AFFYMETRIX®

Data Sheet

GeneChip® Human Mapping 100K Set

The GeneChip® Human Mapping 100K Set is part of a growing family of products for copy number analysis. This two-array set provides the highest resolution copy number analysis using the proven, easy-to-use GeneChip® Mapping Assay. With a mean marker distance of 26 kb, the Mapping 100K Set provides the highest genomic coverage, giving researchers more power to detect changes in chromosomal copy number. The Mapping 100K Set is the only technology to provide both copy number and allele-specific information, allowing researchers to discover copy neutral changes.

Introduction

The GeneChip® Mapping 100K Set is part of a growing family of products for copy number analysis. It uses the same proven technology as the GeneChip Mapping 500K Array Set and the GeneChip Mapping 10K Array, and enables researchers to generate 100,000 genotypes with the easy-to-use, one-primer GeneChip® Mapping Assay, which has been validated in over 50 published studies.

PROVEN ASSAY

The GeneChip Mapping Assay for the Mapping 100K Set builds on the proven and simple approach for reducing complexity of the genome that is employed by the Mapping 10K Array.

Total genomic DNA (250 ng) is digested with a restriction enzyme (XbaI or HindIII) and ligated to adaptors that recognize the cohesive four basepair (bp) overhangs. All fragments resulting from restriction enzyme digestion, regardless of size, are substrates for adaptor ligation. A generic primer that recognizes the adaptor sequence is used to amplify adaptor ligated DNA fragments. PCR conditions have been optimized to preferentially amplify fragments in the 250 to 2,000 bp size range. The amplified DNA is then fragmented, labeled, and hybridized to the Mapping 100K Set.

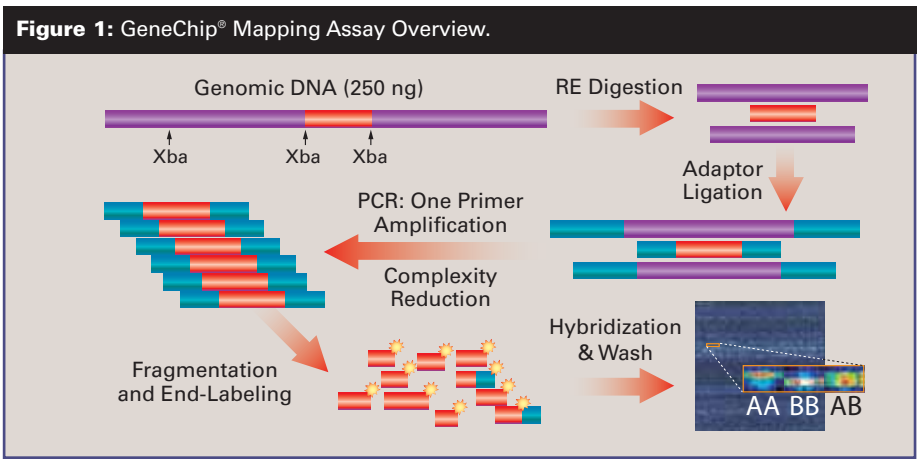
USE LESS DNA STARTING MATERIAL

Each array in the GeneChip Mapping 100K Set requires only 250 ng genomic DNA as starting material, making the most of precious sample resources.

HIGH RESOLUTION COVERAGE OF GENOME

The Mapping 100K Set is comprised of two arrays, each with greater than 50,000 SNPs. One array uses the XbaI restriction enzyme, while the second uses HindIII. Together, the family of GeneChip Mapping products offers researchers solutions for genotyping 10,000, 50,000, 100,000, 250,000, or 500,000 SNPs for a variety of applications, including linkage and association studies,

Figure 1: GeneChip® Mapping Assay Overview.



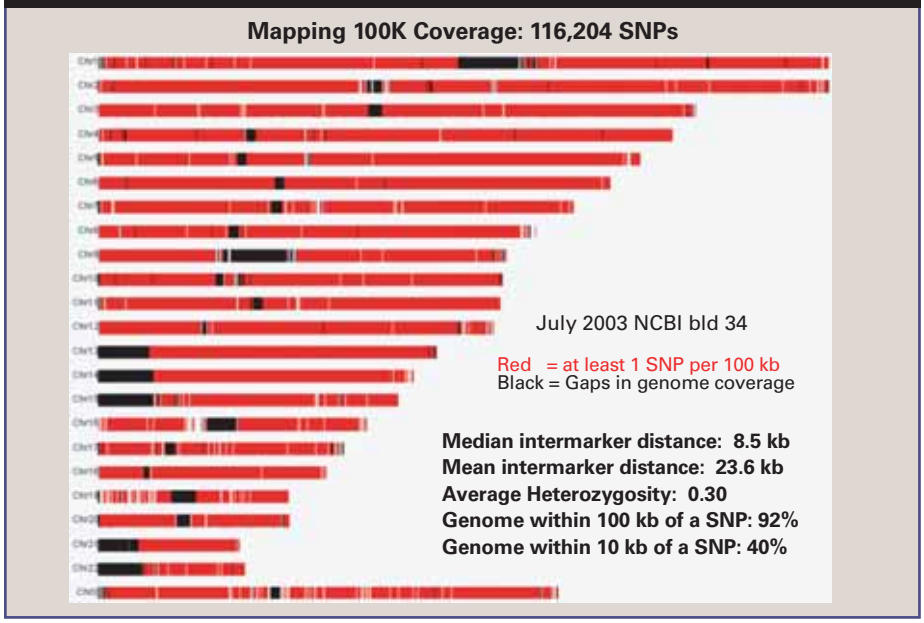
as well as cancer and population genetics. The new GeneChip Mapping 500K Array Set provides more power to understand the genetic differences between individuals, and is the recommended product for association studies.

Approximately half of the SNPs on the Mapping 100K Set are from the public domain, while the other half are from the SNP database discovered by Perlegen Sciences, Inc.

All SNPs on the Mapping 100K Set went through a rigorous screening and validation process. The optimal SNPs in this group were selected based on accuracy, call rate, and physical distribution across the genome. The median physical distance between SNPs is 8.5 kb and the average distance between SNPs is 23.6 kb. The average heterozygosity of these SNPs is 0.30.

The Mapping 100K Set provides broad coverage of the human genome: 92 percent of the genome is within 100 kb of a SNP, 83 percent of the genome is within 50 kb of a SNP, and 40 percent of the genome is within 10 kb of a SNP. For more information about the SNPs on the arrays, please visit the NetAffx™ Analysis Center.

Figure 2: Genome Coverage of Mapping 100K SNPs by chromosome. Black areas represent gaps in the human genome sequence, primarily centromeres and telomeres.



HIGHLY ACCURATE AND REPRODUCIBLE

In order to determine the accuracy and reproducibility of the Mapping 100K Set, 30 Caucasian trios were genotyped with four different technologies as part of the HapMap Project. These reference genotypes were used to measure concordance of genotypes generated with the Mapping 100K Set. 13,056 SNPs overlapped between the HapMap public data release #4 and the Mapping 100K Set. Concordance of genotypes generated with the Mapping 100K Set with the reference data was 99.73 percent across greater than 1.1 million genotypes. Mendelian inheritance consistency was measured at 99.97 percent over ten trios. Reproducibility was measured at 99.99 percent when calculated for three different individuals with five replicates each. Genotype calls were made with a confidence score cut off of 0.25 for all of the measurements above.

AUTOMATED GENOTYPE CALLS IN SOFTWARE

The GeneChip Mapping 100K Set is used in conjunction with GeneChip® Genotyping Analysis Software (GTYPE) 4.0, which uses an automated genotype-calling algorithm that provides a confidence score for each individual genotype.

Customers can expect a call rate of greater than or equal to 95 percent when using DNA of reasonable quality. Internal validation studies using over 354 samples have demonstrated a call rate of over 99 percent over a variety of populations, including African-American, Asian, and Caucasian.

GTYPE 4.0 SUPPORTS A VARIETY OF MAPPING STUDIES

GTYPE 4.0 is part of the GeneChip Mapping Array System, specifically designed to give highly accurate, automated SNP allele calls for the GeneChip Mapping Arrays. In addition, GTYPE was developed to enable workflows for a variety of SNP mapping applications.

SOFTWARE ENABLES ANALYSIS OF BOTH COPY NUMBER AND ALLELE SPECIFIC INFORMATION

The GeneChip® Chromosome Copy Number Analysis Tool (CNAT) implements an algorithm to identify genome-wide chromosomal gains and losses using the GeneChip Mapping Arrays and Mapping Assay.[1] The Mapping 100K Set, in combination with CNAT, enables researchers to detect copy number alterations, loss of heterozygosity (LOH), and genotypes in a single experiment.

ADJUSTABLE ALGORITHM ALLOWS 99% CALL RATES AT 99% ACCURACY

GTYPE 4.0 uses an automated, model-based algorithm (called the Dynamic Model), which uses a *p*-value-based confidence score to make genotype calls. In a study conducted at Affymetrix, the default value of 0.25 gives a call rate of 99.14 percent for 30 HapMap Trios (90 Caucasian samples). Concordance was 99.73 percent for 13,056 SNPs in common between the Mapping 100K Set and the HapMap public data release #4. The confidence score can be adjusted to allow genotyping with either greater accuracy or higher call rates, depending on what is needed for the application of interest. At a confidence score of 0.4, the experiment above has a call rate of 99.65 with 99.68 concordance. At a confidence score of 0.05, the call rate is 95.96 percent, with 99.77 percent concordance.

SNP ANNOTATION AVAILABLE THROUGH THE NETAFFX™ ANALYSIS CENTER

Extensive annotation for each SNP is provided in both GTYPE 4.0 and in the NetAffx™ Analysis Center. This annotation combines data from multiple sources within the public domain and consolidates them into a single database, providing a level of standardization that facilitates collaboration and sharing of data. Open access to SNP lists and annotations, as well as to research methods and validation studies, provides researchers with the greatest flexibility in their research, and enables them to troubleshoot results or follow-up genotyping experiments with downstream analyses.

SNP annotation includes dbSNP ID, nearest microsatellite markers, nearest gene, physical map location, cytoband, genetic map location, and allele frequencies in multiple populations.

ARRAY SPECIFICATIONS

Each array in the Mapping 100K Set includes more than 2.5 million features, each consisting of more than one million copies of a 25-bp oligonucleotide probe of a defined sequence, synthesized in parallel by proven photolithographic manufacturing. Each SNP is interrogated by 10 probe quartets where each probe quartet is comprised of a Perfect Match and a Mismatch probe for each allele. In total, there are 40 different 25 bp oligonucleotides per SNP.

NEW APPLICATIONS BUILT FOR THE GENECHIP® 2.0 PLATFORM

The GeneChip Mapping 100K Set runs on the GeneChip® 2.0 Platform and uses the GeneChip® Scanner 3000 with the complimentary High-Resolution Update, the GeneChip® Fluidics Station 450, and GTYPE 4.0. The Mapping 100K Set is also compatible with the previous generation GeneChip Fluidics Station 400.

REAGENT KITS VALIDATED AS PART OF THE MAPPING 100K SET

Two reagent kits are included with the Mapping 100K Set. One kit is specific to the XbaI restriction enzyme while the other is designed for the HindIII restriction enzyme. Both kits contain validated and qualified reagents for the most critical steps in the GeneChip Mapping Assay. This includes the PCR primer and adaptor necessary to selectively amplify a portion of the human genome, reagents to fragment and label the PCR products, and several control reagents. The kit was developed and validated during the process of developing the Mapping 100K Set.

REFERENCE GENOMIC DNA SERVES AS A PROCESS CONTROL

Each assay kit contains a sample of human genomic DNA to serve as a control for the entire process from DNA to data, as well as for troubleshooting. In addition, Affymetrix provides the consensus genotypes for this sample from nine independent replicates.

BUILT-IN CONTROLS TO CROSS-CHECK FOR CONSISTENCY

Thirty-one SNPs on both the XbaI and HindIII arrays serve as built-in controls for the array set. These controls allow researchers to cross-check genotypes from the same sample on each array to verify that both arrays remain together through array preparation protocols and data analysis.

REAGENT KIT PACKAGING DESIGNED TO MINIMIZE DNA CROSS CONTAMINATION

As with all PCR applications, DNA contamination is a concern, as it can lead to genotyping errors and, therefore, a reduction in genetic power. Each GeneChip Mapping 100K Assay Kit is subdivided into three boxes to support a recommended workflow designed to minimize the possibility of DNA contamination. Additionally, the GTYPE 4.0 software provides a report to help identify samples that may have otherwise undetected DNA contamination.

SAMPLE THROUGHPUT

With a standard instrument configuration of one scanner and four fluidics stations with four runs per day, a user can process 3.2 million genotypes per day. The modular GeneChip System can be easily expanded to accommodate high-throughput needs enabling tens of thousands of samples per year. For example, fluidics stations can be daisy-chained together, and GeneChip Scanner 3000s and AutoLoaders can be added to accommodate higher levels of throughput. While the fluidics stations and scanners are centrally controlled by the GeneChip® Operating Software (GCOS) platform, the GCOS Server is also available for moderate- to high-throughput analysis capabilities.

100,000 SNPS. PART OF A GROWING FAMILY FOR COPY NUMBER ANALYSIS.

The Mapping 100K Set allows 100,000 genotypes in a single experiment, and is the only technology to provide both copy number and allele specific information, allowing researchers to discover copy neutral changes. The high resolution, ease of use of the assay, and low cost per sample now make it possible to be used in copy number as well as linkage analysis and association studies.

REFERENCES

[1] **Whole Genome DNA Copy Number Changes Identified by High Density Oligonucleotide Arrays**
Huang, J., *et al. Human Genomics* **1**(4):287-99 (2004).

Genotyping over 100,000 SNPs on a Pair of Oligonucleotide Arrays
Matsuzaki, H., *et al. Nature Methods* **1**:109-111 (2004).

Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma
Levi A. Garraway, *et al. Nature* **436**(7047):117-22 (2005 Jul 7).

Genome Wide Linkage Analysis of Bipolar Disorder Using High Density Single Nucleotide Polymorphisms (SNP) Genotyping Arrays: A Comparison with Microsatellite Markers and the Finding of a Significant Linkage to Chromosome 6q22
Middleton, F.A., *et al. American Journal of Human Genetics* **5**:886-97 (2004).

Genomic Alterations in Cultured Human Embryonic Stem Cells
Maitra *et al. Nature Genetics* (2005)

Homozygous Deletions and Chromosome Amplifications in Human Lung Carcinomas Revealed by Single Nucleotide Polymorphism Array Analysis
Xiaojun Zhao, *et al. Cancer Res* **65**(13) (2005).

Critical Specifications

Number of SNPs	116,204
Number of Array	2
DNA Required/Array	250 ng
Call Rate	≥95%
Reproducibility	99.96%
Observed HapMap Concordance	99.73%
Average MAF	0.22
Average Heterozygosity	0.30
PCR Primers	1 per sample
PCR Reactions/Array	3 per sample
Percent of Genome Within 100 kb of a SNP	92%
Instrumentation	GeneChip® 2.0 Instrument Platform
Throughput	3.2 million genotypes per day with four GeneChip® Fluidics Stations 450

GeneChip® Mapping 100K Set Components

Adaptor, Xba, or Adaptor, Hind	Two annealed oligonucleotides specific for ligation to the Xbal or HindIII restriction site.
PCR Primer 001	PCR primer to amplify ligated genomic DNA
Reference Genomic DNA, 103	Human genomic DNA control, with consensus genotypes
GeneChip® Fragmentation Reagent	DNaseI enzyme, formulated to fragment purified PCR amplicons
10X Fragmentation Buffer	Buffer for fragmentation reaction
GeneChip® DNA Labeling Reagent (7.5 mm)	Proprietary biotin-labeled reagent for end-labeling fragmented PCR amplicons
Terminal Deoxynucleotidyl Transferase	Enzyme used to end-label fragmented PCR amplicons with the GeneChip® DNA Labeling Reagent
5X Terminal Deoxynucleotidyl Transferase Buffer	Buffer for labeling reaction
Oligo Control Reagent, 0100	Mixture of five biotin-labeled oligonucleotides, which hybridize to control regions (gridding and array controls) on the GeneChip® Mapping 50K Arrays

Ordering Information

GeneChip® Human Mapping 100K Set and Assay Kit

GeneChip® Human Mapping 50K Array Xba 240

900518 *Contains 30 Mapping 50K Xba Arrays*

GeneChip® Human Mapping 50K Array Hind 240

900523 *Contains 30 Mapping 50K Hind Arrays*

Supporting Products

GeneChip® Mapping 50K Xba Assay Kit

900520 *Sufficient for 30 reactions*

GeneChip® Mapping 50K Hind Assay Kit

900521 *Sufficient for 30 reactions*

GeneChip® Genotyping Analysis Software (GTYPE)

690051 *GeneChip® Genotyping Analysis Software*

AFFYMETRIX, INC.

3380 Central Expressway
Santa Clara, CA 95051 USA
Tel: 1-888-DNA-CHIP (1-888-362-2447)
Fax: 1-408-731-5441
sales@affymetrix.com
support@affymetrix.com

AFFYMETRIX UK Ltd

Voyager, Mercury Park
Wycombe Lane, Wooburn Green,
High Wycombe HP10 0HH
United Kingdom
UK and Others Tel: +44 (0) 1628 552550
France Tel: 0800919505
Germany Tel: 01803001334
Fax: +44 (0) 1628 552585
saleseurope@affymetrix.com
supporteurope@affymetrix.com

AFFYMETRIX JAPAN K.K.

Mita NN Bldg., 16 F
4-1-23 Shiba, Minato-ku,
Tokyo 108-0014 Japan
Tel: +81-(0)3-5730-8200
Fax: +81-(0)3-5730-8201
salesjapan@affymetrix.com
supportjapan@affymetrix.com

www.affymetrix.com Please visit our web site for international distributor contact information.

For research use only. Not for use in diagnostic procedures.

Part No. 701674 Rev. 3