

Analyzing Copy Number Variation with the Infinium® Whole-Genome Genotyping Assay

The Infinium Assay and the Illumina® DNA Analysis product line provide powerful tools for analyzing copy number variants in the study of human disease.

INTRODUCTION

Genome profiling for chromosomal aberrations such as amplifications and deletions is a crucial element of cancer biology as well as the analysis of chromosomal abnormalities associated with congenital disorders. The study of DNA copy number variations (CNVs) has become an essential component in the field of population genetics and whole-genome association studies.

Numerous regions in the human genome contain large-scale copy number variations such as segmental duplications. Although the biological significance of these regions is just coming to be understood, they

are thought to contribute to population diversity, disease, and gene expression levels.

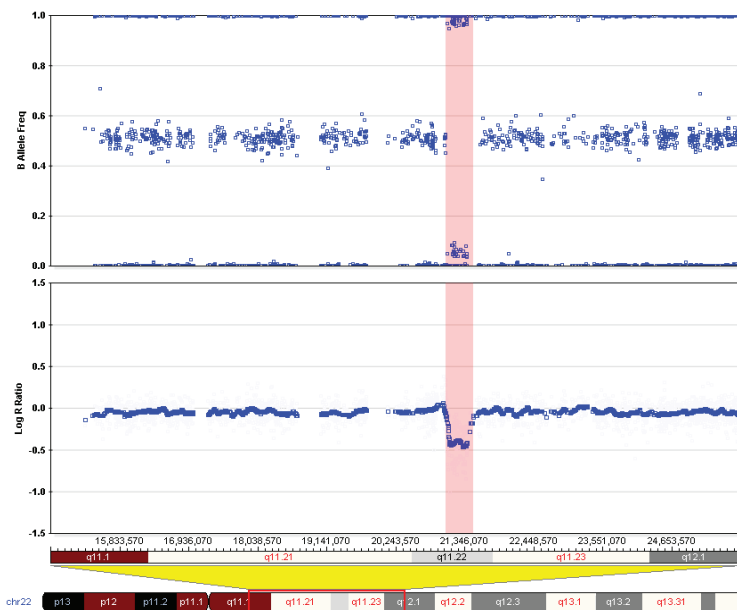
Data generated from the Illumina DNA Analysis BeadChips can be used to analyze both novel and commonly reported copy number variants in the human genome. The

high density of markers on the DNA Analysis BeadChips allows precise estimation of CNV breakpoints across multiple individuals while targeting a high percentage of the most commonly reported CNV regions.

HIGHLIGHTS OF THE INFINIUM ASSAY AND CNV ANALYSIS

- Comprehensive CNV Coverage
Enables analysis of novel and commonly reported regions of CNV
- Simple Workflow
PCR-free protocol using a single BeadChip for each sample
- High-Quality Data
Infinium Assay provides high call rates and precise copy number measurements
- Intelligent SNP Selection
Genome-wide coverage of all populations with up to 650,000 markers at down to 2.5kb median marker spacing

FIGURE 1: COPY NUMBER VARIANT IN A HAPMAP SAMPLE DETECTED ON THE HUMANHAP550 GENOTYPING BEADCHIP



A CNV of size ~400kb on Chromosome 22 is noted by a loss of heterozygosity in the B allele frequency and a deflection downward to -0.5 in the log R ratio. Both parameters indicate a hemizygous loss in copy number from two to one. In addition to other genes, this CNV region contains the genes SUHW2 and PRAME, which are both known to lie within a commonly reported CNV^{1,2}. These profiles were generated using the Illumina Genome Viewer within BeadStudio. For the log R ratio, a 100kb moving average is shown.

TABLE 1: DNA ANALYSIS MARKER COVERAGE OF ~2,700 CNV REGIONS FROM THE DATABASE OF GENOMIC VARIANTS*

Total	HumanHap300-Duo	HumanCNV370-Duo	HumanHap550	HumanHap650Y
CNV regions covered with one or more SNPs	89%	98.5%	94%	94%
Total SNPs lying in reported Toronto CNV regions	> 52,000	> 74,000	> 96,000	> 111,000

*From build 36, February 2007

INFLUENCE OF CNV REGIONS ON THE GENOME

There are many ways in which CNV can affect the human genome. First, a dosage-sensitive gene may be encompassed by a structural variant. Changes in the copy number level of this region may alter the gene dosage. Second, a hemizygous deletion (reduction from two copies to one) of a CNV in a gene region can unmask a recessive mutation on a homologous chromosome. Alternatively, genes overlapping structural variants can be disrupted by inversion, translocation, deletion, or copy number variant breakpoints. Last, there may also be positional effects whereby regulatory elements located within a CNV region can be altered or deleted³.

BEADCHIP CONTENT

Illumina’s HumanHap300-Duo, HumanCNV370-Duo, HumanHap550, and HumanHap650Y Genotyping BeadChips, in conjunction with the powerful Infinium Assay^{4,5}, interrogate up to ~650,000 loci efficiently and accurately on a single BeadChip. The Infinium Assay uses a single-tube, whole-genome amplification method that does not require PCR.

The median SNP spacing for each product remains relatively constant along the length of the genome, enabling high-powered loss of heterozygosity (LOH) and copy number studies. For the HumanCNV370-

Duo, HumanHap300, HumanHap550, and the HumanHap650Y, the median SNP spacing is ~5kb, ~4.9kb, ~2.8kb, and ~2.0kb, respectively.

Targeting New CNV Regions with the HumanCNV370-Duo

The HumanCNV370-Duo contains the same content as the HumanHap300-Duo but targets an additional ~11,000 CNV regions, not available from public databases, for a total of over 370,000 markers. These regions include segmental duplications, megasatellites, and regions lacking SNPs. These CNV regions have been targeted with SNPs and non-polymorphic probes. On average, there are three markers for CNV regions of less than 30kb. For larger regions, one marker was placed every 10kb.

ILLUMINA CNV ANALYSIS

BeadStudio software enables effective scanning of data generated from Illumina BeadChips for structural aberrations and copy number variants. Illumina has shown that copy number analysis is optimal when using two metrics. First is the B allele frequency (allelic composition) measurement based on theta intensity values, which has extremely low noise levels and high signal-to-noise ratios. The second is the log R ratio, a parameter based on the normalized intensity values, which also has low noise levels⁶. The B allele frequency is used to determine the presence of

CNVs and the log R ratio is used to determine the nature of the CNVs (Figure 1, 2).

Due to the rapid adoption of array-CGH and SNP genotyping arrays for the detection of copy number variation in the human genome, the number of reported CNV regions has dramatically increased. Some of these regions are currently being assembled in the Database of Genomic Variants⁷, the objective of which is to provide a continually updated comprehensive summary of structural variation in the human genome⁸. A significant proportion of the markers on each HumanHap BeadChip lies within these reported CNV regions, allowing assessment of potential copy number variance in ongoing disease-association and copy number studies.

COVERAGE OF THE DATABASE OF GENOMIC VARIANTS

To determine how well each BeadChip product covers the Database of Genomic Variants, all ~2,700 regions of the database were downloaded in build 36 coordinates (February 2007).

HumanHap300-Duo

Over 52,000 SNPs on the HumanHap300-Duo lie directly within a reported CNV region (Table 1). Of the 2,714 regions, 2,406 (89%) contain at least one SNP per segment. The average number of SNPs per region size varies

(Figure 2A). For example, for a CNV region sized between 100 and 200kb, the HumanHap300-Duo contains an average of 19 SNPs. Therefore, the average SNP density is one SNP every ~7.9kb for these regions.

HumanCNV370-Duo

Over 79,000 SNPs on the HumanCNV370-Duo lie directly within a reported CNV region (Table 1). Of the 2,714 regions, 2,689 (99%) have at least one SNP per segment. The average number of markers per region varies (Figure 2B). For example, for a CNV region from 100–200kb, the HumanCNV370-Duo contains an average of 25 markers. The average marker density for this region is one marker every 6kb.

HumanHap550

Over 96,000 SNPs on the HumanHap550 lie directly within a reported CNV region (Table 1). Of the 2,714 regions, 2,555 (94%) have at least one SNP per segment. The average number of SNPs per region size varies (Figure 2C). For example, for a CNV region between 100 and 200kb, the HumanHap550 contains an average of 33 SNPs. Therefore, the average SNP density in these regions is one SNP every ~4.5kb.

HumanHap650Y

Over 111,000 SNPs on the HumanHap650Y lie directly within a CNV region (Table 1). Of the 2,714 regions, 2,562 (94%) have at least one SNP per segment. The average number of SNPs per region size varies (Figure 2D). For example, for a CNV region between 100 and 200kb, the Human650Y contains an average of 39 SNPs. As a result, the average SNP density in these regions is ~3.75kb.

FIGURE 2: PRODUCT COVERAGE OF DATABASE OF GENOMIC VARIANTS (DGV)

FIGURE 2A: AVG. MARKER COVERAGE PER CNV SIZE RANGE: HUMANHAP300*

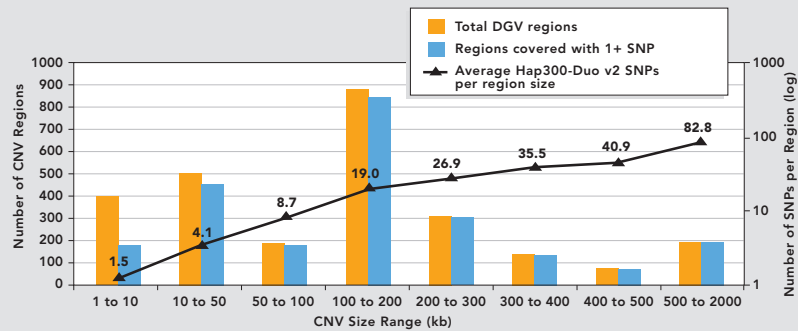


FIGURE 2B: AVG. MARKER COVERAGE PER CNV SIZE RANGE: HUMANCNV370-DUO*

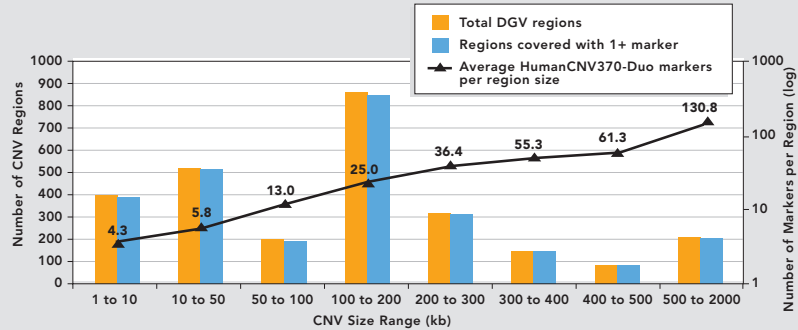


FIGURE 2C: AVG. MARKER COVERAGE PER CNV SIZE RANGE: HUMANHAP550*

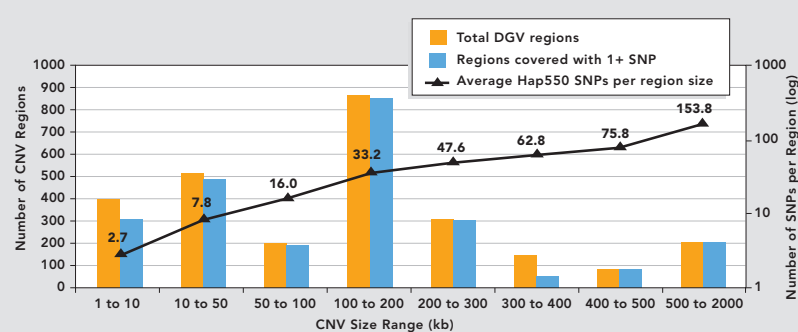
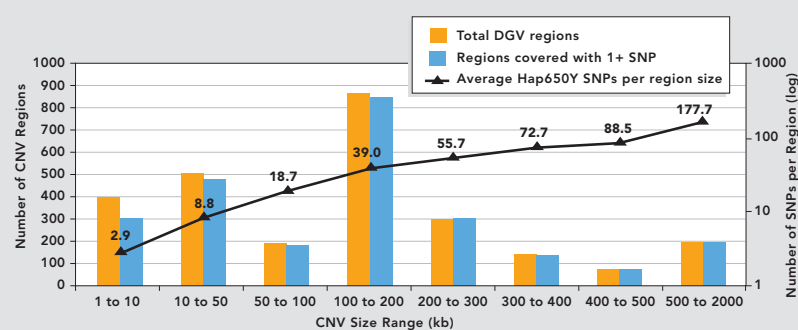


FIGURE 2D: AVG. MARKER COVERAGE PER CNV SIZE RANGE: HUMANHAP650Y*



* These graphs demonstrate coverage of the Database of Genomic Variants only. The HumanCNV370-Duo targets an additional ~9K CNV regions.

CUSTOM COPY NUMBER SOLUTIONS

Illumina also offers several custom copy number products for follow-on studies or custom selection of SNPs in CNV regions. The HumanHap300-Duo+ and HumanHap550+ allow for the addition of up to 60,800 or 121,000 custom loci, respectively. Multi-Sample iSelect™ Infinium Genotyping allows for the addition of 7,600 to 60,800 custom loci on a single BeadChip.

ILLUMINA SOLUTIONS FOR CNV ANALYSIS

With no changes to SNP content and assay format, Illumina DNA Analysis BeadChips support investigation of commonly reported copy number variants in the Human Genome on a single BeadChip. On the HumanCNV370-Duo, up to 98% of the most commonly reported copy number variants can be analyzed at a high density using the Infinium

Whole-Genome Genotyping Assay. In addition, this BeadChip targets CNV regions not available from public databases, such as segmental duplications. The combination of the flexible Infinium Assay, high-quality data, and coverage of commonly reported CNV regions enables the integrated analysis of genotyping and copy number data in disease-association CNV studies.

ORDERING INFORMATION

CATALOG NO.	PRODUCT	# SAMPLES	# BEADCHIPS	DESCRIPTION
WG-31-311	HumanHap300-Duo Whole-Genome Genotyping BeadChip	8	4	Each HumanHap300-Duo Genotyping BeadChip can process two samples and genotype > 318,000 loci per sample.
WG-31-313		48	24	
WG-31-314		96	48	
WG-31-321	HumanCNV370-Duo DNA Analysis BeadChip	8	4	Each HumanCNV370-Duo DNA Analysis BeadChip can process two samples and analyze > 370,000 loci per sample.
WG-31-322		48	24	
WG-31-323		96	48	
WG-30-521	HumanHap550 Whole-Genome Genotyping BeadChip	8	8	Each HumanHap550 Genotyping BeadChip can process one sample and genotype > 550,000 loci.
WG-30-522		24	24	
WG-30-523		96	96	
WG-30-621	HumanHap650Y Whole-Genome Genotyping BeadChip	8	8	Each HumanHap650Y Genotyping BeadChip can process one sample and genotype > 655,000 loci.
WG-30-622		24	24	
WG-30-623		96	96	

REFERENCES

- (1) Locke DP, et al. (2006) Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. *Am J Hum Genet* 79: 275-90.
- (2) Sebat J, et al. (2004) Large-scale copy number polymorphism in the human genome. *Science* 305: 525-8.
- (3) Feuk L, et al. (2006) Structural variation in the human genome. *Nat Rev Genet* 7: 85-97.
- (4) Gunderson KL, et al. (2005) A genome-wide scalable SNP genotyping assay using microarray technology. *Nat Genet* 37: 549-54.
- (5) Steemers FJ, et al. (2006) Whole-genome genotyping with the single-base extension assay. *Nat Methods* 3: 31-3.
- (6) Peiffer DA, et al. (2006) High-resolution genomic profiling of chromosomal aberrations using Infinium Whole-Genome Genotyping. *Genome Res* 16: 1136-48.
- (7) <http://projects.tcag.ca/variation/>
- (8) Iafrate AJ, et al. (2004) Detection of large-scale variation in the human genome. *Nat Genet* 36: 949-51.
- (9) Redon R, et al. (2006) Global variation in copy number in the human genome. *Nature* 444: 444-54.

ADDITIONAL INFORMATION

Visit our website or contact us at the address below to learn more about CNV analysis with Infinium Whole-Genome Genotyping and custom Infinium products.

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
www.illumina.com

FOR RESEARCH USE ONLY

© 2007 Illumina, Inc.
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. 370-2006-028 13Mar07

