Evaluation of NHANES DNA Bank Specimens with High-Throughput Single Nucleotide Genotyping Arrays

Christopher N. Greene, Lisa M. Keong, Patricia W. Mueller

Division of Laboratory Sciences, National Center for Environmental Health
Centers for Disease Control and Prevention, Atlanta, Georgia

Objective

The purpose of this study is to determine the suitability of the NHANES III and NHANES 99-02 DNA banks for use with the latest-generation Affymetrix Genome-Wide 5.0 and Genome-Wide 6.0 high-density Single Nucleotide Polymorphism (SNP) mapping microarrays.

Background

DNA repositories have been collected from the National Health and Nutrition Examination Survey (NHANES) NHANES III and NHANES 99-02 to add a genetic component capability to studies of U.S. public health.

The NHANES III DNA bank consists of unpurified lymphocyte cell lysates. These samples have variable DNA concentration and quality. The NHANES 99-02 DNA bank is available as highly-purified genomic DNA.

The Affymetrix Genome-Wide (GW) 5.0 and GW 6.0 Mapping Arrays genotype up to 906,000 individual SNPs and greater than 900,000 copy number variants. We are evaluating the suitability of the NHANES III and 99-02 DNA bank specimens for use with these high-throughput genetic assays.

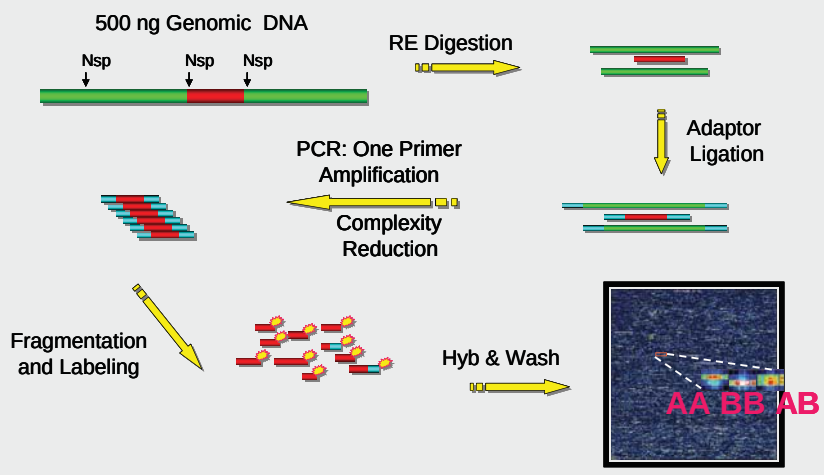
Research Design and Methods

Samples: A set of 44 NHANES III samples were genotyped with the Genome-Wide 6.0 array, including 2 sets of blinded repeats, and 8 NHANES 99-02 samples, including one blind repeat, were genotyping with the GW 5.0 and GW 6.0 arrays.

Microarray Genotyping: NHANES III samples were whole-genome amplified (WGA) by multiple displacement amplification while NHANES 99-02 samples were normalized to working concentrations. Samples were processed according to standard protocol and genotypes called by the BRLMM and Birdseed algorithms for the GW 5.0 and GW 6.0 mapping arrays, respectively.

Data Analysis: Genotype call rates for all samples and arrays were determined and genotype concordance of blinded sample repeats were calculated to assess genotype data quality.

Affymetrix Assay Overview



Affymetrix Genome-Wide 5.0 and 6.0 Content

Genome-Wide 5.0 Array
- 500,568 SNP probes – random distribution
- 420,000 Non-polymorphic CNV Probes

Genome-Wide 6.0 Array
- 906,600 SNP probes
  - 482,000 SNPs from 500K and 5.0 – random
  - 424,000 additional SNPs
    - "Tag" SNPs
    - Y Chromosome and Mitochondrial SNPs
    - Recombination hot-spots
- 946,000 Non-polymorphic CNV probes

NHANES III Genome-Wide 6.0 Mapping Array Call Rates

Samples	Median Call Rate	Minimum Call Rate	Maximum Call Rate
NHANES III DNA Bank: 44 WGA DNAs	98.50%	97.44%	99.15%
Affymetrix Control: 4 Genomic DNAs	99.06%	98.76%	99.20%

****SNPs with Call Rate >85%: 889,966 (98.17%)**

NHANES 99-02 Genome-Wide 5.0 and 6.0 Mapping Array Call Rates

Samples	Median Call Rate	Minimum Call Rate	Maximum Call Rate
Genome-Wide 5.0: 8 Genomic DNAs	99.58%	99.33%	99.74%
Genome-Wide 6.0: 8 Genomic DNAs	99.62%	99.41%	99.78%

Affymetrix Genome-Wide 5.0 and 6.0 Reproducibility

NHANES III GW 6.0 Repeats	Calls in Both	Equal Calls	Concordance
18CZQM:18CZQN	882,984	870,865	98.63%
18CZQE:18CZQG	884,843	876,676	99.08%
Affymetrix Control	895,802	892,937	99.68%
Combined	2,663,629	2,604,478	99.13%

NHANES 99-02 GW 5.0 Repeats	Calls in Both	Equal Calls	Concordance
001DLFTZ-001DLFT2	439,689	438,732	99.78%
001DLFT1-001DLFT3	440,276	439,531	99.83%
Combined	879,965	878,263	99.81%

NHANES 99-02 GW 6.0 Repeats	Calls in Both	Equal Calls	Concordance
001DLFTZ-001DLFT2	906,387	902,845	99.61%
001DLFT1-001DLFT3	905,126	901,157	99.56%
Combined	1,811,513	1,804,002	99.59%

Results

- Genotype call rates for the Affymetrix Genome Wide 5.0 and Genome-Wide 6.0 arrays were in excess of 99% for both the NHANES III and NHANES 99-02 samples.
- Genotype concordance measurements were likewise greater than 99% for both the NHANES III and NHANES 99-02 samples for both the Genome-wide 5.0 and 6.0 arrays.
- Whole Genome-Amplified NHANES III DNA bank cell lysate samples performed comparably to NHANES 99-02 DNA bank and control genomic DNA samples.

Conclusions

This proof of principle evaluation indicates that the NHANES DNA banks are suitable for genotyping with the Affymetrix high-density DNA mapping microarrays. These results highlight the potential for using the NHANES DNA bank in whole genome association studies to confirm known and to identify novel genetic risk factors that impact health measures that are important to U.S. public health.

Acknowledgements

CCEHIP/NCEH/DLS
Peg Gallagher
Miyono Hendrix
Patricia Mueller

CCID/NCPDCID/DSR
Robin Scarborough
Irshad Sulaiman

NCHS
Chris Sanders

CCID/NCIRD/ID
Natalie McDonald

CONTACT INFORMATION: Christopher Greene, Ph.D. CGreene@cdc.gov 770.488.4845