Genome-Wide DNA Analysis BeadChips
Up To 50% More Coverage of Common and Rare Variants

Introduction

Genome-wide association studies (GWAS) are a proven tool that enables researchers to achieve a greater understanding of human health and disease. By evaluating the whole-genome genotypes of a large number of DNA samples, researchers can identify the causative alleles for many human traits and diseases. To increase the power of GWAS, the human genetics community is greatly expanding the catalog of identified variants. Leveraging this new information, Illumina's Omni family of microarrays provide unprecedented coverage of the human genome, enabling researchers to examine the role of rarer variants and equipping them for success in the next generation of GWAS.

Omni microarrays feature content derived from the most recent discoveries, including the 1000 Genomes Project (1kGP). A collaborative effort from research institutions around the globe, the 1kGP is using the power of next-generation sequencing to identify human genetic variants that occur at any appreciable frequency (≥ 1% MAF, or minor allele frequency) across diverse populations. With this vast amount of information, Illumina scientists estimate that roughly 5 million markers will be needed for microarrays to capture genome-wide variation down to 1% MAF.

To enable the fastest access to new data from the 1kGP, Illumina has developed a product roadmap for Omni microarrays that allows researchers to begin their next-generation GWAS now. Starting from any Omni array (Figure 1), researchers can continually add the latest content from the 1kGP to their study in a step-wise manner and progressively build up to the full 5 million variants. For GWAS follow-up studies, custom iSelect® BeadChips can be easily developed with up to 200,000 markers targeting any loci across the genome, including any new variants identified by the 1kGP.

Proven Genotyping and CNV Platform

The Omni family of microarrays is powered by the whole-genome Infinium® HD Assay, the industry’s most trusted, proven DNA analysis platform for genotyping and copy number variation (CNV) studies. The assay is deployed with Illumina's proprietary BeadArray™ technology. With the Infinium platform, Omni arrays deliver a high degree of flexibility, enabling a number of sample formats and a wide multiplex range that allows researchers to profile over 2.5 million markers per sample. Infinium HD BeadChips have low DNA input requirements, expanding the range of sample sources that can be used for a study. Genetic researchers world-wide 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.

The Omni2.5 Path

The HumanOmni2.5 BeadChip (Omni2.5) is a ground-breaking product for genomic research, the first commercial microarray designed to be maximally powerful for 1kGP data. This array features ~2.5 million markers that capture genomic variation down to 2.5% MAF and delivers the highest coverage rates across diverse populations. With respect to the 1kGP data set (2.5% MAF), the Omni2.5 provides up to 50% greater coverage of the genome than any previous commercial product. Using the proven iScan or HiScan™SQ System, this four-sample BeadChip offers high throughput to support large-scale studies. Optimized tag SNP content and extremely dense marker spacing (median gap = 0.63 kb) enables a broad range of study types, providing extremely high resolution for copy number variation (CNV) and other structural variation applications.

Researchers starting the Omni Roadmap with the Omni2.5 will complete the 5 million marker set with the Omni2.5S, which will provide a unique set of ~2.5 million markers derived from future releases of the 1kGP, including coverage of rare variants down to 1% MAF.

The Omn1 Path

The four-sample HumanOmni1-Quad BeadChip (Omni1) features content derived from all three HapMap phases, along with strategically selected markers chosen to target high-value regions of the genome associated with human disease. Each BeadChip features over one million available assays per sample, containing carefully chosen markers that capture genomic variation down to 5% MAF. High-density markers, having a median spacing of 1.5 kb, ensure a high level of resolution for CNV identification. Specialized content includes ~18,000 SNPs targeting four 1Mb regions known to be associated to three or more human diseases; over 38,000 SNPs predicted to be nonsynonymous; 62,000 SNPs covering an additional 100 intervals surrounding published peak markers from the NHGRI GWAS database; and the remaining 950 top single-marker associated SNPs from the GWAS database.
Researchers starting the Omni Roadmap with the Omni1 array can add an additional ~1.2 million markers with the HumanOmni1S (Table 2), providing high coverage of low-frequency alleles down to ~2.5% MAF from the first releases of 1kGP. The final 2.5 million markers will be available on the Omni2.5S, which will provide comprehensive coverage of rare alleles down to 1% MAF derived from the complete 1kGP.

### The OmniExpress Path

The HumanOmniExpress BeadChip (OmniExpress) delivers excellent power for GWAS, providing high sample throughput and comprehensive genomic content at the industry's best price. This twelve-sample BeadChip is the ideal solution for processing the greatest number of samples within a given budget. Optimized tag SNP content from all three HapMap phases has been strategically selected to capture the greatest amount of common SNP variation (> 5% MAF) and drive the discovery of novel associations with traits and diseases.
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.

Like the Omni1 Path, researchers starting the Omni Roadmap with either the OmniExpress or OmniExpress+ array can add an additional ~1.2 million markers with the Omni1S. The content on the Omni1S is derived from the first releases of the 1kGP, providing high coverage of low frequency alleles down to ~2.5% MAF. The third array on this path is the Omni2.5S, which will deliver the final 2.5 million markers from the complete 1kGP, providing comprehensive coverage of rare alleles down to 1% MAF.

---

### Table 1: Omni BeadChip Performance Parameters

<table>
<thead>
<tr>
<th></th>
<th>OmniExpress</th>
<th>Omni1</th>
<th>Omni2.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Markers</td>
<td>731,442</td>
<td>1,138,747</td>
<td>2,443,179</td>
</tr>
<tr>
<td>Number of Samples</td>
<td>12</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>DNA Requirement</td>
<td>200 ng</td>
<td>200 ng</td>
<td>200 ng</td>
</tr>
<tr>
<td>Assay</td>
<td>Infinium HD</td>
<td>Infinium HD</td>
<td>Infinium HD</td>
</tr>
<tr>
<td>Instrument Support</td>
<td>HiScanSQ or iScan</td>
<td>HiScanSQ, iScan, BeadArray Reader</td>
<td>HiScanSQ or iScan</td>
</tr>
<tr>
<td>Sample Throughput*</td>
<td>&gt; 1,400 samples / week</td>
<td>~576 samples / week</td>
<td>~480 samples / week</td>
</tr>
<tr>
<td>Scan Time / Sample</td>
<td>~5 minutes</td>
<td>~13 minutes</td>
<td>~15 minutes</td>
</tr>
<tr>
<td>% Variation Captured ((r^2 &gt; 0.8))</td>
<td>HapMap MAF &gt; 5%</td>
<td>1kGP MAF &gt; 2.5%</td>
<td>HapMap MAF &gt; 5%</td>
</tr>
<tr>
<td>CEU</td>
<td>0.91</td>
<td>0.65</td>
<td>0.94</td>
</tr>
<tr>
<td>CHB + JPT</td>
<td>0.91</td>
<td>0.74</td>
<td>0.93</td>
</tr>
<tr>
<td>YRI</td>
<td>0.66</td>
<td>0.36</td>
<td>0.78</td>
</tr>
<tr>
<td>Data Performance</td>
<td>Call Rate (average)</td>
<td>99.84% / &gt; 99%</td>
<td>99.97% / &gt; 99%</td>
</tr>
<tr>
<td>Reproducibility</td>
<td>99.99% / &gt; 99.9%</td>
<td>100% / &gt; 99.9%</td>
<td>99.99% / &gt; 99.9%</td>
</tr>
<tr>
<td>Log R Dev</td>
<td>0.15 / &lt; 0.30(^a)</td>
<td>0.13 / &lt; 0.30(^a)</td>
<td>0.13 / &lt; 0.30(^a)</td>
</tr>
<tr>
<td>Spacing (kb)</td>
<td>3.98 / 2.17 / 9.23</td>
<td>2.63 / 1.25 / 6.39</td>
<td>1.19 / 0.64 / 2.76</td>
</tr>
</tbody>
</table>

### Marker Categories

<table>
<thead>
<tr>
<th></th>
<th>Number of Markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of SNPs with 10kb of RefSeq genes</td>
<td>392,511</td>
</tr>
<tr>
<td>Nonsynonymous SNPs (NCBI annotated)</td>
<td>15,080</td>
</tr>
<tr>
<td>MHC / ADME</td>
<td>7,481 / 16,673</td>
</tr>
<tr>
<td>Sex Chromosome (X / Y / PAR Loci)</td>
<td>18,159 / 1,679 / 572</td>
</tr>
<tr>
<td>Mitochondrial</td>
<td>18,159 / 1,679 / 572</td>
</tr>
</tbody>
</table>

\(^a\) Estimate assumes one iScan system, one AutoLoader2, one Tecan Robot, and a five-day work week.

\(^b\) Values are derived from reference samples.

---

**GWAS Follow-Up Studies**

Customized iSelect BeadChips can be easily developed to fit any experimental design, allowing researchers to develop the ideal solution for their loci multiplexing and sample throughput requirements. Illumina’s iSelect custom genotyping platform offers all of the benefits of standard Infinium products, including industry-leading data quality and call rates, streamlined workflow, and rational SNP selection with access to the entire genome. Custom products can be deployed on either the 12-sample format (60,801 to 200,000 attempted bead types), or the 24-sample format (3,000 to 68,000 attempted bead types). Convenient online tools and Illumina representatives are available to help researchers design and select markers that best suit any research goals.
a high $r^2$ between two SNPs indicates high correlation, making these each BeadChip. Power to identify associations, while reducing the SNP redundancy on from each haplotype block, Illumina scientists can ensure maximum given phenotype. By strategically selecting the most powerful tag SNP in the industry and maximizing the likelihood of finding true associations for a ing the tag SNP approach, using the highest average $r^2$ values in the Analysis products offer unparalleled genomic coverage by leverag- ing 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 from each haplotype block, Illumina scientists can ensure maximum power to identify associations, while reducing the SNP redundancy 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 very cutting-edge at the time, the HapMap universal 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, starting with the Omni2.5, Omni arrays are the first commercial products to maintain greater than 80% coverage with respect to 1kGP data (2.5% MAF). Omni arrays provide up to 50% greater coverage of common and rare variants over any previous array product.

Multi-Use Sample Preparation
For researchers using Omni microarrays, Illumina has developed Infinium Multi-Use Sample Preparation kits. These kits require only a single sample amplification, which can be used and then stored for supplemental Omni Roadmap arrays as they become available. Illumina testing has shown that the samples will continue to produce high-quality data over multiple usages without any deleterious effects from storage periods. The DNA input requirement for the multi-use kit is 750 ng at 50 ng/μl. Both multi-use and single-use sample preparation workflows are available for each Omni Roadmap path. Whichever workflow is selected upon entry to the Omni Roadmap must be maintained for all supplemental arrays as well. That is, a customer cannot begin the Omni Roadmap with a multi-use product and then add supplemental content with a single-use product, or vice versa.

Structural Variation Analysis
Structural variation, including copy number variants and copy neutral variants (inversions and translocations), are 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 HD 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.

Proven Technology
The combination of Illumina's 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’s 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 HD 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.

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 that form haplotype blocks, 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 correlation between SNPs is commonly described by $r^2$, where a high $r^2$ between two SNPs indicates high correlation, making these SNPs good proxies for each other. At a maximum $r^2 = 1$, two SNPs are in perfect Linkage Disequilibrium (LD) and can serve as exact prox- ies for each other; therefore, only one SNP needs to be genotyped to know the genotype of the others with high certainty. Illumina DNA Analysis products offer unparalleled genomic coverage by leverag-
Assay Chemistry

The Infinium HD 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 (Figure 3). Likewise, the data acquisition process and analysis are the same.

The Infinium HD 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. 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.

These charts show the percentage of genomic variation captured by Illumina whole-genome arrays and other competing products for three populations—CEU, CHB/JPT, and YRI. Prior to the Omni2.5, all whole-genome arrays were derived from HapMap data. For simplicity, these HapMap-based arrays have been consolidated to show the range of variation they capture. These ranges have been calculated for three datasets: HapMap > 5% MAF, 1kGP > 5% MAF, and 1kGP > 2.5% MAF.

With respect to HapMap data (green bars), all arrays show a high percentage of variation captured, exceeding the industry standard, 80% (red line). However, as the reference data sets are expanded to include 1kGP data (orange and blue bars), the coverage ranges of the HapMap-based arrays drop precipitously. In contrast, the Omni2.5 has been designed to be maximally powerful across all three reference data sets for all populations.
Genotype Calling

The Infinium HD 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. Illumina’s proven Gentrain2 algorithm accurately and efficiently identifies cluster patterns of genotyping samples and reports summary statistics. These statistics are used for downstream genotype calling and copy number variation (CNV) analysis.

BeadArray Scanners and Automation Systems

Omni microarrays are compatible with the Illumina iScan and HiScanSQ systems. These cutting-edge array scanners feature high-performance lasers and powerful optical systems that enable rapid scan times and precise assay detection. The HiScanSQ system can also perform Illumina sequencing by synthesis chemistry, the world’s most widely adopted next-generation sequencing platform.

A convenient modular design enables researchers to easily build out the system 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

Illumina’s GenomeStudio® Data Analysis Software 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.

Bedline 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.

iControlDB
Illumina hosts a database of genotypic and phenotypic data generated by researchers using Illumina genotyping products, which can be used to supplement controls in case-control association studies. These control samples can also be used as a scaffold to align data from different experiments for genotype imputation, enabling researchers to increase the power of an association study and decrease overall project costs.

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 HD 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.