

# Human1M DNA Analysis BeadChip

The Human1M DNA Analysis BeadChip interrogates more than 1,000,000 SNPs on a single array, combining superior genomic coverage and power for whole-genome association and copy number studies.

## INTRODUCTION

The Illumina Human1M DNA Analysis BeadChip uses the powerful and proven Infinium® Assay to profile human genomes at more than one million loci. The Infinium Assay<sup>1,2</sup> uses a single-tube, whole-genome amplification method that does not require PCR and enables intelligent marker selection of virtually any SNP or probe across the human genome<sup>3,4</sup>. Tag SNPs and other SNPs from the International HapMap Project and

NCBI's dbSNP have been selected for the Human1M BeadChip to increase genomic coverage and probe uniformity across the genome. As proxies for many other SNPs, the use of tag SNPs greatly increases the power of genome-wide association studies.

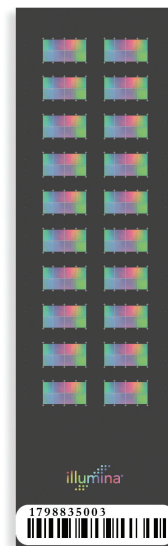
The content of the Human1M BeadChip is focused on tag SNPs, SNPs in genes, and SNPs and non-polymorphic markers in known and novel copy number variation (CNV) regions. In addition to ~950,000 genome-spanning tag SNPs and ~100,000 additional non-HapMap SNPs, the Human1M BeadChip contains more than 565,000 SNPs in and near coding regions of the genome such as nsSNPs, promoter regions, 3' and 5' UTRs, as well as dense coverage in ADME and MHC regions. For identification of CNVs, there are ~260,000 markers located in novel and reported<sup>5</sup> copy number polymorphic regions. In sum, the vast panel of markers on the Human1M BeadChip provides powerful whole genome analysis for both association studies and structural variation research.

Additionally, current Human-Hap550 researchers have the flexibility to upgrade their studies to include the exact content as on the Human1M BeadChip with the highly gene-centric Human450S DNA Analysis BeadChip.

### HIGHLIGHTS OF THE HUMAN1M BEADCHIP

- **Superior SNP content:** First platform to offer more than one million SNPs on a single array
- **Novel CNV content:** Targets 14,000 known and novel CNV regions with SNPs and probes
- **Comprehensive coverage:** Superior genome-wide coverage of CEU, CHB+JPT, and YRI HapMap populations
- **High-value content:** High SNP density in genes, ADME, and MHC regions
- **Full software support:** Integrated analysis of both SNPs and probes
- **Simple workflow:** PCR-free protocol using a single BeadChip for each sample

FIGURE 1: HUMAN1M DNA ANALYSIS BEADCHIP



## COMPREHENSIVE GENOMIC COVERAGE

The exceptionally comprehensive set of markers on the Human1M BeadChip provides access to dense genome-wide tag SNP coverage as well as additional content targeted to high-value genomic regions of interest. The Human1M BeadChip contains the same content as the HumanHap550, plus over 450,000 additional markers to further expand coverage across the genome and provide dense coverage in genes and other functional regions.

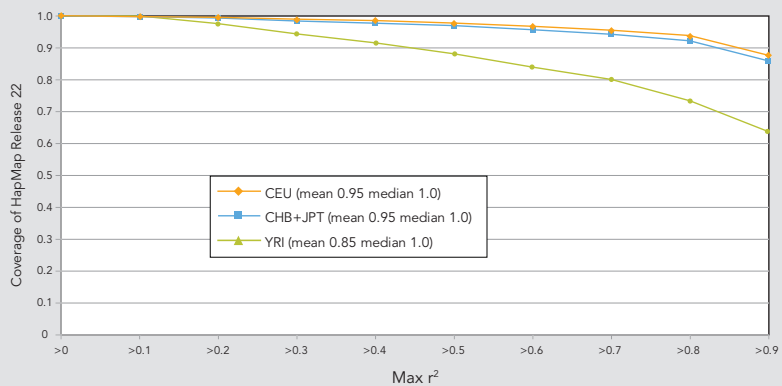
Illumina scientists have created a panel of intelligently selected tag SNPs that provide outstanding

power for probing genetic variation. Tag SNPs on the Human1M BeadChip were selected from data produced by the International HapMap Project<sup>6</sup>. The Phase II HapMap database contains over two million common SNPs—those with a minor allele frequency (MAF)  $\geq 0.05$  in each population studied (Caucasian [CEU], Han Chinese/Japanese [CHB+JPT], and Yoruba [YRI]). Illumina uses an algorithm relying on linkage disequilibrium (LD) statistics to identify the most informative tag SNPs that capture the majority of common variation in all four HapMap populations. Tag SNPs selected from HapMap populations, in turn, have been shown to be broadly transferable for association studies using samples from many independent populations<sup>7</sup>. This rational marker design process results in the Human1M BeadChip providing coverage of 94% of CEU, 93% of CHB+JPT, and 74% of YRI HapMap Phase II loci at an  $r^2$  value  $\geq 0.8$  (Figure 2).

Additional SNPs have been selected in and around genes to increase the gene coverage and other probes are located in SNP deserts to fill in coverage gaps. The uniform genome-wide coverage results in a mean spacing between markers of 2.7kb (median = 1.7kb) and very few large gaps for high-resolution CNV identification (Figure 3).

The Human1M BeadChip is designed to interrogate regions of CNV, in addition to providing thorough genomic coverage. This includes difficult to analyze regions like megasatellites and segmental duplications, which are targeted with both SNPs and non-polymorphic probes. The Human1M BeadChip covers 3,298 of the 3,644 CNV regions reported in the Database of Genomic Variants<sup>5</sup> with

FIGURE 2: GENOMIC COVERAGE BY POPULATION



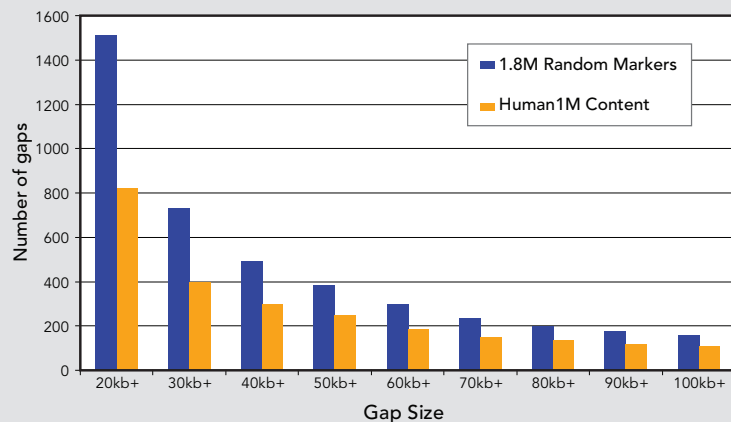
The Human1M BeadChip content covers the majority of HapMap common variation in three distinct populations.

more than 206,000 markers. Developed in collaboration with deCODE Genetics, the Human1M BeadChip contains more than 52,000 additional markers for regions likely to contain undiscovered CNV, including segmental duplications, megasatellites, and regions lacking SNPs. These novel CNV probes and the dense uniform genome-wide coverage support unbiased discovery and analysis of copy number polymorphisms.

Since SNPs and CNV markers on the Human1M BeadChip share the same design parameters, all markers are analyzed together in a fully integrated fashion using Illumina BeadStudio software.

In addition to the broad coverage crucial for all whole-genome association studies, the Human1M BeadChip also targets other high-value content. The Human1M BeadChip has more than 10,000 SNPs and non-polymor-

FIGURE 3: ILLUMINA INTELLIGENT MARKER SELECTION MINIMIZES GAPS



The intelligent marker selection used for the Human1M BeadChip results in substantially fewer genome-wide inter-marker gaps compared to randomly selected markers.

TABLE 1: HUMAN1M BEADCHIP MARKER DESIGN

CONTENT	NUMBER
All markers	1,072,820
SNPs in genes	565,718
nsSNPs	23,288
SNPs in reported <sup>5</sup> CNV regions	206,665
Probes and SNPs in new CNV regions	~52,000
SNPs in mitochondrial genome	163
X-chr/Y-chr/PAR loci	40,097/2,283/686
SNPs within 10kb of a known ADME related gene	15,468
SNPs in MHC region (defined by de Bakker et al. <sup>10</sup> )	10,073

phic probes focused in the MHC region. The 4.3Mb MHC region contains a high density of genes, many of which have been associated with various autoimmune and infectious diseases, such as Type-I diabetes, multiple sclerosis, malaria, and HIV<sup>8</sup>. The dense coverage facilitates accurate genotyping of this complex region. ADME (absorption, distribution, metabolism, and excretion) studies and pharmacogenomics research to improve drug efficacy is supported with more than 15,000 markers in ~200 important gene regions related to drug metabolism.

#### HIGH QUALITY DATA

All of the assays on the Human1M DNA Analysis BeadChip use standard Infinium chemistry. Rigorous functional testing ensures strong and reproducible performance. Illumina has assessed data quality by analyzing a diverse panel of HapMap samples. Successful whole-genome association studies depend, in part, on the high call rates that Illumina DNA Analysis BeadChips exhibit. Since complex disease traits often have relatively small gene effects, potential associations may be missed if an assayed SNP, in LD with a disease SNP, has a low call rate. Data acquired from the Human1M

DNA Analysis BeadChip show strong concordance with the International HapMap Project (99.66% for >950,000 loci, Table 2). Additionally, the Human1M BeadChip provides precise copy number metrics (Table 2) with low overall noise levels (log R ratio

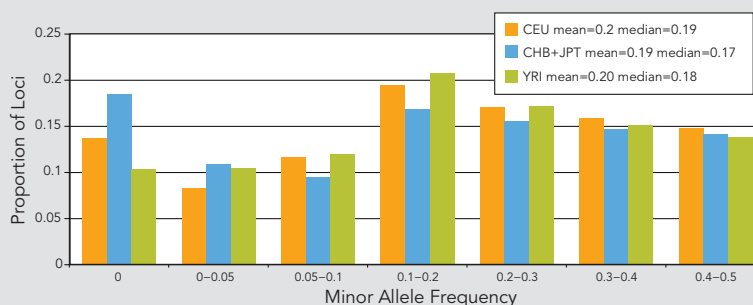
stdev = 0.19), which allows for the reliable detection of single changes in copy number levels.

#### SOFTWARE AND SERVICES

Illumina's BeadStudio analysis software package offers integrated copy number tools and a Genome Viewer. BeadStudio also offers an open plug-in interface that accepts other downstream analysis tools created in collaboration between Illumina and various partners. BeadStudio plug-ins are available for download from illumina•Connect<sup>9</sup>.

Importantly, all of the SNPs and probes on the Human1M BeadChip are created using the same rational design strategy to maximize genome-wide SNP and copy number

FIGURE 4: DISTRIBUTION OF MINOR ALLELE FREQUENCY BY POPULATION



The Human1M BeadChip provides access to loci having a wide range of minor allele frequencies in unrelated individuals in three distinct HapMap populations.

TABLE 2: HUMAN1M BEADCHIP GENOTYPING DATA QUALITY ON 125 DNA SAMPLES (7 REPLICATES, 28 TRIOS)

GENOTYPING PARAMETER	PERCENT	SPECIFICATION
Call Frequency	99.66%	>99% average
Reproducibility	99.99%	>99.9%
Mendelian Inconsistencies	0.05%	<0.1%
HapMap Concordance	99.66%	
CNV ANALYSIS PARAMETER	STDEV	PRODUCT AVERAGE
log R Ratio	0.19	~0.20
B Allele Frequency	0.04*	~0.04

\*Heterozygotes only

measurements. Consequently, all markers can be analyzed together with BeadStudio software, without changes to the normalization algorithm. Because of this continuity across all BeadChips, Illumina allows completely integrated genotyping and copy number studies.

As with all of Illumina's standard products, an optional Laboratory Information Management System (LIMS) and robotic automation are available to accurately and efficiently track samples throughout analysis. Additionally, Illumina's in-house FastTrack Genotyping Service is available to analyze samples in a timely fashion at a reasonable cost using the Human1M BeadChip.

#### ILLUMINA DNA ANALYSIS SOLUTIONS

The high quality data and low per-sample cost of the Human1M BeadChip offers researchers unprecedented DNA analysis and genotyping power. The greater than 1,000,000 markers on each BeadChip provide comprehensive genome-wide coverage, targeted dense coverage of high-value regions, as well as coverage of unstable regions that traditionally have been difficult to target.

Researchers who have already begun genotyping studies with the HumanHap550 BeadChip can expand their studies to incorporate the new content by using the Human450S BeadChip. The combined content of these two BeadChip arrays is identi-

cal to that of the Human1M BeadChip.

The combination of Illumina's proprietary assay technologies and intelligent content deployment delivers the most comprehensive portfolio of DNA analysis solutions now available.

#### ORDERING INFORMATION

CATALOG NO.	PRODUCT	SAMPLES	DESCRIPTION
WG-30-1001	Human1M DNA Analysis kit	8	Each Human1M DNA Analysis BeadChip can process one sample and assay >1,000,000 loci. Each package contains eight BeadChips, along with reagents for amplifying, fragmenting, hybridizing, labeling, and detecting eight whole-genome genotyping samples.
WG-30-1002	Human1M DNA Analysis kit	24	Each package contains BeadChips and reagents for processing 24 samples.
WG-30-1003	Human1M DNA Analysis kit	96	Each package contains BeadChips and reagents for processing 96 samples.

#### REFERENCES

- (1) Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS (2005) A genome-wide scalable SNP genotyping assay using microarray technology. *Nat Genet* 37: 549-554.
- (2) Steemers FJ, Chang W, Lee G, Barker DL, Shen R et al. (2006) Whole-genome genotyping with the single-base extension assay. *Nat Methods* 3: 31-33.
- (3) Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L et al. (2004) Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 74: 106-120.
- (4) Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E et al. (2005) Whole-genome patterns of common DNA variation in three human populations. *Science* 307: 1072-1079.
- (5) Database of Genomic Variants. <http://projects.tcag.ca/variation> (Build 36) Accessed July, 2007.
- (6) The International HapMap Consortium (2003) The International HapMap Project. *Nature* 426: 789-796.
- (7) de Bakker PI, Burt NP, Graham RR, Guiducci C, Yelensky R et al. (2006) Transferability of tag SNPs in genetic association studies in multiple populations. *Nat Genet* 38: 1298-1303.
- (8) Traherne JA, Horton R, Roberts AN, Miretti MM, Hurles ME et al. (2006) Genetic analysis of completely sequenced disease-associated MHC haplotypes identifies shuffling of segments in recent human history. *PLoS Genet* 2: e9.
- (9) Illumina•Connect. Available: <http://www.illumina.com/IlluminaConnect>.
- (10) de Bakker PI, McVean G, Sabeti PC, Miretti MM, Green T et al. (2006) A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat Genet* 38(10):1166-1172.

#### ADDITIONAL INFORMATION

Visit our website or contact us at the address below to learn more about Illumina DNA Analysis BeadChips.

**Illumina, Inc.**

**Customer Solutions**

9885 Towne Centre Drive  
San Diego, CA 92121-1975  
1.800.809.4566 (toll free)  
1.858.202.4566 (outside the U.S.)  
[techsupport@illumina.com](mailto:techsupport@illumina.com)  
[www.illumina.com](http://www.illumina.com)

#### FOR RESEARCH USE ONLY

© 2007 Illumina, Inc. All rights reserved.

Illumina, Solexa, Making Sense Out of Life, Oligator, Sentrix, GoldenGate, DASL, BeadArray, Array of Arrays, Infinium, BeadXpress, VeraCode, IntelliHyb, iSelect, and CSPro are registered trademarks or trademarks of Illumina. All other brands and names contained herein are the property of their respective owners.  
Pub. No. 370-2007-012 16Aug07

