The CDC/NCI NHANES III Collaborative Genomics Project

The Genomics and Population Health Research Team
National Office of Public Health Genomics
Nicole F. Dowling, Man-huei Chang, Ramal Moonesinghe, Renée M. Ned, Ajay Yesupriya, Mechele A. Lynch, Diane T. Smelser, Lyna Zhang

Background:

- Understanding the role of genomic variation in human health and disease burden holds great promise for health promotion and disease prevention.
- The Third National Health and Nutrition Examination Survey (NHANES III), a weighted and minority-enriched population-based survey of the United States, includes data from detailed household interviews and clinical, laboratory and radiologic examinations.
- The NHANES III DNA bank (1991-1994) contains specimens from 7,159 survey participants aged 12 years and older and is maintained at the National Center for Environmental Health.
- Together, the survey and genetic data from NHANES III are a unique and unparalleled resource for epidemiological and genetic research.
- NOPHG established the CDC/NCI NHANES III Genomics Working Group in 2002 to formulate a plan for genetic analyses of NHANES III DNA bank samples.

CDC/NCI NHANES III Genomics Working Group (70+ members)

Centers for Disease Control and Prevention (CDC)

- Office of the Director
- National Office of Public Health Genomics (NOPHG)
- National Center for Infectious Diseases
- National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention
- National Center for Immunization and Respiratory Diseases
- National Center for Environmental Health (NCEH)
- National Institute for Occupational Safety and Health (NIOSH)
- National Center for Chronic Disease Prevention and Health Promotion
- National Center for Birth Defects and Developmental Disabilities
- National Center for Health Marketing
- National Center for Health Statistics (NCHS)
- Agency for Toxic Substances and Disease Registry

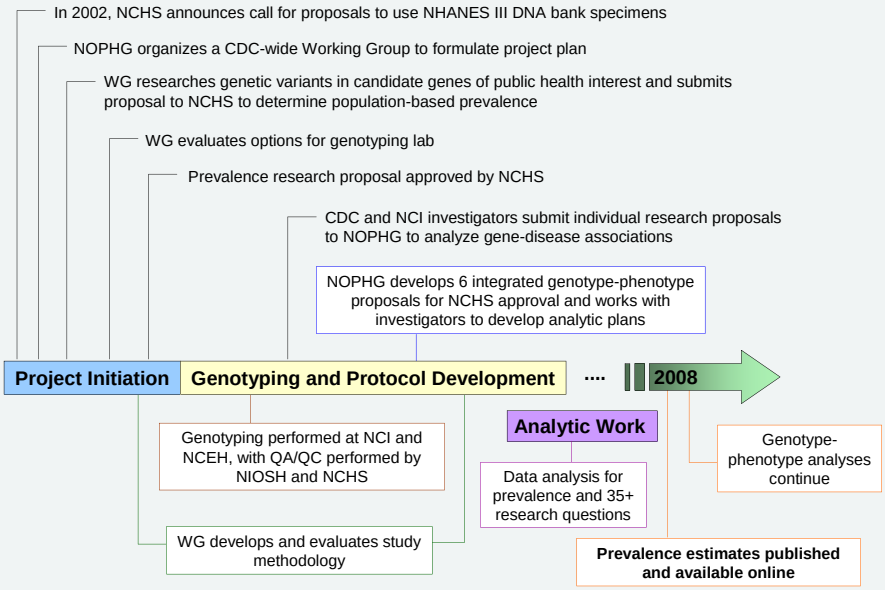
National Cancer Institute (NCI)

- Core Genotyping Facility
- Division of Cancer Epidemiology and Genetics

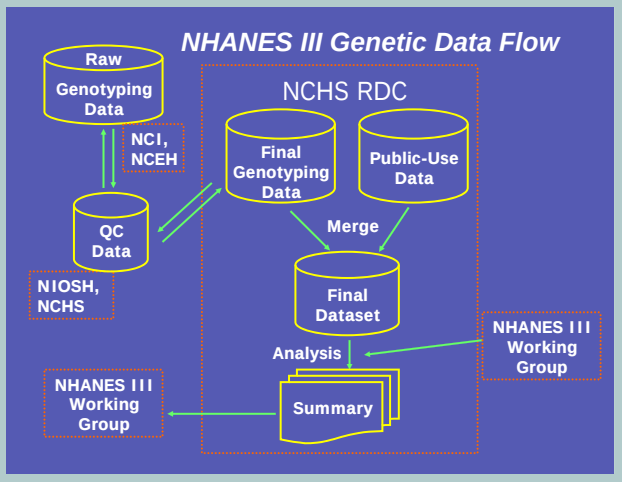
Study Objectives:

- To determine the prevalence of genotypes of public health importance in the U.S. population and its subgroups
- To evaluate the relationships between genomic variation and clinical, behavioral, environmental, and biochemical measures

NHANES III Project Overview



NHANES III Genetic Data Flow



Methods:

Selection of candidate genes and variants with:

- Known or hypothesized association with disease(s) of public health importance
- Role in cellular/physiologic pathways affecting multiple diseases, or evidence of gene-environment or gene-gene interactions
- Identified functional variants that are relatively common (i.e., >2% frequency)

Genotyping and Quality Assurance/Quality Control (QA/QC)

- Genotyping facilities:
  - The Core Genotyping Facility, NCI
  - The Division of Laboratory Sciences, CDC/NCEH
- Genotyping methods:
  - Taqman and Epoch MGB Eclipse assays for majority of the variants
- QA/QC guidelines established by NCHS

Data Analysis

- Conducted in the Research Data Center (RDC) at NCHS by using SAS-Callable SUDAAN 9.01 and SAS 9.1
- Since NHANES III is a multistage complex sample survey, all statistical analyses account for sample weights and the survey design

Results:

- 90 variants in 50 genes are available for prevalence estimation and genotype-phenotype association studies
- Prevalence data will be published soon and available online in the Human Genome Epidemiology (HuGE) Navigator
- Six large proposals were compiled (including 35-40 unique research questions):
  - Asthma/Chronic Obstructive Pulmonary Disease
  - Cardiovascular Disease
  - Diabetes, Obesity, and Metabolic Syndrome
  - Infectious Disease
  - Lead Exposure and Toxicity
  - Reproductive Health, Osteoporosis, and Cancer
- Genetic association studies are underway, and presentations/publications of findings are in progress

Conclusion:

- The nationally representative prevalence data can be used in evaluating the genetic epidemiology of complex diseases.
- The assessment of the impact of genetic variation in common diseases and the measurement of gene-gene and gene-environment interactions is critical for the translation of genomic information into clinical and public health practices.

Next Steps:

- Further examine the genetic substructure of the U.S. population
- Launch the **Beyond Gene Discovery (BGD) Initiative**
  - A collaboration with public, private, and academic partners
  - Purposes are to:
    - Create comprehensive population-based prevalence estimates for single nucleotide polymorphisms (SNPs) and other variants, creating the genomic profile of the U.S. population
    - Comprehensively identify the associations among variations in genotype, phenotype, and risk factors in a nationally representative sample of the population
  - Goal is to analyze more than 1 million genetic variants
  - Include specimens from NHANES III and NHANES 1999-2002 (approximately 15,000 survey participants in total)