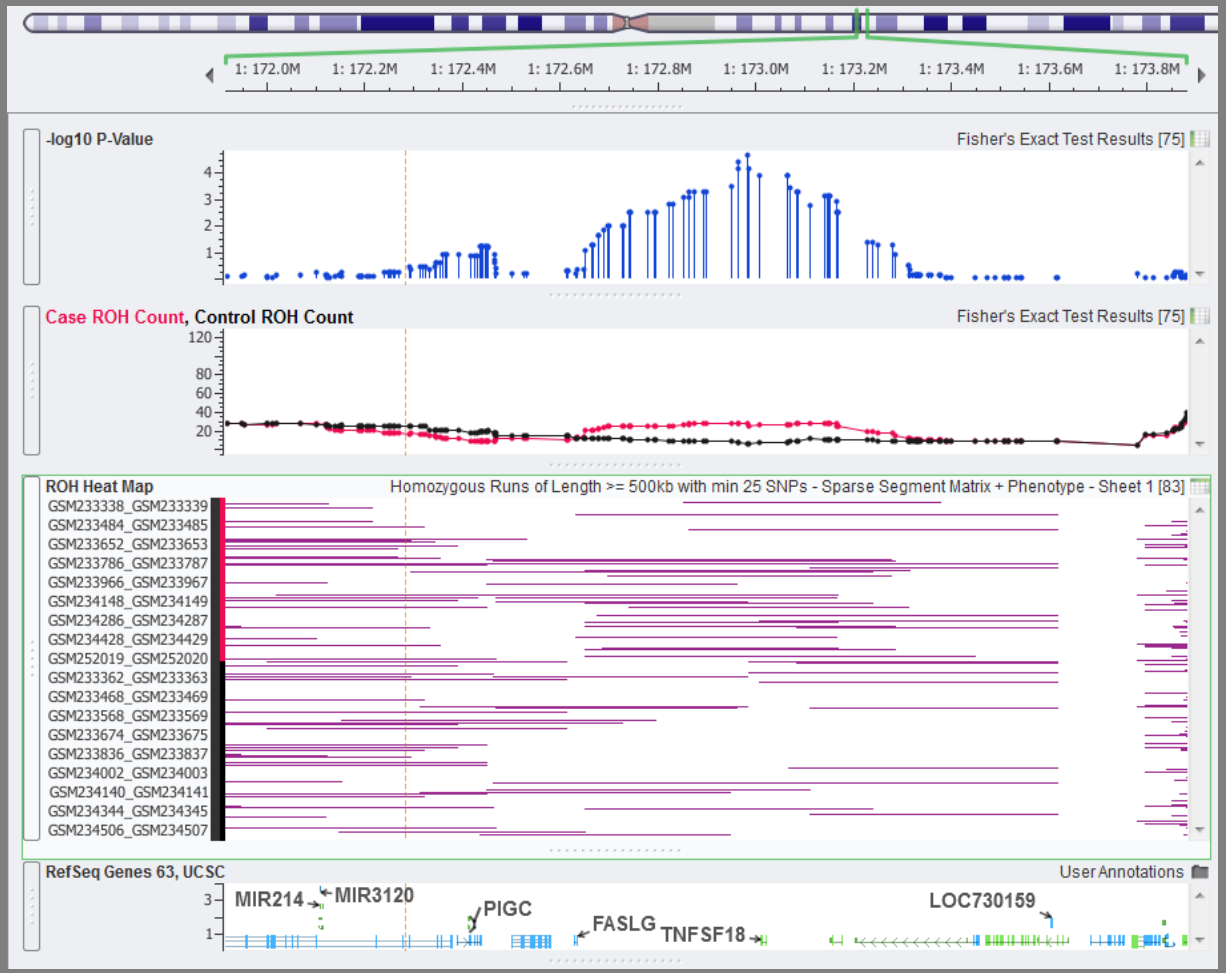


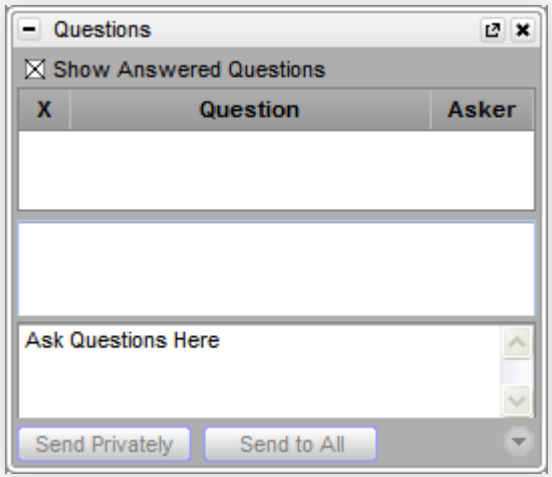
Getting More from GWAS

February 11, 2015

Bryce Christensen
Director of Services

Questions during the presentation

Use the Questions pane in your GoToWebinar window



About Golden Helix

Leaders in Genetic Analytics

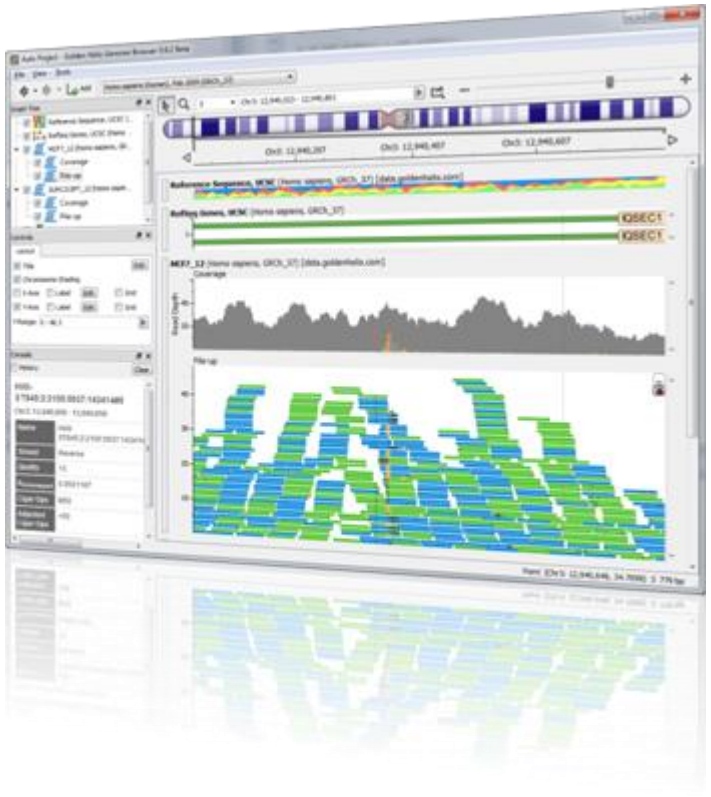
- Founded in 1998
- Multi-disciplinary: computer science, bioinformatics, statistics, genetics
- Software and analytic services
- Hundreds of literature citations

GOLDEN HELIX
Accelerating the Quest for Significance™

GenomeBrowse

- Powerful visualization software for DNA and RNA sequencing data
- Supports most standard bioinformatics file formats
- Fast and responsive for interactive analysis
- Intuitive controls
- Stream data from the cloud and from your own remote data servers

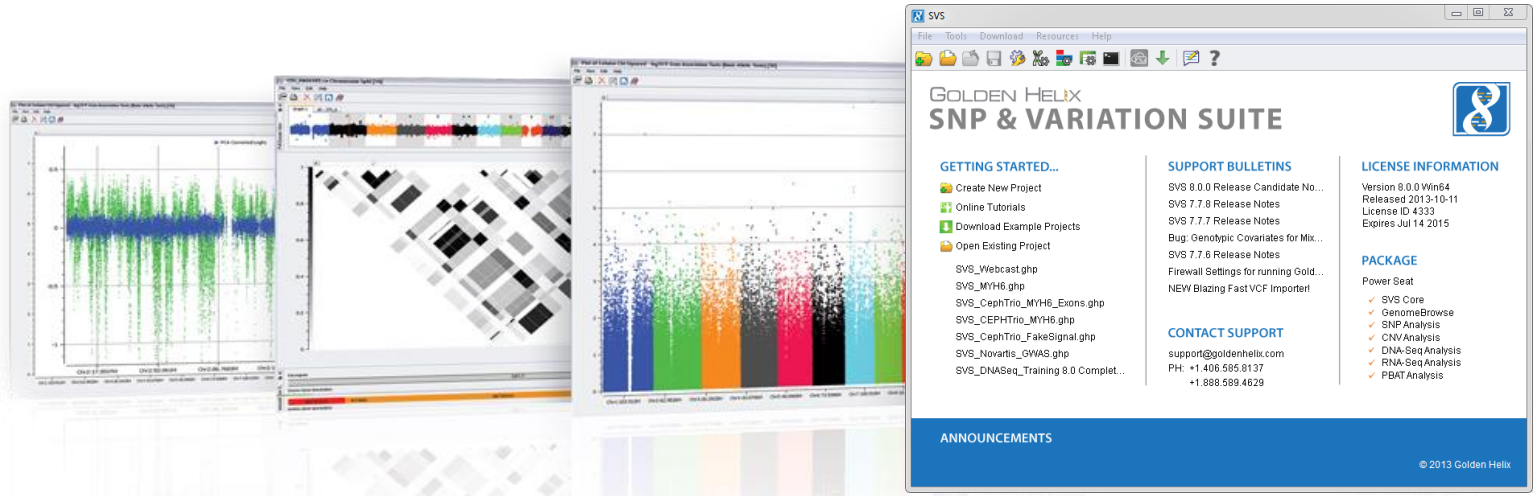
Golden Helix GenomeBrowse™



VarSeq

- **Powerful environment for annotation, filtering and visualization of DNAseq data**
- **Intuitive interface**
- **Repeatable workflows**
- **Optimized for clinical applications**

SNP & Variation Suite (SVS)



Core Features

- Powerful Data Management
- Rich Visualizations
- Robust Statistics
- Flexible

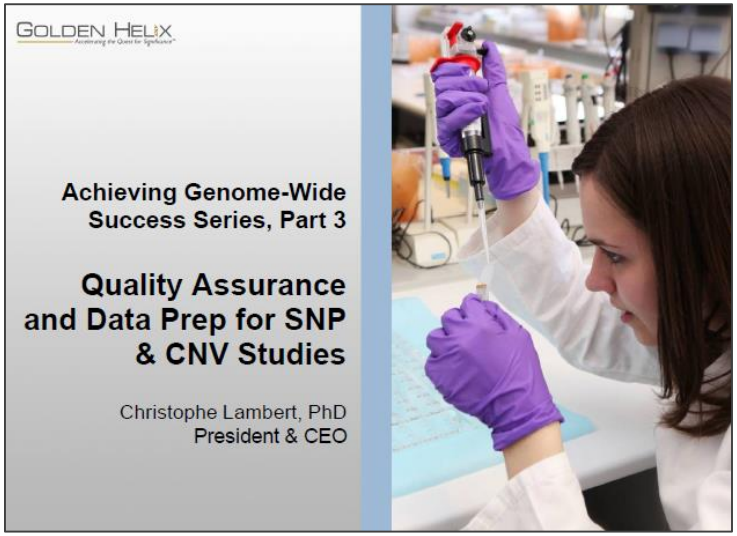
Applications

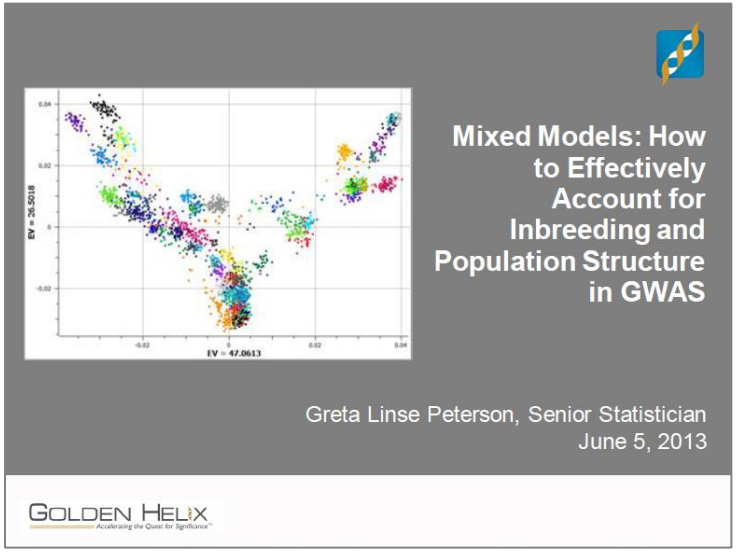
- Genotype Analysis
- DNA sequence analysis
- CNV Analysis
- RNA-seq differential expression

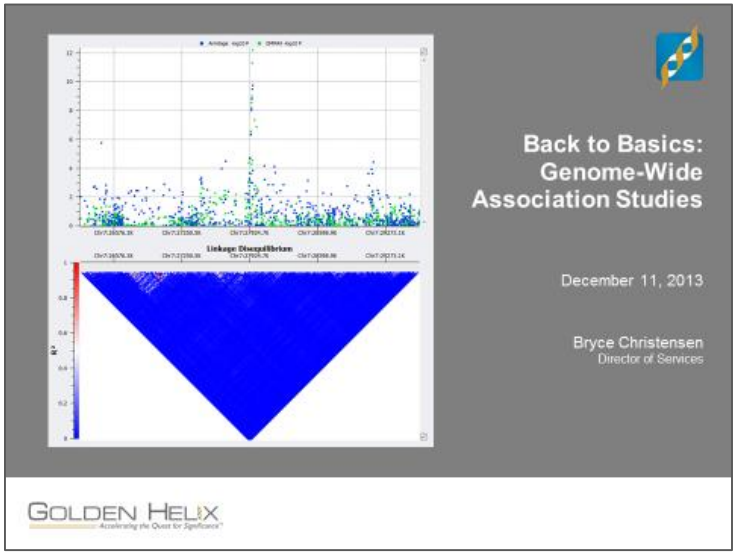
Approximate Agenda

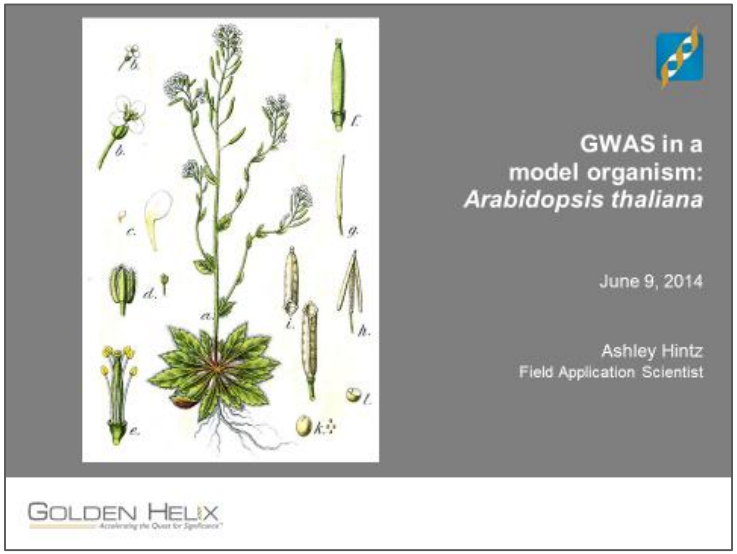
1. GWAS Background
2. Haplotypes and Haplotype Testing
3. Runs of Homozygosity
4. Q&A

Previous GWAS Webcasts Online

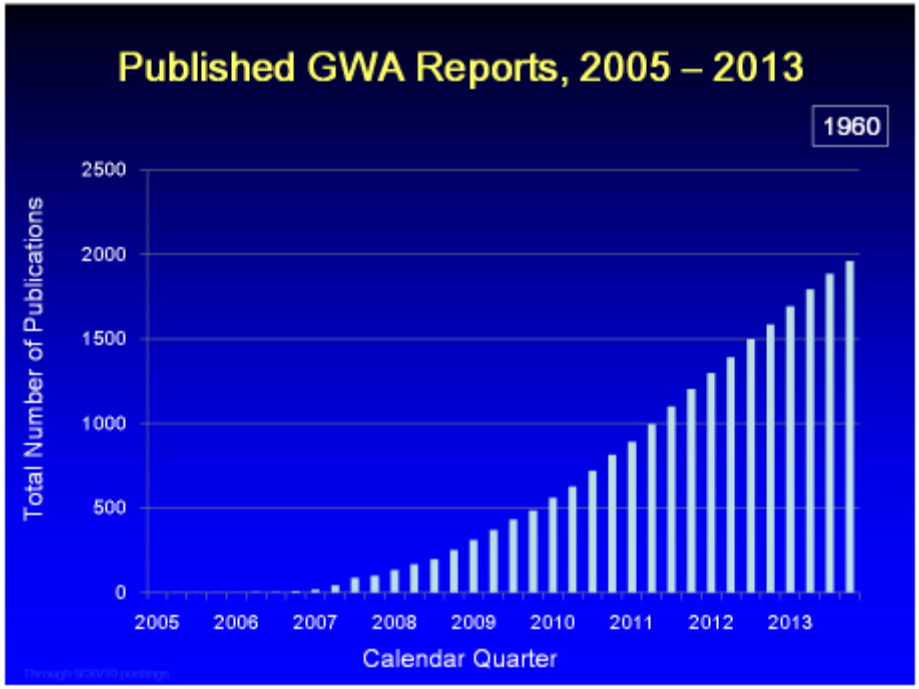








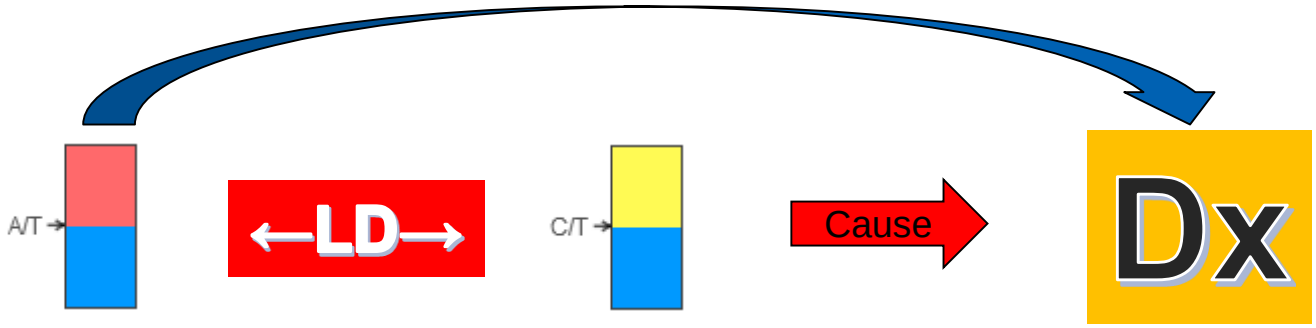
The GWAS Era



Slide Credit: Teri Manolio

- Array-based GWAS has been the primary technology for gene-finding research for the past decade
- Most published results are common variants with small effect on phenotype

GWAS is based on tag-SNPs



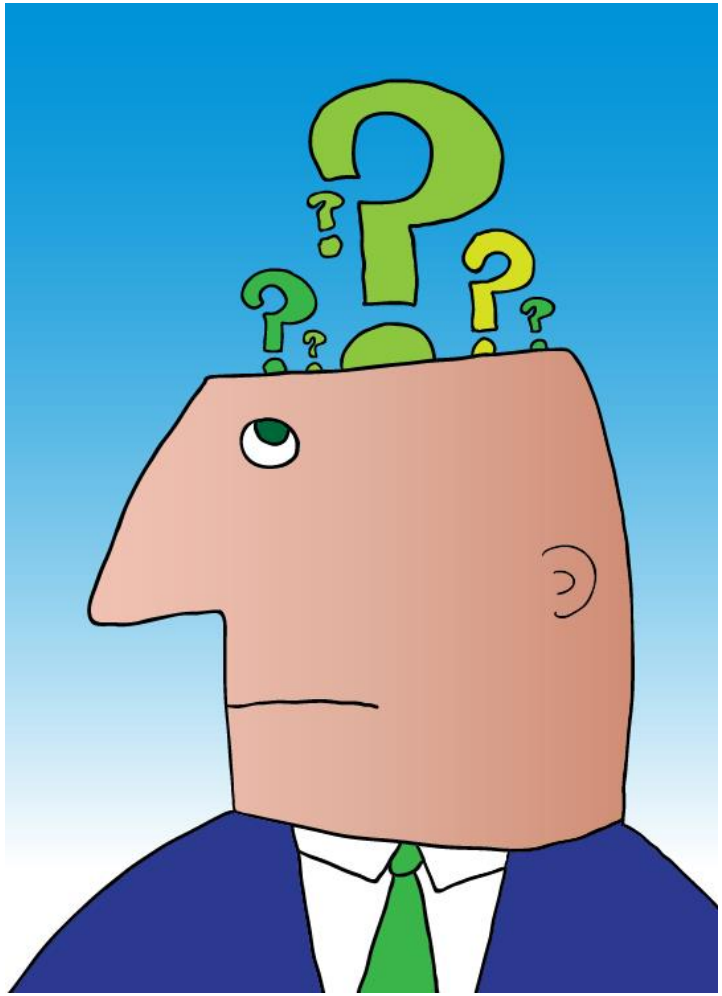
A. Observed GWAS SNP

B. Untyped Causal SNP

C. Disease Outcome

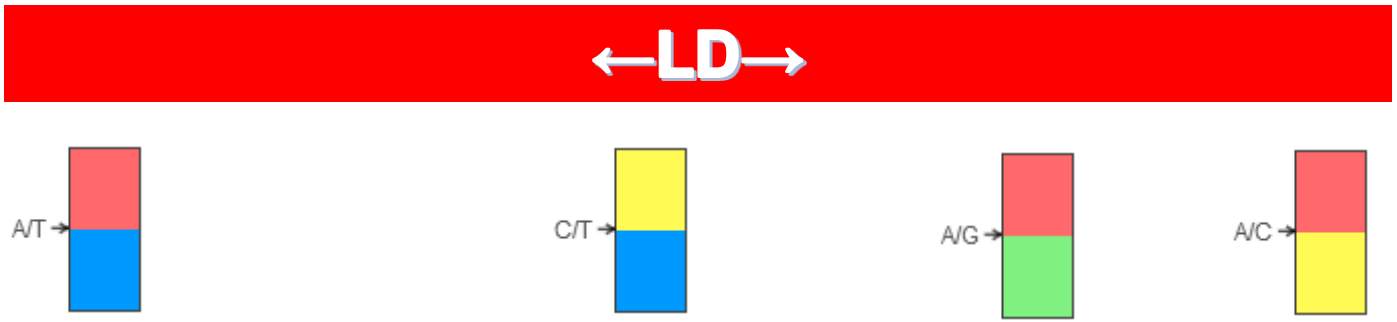
- **GWAS tests measure the relationship between A and C, assuming that B won't be tested on the array**
- **Tagging known variation is major consideration of GWAS array design**
- **But some variants may still be poorly represented, leading to potential type-2 errors and "missing heritability"**

How to get more from GWAS



- Can we learn more from existing data?
- Collecting more samples or investing in NGS isn't always possible.
- Combinations of neighboring markers may give a better representation of some un-genotyped SNPs.
- Haplotypes capture greater allelic diversity than individual SNPs.
- Runs of homozygosity (ROH) can capture trait associations that aren't specific to a particular allele.

From tag-SNPs to Haplotypes



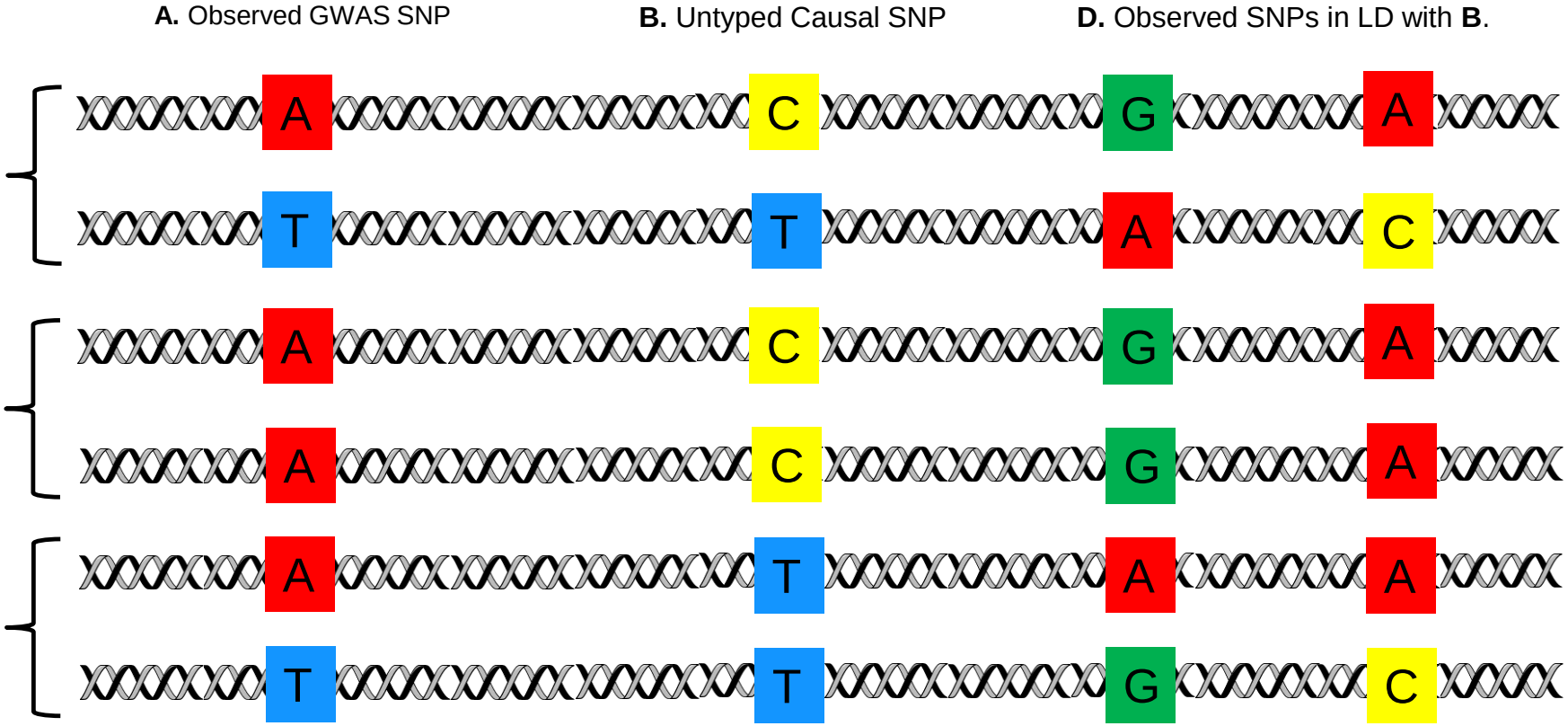
A. Observed GWAS SNP

B. Untyped Causal SNP

D. Additional GWAS SNPs in LD with **B**.

- **Haplotype: Set of sequential alleles found on a single chromosome**
- **Haplotypes tend to be conserved across generations and populations**
  - Linkage Disequlibrium (LD)
  - “LD Blocks”
- **Haplotypes may serve as markers for disease susceptibility loci**

Inference from Haplotypes



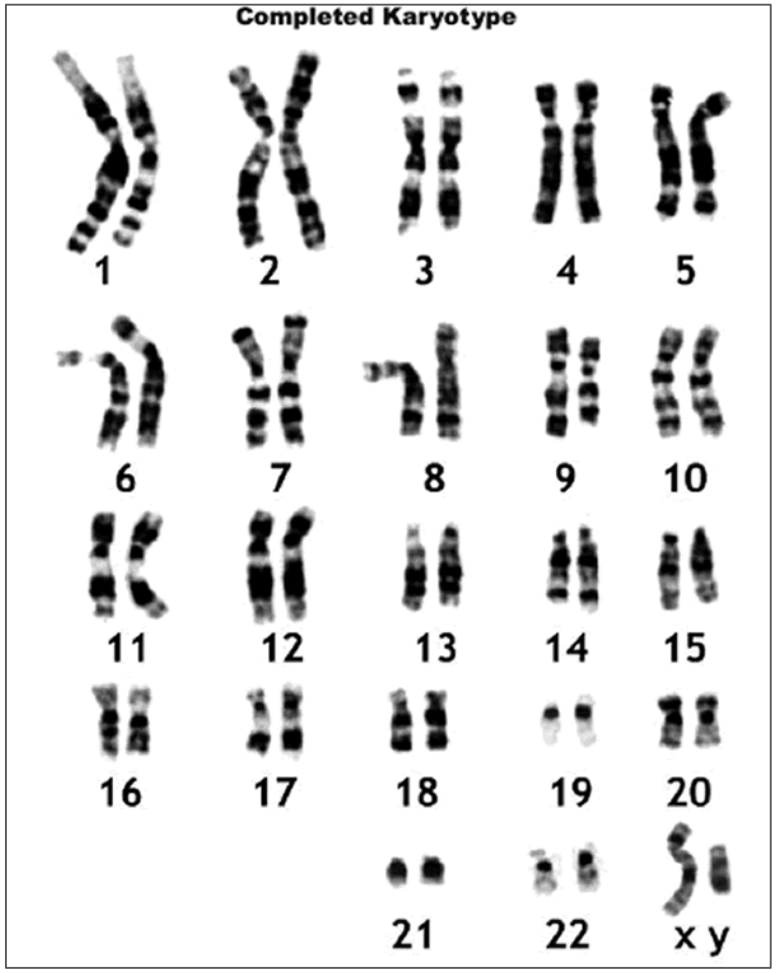
- **In this example, the C allele at position "B" only occurs when the 3 observed alleles on the haplotype are A-G-A.**
- **The A-G-A haplotype can serve as a marker for the causal C allele**
- **This is also the theoretical basis for genotype imputation**

GOLDEN HELIX

SNP & VARIATION SUITE

Demonstration

Haplotype Estimation

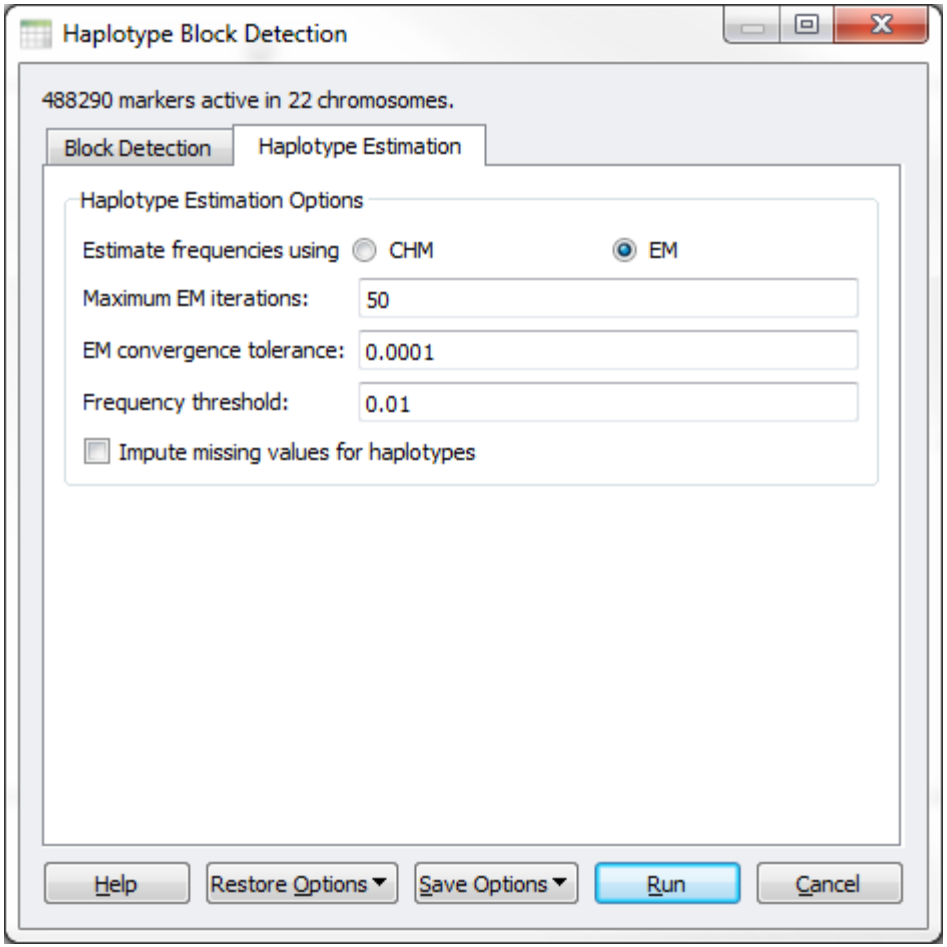


pixshark.com

- Gametic phase is generally unknown in GWAS data
- Haplotype inference, or “phasing,” is based on probability
- Allele frequency and co-occurrence rate in unphased genotypes inform phasing algorithm
- Accuracy improves with larger sample sizes
- Family information is very valuable if available

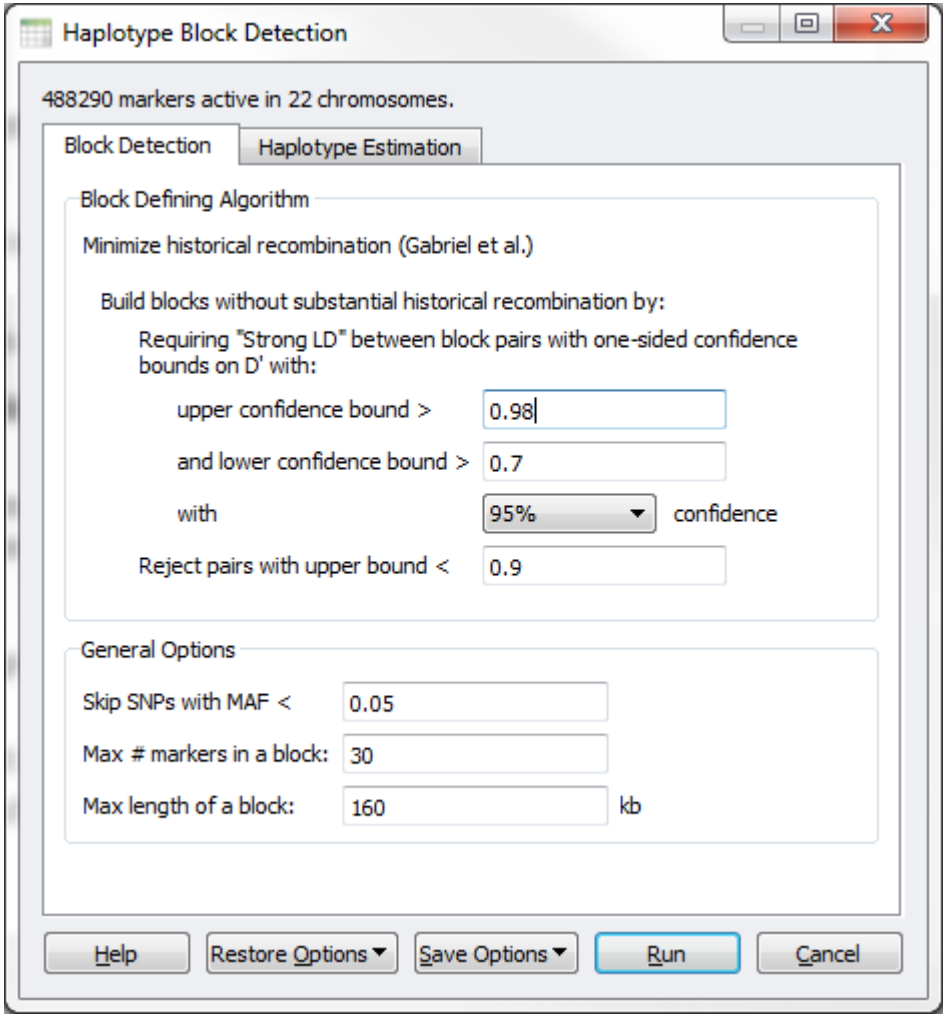
Haplotype Estimation in SVS

- **Expectation Maximization**
  - MLE of sample haplotype probability
  - Iterative process
  - Default method in SVS
- **Compound Haplotype Method**
  - Directly computed
  - Faster
  - Incorporates Hardy-Weinberg correction

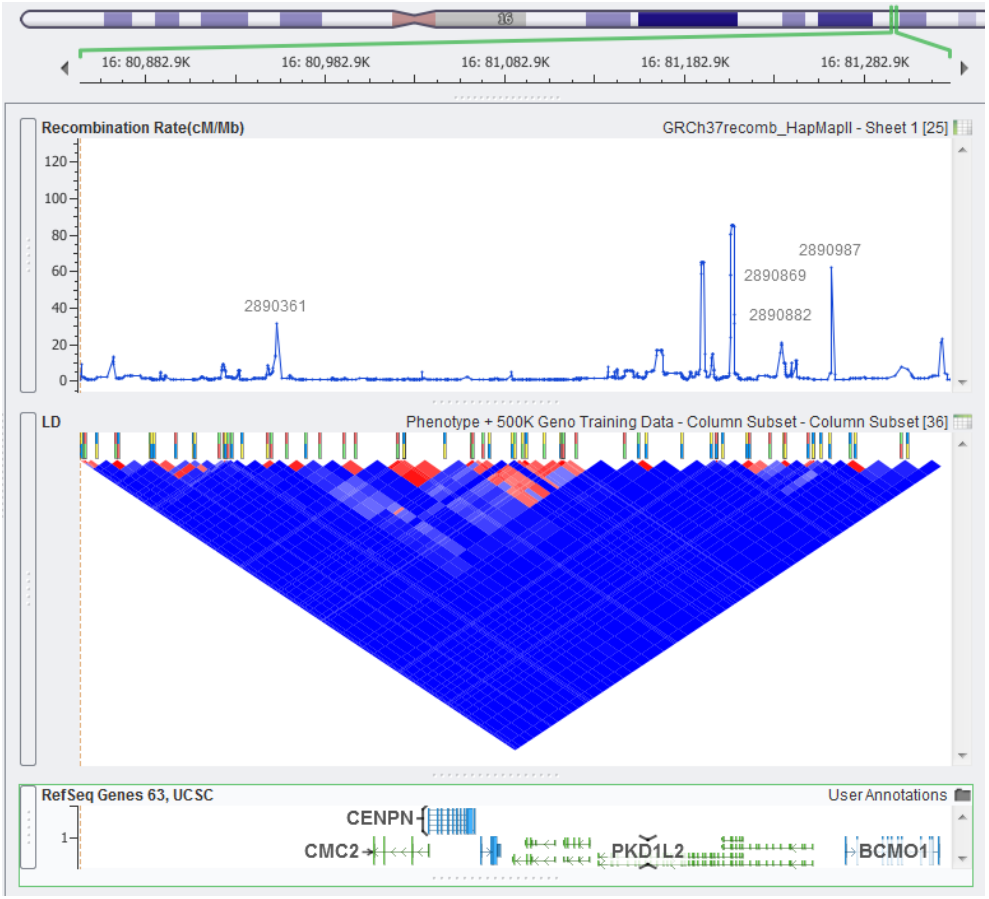


Options to Identify Haplotypes in SVS

- **Manual identification of marker blocks in GenomeBrowse LD plots.**
  - Useful for following up on regions of interest
- **Automated detection of LD blocks throughout the genome**
  - Algorithmic approach to identify marker blocks with minimal historic recombination
- **Sliding Window selection**
  - Window based on number of SNPs or region size in kilobase pairs.
  - Window size selection may affect results.
  - Exhaustive, “brute force” approach.



Two Haplotype Association Test Options in SVS



- **Haplotype Association Tests**
  - Binary traits
  - Chi-Square test
  - Test results per haplotype or per block
- **Haplotype Trend Regression**
  - Quantitative or binary traits
  - Allows covariate adjustment
  - Test results per block
  - Detailed output about individual haplotypes

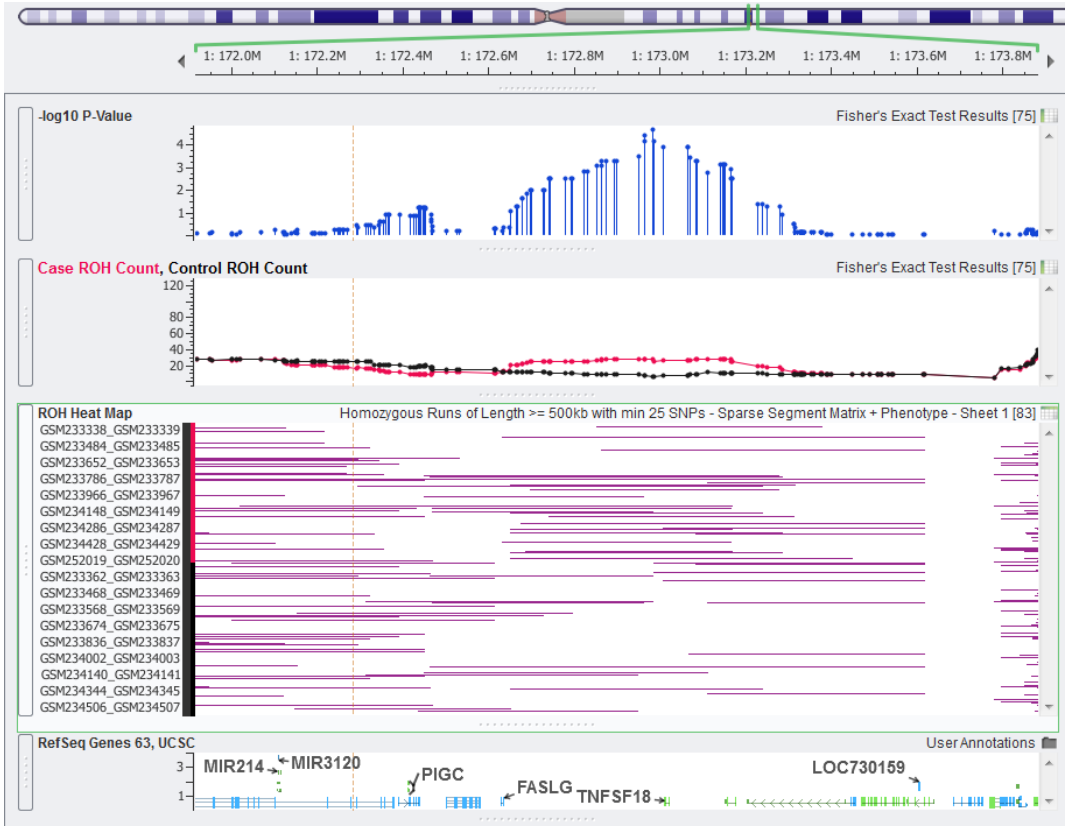
GOLDEN HELIX

SNP & VARIATION SUITE

Demonstration

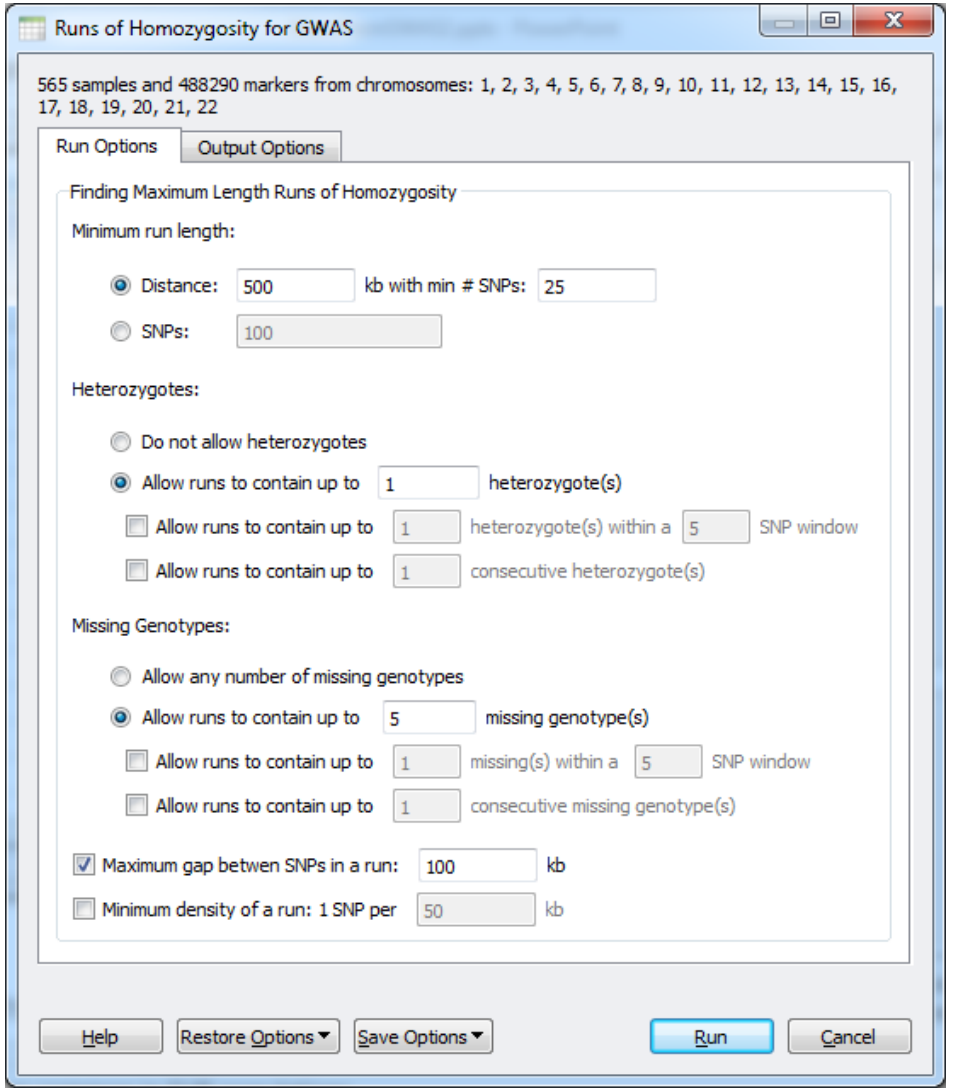
Runs of Homozygosity

- ROH identifies homozygous chromosomal segments of a certain length
- May be a marker for autozygosity and/or recessive trait inheritance
- ROH analysis does not require samples to have the same alleles
- Associates trait with a locus, but not a particular allele



ROH Options in SVS

- **Identify ROH segments by minimum number of SNPs or by minimum length**
- **Options to allow for missing data and sporadic heterozygous genotypes**
- **Maximum gap between SNPs**
- **Minimum SNP density**



ROH Outputs in SVS

- **ROH segment list**
  - Tip: Format is similar to CNV segment list
  - Some CNV output functions are very useful here!
- **Binary ROH status**
  - Merge this with phenotypes and use for association tests
- **Clusters of runs**
  - Multiple options to identify regions where a specified minimum number of samples have overlapping ROH segments.
  - “Optimal” clusters
  - Haplotype similarity clusters
  - May be used for association testing purposes

GOLDEN HELIX

SNP & VARIATION SUITE

Demonstration

Summary

- GWAS is useful for associating traits with common variants
- Haplotype analysis and other methods that consider multiple SNPs may reveal associations that are not evident in standard GWAS tests.
- ROH analysis can be useful to identify recessive trait loci and other genomic features
- SVS is a powerful platform for analysis of GWAS data

msue.anr.msu.edu

Questions or more info:

- Email info@goldenhelix.com
- Request an evaluation of the software at www.goldenhelix.com
- Check out our abstract competition!

Questions?

Use the Questions pane in your GoToWebinar window