

SOLAR-Eclipse

***a genetic imaging research
tool***

Kochunov Peter, PhD, DABMP

Maryland Psychiatric Research Center, University of
Maryland

And

Texas Biomedical Research Institute

Overview

- Historical perspective
- Downloading and installing SOLAR-Eclipse
- Creating a solar analysis directory
 - pedigree files
 - Marker files
 - phenotype files
- Common analyses types
 - Univariate and bivariate analyses of variance, linkage and association
- Using Solar-Eclipse for Mega-and-Meta analyses
 - Additive Genetic Analyses (Heritability)
 - Analyses of fixed genetic factors (Association)
- Executing solar in parallel environment

SOLAR-Eclipse

- Extension of SOLAR for imaging genetic research
- Developed for multiplatform (pc/mac/linux)
 - Genetic analysis of discrete and continuous traits
 - Supports polygenic, quantitative trait and GWAS analyses in related and unrelated samples
 - Supports uni-and-multivariate analyses
 - Supports discrete and continuous covariates
- Implements functionality of three genetic tools
 - MENDEL, FISHER and SEARCH
- MENDEL and FISHER
 - Developed from 1960 to late 1980s by
 - Kenneth Lange, Michael Boehnke, Daniel Weeks
 - UCLA and University of Michigan, Ann Arbor

MENDEL

- Intended for
 - Gene mapping calculation
 - QTL analyses
 - Pedigree segregation
 - Multipoint/Linkage Quantitative Trait mapping
 - Allele frequency estimation
 - Paternity testing
 - Genetic Counseling for disorders such as
 - Cystic Fibrosis
 - Duchenne Dystrophy and others

FISHER

- Genetic analysis of quantitative traits whose variability explained by
 - polygenic inheritance
 - environmental forces
- Provided functions for additive genetic analysis of continuous polygenic traits
 - Heritability
 - Genetic correlation

MENDEL and FISHER

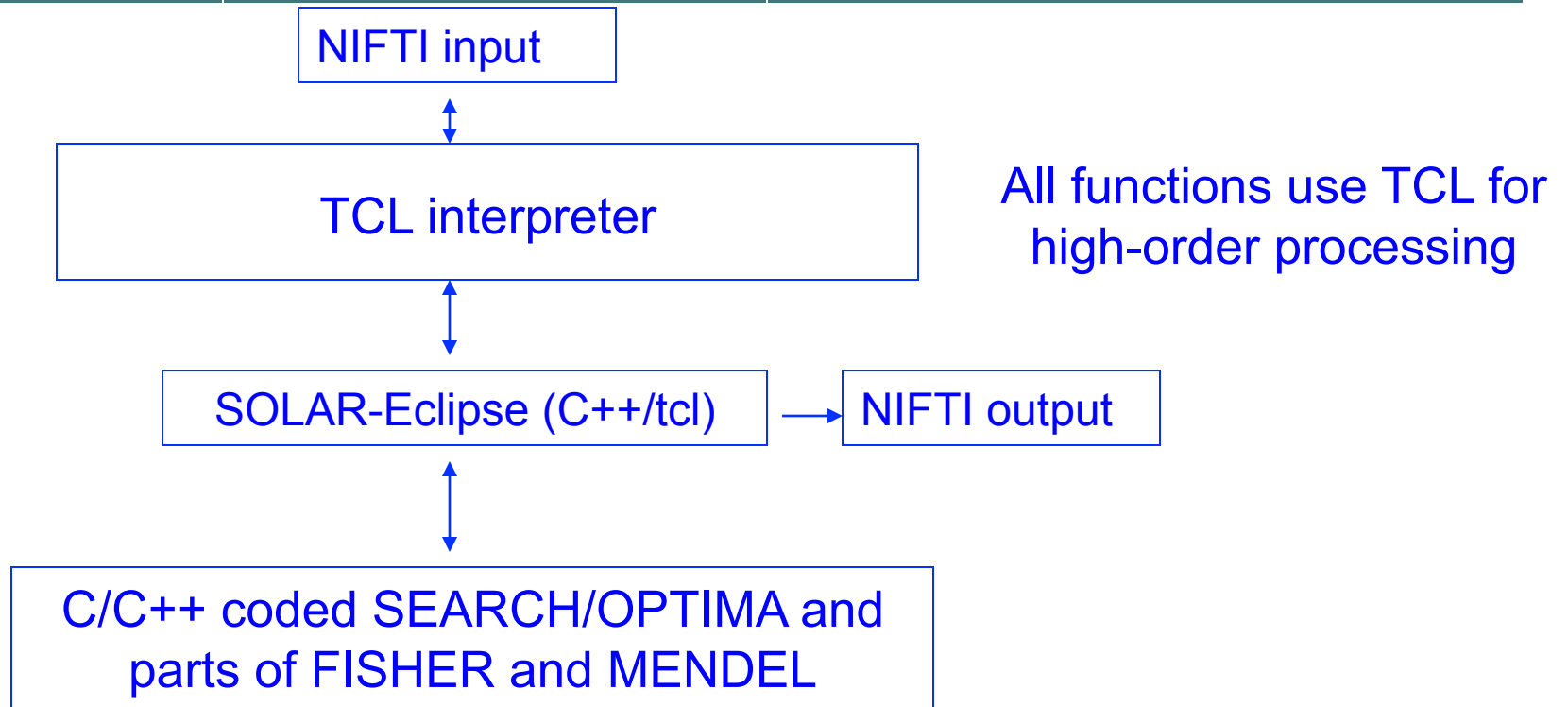
- Use genetic likelihood calculations
 - Variability in trait is considered to be multivariate normal
 - Multivariate and univariate analyses are supported
 - Proband corrections and robust outlier detection is implemented
- Use maximum-likelihood estimation (MLE) called SEARCH

SEARCH algorithm

- The MLE engine for MENDEL/FISHER

- An optimization routine
 - Works within user-set bounds and limits
 - Maximizes a cost function
 - Samples a function over a grid
 - Implemented the least square and use-defined estimators
- John has optimized it for discrete traits “OPTIMA”

What is SOLAR-Eclipse?



Progress since last year.

- Polyclass functionality is now standard
 - Combined analyses of multiple pedigrees
 - Neda's talk will discuss this
- Changes in data handling
 - Inverse normalization can be performed before or after covariate regression
- SOLAR-Eclipse WEB version is available!
- Great speed up is coming up – ready for WGS
 - Single-step estimators of heritability and association
 - GPU acceleration

Video Tutorials on Solar-Eclipse



<http://Mdbrain.org/solareclipse>

Downloading and installation

- Available at
http://www.nitrc.org/projects/se_linux/
- Installation is simple – untar
- Need to get software key/license
 - Email solar@txbiomedgenetics.org
 - One per user
- Website (WIP)
 - <http://www.mdbbrain.org/solareclipse/>
 - Instruction videos, how-to's, example files
 - Main solar manuals are at
 - <http://solar.txbiomedgenetics.org/doc/>

When you type “solar”

- TCL interpreter starts
- When you type a command like “polygenic”
 - TCL interpreter reads solar.tcl file
 - Codes all the popular functions such as polygenic
 - Executes “polygenic” function
 - analyzes the arguments and existing global variables
 - Chooses among several analyses e.g. univariate vs. multivariate
 - Calls upon compiled work for computationally-intensive procedures

What is TCL and its role in Solar

- Tool Command Language
- Similar to shell scripts such as bash and sh
- All solar command are TCL functions
- You can examine/modify any command by
 - `showproc command`
 - `showproc command new_command.tcl`
 - It will create a copy of original TCL script in your directory
 - Open it with a text editor to see what it does

SOLAR working directory

- All projects are organized by “directory”
 - Solar reads the default files from the directory it is started from
 - Including pedigree and default phenotype files
- Directory should contain the pedigree file
 - The kinship matrix
 - Definition of the familial ties among the subjects

Coding the pedigree is Simple®

- Make in excel and save as a csv file

- Mac users should save it in Windows' CSV format

- Pedigree file defines

- Subject ID
- IDs for Father and Mother, including founders
- Sex: 2 and 1 or M and F.
- Family ID
- Monozygotic twin label
- Class – used for mega-genetic analysis

Defining pedigree

- Each “active” subject has to have
 - ID
 - Family id
 - Gender
 - Parent IDs, even if you don’ t use parents
 - Even if the parents aren’ t known
 - A label for being a twin
- Additional variables
 - Class- specifies that subjects belong to a different pedigrees
 - House – specifies household, to study enviromental effects

SOLAR-Eclipse Video Library

Example pedigree for UCLA twins

ID	FA	MO	SEX	FAMID	MZTWIN	CLASS
UCLA8003201	UCLA8003204	UCLA8003203	2	UCLA80032	UCAL8003	1
UCLA8003202	UCLA8003204	UCLA8003203	2	UCLA80032	UCAL8003	1
UCLA8004101	UCLA8004104	UCLA8004103	2	UCLA80041	UCAL8004	1
UCLA8004102	UCLA8004104	UCLA8004103	2	UCLA80041	UCAL8004	1
UCLA8005001	UCLA8005004	UCLA8005003	2	UCLA80050	UCAL8005	1
UCLA8005002	UCLA8005004	UCLA8005003	2	UCLA80050	UCAL8005	1
UCLA8018001	UCLA8018004	UCLA8018003	1	UCLA80180	UCAL8018	1
UCLA8018002	UCLA8018004	UCLA8018003	1	UCLA80180	UCAL8018	1
UCLA8033301	UCLA8033304	UCLA8033303	2	UCLA80333	UCAL8033	1
UCLA8033302	UCLA8033304	UCLA8033303	2	UCLA80333	UCAL8033	1
UCLA8041401	UCLA8041404	UCLA8041403	2	UCLA80414	UCAL8041	1
UCLA8041402	UCLA8041404	UCLA8041403	2	UCLA80414	UCAL8041	1
UCLA8004104			1	UCLA80041		1
UCLA8004103			2	UCLA80041		1
UCLA8003304			1	UCLA80033		1
UCLA8003303			2	UCLA80033		1
UCLA8003204			1	UCLA80032		1
UCLA8003203			2	UCLA80032		1
UCLA2182904			1	UCLA21829		1
UCLA2182903			2	UCLA21829		1

Defined twins

Defined parents

Reading in the csv pedigree file

```
heron : peterk
File Edit View Scrollback Bookmarks Settings Help

[peterk@heron pedigree]$
[peterk@heron pedigree]$
[peterk@heron pedigree]$
[peterk@heron pedigree]$ solar
Starting Solar

SOLAR version 7.0.3 (Experimental), last updated on November 05, 2012
Copyright (c) 1995-2012 Texas Biomedical Research Institute
Enter help for help, exit to exit, doc to browse documentation.

solar> ls
gobs_ucla_pedigree.csv
Checking that the cvs pedigree is there
solar> load pedi gobs_ucla_pedigree.csv
Loading pedigree
Loading pedigree data from the file gobs_ucla_pedigree.csv
solar> ls
gobs_ucla_pedigree.csv house.gz pedigree.info pedindex.cde pedindex.out phi2.gz
A successfully loaded pedi should have all these files created in the directory
solar> █
```

heron : peterk peterk : scp ...rk : dti_atlas_maski

Import of existing pedigrees

- Use Pedsys (<http://pedsys.txbiomedgenetics.org>)
- To import from other packages
 - S.A.G.E.
 - REGC
 - CRI-MAP
 - PAP
 - IBDMAT and others
- More information
 - <http://solar.txbiomedgenetics.org/doc/04.chapter.html>

Creating a SNP file

- SOLAR uses a simple marker format.
 - CSV file with a header
 - ID/EGO column is used to define subject ids
- It would look like
 - ID, rs429358, rs839523, rs1799945
 - PF0132, 1, 2, 1
 - PF0133, 2,2,2
 - PF0134,1,2,2

Reading PLINK genotype files

- Work-in-progress
- Have a converter – contact me or John
- ENIGMA-2 offers SNP extractor scripts that will create solar genotypes from a list of snps.

Creating phenotype files

- SOLAR phenotype files are in CSV format
- Header includes either ID or EGO
 - Identifiers for ID field
- Typical file for a continuous trait is
 - ID, Age, Sex, Gmthickness, FA, FLAIR
 - PH0001, 25, M, 2.5, 0.56, 1.193
 - PH0002, 39, F, 2.34, 0.49, 2.141

Phenotypes to incorporate imaging data

- The CSV header should identify the format of the trait using semicolon
 - GMdensity:NIFTI
 - GMthickness:GIFTI
 - Do we need support for other format?
- File name is provided for each subject
 - Semicolon can be used for volume identification in 4D file

An example of a phenotype file with binary traits

mrid	id	age	sex	AverageFA	FA:NIFTI	
EH0109	A32067	52	M	3420	gobs_fa.nii.gz:1	
EH0110	A32195	61	F	4318	gobs_fa.nii.gz:2	
EH0111	A01685	51	F	4331	gobs_fa.nii.gz:3	
EH0112	A11124	39	M	4174	gobs_fa.nii.gz:4	
EH0116	A32020	62	M	4139	gobs_fa.nii.gz:5	
EH0117	A32021	61	F	4386	gobs_fa.nii.gz:6	
EH0118	A32098	37	F	4354	gobs_fa.nii.gz:7	

Performing processing on individual voxels

- Specific voxel(s) need to be defined before processing
 - Individual voxels can be defined as
-

- voxel X:Y:Z
- solar> help voxel

Purpose: To set and save current voxel position

Usage: voxel [<voxel-value>]

<voxel-value> is 3 coordinates delimited by colons as x:y:z
for example, 12:8:23

If no voxel-value is specified, the current voxel is returned.

If no current voxel has been defined, an error is raised.

If a voxel has been defined, it is written to model files.

The current voxel can also be set with the mask
command, and
that is the general way it should be done.

Performing processing on set of voxels

- Mask command defines the set of voxels of the same intensity
- solar> help mask

Purpose: To read image mask file and set current voxel

Usage: mask [<filename>] [-intensity <intensity>] [-index <index>]

mask -next

mask -delete

<filename> is the name of the file containing the mask

<intensity> is the integer value that defines this mask

<index> is position within the set of mask-defined voxels

-next specifies to advance to the next mask-defined voxel

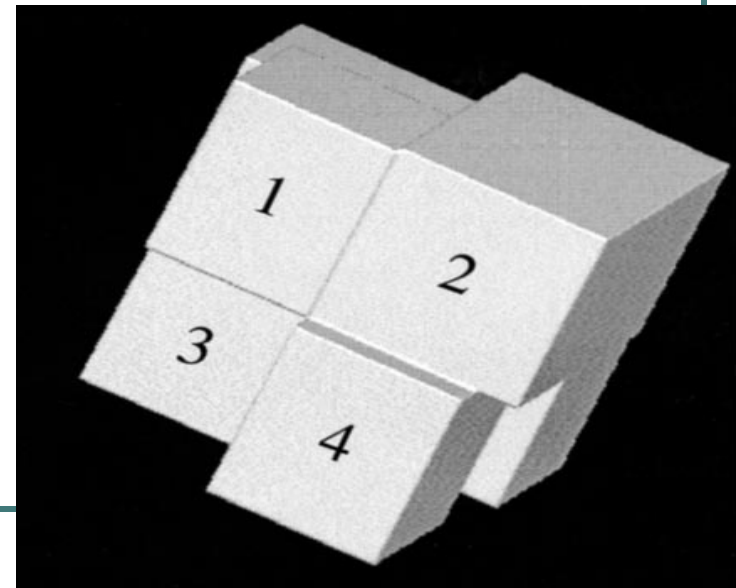
-delete deletes the mask and frees all related storage

Defining output (WIP)

- Imaging trait analysis can be saved into nifti multi-volume file
- **outputvolume [<filename>]**
 - Copies header information from the trait file
 - Processing of each voxel is stored in the corresponding coordinate space
 - [0] – heritability
 - [1] – standard error
 - [2] – probability
 - [3] - % variance explained by covariates
 - Similar format will be used for the genetic correlation

Executing it in a parallel environment

- Masking for simplified sge-execution
- Use a mask to parcelate space
- Save outputs in different files
- Submit them as separate jobs
- Add files for the final result!



Mega-and-Meta genetic analyses

- SOLAR engine is flexible and powerful

- Works with pedigrees of random length and complexity
 - Families
 - Twins
 - Unrelated
- Separate pedigrees can be combined at raw-data state
 - Merge small populations into a synthetic pedigree
 - Perform genetic analysis as if working in large extended families.

Advantages Mega vs. Meta genetic analysis

- Advantages of Mega-genetic analysis:
 - More “powerful” than meta-analysis
 - Test of significance is performed on higher degree of freedoms
 - Smaller samples can be combined into a single set
 - Less weighted by discrepancies in subject numbers and less affected by variance in standard deviation among cohorts
 - Can be used for testing heterogeneity of genetic effects
- Advantages of Meta-genetic analysis:
 - Easier to perform and is readily accepted
 - “Synthetic” heritability and fixed effects may not be well defined due to population differences

How does one proceed?

- Pedigrees are merged with “class” field
 - Class defines all subjects belonging to the same group
 - Sporadic model is fit separately for each “class”
 - Individual datasets are inverse-Gaussian normalized
 - Combine data by ensuring equal distribution and variance of the trait
 - Several polygenic tests are performed
 - Test of significance for each pedigree
 - Cross-wise test of difference in genetic effects for each pedigree
 - Test of significance for a combined pedigree

Polyclass modeling function

- Polyclass function will model additive and fixed genetic effects across pedigree
 - It performed per-pedigree modeling of genetic effect
 - Useful for meta-analysis and analysis of site-specific effects
 - Performs combined modeling of genetic effect
 - Statistical power of discovery is greatly improved
 - Genetic effects are “generalized” across cohorts

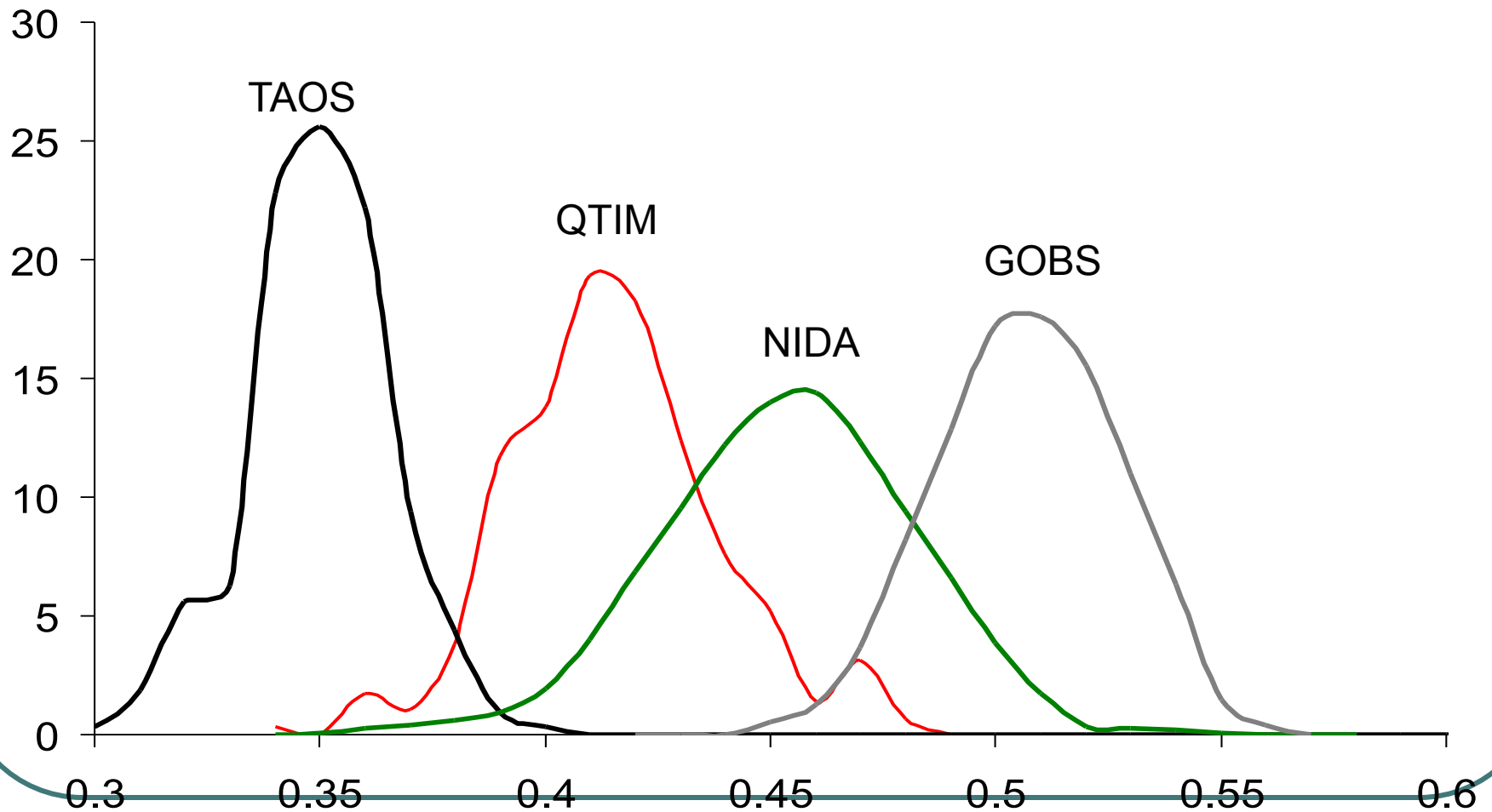
Example: mega-association analysis

- To perform association analysis across more than one cohort (pedigree)
- This example uses 4 cohorts
 - GOBS (N=800) <- families
 - QTIM (N=600) <- twins
 - NIDA-IRP (N=400) <-unrelated
 - TOAS (N=300) <- twins and siblings.

Mega-genetic analysis is simple

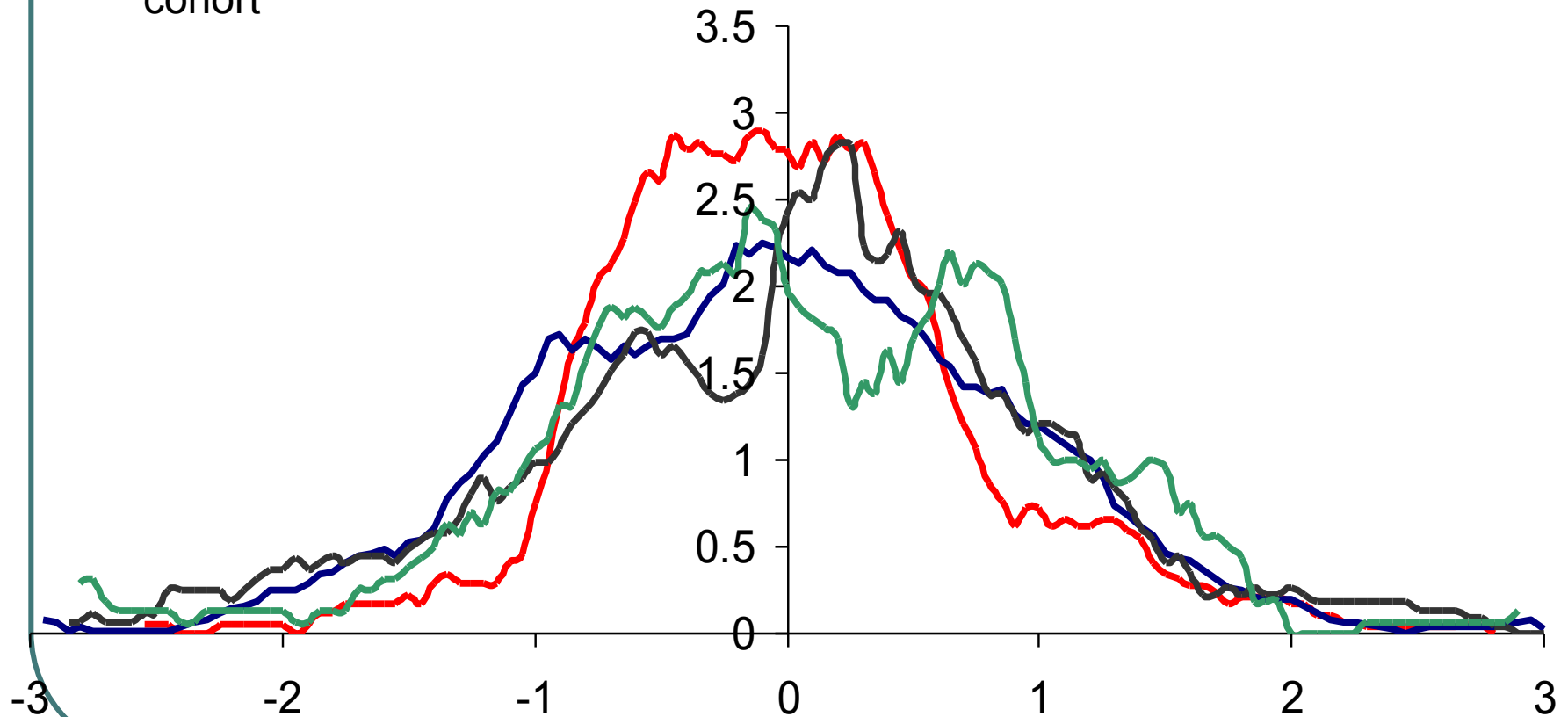
- polyclass 0-3 –maxsnp rs6675281 –
intrait
 - This will model association between the trait and rs6675281
 - separately in classes 0, 1, 2, 3
 - -intrait option will use inverse Gaussian for trait normalization

The analysis: FA values (raw)

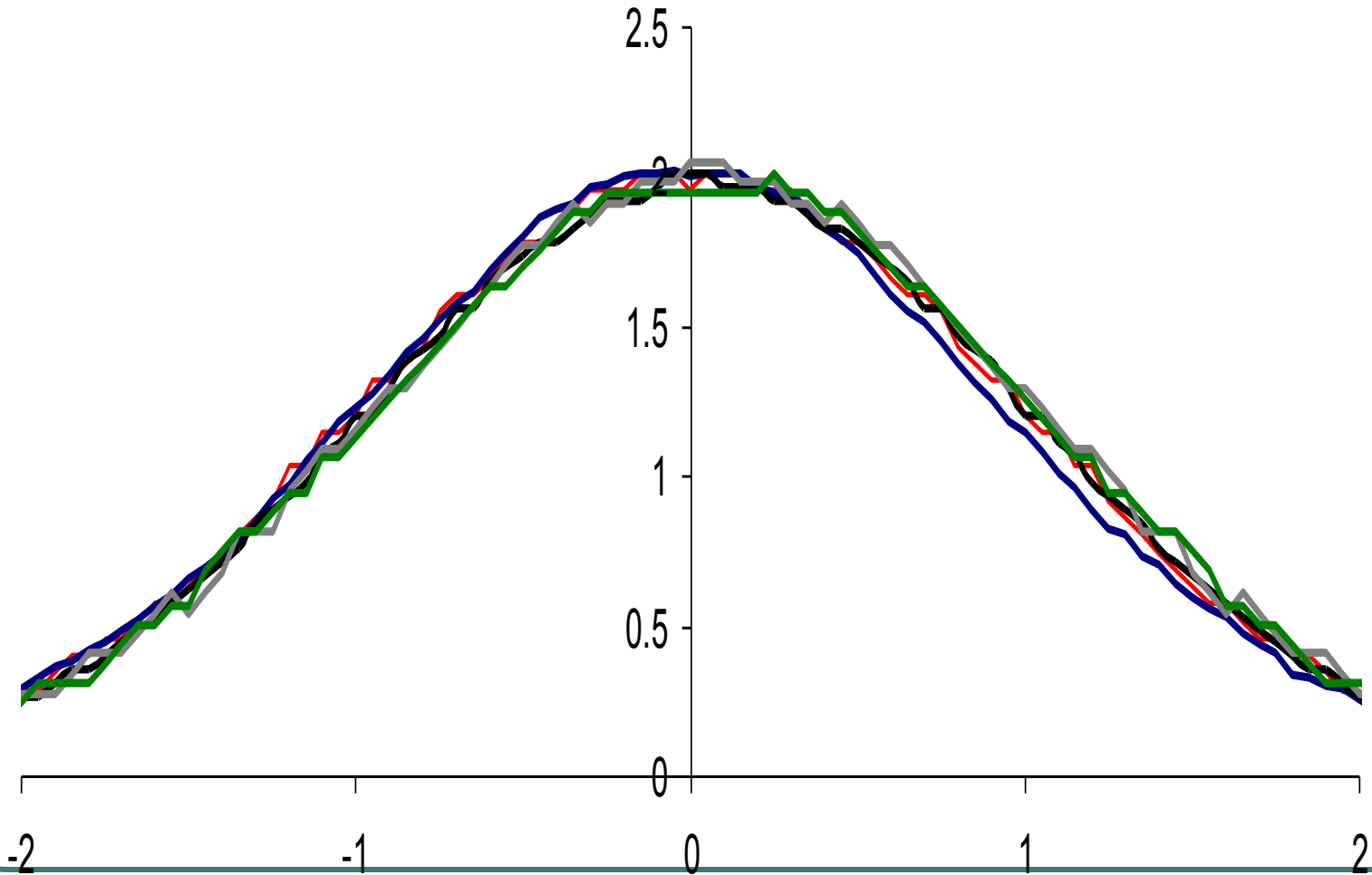


Sporadic model is fit per cohort

Regression of age, gender and ethnicity effects separately per cohort



Inverse normalization performed per cohort.

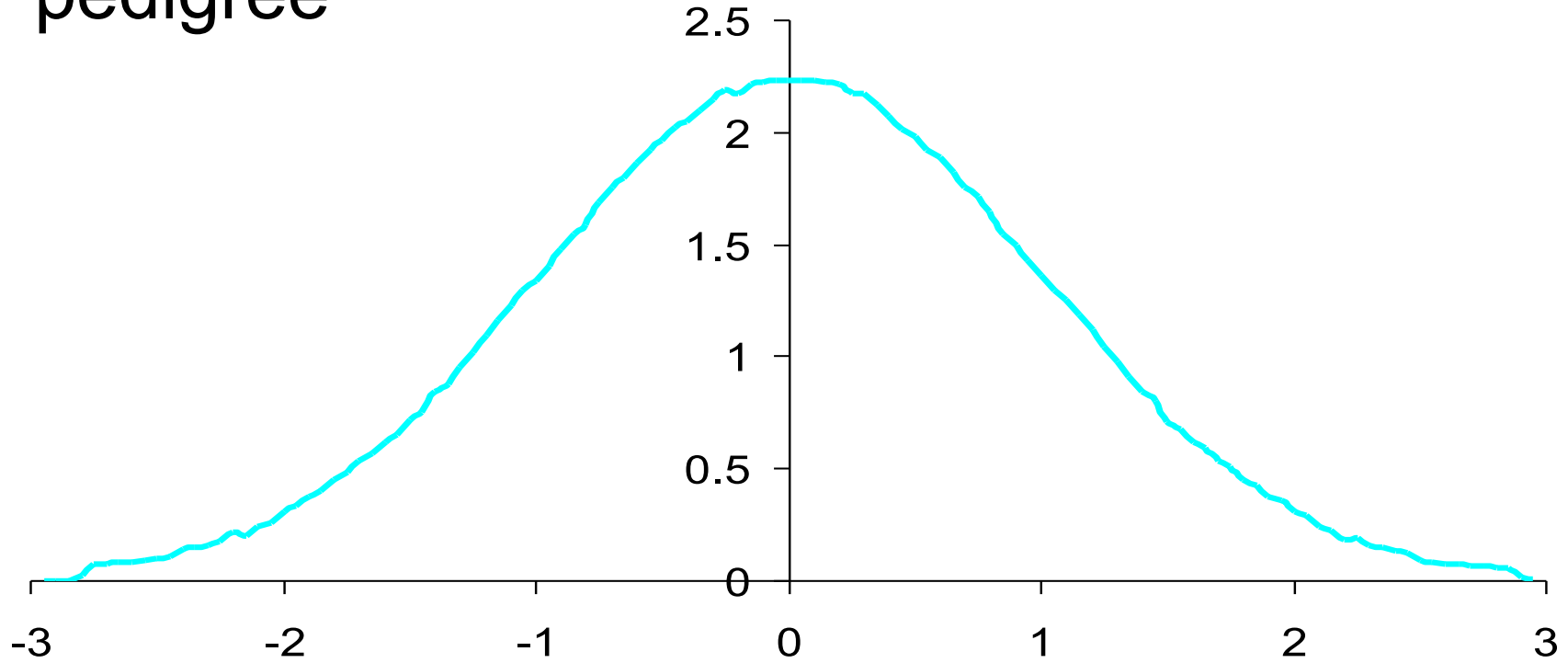


Mega-genetic GWAS is simple

- polyclass 0-3 –maxsnp rs6675281 –
intrait –comb
 - This will model association between the trait and rs6675281
 - separately in classes 0, 1, 2, 3
 - -intrait option will use inverse Gaussian for trait normalization
 - –comb will provide for mega-analysis

All three preceding steps are performed

- Populations are combined into a super pedigree



SOLAR-Eclipse web-version

- Developed with HPC team
- Large projects can put SOLAR-Eclipse front end for their sample
- Users can run the full range of analyses w/o having access to raw data
- Provide immediate access for discovery/replication analyses

CONNECTOMEdb

DashboardPublished Subject Keys

Search

⚠️ Plugin Required!

A browser check indicates that you do not have the Aspera Connect plug Project is using an Aspera server to dramatically boost data transfer speed using any of ConnectomeDB's services, including downloading. ([Need help](#))

Trait FABody

Covariates age sex

Inverse normalization? ☒

Submit

Logged in as: [kochur](#)

* Summary of Results *

Pedigree: hcp_pedi.csv
Phenotypes: combined_data.csv
Trait: FABodyINORM

Individuals: 214

H2r is 0.8458364 p = 1.5415431e-10 (Significant)
H2r Std. Error: 0.0522210

age p = 0.9970146 (Not Sig., but fixed)
sex p = 0.0007807 (Significant)

Proportion of Variance Due to All Final Covariates Is
0.0623123

Output files and models are in directory /home/xnat/data/user/resources/SolarEcl:
Summary results are in /home/xnat/data/user/resources/SolarEclipse-8fa0dc8bea32/
Loglikelihoods and chi's are in
/home/xnat/data/user/resources/SolarEclipse-8fa0dc8bea32/polygenic.logs.out
Best model is named poly and null0 (currently loaded)
Final models are named poly, spor, nocovar
Constrained covariate models are named no

Residual Kurtosis is -0.4497, within normal range

Great Speed Up

- Speed up by 1,000,000 fold
- Allow for “real time” gwas in related samples
- Allow for WGS analysis
 - And analysis of genotypes imputed to WGS
- Make use of GPU technology
 - Greatly reduce processing costs <\$1 per core

Great speed – Single step approximators

- Work in collaboration with Tom Nichols
- Based on the work by John Blangero
- Use EVD transform for single-step approximation of the heritability GWAS
- Allows for permutation inference
- About 1,000 faster than MLE

Great Speed – GPU processing

- Collaboration with Da Mota and Froulin
- GPUs allow for massively parallel computations
 - Voxel-wise heritability
 - GWAS and WGS analyses
- GPU cards offer ~2500 cores
 - Large on-board ram up to 12GB
- GWAS in 1,000 HCP subjects and 10^6 genotypes
 - About 5 seconds.

Downloading solar-eclipse

- We are always in need of few brave souls/testers
 - http://www.nitrc.org/projects/se_linux/
 - Stability should improve in the next six month
- Best to contact me to discuss the project
 - pkochunov@gmail.com

Next year's workshop

- Discussion of new features
 - Single step approximators will be ready
 - Permutation based inference should be ready
 - The GPU version should be in beta
- Empirical Kinship Kernel analyses
 - Assembling kinship matrix from GWAS/WGS
 - Linkage based on empirical kinship?

Acknowledgements

- Collaborative effort of
 - John Blangero
 - Charles Peters
 - David Glahn
 - Tom Nichols
 - Bennett Landman
 - Neda Jahanshad
 - Paul Thompson
- Supported by EB015611, P01HL045522, R37MH059490, R01MH078111, R01MH0708143 and R01MH083824