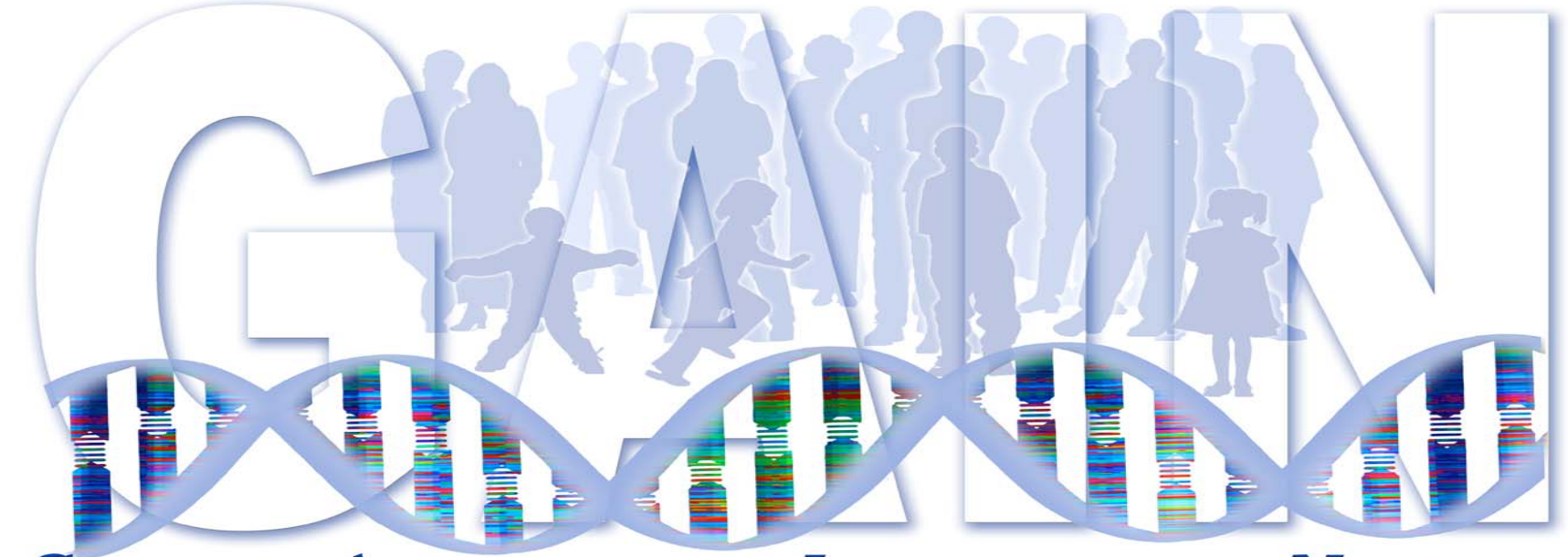
Lessons Learned, Action Items, Next Steps



What Are We Looking For?

- Common variants– this is what GAIN is designed to find
- Other possibilities:
  - Copy number variants
  - Rare variants with high heterogeneity
  - Functional variants (possibly larger effect sizes than with marker SNPs)
- Gene-gene and gene-environment interactions

Issues Related to Genotyping

- Genotyping QC pipeline is really cool and should be written up and disseminated
- Rare minor alleles present multiple QC challenges
  - Genotyping platforms that deal with these are urgently needed
  - Imputation boosts power for rare SNPs, but performs worse
- TDT is not immune to bias: genotyping bias rather than selection bias
- Training and refining of Birdseed algorithm had significant impact on quality and completeness

Issues Related to Analysis

- Interactions– is anyone looking?
- Combining scans for different diseases: disease cases based on pathophysiology, controls based on ancestral origin
- Bayes Factors “correct” p-values for low sample size and power
  - Interpreting p-values across studies of different sizes isn’t wise
  - Jonathan really likes Bayes Factors
- Not for the faint-hearted nor foolish

Copy Number Variants

- Need to refine calling methods for these regions
- Need analytic tools that deal with more than 3 genotypes at a locus
- Need better detection of CNVs
- Need to analyze SNPs and CNPs together

Population Stratification

- Cryptic relatedness, especially half-sibs, really skews a principal components analysis
- Some people participate in more than one study (socially responsible individuals)
- May be a heritable trait (first-degree relatives)
- Selection of SNPs for second stage: 7-13% are different if correct for PCA

Phenotypes

- “Sub-phenotypes” likely to have different GWA signals
  - Broad
  - Narrow
- Genetics may help to refine phenotypes

Need for Collaboration

- As always, larger samples needed
  - Increased power
  - Diversity across ancestral backgrounds and environmental exposures
  - Across phenotypes– shared genetic factors, “free phenotypes”
- What do we do when we run out?

Questions/Recommendations for NIH in Developing GWAS Policies

- Educational information for public, investigators
- How to deal with follow-up studies in terms of data deposition
- Clearer guidance on “exceptions” to data sharing: case-by-case with funding Institute
- Better examples of acceptable consent forms

Major Action Items, 10/18/07

- Write up genotyping QC methods and results
- Fix over-transmission of major allele in TDT
- Apply alternative calling algorithms to GAIN platforms and compare association results
- Compare six imputation methods and dare to choose a winner
- Develop BF that take covariates into account
- Calculate and disseminate Bayes Factors and compare association results
- Analyze SNPs and CNPs together

Major Action Items, 10/18/07

- Look for cryptic relatedness and socially responsible individuals
- May want to correct for PCA in selecting SNPs for second stage genotyping
- Develop educational materials for lay public
- Figure out how to combine GAIN control groups

Recommendations for Database 11/6/06

- ✓ •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
- ? •Tools needed to make cluster files more accessible to investigators

Other Issues in 11/06

- Pre-computed analyses: major concerns about scientific validity, caveats that pre-computes may differ from meticulously done analyses by those who know data best

Provide Best (Better/Good) Practices for Genome-Wide Association Field (11/06)

- 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

Issues Related to Data Sharing

- ACD Working Group to focus on requests that are difficult to resolve or denied
- Need for information/point of contact for:
  - Public: explain value of this research
  - Participants: from PIs how/as appropriate
  - Investigators submitting data: what to do
  - Investigators requesting data: what to do
- Unresolved issues:
  - Examine group harms as potential concern
  - Develop broad data-sharing consents
  - Return of results

Issues Related to Calling Algorithms

- Active area of productive research and clever names
- CHIAMO arguably provides measurable improvements over contemporary algorithms
- Training and refining of Birdseed algorithm had significant impact on quality and completeness
- Similar training and refining of Perlegen algorithm likely to address problem of over-transmission of major allele
- Look at SNPs that perform variably (“slide around”) across platforms for “fragile” genomic regions

Issues Related to Analysis

- Interactions– is anyone looking
- Combining scans for different diseases
  - Search for groups of disease cases that might logically be combined based on pathophysiology (autoimmune diseases in WTCCC)
  - Disease cases and other control groups that can be combined for disease-free control group or comparison cohort
- Not for the faint-hearted nor foolish

Issues Related to Analysis (2)

- Bayes Factors “correct” p-values for low sample size and power
  - Interpreting p-values across studies of different sizes isn’t wise
  - Jonathan really likes Bayes Factors
- Questions remaining:
  - How best to parameterize models
  - Need to develop BF that take covariates into account

Lessons Learned from Ongoing Studies

- Can we refine phenotype based on genotyping results?
- Many traits for “free” once you do GWA genotyping