Prioritizing Studies for Whole Genome Association Genotyping

Panel V

Multi-IC Symposium on Application of Genomic Technologies to Population-Based Studies

Christopher J. O'Donnell, MD MPH
DHHS/NIH/NHLBI

Panel V: Prioritizing Studies for WGA Genotyping

- **Determining when WGA genotyping is appropriate**
- **Determining when replication/expansion of WGA genotyping is needed**
- **Need for diversity in population genetic and environmental backgrounds**
- **Appropriate studies to follow-up on WGA findings**
- Assuring appropriate DNA collection, proper human subjects protection, appropriate quality of DNA samples
- Pooling of cohort DNA samples, genotype data, and phenotypes; assuring that phenotype criteria allow pooling
- Determining sufficient power to detect an association
- Pooling DNA samples before genotyping to reduce N of genotype assays that needs to be done

Working Group on GWA in NHLBI Cohorts: September 2005

Working Group Discussion Areas

September 12, 2005, NHLBI

- Criteria for Selecting Cohorts to be Genotyped
- Informatics Needs/Data Management
- Statistical Analysis Issues
- Data Sharing, Access, Consent, Confidentiality, Reporting
- Approaches to Genotyping, Assessment of Platforms, Quality Control
- Sample Acquisition and Types

Summary at www.nhlbi.nihl.gov

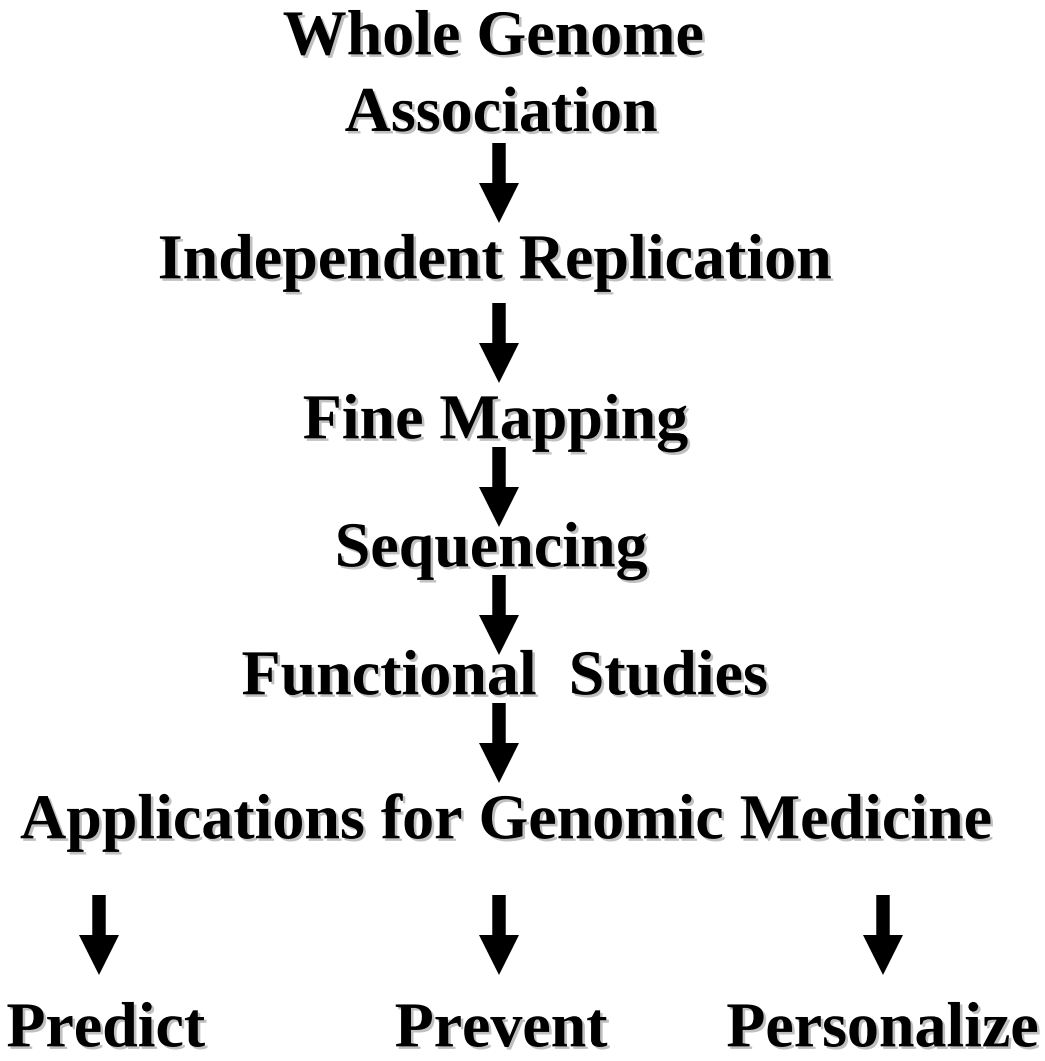
Working Group Members

Eric Boerwinkle, Ph.D., Gregory L. Burke, M.D., M.S., Aravinda Chakravarti, Ph.D., Christopher G. Chute, M.D., Dr.P.H., Mark J. Daly, Ph.D., Seth E. Dobrin, Ph.D., Kelly A. Frazer, Ph.D., Stacey Gabriel, Ph.D., Gary Gibbons, M.D., Gerardo Heiss, M.D., Eduardo Marbán, M.D., Betty S. Pace, M.D., Susan Redline, M.D., M.P.H., Christine Seidman, M.D., Belgaum S. Thyagaraja, Ph.D., Russell P. Tracy, Ph.D., Bruce S. Weir, Ph.D., Scott T. Weiss, M.D., M.S.

Working Group on WGAS in NHLBI Cohorts: Recommendations

- Large-scale GWA projects should proceed.
- Existing cohorts with large amount of phenotype and exposure data should be used.
- >1 cohorts should be included, to catalyze data and repository harmonization and maximize diversity by age, sex, race, ethnicity; consider opportunities for replication.
- Well-defined, quantified phenotypes of public health importance should be analyzed.
- **Immediate access should be provided** to data held in a centralized data repository.
- A mixture of study designs should be supported-- studies of families, trios, & unrelated individuals.
- Information about cohorts should be quickly disseminated, allowing easy application for data.

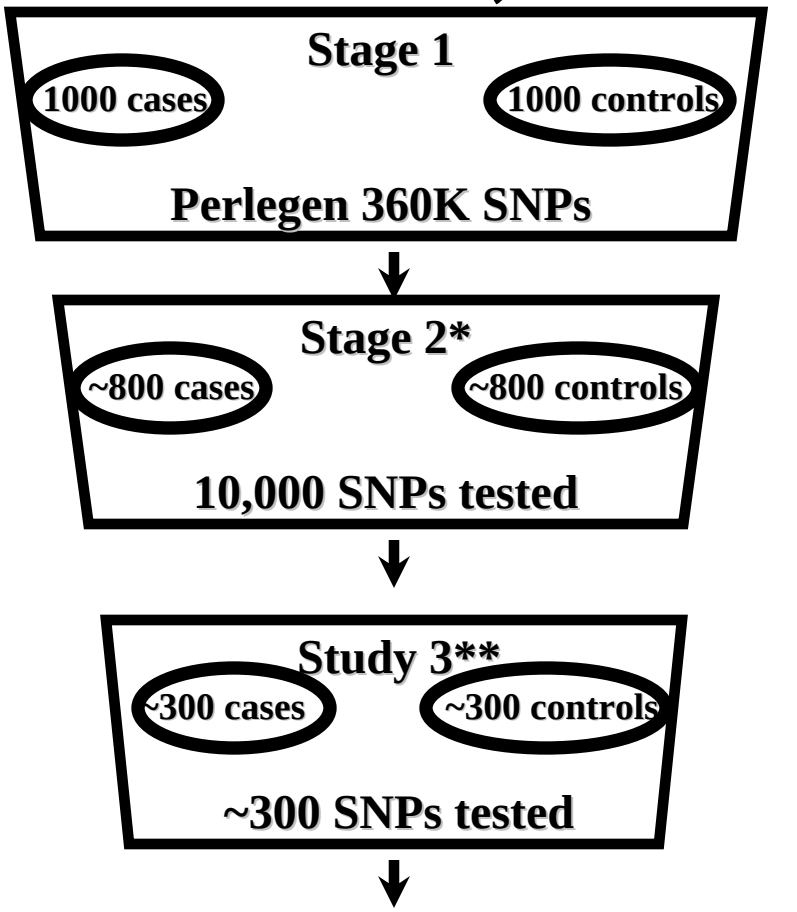
Translation from Whole Genome Association to Genomic Medicine



NHLBI WGA Portfolio: Planned and Ongoing WGA Studies

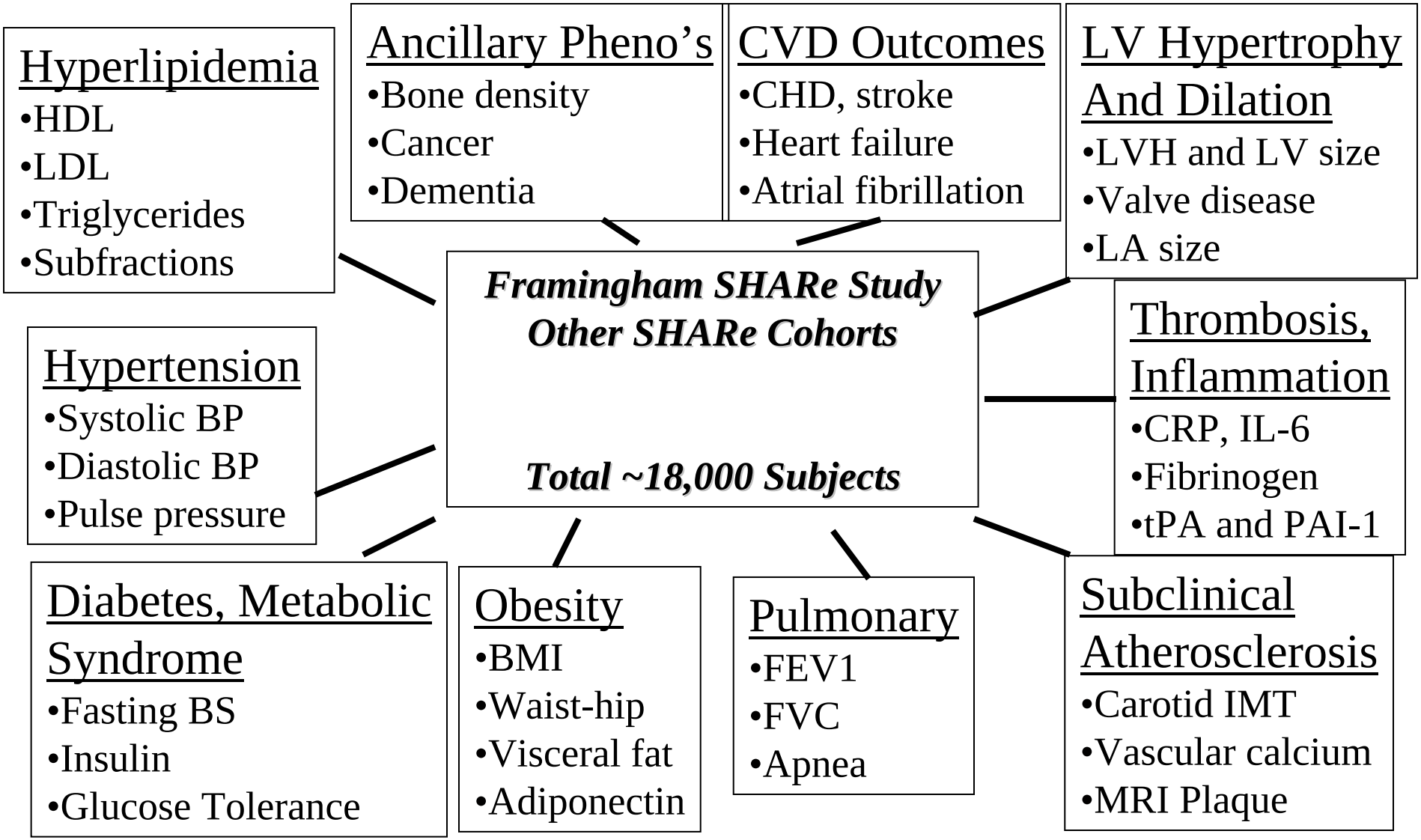
- SNP Typing for Association with Multiple Phenotypes in Existing Epidemiology Data (STAMPEED)
  - Up to 6 WGAS's for heart, lung, blood and sleep disorders
  - RFA-HL-06-012
- WGAS in Women's Health Initiative (WHI)
  - Staged design with pooled cases vs controls in Stage 1
  - Three phenotypes: CHD, stroke and breast cancer
- Framingham Heart Study: Framingham SNP Health Association Resource (SHARe) + other SHARe Projects
  - WGAS in 9000 subjects from three generations of FHS
  - Collaboration between NHLBI, BU and NCBI
  - Phenotypes: **all available** contract & ancillary study data
  - 2 additional cohorts ~7200 subjects (RFP NHLBI-PB-2006-091)

Pooled WGA Scans in the WHI for CHD, Stroke, Breast CA



- Stage 1: Pooled DNA, threshold p<.02
- Stage 2: Individual DNA samples, p=.0004 or p<.02 in both Stage 1 & 2
- Stage 3: Individual DNA samples, threshold p<.00001 for 360K SNPs yields 3.6 FP's

SHARe WGA Studies in Framingham and Other NHLBI Population-Based Cohorts



Simple WGAS Mathematics

9000 subjects
x 500,000 SNPs
x 1000 Phenotypes
x Two Adjustments (Age-adjusted and Multivariable)
x Four Genetic Models

= 36,000,000,000,000 Association Tests

Features of Framingham SHARe

- All participants provided written informed consent for genetic research, genetic data sharing (DNA Committee) in place for a decade, ongoing dialogue with participants
- Consent includes questions regarding opt-in for genetics, commercial use, other disease areas (eg, cancer)
- DNA samples backed up by “cell lines“ (majority)
- SHARe Oversight Committee and Steering Committee
- Community Ethics Advisory Board and OSMB consulted
- Participants informed via newsletters and focus group
- Data Access Committee authorizes distribution of data after review of application, IRB, distribution agreement
- NCBI provides access to authorized Users
- 12 month embargo on publications except for FHS Investigators and collaborators

NHLBI WGAS Portfolio: Design, Analysis and Follow-up Genotyping

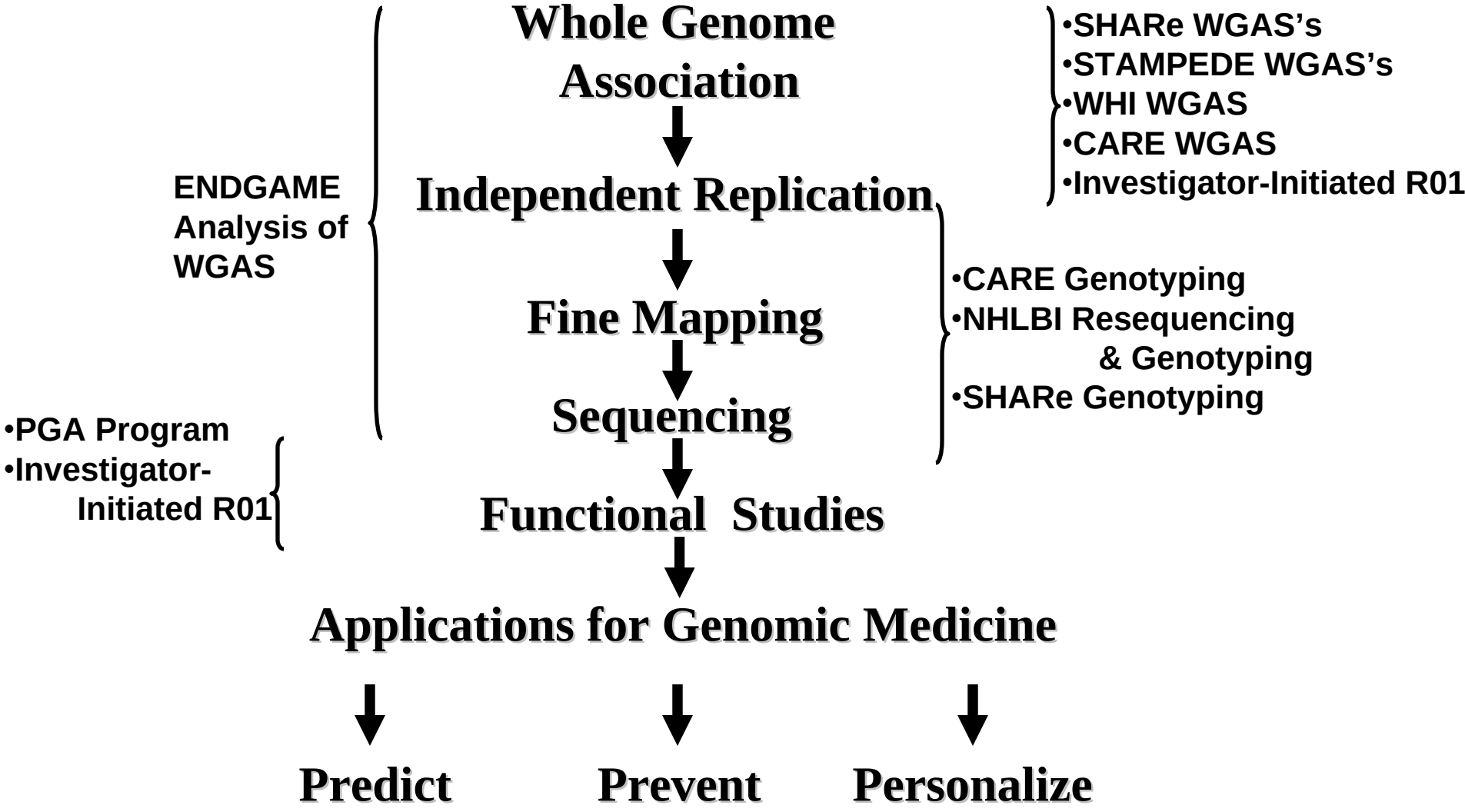
- Design and Analysis of WGAS
  - ENhancing Development of WGAS MEthods (ENDGAME RFA-HL-05-011)
  - 11 projects on methods of analysis for WGAS data
- Large-Scale Genotyping
  - Candidate Gene Association Resource (CARE)
  - Genotyping of up to 50,000 subjects from 8 cohorts:
    - MESA
    - ARIC
    - CHS
    - CARDIA
    - Cooperative Study for Sickle Cell Cohort
    - Sleep Heart Health Study
    - Jackson Heart Study
    - Framingham Heart Study
  - Broad Institute (Cambridge, MA) contracted to conduct:
    - Modest-sized WGAS (n=1500)
    - Genotype ~10 SNPs in ~1700 candidate genes
    - Plan for follow-up of available WGAS findings

Other NHLBI Genetic and Genomic Resources Relevant to WGAS

NHLBI Programs and Services:

- Resequencing and Genotyping Service
- Programs for Genomics Applications (PGA)
- Program on Interactions of Genes and Environment in Shaping Risk Factors for Heart, Lung, Blood and Sleep Disorders (PROgram for GENetic Interaction, PROGENI)

Translation from Whole Genome Association to Genomic Medicine



Panel V: Prioritizing Studies for WGA Genotyping

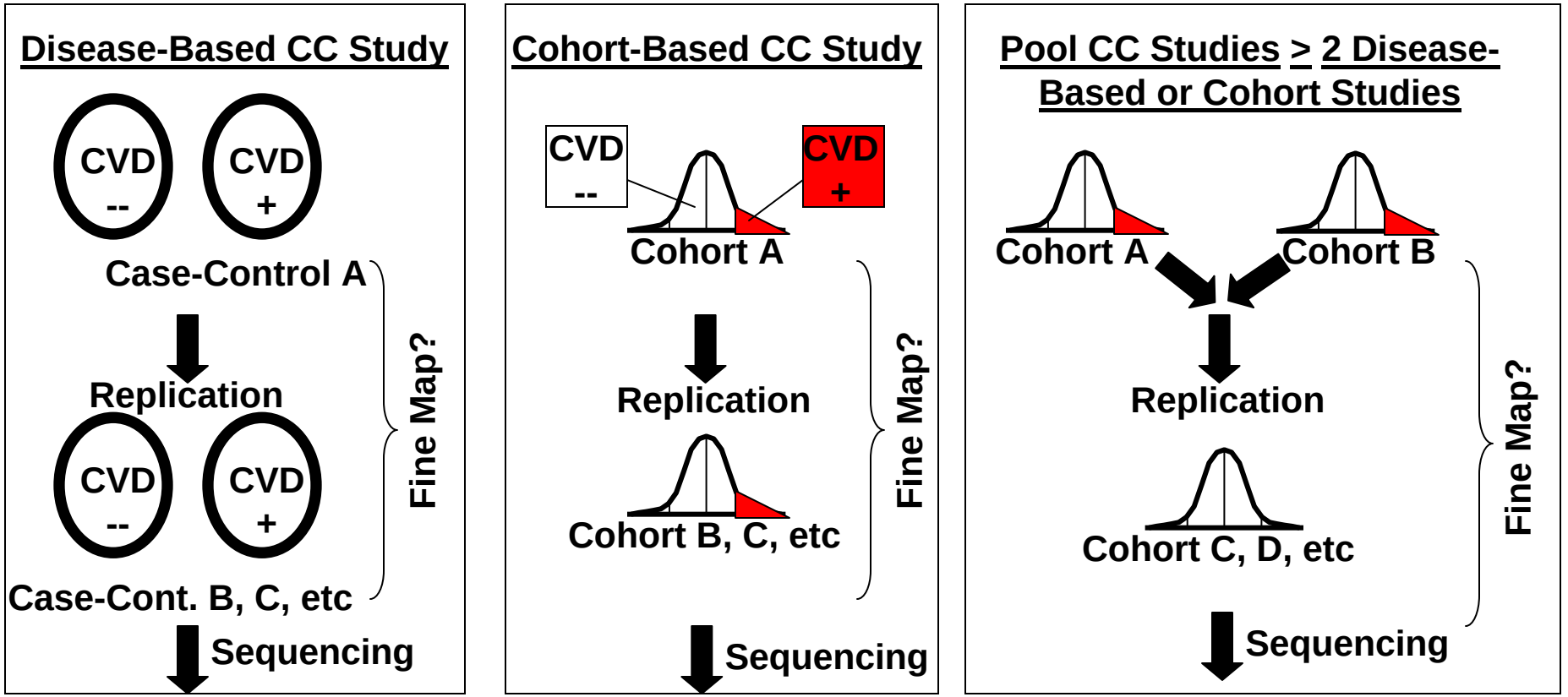
- Determining when WGA genotyping is appropriate
- Determining when replication/expansion of WGA genotyping is needed
- Need for diversity in population genetic and environmental backgrounds
- Appropriate studies to follow-up on WGA findings

Determining that WGAS is Appropriate for HLBS Phenotypes

Types of cohorts and types of traits

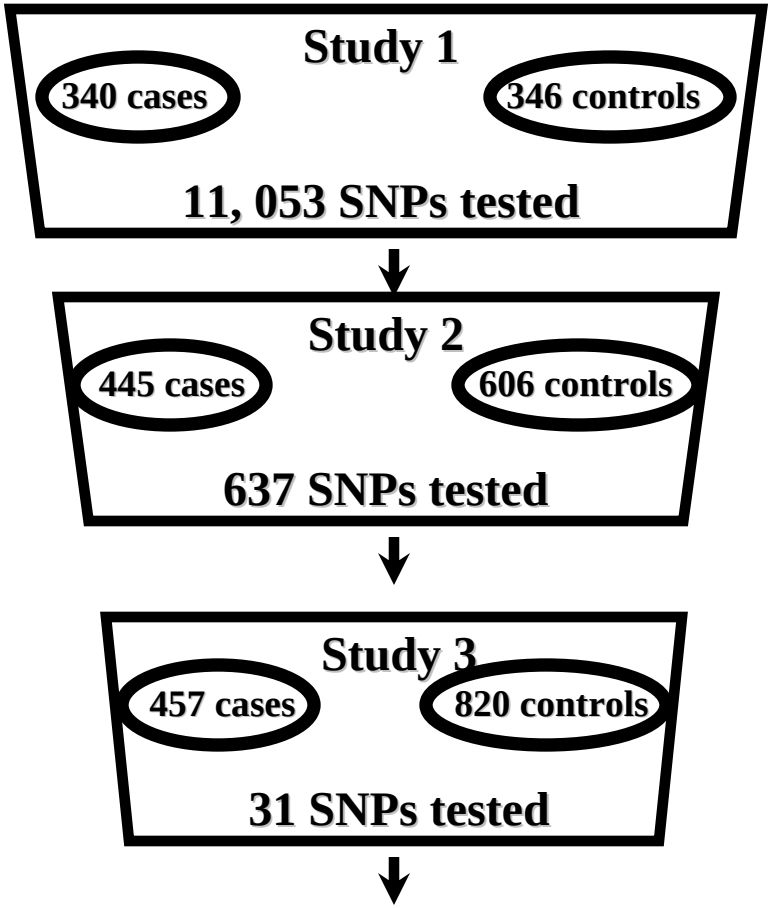
- Cohorts for Study
  - Population-based: prospective versus cross-sectional
  - Disease-based: case-control
- Traits for Study
  - Qualitative
    - Heart, lung, blood or sleep outcomes
    - Extremes of quantitative traits
  - Quantitative
    - Examination measures, risk factors, biomarkers
    - Subclinical disease measures

WGAS: Qualitative Traits (eg, CVD) in Case-Control, Population-Cohorts



Functional Studies

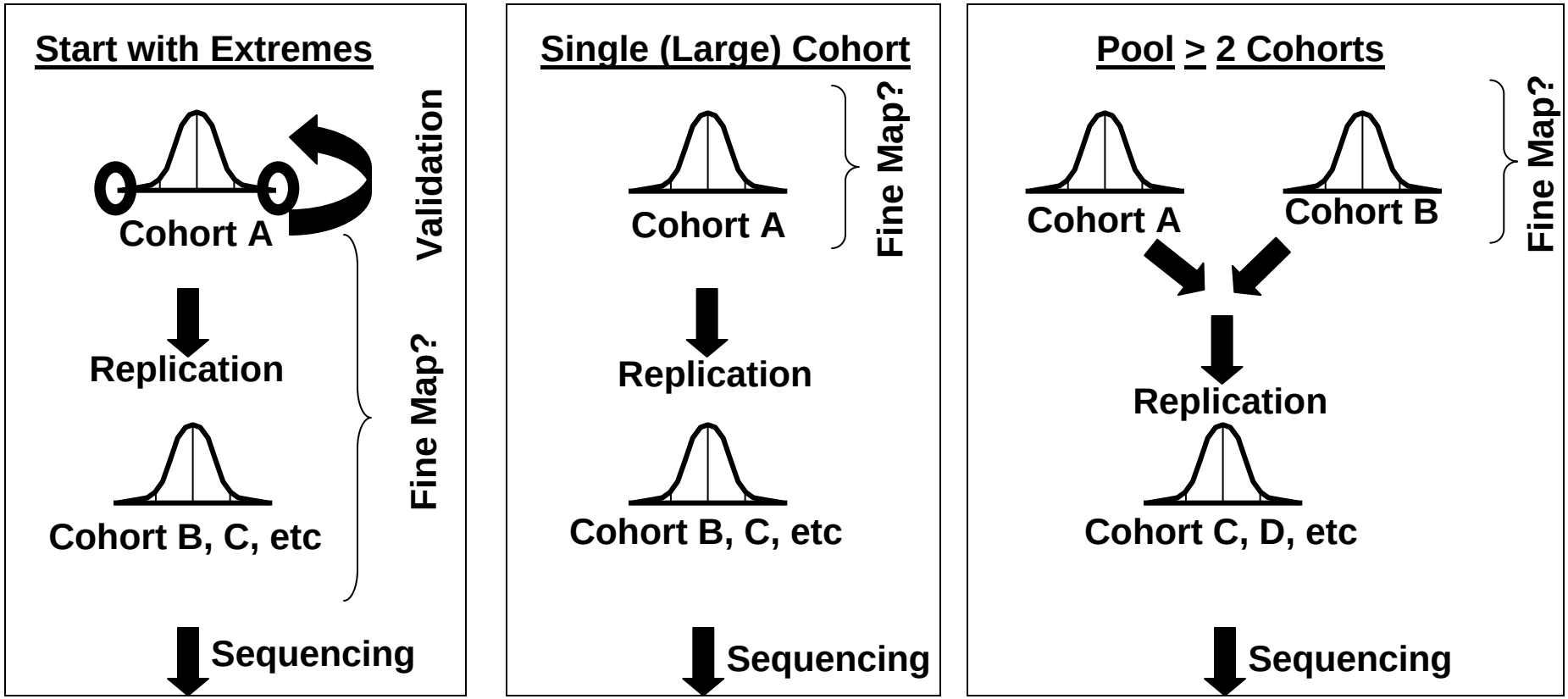
Case-Control Studies of MI: Survey of Missense Gene Variants



- Study 1: Pooled DNA, threshold p<.05
- Study 2: Pooled DNA, threshold p<.05 in both Studies 1 and 2
- Study 3: Individual DNA samples, threshold FDR<10% for accepting SNPs

Shiffman et al. AJHG 2005;77:596.

WGAS: Quantitative Traits (eg, HDL-Chol) in Population-Based Cohorts



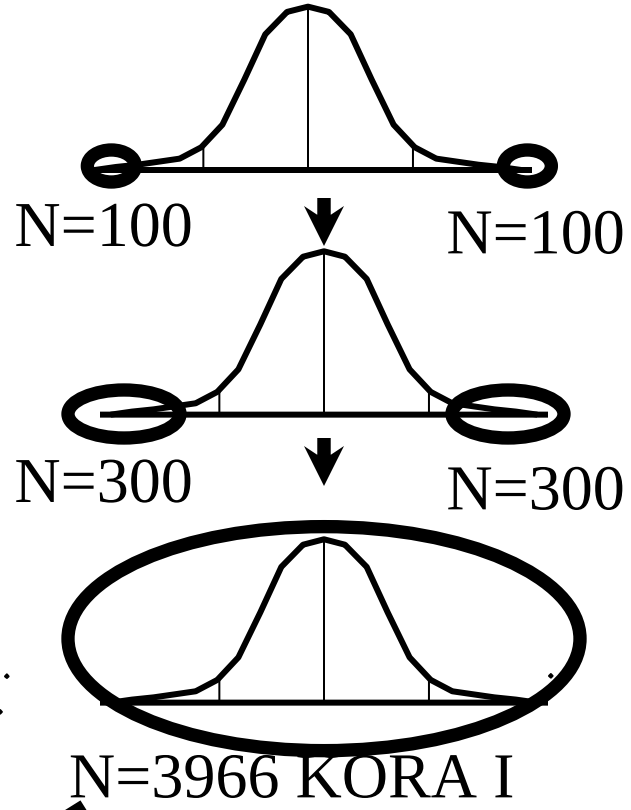
Functional Studies

Screen Extremes, Replicate in Cohorts: *CAPON* Variant (*NOS1* Regulator) and QTc

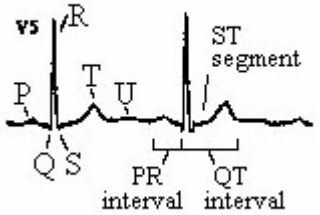
Stage I:~100K SNPs, Women Only KORA I

Stage II: Top 10 SNPs, Women Only KORA I

Stage III: Top 7 SNPs, Men & Women KORA I



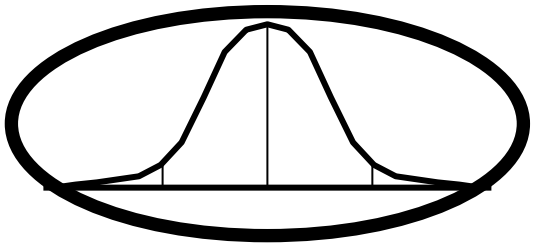
Phenotype: Age-, sex- and RR-adjusted QT interval



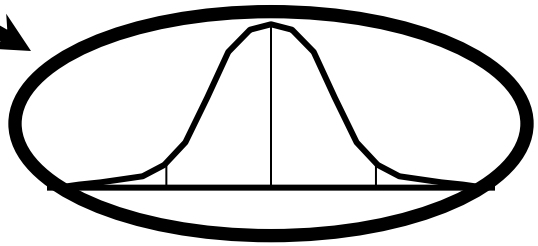
Arking DE et al. Nature Genetics 2006; epub.

NOS1AP (CAPON)

QTc_5.3
FGFR2
QTc_14.1
KCNK1
ITPR1
CACNA2D1



N=2646 KORA II



N=1805 Fram. Heart Study

100K Quantitative Trait Study in Cohort of Families: *INSIG2* Variant and BMI/Obesity

A Common Genetic Variant Is Associated with Adult and Childhood Obesity

Alan Herbert,[1*] Norman P. Gerry,[1] Matthew B. McQueen,[2] Iris M. Heid,[3,4] Arne Pfeufer,[5,6] Thomas Illig,[3,4] H.-Erich Wichmann,[3,4,7] Thomas Meitinger,[5,6] David Hunter,[2,8,9] Frank B. Hu,[2,8,9] Graham Colditz,[8,9] Anke Hinney,[10] Johannes Hebebrand,[10] Kerstin Koberwitz,[6,10] Xiaofeng Zhu,[11] Richard Cooper,[11] Kristin Ardlie,[12] Helen Lyon,[13,14,15] Joel N. Hirschhorn,[13,14,15] Nan M. Laird,[16] Marc E. Lenburg,[1] Christoph Lange,[9,13] Michael F. Christman[1*]

Obesity is a heritable trait and a risk factor for many common diseases such as type 2 diabetes, heart disease, and hypertension. We used a dense whole-genome scan of DNA samples from the Framingham Heart Study participants to identify a common genetic variant near the *INSIG2* gene associated with obesity. We have replicated the finding in four separate samples composed of individuals of Western European ancestry, African Americans, and children. The obesity-predisposing genotype is present in 10% of individuals. Our study suggests that common genetic polymorphisms are important determinants of obesity.

Obesity is associated with an increased risk of type 2 diabetes mellitus, heart disease, metabolic syndrome, hypertension, stroke, and some forms of cancer (*1*). It is commonly assessed by calculating an individual's body mass index (BMI) [weight/(height)^2 in kg/m^2] as a surrogate measurement. Individuals with a BMI $\geq$ 25 kg/m^2 are classified as overweight, and those with a BMI $\geq$ 30 kg/m^2 are considered obese. Having a BMI over 25 kg/m^2 increases the risk of death (*2*). Presently, 65% of Americans are overweight and 30% are obese (*3*). Genetic factors contribute significantly to the etiology of obesity (*4*, *5*), with estimates of the heritability of BMI ranging from 30 to 70% (*6*–*9*).

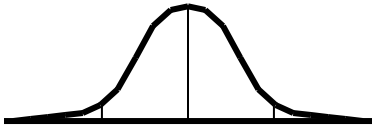
To identify common genetic variants associated with elevated BMI, we have studied individuals from the National Heart, Lung, and Blood Institute (NHLBI)–Framingham Heart Study (FHS) (*10*). The participants were enrolled from the community without being selected for a particular trait or disease and were followed over 24 years (table S1). In this population, heritability estimates for BMI range between 37 and 54% (*11*, *12*).

Using families from this sample, we performed a genome-wide association analysis, using a testing strategy for quantitative traits in

[1]Department of Genetics and Genomics, Boston University Medical School, E613, 715 Albany Street, Boston, MA 02118, USA. [2]Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA. [3]Institute of Epidemiology, [4]KORA Group, [5]Institute of Human Genetics, Forschungszentrum für Umwelt und Gesundheit, D-85764 Neuherberg, Germany. [6]Institute of Human Genetics, Technical University Munich, D-81671 Munich, Germany. [7]Institute for Medical Informatics, Biometry and Epidemiology, Ludwig Maximilians University, Munich, Germany. [8]Nurses Health Study, [9]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, 181 Longwood Avenue, Boston, MA 02115, USA. [10]Department of Child and Adolescent Psychiatry of the University of Duisburg-Essen, D-45147 Essen, Germany. [11]Department of Preventive Medicine and Epidemiology, Loyola University Medical Center, Maywood, IL 60153, USA. [12]Genomics Collaborative, SeraCare Life Sciences, Cambridge, MA, USA. [13]Program in Genomics and Divisions of Genetics and Endocrinology, Children's Hospital, Boston, MA 02115, USA. [14]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. [15]Broad Institute of Harvard and MIT, Cambridge, MA 02139, USA. [16]Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, MA 02115, USA.

*To whom correspondence should be addressed. E-mail: aherbert@bu.edu (A.H.); mfc@bu.edu (M.F.C.)

Framingham 100K Screen Identifies Prior Association with Factor 7 Locus



Framingham Offspring (Gen2)

~1800 Unrelated ~1350 in Families

Circulating Factor VII Antigen
Phenotype: Covariate-Adjusted Level
Genotype: htSNPs spanning F7 Locus

VARIANT	ROLE	minor allel freq	Individual p value
rs2146751	5' Flanking	0.33	<0.0001
rs10665	5' Flanking	0.13	<0.0001
rs964617	5' Flanking	0.24	<0.0001
rs1755685	5' Flanking	0.14	<0.0001
rs762636	5' Flanking	0.22	<0.0001
rs6039	Intronic	0.14	<0.0001
rs1475931	Intronic	0.23	0.06
rs6046*	Coding Arg/Gly	0.13	<0.0001

***Explains ~10% of Variation**

0.05≤P<1.0 | 0.01≤P<0.05 | P<0.01

Kathiresan et al. ATVB 2006.

Circulating Factor VII Antigen
Phenotype: Covariate-Adjusted Level
Genotype: 100K Affymetrix CHIP

Of 84633 tests for adjusted FVII trait, strongest association is a SNP (P=5*10-16) highly correlated with Arg/Gln F7

Determining that WGAS is Appropriate for HLBS Phenotypes

Phenotypes

- Disease phenotypes relevant to public health
  - Common, complex cardiovascular, lung and blood dz
    - Clinical outcomes: eg, CVD, COPD, OSA
    - Quantitative phenotypes: eg, SBP, LDL-C
  - Rarer disease phenotypes: eg, HCM, SSD, CF
- Intermediate phenotypes that predict disease
- Significant genetic component
- Measurable, reproducible
- Prediction or treatment of phenotype may prevent future occurrence of disease

Determining that WGAS is Appropriate for HLBS Phenotypes

Program	Phenotype	Criteria
FHS SHARe	HLBS+Endo, Bone, CA Risk Factors,Biomarkers Subclinical Disease Clinical Disease	Major Diseases Longitudinal Large N 3 Gen Families Well-phenotyped
STAMPEED	HLBS Diseases Quantitative Traits or Disease Traits	Major Diseases Peer Review Well-phenotyped
WHI	CHD Stroke Breast CA	Major Diseases Large N Well-phenotyped

Diversity in Population Genetic and Environmental Backgrounds in WGAS

- Given differences in LD noted in HapMap for major ethnic groups, general approach is to confine WGAS to cohorts with adequate sample size within a single genetic background (eg, Caucasian or African American), then test in other backgrounds
- Fortunately, a number of NHLBI cohorts have been ascertained based upon major ethnic groups

Family Blood Pressure Program Networks, Ethnic Diversity

Network	Population	Major Phenotype Interests
GenNet	African American, Hispanic, and Caucasian: N~2000	Pre-hypertension, biochemical testing
GENOA	African American, Hispanic, and Caucasian: N~5000	Context dependent genetic effects
HyperGEN	African-American, Caucasian: Sib pairs w/ HTN (N= 1,836) Offspring, no HTN (N=1200)	Intermediate phenotypes for HTN in pedigrees
SAPPHIRe	Chinese, Japanese ancestry: 1200 Sib pairs, HTN families, ½ concordant, ½ discordant	Insulin resistance syndrome, metabolic syndrome

Factors Determining Whether to Attempt to Replicate WGAS Findings

- Power, False Discovery Rate
- Phenotype
  - Quality, Reproducibility
  - Heritability
  - Heterogeneity
- Strength of Association
- Potential Generalizability to Other Populations
  - Genotypic Background: Ethnicity
  - Environmental Background
- Prior Probability, Prior Knowledge
  - Known Candidate Locus or Linkage Peak
  - Biological Plausibility

Factors Determining Whether Replication is Valid

- How do we Define:
  - “Validation” vs “follow-up” vs “replication”
- When has Replication been Achieved? No firm agreement on criteria other than “more is better”. Criteria might include:
  - Statistical significance and pre-test probability
  - Similar ascertainment and phenotype definition in replication study(s)
  - Control for confounding, eg pop’n stratification
  - Comparisons of ≥1 genotyping platform
  - Number & sample size of replication(s)
  - Consistency in magnitude, direction of genetic effect
- How Large?
  - Again, not clear what size, but certainly replication studies should be powered to reliably be able to exclude a null finding for the initially observed magnitude of effect and genetic model.

Appropriate Studies to Follow-Up on WGAS Findings

- Wealth of Resources from NHLBI Cohorts
- Eg, CARE, anticipated total N~50,000:
  - Multiethnic Study of Atherosclerosis (MESA)
  - Atherosclerosis Risk in Communities (ARIC)
  - Cardiovascular Health Study (CHS)
  - Coronary Artery Risk Devel. in Young Adults (CARDIA)
  - Framingham Heart Study (FHS)
  - Jackson Heart Study (JHS)
  - Sleep Heart Health Study (SHHS)
  - Cooperative Study for Sickle Cell (CSSC)

Panel V: Prioritizing Studies for WGA Genotyping--Additional Considerations

- Criteria for assuring appropriate DNA collection, proper human subjects protection, and appropriate quality of DNA samples
- Pooling of cohort DNA samples, genotype data, and phenotypes; assuring that phenotype criteria allow pooling
- Determining there is sufficient power to detect an association
- Pooling DNA samples before genotyping to reduce N of genotype assays that needs to be done