Replication and Collaboration

- Three models of collaboration
  - Put data out there (SHARe)
  - Organize consortium prospectively (CARE)
  - Encourage collaboration (STAMPEED), but need to provide infrastructure
- Facilitating rapid replication
  - Lab infrastructure, queuing
  - Availability and type of samples
  - Whom to type
  - Agreement to do very rapidly
  - Established resource

Challenges of Attempting Replication in Different Populations or Settings

- Beware of blaming lack of association on differences in exposures or populations
- Issues in replication can raise intriguing scientific questions
  - Note lack of association of FTO in DGI diabetes scan due to controlling for obesity
  - Differences in population genetics can influence ability to confirm finding, especially with alleles at very high frequency
    - SLC30A8 in African sample due to 97% allele frequency
    - 8q24 rs6983267 risk allele for prostate cancer is 96% in African-Americans

Cross-Study Analyses

- Imputation for cross-platform comparisons essentially solved problem but need to do actual genotyping in lab before reporting
- How to temper over-blown expectations of GWA studies?
- Importance of reliable, standardized phenotyping– can come out of this meeting?
  - Changes over time
  - Major issue 50 years from now
  - Addition of comorbidity data to clinical studies?
- Need to plan in advance for in silico comparisons, to ensure compatibility of formats, genotypes, phenotypes

Follow-Up Studies

- High-throughput sequencing available now, still have need for better methods of targeted sequencing
- When is the right time to develop animal models for effect on phenotype
- Animal models have limits– Huntington's CAG repeat doesn't reproduce phenotype in mouse
- Establishing causality for small effect size loci is challenging
- Methods for detecting non-SNP variation

Follow-Up Studies: Three Types of Sequencing

- Common allele driving your association, may be SNP in LD to those typed and MAF > 3-4%: sequence through area of LD in modest number (n ~ 96)
- May be other haplotypes not in LD affecting function: look in functional unit (gene) outside LD block, still for common SNPs (n ~ 100)
- What if multiple rare alleles affecting function: targeted sequencing in areas not tolerant of variation (n ~ 1,000), esp extreme phenotypes

Data Sharing

- Earlier you can start on formulating data for sharing the better
- dbGaP happy to help you, starting with those willing to share
- Technically easier to put data together from outset rather than trying to connect post facto
- Versioning of datasets will occur
- Culture shift regarding impact on careers– positive aspects in NIA Parkinsons study
- De-identifying tough especially with image data
- Challenges of linking samples in repositories to datasets – should we pursue

Consent/Approval

- Problems with proprietary consents
- Framingham experience with proposed privatization
- Why consents are limited to one condition: IRBs and community concerns
- Consents specifying that results not to be provided may not be “liberating”
- Value of intermediate step of Technical Advisory Group review in GAIN
- How to deal with time limited consent in sharing

Consent/Approval (2)

- Every individual in one and only one consent category for dbGaP– need meticulous records
- No tribal consultation policy estab by NIH
- Value of working with patient organizations
- Community involvement has to be ongoing process, not one-time thing
- Biggest problem is getting controls
- What if there are conflicting consents in one ppt?
- Cultural sensitivities in minorities and foreign settings, religious concerns?
- Acceptability of narrower consents?

Unanswered Questions that May Need Further Investigation

- Utility of convenient freezer controls?
- Exploring the role for weighting for functional or known SNPs
- Can we develop an “alpha spending function” similar to sequential analyses of clinical trials, for GWA?
- What is the value of heritability in justifying GWAS and complex diseases?
- How should public health importance of trait weigh in decisions on replication?

Unanswered Questions that May Need Further Investigation (2)

- Importance and weighting of joint analysis versus any single study?
- How many scans of a given well-studied complex disease do we need?
- Reverse direction of function to association–MDM2 related to TP53 and cancer but not associated?
- What do complex consent forms do to participation rates, especially for controls?

Things We at NIH Can Do

- Making association results available, esp negative ones
- Need repository of negative studies– importance of denominator of studies
- Provide infrastructure funding for collaborations and consortia– critical for young investigators
- Give priority to funding studies that collaborate or are part of consortia
- How do we deal with ‘hostile’ takeovers when academic collaborations are challenged by commercial interests?

Things We at NIH Can Do (2)

- Supplement for reconsent, set timetables
- Support strategic genotyping
- Make sure DNA and samples collected in large-scale studies and trials
- Pre-publication analysis in dbGaP such as imputation, phenotype harmonization
- Can set up for rapid response– set aside repository samples waiting for replication
- Need for standardization of phenotypes for cross-study analyses

Things We at NIH Can Do (3)

- Stimulate better methods of targeted sequencing
- Extensively sequence HapMap samples to identify common SNPs– avoid redundant sequencing attempts
- Prospectively set up data collection process to be suitable for dbGaP
- Need to capture experience of data sharing
- Provide straw person consent elements– general use, sharing, genetics, future use, not discarded

Things that Can Come Out of This Meeting: Action Items

- Catalog GWA studies going on or planned
- Better (or at least standardized) phenotyping…
- Shift playing field to protect young investigators

Many Thanks to…

Stephen Chanock, NCI
Kat Gwinn, NINDS
Thomas Lehner, NIMH
Chris O'Donnell, NHLBI
Jim Ostell, NLM

Maggie Bartlett
Catherine Bennet
Mia Diggs
Eva Kakoza
Mukul Nerurkar
Steve Robinson
Larry Thompson
Susan Vasquez