AFFYMETRIX®

Technical Note

SNP Concordance Across Affymetrix Genotyping Arrays

This Technical Note describes the concordance among Affymetrix genotyping arrays—the Genome-Wide Human SNP Array 6.0, Genome-Wide Human SNP Array 5.0 and Human Mapping 500K Array Set—and provides a detailed description of the SNPs that have been excluded from the next-generation arrays.

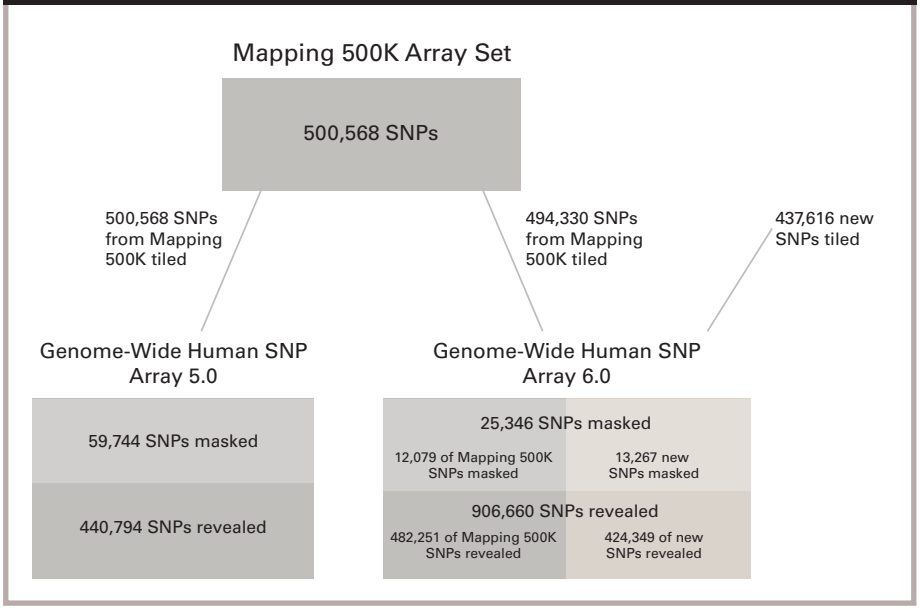
Introduction

The Genome-Wide Human SNP Array 6.0 is the latest addition to Affymetrix' family of genotyping products. The array enables researchers to conduct high-performance analysis of more than 1.8 million markers for genetic variation, including 906,000 SNPs and 946,000 copy number probes. The SNP Array 6.0 demonstrates industry-leading performance, with average call rates greater than 99 percent and near-saturated coverage of human genetic variation.

The SNP Arrays 5.0 and 6.0 were preceded by the Human Mapping 500K Array Set, which contained 500,568 SNPs on two arrays. Some SNPs on the 500K Array Set were excluded in the development of the next-generation SNP Array 6.0. Additionally, some SNPs were retained on the SNP Arrays 5.0 and 6.0 but were masked out in the default library files. One-hundred percent of the SNPs from the 500K Array Set are present on the SNP Array 5.0, and 99 percent are present on the SNP Array 6.0.

Figure 1 presents an overview of the SNP content of the three products. The common SNPs demonstrate very high concordance between platforms—99.9 percent between the SNP Array 6.0 and 5.0, 99.8 percent between the SNP Array 6.0 and 500K Array Set and 99.9 percent between the SNP Array 5.0 and 500K Array Set.

Figure 1: The relationship between the SNP content of the Mapping 500K Array Set and the Genome-Wide Human SNP Arrays 5.0 and 6.0.



Excluded SNPs

The SNP Array 6.0 excludes 6,238 SNPs from the original 500,568 SNPs on the 500K Array Set. The majority of excluded SNPs are in high linkage disequilibrium (LD) with other SNPs on the array. Of the excluded SNPs, 2,157 map to more than one location in NCBI genome build 35 (hg17). In addition, 4,081 excluded SNPs had poor genotyping cluster properties on the SNP Array 5.0 and had a high measure of redundancy with new SNPs on the SNP Array 6.0. The exclusion of SNPs created space in the SNP Array 6.0 design for additional probes for SNPs of higher genetic importance.

A SNP on the 500K Array Set was considered redundant if it had a proxy with a high measure of LD (r^2 of greater than 0.8) among SNPs on SNP Array 6.0 for each of the three HapMap populations (CEU, CHB+JPT and YRI). The identities of the proxy SNPs are provided in three tables. One table per HapMap population[1] is available on the SNP Array 6.0 support webpage (http://www.affymetrix.com/support/technical/byproduct.affx?product=genomewidesnp_6).

Each table maps each excluded SNP to the SNP Array 6.0 SNP that produces the highest measure of LD (max r^2) for the population. Table 1 describes the five columns in the tables. SNPs for which the MAF is less than 5 percent for the population were not assigned a proxy for the population, and have values of NA in the columns. An r^2 value of less than 0.8 occurs for a few cases, where the expected proxy SNP did not make it to the final SNP Array 6.0 design.

Although most of the original SNPs were retained on the new arrays, some of the Mapping 500K SNPs were "masked" on the SNP Array 5.0 and the SNP Array 6.0 in order to produce default SNP sets of the highest quality. Of the 500,568 SNPs on the 500K Array Set, 440,794 are unmasked in the default library for the SNP Array 5.0. The unmasked subset of 500K SNPs on the SNP Array 6.0 is increased to 482,251. Genotypes of masked SNPs can be produced by running the Affymetrix Power Tools (APT) software in conjunction with the "Full" library files. The Full library files are distributed in the library file package alongside the default library files, and can be identified by the use of a .Full suffix in the file name.

APT is a suite of cross-platform APIs implementing algorithms for analyzing and working with Affymetrix GeneChip® arrays. Downloads for the latest APT package are available at www.affymetrix.com/support/developer/powertools/index.affx. Refer to the apt-probeset-genotype manual included in the APT download. Apt-probeset-genotype is the application for making genotype calls using mapping arrays (100K, 500K, Genome-Wide SNP Arrays 5.0 and 6.0). The critical steps for producing genotypes of masked SNPs are 1) specify the path to the Full library files when supplying the arguments for any of the following options: special-snps, cdf-file and chrX-snps; and 2) include the force option. Note that .chp files (the standard Affymetrix genotyping output file for a sample) produced using the force option with APT 1.8.0 (or earlier versions) cannot be loaded into the Genotyping Console software.

Table 1: Columns in tables that map excluded SNPs to SNP Array 6.0 SNPs.

Column Number	Column Name	Description
1	Probe_Set_ID	Affymetrix SNP ID of the excluded SNP
2	Physical_Position	Physical position of the excluded SNP
3	Probe_Set_ID_MaxRsqrSNP	Affymetrix SNP ID of the proxy SNP on SNP Array 6.0
4	Physical_Position_MaxRsqrSNP	Physical position of the proxy SNP
5	Max_Rsqr	r^2 value between SNPs

SNP Concordance

The common SNPs in the default libraries of the 500K Array Set, SNP Array 5.0 and SNP Array 6.0 produce high pair-wise concordance of genotypes (99.8 to 99.9 percent). Table 2 gives the pair-wise concordance of the three products for the 270 HapMap samples. Concordance was computed for sets of common SNPs in the default libraries for the products. Genotype calls were made using the algorithms and defaults provided with the mapping arrays; BRLMM (Mapping 500K), BRLMM-P (SNP Array 5.0) and Birdseed (SNP Array 6.0). SNP concordance is the fraction of the genotype calls for a given SNP, made by a pair of products, which agree over the 270 HapMap samples. "No calls" are excluded from the concordance calculation. Average call rates over the 270 HapMap experiments are given for all SNPs in the default libraries of each of the three products.

The high concordance is shared by essentially all of the common SNPs (Figure 2). Figure 2 shows the cumulative distributions of SNP concordance, where the y-axis is the fraction of SNPs whose concordance is less than the SNP concordance value on the x-axis.

Table 2: Pair-wise concordance of Affymetrix mapping arrays for 270 HapMap samples.

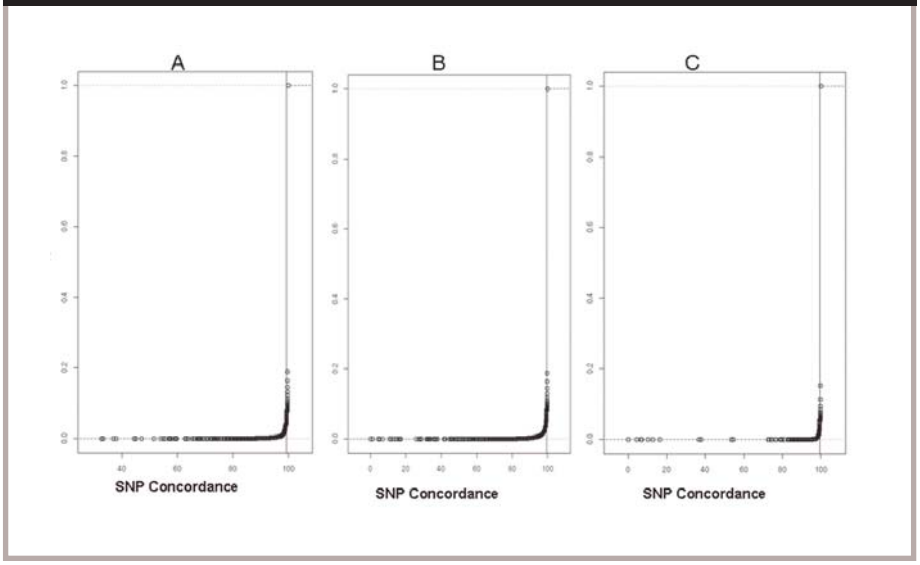
Array 1 vs. Array 2 Comparisons	Avg. Call Rate Array 1	Avg. Call Rate Array 2	N Common SNPs	Percent Concordance
Mapping 500K vs. SNP Array 5.0	99.36	99.71	440,794	99.9
Mapping 500K vs. SNP Array 6.0	99.36	99.86	482,251	99.8
SNP Array 5.0 vs. 6.0	99.71	99.86	436,218	99.9

Conclusion

The Genome-Wide Human SNP Array 6.0 contains nearly twice the content of the SNP Array 5.0 and nearly three times that of the Mapping 500K Array Set. The SNP Array 6.0 added more than 400,000 new SNPs and retained 99 percent of the SNPs on its two predecessors. Most SNPs that were excluded in the transition from the Mapping 500K Array Set to the next-generation SNP Arrays 5.0 and 6.0 are represented by other SNP proxies. Genotypes of SNPs that are masked in the default libraries for SNP Arrays 5.0 and 6.0 can be produced by running the APT software.

The high coverage and performance of the SNP Array 6.0, and high concordance with previous-generation arrays, enables researchers to design high-powered, whole-genome studies that can detect new associations at an unprecedented throughput.

Figure 2: Cumulative distributions of SNP concordance between products. (A) Mapping 500K and SNP Array 5.0 concordance; (B) Mapping 500K and SNP Array 6.0 concordance; (C) SNP Array 5.0 and SNP Array 6.0 concordance.



REFERENCES

1. RedundantCeu_4081.xls (Ceph), RedundantYri_4081.xls (Yoruban), RedundantAsi_4081.xls (CHB+JPT)

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. 702550 Rev. 1