Polygenic risk scores for substance use disorders: Ready for Prime Time?

Rachel L. Kember

Neuroscience Research Training Program, Yale

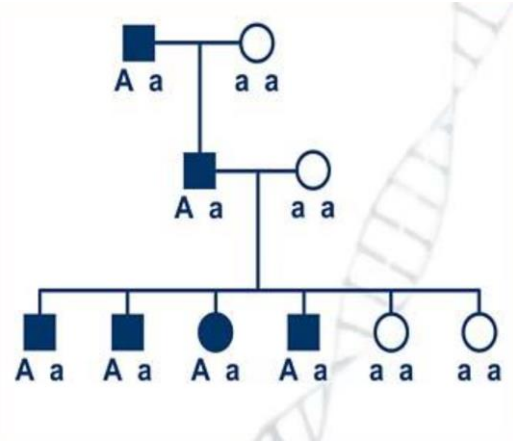
2020

Outline

- History of psychiatric genetics
- GWAS studies
- Biobanks and electronic health records
- Phenome-wide association studies
- Polygenic risk scores
- Conclusions
- Next steps

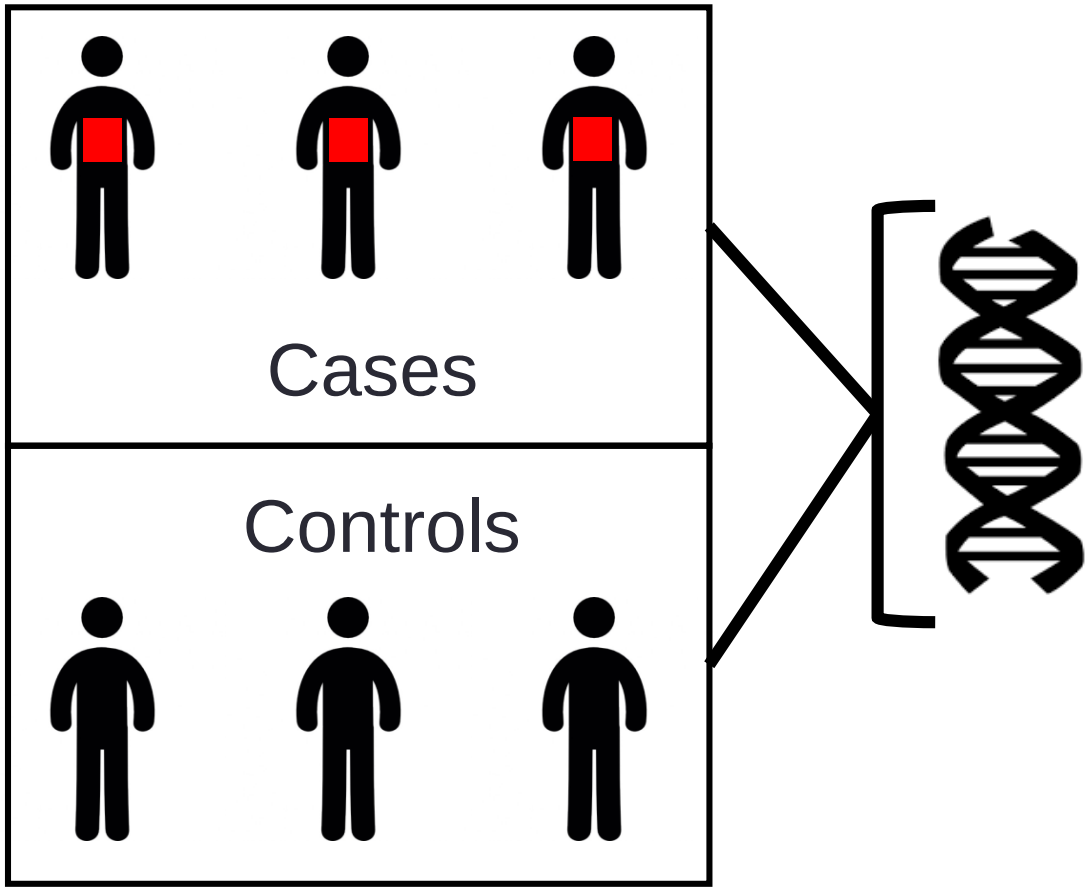
A (brief) history of psychiatric genetics

- Definition: The study of the role of genetics in psychiatric conditions
- Linkage studies
- Candidate gene studies
- Single variant association studies with ‘small’ numbers of cases/controls
- Endophenotypes
- Gene x Environment interactions
- **Small number of associations found**
- **Variants not consistently replicated**



Recent developments - GWAS

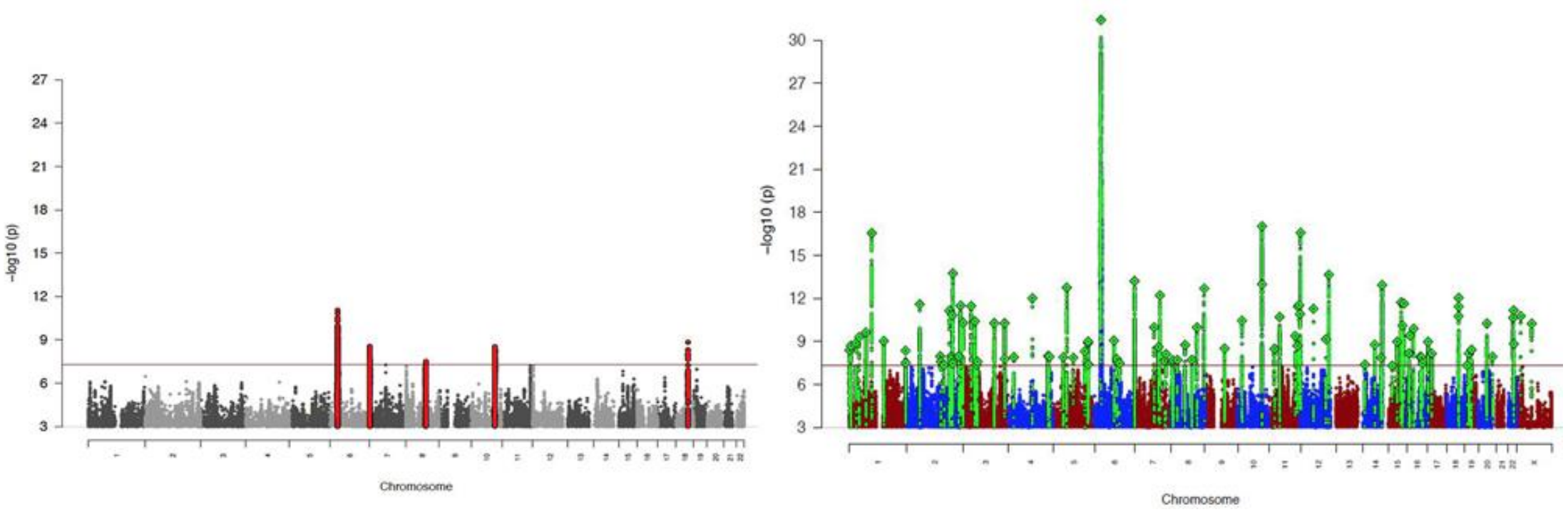
- First GWAS (2005) followed on from the first full human genome sequence (2003)
- Measures multiple common genetic variants and tests each against the phenotype (hypothesis-free)
- Is this variant more or less common in cases compared to controls?



GWAS

Recent developments – large consortia

- Consortia such as the Psychiatric Genomics Consortium have been formed to pool samples from smaller studies
- Larger sample sizes have yielded a greater number of significant results
- Increasing numbers of genetic variants now replicated across studies



Co-morbidities of psychiatric and substance use disorders

Psychiatric

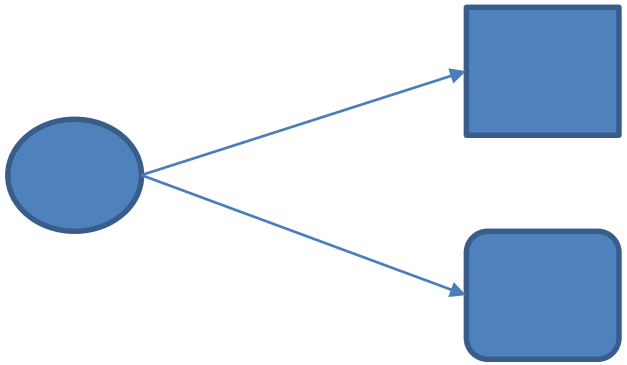
Anxiety disorders
Alcohol and substance use
Attention deficit/hyperactivity disorders
Post traumatic stress disorder
Personality disorder
Eating disorder

Medical

Asthma
Cardiovascular
Hyperlipidemia
Type 2 diabetes
Epilepsy
Thyroid disease
Migraine
Obesity

Pleiotropy

- Variants that affect multiple, unrelated, phenotypes



Horizontal

Vertical

- Phenome-wide association analysis

Solution – large sample sizes, multiple phenotypes (and compromise…….)

BioBanks and Electronic health records

- Commercial and academic BioBanks consisting of hundreds of thousands of samples

Penn Medicine BioBank

Million Veteran Program

UK BioBank

Data used for genetic studies

- Aim: to establish a set of cases and controls for genetic analysis
- Two main methods I will discuss:
  - Use of Phecodes for Phenome-wide association studies
  - Development of EHR derived phenotypes using “domain knowledge”

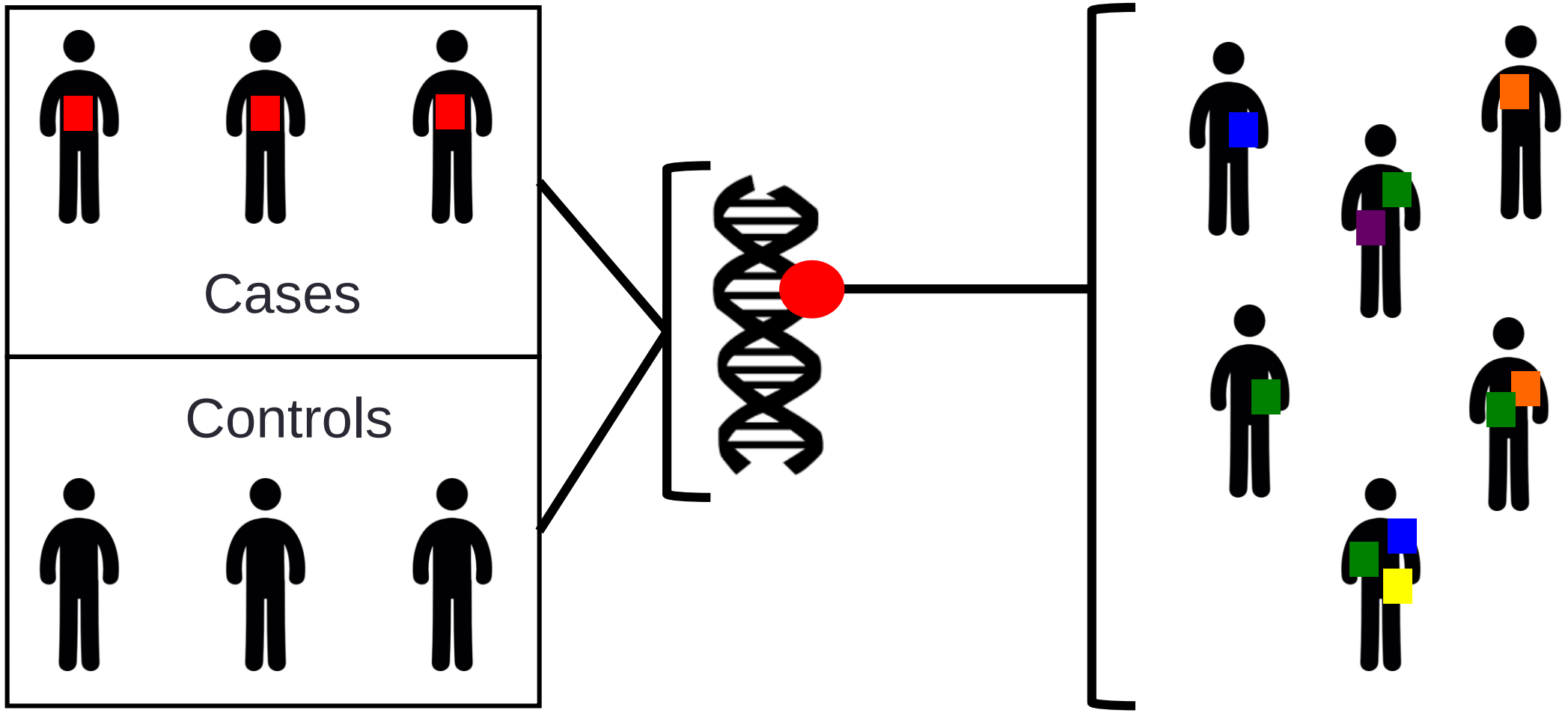
Phenotypes – International Classification of Diseases (ICD) codes

- Each disease has an ICD code
  - Currently ICD9 or ICD10

- Assigned by a physician when evaluating patient
  - In the USA, often used for charging purposes (which can lead to issues)
- “Lifetime Diagnosis”

- Autism: ICD9 - 299.0, ICD10 – F84.0
- Major Depression: ICD9 - 296.3, ICD10 – F33.1
- Cardiovascular disease: ICD9 - 429.2, ICD10 – I51.9

Phenome-wide association studies



Phenotypes: PheCode Mapping

Subj_id	UCSCID	CODE	CODE_DESCR	Freq
14158	PBB12ke4	309	Adjustment disorder with depressed mood	2
14158	PBB12ke4	309.28	Adjustment disorder with mixed anxiety and depressed mood	2
14158	PBB12ke4	F43.23	Adjustment disorder with mixed anxiety and depressed mood	3
14158	PBB12ke4	477.9	Allergic rhinitis, cause unspecified	2
14158	PBB12ke4	V15.06	Allergy to insects and arachnids	2
14158	PBB12ke4	300	Anxiety state, unspecified	8
14158	PBB12ke4	493.9	Asthma, unspecified type, unspecified	12
14158	PBB12ke4	493.92	Asthma, unspecified type, with (acute) exacerbation	3
14158	PBB12ke4	211.3	Benign neoplasm of colon	4
14158	PBB12ke4	211.4	Benign neoplasm of rectum and anal canal	3
14158	PBB12ke4	226	Benign neoplasm of thyroid glands	4
14158	PBB12ke4	373.2	Chalazion	2
14158	PBB12ke4	K59.00	Constipation, unspecified	5
14158	PBB12ke4	564	Constipation, unspecified	6
14158	PBB12ke4	786.2	Cough	3
14158	PBB12ke4	V65.42	Counseling on substance use and abuse	2
14158	PBB12ke4	729.82	Cramp of limb	2
14158	PBB12ke4	787.91	Diarrhea	6
14158	PBB12ke4	722	Displacement of cervical intervertebral disc without myelopathy	2
14158	PBB12ke4	Z09	Encounter for follow-up examination after completed treatment for conditions other than malignar	2
14158	PBB12ke4	530.81	Esophageal reflux	25
14158	PBB12ke4	I10	Essential (primary) hypertension	3
14158	PBB12ke4	V67.51	Follow-up examination, following completed treatment with high-risk medication, not elsewhere cl	7
14158	PBB12ke4	787.6	Full incontinence of feces	6
14158	PBB12ke4	K21.9	Gastro-esophageal reflux disease without esophagitis	3
14158	PBB12ke4	306.4	Gastrointestinal malfunction arising from mental factors	3
14158	PBB12ke4	240.9	Goiter, unspecified	2
14158	PBB12ke4	784	Headache	2
14158	PBB12ke4	E78.5	Hyperlipidemia, unspecified	2
14158	PBB12ke4	E87.6	Hypokalemia	2
14158	PBB12ke4	276.8	Hypopotassemia	7
14158	PBB12ke4	780.52	Insomnia, unspecified	6

Phenotypes: PheCode Mapping

Subj_id	UCSCID	CODE	CODE_DESCR	Freq
14158	PBB12ke4	309	Adjustment disorder with depressed mood	2
14158	PBB12ke4	309.28	Adjustment disorder with mixed anxiety and depressed mood	2
14158	PBB12ke4	F43.23	Adjustment disorder with mixed anxiety and depressed mood	3
14158	PBB12ke4	477.9	Allergic rhinitis, cause unspecified	2
14158	PBB12ke4	V15.06	Allergy to insects and arachnids	2
14158	PBB12ke4	300	Anxiety state, unspecified	8
14158	PBB12ke4	493.9	Asthma, unspecified type, unspecified	12
14158	PBB12ke4	493.92	Asthma, unspecified type, with (acute) exacerbation	3
14158	PBB12ke4	211.3	Benign neoplasm of colon	4
14158	PBB12ke4	211.4	Benign neoplasm of rectum and anal canal	3
14158	PBB12ke4	226	Benign neoplasm of thyroid glands	4
14158	PBB12ke4	373.2	Chalazion	2
14158	PBB12ke4	K59.00	Constipation, unspecified	5
14158	PBB12ke4	564	Constipation, unspecified	6
14158	PBB12ke4	786.2	Cough	3
14158	PBB12ke4	V65.42	Counseling on substance use and abuse	2
14158	PBB12ke4	729.82	Cramp of limb	2
14158	PBB12ke4	787.91	Diarrhea	6
14158	PBB12ke4	722	Displacement of cervical intervertebral disc without myelopathy	2
14158	PBB12ke4	Z09	Encounter for follow-up examination after completed treatment for conditions other than malignar	2
14158	PBB12ke4	530.81	Esophageal reflux	25
14158	PBB12ke4	I10	Essential (primary) hypertension	3
14158	PBB12ke4	V67.51	Follow-up examination, following completed treatment with high-risk medication, not elsewhere cl	7
14158	PBB12ke4	787.6	Full incontinence of feces	6
14158	PBB12ke4	K21.9	Gastro-esophageal reflux disease without esophagitis	3
14158	PBB12ke4	306.4	Gastrointestinal malfunction arising from mental factors	3
14158	PBB12ke4	240.9	Goiter, unspecified	2
14158	PBB12ke4	784	Headache	2
14158	PBB12ke4	E78.5	Hyperlipidemia, unspecified	2
14158	PBB12ke4	E87.6	Hypokalemia	2
14158	PBB12ke4	276.8	Hypopotassemia	7
14158	PBB12ke4	780.52	Insomnia, unspecified	6

Phenotypes: PheCode Mapping

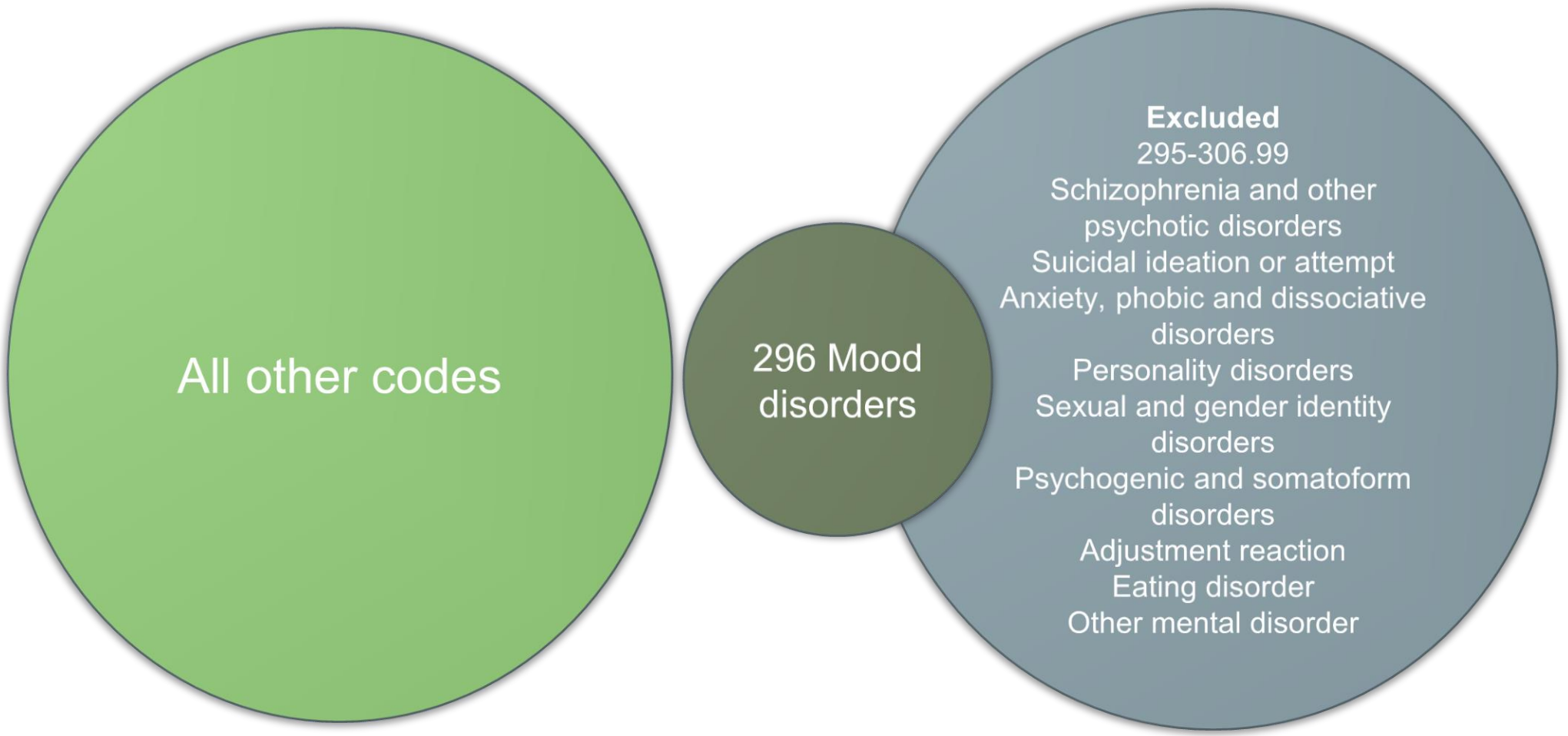
PheWAS Resources

PheCode Map with ICD-9 Codes

ICD9	ICD9 String	PheCode	Phenotype
309	description	code	phenotype
309	Adjustment reaction	304	Adjustment reaction
309.0	Adjustment disorder with depressed mood	304	Adjustment reaction
309.1	Adjustment reaction with prolonged depressive reaction	304	Adjustment reaction
309.2	Adjustment reaction with predominant disturbance of other emotions	304	Adjustment reaction
309.21	Separation anxiety disorder	313	Pervasive developmental disorders
309.22	Emancipation disorder of adolescence and early adult life	304	Adjustment reaction
309.23	Specific academic or work inhibition	304	Adjustment reaction
309.24	Adjustment disorder with anxiety	304	Adjustment reaction
309.28	Adjustment disorder with mixed anxiety and depressed mood	304	Adjustment reaction
309.29	Other adjustment reactions with predominant disturbance of other emotions	304	Adjustment reaction
309.3	Adjustment disorder with disturbance of conduct	304	Adjustment reaction
309.4	Adjustment disorder with mixed disturbance of emotions and conduct	304	Adjustment reaction
309.8	Other specified adjustment reactions	304	Adjustment reaction
309.81	Posttraumatic stress disorder	300.9	Posttraumatic stress disorder
309.82	Adjustment reaction with physical symptoms	304	Adjustment reaction
309.83	Adjustment reaction with withdrawal	304	Adjustment reaction
309.89	Other specified adjustment reactions	304	Adjustment reaction
309.9	Unspecified adjustment reaction	304	Adjustment reaction

Denny et al., Nature Biotechnology, 2013

Phenome-wide association studies



Adapted from Bush et al., 2016

EHR derived phenotypes

- Use of “domain knowledge”
  - (i.e. a physician who understands the disease or trait phenotype)

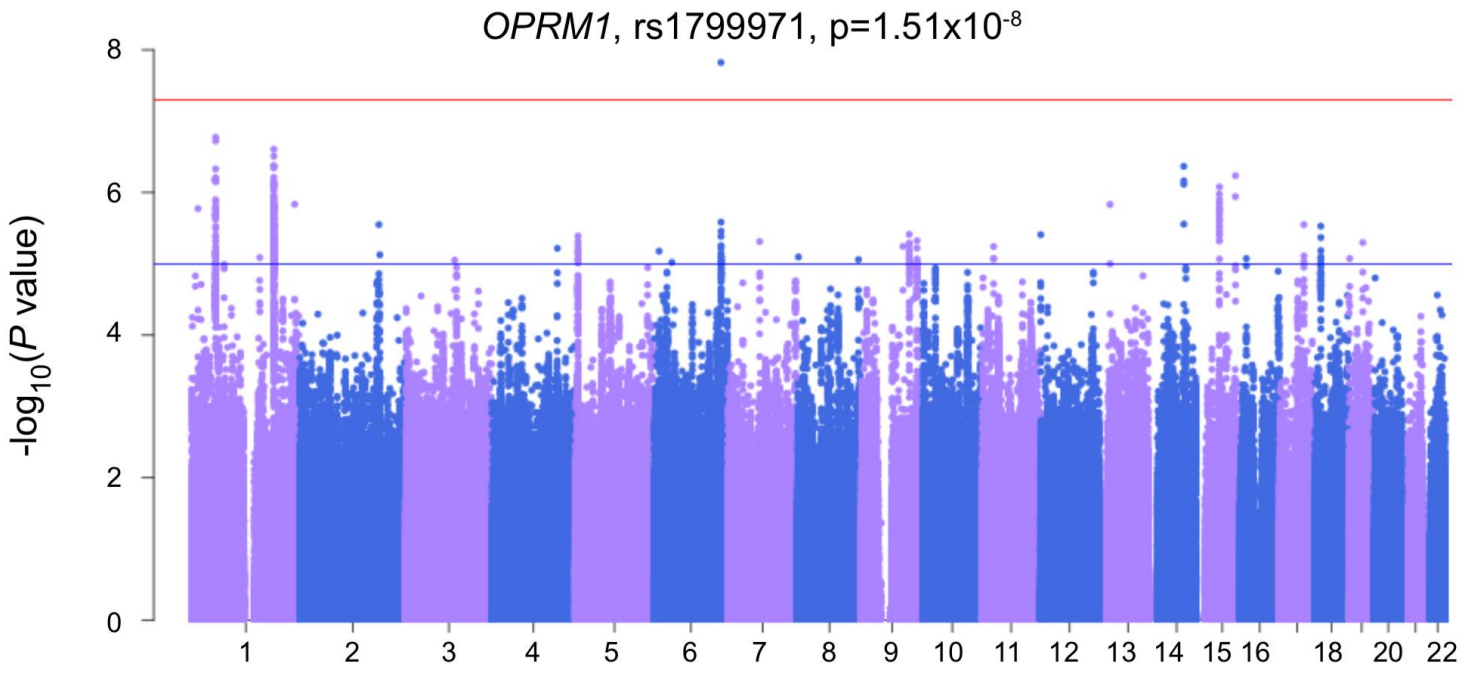
- Uses defined groups of ICD codes, possibly with the addition of other EHR data

- Example:
  - 2 ICD-9 or ICD-10 codes for bipolar disorder on separate days as outpatient
  - 1 ICD-9 or ICD-10 code for bipolar disorder as inpatient
  - Plus: Medication prescribed for bipolar disorder

Genetic liability for Opioid Use Disorder

OUD GWAS

- GWAS of opioid use disorder in MVP, Yale-Penn and SAGE studies
- 10,544 OUD cases and 72,163 opioid-exposed controls
- Single SNP reached genome-wide significance
- OR=1.07

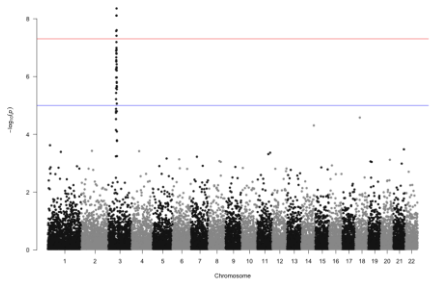


Zhou et al., https://doi.org/10.1101/19007039

Polygenic Risk Scores

- Not all of the heritability is explained by the significant GWAS SNP
- SNPs that are non-significant contain real signal
  - Why are they not significant?
  - Very small effect sizes, stringent multiple-testing correction
- What if we want to predict the phenotype in a different sample?
  - Calculate polygenic risk scores!

Polygenic Risk Scores



β_A=0.02 ⟵ AA 2x.02 + AG 1x.02 +

β_G =-0.04 ⟵ GT 1x-.04 + TT 0x-.04 +

β_C =-0.05 ⟵ CG 1x-.05 + CC 2x-.05 +

β_T =0.09 ⟵ AT 1x.09 + AA 1x.09

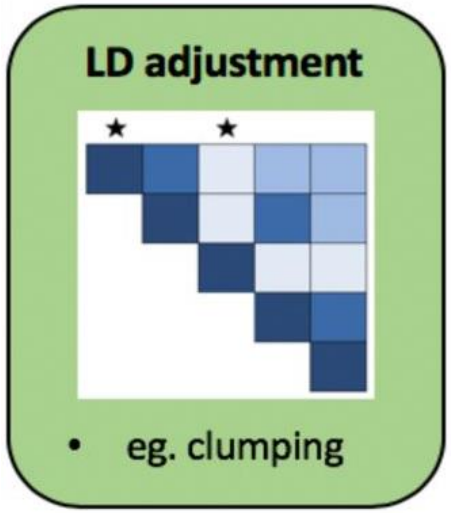
PRS = 0.04 PRS = 0.01

Penn Medicine BioBank (PMBB)

- Provides researchers with centralized access to a large number of blood and tissue samples with attached health information
- Facility banks blood specimens (i.e., whole blood, plasma, serum, buffy coat, and DNA isolated from leukocytes) and tissues (i.e.,formalin-fixed paraffin embedded, fresh and flash frozen)
- ~ 60,000 individuals
- Multiple ancestries

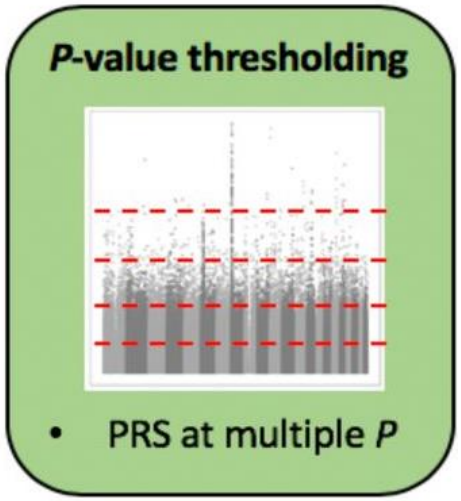
PRS methods

- Used summary statistics provided by Hang Zhou from OUD meta-analysis (Zhou et al., https://doi.org/10.1101/19007039)
- Used clumping/thresholding with a number of p-value cut offs (9 scores)



Clumps SNPs into LD blocks (SNPs that 'travel together')

Selects the most significant SNP from that LD block based on p-value



Thresholds: p ≤ 0.000001, 0.00001, 0.0001, 0.001, 0.01, 0.05, 0.1, 0.5, 1

Choi et al., BioRxiv

Determining the best PRS

- To determine best PRS, tested for association of PRS with OUD phenotype
- OUD phenotype determined by ICD9 and 10 codes (summary table from Zhou et al.)
- ICD9 and 10 codes restricted to subset of encounters that represent encounters with a physician
- In 52,354 PMBB individuals, 566 have at least 1 code for OUD
- In 10,182 EUR individuals with genetic data, 85 have at least 1 code for OUD (64 males, 21 females, mean age=62.2)
- Logistic regression model to test for association between PRS and OUD phenotype, with age, sex and PCs 1-10 as covariates

Determining the best PRS

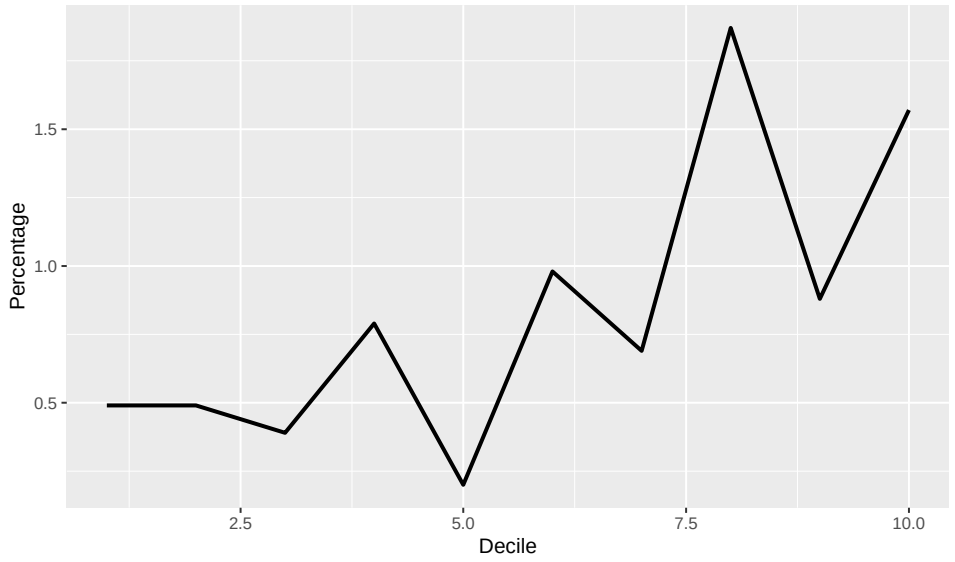
PRS method	Parameter	OR (95% CI)	P	AUC
Clumping/thresholding	$p<1x10^{-6}$	0.84 (0.68-1.04)	0.1178	0.69
	$p<1x10^{-5}$	0.99 (0.80-1.23)	0.9212	0.687
	$p<1x10^{-4}$	1.20 (0.96-1.48)	0.1032	0.6911
	$p<1x10^{-3}$	1.19 (0.95-1.47)	0.1233	0.6903
	p<0.01	1.38 (1.11-1.72)	0.0032	0.708
	p<0.05	1.55 (1.25-1.92)	$7.49x10^{-5}$	0.7222
	p<0.1	1.52 (1.22-1.89)	0.0002	0.719
	p<0.5	1.51 (1.22-1.88)	0.0002	0.7149
	p<1	1.50 (1.20-1.86)	0.0003	0.7143

Determining the best PRS

PRS method	Parameter	OR (95% CI)	P	AUC
Clumping/thresholding	$p<1x10^{-6}$	0.84 (0.68-1.04)	0.1178	0.69
	$p<1x10^{-5}$	0.99 (0.80-1.23)	0.9212	0.687
	$p<1x10^{-4}$	1.20 (0.96-1.48)	0.1032	0.6911
	$p<1x10^{-3}$	1.19 (0.95-1.47)	0.1233	0.6903
	p<0.01	1.38 (1.11-1.72)	0.0032	0.708
	p<0.05	1.55 (1.25-1.92)	$7.49x10^{-5}$	0.7222
	p<0.1	1.52 (1.22-1.89)	0.0002	0.719
	p<0.5	1.51 (1.22-1.88)	0.0002	0.7149
	p<1	1.50 (1.20-1.86)	0.0003	0.7143

Case prevalence

- Split PRS into deciles
- Calculated case prevalence per decile
- Compared top 10% of PRS to rest (90%): OR=2.05 (1.17-3.57), p=0.012

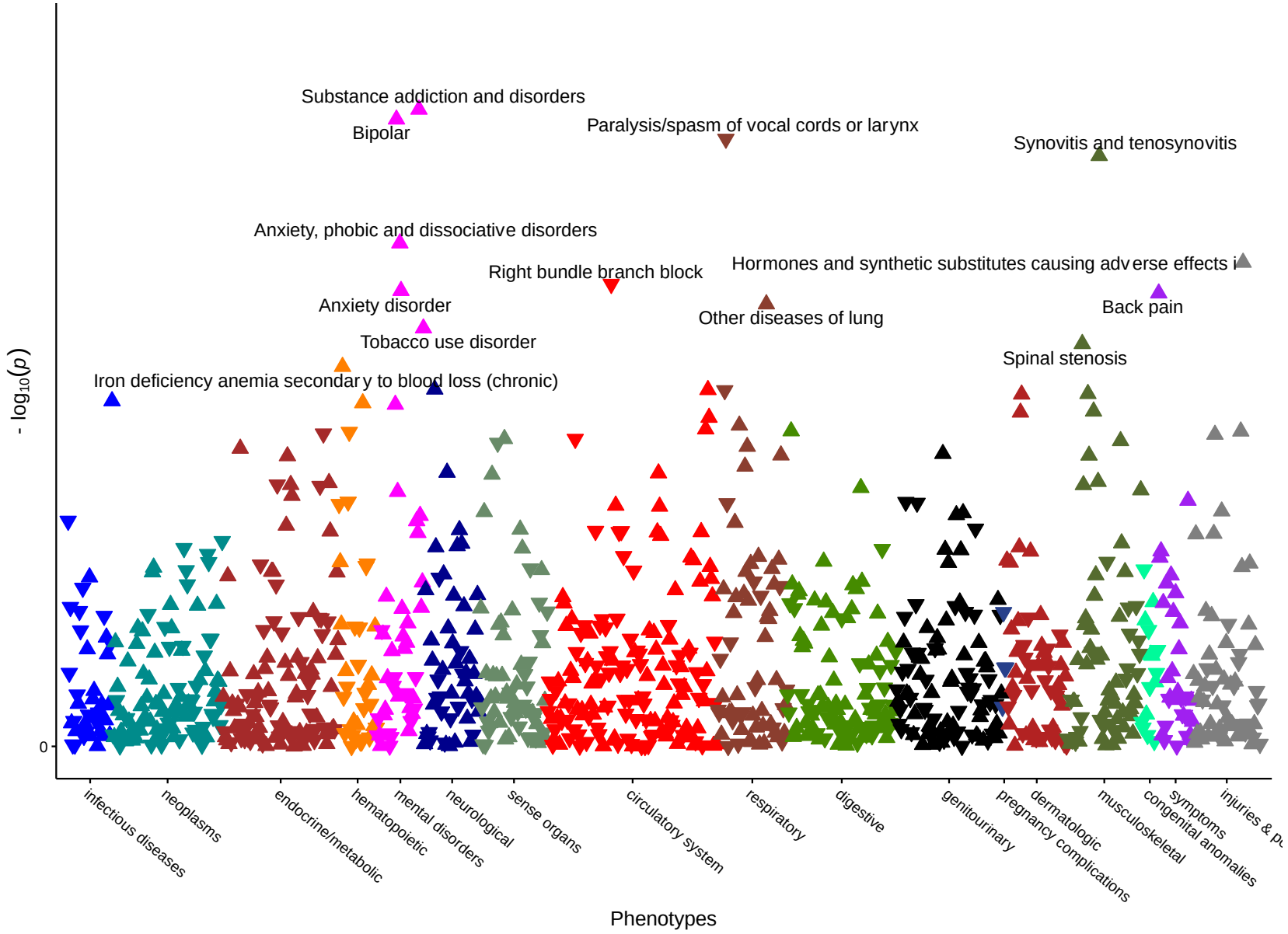


Decile	# cases	Percentage
1	5	0.49
2	5	0.49
3	4	0.39
4	8	0.79
5	2	0.20
6	10	0.98
7	7	0.69
8	19	1.87
9	9	0.88
10	16	1.57

Clinical relevance

- Current clinical relevance is limited
- The PRS is associated with the phenotype, but is not predictive in a naïve patient – we can't use this to label individuals as 'cases' and 'controls'
- As GWAS sample sizes increase, we expect that genetic variants associated with disease will be identified with more accurate effect sizes, allowing us to create PRS that are more accurate
- Meanwhile, we can explore associations with other phenotypes to help us understand more about the disorder

PheWAS of PRS



Genetic liability for Alcohol Consumption

Million Veteran Program

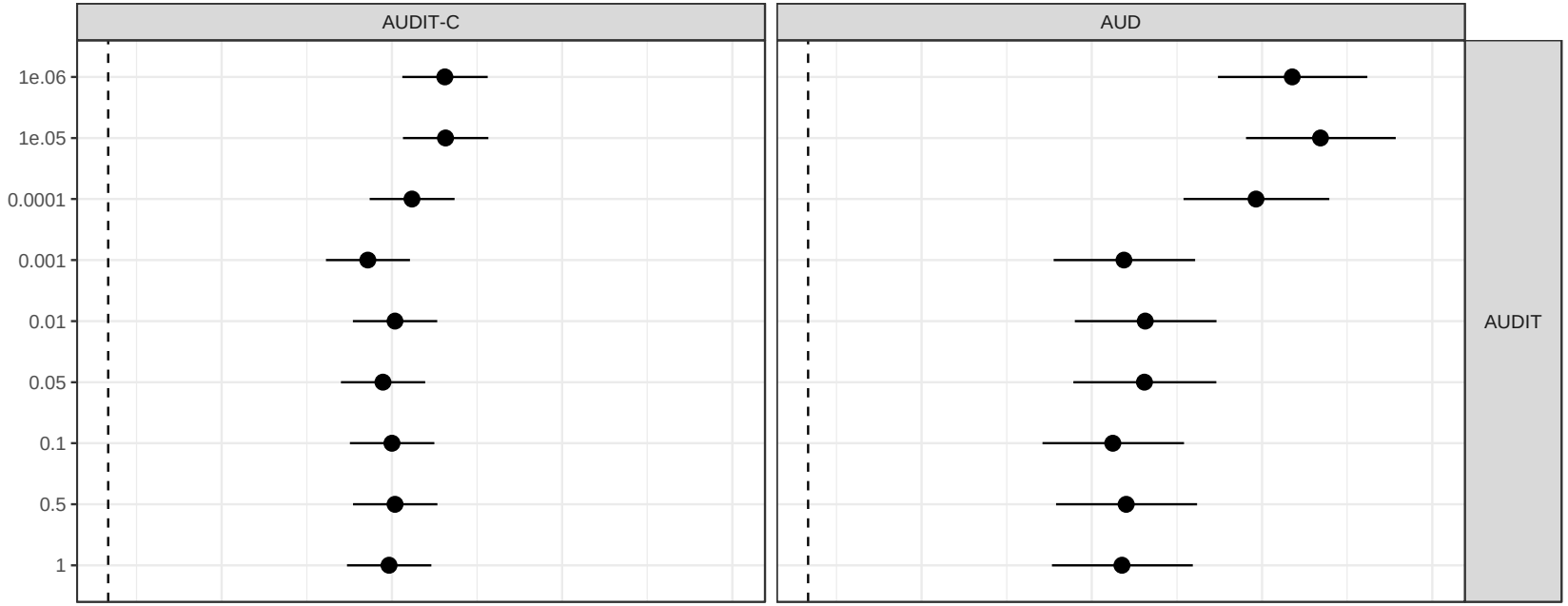
- Single cohort
- Large sample size (300,000 for current analysis, 600,000 for next analysis)
- Multiple ancestries
- Longitudinal repeated measures from EHR
- Alcohol consumption measure and AUD diagnosis
  - Age adjusted mean Alcohol Use Disorders Identification Test-Consumption (AUDIT-C)
  - AUD by ICD 9/10

PRS methods

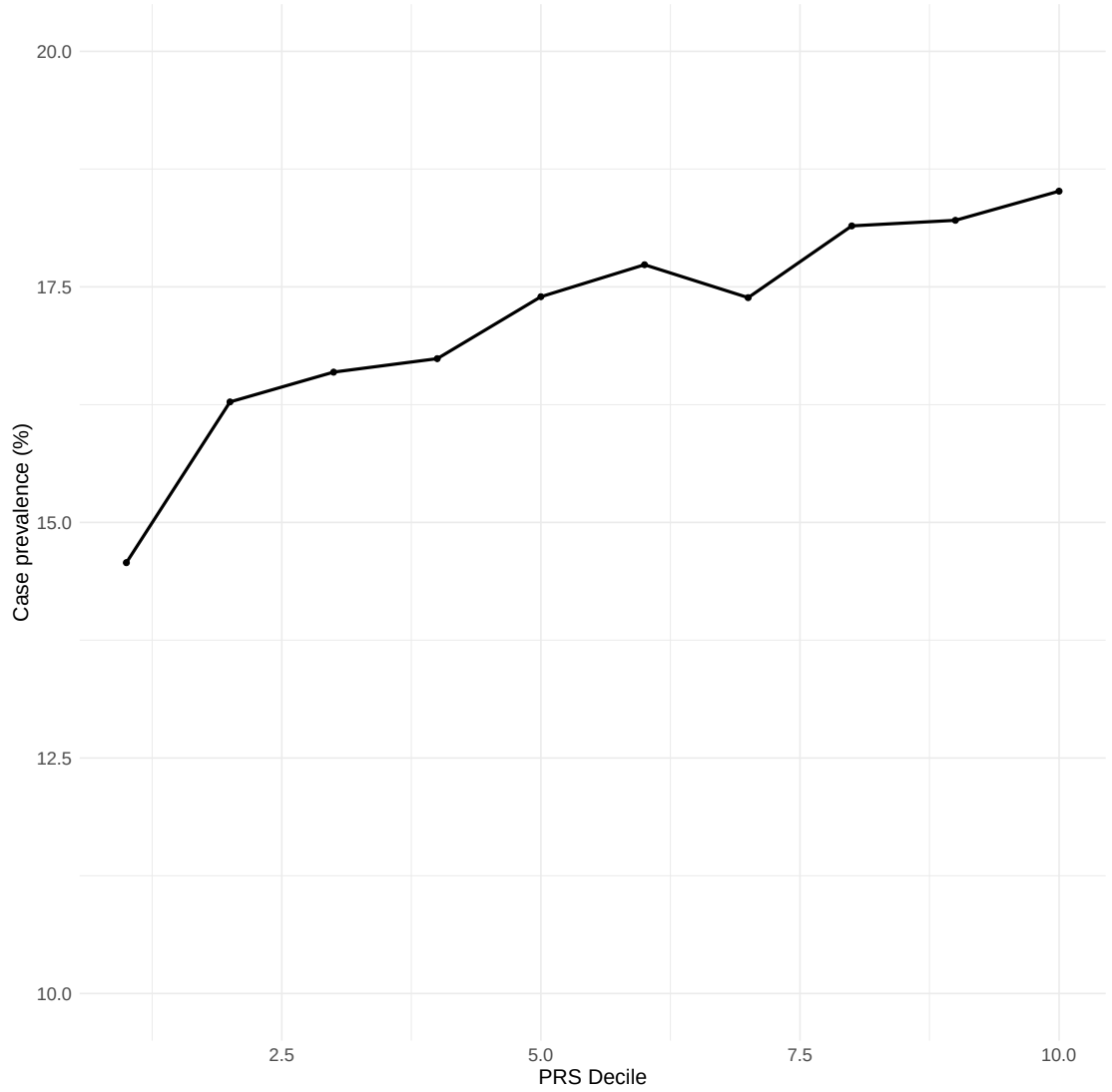
- Summary statistics from a genome-wide association study performed in the UK Biobank for the alcohol use disorders identification test (AUDIT) were used to construct polygenic risk scores (PRS).
- PRS were created for 209,020 European ancestry individuals using the clumping and thresholding method.
- P-value informed clumping was performed using 1000 Genomes European individuals as the LD background, with an r2 = 0.1 and a distance threshold of 250kb.
- Risk scores were calculated for nine different p-value thresholds ($p \leq 0.000001, 0.00001, 0.0001, 0.001, 0.01, 0.05, 0.1, 0.5, 1$).

AUDIT PRS

- AUDIT PRS was significantly associated with AUDIT-C (OR=1.06, $p=4.6x10^{-57}$) and AUD (OR=1.09, $p=1.4x10^{-44}$).



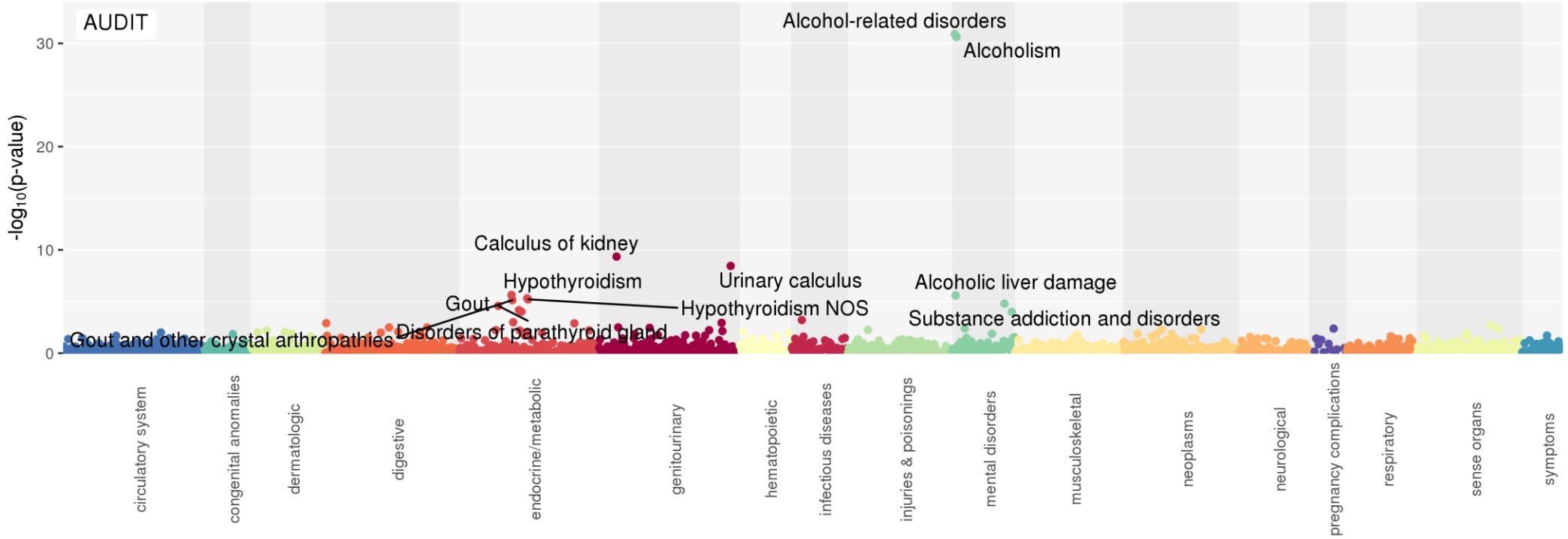
AUDIT PRS – case prevalence of AUD per decile



18.5% of individuals in the top decile of PRS had AUD compared to 14.6% of individuals in the lowest decile of PRS.

AUDIT PheWAS

- Positive phenotypic associations with alcohol-related disorder, alcoholic liver damage, and substance addiction and disorders
- Negative phenotypic associations with calculus of kidney, urinary calculus, gout, hypothyroidism, and hyperglyceridemia



Conclusions

- GWAS studies have identified variants associated with substance use disorders
- Polygenic risk scores can explain a larger amount of phenotypic variation than single SNPs alone
- PRS are associated with the expected phenotypes in an independent sample
- PRS can identify secondary phenotypes associated with genetic liability for disorder
- However, clinical relevance is currently limited – not ‘prime time ready’ just yet

Next steps

- Larger/better GWAS (including multiple ancestries) – will allow us to create more powerful PRS
- Test association with intermediate phenotypes, environment
  - Is genetic liability for opioid use disorder associated with variation in opioid neurotransmission?
  - Positron emission tomography (PET) neuroimaging
  - Binding potential of the mu-opioid receptor - [11C]-carfentanil, a PET tracer
- Incorporate PRS into clinical prediction models
  - May look different for different disorders

Acknowledgements

Henry Kranzler
Hang Zhou
Amy C. Justice
Joel Gelernter
Daniel Rader
Scott Damrauer
JoEllen Weaver
Regeneron Genetics Center
Participants of the Penn Medicine BioBank
Participants of the Million Veteran Program