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
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:
  - 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

- 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 by variation in a gene.
- Do not have predictable patterns of inheritance.
- Spectrum of genetic effects is broad impacting proteins directly and directly; effect sizes are small.
Alcoholism Is A Complex Disease

GWAS suggests many variants of small effect.

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,DβH, 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 \( h^2_{\text{SNP}; \text{s.e.}} \) of AD

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Current Study ( h^2_{\text{SNP}; \text{s.e.}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD Diagnosis</td>
<td>0.300 (0.136) (^a)</td>
</tr>
<tr>
<td>AD factor score</td>
<td>0.307 (0.130) (^a)</td>
</tr>
</tbody>
</table>

### 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.

---

Palmer et al., 2015
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
- $5 \times 10^{-8}$ 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.

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).

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

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

- 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 2**

**Genetic Imputation**

- **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

**Pooled Data**

- 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

Final Quality Control: MAF > 0.01, Hardy-Weinburg equilibrium p-value <0.0001, call rate > 0.99, individual missingness > 0.10, imputation r² > 0.5
SVS Data Manipulation

• Step 1
  – Identify target variants across platforms
  – Import 1000 Genomes Reference Data (1GKP)
  – 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
SVS Data Manipulation

<table>
<thead>
<tr>
<th>Marker name</th>
<th>Marker map info</th>
<th>Sample info</th>
<th>Marker calls</th>
</tr>
</thead>
<tbody>
<tr>
<td>360</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>385</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>390</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>437</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>461</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>470</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>482</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>498</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>511</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>520</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>567</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>568</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>569</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>570</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>577</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>702</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>705</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>730</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>741</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>743</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>775</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>780</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>781</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>828</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>876</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>902</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>904</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>906</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>916</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>917</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>987</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>988</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>989</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>
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
### SVS Data Manipulation

- **Combing 1KGP and QC’d sample data:**

[Image of spreadsheet showing genetic data with annotations for LD pruned SNPs and combined samples.]
Ancestry Determination

sample data

EUR

AFR
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.

<table>
<thead>
<tr>
<th>Study (original N)</th>
<th>Population</th>
<th>Final N</th>
<th># SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heroin C4 (N=477)</td>
<td>EUR</td>
<td>374</td>
<td>514171</td>
</tr>
<tr>
<td>Heroin C1 (N=874)</td>
<td>EUR</td>
<td>660</td>
<td>513392</td>
</tr>
<tr>
<td>Heroin C2 (N=142)</td>
<td>EUR</td>
<td>131</td>
<td>510886</td>
</tr>
<tr>
<td>Heroin C2-GWASRel6 (N=2440)</td>
<td>EUR</td>
<td>2216</td>
<td>293504</td>
</tr>
<tr>
<td>Heroin C3 (N=2008)</td>
<td>EUR</td>
<td>1417</td>
<td>514114</td>
</tr>
<tr>
<td>Heroin C4-YaleCases (N=469)</td>
<td>EUR</td>
<td>420</td>
<td>320248</td>
</tr>
<tr>
<td>OZALC (N=6,701)</td>
<td>EUR</td>
<td>6052</td>
<td>296951</td>
</tr>
<tr>
<td>YALE (N=2,909)</td>
<td>EUR</td>
<td>711</td>
<td>758107</td>
</tr>
<tr>
<td>YALE (N=2,909)</td>
<td>AFR</td>
<td>1541</td>
<td>844068</td>
</tr>
<tr>
<td>SAGE (N=4,316)</td>
<td>EUR</td>
<td>2434</td>
<td>825419</td>
</tr>
<tr>
<td>SAGE (N=4,316)</td>
<td>AFR</td>
<td>1018</td>
<td>909,846</td>
</tr>
<tr>
<td>Heroin C3 (N=2008)</td>
<td>AFR</td>
<td>1</td>
<td>42205</td>
</tr>
<tr>
<td>Heroin C4 (N=477)</td>
<td>AFR</td>
<td>1</td>
<td>43577</td>
</tr>
</tbody>
</table>
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
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).
• 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.
Estimate SNP-based heritability ($h^2_{SNP}$) of a trait:

- Amount of phenotypic variation (VP) that is due to additive genetic variation (VA) 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 (h²_{SNP})
    2. An effect representing unmeasured environmental variance, genetic effects that were not captured (i.e. by the genotyping array), and random noise.

\[
\begin{align*}
  y &= Xb + g + e \\
  V &= A \sigma^2_g + I \sigma^2_e
\end{align*}
\]

• 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.
CFA: Equal Forms Supported

<table>
<thead>
<tr>
<th>Model Info</th>
<th>AD_{AA}</th>
<th>AD_{EA}</th>
</tr>
</thead>
<tbody>
<tr>
<td>X^2(14)</td>
<td>259.80</td>
<td>302.65</td>
</tr>
<tr>
<td>p</td>
<td>&lt;0.001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>CFI</td>
<td>0.986</td>
<td>0.995</td>
</tr>
<tr>
<td>RMSEA [CI]</td>
<td>0.09[0.08,0.10]</td>
<td>0.06[0.05,0.06]</td>
</tr>
</tbody>
</table>
GREML: Genetic Effects on AD factors

A → AD_{AA} (0.55***)

A → AD_{EA} (0.45**)

A → AD_{AA} (0.76*)

*- 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: $h^2_{SNP_{EA}} = 0.20$

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

Recall: $h^2_{SNP_{AA}} = 0.30$

BayesR $h^2_{SNP} = 0.034$
$\sim N$ 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.