

ESTIMATION AND CHARACTERIZATION OF THE SNP-HERITABILITY OF ALCOHOL DEPENDENCE IN SUBJECTS OF EUROPEAN AND AFRICAN ANCESTRY

RHC Palmer, PhD

LA Brick, JE McGeary, VS Knopik, MC Keller

RIH | Division for Behavioral Genetics

Brown University | Department of Psychiatry and Human Behavior



Rhode Island Hospital
A Lifespan Partner



BROWN

Outline of Presentation

- Consequences & mechanisms of alcohol harm
- Approach for genomewide comparison of effects across ancestral populations.
- Summary of findings



Alcohol Misuse – A Global Issue


- Regular alcohol consumption is a risk factor for increased mortality.
 - In 2012, 3.3M deaths were attributable to alcohol consumption.
 - It is the 4th leading cause of death in USA.



Mechanisms of Harm

- There are three direct mechanisms of alcohol harm:
 1. **Toxic effects** on organs and tissues;
 2. **Intoxication**, leading to impairment of physical coordination, consciousness, cognition, perception, affect or behavior;
 3. **Dependence**, whereby the drinker's self-control over his or her drinking behavior is impaired.

Definition of Alcohol Dependence

- A maladaptive pattern of alcohol abuse leading to clinically significant impairment or distress as described by these seven symptoms: 
- DSM-IV criteria
- 3+ (within 12 months)
 - Tolerance
 - Withdrawal
 - Drinking longer than intended
 - Failure to quit drinking
 - Much time spent using/recovering from alcohol
 - Social/occupational activities foregone
 - Drinking despite physiological/psychological problems

Genetics of Alcohol Dependence

- Alcohol dependence runs in families.
- Genetic differences between individuals account for ~50% of the risk for alcohol dependence.
 - Genetic difference can increase or decrease a person's risk.
 - No such thing as an "alcoholism gene".
- Familial, psychological, and sociocultural factors are also very important.

Causes of Disease

Monogenic diseases

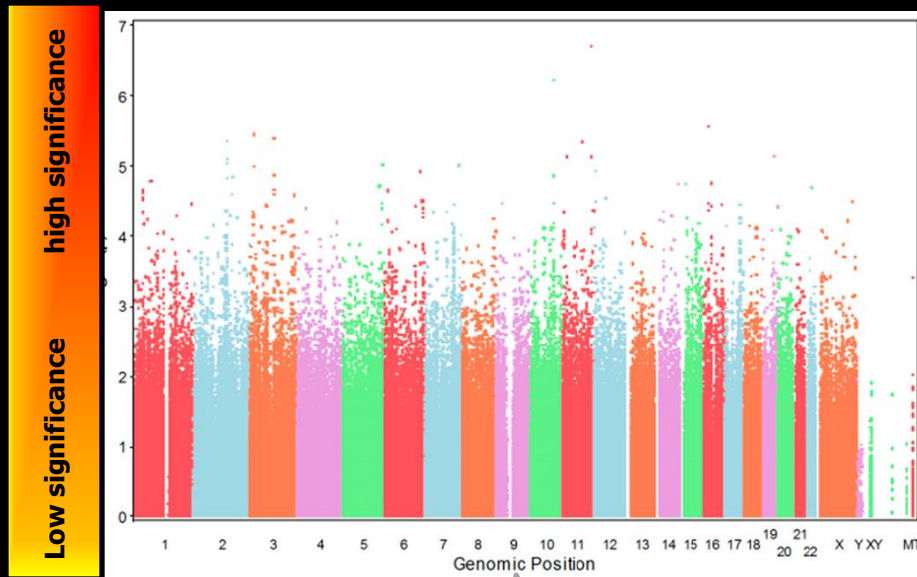
- Strongly influenced by variation in a single gene.
- Classic patterns of inheritance within families.
 - Inheritance conforms to Mendelian principles.
 - Occurrence is rare.
- Genetic variants typically have large effects, altering/reducing function or stability of proteins(s).
- E.g., PKU or HD

Complex diseases

- Strongly influenced by variation within multiple genes; can be caused variation in a gene.
- Do not have predictable patterns of inheritance
- Spectrum of genetic effects is broad impacting proteins directly and indirectly; effect sizes are small.

Alcoholism Is A Complex Disease

GWAS suggests many variants of small effect.



Heath et al., 2011

Variants across genome

Many genes* linked to AD

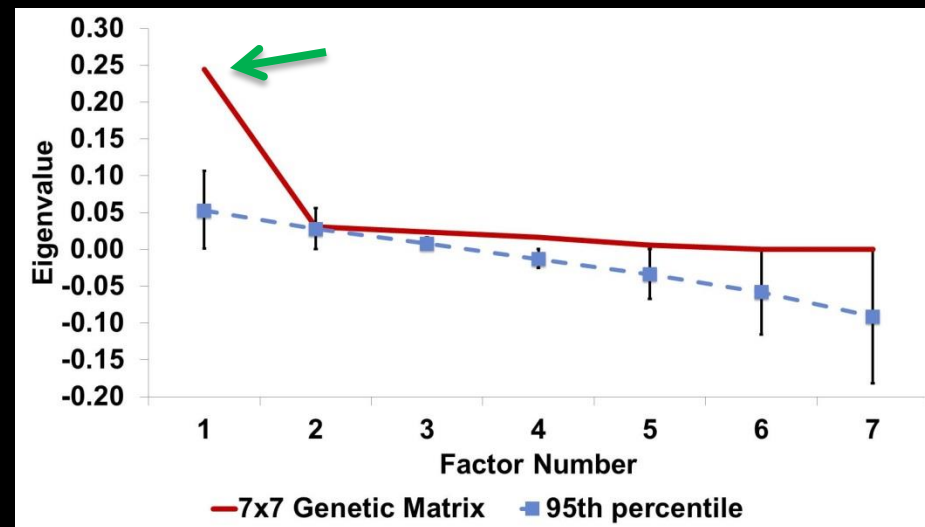
- *ALDH2* (aldehyde dehydrogenase), *ADH1B*, *ADH1C*, *ADH4*,
- *CHRM2*, *nACHRs* *A3A5B4*
- *OPRK1*, *OPRM1* (opioid), *PDYN*
- *5-HTTLPR*
- *NMDAR1*, *NMDAR2B*
- *GABA-A: α2, β1, β3, γ3*
- *GABA-B*
- *MAO-A*, *MAO-C*, *DBH*, *COMT*
- *DAT* (*SLC6A3*), *DRD2*, *DRD4*
- *GRIK1* (glutamate)

* - listed genes have both positive and negative association findings, and should be carefully interpreted.

Common Variants Influence Alcoholism

SNP heritability (h^2_{SNP} ; s.e.) of AD	
Phenotype	Current Study h^2_{SNP} (s.e.)
AD Diagnosis	0.300 (0.136) ^a
AD factor score	0.307 (0.130) ^a
DSM-IV AD Symptoms	
Sx 1: Tolerance	0.242 (0.129) ^a
Sx 2: Withdrawal	0.281 (0.174)
Sx 3: Using longer than intended	0.324 (0.158) ^a
Sx 4: Failure to quit	0.197 (0.146)
Sx 5: Great time spent using/recovering	0.072 (0.104)
Sx 6: Activities foregone	0.199 (0.091) ^a
Sx 7: Continued use despite problems	0.237 (0.109) ^a

Analysis of the genetic covariance of AD symptoms suggests a single factor.



Inability to localize important variants

Possible reasons include:

- 1. Studies are underpowered to detect small effects.**
- 2. Clinical phenotypes lack disease sensitivity.**
- 3. Failure to fit model using all SNPs simultaneously.**
 - a) Provides less biased SNP-effects
- 4. Studies are biased toward a singular population.**
 - a) Heterogeneity in allele frequency across ancestral groups affects power for different markers.
 - b) More than 90% of research into genetic causes of alcohol dependence focus on people of European descent.

Recent GWAS of AD in EAs & AAs

Gelernter et al., 2014

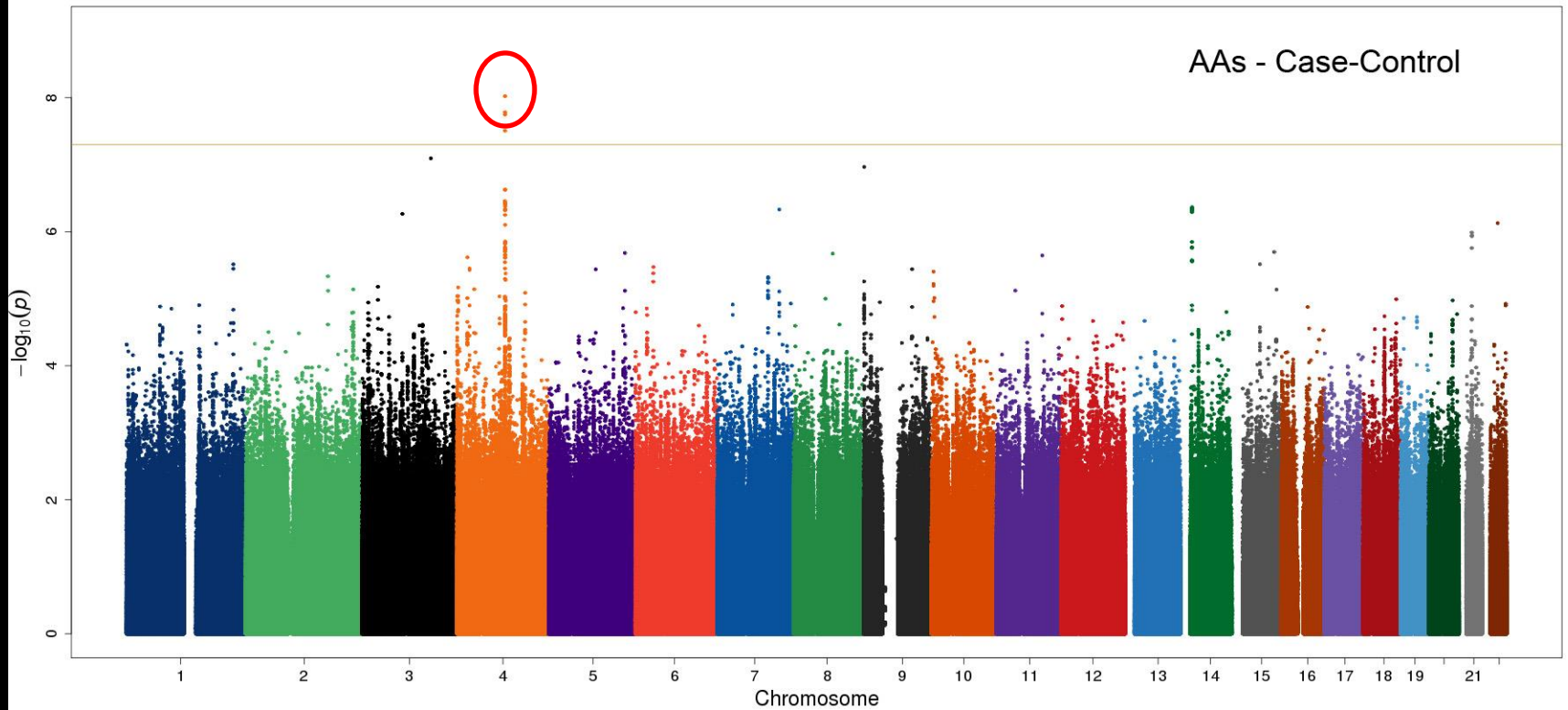
- GWAS of AD in people of European (EA) and African ancestry (AA).
- Large sample (n=16,087)
 - GWAS Discovery (9,758)
 - Multiple replication data
- 5×10^{-8} significance threshold

Findings

- Novel SNPs were found
 - Chromosome 4 ADH gene cluster
 - PDLIM5 (PDZ and LIM Domain 5)
 - METAP1 (methionyl aminopeptidase 1)
 - LOC100507053 (a lncRNA gene)
 - ADH1B and ADH1C
 - Chromosomes 2, 5, 9, 19
- Evidence for biological convergence as similar gene loci were observed across EA & AA.
- Most significant SNPs were replicated in independent samples.

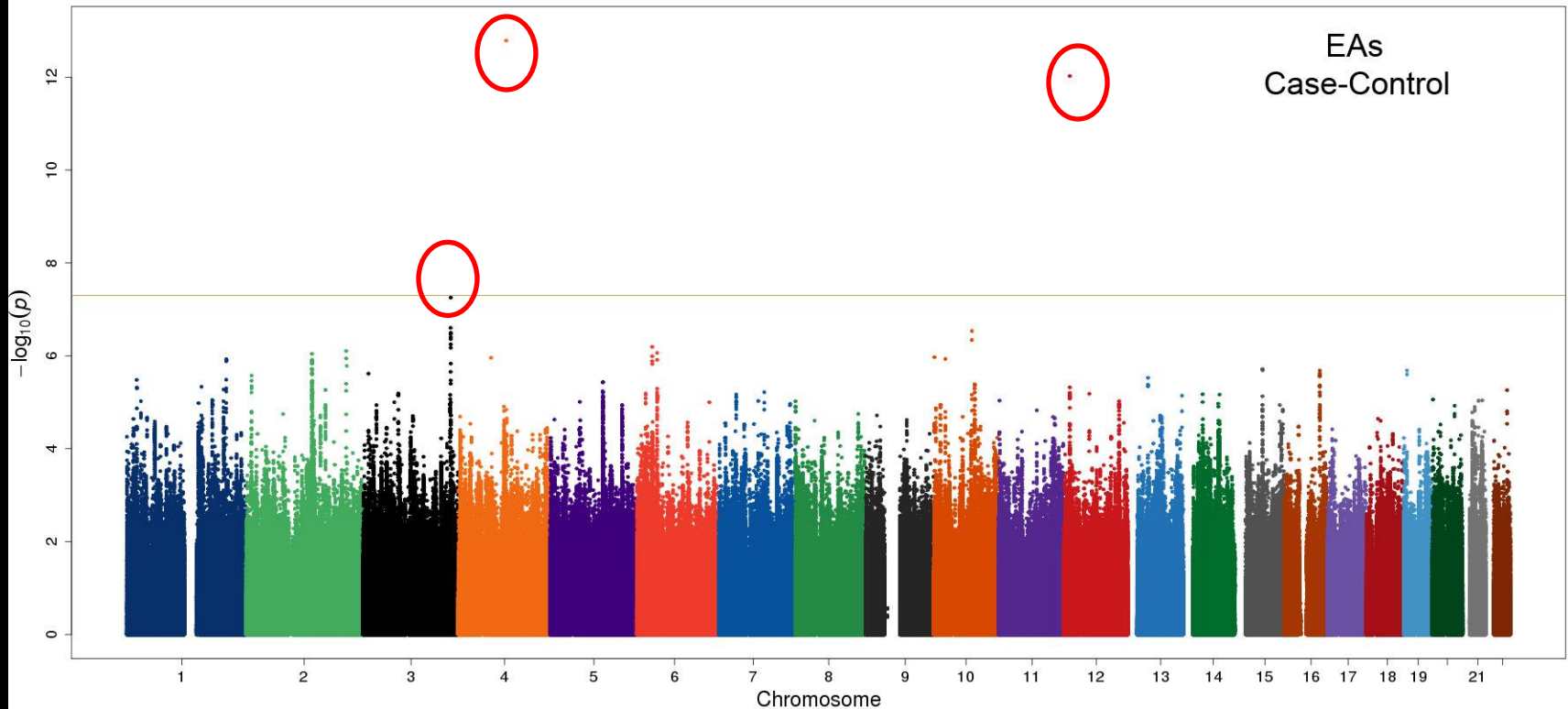
GWAS of AD: AA Results

Gelernter et al., 2014



GWAS of AD: EA Results

Gelernter et al., 2014



Current Study

Alcohol Dependence

- 3+ (within 12 months)
 - Tolerance
 - Withdrawal
 - Drinking longer than intended
 - Failure to quit drinking
 - Much time spent using/recovering from alcohol
 - Social/occupational activities foregone
 - Drinking despite physiological/psychological problems

Determine whether similar genetic factors influence alcohol dependence in EAs and AAs?

Goals of the Current Study

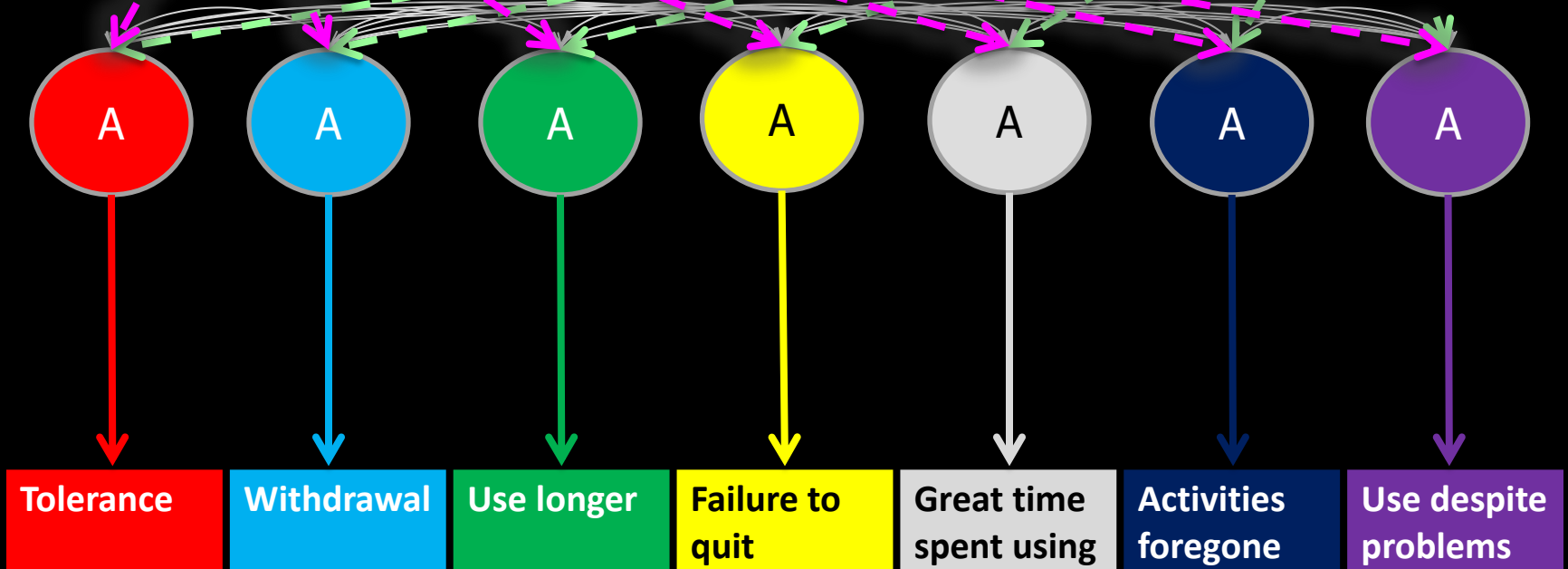
1. Investigate the extent to which additive genetic variance tagged by common SNPs explain variation in alcohol dependence.

h^2_{SNPAA}

r_{G-SNP}

h^2_{SNPAA}

2. Investigate whether these markers are shared across two populations, Eas and AAs.



Data manipulation, imputation, and analytics

METHODS

Study Samples

- Phenotype and genotype data were pooled across four studies (N~20,500 individuals).

Study (original N)	Population	Original N	# SNPs	Chip
Heroin GWAS (N=477)	EUR	6487	601273	Human 610
			592839	Human 660
			373339	Human CNV 370
Australian GWAS (OZ-ALC) (N=6,701)	EUR	6775	370404	Human CNV 370
Alcohol Dependence GWAS in Europeans and African Americans	EUR	2909	1051295	Human Omni 1
SAGE (N=4,316)	EUR	4121	109365	Human 1

- Analyses focused on ~2.2M SNPs across the various Illumina arrays.

Study Phenotypic Info

- **Assessments:** AD diagnosis and DSM-IV AD symptoms
 - Responses were limited to individuals who have been exposed to alcohol (and possibly other drugs).
 - The effective sample for all analyses included individuals with and without a lifetime diagnosis of alcohol dependence.

Data Manipulation Guide

Step 1

DATA PREPARATION

#SNPs = 2,050,463

Admixed populations from dpGAP studies:
N = 20,486

- Study of Addiction: Genetics and Environment (N=4,316)
- Australian twin-family study of alcohol use disorder (N=6,775)
- Alcohol Dependence GWAS in European- and African Americans (N=2,909)
- Genome-Wide Association Study of Heroin Dependence (N=6,486).

Identify and separate largest ancestral groups

- **STEP 1.** Conduct Quality Control (call rate < 0.95, minor allele frequency [MAF] > 0.10) and Strand Orientation
- **STEP 2.** Merge with 1000 Genomes Reference Panel
- **STEP 3.** Principal Components Analysis and Multidimensional Scaling to identify individuals within each ancestral group and remove outliers

Step 2

GENETIC IMPUTATION

African
N=2,196

Quality Control (call rate < 0.95, MAF > 0.01) and Strand Orientation

Separate by Chromosome

European
N = 6,515

Quality Control (call rate < 0.95, MAF > 0.01) and Strand Orientation

Separate by Chromosome

POOLED DATA

N=8,711

#SNPs = 1,656,234

Final Quality Control: MAF > 0.01, Hardy-Weinberg equilibrium p-value < 0.0001, call rate > 0.99, individual missingness > 0.10, imputation $r^2 > 0.5$

SVS Data Manipulation

- Step 1
 - Identify target variants across platforms
 - Import 1000 Genomes Reference Data (1KGP)
 - Prep 1KGP for ancestry determination
 - Check strand orientation in sample data
 - Integrate 1KGP with sample data
 - Estimate ancestry for sample data and select desired groups.

SVS Data Manipulation

- Raw sample data for SAGE

GENEVA_SAGE_ARC_GENO_FINAL_ZEROED BED Dataset - Sheet 1 [4]

File Edit Select DNA-Seq Genotype Numeric RNA-Seq GenomeBrowse Plot Scripts Help

All: 284 x 1,040,113
Active: 284 x 1,040,113

Unsort	C 1	C 2	C 3	C 4	B 5	B 6	G 7	G 8	G 9	G 10	G 11	G 12	G 13	G 14	G 15	G 16	G	
Map	Patients	Family ID	Patient ID	Father ID	Mother ID	Sex	Affection Status	rs12354060	rs4477212	rs2185539	rs6681105	rs12184279	rs11240767	rs12564807	rs3094315	rs3131972	rs3115860	rs
Chromosome								1	1	1	1	1	1	1	1	1	1	
Position								10004	72017	556738	581938	707348	718814	724325	742429	742584	743268	
1	360	?	360	?	?	0	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	C_T	C_T	G_T	
2	385	?	385	?	?	1	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	C_T	C_T	G_T	
3	390	?	390	?	?	1	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	T_T	C_C	T_T	
4	437	?	437	?	?	0	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	T_T	C_C	T_T	
5	461	?	461	?	?	0	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	C_T	C_T	G_T	
6	470	?	470	?	?	0	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	C_T	C_T	??	
7	482	?	482	?	?	1	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	C_T	C_T	G_T	
8	498	?	498	?	?	0	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	C_T	C_T	G_T	
9	511	?	511	?	?	0	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	T_T	C_C	T_T	
10	520	?	520	?	?	0	?	G_G	A_A	C_C	??	C_C	C_C	A_A	C_C	T_T	G_T	
11	567	?	567	?	?	0	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	T_T	C_C	T_T	
12	568	?	568	?	?	0	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	T_T	C_C	T_T	
13	616	?	616	?	?	1	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	T_T	C_C	T_T	
14	653	?	653	?	?	1	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	T_T	C_C	T_T	
15	679	?	679	?	?	1	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	C_T	C_T	G_T	
16	694	?	694	?	?	1	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	C_T	C_T	G_T	
17	699	?	699	?	?	1	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	C_T	C_T	G_T	
18	702	?	702	?	?	1	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	T_T	C_C	T_T	
19	705	?	705	?	?	0	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	T_T	C_C	T_T	
20	730	?	730	?	?	1	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	T_T	C_C	T_T	
21	741	?	741	?	?	0	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	C_T	C_T	G_T	
22	743	?	743	?	?	0	?	G_G	A_A	C_C	T_T	C_C	C_T	A_A	C_T	T_T	G_T	
23	775	?	775	?	?	0	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	C_T	C_T	G_T	
24	780	?	780	?	?	0	?	G_G	A_A	C_C	??	C_C	C_T	A_A	C_C	T_T	??	
25	784	?	784	?	?	1	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	C_T	C_T	G_T	
26	828	?	828	?	?	1	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	T_T	C_C	T_T	
27	876	?	876	?	?	1	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	C_T	C_T	T_T	
28	902	?	902	?	?	1	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	T_T	C_C	T_T	
29	906	?	906	?	?	0	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	T_T	C_C	T_T	
30	907	?	907	?	?	0	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	T_T	C_C	T_T	
31	916	?	916	?	?	1	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	T_T	C_C	T_T	
32	978	?	978	?	?	0	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	C_C	T_T	??	
33	989	?	989	?	?	1	?	G_G	A_A	C_C	T_T	C_C	C_C	A_A	T_T	C_C	T_T	

SVS Data Manipulation

Marker name

Marker map info

Sample info

Unsort	C 1	C 2	C 3	C 4	B 5	B 6	G 7	G 8	G 9	G 10	G 11	G 12	G 13	G 14	G 15	G 16	G 17	
Map	Patients	Family ID	Patient ID	Father ID	Mother ID	Sex	Affection Status	rs12354060	rs4477212	rs2185539	rs6681105	rs12184271	rs11240767	rs12564807	rs3094315	rs3131972	rs3115860	rs
Chromosome							1	1	1	1	1	1	1	1	1	1	1	
Position							10004	72017	556738	581938	707348	718814	724325	742429	742584	743268		
1	360	?	360	?	?	0	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	C,T	C,T	G,T	
2	385	?	385	?	?	1	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	C,T	C,T	G,T	
3	390	?	390	?	?	1	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	T,T	C,C	T,T	
4	437	?	437	?	?	0	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	T,T	C,C	T,T	
5	461	?	461	?	?	0	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	C,T	C,T	G,T	
6	470	?	470	?	?	0	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	C,T	C,T	?	?
7	482	?	482	?	?	1	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	C,T	C,T	G,T	
8	498	?	498	?	?	0	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	C,T	C,T	G,T	
9	511	?	511	?	?	0	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	T,T	C,C	T,T	
10	520	?	520	?	?	0	?	G,G	A,A	C,C	?	C,C	C,C	A,A	C,C	T,T	G,T	
11	567	?	567	?	?	0	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	T,T	C,C	T,T	
12	568	?	568	?	?	0	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	T,T	C,C	T,T	
13	616	?	616	?	?	1	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	T,T	C,C	T,T	
14	653	?	653	?	?	1	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	T,T	C,C	T,T	
15	679	?	679	?	?	1	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	C,T	C,T	G,T	
16	694	?	694	?	?	1	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	C,T	C,T	G,T	
17	699	?	699	?	?	1	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	C,T	C,T	G,T	
18	702	?	702	?	?	1	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	T,T	C,C	T,T	
19	705	?	705	?	?	0	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	T,T	C,C	T,T	
20	730	?	730	?	?	1	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	T,T	C,C	T,T	
21	741	?	741	?	?	0	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	C,T	C,T	G,T	
22	743	?	743	?	?	0	?	G,G	A,A	C,C	T,T	C,C	C,T	A,A	C,T	T,T	G,T	
23	775	?	775	?	?	0	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	C,T	C,T	G,T	
24	780	?	780	?	?	0	?	G,G	A,A	C,C	?	C,C	C,T	A,A	C,C	T,T	?	?
25	784	?	784	?	?	1	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	C,T	C,T	G,T	
26	828	?	828	?	?	1	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	T,T	C,C	T,T	
27	876	?	876	?	?	1	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	C,T	C,T	T,T	
28	902	?	902	?	?	1	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	T,T	C,C	T,T	
29	906	?	906	?	?	0	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	T,T	C,C	T,T	
30	907	?	907	?	?	0	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	T,T	C,C	T,T	
31	916	?	916	?	?	1	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	T,T	C,C	T,T	
32	978	?	978	?	?	0	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	C,C	T,T	?	?
33	989	?	989	?	?	1	?	G,G	A,A	C,C	T,T	C,C	C,C	A,A	T,T	C,C	T,T	

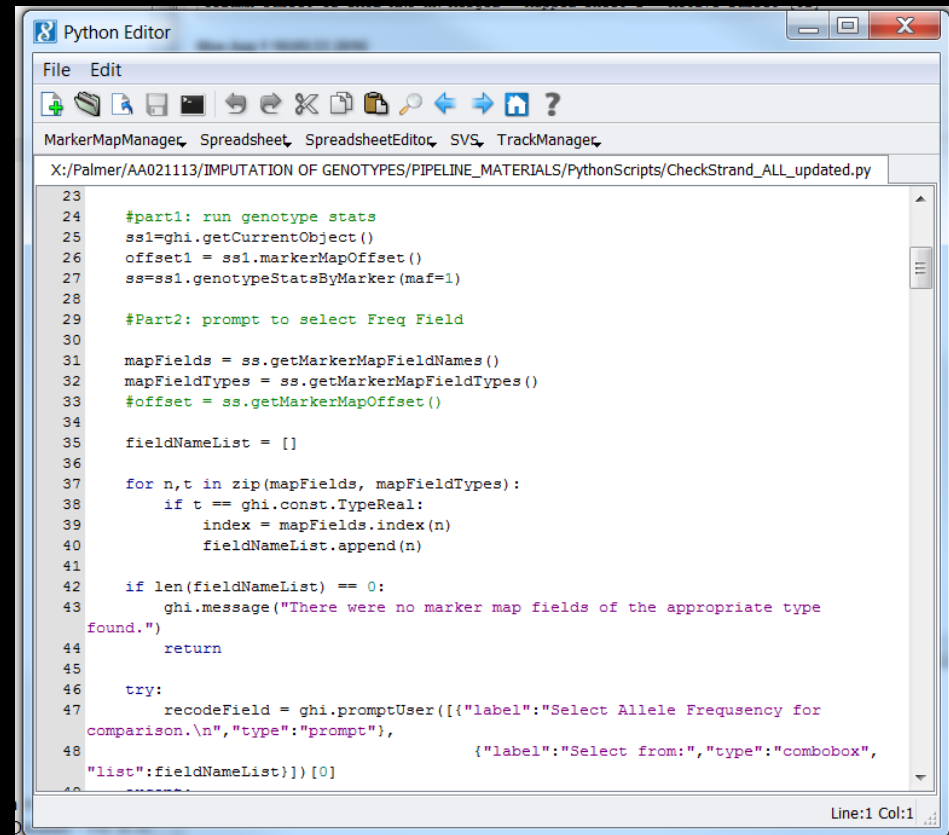
Marker calls

SVS Data Manipulation

- Obtain 1KGP reference panel and filter on target variants.
 - MAF > 5%
 - Call Rate > 99%
 - LD prune (r^2 threshold of 0.5)
- Obtain sample information on 1KGP participants.
 - Super-population classification
 - African (AFR)
 - Americas (AMR)
 - East Asian (EAS)
 - European (EUR)
 - South Asian (SAS)

SVS Data Manipulation

- Apply 1KGP marker map containing ancestral population specific allele frequencies to sample data.
- QC:
 - CR < .95
 - MAF < .1
- Check strand orientation



```
Python Editor
File Edit
MarkerMapManager Spreadsheet SpreadsheetEditor SVS TrackManager
X:/Palmer/AA021113/IMPUTATION OF GENOTYPES/PIPELINE_MATERIALS/PythonScripts/CheckStrand_ALL_updated.py
23
24 #part1: run genotype stats
25 ss1=ghi.getCurrentObject()
26 offset1 = ss1.markerMapOffset()
27 ss=ss1.genotypeStatsByMarker(maf=1)
28
29 #Part2: prompt to select Freq Field
30
31 mapFields = ss.getMarkerMapFieldNames()
32 mapFieldTypes = ss.getMarkerMapFieldTypes()
33 #offset = ss.getMarkerMapOffset()
34
35 fieldNameList = []
36
37 for n,t in zip(mapFields, mapFieldTypes):
38     if t == ghi.const.TypeReal:
39         index = mapFields.index(n)
40         fieldNameList.append(n)
41
42 if len(fieldNameList) == 0:
43     ghi.message("There were no marker map fields of the appropriate type
44 found.")
45     return
46
47 try:
48     recodeField = ghi.promptUser({"label":"Select Allele Frequency for
49 comparison.\n","type":"prompt"}, {"label":"Select from:","type":"combobox",
50 "list":fieldNameList})[0]
```


SVS Data Manipulation

- Combing 1KGP and QC'd sample data:

LD pruned SNPs

Unsort	C 1	C 2	C 3	C 4	B 5	B 6	G 7	G 8	G 9	G 10	G 11	G 12	G 13	G 14	G 15	G 16	G 17
Map	Patients	Family ID	Patient ID	Father ID	Mother ID	Sex	Affection Status	rs12124819	rs1806509	rs7537756	rs4040604	rs1110052	rs28705211	rs6696281	rs28504611	rs2340592	rs1891910
4299	4364	?	4364	?	?	1	?	A_A	?	A_G	?	G_G	C_G	?	C_C	?	G_G
4300	1486	?	1486	?	?	1	?	A_A	A_C	A_A	G_T	G_T	G_G	C_C	C_T	A_G	A_G
4301	575	?	575	?	?	0	?	A_A	A_A	A_A	T_T	T_T	G_G	C_C	C_C	G_G	G_G
4302	621	?	621	?	?	1	?	A_A	C_C	A_A	G_G	G_G	G_G	C_T	C_T	A_G	G_G
4303	3704	?	3704	?	?	1	?	A_A	C_C	G_G	T_T	G_T	G_G	C_C	C_C	A_G	G_G
4304	3638	?	3638	?	?	0	?	A_G	A_C	A_G	T_T	G_T	G_G	C_C	C_C	A_G	G_G
4305	1188	?	1188	?	?	0	?	A_G	A_C	A_A	T_T	T_T	G_G	C_C	C_C	G_G	A_G
4306	738	?	738	?	?	0	?	A_A	C_C	A_A	G_G	G_G	G_G	C_T	T_T	A_G	G_G
4307	233-1204	233	1204	?	?	1	?	A_G	C_C	A_G	T_T	G_T	C_G	C_C	C_C	A_G	G_G
4308	3318	?	3318	?	?	1	?	A_G	A_A	A_A	?	?	G_G	C_C	C_C	G_G	G_G
4309	4349	?	4349	?	?	1	?	A_G	A_C	A_G	T_T	G_T	C_G	C_C	C_C	A_G	A_G
4310	4521	?	4521	?	?	0	?	A_A	A_A	A_A	T_T	T_T	G_G	C_C	C_C	G_G	G_G
4311	1826	?	1826	?	?	1	?	A_G	A_A	A_A	G_G	G_G	G_G	C_T	C_C	A_G	G_G
4312	1272	?	1272	?	?	0	?	A_A	A_C	A_A	G_T	G_T	G_G	C_T	C_T	A_G	G_G
4313	3284	?	3284	?	?	0	?	A_A	A_A	A_A	G_T	G_T	G_G	C_C	C_T	G_G	G_G
4314	3976	?	3976	?	?	0	?	A_A	C_C	A_G	G_T	G_G	G_G	C_T	T_T	A_A	A_G
4315	1556	?	1556	?	?	0	?	A_A	A_A	A_A	T_T	G_T	C_G	C_C	C_C	G_G	G_G
4316	1980	?	1980	?	?	1	?	A_G	A_A	A_A	T_T	T_T	G_G	C_C	C_C	G_G	A_G
4317	HG00096	?	?	?	?	?	?	A_A	A_C	A_A	T_T	G_T	C_G	C_T	C_C	A_G	A_G
4318	HG00097	?	?	?	?	?	?	A_A	C_C	A_G	T_T	G_G	C_C	C_C	C_C	A_G	G_G
4319	HG00099	?	?	?	?	?	?	A_A	A_A	A_A	T_T	T_T	G_G	C_C	C_C	G_G	G_G
4320	HG00100	?	?	?	?	?	?	A_G	A_C	A_G	T_T	T_T	G_G	C_C	C_C	A_G	G_G
4321	HG00101	?	?	?	?	?	?	A_A	A_A	A_A	T_T	G_T	C_C	C_C	C_C	G_G	A_G
4322	HG00102	?	?	?	?	?	?	A_G	A_A	A_A	T_T	T_T	G_G	C_C	C_C	G_G	A_G
4323	HG00103	?	?	?	?	?	?	A_G	A_C	A_A	T_T	T_T	G_G	C_C	C_C	G_G	G_G
4324	HG00105	?	?	?	?	?	?	A_G	A_C	A_A	T_T	T_T	G_G	C_T	C_C	G_G	G_G
4325	HG00106	?	?	?	?	?	?	A_G	A_C	A_A	T_T	G_T	C_G	C_C	C_C	A_G	G_G
4326	HG00107	?	?	?	?	?	?	A_G	A_A	A_A	T_T	T_T	G_G	C_C	C_C	G_G	A_A
4327	HG00108	?	?	?	?	?	?	A_A	A_A	A_A	T_T	T_T	G_G	C_C	C_C	G_G	G_G
4328	HG00109	?	?	?	?	?	?	A_G	A_C	A_A	T_T	T_T	C_G	C_T	C_C	A_G	G_G
4329	HG00110	?	?	?	?	?	?	G_G	A_C	A_G	T_T	G_T	C_G	C_C	C_C	A_G	A_G
4330	HG00111	?	?	?	?	?	?	A_A	A_A	A_A	T_T	T_T	G_G	C_C	C_C	G_G	G_G
4331	HG00112	?	?	?	?	?	?	G_G	A_A	A_A	T_T	T_T	G_G	C_C	C_C	G_G	A_G
4332	HG00113	?	?	?	?	?	?	G_G	A_A	A_A	T_T	G_T	C_G	C_C	C_C	A_G	G_G
4333	HG00114	?	?	?	?	?	?	A_A	A_C	A_A	T_T	T_T	G_G	C_C	C_C	G_G	G_G

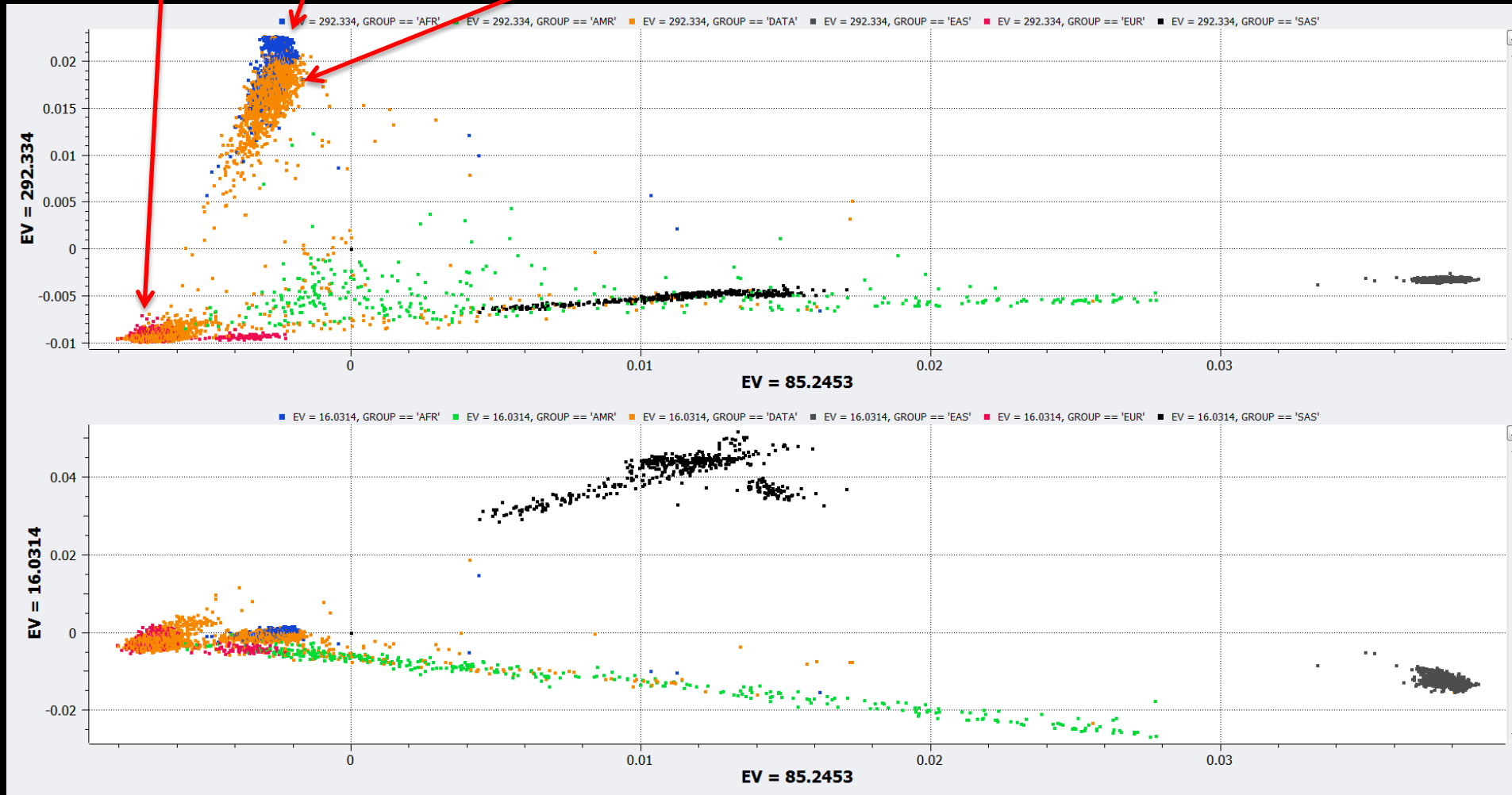
Combined samples

Ancestry Determination

EUR

AFR

sample data



Ancestry Selection

- First, compute threshold for first principal component (PC; separate largest ancestral groups).
 - EUR and AFR.
 - Retain sample data that falls within two standard deviation of the mean of the first PC in the ancestral population.
- Second, use Multidimensional Scaling to reduce stratification within EUR and AFR subgroups.
 - Use the first 3 PCs within each group to remove multidimensional outliers.

End of Step 1

- Resulting samples after ancestry determination.

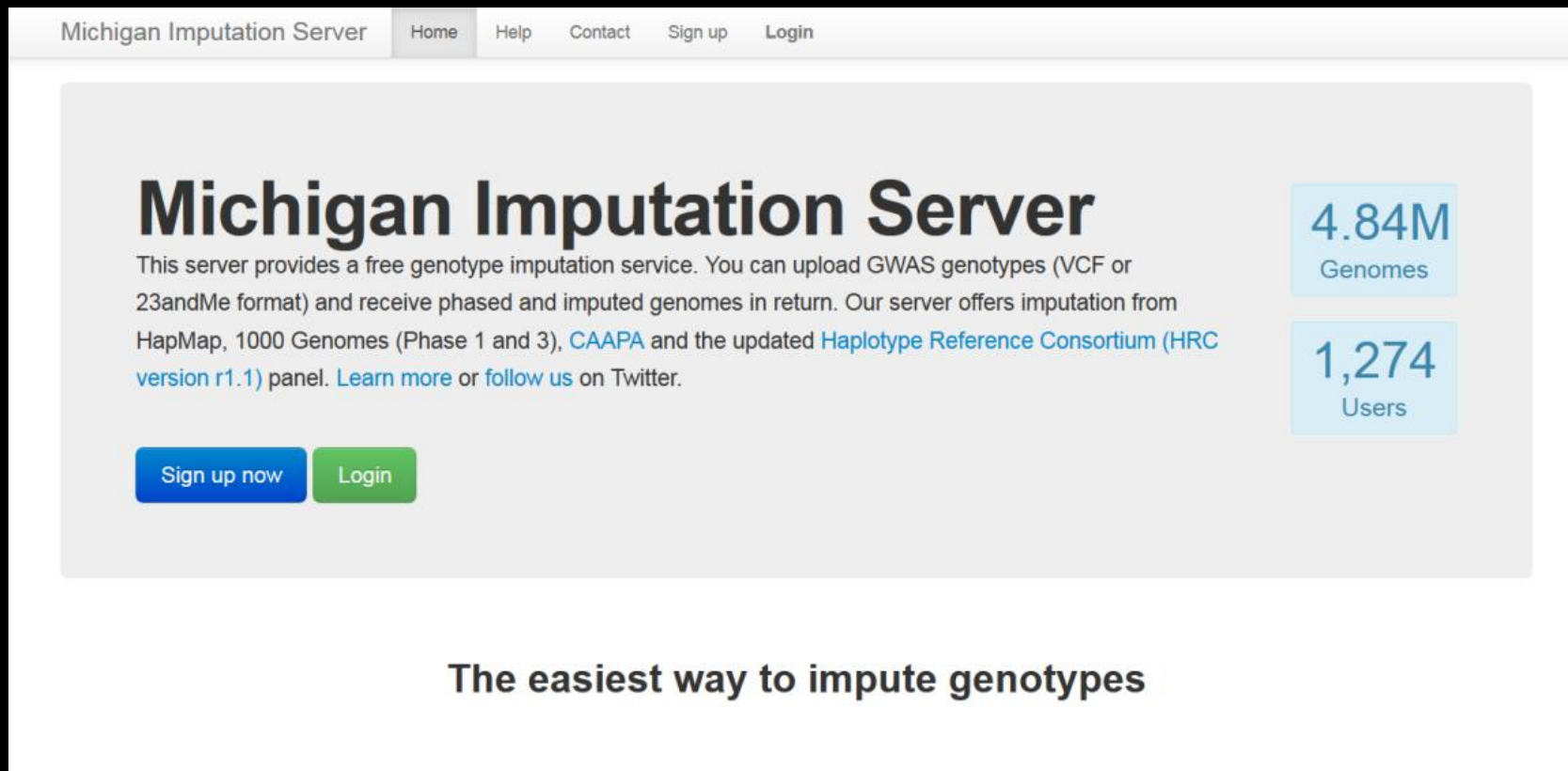
Study (original N)	Population	Final N	# SNPs
Heroin C4 (N=477)	EUR	374	514171
Heroin C1 (N=874)	EUR	660	513392
Heroin C2 (N=142)	EUR	131	510886
Heroin C2-GWASRel6 (N=2440)	EUR	2216	293504
Heroin C3 (N=2008)	EUR	1417	514114
Heroin C4-YaleCases (N=469)	EUR	420	320248
OZALC (N=6,701)	EUR	6052	296951
YALE (N=2,909)	EUR	711	758107
YALE (N=2,909)	AFR	1541	844068
SAGE (N=4,316)	EUR	2434	825419
SAGE (N=4,316)	AFR	1018	909,846
Heroin C3 (N=2008)	AFR	1	42205
Heroin C4 (N=477)	AFR	1	43577

Imputation Preparation

- Subset original sample file using the determined ancestral groups.
- Conduct QC and strand orientation check
 - CR > 95%
 - MAF > 1%
 - Individual missingness > 95% per chromosome
 - Strand Check within ancestral group
- Save each chromosome as a separate *.vcf file

Step 2: Imputation Description

- Upload individual *.vcf files to Michigan Imputation Server



The screenshot shows the homepage of the Michigan Imputation Server. At the top, there is a navigation bar with links for Home, Help, Contact, Sign up, and Login. The main content area features the title "Michigan Imputation Server" in large, bold black text. Below the title, a paragraph describes the service: "This server provides a free genotype imputation service. You can upload GWAS genotypes (VCF or 23andMe format) and receive phased and imputed genomes in return. Our server offers imputation from HapMap, 1000 Genomes (Phase 1 and 3), CAAPA and the updated Haplotype Reference Consortium (HRC version r1.1) panel. Learn more or follow us on Twitter." To the right of this text, there are two light blue boxes: the top one displays "4.84M Genomes" and the bottom one displays "1,274 Users". Below the main text, there are two buttons: a blue "Sign up now" button and a green "Login" button. At the bottom of the page, the slogan "The easiest way to impute genotypes" is centered in a bold, dark grey font.

Michigan Imputation Server Home Help Contact Sign up Login

Michigan Imputation Server

This server provides a free genotype imputation service. You can upload GWAS genotypes (VCF or 23andMe format) and receive phased and imputed genomes in return. Our server offers imputation from HapMap, 1000 Genomes (Phase 1 and 3), CAAPA and the updated Haplotype Reference Consortium (HRC version r1.1) panel. [Learn more](#) or [follow us](#) on Twitter.

4.84M
Genomes

1,274
Users

Sign up now Login

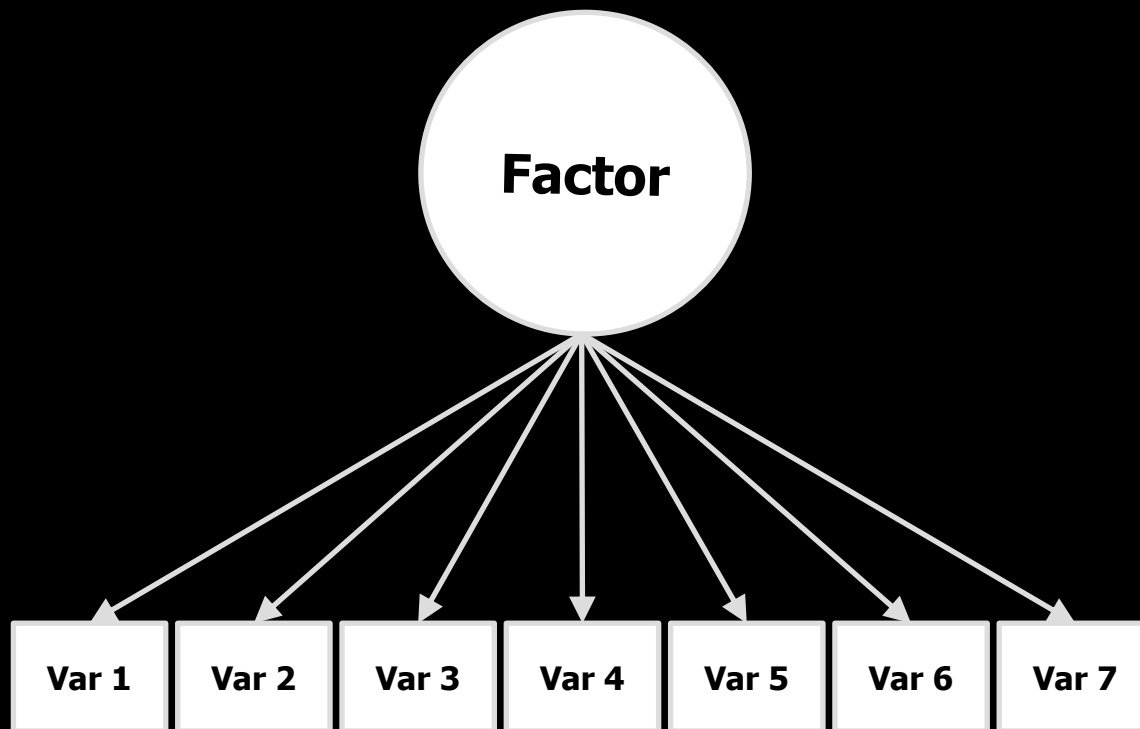
The easiest way to impute genotypes

Step 3: Analytical Approach

- **Approach:**
 - **Common Pathway Model**
 - Exploratory and confirmatory factor analysis of AD symptoms.
 - Quantification of SNP heritability of identified factor(s)
 - **Test for invariance across ethnic groups**
 - **GREML within & between ethnic groups**
 - DeCandia et al., 2013
 - **Study Covariates included:**
 - age, gender, study origin (COGA, COGEND, FSCD), and ancestry (using 3 ancestral principal components).

Common Pathway Model

- **A multivariate statistical model that explores whether common genetic and environmental factors influence all observed variables via a single psychometric factor, or underlying latent liability.**



Invariance Testing

Is the same construct being measured in EAs & AAs?

1. Equal form: Test if the number of factors are identical across groups.
2. Equal loadings: Test if factor loadings are equal across groups.
3. Equal thresholds: Tests if the item thresholds are equal across groups.
4. Equal residual variances: Tests if the residual variances of the observed scores not accounted for by the factors are equal across groups.

GREML Overview (Yang et al., 2010)

- Estimate SNP-based heritability (h^2_{SNP}) of a trait:
 - Amount of phenotypic variation (V_P) that is due to additive genetic variation (V_A) among individuals in a population

$$h^2_{SNP} = \frac{V_A}{V_P}$$



e.g. human height has a heritability of ~ 0.80 from twin/family studies, and a SNP-based heritability of 0.45 from genome-wide SNPS

GREML Overview

- Mixed linear model:
 - Decompose the phenotypic variance into two components:
 1. A random effect representing the additive genetic variance of all measured SNPs (h^2_{SNP})
 2. An effect representing unmeasured environmental variance, genetic effects that were not captured (i.e. by the genotyping array), and random noise.

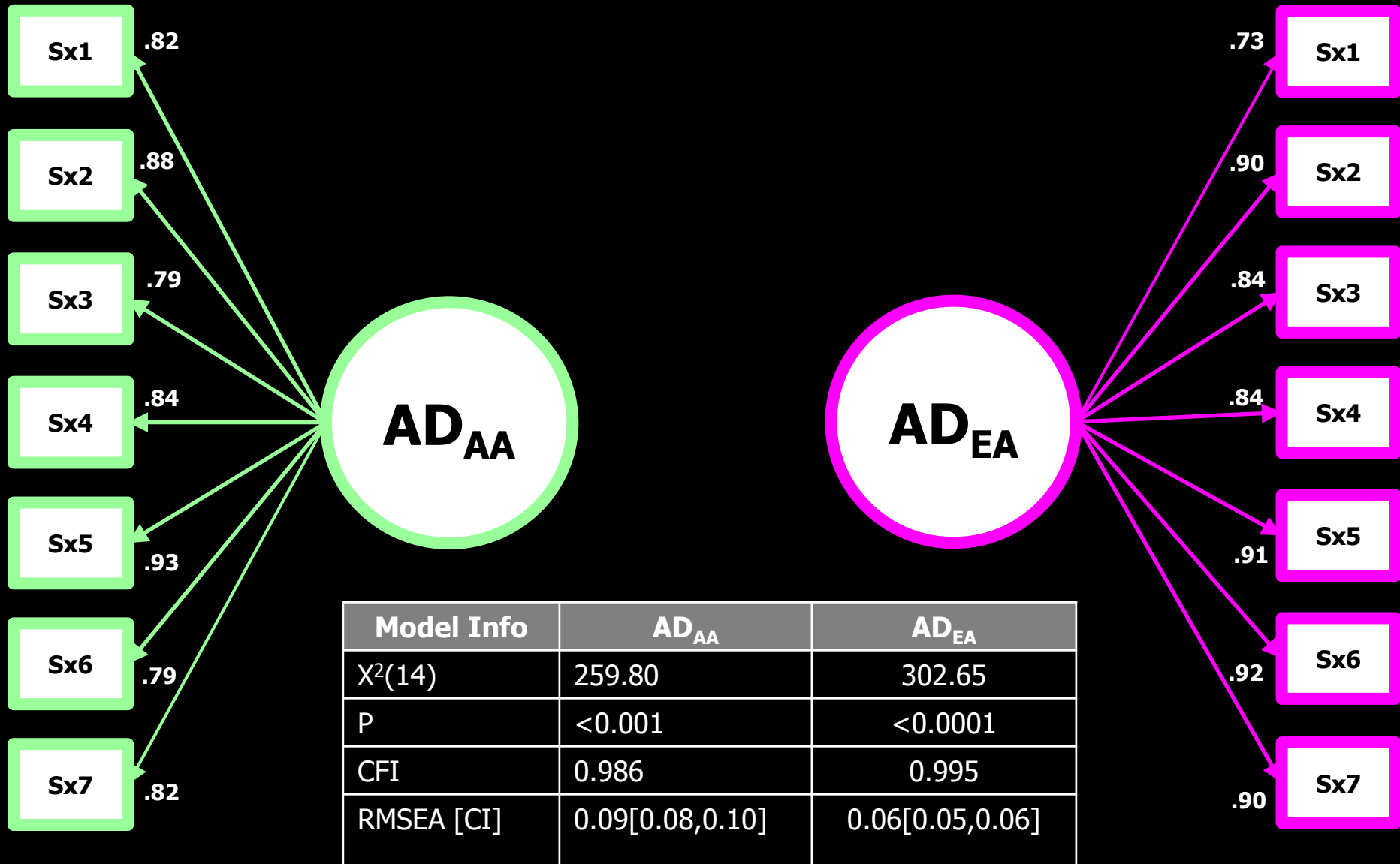
$$y = Xb + g + e$$

$$V = AS_g^2 + IS_e^2$$

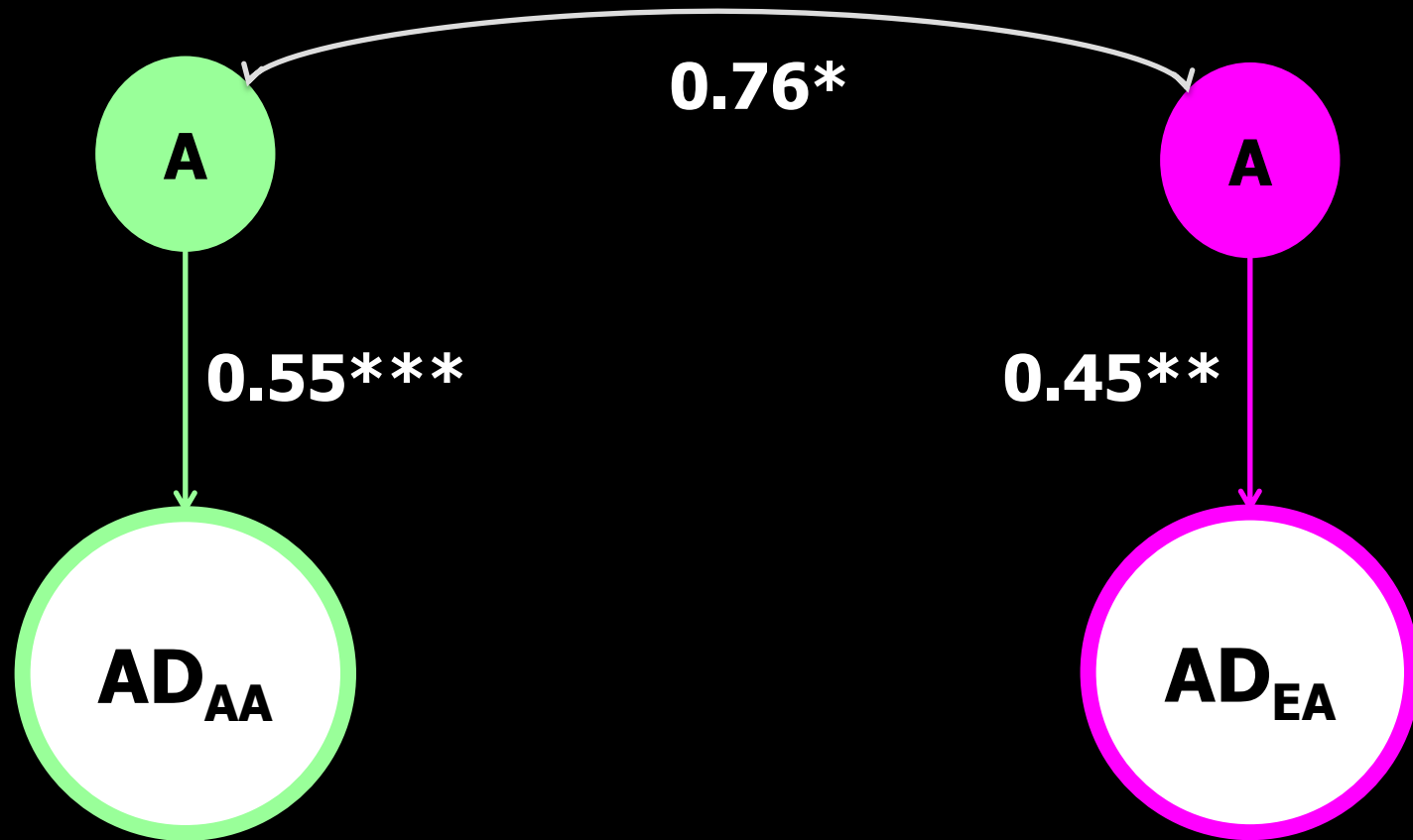
- We incorporate fixed effects: sex, age, ancestral principle components as covariates.
- The bivariate model estimates the genetic covariance between two traits that can be captured by all SNPs.

RESULTS

CFA: Equal Forms Supported



GREML: Genetic Effects on AD factors



*- $P < 0.05$, ** - $P < 0.01$, *** - $P < 0.001$

Summary

- **Similar SNP-based heritability estimates for individuals of European and African ancestry.**
- **A large genetic correlation that provides evidence for overlapping genetic factors influencing AD in EAs and AAs.**
- **Simultaneous estimation of SNP effects may be useful, but requires careful specification and interpretation.**
- **Follow-up work to improve model specification and identification of variants*.**

Future Directions:

Dissect genetic variance using Bayesian mixture models

Use four zero-mean normal distributions of SNP effects (0=Null effect, 10^{-4} = polygenic effect, 10^{-3} = small effect, 10^{-2} = moderate effect).

Recall: $h^2_{\text{SNP}_{EA}} = 0.20$

BayesR $h^2_{\text{SNP}} = 0.032$
 $\sim N$ SNPs = 4767

Recall: $h^2_{\text{SNP}_{AA}} = 0.30$

BayesR $h^2_{\text{SNP}} = 0.034$
 $\sim N$ SNPs = 1745

Disclosure & Acknowledgements

Funding sources:

- NIAAA: K01 AA021113 [Palmer]
- NIDA: R01 DA023134 [Knopik]
- NIMH: R01MH100141 [Keller]

Data Providers

- Database of Genotypes & Phenotypes*
 - U01 HG004422
 - U10 AA008401
 - P01 CA089392
 - R01 DA013423
 - R01 DA019963.



GOLDEN HELIX



Rhode Island Hospital
A Lifespan Partner



BROWN