

Getting More from GWAS

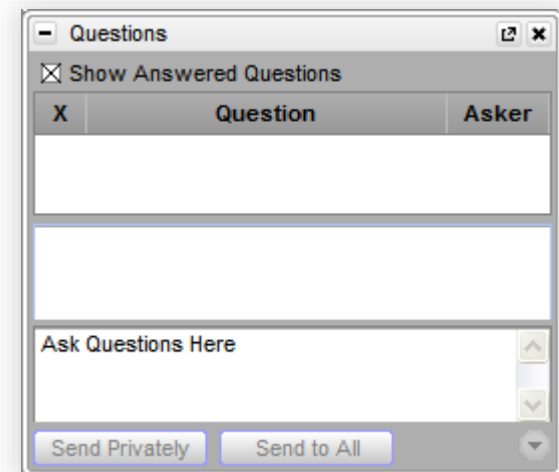
February 11, 2015

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

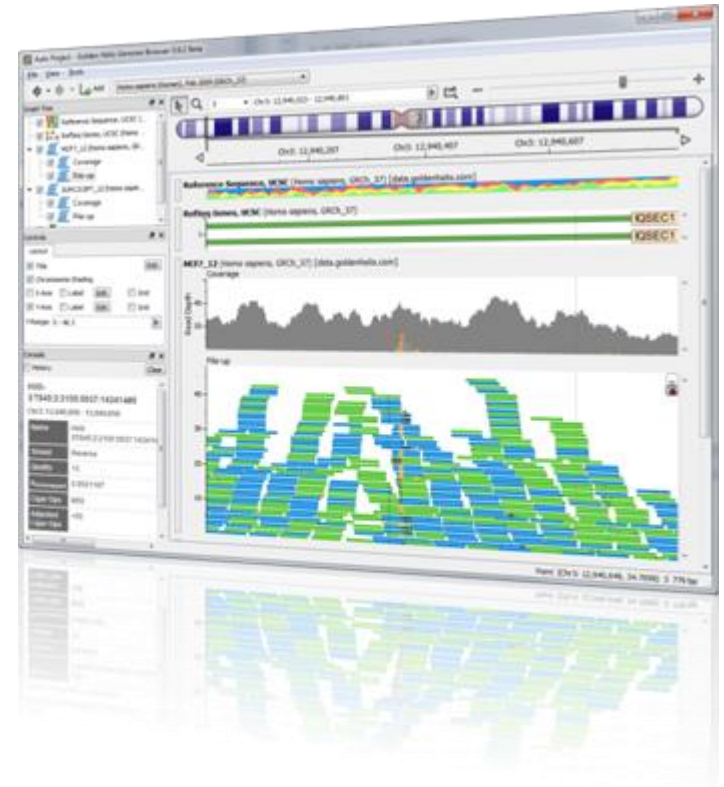
DISCOVERY DR

ENTERPRISE BLVD

GenomeBrowse



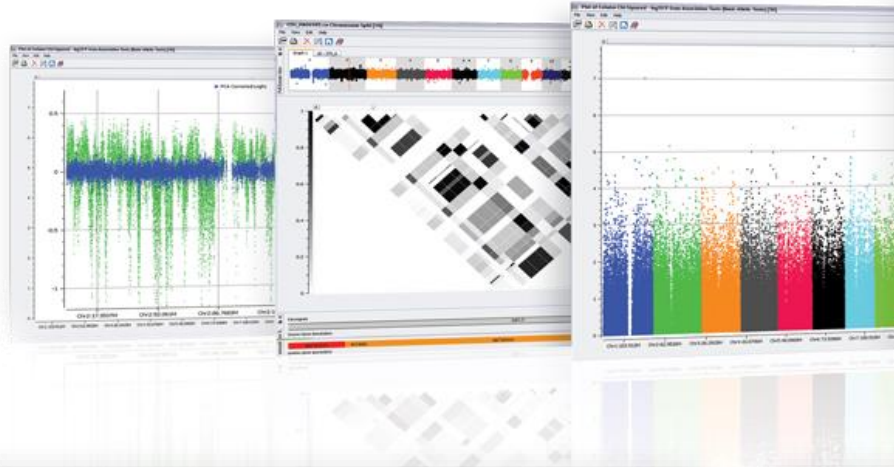
- Powerful visualization software for DNA and RNA sequencing data
- Supports most standard bioinformatics file formats
- Fast and responsive for interactive analysis
- Intuitive controls
- Stream data from the cloud and from your own remote data servers





- 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



GOLDEN HELIX
Accelerating the Quest for Significance™

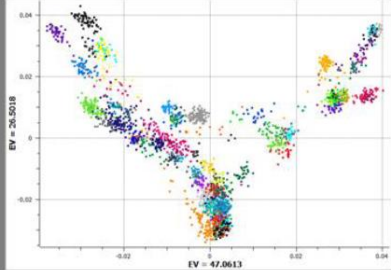
Achieving Genome-Wide Success Series, Part 3

Quality Assurance and Data Prep for SNP & CNV Studies

Christophe Lambert, PhD
President & CEO



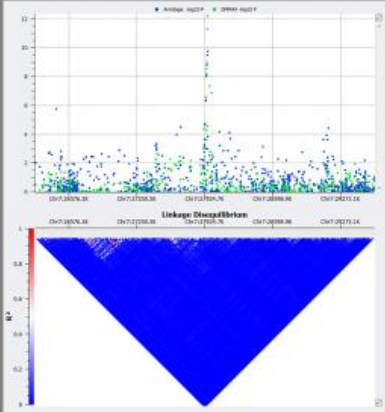
GOLDEN HELIX
Accelerating the Quest for Significance™



Mixed Models: How to Effectively Account for Inbreeding and Population Structure in GWAS

Greta Linse Peterson, Senior Statistician
June 5, 2013

GOLDEN HELIX
Accelerating the Quest for Significance™




Back to Basics: Genome-Wide Association Studies

December 11, 2013

Bryce Christensen
Director of Services

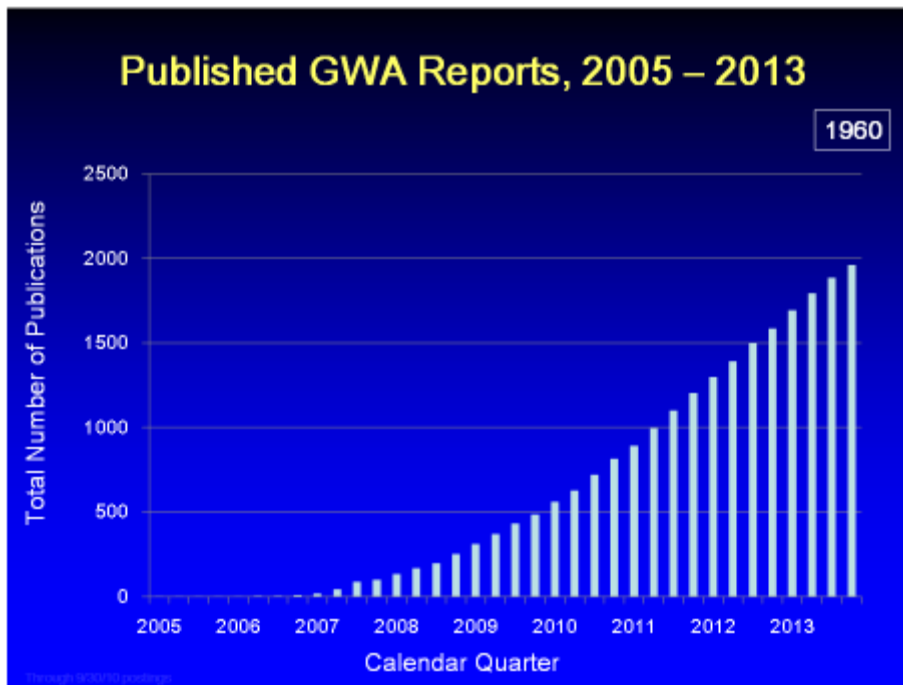
GOLDEN HELIX
Accelerating the Quest for Significance™



GWAS in a model organism: Arabidopsis thaliana

June 9, 2014

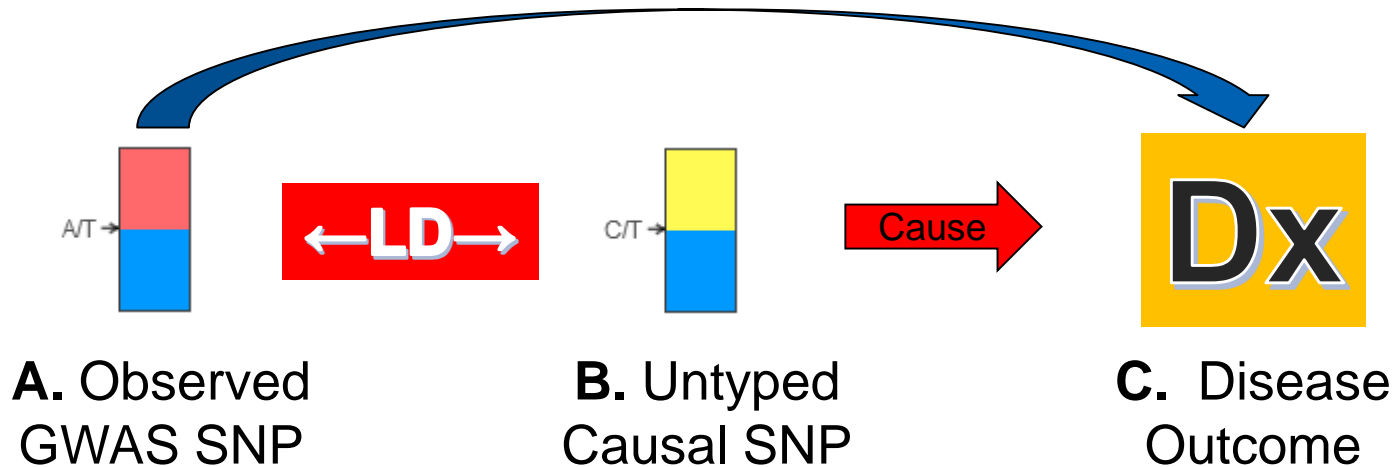
Ashley Hintz
Field Application Scientist



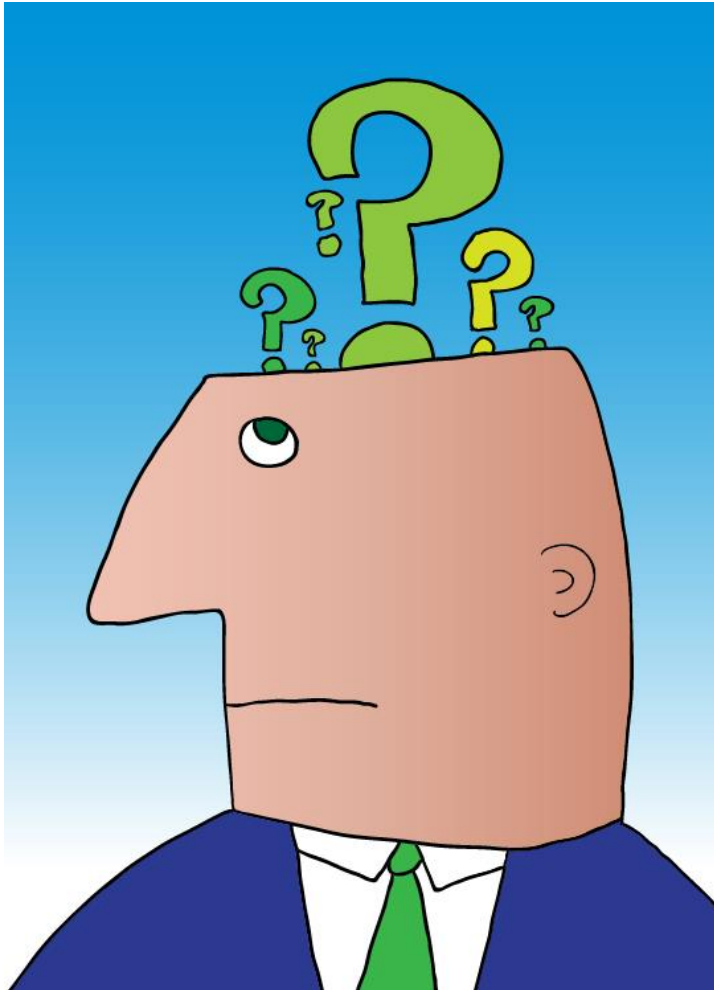
Slide Credit: Teri Manolio

- 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



- **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"**

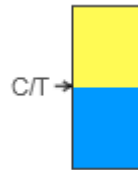


- Can we learn more from existing data?
- Collecting more samples or investing in NGS isn't always possible.
- Combinations of neighboring markers may give a better representation of some ungenotyped SNPs.
- Haplotypes capture greater allelic diversity than individual SNPs.
- Runs of homozygosity (ROH) 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 Disequilibrium (LD)
 - “LD Blocks”
- **Haplotypes may serve as markers for disease susceptibility loci**

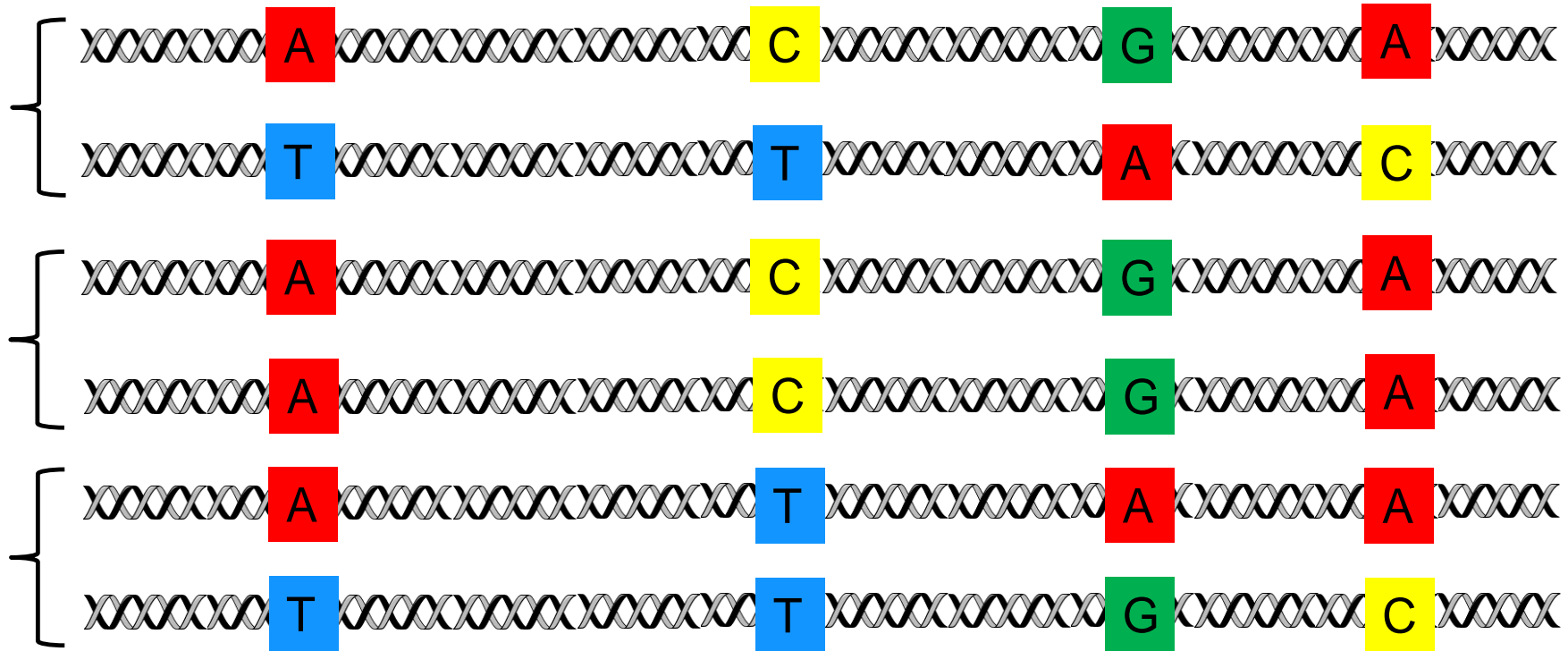
Inference from Haplotypes



A. Observed GWAS SNP

B. Untyped Causal SNP

D. Observed SNPs in LD with B.



- 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

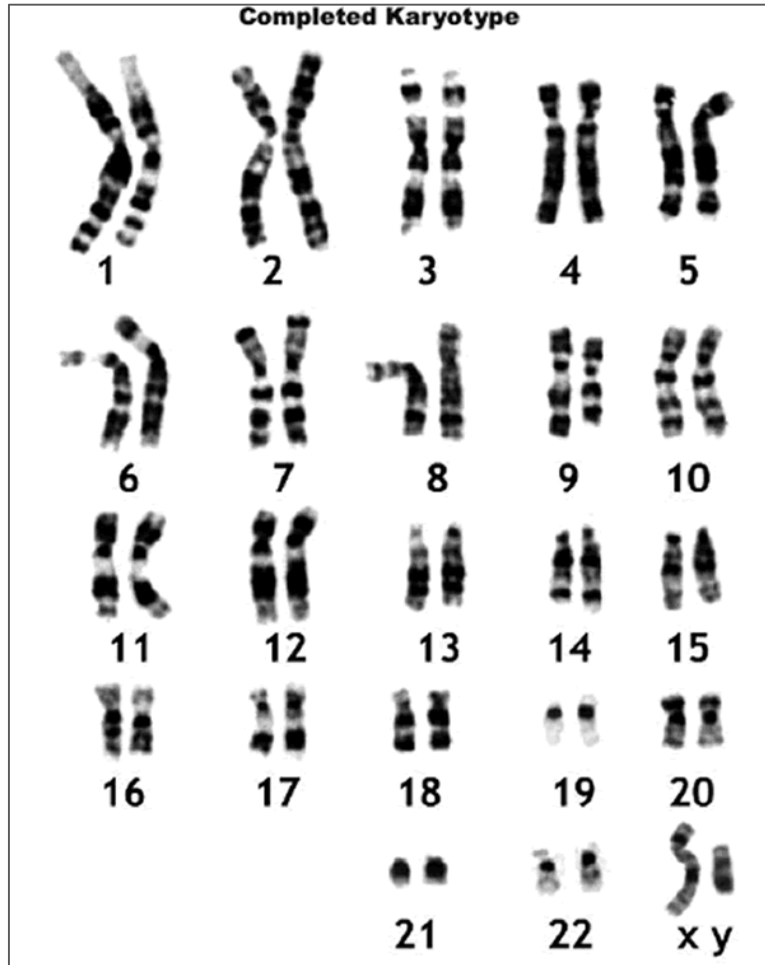


GOLDEN HELIX

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

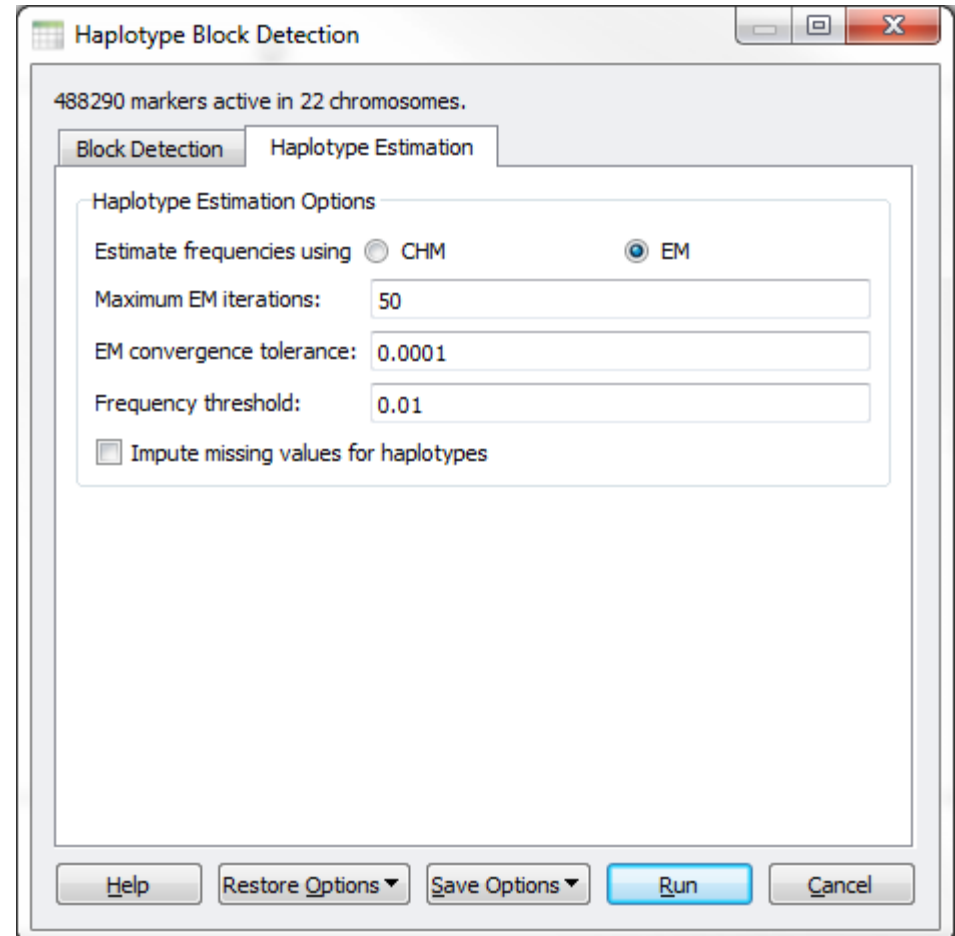


- **Expectation Maximization**

- 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.

Haplotype Block Detection

488290 markers active in 22 chromosomes.

Block Detection | Haplotype Estimation

Block Defining Algorithm

Minimize historical recombination (Gabriel et al.)

Build blocks without substantial historical recombination by:

Requiring "Strong LD" between block pairs with one-sided confidence bounds on D' with:

upper confidence bound > 0.98

and lower confidence bound > 0.7

with 95% confidence

Reject pairs with upper bound < 0.9

General Options

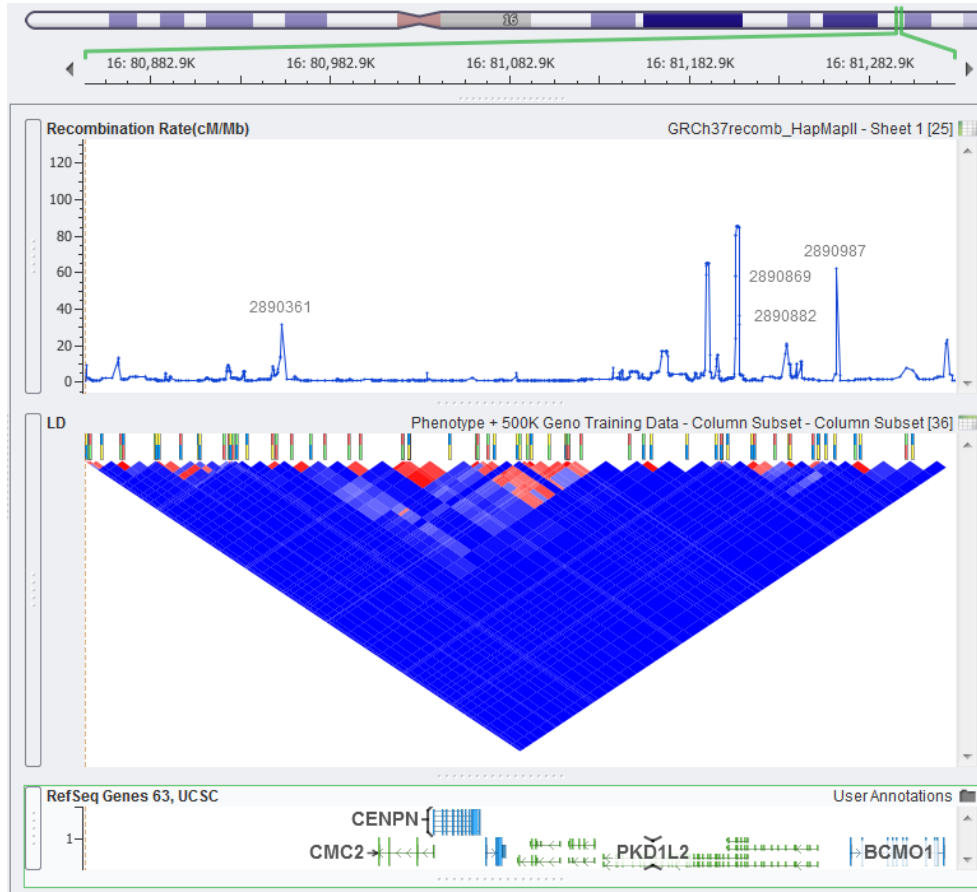
Skip SNPs with MAF < 0.05

Max # markers in a block: 30

Max length of a block: 160 kb

Help | Restore Options | Save Options | Run | Cancel

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

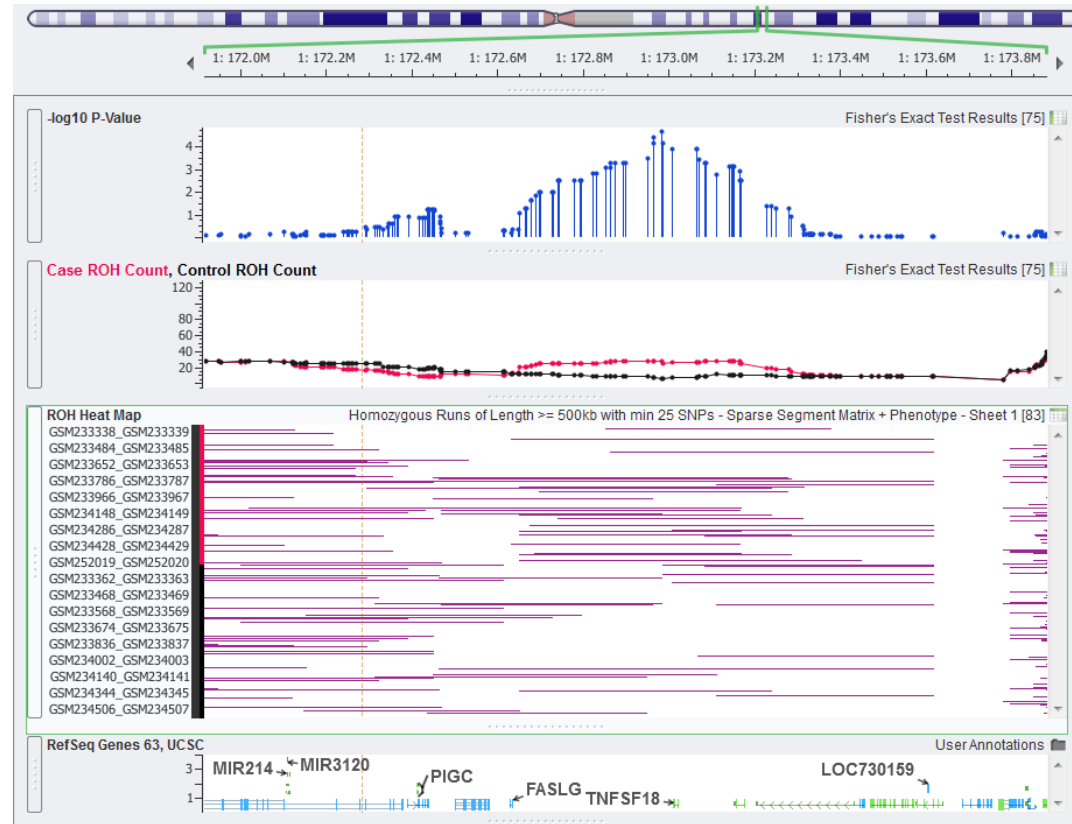
SNP & VARIATION SUITE

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

The screenshot shows the 'Runs of Homozygosity for GWAS' dialog box. At the top, it displays '565 samples and 488290 markers from chromosomes: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22'. The dialog has two tabs: 'Run Options' (selected) and 'Output Options'. Under 'Run Options', the section 'Finding Maximum Length Runs of Homozygosity' contains the following settings:

- Minimum run length:**
 - Distance: 500 kb with min # SNPs: 25
 - SNPs: 100
- Heterozygotes:**
 - Do not allow heterozygotes
 - Allow runs to contain up to 1 heterozygote(s)
 - Allow runs to contain up to 1 heterozygote(s) within a 5 SNP window
 - Allow runs to contain up to 1 consecutive heterozygote(s)
- Missing Genotypes:**
 - Allow any number of missing genotypes
 - Allow runs to contain up to 5 missing genotype(s)
 - Allow runs to contain up to 1 missing(s) within a 5 SNP window
 - Allow runs to contain up to 1 consecutive missing genotype(s)
- Maximum gap between SNPs in a run: 100 kb
- Minimum density of a run: 1 SNP per 50 kb

At the bottom, there are buttons for 'Help', 'Restore Options', 'Save Options', 'Run', and 'Cancel'.



- **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



- **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

