Epistatic Interactions for Brain eGWAS in Alzheimer's Disease

Mariet Allen and Curtis Younkin, Mayo Clinic in Florida

Introduction

Complex genetic diseases, such as, Alzheimer's Disease (AD) are likely influenced by multiple genetic and environmental factors.

We hypothesize that some of these factors influence risk for disease through effects on gene expression.

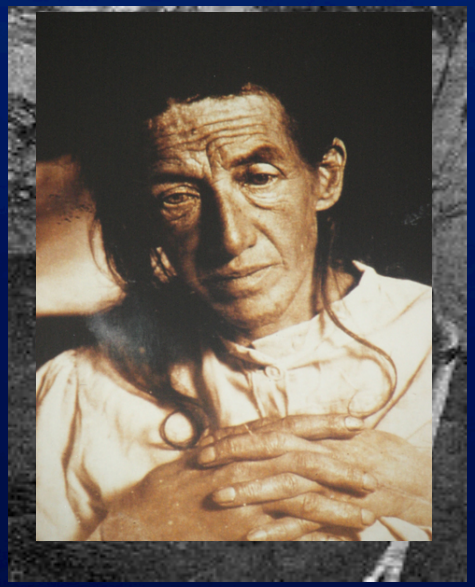
The goal of the current study is to identify pairs of genetic variants that influence brain gene expression levels: epistasis.

This is the first study to analyze brain gene expression data using an epistasis approach.

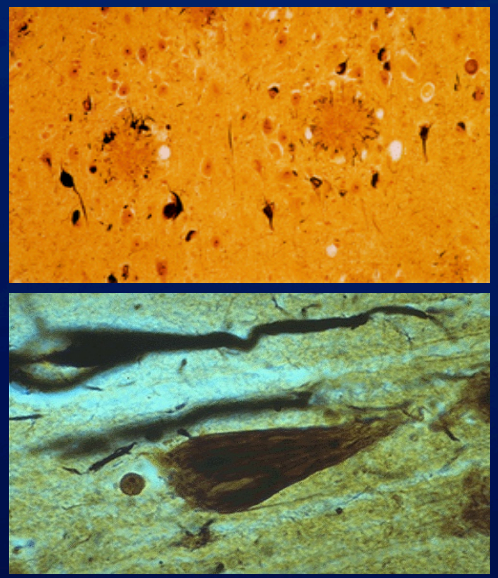
MAYO CLINIC

Background – Why this matters.

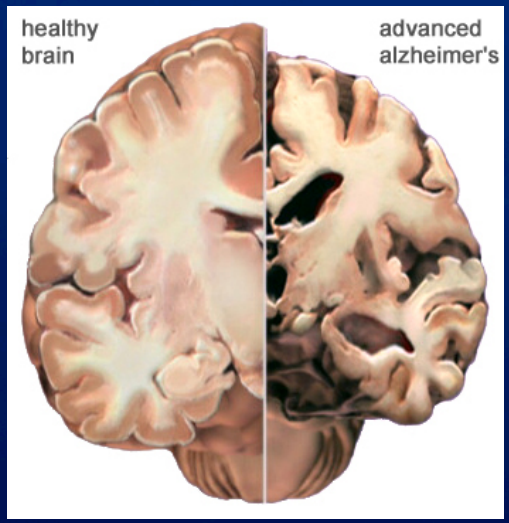
Alzheimer's Disease is the most common type of dementia with a specific neuropathology.



Memory, language and other cognitive problems NOT normal for age.



Abnormal accumulation of extracellular amyloid β (senile plaques) and intracellular tau (neurofibrillary tangles).



Atrophy of brain tissue

MAYO CLINIC

A. Alzheimer, Allgemeine Zeitschrift fur Psychiatrie und Psychisch-gerichtliche Medizin. 1907; www.alz.org; Adams and Victor's Neurology.

Incidence

Alzheimer's Disease is a Deadly Epidemic

5.2 million Americans have Alzheimer's disease (AD).

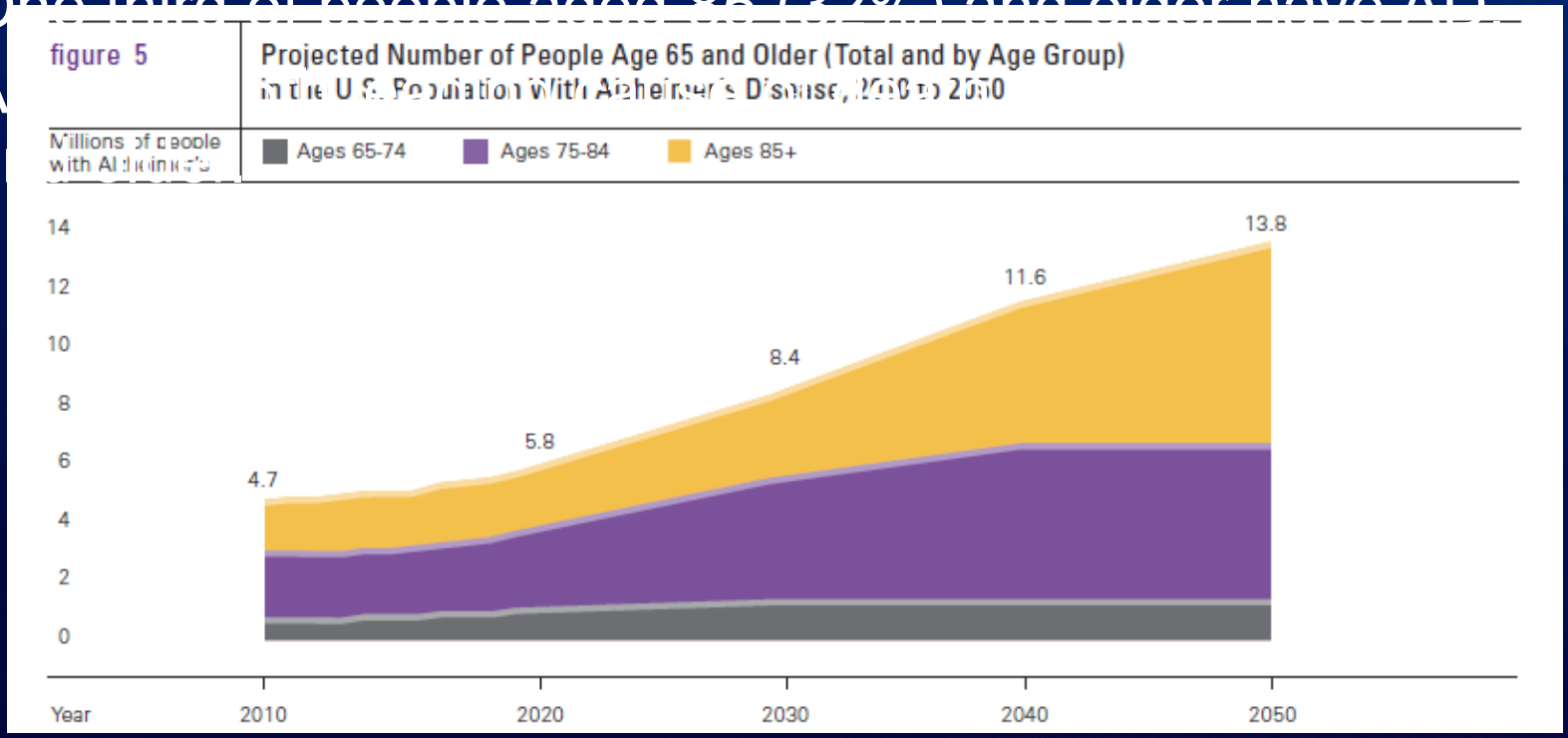
1 in 9 people aged 65 and older have AD (11%).

One third of people aged 85 (32%) and older have AD.

A

a

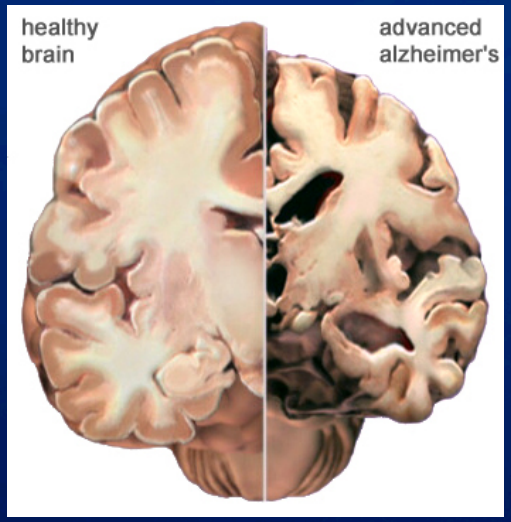
figure 5 Projected Number of People Age 65 and Older (Total and by Age Group) in the U.S. Population With Alzheimer's Disease, 2010 to 2050



2014 Alzheimer Disease Facts and Figures

Treatments

- Pharmacological:
  - Five FDA approved drugs are available that can provide temporary relief of the symptoms in some patients.
  - None of the current therapies for AD slows or stops the disease process.
- Non Pharmacological
  - Aimed at improving quality of life by managing symptoms.
  - Include, physical therapy, reminiscence therapy, cognitive stimulation.
  - Do not cure or prevent the disease.



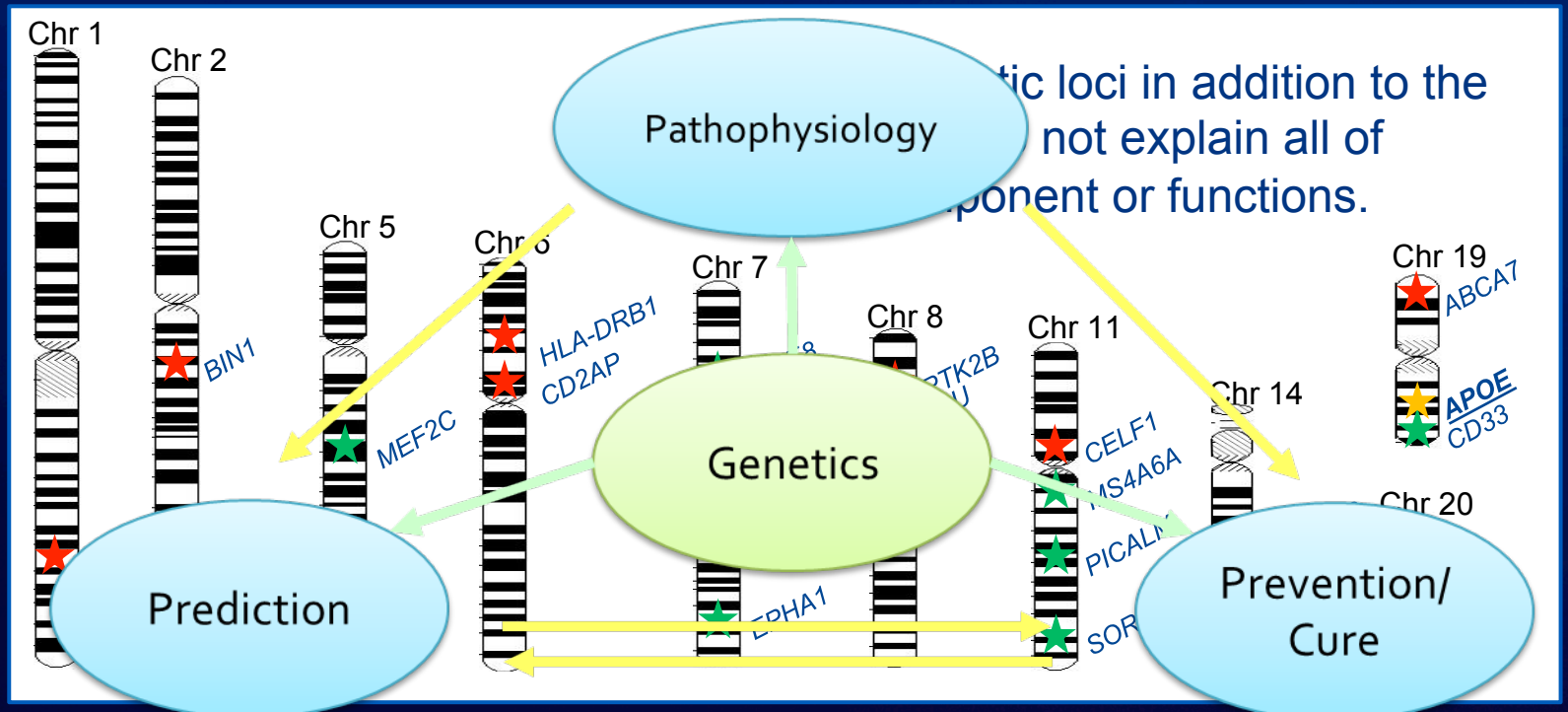
MAYO CLINIC

2014 Alzheimer Disease Facts and Figures

Genetics

Mutations in 3 genes: *APP, PSEN1* and *PSEN2* known to cause rare early onset familial AD.

Up to 80% of risk for Late-Onset AD (LOAD) is predicted to be accounted for by genetics.



Approach

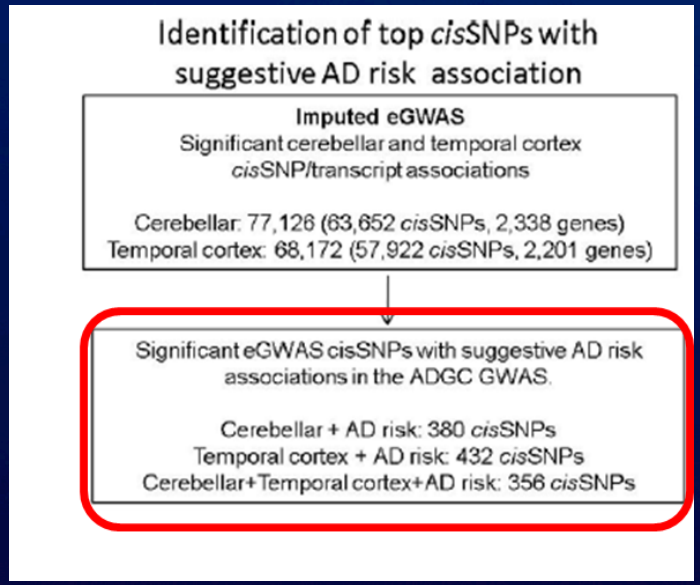
Can leverage endophenotypes, such as gene expression levels, to

- identify additional genetic factors.
- determine mechanism of action.

OPEN ACCESS Freely available online PLOS GENETICS

Brain Expression Genome-Wide Association Study (eGWAS) Identifies Human Disease-Associated Variants

Fanggeng Zou[1☯], High Seng Chai[2☯], Curtis S. Younkin[1☯], Mariet Allen[1☯], Julia Crook[3], V. Shane Pankratz[2], Minerva M. Carrasquillo[1], Christopher N. Rowley[1], Asha A. Nair[2], Sumit Middha[2], Sooraj Maharjan[2], Thuy Nguyen[1], Li Ma[1], Kimberly G. Malphrus[1], Ryan Palusak[1], Sarah Lincoln[1], Gina Bisceglio[1], Constantin Georgescu[1], Naomi Kouri[1], Christopher P. Kolbert[4], Jin Jen[4], Jonathan L. Haines[5], Richard Mayeux[6], Margaret A. Pericak-Vance[7], Lindsay A. Farrer[8], Gerard D. Schellenberg[9], Alzheimer's Disease Genetics Consortium[¶], Ronald C. Petersen[10], Neill R. Graff-Radford[11], Dennis W. Dickson[1], Steven G. Younkin[1], Nilüfer Ertekin-Taner[1,11]*



Neurology®

Novel late-onset Alzheimer disease loci variants associate with brain gene expression

Mariet Allen, Fanggeng Zou, High Seng Chai, et al.

Neurology 2012;79;221-228 Published Online before print June 20, 2012
DOI 10.1212/WNL.0b013e3182605801

MAYO CLINIC

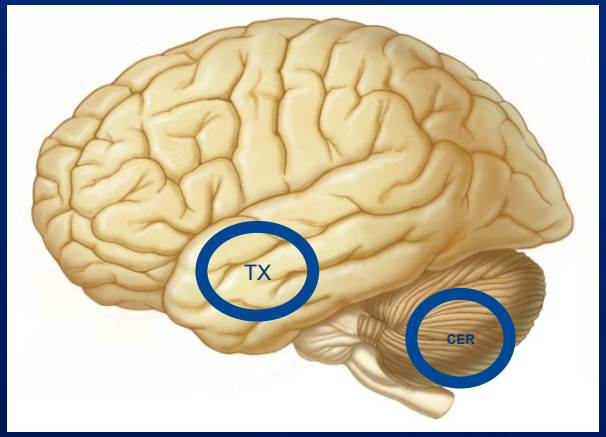
Approach: Samples

	Temporal Cortex		Cerebellum	
RNA and DNA	202 AD	197 Non-AD	197 AD	177 Non-AD
Females (%)	108 (53%)	78 (40%)	101 (51%)	63 (36%)
Age: Mean (SD)	73.6 (5.5)	71.6 (5.6)	73.6 (5.6)	71.7 (5.5)
RIN: Mean (SD)	6.3 (0.9)	6.9 (1.0)	7.2 (1.0)	7.2 (1.0)

Group 1 Group 2 Group 4 Group 5

Group 3 Group6

Replication



Frozen Brain Tissue from the Mayo Clinic Brain Bank (Dr. Dennis Dickson)

MAYO CLINIC

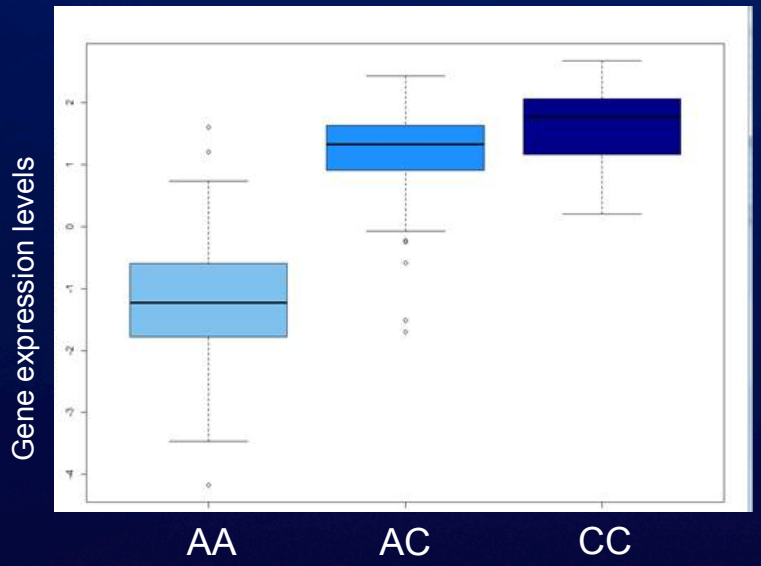
Approach – Genotypes and Phenotypes.

- Genotypes for >300,000 common SNPs distributed throughout the genome. For each subject 1 of 3 possible genotypes at each SNP is obtained: Carrasquillo et al, Nature Genetics 2009.

SNP A

AA	AG	GG

- Phenotypes are gene expression levels (mRNA) measured using Illumina Whole-Genome DASL: 24,526 probes (18,401 genes). Zou et al, PLoS Genetics 2012.

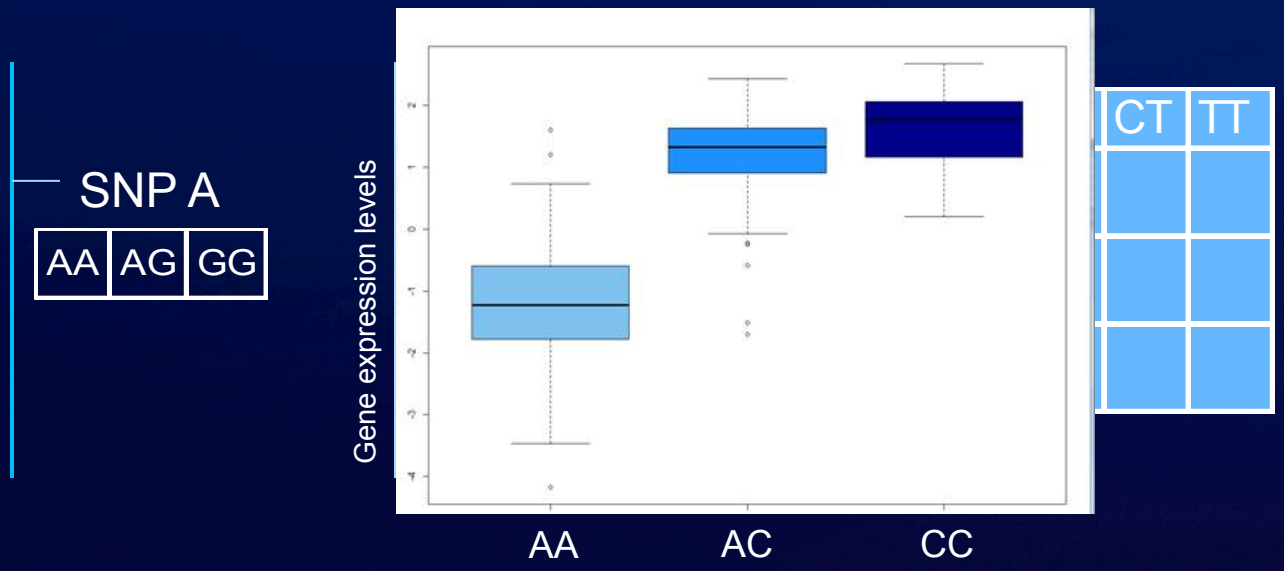


MAYO CLINIC

Why epistasis?

Single SNP/single phenotype approach is simplistic and cannot fully explain the known heritability of various diseases and phenotypes studied.

Epistasis allows for the study of interaction effects of pairs of SNPs on a given phenotype and can uncover *additional genetic factors* that influence gene expression and disease.



Key Challenges

Computation resources: Increase quadratically with the number of SNP interactions being considered. How do we to compute analysis of ~300,000 x ~300,000 epistatic interactions for 24,000 phenotypes?

Accounting for covariates: statistical applications that facilitate analysis of epistatic interactions do not allow for incorporation of covariates in regression analysis.

Storage: epistasis analysis of the scale described here generates large amounts of data which must be stored and organized.

MAYO CLINIC

Why BlueWaters?

Computation Resources:

Home Clusters	BlueWaters
Software: PLINK	**Software**: Fast-Epistasis
Phenotypes: 1 at a time	**Phenotypes**: 32 at a time
Estimated time: 75 hours/phenotype	**Actual Time**: < 2 days for all phenotypes

Storage:

- Simultaneous FastEpistasis computation on increasing number of phenotypes quickly saturates the aggregate disk I/O on standard academic clusters.
- Intermediate files generated quickly add up to hundreds of terabytes per analysis.
- Easily handled by Blue Waters' petabyte storage facility.

MAYO CLINIC

Data preparation workflow

1. GWAS genotypes: 318,257 SNPs

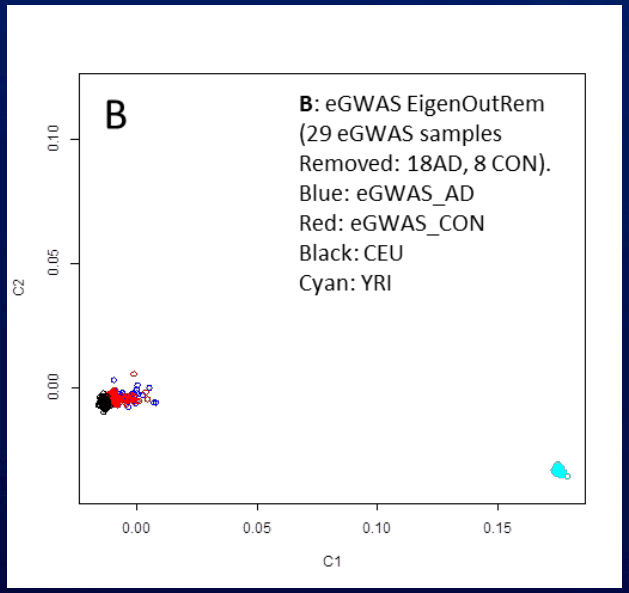
2. GWAS Post QC: 304,624 SNPs

3. Diploid, Common, High Quality: 273,848 SNPs

4. LD Prune: 223,632 SNPs

5. Remove population outliers.

6. Generate Residuals for Phenotypes (Covariates)



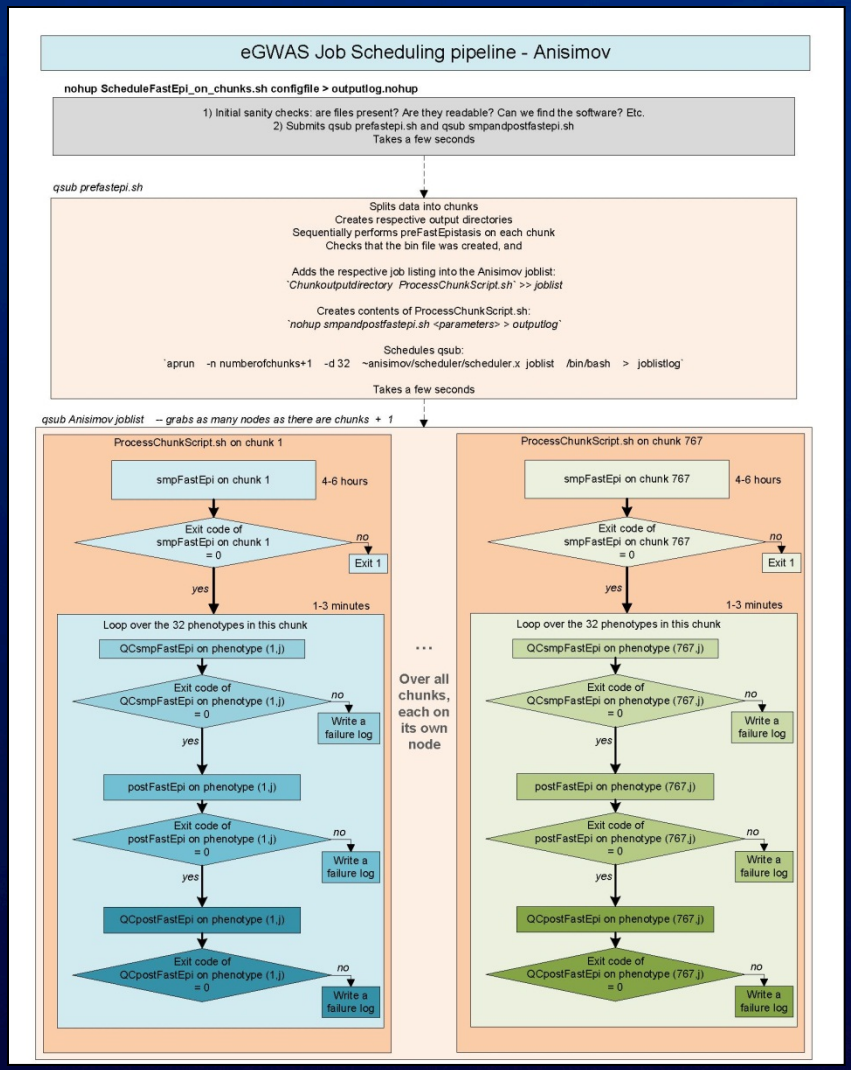
MAYO CLINIC

Accomplishments

NCSA senior scientist Liudmila Mainzer

- determined that Fast-Epistasis runs most optimally with 32 phenotypes at a time.
- designed code to launch Fast-Epistasis on BlueWaters.

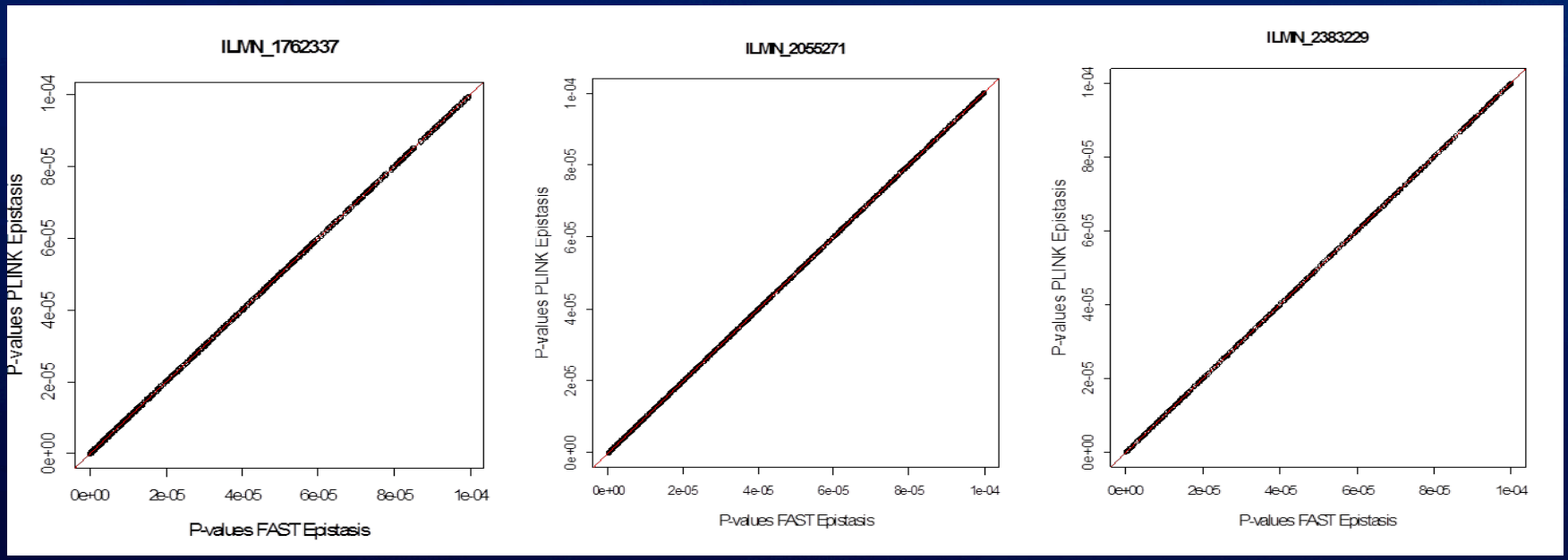
Fast Epistasis author Thierry Schuepbach, is collaborating with us to make further improvements to the application.



Accomplishments

PLINK and Fast-Epistasis give the same results

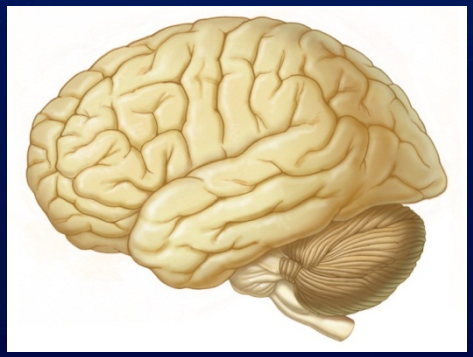
Have successfully completed the first of 6 analysis – Temporal Cortex AD samples



MAYO CLINIC

Future Directions

- Complete analysis of 2 additional groups of data Non-AD and AD+Non-AD.
- Completion of analysis of same 3 groups using expression data from Cerebellum.
- Filtering of results by counts for genotypes.
- Analysis of many hundreds of additional samples using gene expression data collected using RNAseq.



MAYO CLINIC

Acknowledgements

Patients and families who donated their brains for research.

Mayo Clinic
Nilufer Ertekin-Taner
Curtis Younkin
Minerva Carrasquillo
Fanggeng Zou
Julia Crook
Steven G Younkin
Travis Drucker

UIUC and NCSA
Luda Mainzer
Victor Jongeneel
Gloria Rendon
Jenny Drnevich Zadeh

Swiss Institute Bioinformatics
Thierry Schuepbach

Research supported by: R01 AG032990, R01 NS080820, P50 AG016574, U01 AG046139, PSP support for data analysis on Bluewaters.

MAYO CLINIC