Genome-Wide DNA Analysis BeadChips

Offering a combination of powerful content and unprecedented flexibility for experimental design.

Introduction

Whole-genome genotyping (WGGT) microarrays have been used successfully for almost a decade to identify regions of the human genome that contribute to disease susceptibility and phenotype traits. The two primary applications of these tools, genome-wide association studies (GWAS) and copy number variation (CNV) identification, have enabled researchers to achieve a greater understanding of how genetics contributes to human health and disease progression.

In just a few short years, the research community has identified thousands of trait- and disease-associated loci, and published hundreds of peer-reviewed papers using this technology. Researchers identify causative alleles by genotyping a large number of DNA samples, testing for allele frequency shifts between affected and unaffected individuals. The power to detect these associated alleles increases with the addition of new and more complete data gathered from next-generation sequencing studies. By leveraging this new content, Illumina Omni family of microarrays provides unprecedented coverage of the human genome, empowering researchers to evaluate the role of common and rare variants like never before. Much of the selected content has been derived from the 1000 Genomes Project (1kGP), a collaborative sequencing effort by research institutions around the world that seeks to identify human genetic variants occurring at any appreciable frequency across diverse populations.

Omni Family of Microarrays

The Omni family of microarrays offers a series of complementary and additive array options for a range of study objectives (Figure 1). Each primary array features a unique set of markers designed to target a specific minor allele frequency (MAF) range (Table 1). Supplemental arrays allow investigators to add additional rarer content to their studies as their research objectives evolve (Table 2).

From the flagship HumanOmni5 (Omni5) array targeting variants down to 1% MAF, to the 12-sample HumanOmniExpress (OmniExpress) array, the Omni family of arrays provides the flexibility to meet a variety research goals and budget needs. Semi-custom options allow researchers to increase the power of their studies by tailoring the arrays with novel variants from their own sequencing studies. For follow-up studies, fully custom iSelect® BeadChips can be easily developed with up to 1 million markers targeting any loci across the genome.

The Omni5

Get it all in one experiment—the Omni5 BeadChip delivers the complete set of Omni family markers on one array, plus the ability to add 500K custom markers. Content includes powerful tagSNPs selected from the International HapMap and 1000 Genomes Projects that target genetic variation down to 1% MAF, along with focused, high-value content covering important regions of the genome such as the MHC, ADME genes, and nsSNPs. The option to add 500K

Figure 1: Omni Family of Microarrays

OmniExpress	Omni1S	Omni2.5-8	Omni2.5S-8	Omni5
Highest throughput, exceptional price, common variation coverage down to 5% MAF.	Supplementary array, rare variation coverage down to 2.5% MAF.	Comprehensive common and rare variation coverage down to 2.5% MAF from the 1kGP.	Supplementary array, rare variation coverage down to 1% MAF.	Near complete common and rare variation coverage down to 1% MAF from the 1kGP.

Omni arrays provide flexibility for timing and budget to help investigators effectively achieve their research goals.

Table 1: Omni BeadChip Performance Parameters

	OmniExpress	Omni2.5	Omni5
Number of Fixed Markers	730,525	2,379,855	4,301,331
Available Custom Markers	up to 200,000	n/a	up to 500,000
Number of Samples	12	8	4
DNA Requirement	200 ng	200 ng	400 ng
Assay	Infinium HD	Infinium LCG	Infinium LCG
Instrument Support	HiScan or iScan	HiScan or iScan	HiScan or iScan
Sample Throughput*	> 1,400 / week	~1,067 samples / week	> 460 samples / week
Scan Time / Sample	5 minutes	6.5 minutes (HiScan) 11.4 minutes (iScan)	15 minutes (HiScan) 25 minutes (iScan)

% Variation Captured (r^2 > 0.8)	1kGP† MAF > 5%	1kGP† MAF > 1%	1kGP† MAF > 5%	1kGP† MAF > 1%	1kGP† MAF > 5%	1kGP† MAF > 1%
CEU	0.73	0.58	0.83	0.73	0.87	0.83
CHB + JPT	0.74	0.62	0.83	0.73	0.85	0.76
YRI	0.40	0.25	0.65	0.51	0.71	0.58

Data Performance	Value¨ / Product Specification		
Call Rate (average)	99.84% / > 99%	99.65% / > 99%	99.9% / > 99%
Reproducibility	99.99% / > 99.9%	99.99% / > 99.9%	99.99% / > 99.9%
Log R Dev	0.15 / < 0.30‡	0.12 / < 0.30‡	0.12 / < 0.30**

Spacing	Mean / Median / 90th%		
Spacing (kb)	4.0 / 2.1 / 9.3	1.19 / 0.64 / 2.76	0.68 / 0.36 / 1.57

Marker Categories	Number of Markers		
Number of SNPs with 10kb of RefSeq genes	392,197	1,231,382	2,311,849
Nonsynonymous SNPs (NCBI annotated)	15,062	41,900	84,004
MHC / ADME	7,459 / 16,649	19,238 / 27,335	43,904 / 43,615
Sex Chromosome (X / Y / PAR Loci)	18,055 / 1,409 / 471	55,208 / 2,561 / 418	113,213 / 2,498 / 511
Mitochondrial	0	256	267

* Estimate assumes one iScan system, one AutoLoader2, one Tecan Robot, and a five-day work week.

† Compared against June 2011 1kGP data release.

¨ Values are derived from reference samples.

‡ Value expected for typical projects, excluding tumor samples or any samples prepared not following standard Illumina protocols.

custom markers allows researchers to tailor the BeadChip for targeted applications and population-focused or disease-specific studies. With an average marker spacing of only 680 base pairs, the Omni5 is the industry leader for genotyping and CNV detection.

The Omni2.5

The HumanOmni2.5-8 BeadChip (Omni2.5) features ~2.3 million markers that capture genomic variation down to 2.5% MAF. Optimized tag SNP content and dense marker spacing (mean spacing = 1.2 kb) enables a broad range of study types, including high resolution for CNV and other structural variation applications.

Researchers starting their studies with the Omni2.5 can supplement their data with the HumanOmni2.5S (Omni2.5S), which provides a unique set of ~2 million markers derived from the 1kGP, including coverage of rare variants down to 1% MAF and the ability to customize with up to 500K markers.

The OmniExpress

The HumanOmniExpress BeadChip (OmniExpress) delivers excellent power for common-variant GWAS, providing high sample throughput at the industry's best price. This 12-sample BeadChip is the ideal solution for processing the greatest number of samples within a given budget. Optimized tag SNP content from all three phases of the

HapMap project has been strategically selected to capture the greatest amount of common SNP variation (> 5% MAF). For researchers that need a more customized solution, up to 200,000 markers can be added with OmniExpress+ BeadChip. This option provides the same base content as the OmniExpress BeadChips, but allows researchers to include selected markers unique for their study.

Researchers starting with either the OmniExpress or OmniExpress+ array can add an additional ~1.2 million markers with the HumanOmni1S (Omni1S). The content on the HumanOmni1S is derived from the pilot releases of the 1kGP, providing high coverage of low frequency alleles down to ~2.5% MAF.

Customized Follow-Up For Targeted Studies

Customized iSelect BeadChips can be easily developed to fit any experimental design, allowing researchers to develop an ideal selection of markers for any budget and throughput requirement. The Illumina iSelect custom genotyping platform offers all of the benefits of standard Infinium® products, including industry-leading data quality and call rates, streamlined workflow, and informed SNP selection, with the flexibility to access virtually the entire genome.

Custom products can be deployed on either the 4-sample (200,001 to 1,000,000 attempted bead types), 12-sample (60,801 to 200,000 attempted bead types), or 24-sample (3,000 to 68,000 attempted bead types) format. Convenient online tools and Illumina representatives are available to help researchers design and select markers that best suit any research goals.

Intelligent tag SNP Content

Illumina's proven tag SNP approach for selecting BeadChip content allows the most informative markers from the 1kGP data set to be included. The power of a tag SNP approach stems from the inherent correlation among markers, which allows the selection of one highly correlated marker to serve as a proxy for a number of additional highly correlated markers across the genome.

The relationship between markers is commonly measured by correlation coefficient, r^2. A large r^2 value between two markers indicates that they are highly correlated, making them good proxies for each other. At a maximum $r^2 = 1$, two markers are in perfect Linkage Disequilibrium (LD) and can serve as exact proxies for each other, so that only one SNP needs to be genotyped to know the genotype of the other with high certainty. Illumina DNA Analysis products offer unparalleled genomic coverage by leveraging the tag SNP approach, using the highest average r^2 values in the industry and maximizing the likelihood of finding true associations for a given phenotype. By strategically selecting the most powerful tag SNP, Illumina scientists can ensure maximum power to identify associations, while reducing the SNP redundant information on each BeadChip.

Maximized Genomic Coverage

Genomic coverage is a key metric for any whole-genome microarray; it indicates the percent of variation captured on the array at an LD of $r^2 > 0.8$. Prior to the 1kGP, coverage statistics were based on the catalog of variants identified from the International HapMap project. While it was cutting-edge at the time, the HapMap universal

Table 2: Omni Supplimental BeadChip Performance Parameters

	Omni1S		Omni2.5S	
Number of Markers	1,185,076		2,015,318	
Number of Samples	8		8	
DNA Requirement	200 ng		200 ng	
Assay	Infinium HD		Infinium LCG	
Instrument Support	HiScanSQ™ or iScan		HiScan® or iScan	
Sample Throughput*	~960 samples / week		> 1067 samples / week	
Scan Time / Sample	~7.5 minutes		6.5 minutes	
% Variation Captured ($r^2 > 0.8$)	**1kGP† MAF > 5%**	**1kGP† MAF > 1%**	**1kGP† MAF > 5%**	**1kGP† MAF > 1%**
CEU	0.65	0.58	0.61	0.61
CHB + JPT	0.65	0.57	0.63	0.56
YRI	0.37	0.30	0.36	0.29
Data Performance	**Value** / Product Specification**			
Call Rate (average)	99.8% / > 99%		99.95% / > 99%	
Reproducibility	99.8% / > 99.9%		100% / > 99.9%	
Log R Dev	0.17 / < 0.30‡		0.095 / < 0.30‡	
Spacing	**Mean/ Median/ 90th%**			
Spacing (kb)	2.47 / 1.27 / 5.68		1.45 / 0.79 / 3.43	
Marker Categories	**Number of Markers**			
SNPs within 10kb of RefSeq genes	586,877		1,160,001	
Nonsynonymous SNPs (NCBI annotated)	5,641		57,360	
MHC / ADME	1,716 / 7,429		34,179 / 18,365	
Sex Chromosome (X / Y / PAR Loci)	26,451 / 319 / 0		66,578 / 154 / 76	
Mitochondrial	93		31	

* Estimate assumes one HiScan system, one AutoLoader2, one Tecan Robot, and a five-day work week.

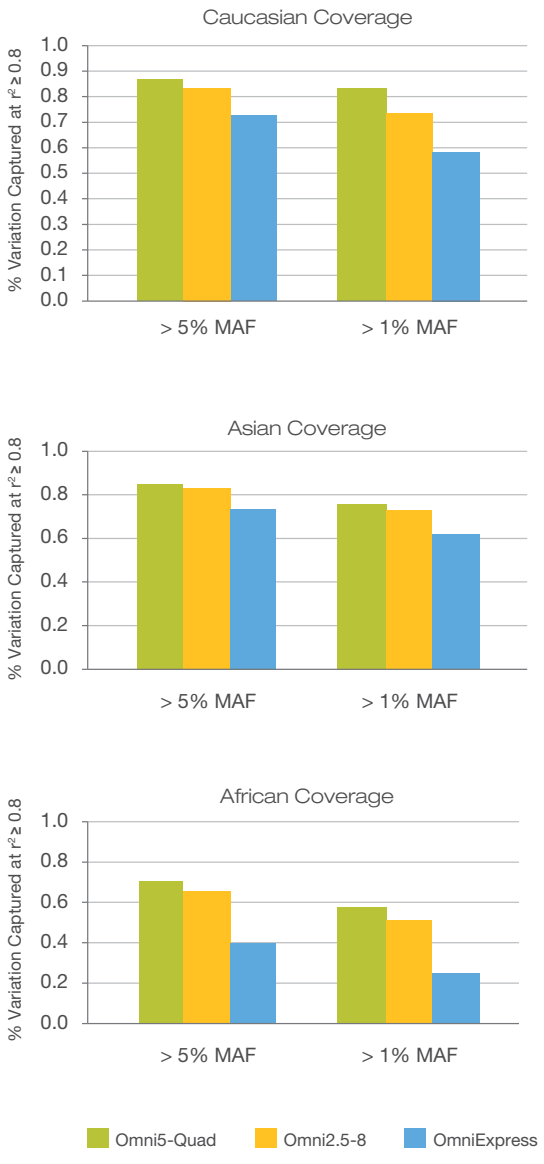
† Compared against June 2011 1kGP data release.

** Values are derived from reference samples.

‡ Value expected for typical projects, excluding tumor samples or any samples prepared not following standard Illumina protocols.

reference database, as we now know, only offered limited information about extent of genetic variation. By the end of the project, it contained ~3.5 million variants, targeting MAFs > 5%. In light of the more comprehensive data available from the 1kGP, the reference point for coverage statistics must be adjusted.
As shown in Table 1 and Figure 2, Omni arrays offer greater than 80% coverage of variants with respect to 1kGP data (MAF > 1%).

Figure 2: Maximized Genomic Coverage



Omni whole-genome microarrays provide superior coverage of common and rare variants across Caucasian, Asian, and African populations. The Omni5 array provides the highest coverage of vairants down to 1% MAF. With the option of selecting 500K additional custom variants on the Omni5, researches can use tagSNPs to increase coverage by up to 10%.

Structural Variation Analysis

Structural variation is thought to be a significant contributor to the genetic basis of human disease. Dense genome-wide coverage on Omni microarrays, coupled with the sensitive Infinium assay, offer researchers a powerful tool for structural variation analysis. The assay delivers very high signal-to-noise ratios and low overall noise levels, which are ideal for precise structural variation analysis. Whether it's genotype calling, structural variation analysis, or both, Omni arrays provide a single solution for any course of genetic research.

Superior Data Quality

The Omni family of microarrays is powered by the Infinium assay, the industry's most trusted, proven DNA analysis platform for both genotyping and CNV studies. The assay is deployed using Illumina proprietary BeadArray™ technology, which allows Omni arrays to deliver a high degree of flexibility, enabling a number of sample formats and a wide multiplex range.

Infinium BeadChips have low DNA input requirements, expanding the range of sample sources that can be used for a study. Genetic researchers worldwide have embraced this technology to catalyze many revolutionary discoveries in disease research and have amassed a vast publication record. Infinium products deliver exceptionally high-quality data with respect to call rates (average > 99%), reproducibility (> 99.9%), and low sample redo rates (Table 1). With such high data quality, the assay minimizes the number of false positives, allowing researchers to avoid time-consuming and frustrating extra analysis and expensive follow-up studies on erroneous associations. High signal-to-noise ratios and low overall noise levels allow for precise, reliable copy number analysis.

Proven Technology

The combination of the Illumina well-proven BeadArray platform, assay technology, and proprietary algorithms present a powerful solution for genetic analysis, delivering the highest quality and most convenient user experience.

BeadArray Manufacturing

Illumina BeadArray technology is based on small silica beads that self assemble in microwells on planar silica slides. Each bead is covered with hundreds of thousands of copies of a specific oligonucleotide that act as the capture sequences in the Infinium assay. Once the beads have self assembled, a proprietary decoding process maps the location of every bead, ensuring that each one is individually quality controlled. The result of this manufacturing process is that every BeadChip undergoes rigorous testing to assure the highest possible quality standards.

Assay Chemistry

The Infinium assay can be scaled to unlimited multiplexing without compromising data quality, unlike many alternative PCR-dependent assays. The simple, streamlined workflow is common across all products, no matter how many SNPs are being interrogated. Likewise, the data acquisition process and analysis are the same. The Infinium assay protocol features single-tube sample preparation and whole-genome amplification without PCR or ligation steps, significantly reducing labor and sample handling errors. After hybridizing unlabeled

DNA sample to the BeadChip, two-step allele detection provides high call rates and accuracy (Figure 3). Selectivity and specificity are accomplished in two steps. Target hybridization to bead-bound 50-mer oligos provides high selectivity while enzymatic single-base extension provides powerful specificity. The single-base extension also incorporates a labeled nucleotide for assay readout. The staining reagent is optimized to provide a higher signal, and more balanced intensities between red and green channels. These features contribute to industry-leading accuracy, high call rates, and copy number data with lower noise.

Genotype Calling

The Infinium assay produces two-color readouts (one color for each allele) for each SNP in a genotyping study. Intensity values for each of the two-color channels, A and B, convey information about the allelic ratio at a single genomic locus (Figure 4). Typical studies incorporate values for a large number of samples (hundreds to tens of thousands) to ensure significant statistical representation. When these values are appropriately normalized and plotted, distinct patterns (or clusters) emerge, in which samples that have identical genotypes at an assayed locus exhibit similar signal profiles (A and B values) and aggregate in clusters. For diploid organisms, bi-allelic loci are expected to exhibit three clusters (AA, AB, and BB).

Genotype calls are based upon information derived from a standard cluster file, which provides statistical data from a representative sample set. This enables genotypes to be called by referencing assay signal intensities against known data for a given locus. Since the call accuracy is tied to the quality of the cluster data, having an efficient and robust clustering algorithm is essential for accurate genotyping. The Illumina proven Gentrain2 algorithm accurately and efficiently identifies cluster patterns of genotyping samples and reports summary statistics. These statistics are used for downstream genotype calling CNV analysis.

BeadArray Scanners and Automation Systems

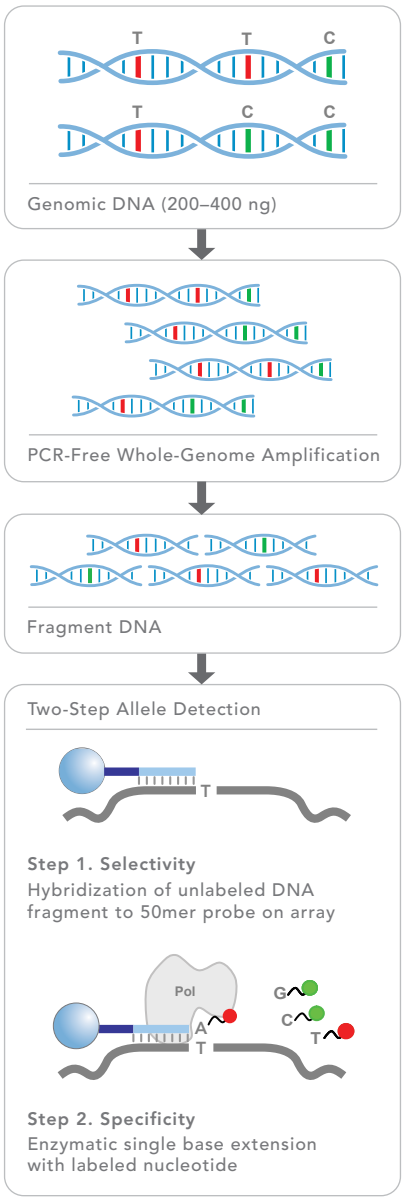
Omni micorarrays are compatible with the Illumina iScan and HiScan systems. These cutting-edge array scanners feature high-performance lasers and powerful optical systems that enable rapid scan times and precise assay detection. A convenient modular design enables researchers to easily build out the systems for evolving research needs. An optional Laboratory Information Management System (LIMS) is available to accurately and efficiently track samples. Robotic automation capabilities can be added to improve throughput for labs processing large numbers of samples.

Data Analysis Software

GenomeStudio® Data Analysis Software by Illumina offers integrated genotyping and copy number tools and a graphical Genome Viewer. GenomeStudio has an open plug-in interface to integrate third-party applications for more downstream data analysis options. Beeline Software provides a direct path to project creation and sample management for large array experiments. The time required for data analysis is reduced by flexible allele calling and data filtering prior to entry into GenomeStudio.

The illumina•Connect program leverages this open architecture and has made numerous plug-ins available to support genotyping and copy number analysis.

Figure 3: Infinium HD Assay Chemistry



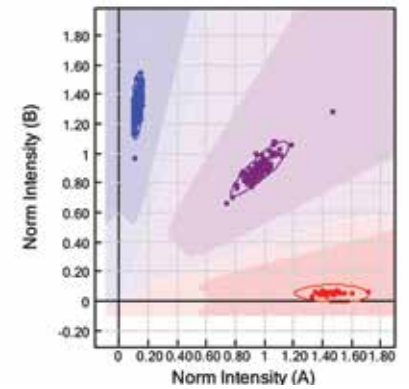
Services

Illumina FastTrack Genotyping Services are available to analyze samples in a timely fashion at a reasonable cost using any Infinium DNA Analysis BeadChip. This option allows researchers to acquire high-quality data for limited studies or before purchasing their own equipment.

Product Summary

Leveraging the proven Infinium assay, powerful BeadArray technology, an advanced tag SNP selection strategy, and the latest genomic content from the 1kGP, Omni microarrays offer unprecedented access to the human genome and enable a range of new hypotheses that will fuel the next wave of ground-breaking discoveries.

Figure 4: Typical Diploid Genotyping Plot



Three clusters of points can be seen for this example locus. The red, purple, and blue regions represent the AA, AB, and BB clusters, respectively.

Ordering Information

BeadChip Kit	Catalog No.
HumanOmniExpress-12 DNA Analysis Kit v1 (48 Samples)	WG-312-1120
HumanOmniExpress-12 DNA Analysis Kit v1 (288 Samples)	WG-312-1121
HumanOmniExpress-12 DNA Analysis Kit v1 (1152 Samples)	WG-312-1122
HumanOmni1S-8 DNA Analysis Kit v1 (16 Samples)	WG-311-1114
HumanOmni1S-8 DNA Analysis Kit v1 (48 Samples)	WG-311-1115
HumanOmni1S-8 DNA Analysis Kit v1 (96 Samples)	WG-311-1116
HumanOmni1S-8 DNA Analysis Kit v1 (384 Samples)	WG-311-1117
HumanOmni2.5-8 DNA Analysis Kit v1 (16 Samples)	WG-311-2511
HumanOmni2.5-8 DNA Analysis Kit v1 (48 Samples)	WG-311-2512
HumanOmni2.5-8 DNA Analysis Kit v1 (96 Samples)	WG-311-2513
HumanOmni2.5-8 DNA Analysis Kit v1 (384 Samples)	WG-311-2514
HumanOmni2.5S-8 DNA Analysis Kit v1 (16 Samples)	WG-311-2505
HumanOmni2.5S-8 DNA Analysis Kit v1 (48 Samples)	WG-311-2506
HumanOmni2.5S-8 DNA Analysis Kit v1 (96 Samples)	WG-311-2507
HumanOmni2.5S-8 DNA Analysis Kit v1 (384 Samples)	WG-311-2508
HumanOmni5-Quad DNA Analysis Kit v1 (16 Samples)	WG-311-5001
HumanOmni5-Quad DNA Analysis Kit v1 (48 Samples)	WG-311-5002
HumanOmni5-Quad DNA Analysis Kit v1 (96 Samples)	WG-311-5003
HumanOmni5-Quad DNA Analysis Kit v1 (384 Samples)	WG-311-5004

Illumina, Inc. • 9885 Towne Centre Drive, San Diego, CA 92121 USA • 1.800.809.4566 toll-free • 1.858.202.4566 tel • techsupport@illumina.com • illumina.com

FOR RESEARCH USE ONLY

 Pub. No. 370-2010-027 Current as of 31 January 2013

illumina®