Publications

Key Publications on the SNP Array 5.0/6.0 Platform for Whole-genome Association Studies

- Myocardial Infarction Genetics Consortium [Kathiresan S., *et al*.]. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nature Genetics* 41:334-341 (2009).

  Key findings:
  - This publication describes the first comprehensive genome-wide association study with the Genome-Wide Human SNP Array 6.0, testing single nucleotide polymorphisms (SNPs) and copy number variants (CNVs) for association with early-onset myocardial infarction in 2,967 cases and 3,075 controls. The authors also detail the quality control metrics and thresholds that have been used for the analysis of common copy number polymorphisms (CNPs) and rare CNVs. SNPs at nine loci were reproducibly associated with myocardial infarction, but tests of common and rare CNVs failed to identify additional associations with myocardial infarction risk.

- Korn J. M., *et al*. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nature Genetics* 40(10):1253-60 (2008).

  Key findings:
  - This publication describes a new suite of algorithms for integrated genotyping of SNPs, CNPs, and detection of rare CNVs. The SNP Array 6.0 showed an average call rate of 99.47 percent for the HapMap samples and a concordance of 99.74 percent with the Birdseed algorithm. More than 50,000 patient samples have been genotyped at the Broad Institute with the SNP Array 6.0, with an average call rate of greater than 99 percent.
  - Leveraging a previously established map of CNPs, the Canary algorithm showed an average sample call rate of 96.1 percent across 1,177 di-allelic CNPs using the SNP Array 6.0. Birdseye is a powerful algorithm for identifying rare or de novo CNVs. The sensitivity and specificity of Birdseye was simulated in an in silico gender-mixing experiment. The SNP Array 6.0 showed a sensitivity of 97.5 percent for deletions spanning 10 probes, or a mean size of 17 kilobases (kb), demonstrating the power of the SNP Array 6.0 to detect copy number changes.

- McCarroll S. A., *et al*. Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nature Genetics* 40(10):1166-74 (2008).

  Key findings:
  - By characterizing 270 HapMap samples with the SNP Array 6.0, this team developed a map of human CNVs (at 2 kb breakpoint resolution) informed by integer genotypes for 1,320 CNPs. Approximately 80 percent of observed copy number differences between pairs of individuals were the result of common CNPs with an allele frequency of greater than 5 percent, and more than 99 percent derived from inheritance rather than new mutation.
  - Most commonly, di-allelic CNPs were in strong linkage disequilibrium with SNPs, and most low-frequency CNVs segregated on specific SNP haplotypes.

- Erdmann J., *et al*. New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nature Genetics* 41:280-282 (2009).

  Key findings:
  - This publication discusses a three-stage analysis of genome-wide SNP data with the SNP Array 6.0 in 1,222 German individuals with myocardial infarction and 1,298 controls, in silico replication in three additional genome-wide data sets of coronary artery disease (CAD), and subsequent replication in ~25,000 subjects. The study identified one new susceptibility locus for CAD on 3q22.3 with compelling statistical evidence and a second locus on 12q24.31 with suggestive evidence.

- The International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 455:237-241 (2008).

  Key findings:
  - This study focused on rare but highly penetrant structural variation in schizophreniaby examining 3,391 patients with schizophrenia and 3,181 ancestrally matched controls. Using the SNP Array 5.0/6.0 genotyping platform and the Birdseye algorithm to detect rare CNVs, this study supported the primary hypothesis that individuals with schizophrenia have a greater genome-wide burden of CNVs. The data also showed that CNVs in at least three loci act as strong risk factors for schizophrenia in a minority of individuals.

- Weiss, L. A., *et al*. Association between microdeletion and microduplication at 16p11.2 and Autism. *New England Journal of Medicine* 358:667-675 (2008).

  Key findings:
  - This landmark paper describes a screening technique for recurrent de novo autosomal copy number variants that could influence susceptibility to autism or other heritable complex disorders. Using the SNP Array 5.0, SNP Array 6.0, and other microarray technologies, this publication identified a novel, recurrent microdeletion and a reciprocal microduplication that carry substantial susceptibility to autism and appear to account for approximately 1 percent of cases.

- Nishida N., *et al*. Evaluating the performance of Affymetrix SNP Array 6.0 platform with 400 Japanese individuals. *BMC Genomics* 9(1):431 (2008).

  Key findings:
  - This publication looked at the performance and coverage of the SNP Array 6.0 in the Japanese population. The Birdseed algorithm accurately determined the genotype calls of each sample with an overall call rate higher 99.5 percent and a concordance rate higher than 99.8 percent using more than 48 samples. The genome coverage of the SNP Array 6.0 array was calculated to be 75 percent with the 590,248 SNPs in the Japanese population.

- Zheng W., *et al*. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nature Genetics* 41:324-328 (2009).

  Key findings:
  - This study carries out a genome-wide association study among Chinese women to identify risk variants for breast cancer. After analyzing 607,728 SNPs in 1,505 cases and 1,522 controls, the authors selected 29 SNPs for a fast-track replication in an independent set of 1,554 cases and 1,576 controls. This study is a successful example for a project that integrates data from the GeneChip® Human Mapping 500K Array Set and the SNP Array 6.0 in one powerful study.

- Pengyuan L., *et al*. Familial Aggregation of Common Sequence Variants on 15q24-25.1 in Lung Cancer. *Journal of the National Cancer Institute* 100:1326-1330 (2008).

  Key findings:
  - This publication used the Human Mapping 500K Array Set and the SNP Array 6.0 to study the familial risk of lung cancer by genotyping 194 patients with familial lung cancer and 219 cancer-free control subjects. The research identified associations between common sequence variants at 15q24-25.1 and lung cancer. The risk of lung cancer was more than five-fold higher among those subjects who had a family history of lung cancer and two copies of high-risk alleles.

- Hofmann, S., *et al*. Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. *Nature Genetics* 40:1103-1106 (2008)

  Key findings:
  - In this study, the SNP Array 5.0 was used to perform a genome-wide association study on sarcoidosis, comprising 499 cases and 490 controls. The study detected a series of genetic associations, with the strongest association signal mapping to the ANXA11 gene. Extensive fine mapping located the association signal to a region between exon 5 and exon 14 of ANXA11. A common non-synonymous SNP was found to be strongly associated with sarcoidosis.

- Bin X., *et al*. Strong association of de novo copy number mutations with sporadic schizophrenia. *Nature Genetics* 40:880-885 (2008).

  Key findings:
  - This publication demonstrates that de novo copy number mutations are collectively ~8 times more frequent in sporadic cases of schizophrenia (15 of 152) than in unaffected individuals (2 of 159), with an association that is statistically highly significant. This study used the SNP Array 5.0 across 1,077 samples with an average call rate of 99.43 percent. The authors used DNA-Chip Analyzer (dChip)10, 12, 13 and Partek (version 6.3) for whole-genome copy number data analysis.

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

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

P/N F107-2