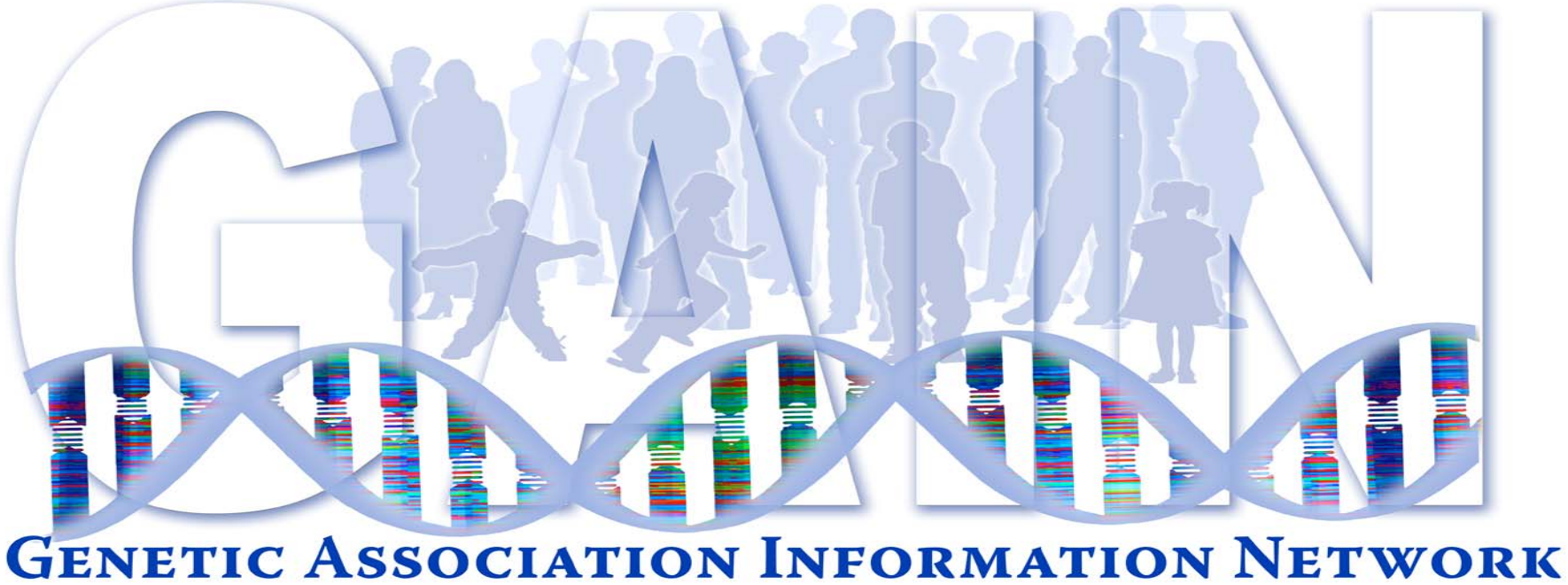
Opportunities for Analysis and Collaboration



Goals of GAIN Collaborations

- Facilitate rapid identification of genes related to complex traits by sharing experience
- Encourage development of open and stimulating scientific consortium for GAIN
- Build bridges across disciplines: many statistical methods are transportable but phenotype specific expertise critical to share
- Identify, address, and resolve quality control and database issues; make database more accessible for “naive” users
- Pursue unique opportunities for cross-study analyses of genotypes and phenotypes

Recommendations for Database

- Tools needed to make cluster files more accessible to investigators
- Flag quality of genotyping data
- Make all data available
- Allow for updating with new phenotyping or genotyping data, versioning with new builds
- Provide links to other databases

Pre-Computed Analyses

- Main purpose likely to be ensuring pre-competitive status of data
- As much as possible should be "scientifically valid"
- Need consultation with study PIs about how and what to show, may differ by study
- Provide caveats that pre-computes may differ from meticulously done analyses by those who know data best
- Should be versioning as analyses become more refined, or published
- QC should be done before pre-computes calculated

Considerations for Collaborative Analyses

- Issues specific to psychiatric genetics: schizophrenia vs. bipolar vs. depression
- Comparability and harmonization of phenotypes: DIGS common to several
- Comparison of family and case-control designs
- Comparison of associations in Caucasian and African-American participants

Considerations for Analyses Beyond Simple Genotype-Phenotype Associations

- Gene–gene interaction
- Gene–environment interaction (?)
- Copy number variants
- Population structure and other confounders

Analysis Discussions

- Each study present periodically on what questions they find challenging
- Analysts present questions they're struggling with and how to address them in GAIN

"Warm-Up" Genotyping Data Set: NINDS Open Access Repository

http://ccr.coriell.org/ninds/

Initial genome wide genotyping in:

- 276 Parkinson Disease (available now)
- 276 Stroke (available Jan 2007)
- 276 Amyotrophic Lateral Sclerosis (Jan 2007)
- 276 Controls
- 200 African Americans

Genotyping completed or planned:

- 109K exon-centric assay (phase I)
- 317K HapMap assay (phase II)
- >99.8% call rate, >450,000,000 unique genotypes
- >99.9% reproducibility (over 19,000,000 replicate genotypes)

Provide Best (Better/Good) Practices for Genome-Wide Association Field

- Standards for genotyping QC
- Standards for study design
- GAIN consortium papers on design, analytic approaches, etc
- Approaches for data sharing: protecting study participants, enhancing validity of outside analysis, protecting investigators' rights

Promote Collaborations with Other Consortia

- ENDGAME
- Cancer Genetic Markers of Susceptibility (CGEMS)
- Genes and Environment Initiative (GEI)
- Framingham SHARe
- Wellcome Trust