

# Prediction of cannabis use disorder severity from genetic and behavioral data

Ariel Ketcherside

Milind Rao

Shikha Prashad

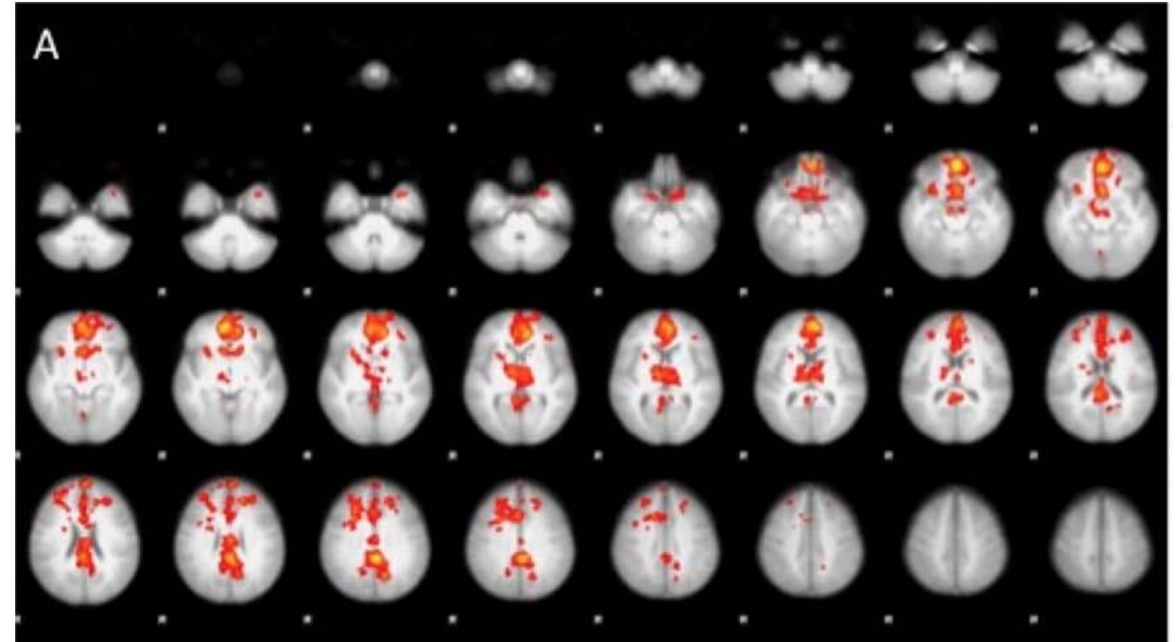
# Substance use disorders (SUDs)

- Continued use despite knowledge of negative consequences
  - Negative health consequences
  - Impacts on social relationships
- Neurological mechanisms of SUDs
  - Involve reward pathway
  - Can be examined through reports of behavior
  - Have a strong genetic component
    - Cannabis: 51–59% heritable (Agrawal et al. 2012)



# Cannabis use disorders (CUDs)

- 10% of cannabis users develop a cannabis use disorder
- Cannabis activates reward circuitry in the same way as other drugs of abuse

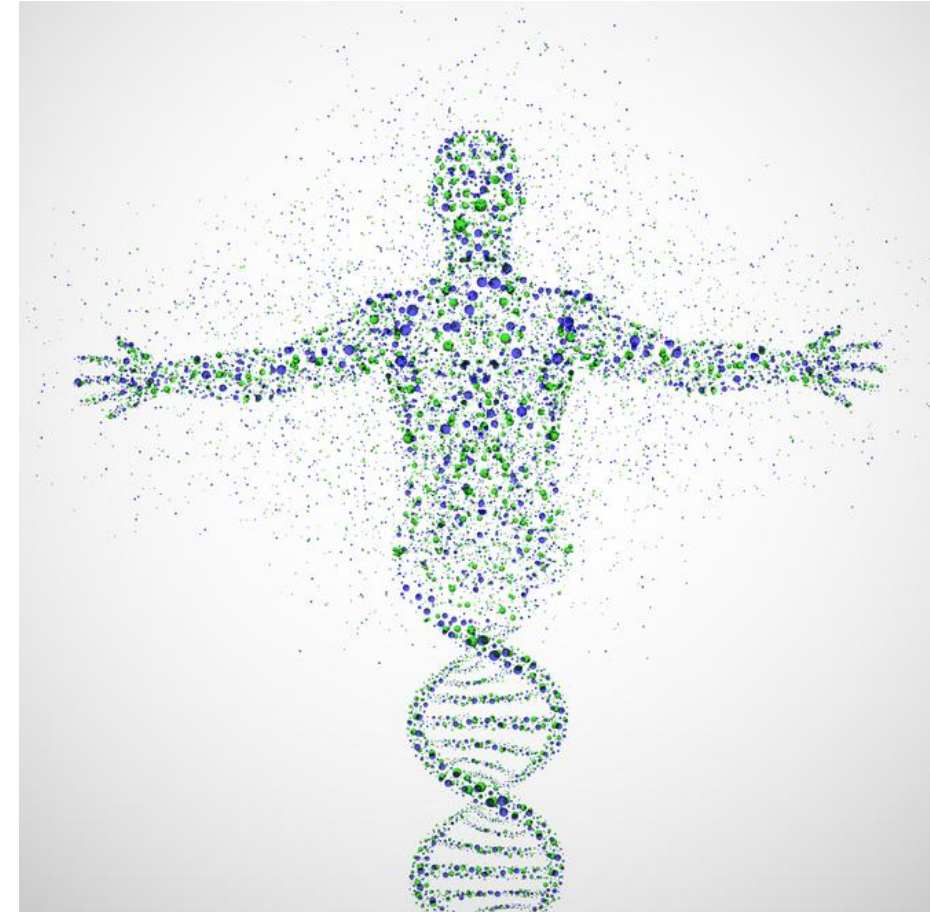


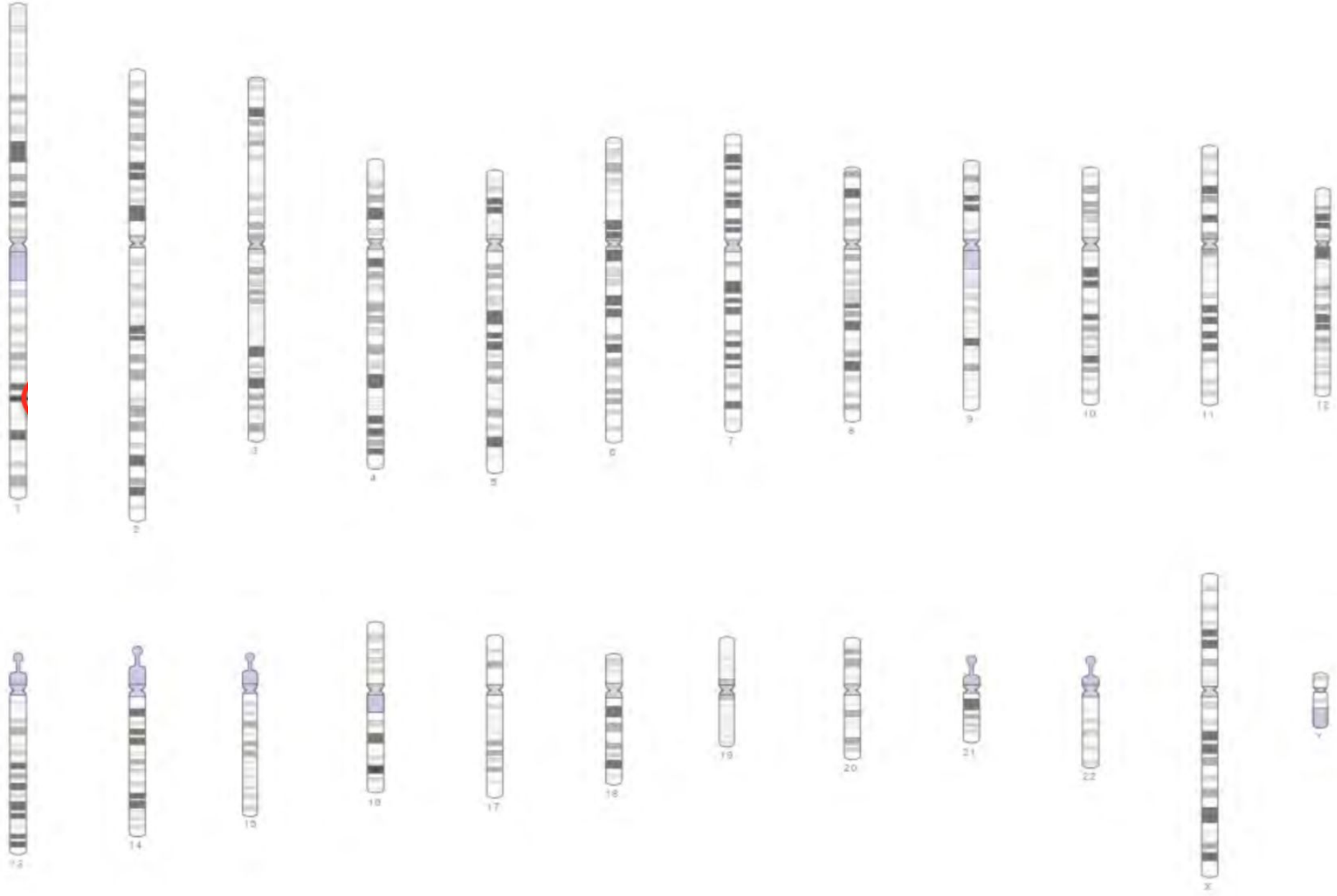
Cannabis users > controls when viewing cannabis-related stimuli

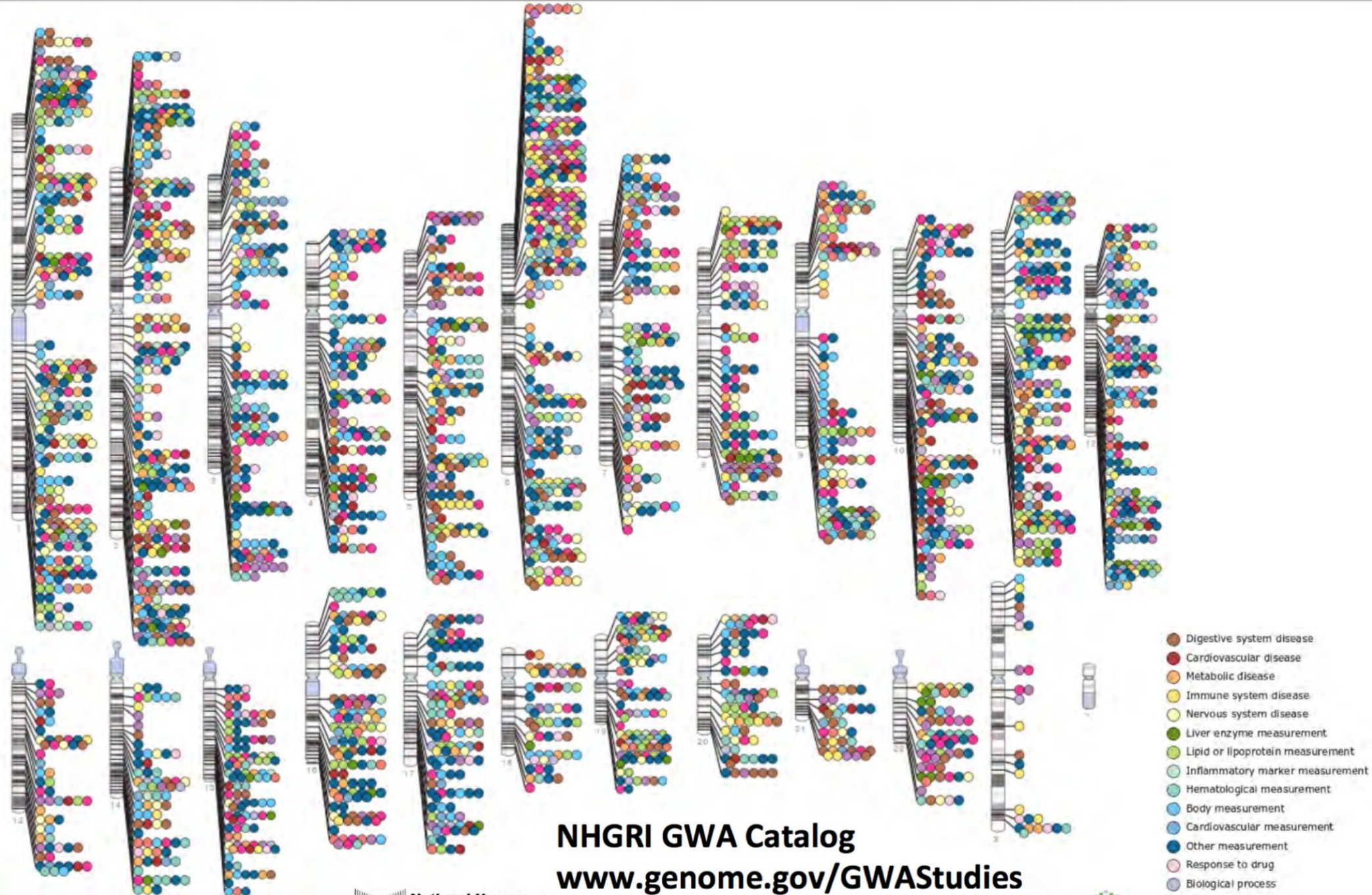
Filbey et al. 2016, *Human Brain Mapping*

# Genetics 101

- Single nucleotide polymorphisms (SNPs) create genetic variability among individuals
- Genetic variability encodes molecular variability that influences neural processes
  - *Motivation*
  - *Reward*
  - *Impulsivity*
- We need a way to sample large quantities of SNP variability







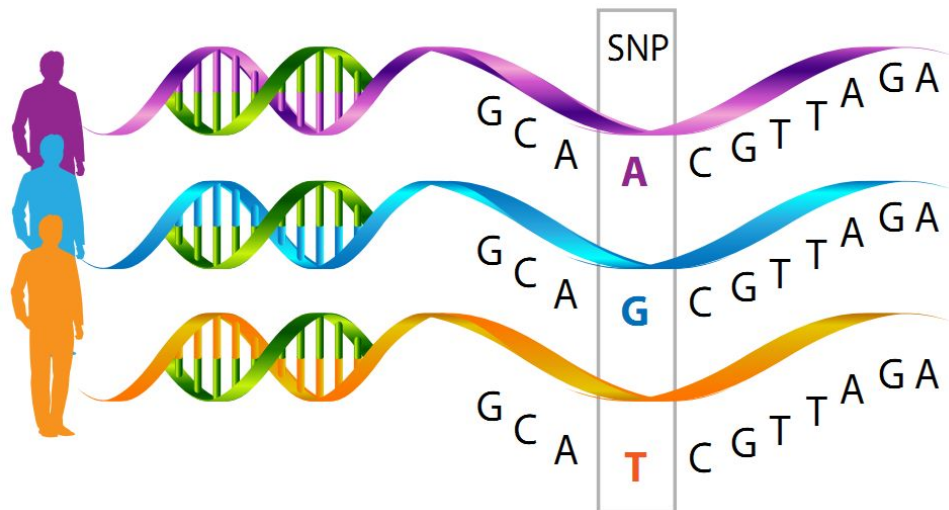
**NHGRI GWA Catalog**  
[www.genome.gov/GWASudies](http://www.genome.gov/GWASudies)

[www.ebi.ac.uk/fgpt/gwas/](http://www.ebi.ac.uk/fgpt/gwas/) EMBL-EBI



# Genome Wide Association Studies (GWAS)

- Single nucleotide polymorphisms
- Measured on high-throughput genotyping plates



BEGIN DATA		BEGIN IMAGE INFO						
Index	Array Row	Array Colu	Spot Row	Spot Colun	Name	ID	X	Y
BEGIN	1	1	1	1	1 'ABCG8	TCHA0013	670	670
ARRAY PATTERN INFO	2	1	1	1	2 'Acetyltran	TCHA0018	860	860
Units	3	1	1	1	3 'ABHD2	TCHA0013	1040	1040
µm	4	1	1	1	4 'Acetyltran	TCHA0018	1230	1230
Array Row:	12	5	1	1	5 'ABHD4	TCHA0013	1400	1400
Array Colu:	4	6	1	1	6 'Acetyltran	TCHA0018	1625	1625
Spot Rows:	24	7	1	1	7 'ABHD6	TCHA0013	1820	1820
Spot Colun:	24	8	1	1	8 'ACHE	TCHA0018	1970	1970
Array Row:	4500	9	1	1	9 'ALOX15	TCHA0050	2140	2140
Array Colu:	4500	10	1	1	10 'AMP-bind	TCHA0055	2325	2325
Spot Row !:	186	11	1	1	11 'ALOXE3	TCHA0051	2515	2515

# Our data

- Our chip assesses 800,000 different loci on the genome
- Marijuana Problem Scale: Measure of cannabis use disorder severity (CUD)
  - *E.g. “how often has marijuana caused you to have fights with family or friends?”*
- Marijuana Craving questionnaire (MCQ)
  - *Questions administered after 3 days of abstinence from cannabis*
  - *E.g. how much would you like to smoke marijuana right now?*
- Participants: 235 cannabis users



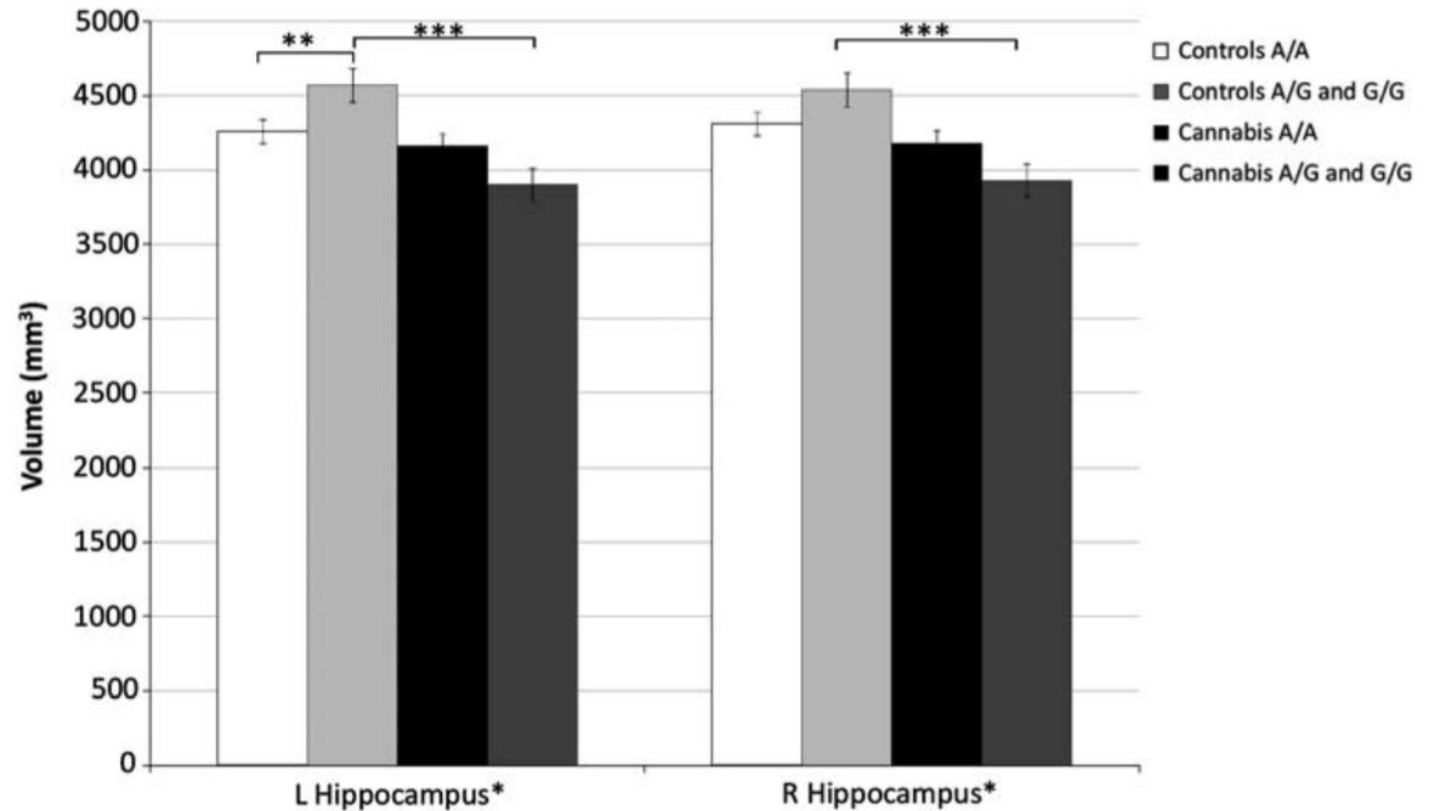
# Aims: Clinical Application

- Aim 1: To predict severity of cannabis use disorder based on their SNPs
  - *Which SNPs contribute most to CUD?*
  - *Are there interaction effects?*
- Aim 2: Determine whether severity can be more accurately predicted by including measures of craving
  - *Are some alleles related to greater reported subjective craving after 3 days of abstinence?*

# Genetics Analysis Approaches

# Candidate allele studies

- One SNP
- Ex: rs2023239 in cannabis users



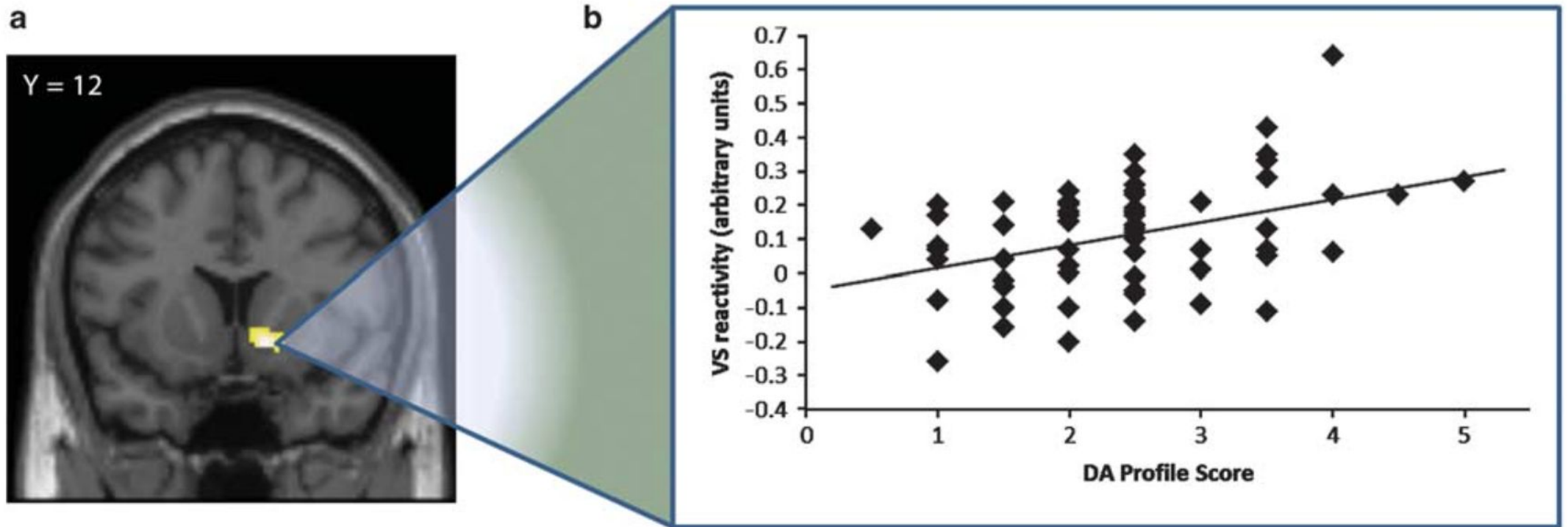
# Multilocus Genetic Profile Approach (MLGP)

- *a priori* collection of specific SNPs
- used to calculate a risk score
- Limited to a linear model which may not encompass size of contribution of SNP

**Table 1** Composition and Distribution of Multilocus Genetic Profile Scores

<b>Polymorphism</b>	<b>Genotypes</b>	<b>N</b>	<b>DA profile score</b>
DRD2 -141C Ins/Del	Del carrier	14	High
	Ins/Ins	55	Low
DAT1 VNTR	9-repeat carrier	35	High
	10/10	34	Low
DRD4 VNTR	7-repeat carrier	42	High
	All others	27	Low
COMT Val <sup>158</sup> Met	Met/Met	12	High
	Val/Met	41	Intermediate
	Val/Val	16	Low
DRD2 TaqIA	C/C	43	High
	C/T	23	Intermediate
	T/T	3	Low

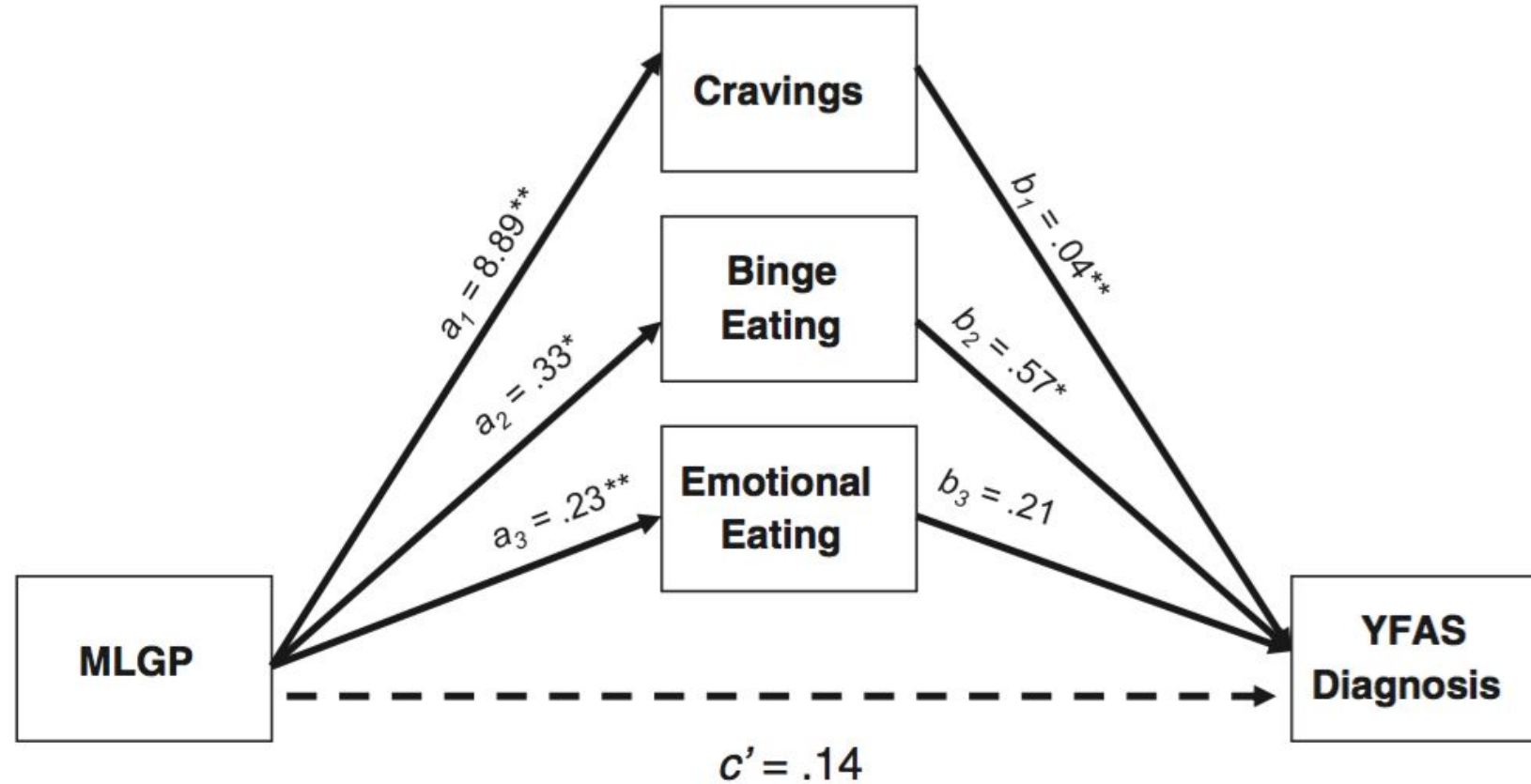
# Multilocus Genetic Profile Approach



# MLGP: Food addiction

Polymorphism	Genotypes	N	DA profile scores
DRD2 Taq1A	A1 +	54	Low
	A1 -	66	High
DRD2 -141C Ins/Del	Del+	39	High
	Ins/Ins	81	Low
DAT1 VNTR	9-repeat	56	High
	10/10	64	Low
COMT Val <sup>158</sup> Met	Met/Met	14	High
	Val/Met	58	Intermediate
	Val/Val	48	Low
C957T	C -	26	High
	C +	94	Low
rs12364283	T +	23	High
	T -	97	Low

# MLGP: Food addiction



Total Indirect effect (bias corrected 95% CI: .08–1.12; Nagelkerke  $R^2=0.58$ ; N=117)

# Few studies can focus on all GWAS loci

Gene Name	Polymorphism*	Model (Major allele > Minor Allele)	Cases vs. Controls (independent samples)	Fixed Effects OR (95% CI)	Random Effects OR (95% CI)	Heterogeneity p-Value	I-Square	Grade**
SLC4A7	rs3278	G > A	1410 vs. 906 (3)	2.34 (1.599-3.420)	2.28 (1.555-3.333)	0.51	0	B
DRD4	48-bp repeat	Other > 7/8 repeats	2324 vs. 1932 (6)	1.44 (1.155-1.804)	1.48 (1.000-2.197)	0.06	52	C
DRD2/ ANKK1***	Taq1A	A2 > A1	6312 vs. 7424 (20)	1.30 (1.192-1.410)	1.38 (1.096-1.733)	< 0.0001	84	C
BDNF	rs6265	G > A	2530 vs. 4126 (9)	1.31 (1.165-1.451)	1.38 (1.056-1.790)	< 0.0001	80	C
CCK	-45 C/T	C > T	860 vs. 2002 (6)	1.34 (1.089-1.650)	1.34 (1.083-1.646)	0.62	0	B
FAAH	rs324420	P > T	498 vs. 1570 (3)	1.38 (1.014-1.875)	1.32 (0.807-2.171)	0.24	28	B
OPRM1	rs1799971	A > G	2846 vs. 4072 (9)	1.24 (1.090-1.410)	1.31 (0.958-1.790)	< 0.0001	80	C
COMT	rs4680	Val > Met	862 vs. 1594 (3)	0.76 (0.634-0.923)	0.82 (0.644-1.051)	0.71	0	B
CNR1	(AAT)n	14 repeats > Other	2304 vs. 2144 (8)	0.76 (0.658-0.878)	0.75 (0.619-0.906)	0.17	32	B
HNMT	rs35953316	Thr > Ile	1540 vs. 1306 (3)	0.76 (0.598-0.975)	0.72 (0.444-1.179)	0.04	70	C
OPRK1	rs702764	A > G	292 vs. 246 (3)	0.62 (0.431-0.901)	0.62 (0.412-0.944)	0.99	0	B
OPRM1	C691G	C > G	796 vs. 786 (3)	0.52 (0.416-0.647)	0.61 (0.330-1.095)	0.0025	83	C

\*Variants were ranked based on the summary ORs. \*\*Degree of 'epidemiological credibility' based on published protocols (A, strong; B, modest; C, weak; see Methods for more details). \*\*\*Researchers previously associated the polymorphism Taq 1A to the *DRD2* gene. However, the polymorphism sits in an exon of the *ANKK1* gene.

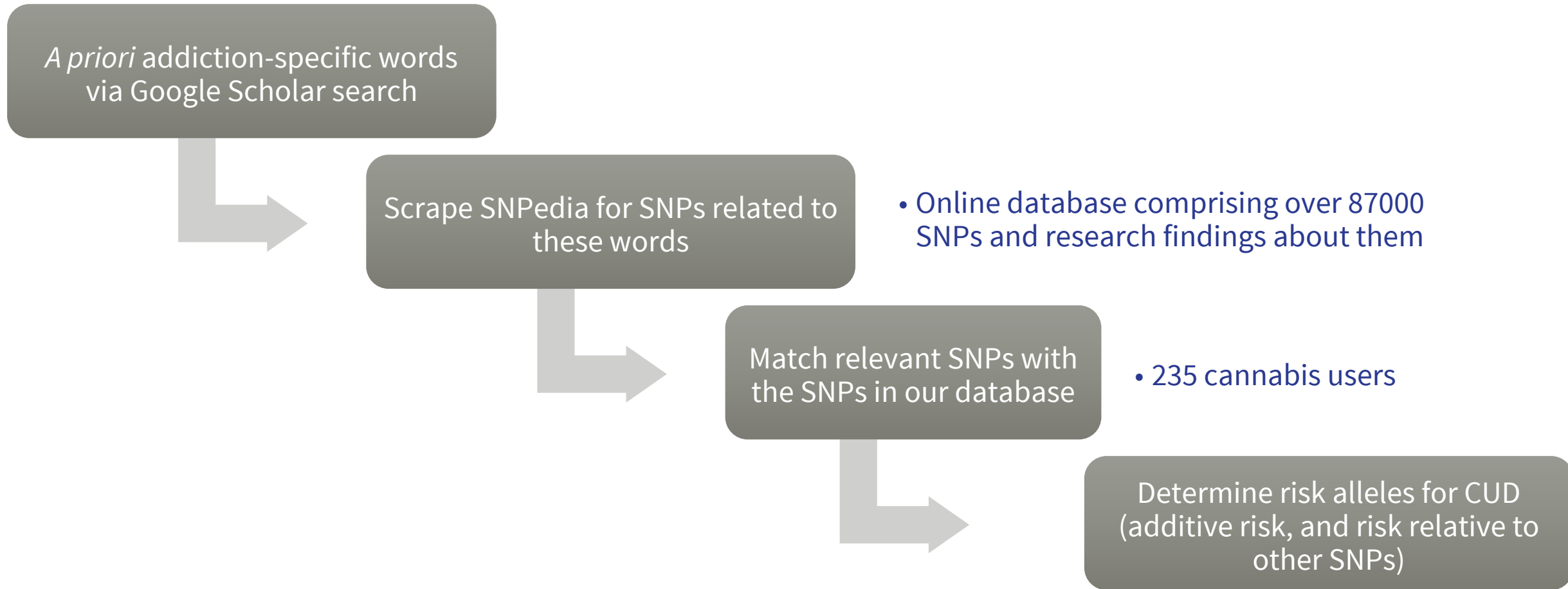


# GWAS Problems

- Large number of alleles can lead to p hacking issues
- Small minor allele frequency → small effect size
- Genotyping currently costs ~\$300

# Methods

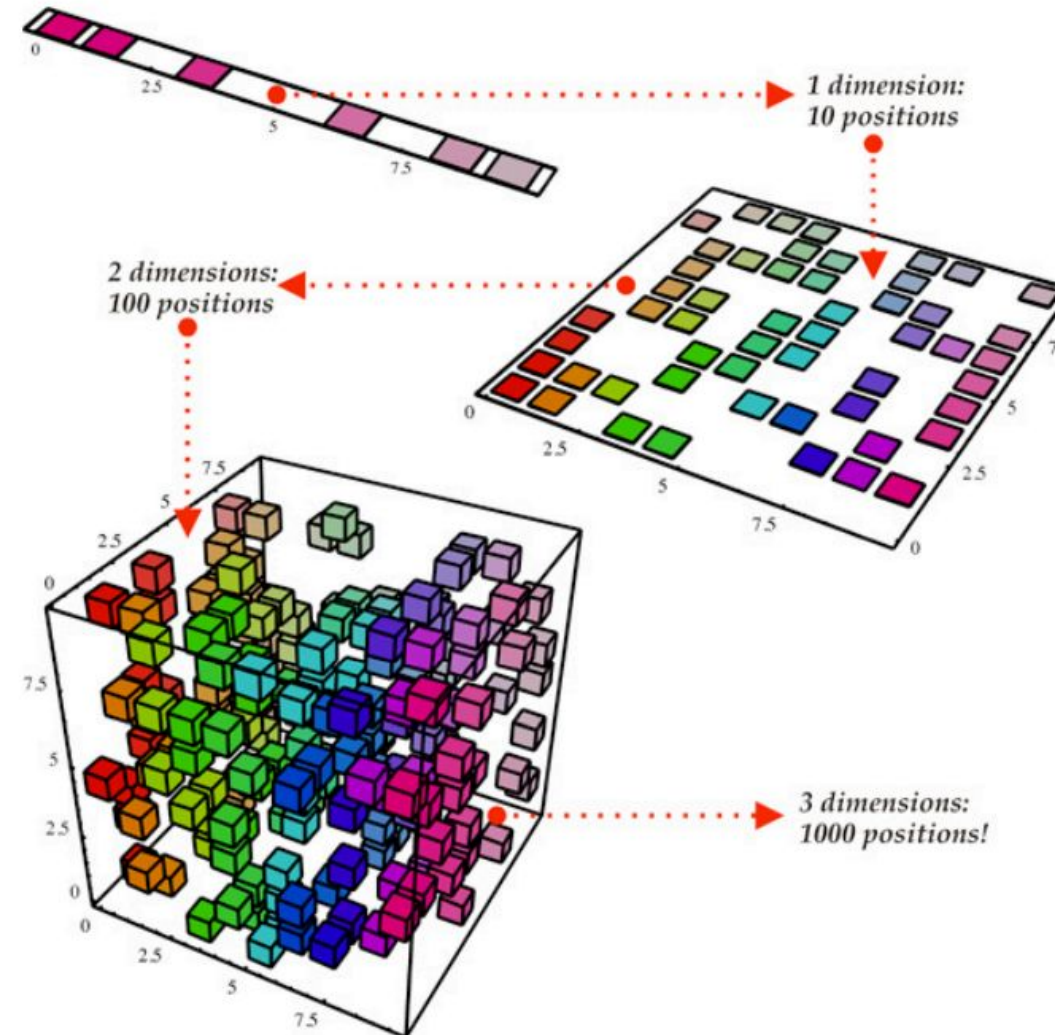
# Work Flow





# Strategic Dimension Reduction

- Previous genetics studies either
  - *Focus on one locus*
  - *Require a large sample size to account for variance in multi-locus models*
  - *In this study, we use the body of extant literature to strategically reduce the volume of data to combine the best of these two approaches*



# Tools: SNPedia



[Aあ](#) [English](#) [Talk](#) [Contributions](#) [Create account](#) [Log in](#)

 [→](#) [Q](#)

Navigation ▾

[Page](#) [Discussion](#) [Edit with form](#) [Edit](#) [History](#)

Have questions? Visit <https://www.reddit.com/r/SNPedia>

## rs2023239

[\[PMID 18705688\]](#) rs2023239 influences [cannabis dependence](#)

[\[PMID 20010552\]](#) Individual and additive effects of the CNR1 and FAAH genes on brain response to marijuana cues

[\[PMID 22669173\]](#) Associations between Cannabinoid Receptor-1 (CNR1) Variation and Hippocampus and Amygdala Volumes in Heavy Cannabis Users

[\[PMID 16741937\]](#) Association study of the CNR1 gene exon 3 alternative promoter region polymorphisms and substance dependence.

[\[PMID 17292652\]](#) No evidence for an involvement of variants in the cannabinoid receptor gene (CNR1) in obesity in German children and adolescents.

[\[PMID 17508995\]](#) Polymorphisms of the dopamine D4 receptor gene (DRD4 VNTR) and cannabinoid CB1 receptor gene (CNR1) are not strongly related to cue-reactivity after alcohol exposure.

[\[PMID 18179391\]](#) Association between single nucleotide polymorphisms in the cannabinoid receptor gene (CNR1) and impulsivity in southwest California Indians.

[\[PMID 18375449\]](#) Endocannabinoid receptor 1 gene variations increase risk for obesity and modulate body mass

Orientation	plus
Stabilized	plus
Make	<a href="#">rs2023239(C;C)</a>
Make	<a href="#">rs2023239(C;T)</a>
Make	<a href="#">rs2023239(T;T)</a>
Reference	<a href="#">GRCh38</a> 38.1/141
Chromosome	6
Position	88150763
Gene	<a href="#">CNR1</a>
is a	<a href="#">snp</a>
is	<a href="#">mentioned by</a>
dbSNP	<a href="#">rs2023239</a>
ebi	<a href="#">rs2023239</a>
HLI	<a href="#">rs2023239</a>
Exac	<a href="#">rs2023239</a>
Varsome	<a href="#">rs2023239</a>
Map	<a href="#">rs2023239</a>
PheGenl	<a href="#">rs2023239</a>
hapmap	<a href="#">rs2023239</a>
1000	<a href="#">rs2023239</a>
genomes	<a href="#">rs2023239</a>

# Tools: SNPedia

Have questions? Visit <https://www.reddit.com/r/SNPedia>

## rs2023239

- [PMID 18705688] rs2023239 influences cannabis dependence
- [PMID 20010552] Individual and additive effects of the CNR1 and FAAH genes on brain response to marijuana cues
- [PMID 22669173] Associations between Cannabinoid Receptor-1 (CNR1) Variation and Hippocampus and Amygdala Volumes in Heavy Cannabis Users
- [PMID 16741937] Association study of the CNR1 gene exon 3 alternative promoter region polymorphisms and substance dependence.
- [PMID 17292652] No evidence for an involvement of variants in the cannabinoid receptor gene (CNR1) in obesity in German children and adolescents.
- [PMID 17508995] Polymorphisms of the dopamine D4 receptor gene (DRD4 VNTR) and cannabinoid CB1 receptor gene (CNR1) are not strongly related to cue-reactivity after alcohol exposure.
- [PMID 18179391] Association between single nucleotide polymorphisms in the cannabinoid receptor gene (CNR1) and impulsivity in southwest California Indians.

Orientation	plus
Stabilized	plus
Make rs2023239(C;C)	
Make rs2023239(C;T)	
Make rs2023239(T;T)	
Reference	GRCh38 38.1/141
Chromosome	6
Position	88150763
Gene	CNR1
is a	snp
is mentioned by	
dbSNP	rs2023239
ebi	rs2023239
HLI	rs2023239
Exac	rs2023239
Varsome	rs2023239
Map	rs2023239
PheGenl	rs2023239
hapmap	rs2023239
1000	rs2023239
genomes	rs2023239

# Determining Risk Alleles

- Interesting alleles selected from behavioral keywords
- Alleles selected with knowledge of specific pathways (eg. Cannabinoid pathways)
- Use the Least Absolute Shrinkage and Selection Operator (LASSO) technique to determine most relevant alleles for determining risk



# Variable Selection Procedure - Lasso

- Let  $b_i$  represent the behaviour score of user  $i$
- $s_j$  is the number of minor alleles in SNP  $j$
- $\alpha_j$  is the weight attached to SNP - a measure of relevance
- We solve the Lasso problem to select a small number of non-zero relevant SNPs

$$\min_{\alpha} \sum_{i \in \text{users}} (b_i - \sum_{j \in \text{SNPs}} \alpha_j s_j)^2 + \lambda \sum_{j \in \text{SNPs}} |\alpha_j|$$

# Learning Interactions

- Interactions: SNP  $s_{j_1}, s_{j_2}$  may not have an effect individually.
- When minor alleles of both present, there may be more of an effect
- We consider a variant of Lasso where we consider the effects of these pairs and find sparse coefficients  $\alpha_{j,j'}$

$$\min_{\alpha} \sum_{i \in \text{users}} (b_i - \sum_{j, j' \in \text{SNPs}} \alpha_{j, j'} s_j s_{j'})^2 + \lambda \sum_{j, j' \in \text{SNPs}} |\alpha_{j, j'}|$$

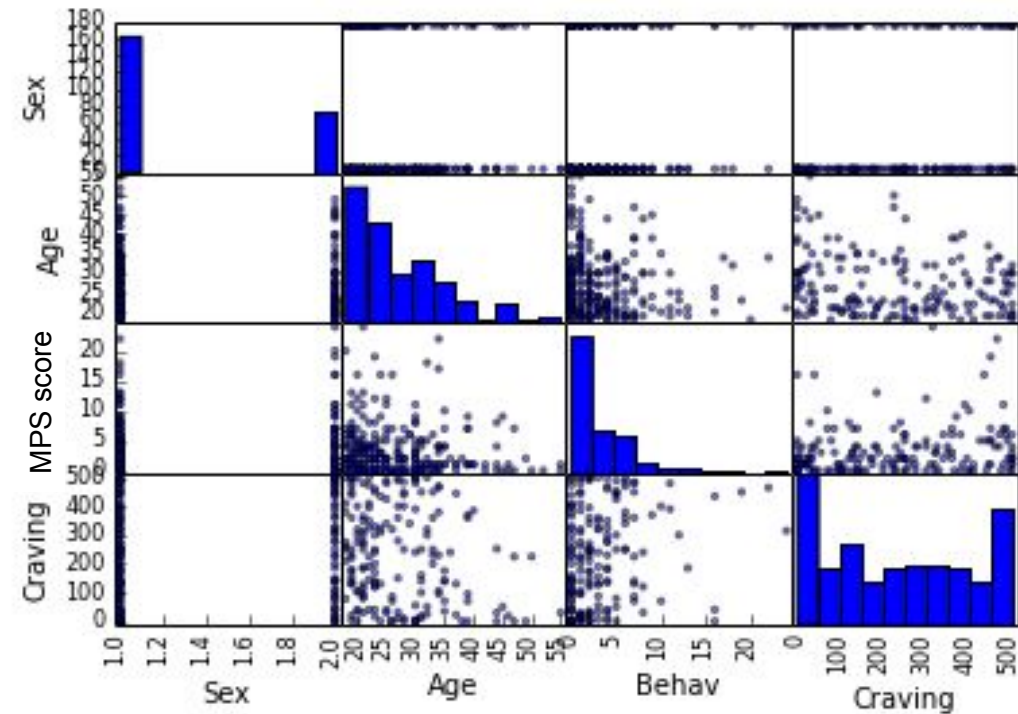
# Expected outcomes

- A solution to getting a signal from high dimensional data in small sample sizes
- These data will delineate genetic variability that can predict CUD

# Preliminary Results

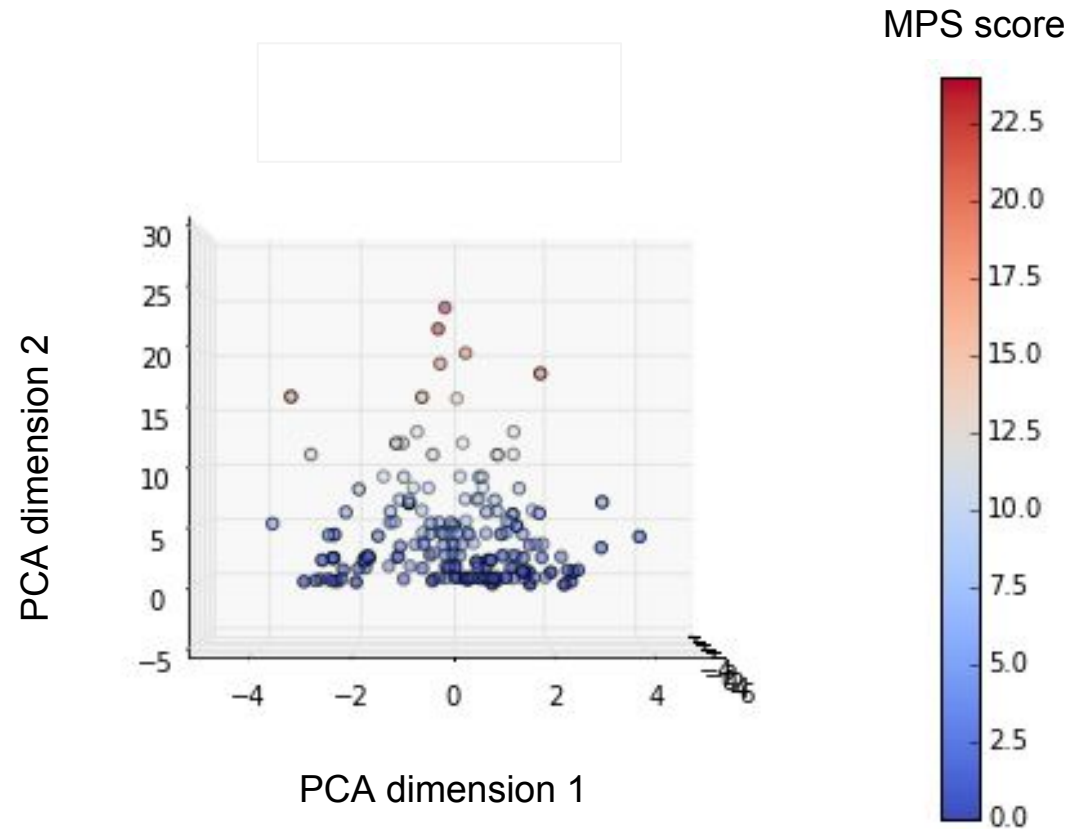
Distributions of variables of interest

- Greater number of males than females in sample
- Craving not correlated with other variables
- Small correlation between craving and MPS score

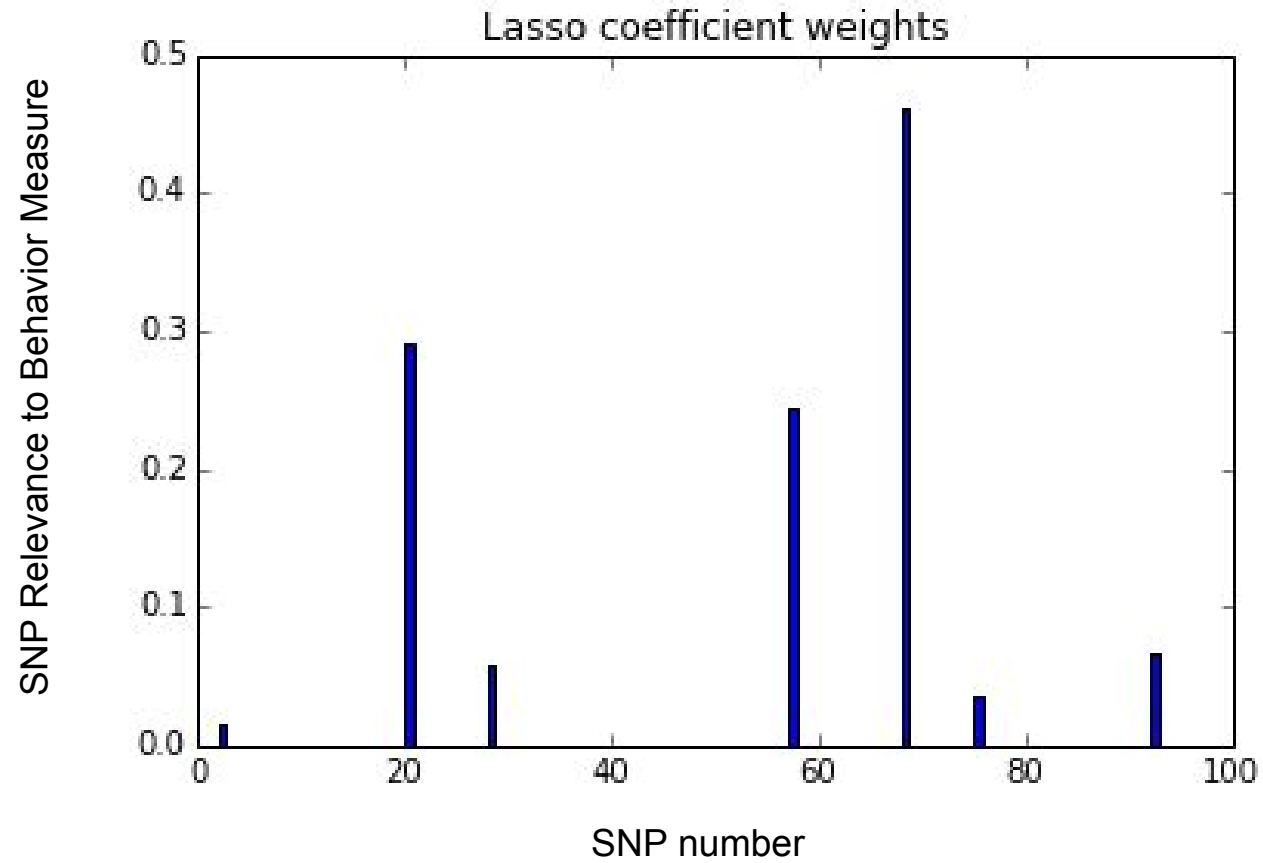


# Preliminary results: Principal component analysis

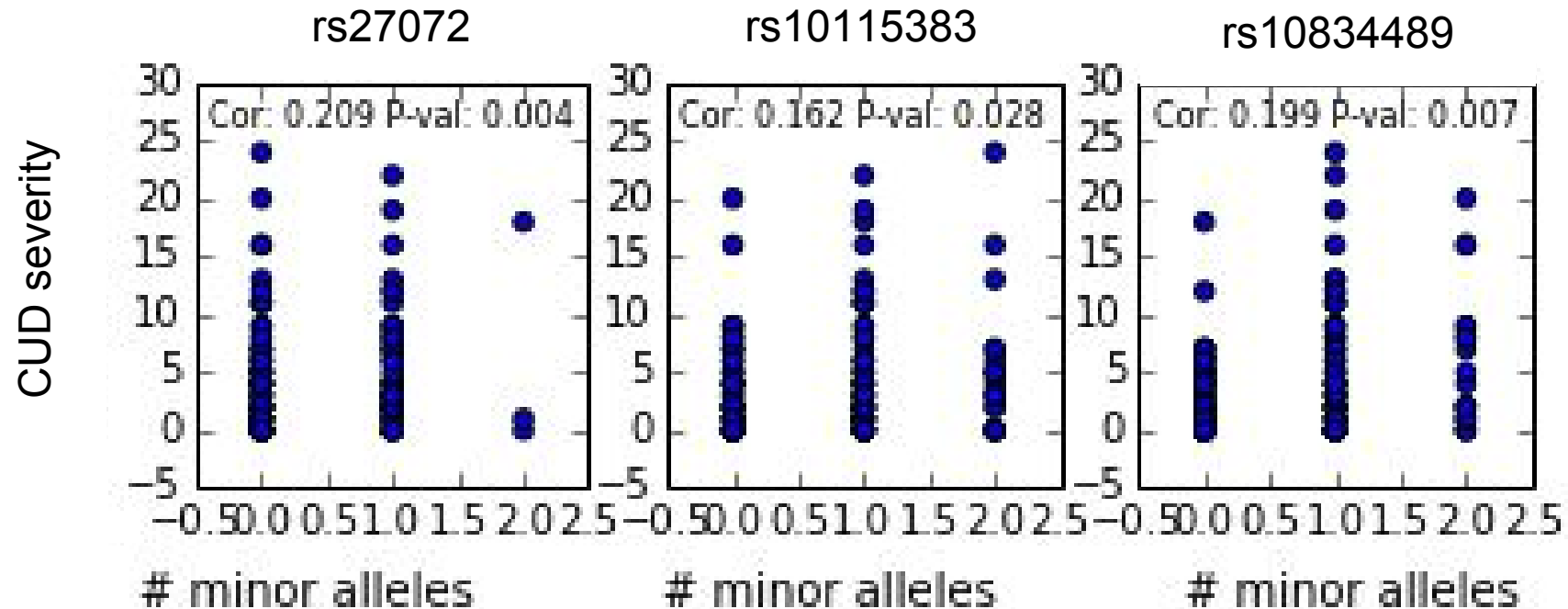
- PC1 - General variability in sample (nothing interesting)
- PC2 - Maps directly onto CUD severity



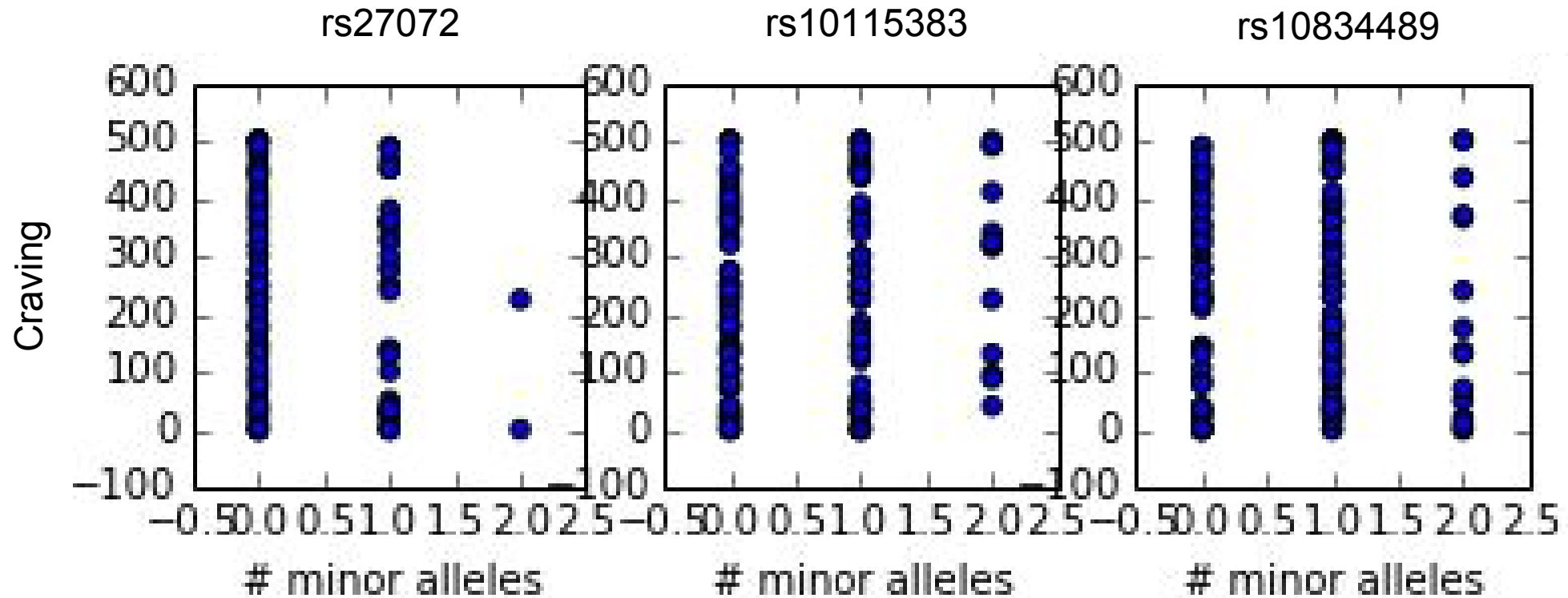
# Preliminary Results: variable selection



# Preliminary Results: CUD severity

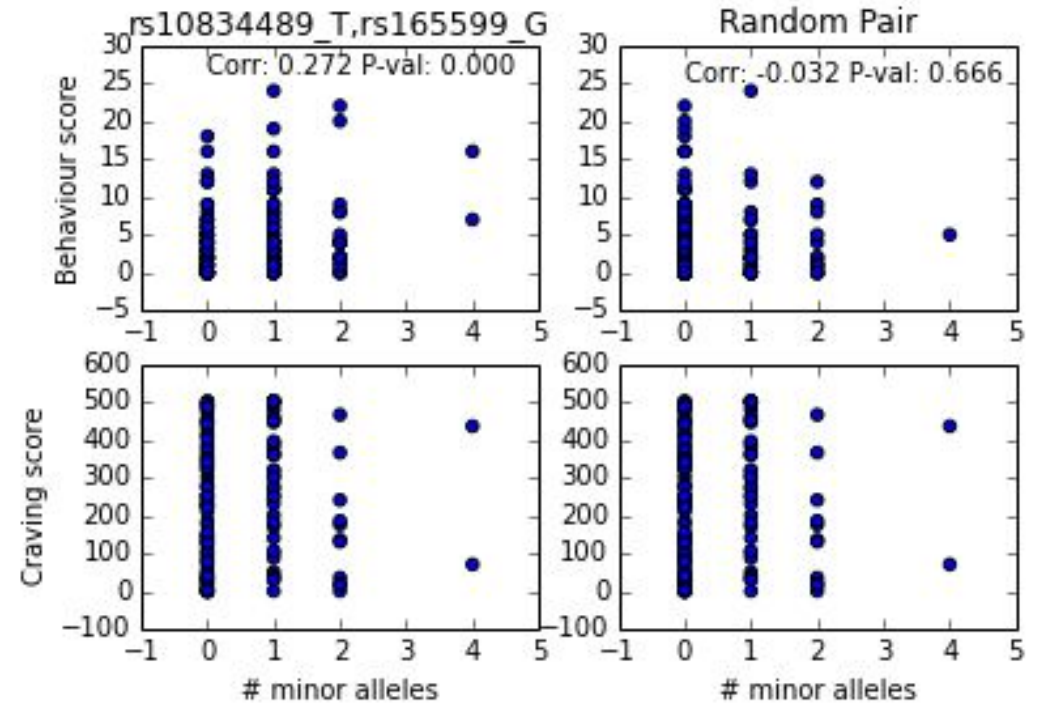
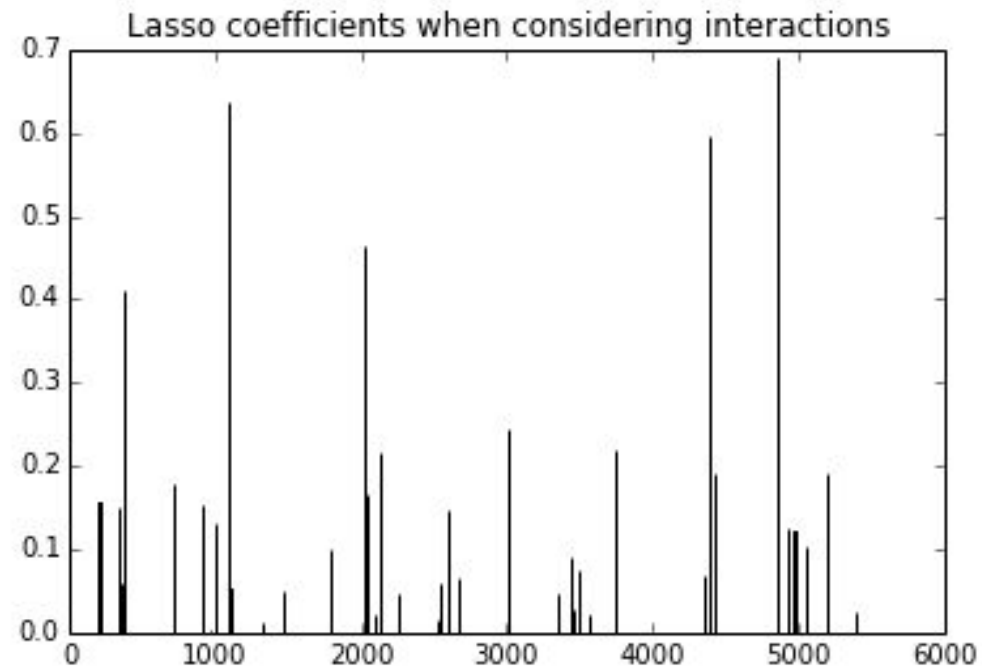


# Preliminary Results: subjective craving





# Preliminary Results - Learning Interactions



# Discussion

# rs27072

- In the dopamine transporter, which takes dopamine from the synapse after it's been released
- Risk allele (T) is the major allele
- Associated with
  - *alcohol withdrawal*
  - *ADHD in children*
  - *Genotype by age of onset interaction in tobacco and alcohol use in teenagers*

# rs165599

- In the Catechol *O*-Methyltransferase (COMT) gene
- COMT degrades dopamine
- G is both the risk allele and the minor allele
- Associated with
  - *anxiety* (Stein et al., 2005; Wray et al., 2008; Hettema et al., 2008)
  - *impulsivity*

# Conclusions

- SNPs related to psychological constructs that are strongly associated with addiction
  - *Negative emotions*
  - *Impulsivity and attention deficits*
- Unknown what degrees different SNPS contribute to severity
- Interactions between SNPS also undetermined
- Challenging to study due to requirement of large sample sizes
- These analyses will
  - *Identify individuals at risk for CUD*
  - *Provide potential solution for smaller studies*

# Future directions

- Determine the most appropriate method of adjusting for multiple comparisons
- Explore how changing parameters in the model changes weights of SNPs and their clinical relevance
- Include SNPs that are susceptible to environmental changes

# Acknowledgements



- The National Science Foundation
- The Center for Science of Information, Purdue University
  - *Brent Ladd*
- The University of Texas at Dallas
  - *Francesca Filbey, PhD*
- Stanford University
  - *Andrea Goldsmith, PhD*



**Stanford**  
**University**