Prevalence of Genetic Variants in the United States: NHANES III (1991-1994)

Chang MH[1], Lindegren ML[2], Butler MA[3], Chanock SJ[4], Dowling NF[1], Gallagher M[5], Moonesinghe R[1], Moore CA[6], Ned RM[1], Reichler M[2], Sanders CL[7], Welch R[8], Yesupriya A[1], Khoury MJ[1] for the CDC/NCI NHANES III Genomics Working Group

[1]National Office of Public Health Genomics, Coordinating Center for Health Promotion, CDC; [2]National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, CDC; [3]National Institute for Occupational Safety and Health, CDC; [4]National Cancer Institute, NIH; [5]National Center for Environmental Health, CDC; [6]National Center on Birth Defects and Developmental Disabilities, CDC; [7]National Center for Health Statistics, CDC; [8]Core Genotyping Facility, Division of Cancer Epidemiology and Genetics, Advanced Technology Program, SAIC Frederick, Inc. NCI

Background

- Background
  - Allele and genotype frequencies are important for understanding the contribution of genetic variation to human disease susceptibility, progression, and outcomes.
  - Population-based prevalence estimates provide the basis for epidemiologic studies of gene-disease associations, for estimating population attributable fractions, and for informing health policy and clinical and public health practice.
- Study aim
  - Determine prevalence of genotypes of public health importance by sex, age, and race/ethnicity in the U.S. population

First population-based estimates of allele and genotype frequencies for the U.S.

Materials and Methods

- DNA samples
  - 7,159 participants aged ≥12 years in the NHANES III DNA bank (1991-1994)
- 90 variants in 50 genes available
- Statistical analysis
  - Conducted analysis with SAS-Callable SUDAAN 9.01 and SAS 9.1
  - Used NHANES III genetic sample weights due to complex survey design
  - Reported allele frequency and genotype prevalence by race/ethnicity, age, and sex
  - Tested differences in allele and genotype frequency using χ^2 test at $p < 0.05$

50 Genes in Major Cellular/Physiologic Pathways

Apoptosis, Cell cycle, Cellular growth and differentiation *CAPN10, IL10, IL1B, IL4, IL4R, ITGB3, PPARG, TGFB1, TNF, VDR*	**DNA Repair** *OGG1, XRCC1*	**Metabolism of free radicals/Oxidative stress** *CAT, NOS2A, NOS3, PON1*
Blood pressure regulation, Cardiac function *ACE, ADRB1, ADRB2, NOS2A, NOS3*	**Hemostasis** *F2, F5, FGB, ITGA2, ITGB3, NOS3, SERPINE1*	**Nutrient Metabolism** *ACE, ADH1B, ADH1C, ADRB1, ADRB2, ADRB3, ALAD, CAPN10, CAT, CBS, CYP1A1, CYP1A2, CYP1B1, CYP2A6, CYP2C19, CYP2C9, CYP2E1, CYP3A4, MTHFR, MTRR, NOS2A, NOS3, NQO1, PPARG, SERPINE1, TNF, VDR*
Cellular adhesion, Cell migration/motility *CCL5, CCR2, CXCL12, F2, FGB, ITGA2, ITGB3, SERPINE1*	**Immunity and Inflammation** *CCL5, CCR2, CXCL12, FCGR2A, IL10, IL1B, IL4, IL4R, MBL2, NOS2A, PPARG, TGFB1, TLR4, TNF, VDR*	**Xenobiotic Metabolism** *ABCB1, ADH1B, ADH1C, ALAD, CYP1A1, CYP1A2, CYP1B1, CYP2A6, CYP2C19, CYP2C9, CYP2E1, CYP3A4, NAT2, NQO1, PON1*

NHANES III Demographic Characteristics

Over-sampling in certain populations

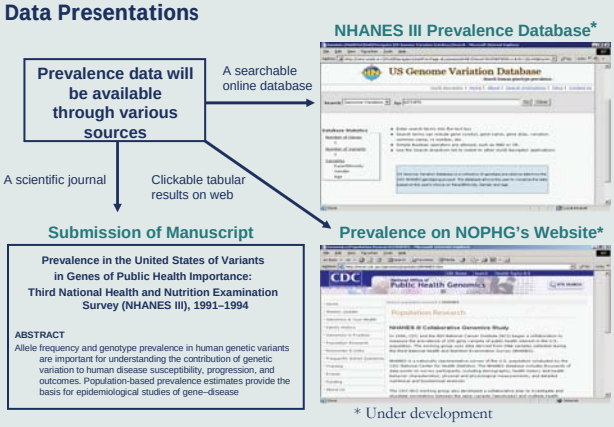
Demographic Characteristics	No. of Subjects	Unweighted Frequency (%)	Weighted Frequency (%)
Total	7,159	100.0	100.0
Sex			
Male	3,102	43.3	48.1
Female	4,057	56.7	51.9
Age (yrs)			
12-19	1,211	16.9	13.6
20-39	2,597	36.3	39.3
40-59	1,552	21.7	27.7
60+	1,799	25.1	19.4
Race/Ethnicity			
Non-Hispanic white	2,630	36.7	73.5
Non-Hispanic black	2,108	29.5	11.7
Mexican-American	2,073	29.0	5.7
Other	348	4.9	9.2

Results

We report the nationally representative estimates and the 95% confidence intervals (CI) of allele and genotype frequencies for 90 variants in 50 genes for each race/ethnicity, age, and sex subgroup in the U.S. population

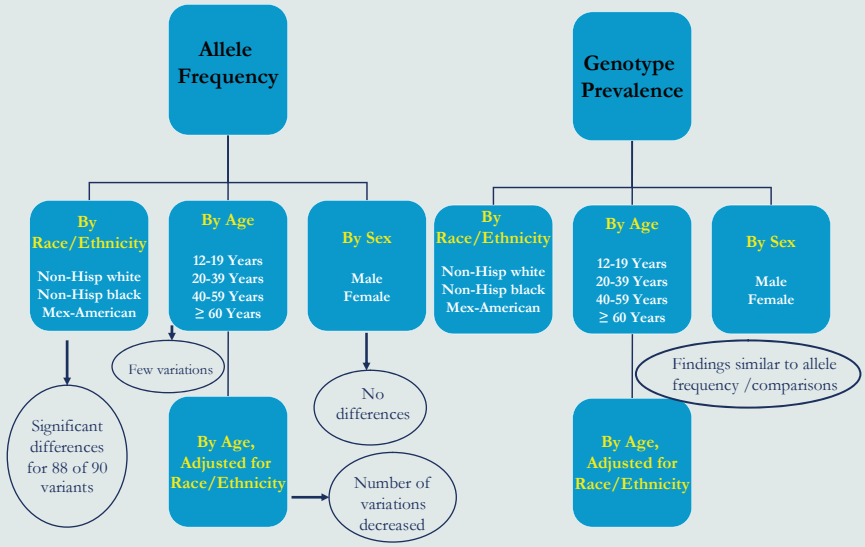
Gene	Variant	Allele	Total		Non-Hispanic White		Non-Hispanic Black		Mexican American		*P* Value
			%	95% CI	%	95% CI	%	95% CI	%	95% CI	
ABCB1	rs1045642	C	52.7	51.0, 54.4	48.4	46.8, 50.0	79.1	77.6, 80.5	55.3	52.8, 57.7	<.0001
		T	47.3	45.6, 49.0	51.6	50.0, 53.2	20.9	19.5, 22.4	44.7	42.3, 47.2	
ACE	rs4646994	Del	53.8	52.2, 55.4	54.6	52.8, 56.3	58.7	56.9, 60.5	46.6	44.5, 48.8	<.0001
		Ins	46.2	44.6, 47.8	45.4	43.7, 47.2	41.3	39.5, 43.1	53.4	51.2, 55.5	

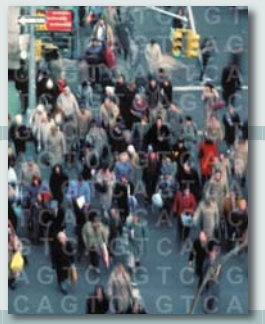
Data Presentations



* Under development

Population-Based Estimates





Conclusion

- These nationally-representative allele and genotype frequency data provide a tremendous resource for future epidemiologic studies in public health in the United States
- Further research
  - Evaluation of the genetic substructure of the U.S. population and subpopulations
  - Determination of associations of the reported genetic variants with health outcomes available in NHANES III (genotype-phenotype associations)
  - Genome-wide prevalence and association studies using the NHANES data (The Beyond Gene Discovery Initiative)