AFFYMETRIX®

Data Sheet

GeneChip® Human Mapping 500K Array Set

The GeneChip® Human Mapping 500K Array Set enables highly powered whole-genome association studies across different populations. It uses a proven and standardized approach with flexible options for data analysis. It is comprised of two arrays which enable genotyping more than 500,000 SNPs with a single primer.

The Mapping 500K Array Set builds on the proven approach used by the Mapping 100K Set for whole-genome association. With over five times the SNP content, and over twice the genetic power, the Mapping 500K Array Set enables truly high-powered, whole-genome association studies.

As was the case with the Mapping 100K Set, data from the Mapping 500K Array Set will be freely available as part of the International HapMap project. This allows users to access and analyze the data to determine for themselves how much genetic power this panel can provide for their experiments.

The proven GeneChip® Mapping Assay, which was validated in over 80 publications in the three years since its launch, is also used with the Mapping 500K Array Set.

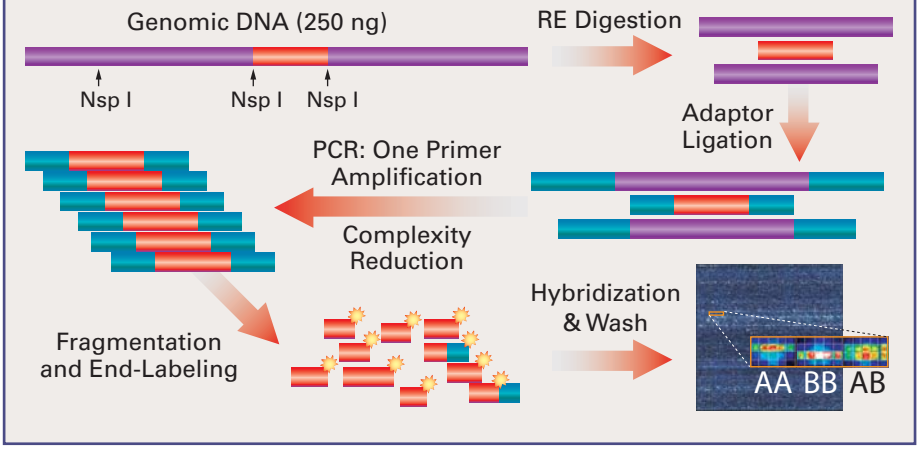
Introduction

The GeneChip® Human Mapping 500K Array Set enables highly powered whole-genome association studies across different populations. It uses the same easy-to-use assay as the GeneChip® Mapping 10K Array and GeneChip® Mapping 100K Set that has been validated in over 80 published studies in less than three years. The Mapping 500K Array Set is comprised of two arrays, each capable of genotyping on average 250,000 SNPs. One array uses the Nsp I restriction enzyme (~262,000 SNPs), while the second uses Sty I (~238,000 SNPs). Together, the family of GeneChip Mapping products offers solutions for researchers genotyping 10,000, 50,000, 100,000, 250,000, or 500,000 SNPs for a variety of applications, including linkage and association studies, as well as cancer and population genetics.

PROVEN ASSAY NOW AVAILABLE IN 96-WELL PROTOCOL

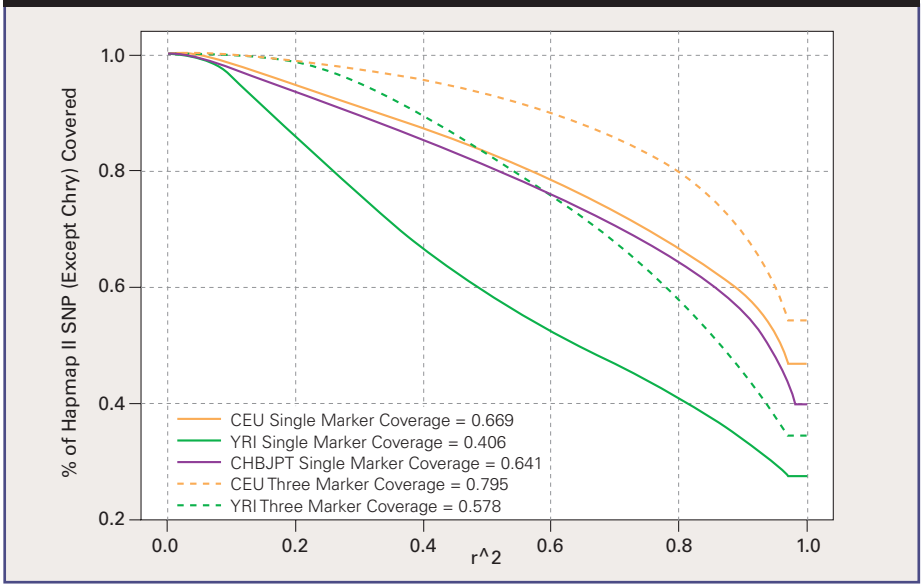
The GeneChip® Mapping Assay for the Mapping 500K Array Set builds on the proven and simple approach for reducing genomic complexity that is employed by the GeneChip Human Mapping 10K Array and the GeneChip Mapping 100K Set. Total genomic DNA (250 ng) is digested with a restriction enzyme (Nsp I or Sty I) and ligated to adaptors that recognize the cohesive four base-pair (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 200 to 1,100 bp size range. The amplified DNA is then fragmented, labeled, and hybridized to a GeneChip

Figure 1: GeneChip® Mapping Assay Overview.



Human Mapping 250K Array. In the latest update of the manual, this protocol is now available for use in 96-well plates, enabling faster and more consistent ramp-up to high-throughput sample processing.

Figure 2: HapMap Phase II coverage of the Mapping 500K Array Set across three populations, using a +/- 200kb sliding window. The x axis reports the r^2 value, and the y axis has the % HapMap Phase II SNPs (MAF>5%) in LD with Mapping 500K Array SNPs at that r^2 value. Three-marker coverage uses a multi-marker technique first developed in Haploview at the Broad Institute. Ceu = Ceph Utah, YRI = Yoruba, CHB/JPT = Chinese/Japanese.



SNP SELECTION AND GENOME COVERAGE

All SNPs on the GeneChip Human Mapping 500K Array Set went through a rigorous screening and validation process. Optimal SNPs were selected and tiled on arrays based on accuracy, call rate, and linkage disequilibrium analysis in three populations across the genome. The median physical distance between SNPs is 2.5 kb and the average distance between SNPs is 5.8 kb. The average heterozygosity of these SNPs is 0.30. Eighty-five percent of the human genome is within 10 kb of a SNP. Figure 2 shows the genetic coverage of the Mapping 500K Array Set across three populations.

GENOTYPES AVAILABLE IN HAPMAP DATABASE

With the release of the Phase II HapMap data, over 70 percent of the SNPs will be genotyped using Affymetrix technology, including data generated at Perlegen Sciences, data generated at Baylor using the Targeted Genotyping Assay, and data from the Mapping 100K Set. Data from the Mapping 500K Array Set will be available in HapMap release 21. Open access to SNP lists and raw genotype data is a critical part of facilitating standardization, as well as giving researchers the ability to determine for themselves the power of the genetic coverage for their experiments. For the most current 500K data on the HapMap samples, please visit the Affymetrix web site: www.affymetrix.com/support/technical/sample_data/500k_hapmap_genotype_data.affx.

SNP ANNOTATION AVAILABLE THROUGH THE NETAFFX™ ANALYSIS CENTER

Extensive annotation for each SNP is provided in both GeneChip® Genotyping Analysis Software (GTYPE) and the NetAffx™ Analysis Center. This annotation combines data from multiple sources within the public domain and consolidates it into a single database. SNP annotation includes dbSNP ID, nearest gene, physical map location, cytoband, and allele frequencies in multiple populations.

500,000 SNPS. HIGHLY POWERED WHOLE-GENOME ASSOCIATION STUDIES

The Mapping 500K Array Set genotypes more than 500,000 SNPs in a single experiment, enabling researchers to conduct highly powered genome-wide studies pertaining to disease genetics, drug response, and linkage disequilibrium. All the SNPs used on the Mapping 500K Array Set have been validated across a wide range of populations. The Mapping 500K Array Set SNP content builds on the Mapping 100K Set which has been successfully used in many whole-genome association studies [1]. The genetic coverage of the Mapping 500K Array Set is shown in Figure 2. Future products will have complementary sets of SNPs to further increase genetic power across many population groups.

USE LESS DNA STARTING MATERIAL

Each array in the Mapping 500K Array Set requires only 250 ng genomic DNA as starting material, making the most of precious sample resources. Whole-genome amplified material prepared by the Qiagen REPLI-g® kits may also be used with the Mapping 500K Array Set.

AUTOMATED GENOTYPE CALLS IN SOFTWARE

The GeneChip Human Mapping 500K Array Set is used in conjunction with GeneChip® Genotyping Analysis Software (GTYPE), which uses an automated, model-based genotype-calling algorithm that provides a confidence score for each individual genotype. The Dynamic Model algorithm (DM), [2] generates a QC call rate for each array. A call rate of greater than or equal to 93 percent (at a DM confidence threshold of 0.33), when using good quality DNA, should be used to determine whether a sample should be repeated or used for downstream analysis. The user may adjust the default confidence score to allow genotyping with either greater accuracy or higher call rates, depending on what is needed for the application of interest [3].

IMPROVED CALL RATES AND ACCURACY WITH NEW GENOTYPE CALLING ALGORITHM

The Bayesian Robust Linear Model with Mahalanobis distance classifier (BRLMM) algorithm is an extension of the Robust

Linear Model with Mahalanobis distance classifier (RLMM) algorithm developed for Mapping 100K Set arrays, and uses data from multiple arrays to make genotype calls [4, 5]. It provides a significant improvement over DM in two important areas: it improves overall performance (call rates and accuracy), and it equalizes the performance of homozygous and heterozygous genotypes. To test the performance of BRLMM, 66 samples from the international HapMap project (33 Caucasians and 33 Yorubans) that were run at a customer site were re-analyzed using BRLMM at a setting of 0.5. The average call rate was 98.8 percent, with a concordance of 99.3 percent. These samples also included 17 trios, on which the Mendelian inheritance consistency was found to be 99.9 percent. Reproducibility was measured as 99.9 percent when calculated for 7 samples run in duplicate at the Affymetrix® Services Laboratory. Each sample in the above analyses passed the 93 percent QC DM 0.33 cutoff score (the average DM 0.33 call rate on these data was 96 percent).

The BRLMM Analysis Tool (BAT) enables analysis using the BRLMM algorithm. BAT is available as a standalone software tool from the Affymetrix web site. The user may adjust the default confidence score to allow genotyping with either greater accuracy or higher call rates, depending on what is needed for the application of interest [5].

Customer-generated data on 66 samples analyzed with BRLMM (0.5)	
Observed call rate	98.8%
Observed concordance with HapMap	99.3%
Observed Mendelian Consistency	99.9

GTYPE FLEXIBLE DATA EXPORT OPTIONS

GTYPE incorporates advanced functionality that supports multiple options for data export, including export by chromosome, filtering by allele frequency, Hardy-Weinberg Equilibrium, and Mendelian-error. GTYPE provides new batch SNP export options allowing researchers to automatically export genotype calls from thousands of samples at a time into tab-delimited files, as well as into third-party-compatible formats such as MERLIN, GeneHunter open source software, and Haploview.

ARRAY SPECIFICATIONS

Each array in the Mapping 500K Array Set includes more than 6.5 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 6- or 10-probe quartets where each probe quartet is comprised of a perfect match and a mismatch probe for each allele. In total, there are 24 or 40 different 25-bp oligonucleotides per SNP.

REAGENT KITS VALIDATED AS PART OF THE MAPPING 500K ARRAY SET

Two reagent kits were developed and validated for use in conjunction with the Mapping 500K Array Set. One kit is specific to the Nsp I restriction enzyme while the other is designed for the Sty I 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. Kits are available for either 30 or 100 reactions.

BUILT-IN CONTROLS TO CROSS-CHECK FOR CONSISTENCY

Fifty SNPs on both the Nsp I and Sty I arrays serve as built-in controls for the Mapping 500K Array Set. GTYPE leverages these controls in the Sample Mismatch Report to cross-check genotypes from the same sample on each array to verify that both arrays remain together from DNA preparation to data analysis. In addition, GTYPE uses additional controls to cross-check Mapping 500K arrays with previous generation arrays to allow customers who are using the Mapping 500K Array Set to increase power in ongoing studies.

REAGENT KIT PACKAGING DESIGNED TO MINIMIZE DNA CONTAMINATION

As with all PCR applications, DNA contamination is a concern, as it can lead to genotyping errors, thereby reducing genetic power. Each GeneChip Mapping 500K Assay Kit is subdivided into three boxes to support a recommended workflow designed to minimize the possibility of DNA contamination. Additionally, the GTYPE software provides a report to help identify samples that may have otherwise-undetected DNA contamination.

SAMPLE THROUGHPUT

With a standard instrument configuration of one GeneChip® Scanner 3000 7G with AutoLoader, and three GeneChip® Fluidics Station 450s with four runs, a user can process more than 10 million genotypes per day. The modular GeneChip System can be easily expanded to accommodate high-throughput needs, enabling analysis of thousands of samples per year. For example, fluidics stations can be daisy-chained together. 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 to collect data from multiple GeneChip Scanner 3000 7G scanners.

REFERENCES

1. Klein R.J., *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science* **308**(5720):385-9 (2005 Apr 15).
2. Di X., *et al.* Dynamic model based algorithms for screening and genotyping over 100K SNPs on oligonucleotide microarrays. *Bioinformatics* **21**(9):1958-63 (2005 May 1).
3. Matsuzaki, *et al.* Genotyping over 100,000 SNPs on a Pair of Oligonucleotide Arrays. *Nature Methods* **1**:109-111 (2004).
4. Nusrat Rabbee and Terence P. Speed. A genotype calling algorithm for Affymetrix SNP array. Bioinformatics Advance Access published online on November 2, 2005.
5. BRLMM: an Improved Genotype Calling Method for the GeneChip® Human Mapping 500K Array Set. Affymetrix, Inc. (*forthcoming*)

Product Information

Number of SNPs	500,568
Number of Arrays	2
DNA required/array	250 ng
QC Call rate (DM 0.33)	≥93%
Expected BRLMM call rate (0.5)	≥96%
Average MAF	0.22
Average heterozygosity	0.30
PCR Primers	1 per sample
PCR Reactions/array	3 per sample
Percent of genome within 10 kb of a SNP	85%
Instrumentation	GeneChip® Scanner 3000 7G with AutoLoader
Throughput	>10 million genotypes per day with three GeneChip® Fluidics Station 450s

GeneChip® Mapping 250K Assay Kit Components

Adaptor, Nsp I or Adaptor, Sty I	Two annealed oligonucleotides specific for ligation to the Nsp I or Sty I restriction site.
PCR Primer 002	PCR primer to amplify ligated genomic DNA
Reference Genomic DNA, 103	Human genomic DNA control, with consensus genotypes
GeneChip® Fragmentation Reagent	DNase I enzyme, formulated to fragment purified PCR amplicons
10X Fragmentation Buffer	Buffer for fragmentation reaction
GeneChip® DNA Labeling Reagent (30mM)	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 250K arrays

Ordering Information

GeneChip® Human Mapping 250K Nsp Array

520330	*Contains 1 Mapping 250K Nsp Array (200 order minimum)*
900768	*Contains 30 Mapping 250K Nsp Arrays*

GeneChip® Human Mapping 250K Sty Array

520331	*Contains 1 Mapping 250K Sty Array (200 order minimum)*
900770	*Contains 30 Mapping 250K Sty Arrays*

GeneChip® Mapping 250K Nsp Assay Kit

900766	*Sufficient for 30 reactions*
900753	*Sufficient for 100 reactions*

GeneChip® Mapping 250K Sty Assay Kit

900765	*Sufficient for 30 reactions*
900754	*Sufficient for 100 reactions*

GeneChip® Genotyping Analysis Software (GTYPE)

690051	*GeneChip® Genotyping Analysis Software*

AFFYMETRIX, INC.

3420 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. 702087 Rev. 4