# Data Aggregation Workshop Jun 5-6

- Deposit (almost) all sequence/exposure/phenot datasets in central database
- Limit (almost) all future sequence/phenot datasets to ppts consenting to broad use
- Public release of summary statistics (variant names, allele freq, OR) should be default
- Data access for dbGaP should be streamlined, including registered users having broad access

# Data Aggregation Workshop Jun 5-6

- New governance procedures for central database should include public disclosure and participation, accountability and penalties
- Central processing of sequence data should be encouraged
- Analysis of aggregate datasets should be supported
- Retrospective harmonization encouraged and efforts captured; future projects include harmonization as part of application and sharing

# Sequencing in Large Sample Collections Background

- Dramatic advances in sequencing technology and genomic understanding have made search for rare LoF variants in large samples achievable
- Very large numbers of cohort studies and other large samples available (http://www.genome.gov/27548522)
- Choosing among them challenging
- Convening workshop/think tank to address

# Goal and Objectives

Goal:

Provide guidance to NIH and the scientific community on the utility of sequencing large sample collections to improve the understanding and treatment of complex diseases.

Objectives:

- Identify the key scientific questions that can be addressed by sequencing
- Define criteria for selecting samples to answer those questions

# Workshop Topics

- Short- and long-term goals for sequencing in complex disease research
- Using genomic variants to guide treatment
- Finding rare variants of large effect related to complex diseases
- Scientific questions addressable by sequencing
  - Finding modifiers of known disease-related variants
  - Finding drug targets
  - Characterizing an individual's genomic risk of complex disease or drug response

# Workshop Topics (continued)

- Informatic imperatives for sequencing in large-scale studies
- Cloud computing for large-scale sequencing
- Criteria for selecting samples to sequence
  - Consent, data access, ability to recontact
  - Breadth vs depth of phenotyping
  - Outcome data, links to electronic medical records
  - Ancestral diversity
  - Participant selection and study design

# 1. Fitting Tools to Job (R Wilson)

- Informatic

# 2. Perils and Promise (P Donnelly)

- Informatic

# 3. Short- and Long-Term Goals (E Boerwinkle)

- Informatic

# 4. Charge and Deliverables (S Chanock)

- Informatic

# 5. Charge and Deliverables (S Chanock)

- Informatic

# 6. Guiding Treatment (M Murray)

- Informatic

# 7. Rare Variants and Disease Risk (E Boerwin)

- Informatic

# 8. Modifiers of Known Variants (M Knowles)

- Informatic

# 9. Finding Drug Targets (J Cho)

- Informatic

# 10. Discussion and Prioritization (Chanock/Lehner)

- Informatic

# 11. Informatic Imperatives (D MacArthur)

- Informatic

# 12. Cloud Computing (N Cox)

- Informatic

# 13. Consent, Access, Re-Contact (G Jarvik)

- Informatic

# 14. Phenotyping (J Buring)

- Informatic

# 15. Outcome Data, EMRs (D Roden)

- Informatic

# 16. Ancestral Diversity (L Jorde)

- Informatic

# 17. Ppt Selection and Design (T Hartge)

- Informatic

# 18. Discussion/Prioritization (E Boerwinkle)

- Informatic

# 19. Key Lessons/Reactions (R Collins)

- Informatic

# 20. Key Lessons/Reactions (M Olson)

- Informatic

# Guidance on Sequencing Large Collections

- Informatic

# Questions Answerable by Sequencing

- Informatic

# Criteria for Selection

- Informatic

# Workshop Topics (continued)

- Informatic imperatives for sequencing in large-scale studies
- Cloud computing for large-scale sequencing
- Criteria for selecting samples to sequence
  - Consent, data access, ability to recontact
  - Breadth vs depth of phenotyping
  - Outcome data, links to electronic medical records
  - Ancestral diversity
  - Participant selection and study design