



# Application Note

## A New Paradigm for Two-stage Association Studies

Affymetrix' high-density, low-cost Genome-Wide Human SNP Array 6.0 enables a completely new approach to the design of whole-genome association studies—a joint analysis of multiple, large sample sets using a whole-genome marker panel.

Researchers have traditionally used a two-stage approach. First, a complete SNP panel is screened in a subset of samples; then a subset of SNPs is screened across a larger number of samples. This approach has several drawbacks, including an under-powered initial screen, SNP drop-out when converting to a subset panel and time delays between stages.

The SNP Array 6.0 allows an integrated approach—analyzing multiple, large sample sets on a single, complete whole-genome panel—thereby increasing the overall genetic power of the study and accelerating the gene discovery process.

### Introduction

Genetic association studies aim to identify DNA variations underlying a disease predisposition, drug response or diagnostic/prognostic outcome. The development of high-density microarrays for genetic variation analysis, such as the Affymetrix® Genome-Wide Human SNP Arrays 5.0 and 6.0, has already enabled studies of common diseases such as diabetes and heart disease<sup>1,2,3</sup>.

Ideally, a genetic association study involves testing for association in a comprehensive, whole-genome panel of up to 1 million SNPs. It also involves screening as many samples as possible, because the underlying genetic associations are often subtle, and in most circumstances the power of the study is directly proportional to the number of samples.

However, the task of screening for genetic associations has traditionally been split into a two-stage effort because the ideal design—a maximally powered screen of all SNPs in all available samples—has typically been too expensive. The usual compromise is to split the study into two stages.

### Drawbacks of a Two-stage Approach

In the first stage, a complete SNP panel is screened in a subset of samples to identify a subset of SNPs that meet a permissive statistical significance threshold. In the second stage, a subset of SNPs showing potential to be associated is typed across a large number of samples. The purpose of the second stage is to attempt to replicate the findings of the first stage. Larger sample numbers are often used in the second stage to more accurately deter-

mine the effect of the disease allele, and to ensure that sufficient power is available to confirm the stage one results.

This two-stage division has three primary drawbacks. First, the initial screen is typically under-powered because the full sample set has been split, thereby increasing the chance that true associations are missed. Secondly, some SNPs are usually lost in the switch to the smaller panel in phase two because of a failure to convert on the new panel, due to technical reasons inherent to all platforms. This can result in a loss of real associations. Lastly, the switch to a custom panel in phase two introduces a delay of six weeks to six months.

Despite these drawbacks, the two-stage approach has traditionally been a sensible choice for association studies. It maximizes power for a fixed study cost whenever the per-sample cost of typing the complete SNP panel is much larger than that of the custom subset panel.

### New Paradigm for Association Studies

Recent advances in technology have enabled Affymetrix to increase the information content of its new Genome-Wide SNP Array 6.0 by nearly a factor of four compared to previous-generation arrays. By leveraging this innovation, as well as superior manufacturing capabilities, Affymetrix has introduced the SNP Array 6.0 at a price point that enables the genetics community to conduct whole-genome association studies without the arbitrary division of a two-stage study.

The ensuing integration enables increases in the overall power—reducing the risk of false negatives in an under-powered stage one—by

increasing the number of samples analyzed on a whole-genome panel. This single-stage approach enabled by the SNP Array 6.0 also accelerates the rate at which studies can be executed.

Work by Skol, *et al.* (2006) and others demonstrates that the integrated, single-phase approach always provides greater power than both separate and joint analysis of the two-phase approach, even after taking into account the more stringent multiple hypothesis testing corrections involved.

The genetics community is already adopting the more powerful, single-phase approach that leverages the ability to run more samples on Affymetrix technology. In several recent publications identifying variants associated with Type 2 Diabetes<sup>1,2,4</sup> researchers have used Affymetrix arrays to integrate the whole-genome screen (stage one) with the statistical replication (stage two), using the same whole-genome marker set in both phases.

The traditional study design paradigm is now being replaced by joint analysis of a larger data set on a whole-genome marker set like the SNP Array 6.0. The high density of markers on the SNP Array 6.0 also contributes to this integrated, whole-genome approach by allowing researchers to directly identify a causative gene without any further fine mapping<sup>1,2,4</sup>. This approach accelerates the process from whole-genome screen to gene discovery considerably—a key factor in the competitive field of genetics.

#### **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

## **Additional Benefits**

Adopting the integrated whole-genome approach to association studies also allows different groups to combine their whole-genome data from different populations. This coordination is not possible when using a marker subset, specific to a particular study, for statistical replication. A whole-genome marker set for replication allows for cross-validation and confirmation of putative associations from individual projects, and increases the overall genetic power of the studies. This approach has been validated in the recent Type 2 Diabetes publications.

Furthermore, an integrated whole-genome association panel enables researchers to identify shared genetic causes among different disorders with overlapping phenotypes. In a recent article by the Diabetes Genetics Initiative of the Broad Institute of Harvard and MIT, Lund University and Novartis Institutes for BioMedical Research, researchers found a key variant that explains susceptibility to Type 2 Diabetes (T2DM). A follow-up publication by DeCode Genetics of Iceland found that the same variant predisposes individuals to a higher risk of myocardial infarction<sup>6</sup> (MI). Given that high blood triglyceride and cholesterol level is commonly high among both T2DM and MI patients, it is not surprising that researchers found the same predisposing alleles. An integrated whole-genome panel is the ideal solution to create such synergy among a multitude of disorders.

#### **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

## **Conclusion**

Analyzing multiple, large sample sets on a single, complete whole-genome panel increases the overall genetic power of association studies and accelerates the time to discovery of functional variants. The SNP Array 6.0 enables researchers to increase genetic power in two ways: by augmenting the number of samples screened on a whole-genome panel, and by allowing the analysis of other cohorts using the same standardized SNP panel. With more than 1.8 million genetic markers on a single microarray, only the SNP Array 6.0 enables this modern study design through high-performance and low-cost genotyping.

#### **REFERENCES:**

1. Saxena, R., *et al.* Genome-Wide Association Analysis Identifies Loci for Type 2 Diabetes and Triglyceride Levels. *Science* Epub 2007.
2. Zeggini, E., *et al.* Replication of Genome-Wide Association Signals in U.K. Samples Reveals Risk Loci for Type 2 Diabetes. *Science* Epub 2007.
3. McPherson, R. A., *et al.* Common Allele on Chromosome 9 Associated with Coronary Heart Disease. *Science Express* Epub May 3 2007.
4. Scott, L. J., *et al.* A Genome-Wide Association Study of Type 2 Diabetes in Finns Detects Multiple Susceptibility Variants. *Science* Epub 2007.
5. Skol, A. D., *et al.* A joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nature Genetics* **38**(3):390 (2006).
6. Helgadottir, A. A., *et al.* Common Variant on Chromosome 9p21 Affects the Risk of Myocardial Infarction. *Science Express* Epub May 3 2007.

#### **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 website for international distributor contact information.**

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

Part No. 702521 Rev. 1

©2007 Affymetrix, Inc. All rights reserved. Affymetrix®, , GeneChip®, HuSNP®, GenFlex®, Flying Objective™, CustomExpress®, CustomSeq®, NetAffx™, Tools To Take You As Far As Your Vision®, The Way Ahead™, Powered by Affymetrix™, and GeneChip-compatible™ are trademarks of Affymetrix, Inc. Products may be covered by one or more of the following patents and/or sold under license from Oxford Gene Technology: U.S. Patent Nos. 5,445,934; 5,700,637; 5,744,305; 5,945,334; 6,054,270; 6,140,044; 6,261,776; 6,291,183; 6,346,413; 6,399,365; 6,420,169; 6,551,817; 6,610,482; 6,733,977; and EP 619 321; 373 203 and other U.S. or foreign patents.