A Walk Through GWAS

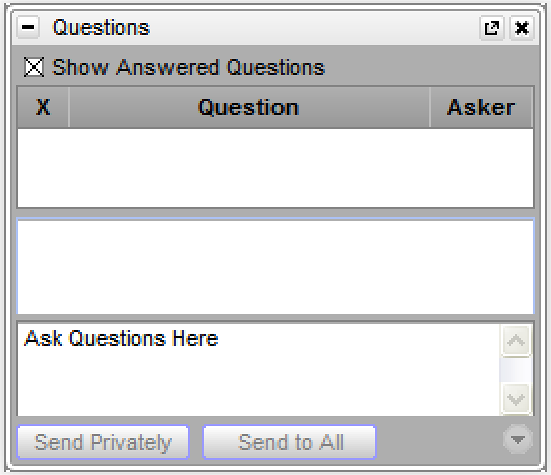
January 20th, 2016

Jami Bartole
Senior Field Application Scientist

GOLDEN HELIX
Accelerating the Quest for Significance™

Questions during the presentation

Use the Questions pane in your GoToWebinar window



Golden Helix – Who We Are

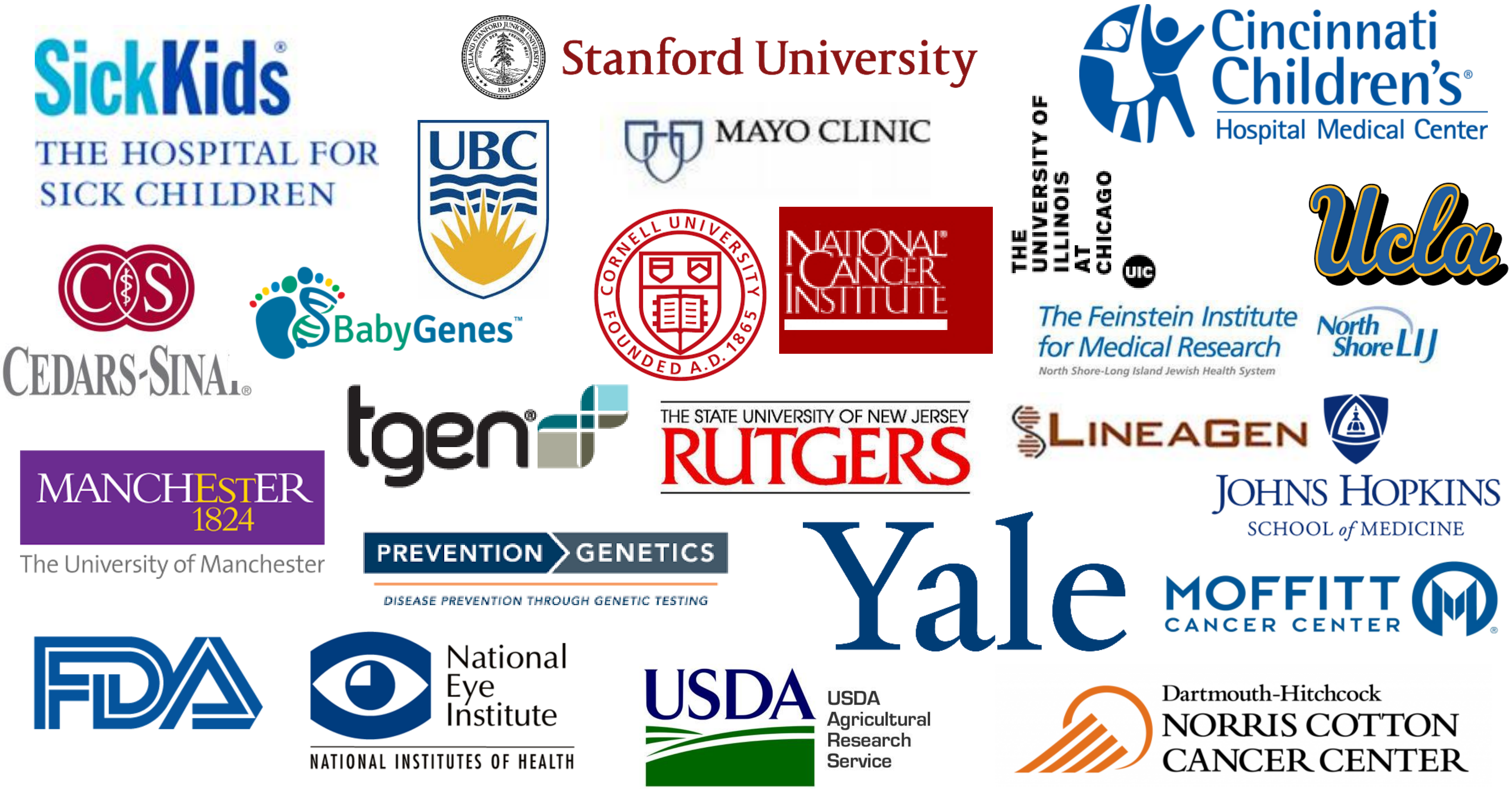
Golden Helix is a global bioinformatics company founded in 1998.

GlaxoSmithKline

We are cited in over 900 peer-reviewed publications

Our Customers

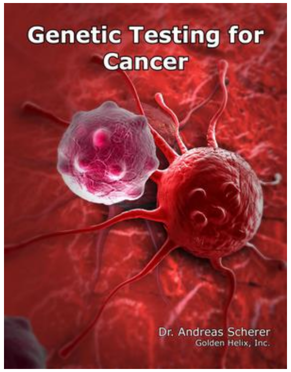
Over 200 organizations world wide, and thousands of users, trust our software.


SickKids®
THE HOSPITAL FOR SICK CHILDREN
Stanford University
Cincinnati Children's®
Hospital Medical Center
UBC
MAYO CLINIC
THE UNIVERSITY OF ILLINOIS AT CHICAGO
UIC
Ucla
CS
CEDARS-SINAI®
BabyGenes™
CORNELL UNIVERSITY FOUNDED A.D. 1865
NATIONAL CANCER INSTITUTE
The Feinstein Institute for Medical Research
North Shore-Long Island Jewish Health System
North Shore LIJ
tgen®
THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS
LINEAGEN
JOHNS HOPKINS
SCHOOL of MEDICINE
MANCHESTER 1824
The University of Manchester
PREVENTION GENETICS
DISEASE PREVENTION THROUGH GENETIC TESTING
Yale
MOFFITT CANCER CENTER
FDA
National Eye Institute
NATIONAL INSTITUTES OF HEALTH
USDA
USDA Agricultural Research Service
Dartmouth-Hitchcock
NORRIS COTTON CANCER CENTER

Golden Helix – Who We Are

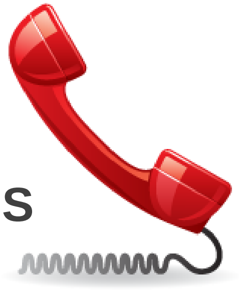
When you choose a Golden Helix solution, you get more than just software

- REPUTATION
- TRUST
- EXPERIENCE



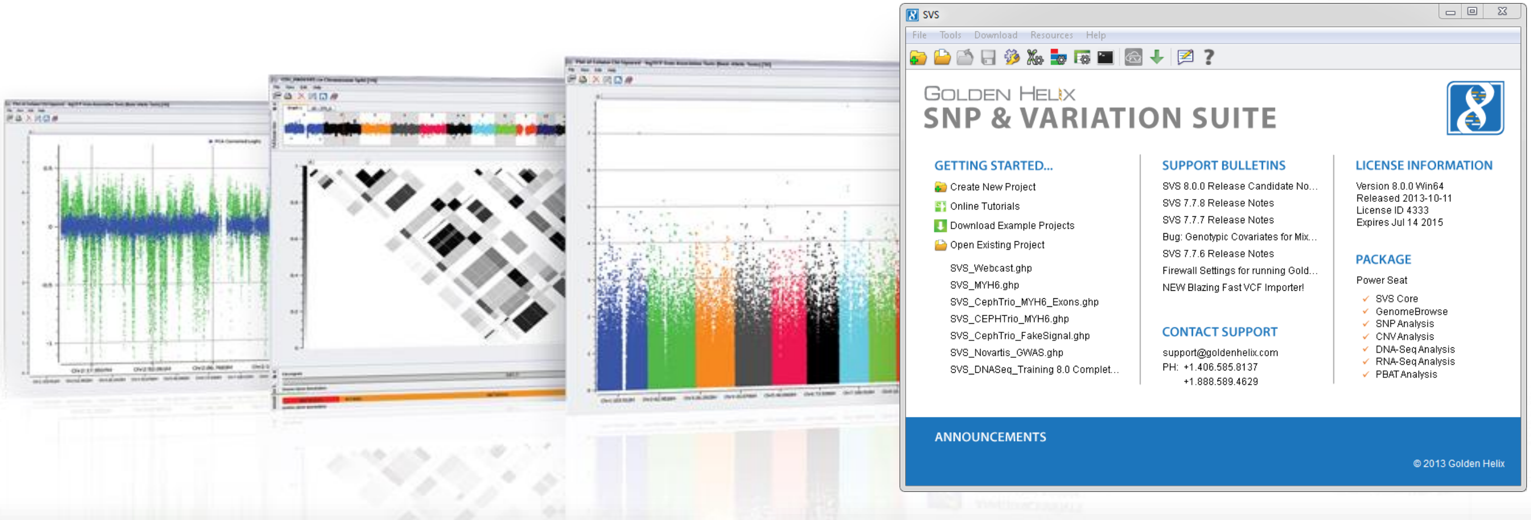
- INDUSTRY FOCUS
- THOUGHT LEADERSHIP
- COMMUNITY

- TRAINING
- SUPPORT
- RESPONSIVENESS



- TRANSPARENCY
- INNOVATION and SPEED
- CUSTOMIZATIONS

SNP & Variation Suite (SVS)



GOLDEN HELIX
SNP & VARIATION SUITE

GETTING STARTED...

- Create New Project
- Online Tutorials
- Download Example Projects
- Open Existing Project

SVS_Webcast.ghp
SVS_MYH6.ghp
SVS_CephTrio_MYH6_Exons.ghp
SVS_CEPHTrio_MYH6.ghp
SVS_CephTrio_FakeSignal.ghp
SVS_Novartis_GWAS.ghp
SVS_DNASeq_Training 8.0 Complet...

SUPPORT BULLETINS

SVS 8.0.0 Release Candidate No...
SVS 7.7.8 Release Notes
SVS 7.7.7 Release Notes
Bug: Genotypic Covariates for Mix...
SVS 7.7.6 Release Notes
Firewall Settings for running Gold...
NEW Blazing Fast VCF Importer!

CONTACT SUPPORT

support@goldenhelix.com
PH: +1.406.585.8137
+1.888.589.4629

LICENSE INFORMATION

Version 8.0.0 Win64
Released 2013-10-11
License ID 4333
Expires Jul 14 2015

PACKAGE

Power Seat

- ✓ SVS Core
- ✓ GenomeBrowse
- ✓ SNP Analysis
- ✓ CNV Analysis
- ✓ DNA-Seq Analysis
- ✓ RNA-Seq Analysis
- ✓ PBAT Analysis

ANNOUNCEMENTS

© 2013 Golden Helix

Core Features

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

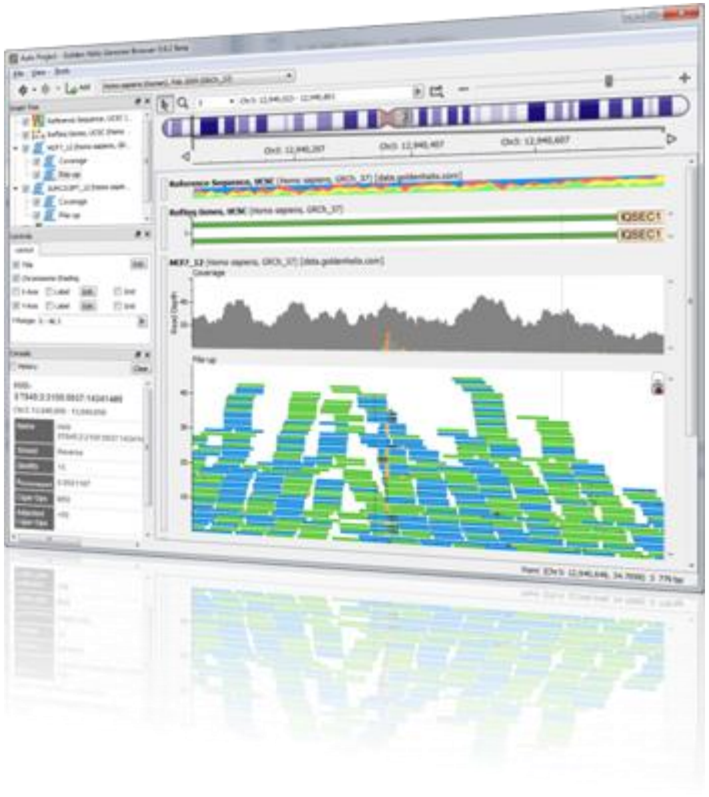
Applications

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

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™



Approximate Agenda

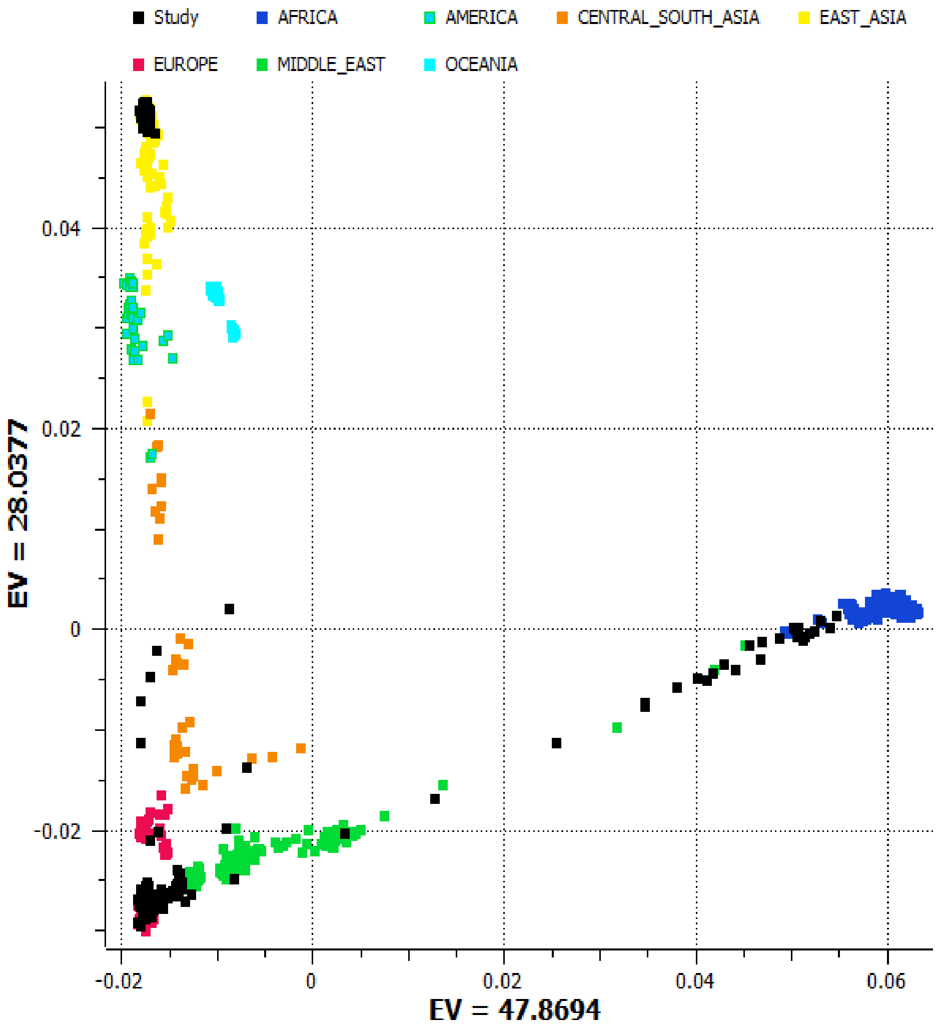
1 Background of GWAS

2 Explore Results from a GWAS Project

3 Population Stratification in Analysis

4 Q&A

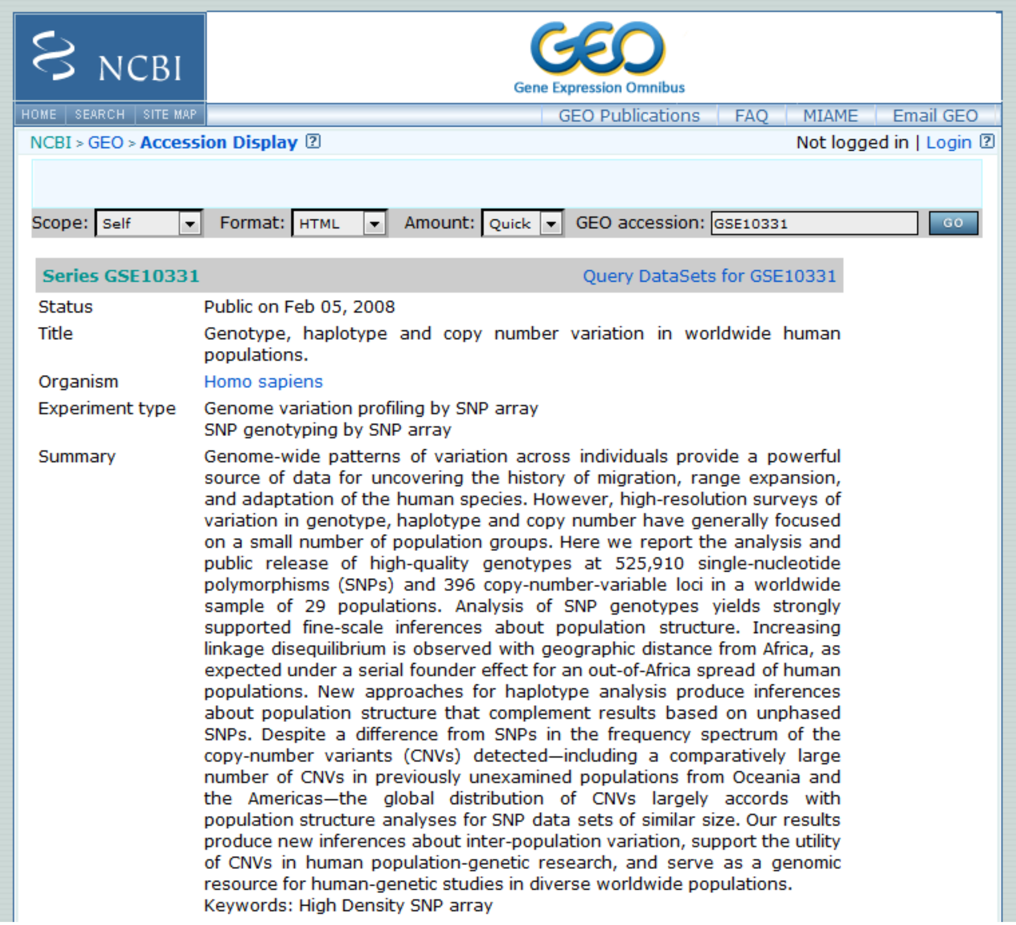
A brief background of GWAS



- First the naïve approaches: Trend Tests, Contingency Tables, Linear/Logistic Regression
- Batch Effects, Population Structure and sharing of controls may violate assumptions of the naïve approaches and result in confounding of results.
- Stratification effects are more pronounced with larger sample sizes.
- Non-independence of samples is especially problematic in agrigenomic applications.

Summary of GWAS Dataset

- 513 individuals
- 29 populations from the Human Genome Diversity Project (HGDP)
- Illumina Infinium HumanHap550 Genotyping BeadChip
- Simulated Case/Control Phenotype



Scope: Self Format: HTML Amount: Quick GEO accession: GSE10331

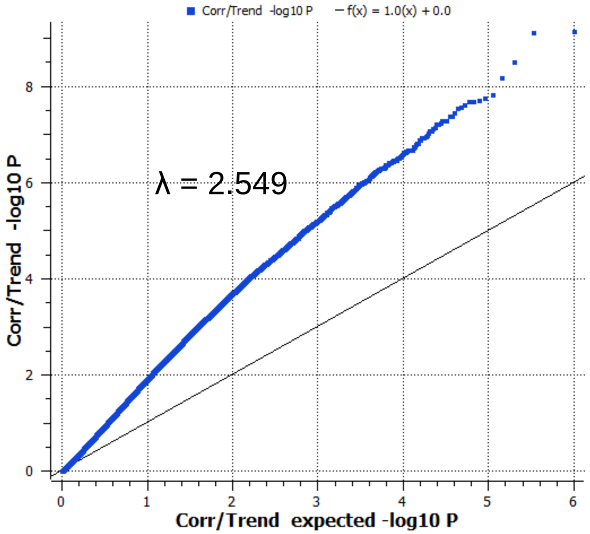
Series GSE10331 Query DataSets for GSE10331

Status	Public on Feb 05, 2008
Title	Genotype, haplotype and copy number variation in worldwide human populations.
Organism	Homo sapiens
Experiment type	Genome variation profiling by SNP array SNP genotyping by SNP array
Summary	Genome-wide patterns of variation across individuals provide a powerful source of data for uncovering the history of migration, range expansion, and adaptation of the human species. However, high-resolution surveys of variation in genotype, haplotype and copy number have generally focused on a small number of population groups. Here we report the analysis and public release of high-quality genotypes at 525,910 single-nucleotide polymorphisms (SNPs) and 396 copy-number-variable loci in a worldwide sample of 29 populations. Analysis of SNP genotypes yields strongly supported fine-scale inferences about population structure. Increasing linkage disequilibrium is observed with geographic distance from Africa, as expected under a serial founder effect for an out-of-Africa spread of human populations. New approaches for haplotype analysis produce inferences about population structure that complement results based on unphased SNPs. Despite a difference from SNPs in the frequency spectrum of the copy-number variants (CNVs) detected—including a comparatively large number of CNVs in previously unexamined populations from Oceania and the Americas—the global distribution of CNVs largely accords with population structure analyses for SNP data sets of similar size. Our results produce new inferences about inter-population variation, support the utility of CNVs in human population-genetic research, and serve as a genomic resource for human-genetic studies in diverse worldwide populations. Keywords: High Density SNP array

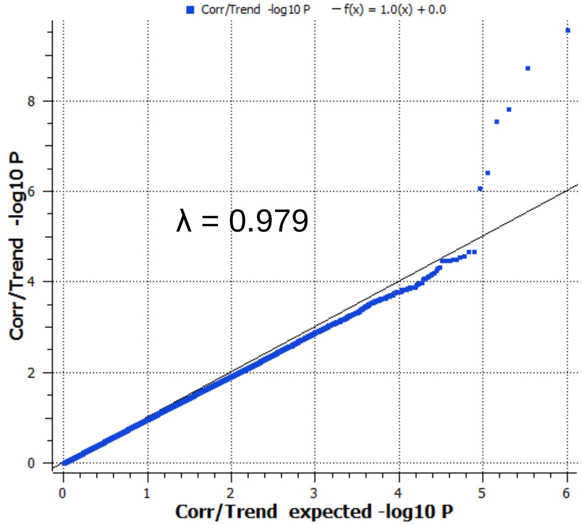
GOLDEN HELIX

SNP & VARIATION SUITE Demonstration

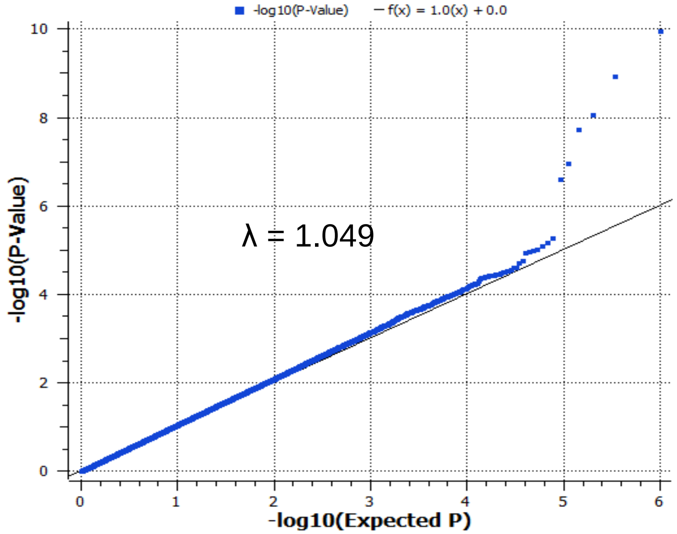
Q-Q Plots



λ = 2.549



λ = 0.979



λ = 1.049

GOLDEN HELIX

SNP & VARIATION SUITE

Demonstration

Overview of Methods

Naïve GWAS

Corr/Trend Test

- Quality Control of Samples and Markers

GWAS + Correcting for Population Stratification

PCA Correction (Eigenstrat Price 2006)

- Direct correction of genotype and phenotype data
- Adding PCs as covariates to regression model

Mixed Model Approach

EMMAX (Kang 2010)

- Using the Genomic Relationship Matrix (IBD) to account for stratification

GOLDEN HELIX

SNP & VARIATION SUITE

Demonstration

Summary

- Performed Basic Association Test
  - Verified contingency table counts
    http://goldenhelix.com/SNP_Variation/scripts/pages/FrequencyTable.html
  - Q-Q Plot to look for inflation of p-values
    http://goldenhelix.com/SNP_Variation/scripts/pages/CalculatePseudoLambda.html

- Examined workflow to determine reasons for inflation of p-values
  - Sample/Marker Statistics
    - Call Rate Histograms
  - Cryptic Relatedness through IBD
    - IBD Heat Maps
  - Population Stratification with PCA
    - PCA Plots (2D & 3D)

- Adjusted Analysis for Population Stratification
  - Using PCA from within Association Testing dialog
  - Using PCs as Covariates in Numeric Regression
  - Using Mixed Linear Model Analysis

Questions or more info:

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