Global Diversity Array-8 v1.0 BeadChip (Figure 1) (Table 1) features up-to-date coverage of clinical research variants associated with disease and pharmacogenomics, and exome content representing diverse populations (Table 2) (Table 3). The Infinium Global Diversity Array-8 v1.0 BeadChip is built on a high-density single nucleotide polymorphism (SNP) global backbone optimized for crosspopulation imputation coverage of the genome (Figure 2). It enables polygenic risk score development and characterization of genetic architecture in diverse populations.

Table 1: Product information[a]

Feature	Description
Species	Human
Total number of markers[b]	1,831,442
Capacity for custom bead types	175,000
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 may vary depending on laboratory and system configurations
b. Variants found on commercial manifest

The combination of a high-density SNP backbone and updated, relevant clinical research variant coverage provides exceptional value per genotype by delivering insights for both discovery and screening applications. The Infinium Global Diversity Array-8 v1.0 BeadChip provides the most cost-effective per variant coverage within the Illumina human array portfolio. It is ideal for precision medicine programs interested in maximizing their return on genotyping investments. Each Global Diversity Array Kit includes convenient packaging containing BeadChips and

Figure 1: Infinium Global Diversity Array-8 v1.0 BeadChip is built on the trusted 8-sample Infinium platform.

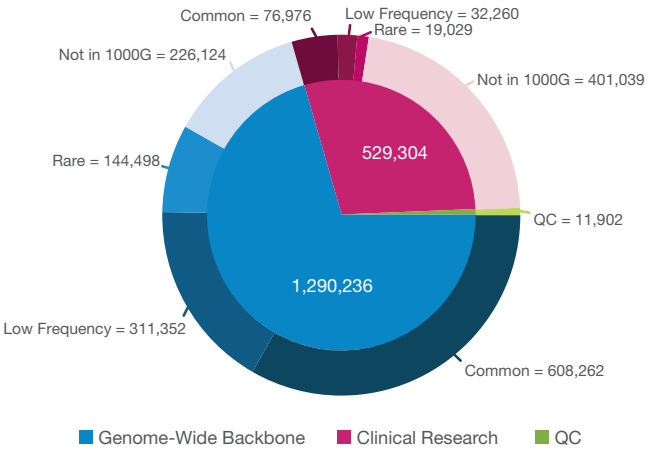

Figure 2: Summary of content—Plotted in the inner pie is the proportion of the array selected for genome-wide coverage, clinical research, and quality control (QC). The outer ring summarizes the weighted reference global allele frequency for unique variants present in the 1000 Genomes Project (1000G).[1] Variants not in 1000G are labeled. Counts represent unique variants.

reagents for amplifying, fragmenting, hybridizing, labeling, and detecting genetic variants using the high-throughput, streamlined Infinium workflow.

Table 2: High-value content

Content	No. of markers[a]	Research application/note	Content	No. of markers	Research application/note
ACMG[2] 59 2016 gene coverage	51,899	Variants with known clinical significance identified from clinical WGS and WES samples	GO[9] CVS genes	318,902	Cardiovascular conditions
ACMG 59 all annotations	25,751		Database of Genomic Variants[10]	1,501,083	Genomic structural variation
ACMG 59 pathogenic	8319		eQTLs[11]	6913	Genomic loci regulating mRNA expression levels
ACMG 59 likely pathogenic	3154		Fingerprint SNPs[12]	780	Human identification
ACMG 59 benign	2178		gnomAD[13] exome	427,536	WES and WGS results from unrelated individuals from various studies
ACMG 59 likely benign	4366		HLA genes[14]	1237	Disease defense, transplant rejection, and autoimmune disorders
ACMG 59 VUS	6077		Extended MHC[14c]	22,089	Disease defense, transplant rejection, and autoimmune disorders
ADME[3] core and extended + CPIC genes	31,176	Drug absorption, distribution, metabolism, and excretion	KIR genes[4]	167	Autoimmune disorders and disease defense
ADME core and extended + CPIC genes +/- 10 kb	37,362	Includes regulatory regions	Neanderthal SNPs[15]	4327	Neanderthal ancestry and human population migration
AIMs[b]	3672	Ancestry-informative markers	Newborn/carrier screening gene coverage	70,698	Genes associated childhood diseases included in the TruSight™ Inherited Disease Sequencing Panel[19]
APOE[4]	102	Cardiovascular disease, Alzheimer's disease, and cognition	NHGRI-EBI GWAS catalog[16]	28,652	Markers from published GWAS
Blood phenotype genes[5]	2928	Blood phenotypes	PharmGKB[17,18] all	4360	Human genetic variation associated with drug responses
ClinVar[6] variants	113,679	Relationships among variation, phenotypes, and human health	PharmGKB level 1A	32	
ClinVar pathogenic	28,821		PharmGKB level 1B	2	
ClinVar likely pathogenic	10,869		PharmGKB level 2A	19	
ClinVar benign	32,355		PharmGKB level 2B	64	
ClinVar likely benign	24,100		PharmGKB level 3	1342	
ClinVar VUS	26,731		PharmGKB level 4	170	
COSMIC[7] genes	1,039,798	Somatic mutations in cancer	RefSeq[20] 3' UTRs	53,278	3' untranslated regions[d]
CPIC[8] all	241	Variants with potential guidelines to optimize drug therapy	RefSeq 5' UTRs	33,738	5' untranslated regions[d]
CPIC-A			RefSeq All UTRs	84,474	Untranslated regions[d]
CPIC-A/B	138		RefSeq	1,137,129	All known genes
CPIC-B	18		RefSeq +/- 10 kb	1,272,757	Regulatory regions[d]
CPIC-C	14		RefSeq Promoters	46,363	2 kb upstream to include promoter regions[d]
CPIC-C/D	103		RefSeq Splice Regions	19,120	Variants at splice sites[d]
CPIC-D	71				

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

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 Knowledge-base; RefSeq: NCBI Reference Sequence Database; UTR: untranslated region; VUS, variant of unknown significance; WES, whole-exome sequencing; WGS, whole-genome sequencing

Chosen by the *All of Us* research program

The Infinium Global Diversity Array-8 v1.0 BeadChip is the array chosen by the *All of Us* Research Program. This program will engage one million or more volunteers living in the United States to contribute their health data over many years to improve health outcomes, fuel the development of new treatments for disease, and catalyze a new era of evidence-based and more precise preventive care and medical treatment. Moreover, one of the program's core values guiding development and implementation is for participants to reflect the rich diversity of the US. The Infinium Global Diversity Array-8 v1.0 BeadChip was built to meet these needs by combining highly optimized multi-ethnic, genome-wide content with curated clinical research variants.

Table 3: Marker information

Marker category			No. of markers
Exonic markers[a]			538,230
Intronic markers[a]			678,231
Nonsense markers[b]			29,227
Missense markers[b]			348,902
Synonymous markers[b]			39,979
MItochondrial markers[b]			1397
Indels[c]			44,172
Sex chromosomes	X	Y	PAR/homologous
	62,617	6478	5497

a. RefSeq - NCBI Reference Sequence Database.[20] Accessed April 2021.

b. Compared against the UCSC Genome Browser.[4] Accessed April 2021.

c. NCBI Genome Reference Consortium, Version GRCh37.[21] Accessed April 2021.
Abbreviations: indel, insertion/deletion; PAR, pseudoautosomal region.

Built through collaboration with leading institutions

The Infinium Global Diversity Array-8 v1.0 BeadChip uses content from the Infinium Multi-Ethnic Global-8 v1.0 BeadChip, a widely used array with adoption by major biobanks. The Infinium Global Diversity Array-8 v1.0 BeadChip contains a robust genome-wide scaffold designed to tag both common and low frequency variants in global populations (minor allele frequency (MAF) > 1%). This scaffold was designed through collaborations with the Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA) and Population Architecture using Genomics and Epidemiology (PAGE).

The Infinium Global Diversity Array-8 v1.0 BeadChip draws from whole-genome sequences not found in 1000G. The array's design leverages more than 1000 whole-genome sequences of African ancestry and populations throughout the Americas, including the US, Caribbean, and Latin and South America.

Exceptional coverage of exonic content

The Infinium Global Diversity Array-8 v1.0 BeadChip includes enhanced tagging in exonic regions and enriched coverage to map GWAS loci with previously identified disease or trait associations with precision. More than 400,000 markers of exome content were gathered from 36,000 individuals of diverse ethnic groups, including African Americans, Hispanics, Pacific Islanders, East Asians, and individuals of mixed ancestry. The Global Diversity Array also features diverse exonic content from the ExAC database,[22] including both cross-population and population-specific markers with either functionality or strong evidence for association (Table 4).

Table 4: Exonic coverage across populations

Population	No. of markers
EUR	305,380
EAS	132,257
AMR	254,594
AFR	241,679
SAS	206,832
EUR/EAS/AMR/AFR/SAS	61,896

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

Exceptional coverage of variants with known disease associations

The Infinium Global Diversity Array-8 v1.0 BeadChip provides coverage of variants selected from the NHGRI-GWAS catalog, representing a broad range of phenotypes and disease classifications (Figure 3). This content provides a powerful opportunity for researchers interested in studying diverse populations to test and validate associations previously found in European populations.

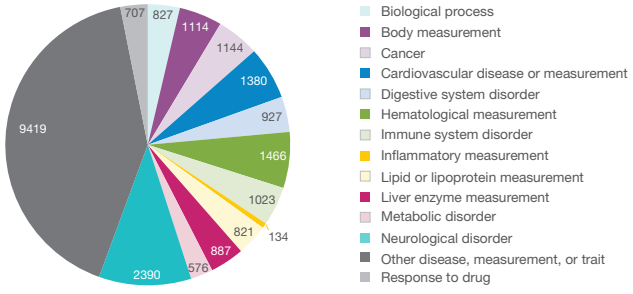

Figure 3: NHGRI disease categories—Global Diversity Array clinical research content features markers across a broad range of disease coategories based on the NHGRI database.

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-8 v1.0 BeadChip provides updated coverage of many of these high value variants contained within annotated databases. Variants included on the array consist of markers with known disease association based on ClinVar, the Pharmacogenomics Knowledgebase (PharmGKB), and the National Human Genome Research Institute (NHGRI)-EBI database (Figure 4). The Infinium Global Diversity Array-8 v1.0 BeadChip also provides imputation-based tagSNPs for HLA alleles, extended MHC region, the KIR gene, and exonic content from the gnomAD database[13] (Table 2).

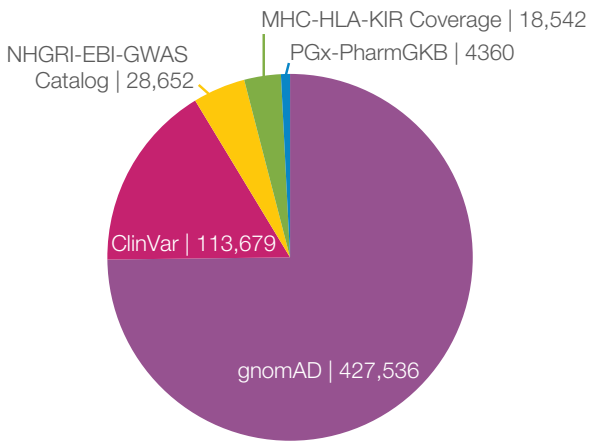

Figure 4: Clinical research content—Content was expertly selected from scientifically recognized databases to create a highly informative array for clinical research applications. Variant counts may be subject to change.

Broad spectrum of pharmacogenomics markers

The Infinium Global Diversity Array-8 v1.0 BeadChip provides coverage of pharmacogenomics variants associated with absorption, distribution, metabolism, and excretion (ADME) phenotypes based on PharmGKB[17] and Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines[8] (Figure 5).

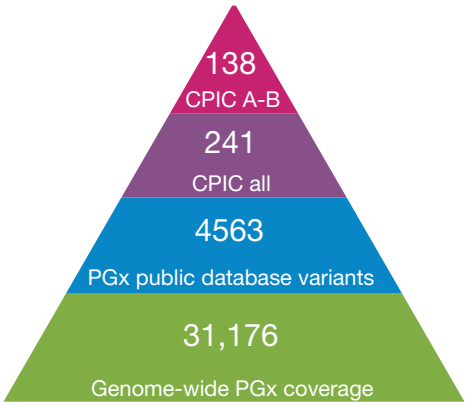

Figure 5: Broad spectrum of pharmacogenomics markers—Clinical research content features an extensive list of pharmacogenomics markers selected based on CPIC guidelines and the PharmGKB database.[16] PGx public database variants, variants annotated in PharmGKB, PharmVar, CPIC; Genome-wide PGx coverage, includes markers located in an extended ADME genes or CPIC level A genes including targeted imputation tag SNPs and CPIC level A copy number variation (CNV) tags.

Extensive coverage of diseases

Clinical research content on the Infinium Global Diversity Array-8 v1.0 BeadChip enables validation of disease associations, risk profiling, preemptive screening research, and pharmacogenomics studies. Variant selection includes a range of pathology classifications based on ClinVar and American College of Medical Genetics (ACMG) annotations[2] (Figure 6A). The BeadChip contains extensive coverage of phenotypes and disease classifications based on ClinVar and the NHGRI-GWAS catalog (Figure 6B).

QC markers for sample tracking

The Infinium Global Diversity Array-8 v1.0 BeadChip includes QC markers for large-scale studies, enabling sample identification, tracking, ancestry determination, and stratification (Figure 7).

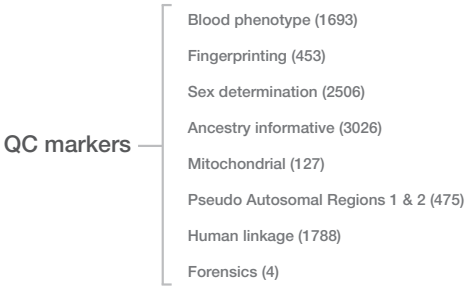

Figure 7: QC markers—QC variants on the array enable various capabilities for sample tracking such as sex determination, continental ancestry, and human identification and more.

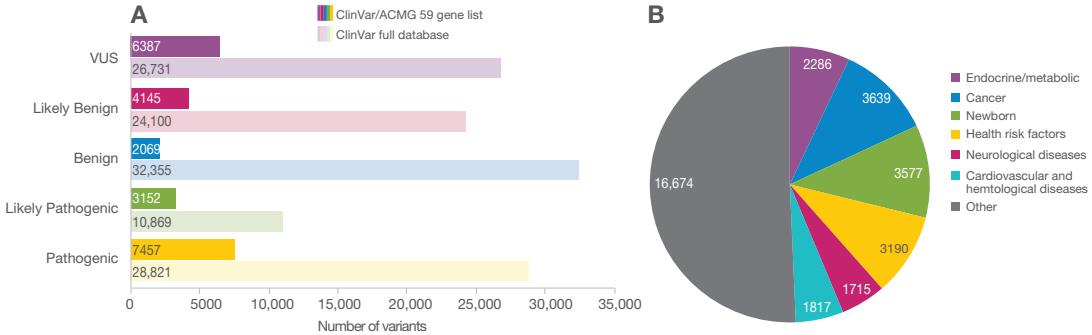

Figure 6: Broad coverage of disease categories—(A) Variants sorted by range of pathology classifications according to ClinVar American College of Medical Genetics (ACMG) annotations. (B) Global Diversity Array clinical research content by category within the ClinVar database. Variant counts may be subject to change.

High imputation performance across ancestries

High imputation accuracy provides increased power to support population-scale disease research and population-specific causal variant detection. Leading disease research consortia involved in the development of the Infinium Global Diversity Array-8 v1.0 BeadChip included population-specific and transethnic tag SNPs to maximize imputation performance, enabling more effective association studies in diverse populations (Table 5, Table 6). The Global Diversity Array backbone maximizes the amount of high quality, valuable information that can be extracted per genotyped sample.

Imputation calculation methodology

Imputation performance is measured by simulating Global Diversity Array-8 genotyped variants on 1000G samples (Table 5, Table 6). A random sample from all 26 global populations of the 1000G were selected, stratified by super population, and variants on the Global Diversity Array-8 were tested. The remaining 1000G samples were treated as the reference (1000G data is already phased using BEAGLE). Minimac3 was used to perform imputation and imputation quality was measured using the correlation r2 from the info file produced by minimac3.

Table 5: Imputation accuracy from 1000G at various MAF thresholds

Population[b]	Imputation accuracy[a]		
	MAF ≥ 5%	MAF ≥ 1%	MAF 1-5%
AFR	0.96	0.93	0.90
AMR	0.95	0.88	0.75
EAS	0.92	0.88	0.77
EUR	0.94	0.91	0.81
SAS	0.96	0.92	0.82

a. Compared against Phase 3, version 5 of the 1000G. www. www.internationalgenome.org. Accessed April 2021. Imputed using minimac3.
b. www.internationalgenome.org/category/population

Table 6: Number of markers imputed at r2 ≥ 0.80 from 1000G[a]

Population[b]	No. of imputated markers		
	MAF ≥ 5%	MAF ≥ 1%	MAF 1–5%
AFR	17,904,224	30,223,608	12,319,384
AMR	13,250,116	18,866,180	5,616,064
EAS	11,064,504	14,116,088	3,051,584
EUR	12,605,568	16,854,634	4,249,066
SAS	13,244,890	18,009,596	4,764,706

a. Compared against Phase 3, version 5 of the 1000G. www. www.internationalgenome.org. Accessed April 2021. Imputed using minimac3.
b. www.internationalgenome.org/category/population

High-throughput workflow

The Infinium Global Diversity Array-8 v1.0 BeadChip uses the proven Infinium 8-sample format that enables laboratories to efficiently scale as needed. 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 8).

For labs interested in quickly scaling or increasing efficiency and operational excellence, the Illumina ArrayLab Consulting Service offers customized solutions.

Trusted, high-quality assay

The Infinium Global Diversity Array-8 v1.0 BeadChip uses trusted Infinium assay chemistry to deliver the same high-quality, reproducible data (Table 7) that Illumina genotyping arrays have provided for over a decade. It is compatible with the Infinium FFPE QC and DNA Restoration Kits,[23] 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 access to genome-wide copy CNV calling.

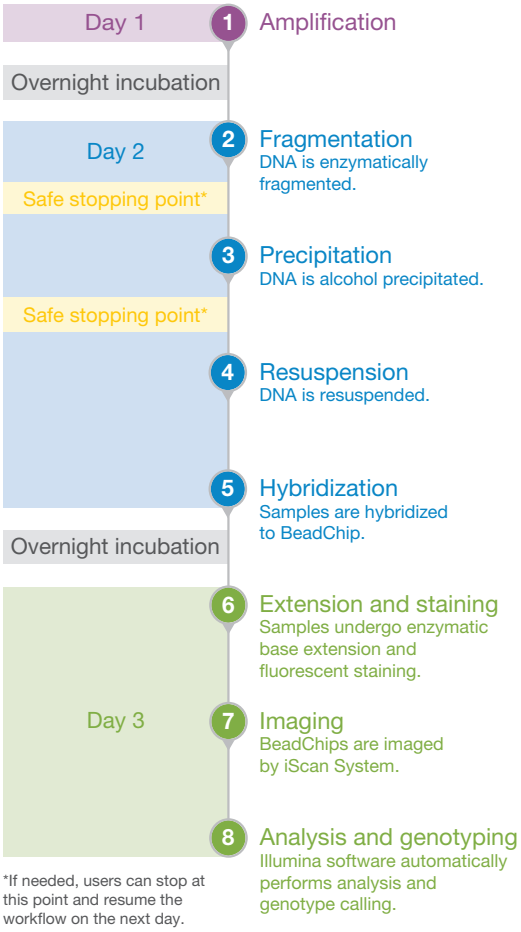

Figure 8: The Infinium workflow provides a rapid three-day workflow with minimal hands-on time.

Table 7: Data performance and spacing

Data performance	Value	Product specification
Call rate	99.7%	> 99.0% Avg
Reproducibility	99.99%	> 99.90%
Log R deviation	0.12	< 0.30 Avg
Spacing		
Spacing (kb)	Mean	Median	90th%
	1.5	0.63	4.0

a. Values are derived from genotyping 2051 HapMap reference samples.
b. Excudes Y chromosome markers for female samples.
c. Based on results from GenTrain sample set.
d. Value expected for typical projects using standard Illumina protocols. Tumor samples and samples prepared by nonstandard protocols are excluded

Summary

Using the iScan System, Infinium assay, and integrated analysis software, the high-density Infinium Global Diversity Array-8 v1.0 BeadChip provides a cost-effective solution for population-scale genetic studies, variant screening, and precision medicine research.

Ordering information

Infinium Global Diversity Array-8 v1.0 BeadChip Kit	Catalog no.
16 samples	20031669
48 samples	20031810
96 samples	20031811
384 samples	20031812
Infinium Global Diversity Array-8+ v1.0 BeadChip Kit[a]	**Catalog no.**
16 samples	20031813
48 samples	20031814
96 samples	20031815
384 samples	20031816

e. Enabled for custom content

Learn more

Infinium Global Diversity Array-8 v1.0 BeadChip and other Illumina genotyping products, illumina.com/techniques/microarrays.html

References

1. The 1000 Genomes Project. 1000genomes.org. Accessed April 7, 2021.
2. ACMG Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing. ncbi.nlm.nih.gov/clinvar/docs/acmg/. Accessed April 7, 2021.
3. PharmaADME Gene List. pharmaadme.org. Accessed April 7, 2021.
4. University of California, Santa Cruz (UCSC) Genome Browser. genome.ucsc.edu. Accessed April 7, 2021.
5. NCBI Reference Sequence Blood Group Antigen Gene Mutation Database. ncbi.nlm.nih.gov/projects/gv/rbc/xslcgi.fcgi?cmd=bgmut/systems. Accessed April 7, 2021.
6. ClinVar Database. ncbi.nlm.nih.gov/clinvar. Accessed April 7, 2021.
7. Catalog of somatic mutations in cancer. cancer.sanger.uk/cosmic. Accessed April 7, 2021.
8. Clinical Pharmacogenetics Implementation Consortium (CPIC). cpicpgx.org. Accessed April 7, 2021.
9. Gene Ontology Consortium. geneontology.org. Accessed April 7, 2021.
10. Database of Genomic Variants. dgv.tcag.ca/dgv/app/home. Accessed April 7, 2021.
11. NCBI eQTL Database. ncbi.nlm.nih.gov/projects/gap/eqtl/index.cgi. Accessed April 7, 2021.
12. The Allele Frequency Database. alfred.med.yale.edu/alfred/snpSets.asp. Accessed April 7, 2021.
13. gnomAD, Genome Aggregation Database. gnomad.broadinstitute.org. Accessed April 7, 2021.
14. 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.
15. Neanderthal Genome Browser. neandertal.ensemblgenomes.org/index.html. Accessed April 7, 2021.
16. National Human Genome Research Institute. genome.gov/. Accessed April 7, 2021.
17. PharmGKB, The Pharmacogenomics Knowledgebase. pharmgkb.org. Accessed April 7, 2021.
18. PharmGKB, Clinical Annotation Levels of Evidence. pharmgkb.org/page/clinAnnLevels. Accessed April 7, 2021.
19. Illumina (2017). TruSight Inherited Disease Sequencing Panel Data Sheet. Accessed April 7, 2021.
20. RefSeq - NCBI Reference Sequence Database. ncbi.nlm.nih.gov/refseq. Accessed April 7, 2021.
21. NCBI Genome Reference Consortium. Version GRCh37. ncbi.nlm.nih.gov/grc/human. Accessed April 7, 2021.
22. Exome Aggregation Consortium (ExAC) Browser. exac.broadinstitute.org. Accessed April 7, 2021.
23. Infinium FFPE QC and DNA Restoration Kit. www.illumina.com/content/dam/illuminamarketing/ documents/products/datasheets/datasheet_FFPE_DNA_restoration.pdf.

illumina®

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

M-GL-00153 v1.0