THE UNIVERSITY OF QUEENSLAND AUSTRALIA | CREATE CHANGE

Introduction to Polygenic Prediction

History, Theory, Methodology & Applications

Jian Zeng

j.zeng@uq.edu.au

THE UNIVERSITY OF QUEENSLAND AUSTRALIA | Institute for Molecular Bioscience

Program in Complex Trait Genomics

Slides credit: Naomi Wray

Acknowledgement of Country

- The University of Queensland (UQ) acknowledges the Traditional Owners and their custodianship of the lands on which we meet.
- We pay our respects to their Ancestors and their descendants, who continue cultural and spiritual connections to Country.
- We recognise their valuable contributions to Australian and global society.

Image: Digital reproduction of *A guidance through time* by Casey Coolwell and Kyra Mancktelow

General Information:

- We are currently located in Building 69

Emergency evacuation point

- Food court and bathrooms are located in Building 63

- If you are experiencing cold/flu symptoms or have had COVID in the last 7 days please ensure you are wearing a mask for the duration of the module

Data Agreement

To maximize your learning experience, we will be working with genuine human genetic data, during this module.

Access to this data requires agreement to the following in to comply with human genetic data ethics regulations

If you haven't done so, please email <ctr-pdg-admin@imb.uq.edu.au> with your name and the below statement to confirm that you agree with the following:

"I agree that access to data is provided for educational purposes only and that I will not make any copy of the data outside the provided computing accounts."

Learning materials

Instructions to access WiFi/desktop/server:

https://suave-pillow-de4.notion.site/Instruction-to-Computing-Resources-dcba658c9a584e6d80a443c5d64042d8?pvs=4

The winter school server is available until **15th July 2024** (2 weeks after the course)

Slides and practical notes for this module:

https://cnsgenomics.com/data/teaching/GNGWS24/module5/

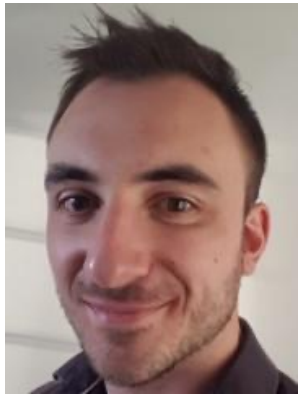
Module structure

- Understand what polygenic scores (PGS) are [Lecture 1]
- How to evaluate PGS prediction accuracy [Lecture/Prac 2]
- Learn the basic and advanced methods to calculate PGS
  - Basic method [Lecture/Prac 1]
  - Best linear unbiased prediction [Lecture/Prac 3]
  - Bayesian methods [Lecture/Prac 4]
  - Summary-data-based methods [Lecture/Prac 5]
- Our pipeline how to generate a PGS from start to end [Lecture/Prac 6]

Approx. 40 min Lecture | 5 min break | 40 min Prac | 5 min break

Jian Zeng

Tian Lin



Valentin Hivert

Fleur Garton

Polygenic scores (PGS)

Polygenic scores (PGS) predict individual genetic values of complex traits using genome variations.

Polygenic risk scores (PRS) are predictors of the genetic susceptibilities of individuals to diseases.

Head in the clouds Head in the sand

Source: Strachan & Read Human Molecular Genetics 3.

History of PGS

2019

GENETICS | HIGHLIGHTED ARTICLE
GENOMIC PREDICTION

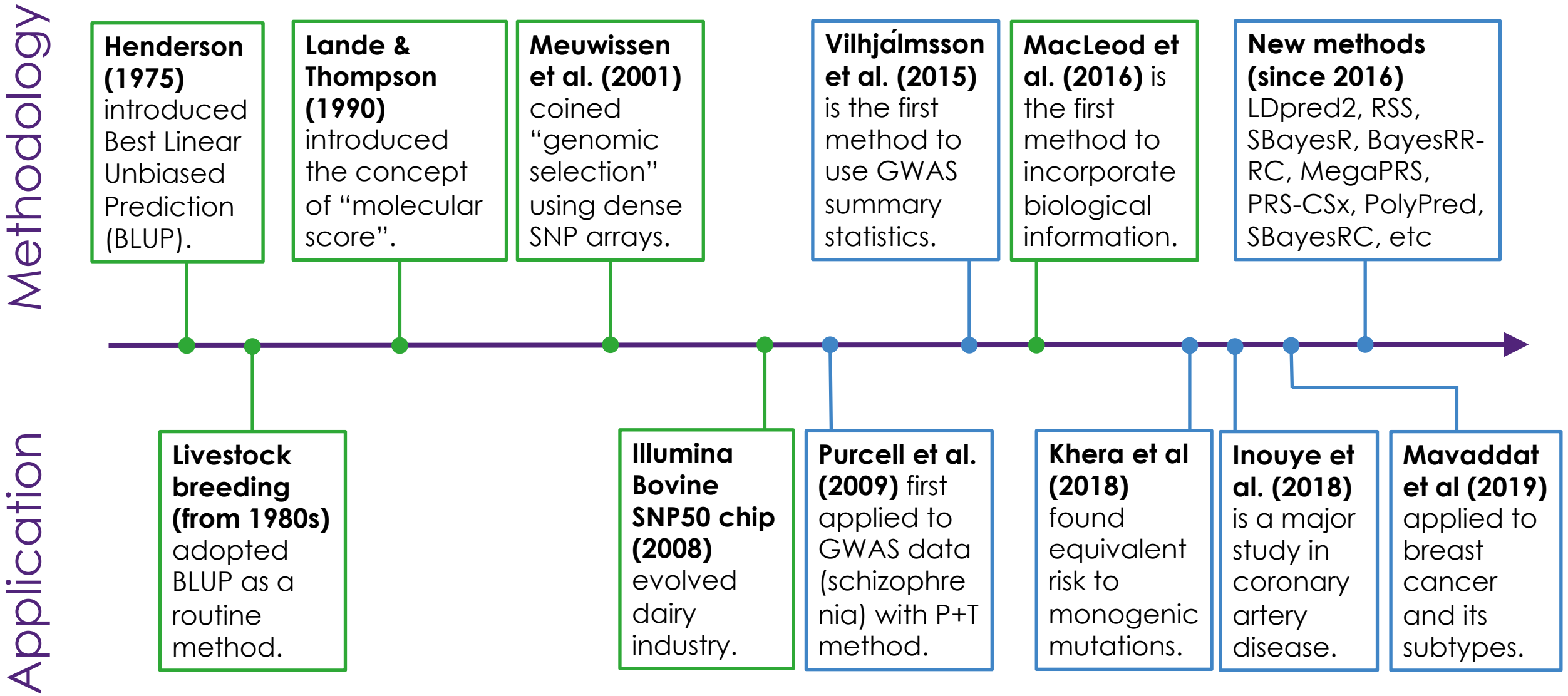
GENOMIC PREDICTION

Complex Trait Prediction from Genome Data: Contrasting EBV in Livestock to PRS in Humans

Naomi R. Wray,*,†,1 Kathryn E. Kemper,* Benjamin J. Hayes,‡ Michael E. Goddard,§,** and Peter M. Visscher*,†

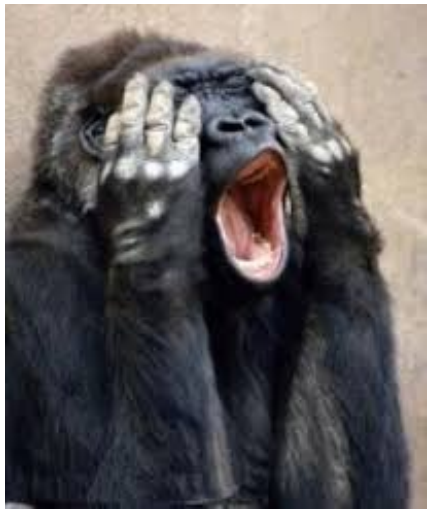
A brief history of PGS in humans & agriculture

The University Of Queensland AUSTRALIA



What's in a name?

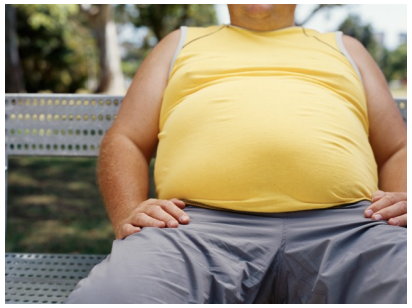
- **PRS**- Polygenic risk score
- **GPRS**- Genomic or genetic profile risk score
- **PGS** -Polygenic score
- **GRS** - Genetic risk score
- **rsPS** – restricted to significant polygenic score
- **gePS** – global extended polygenic score
- **Multi-SNP score** (usually this uses only single nucleotide polymorphisms (SNPs) that are genome-wide significant, hence the same as gePS)
- **MetaGRS** – a PRS constructed from genetic data for the disease/trait of interest plus from other correlated traits
- **MTAG-GRS/PRS** a PRS constructed from GWAS data from multiple correlated traits
- **Genetic score**
- **Genotypic score**
- **Allele score**
- **Profile score**
- **Linear predictor** (this of course is a generic term, but has been used to describe PRS when risk alleles are the only predictors)



Theory and methodology
of polygenic scores (PGS)
are built on
our understanding of
"polygenicity"
in complex traits.

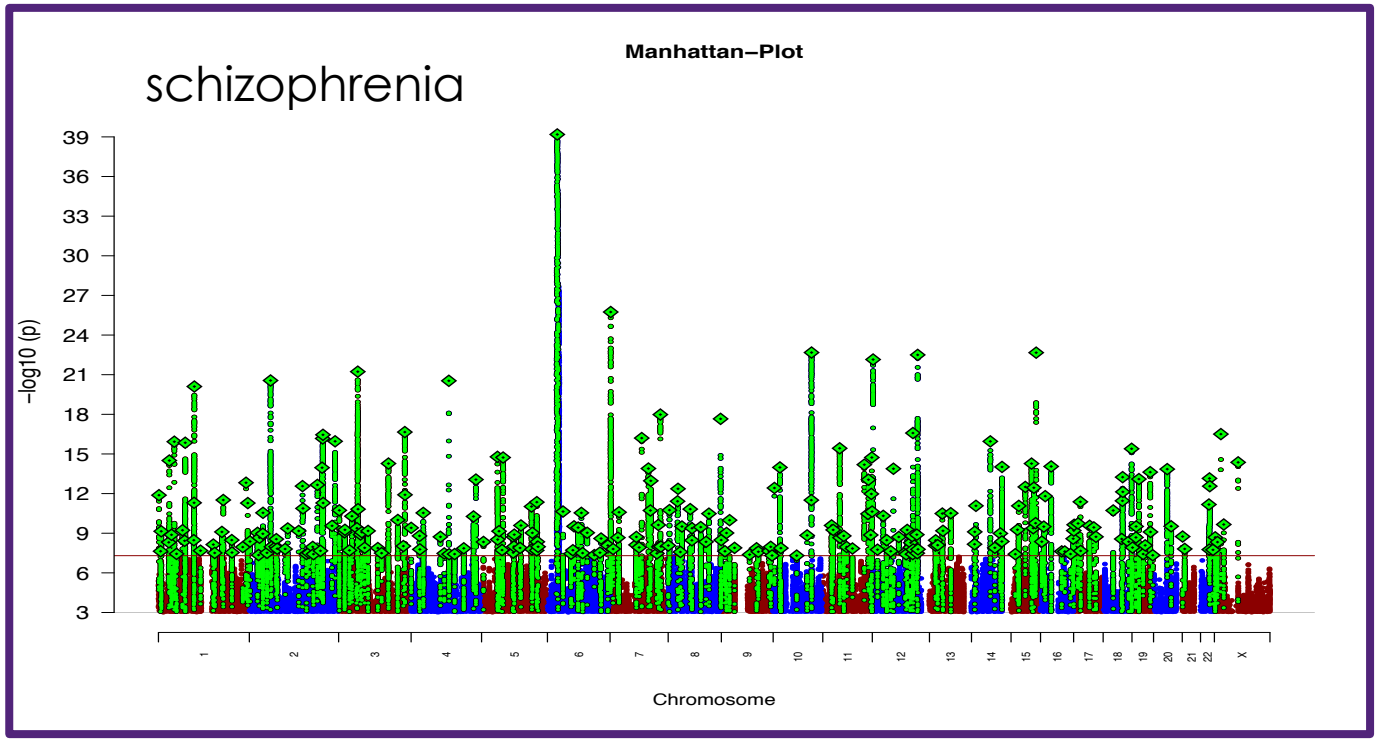
Height

Schizophrenia



Obesity

Common diseases are polygenic



nature

Explore content | About the journal | Publish with us

nature › articles › article

Article | Published: 08 April 2022

Mapping genomic loci implicates genes and synaptic biology in schizophrenia

Vassily Trubetskoy, Antonio F. Pardiñas, Ting Qi, Georgia Panagiotaropoulou, Swapnil Awasthi, Tim B. Bigdeli, Julien Bryois, Chia-Yen Chen, Charlotte A. Dennison, Lynsey S. Hall, Max Lam, Kyoko Watanabe, Oleksandr Frei, Tian Ge, Janet C. Harwood, Frank Koopmans, Sigurdur Magnusson, Alexander L. Richards, Julia Sidorenko, Yang Wu, Jian Zeng, Jakob Grove, Minsoo Kim, Zhiqiang Li, Indonesia Schizophrenia Consortium, PsychENCODE, Psychosis Endophenotypes International Consortium, The SynGO Consortium, Schizophrenia Working Group of the Psychiatric Genomics Consortium

+ Show authors

Nature **604**, 502–508 (2022) | Cite this article

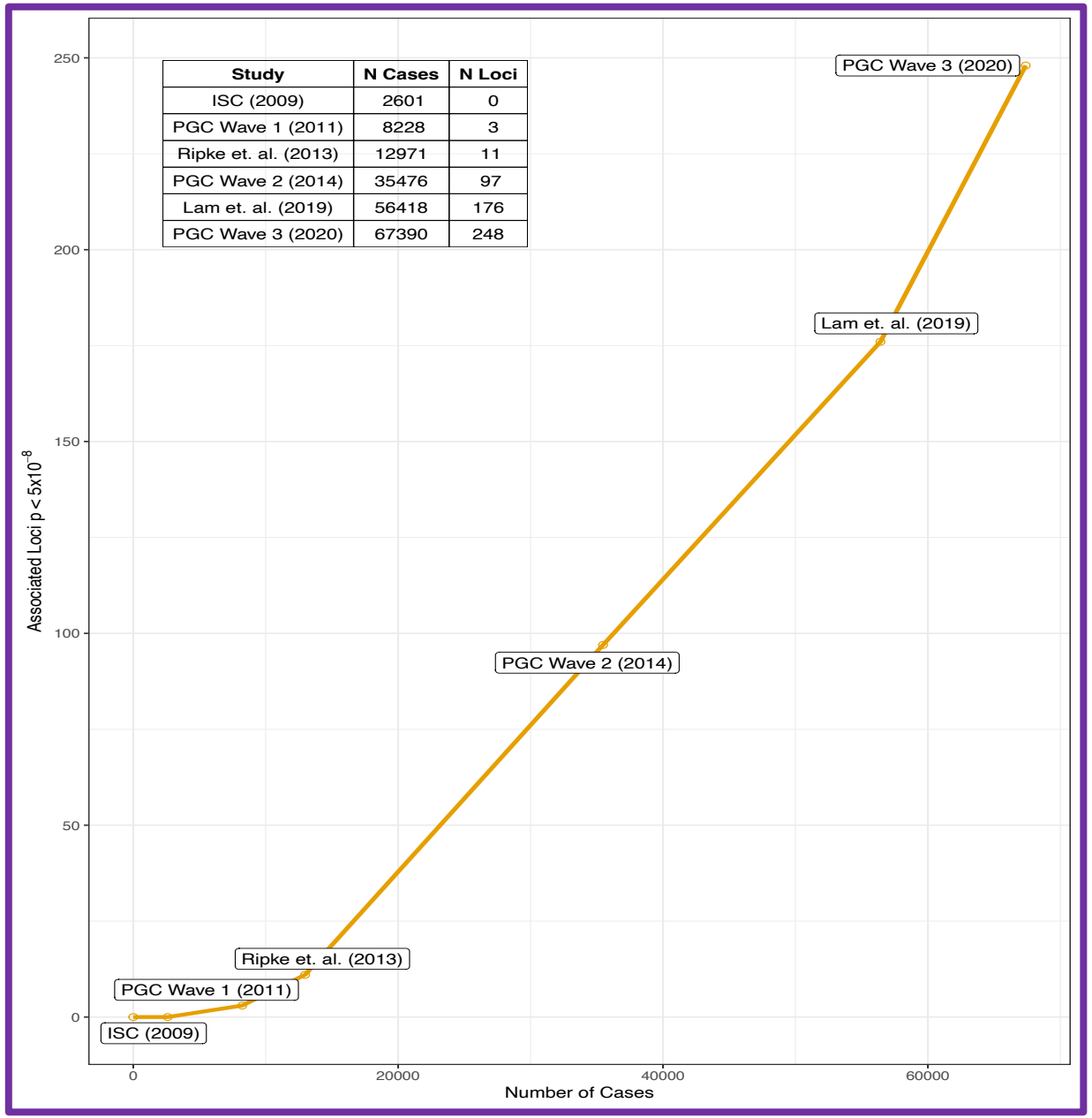
57k Accesses | **321** Citations | **463** Altmetric | Metrics

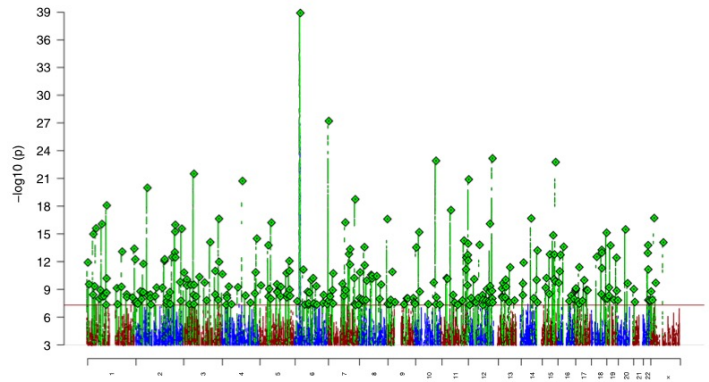
248 risk loci identified at genome-wide significance level.

We predict thousands are associated with schizophrenia.

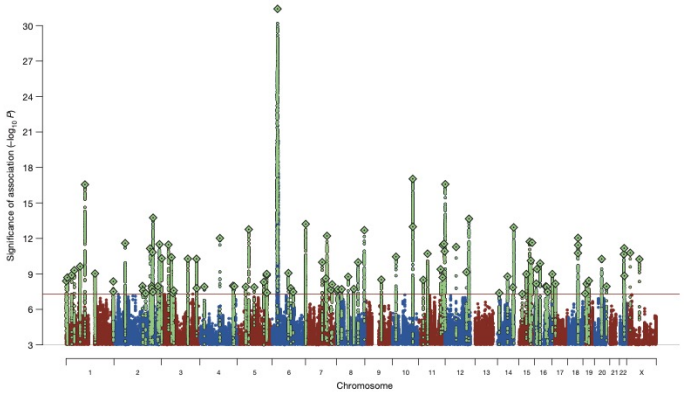
Common diseases are polygenic

Study	N Cases	N Loci
ISC (2009)	2601	0
PGC Wave 1 (2011)	8228	3
Ripke et. al. (2013)	12971	11
PGC Wave 2 (2014)	35476	97
Lam et. al. (2019)	56418	176
PGC Wave 3 (2020)	67390	248

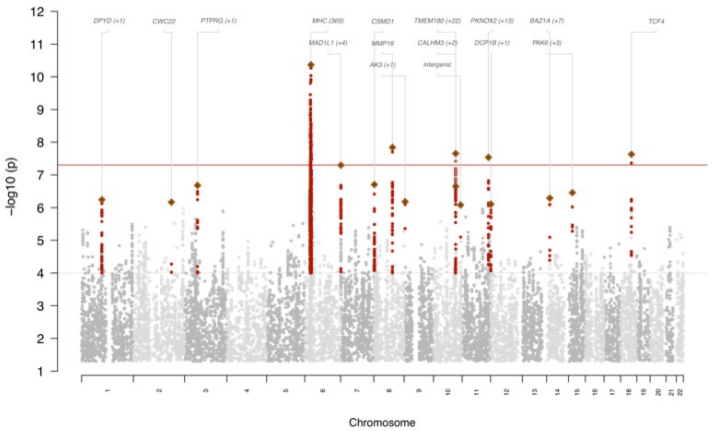




2022 PGC Wave 3

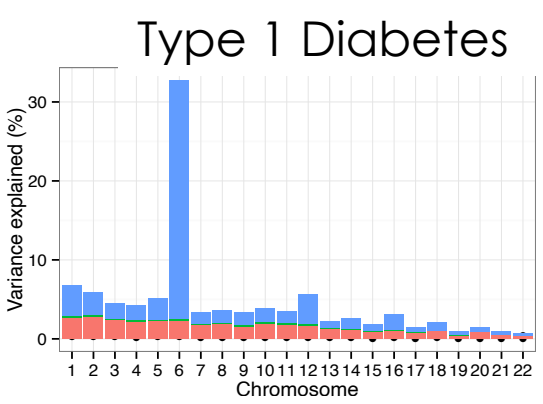


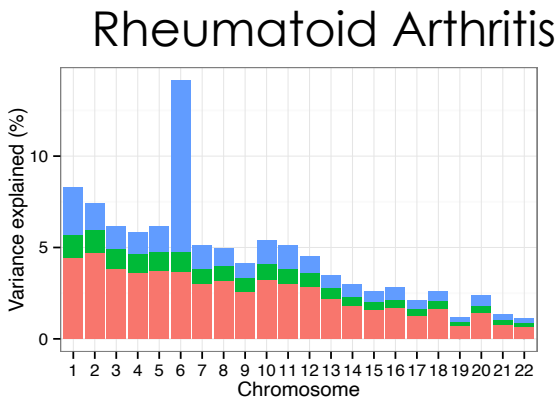
2014 PGC Wave 2

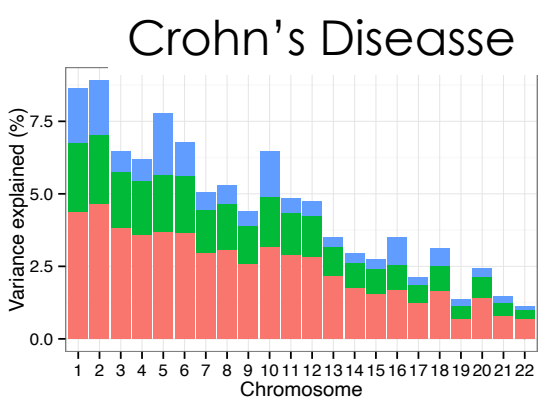


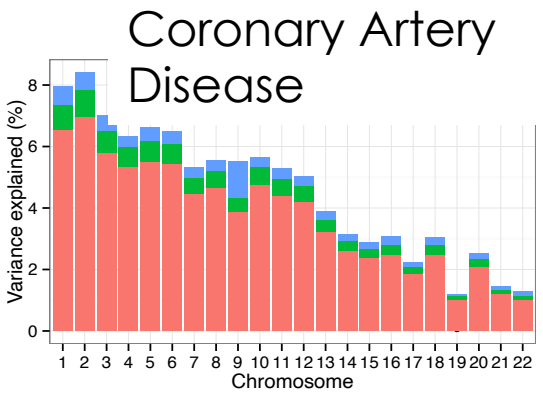
2011 PGC Wave 1

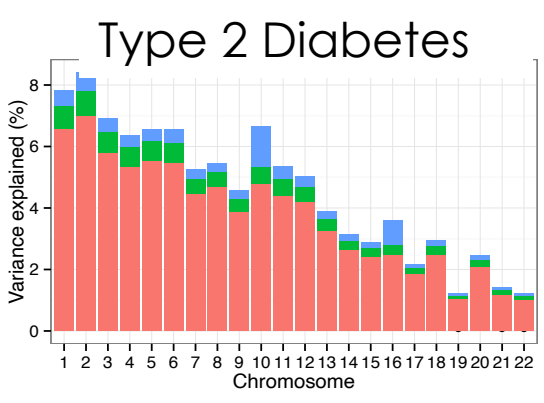
Many polygenic genetic architectures

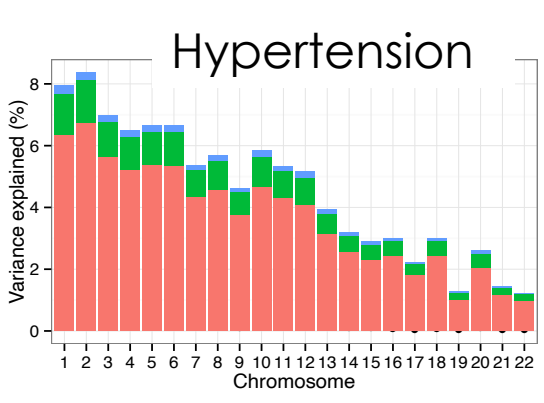


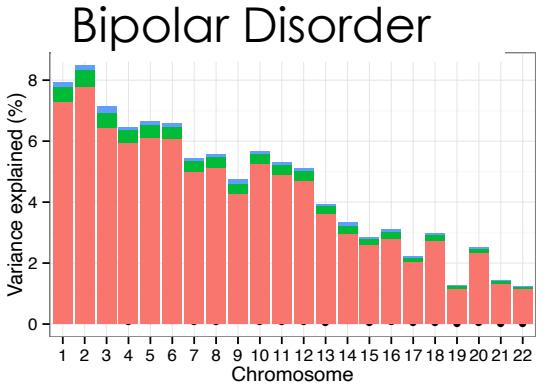


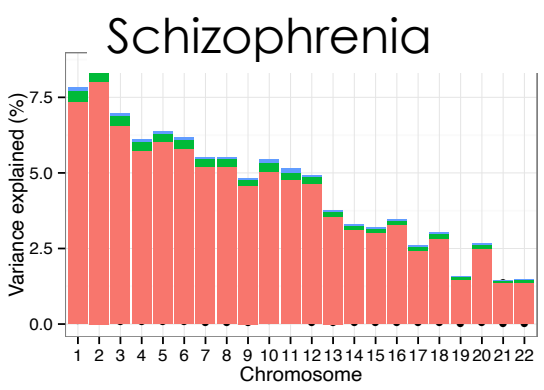


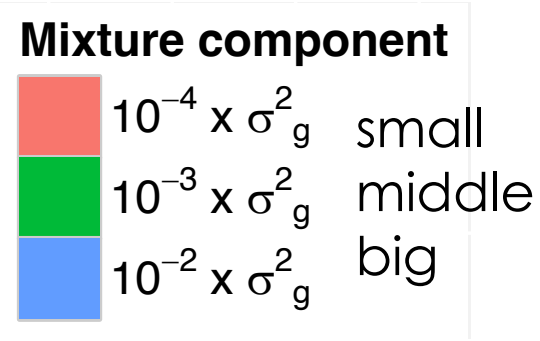


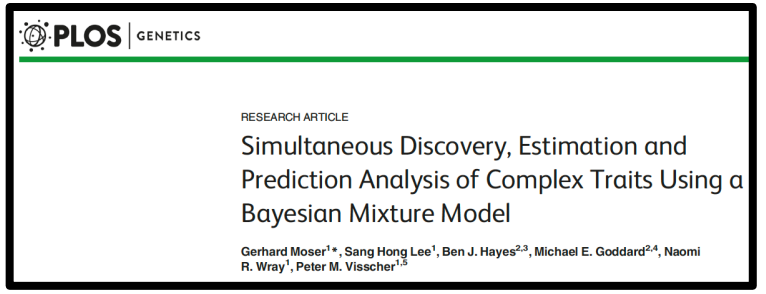











PLOS GENETICS

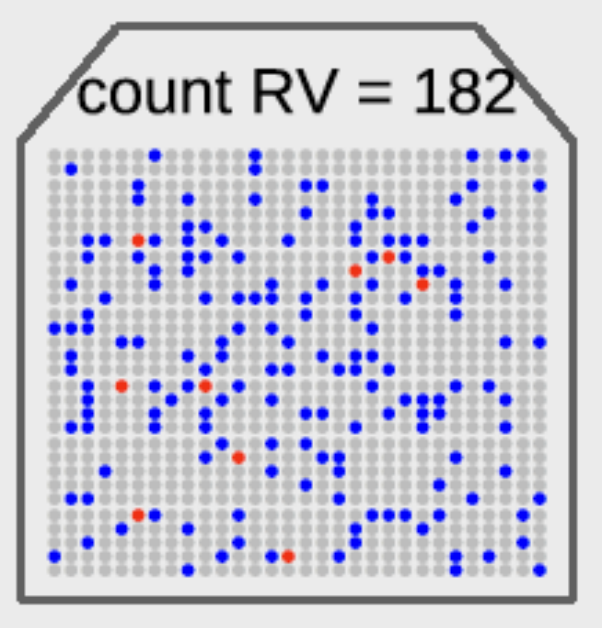
RESEARCH ARTICLE

Simultaneous Discovery, Estimation and Prediction Analysis of Complex Traits Using a Bayesian Mixture Model

Gerhard Moser[1]*, Sang Hong Lee[1], Ben J. Hayes[2,3], Michael E. Goddard[2,4], Naomi R. Wray[1], Peter M. Visscher[1,5]

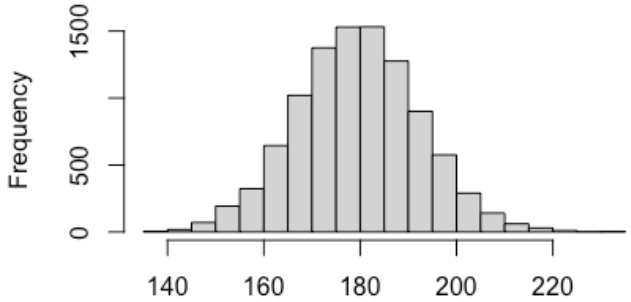
Many DNA variants contribute to genetic risk, and most have very small effects.

Polygenic disease for an individual



900 DNA polymorphic sites

RV =risk variant



Count of RV in population

Frequency of risk variant at each site: 0.1 (p)

Average person 900*2*0.1 = 180 risk variant

Mean +/- 3SD: 142 to 218

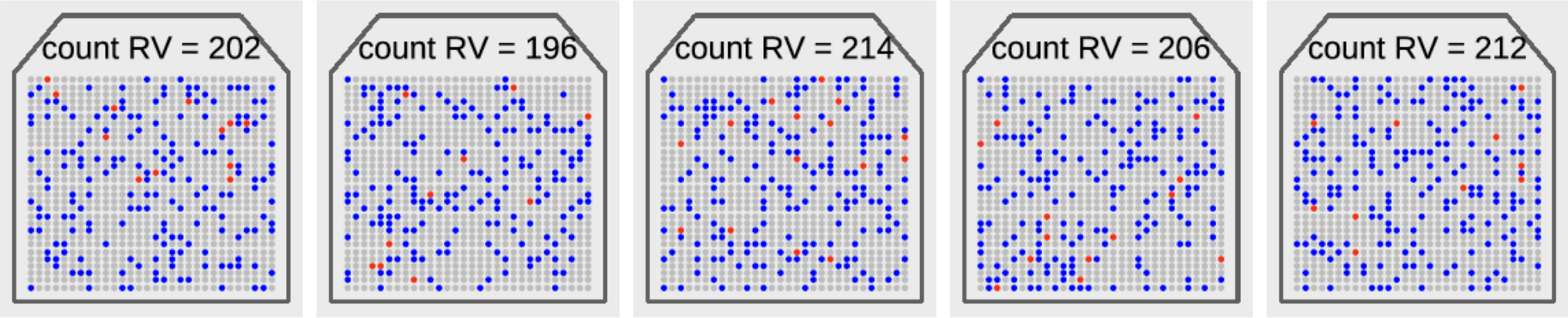
0 Grey: Homozygote no risk alleles (or equivalently 2 protective alleles)
1 Blue : Heterozygote one risk allele (and one non-risk/protective allele)
2 Red: Homozygote two risk alleles

Toy example

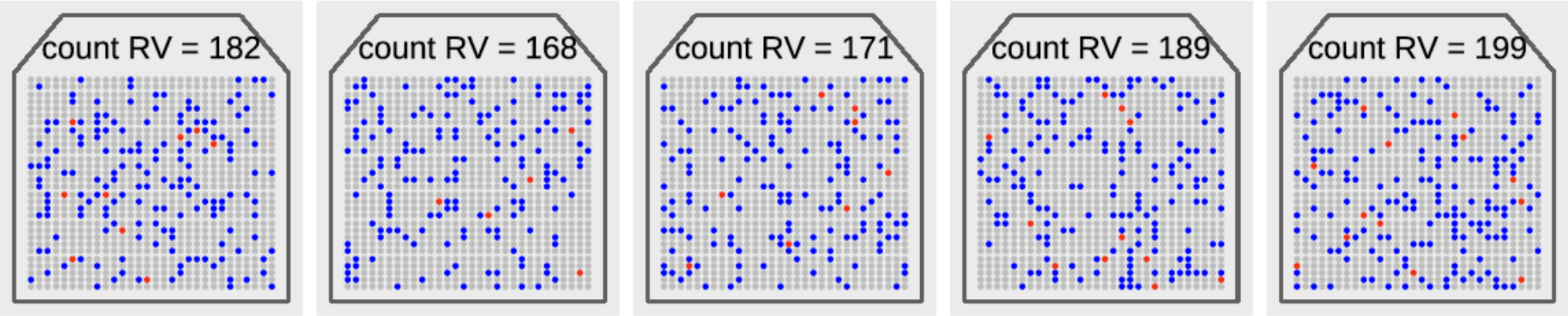
Polygenic disease for an individual

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Affected over lifetime

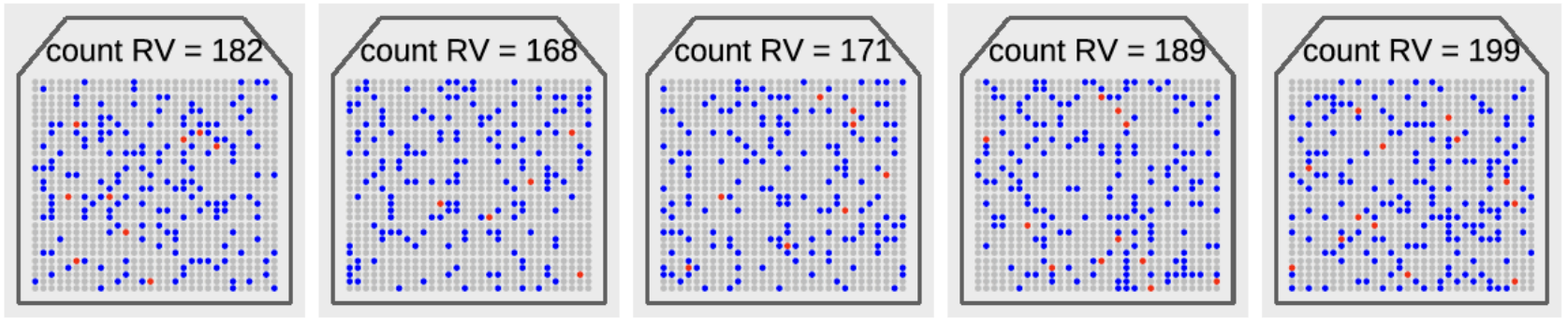


Not affected over lifetime



- We all carry risk variants for all diseases.
- Robustness
- Those affected carry a higher burden.
- Non-genetic factors contribute to risk too
- Each person carries a unique portfolio of risk alleles

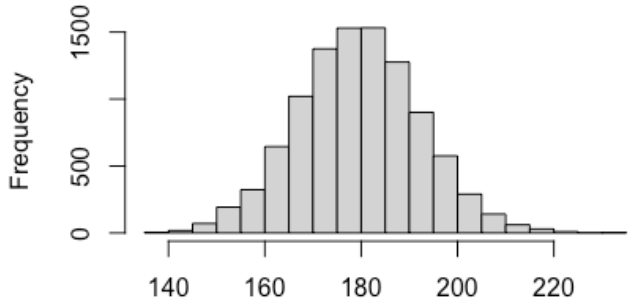
Polygenic score



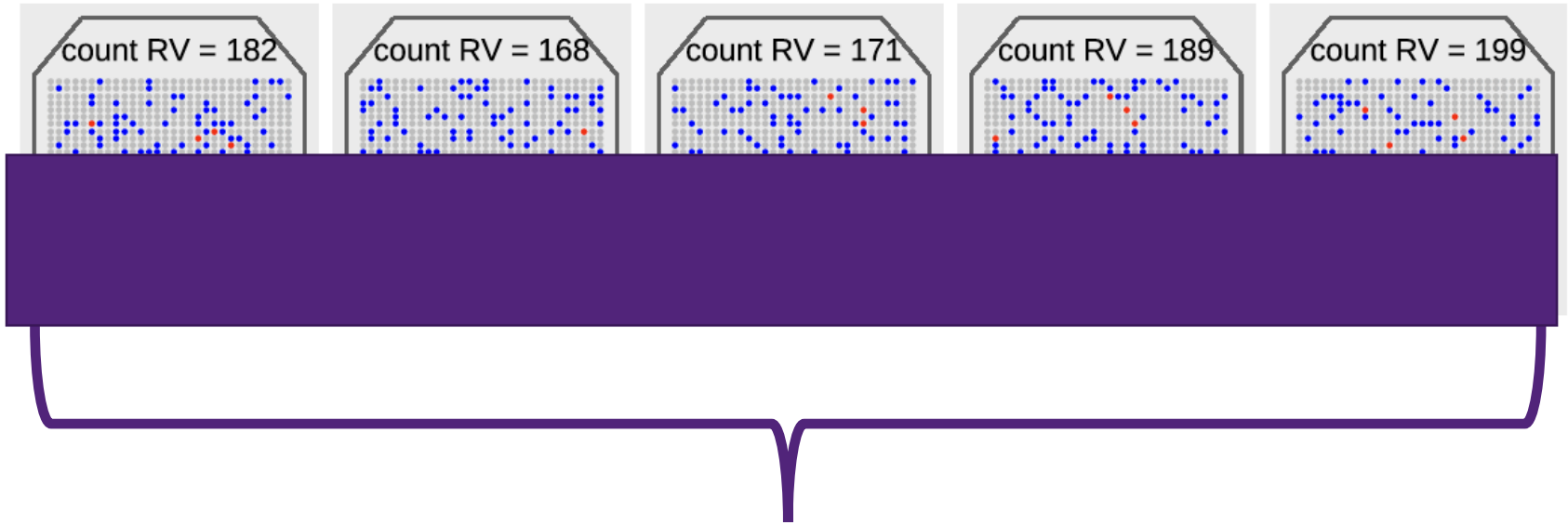
→ "True" polygenic score

Genetic variance between people attributed to all genetic factors V(A)

$h^2 = \frac{V(A)}{V(P)}$ heritability



Polygenic score



→ "True" polygenic score

Not all variants captured on genotyping arrays

Genetic variance between people attributed to all genetic factors V(A)

$h^2 = \frac{V(A)}{V(P)}$ heritability

Genetic variance between people attributed to all genetic factors associated with SNPs on genotyping arrays

$h^2_{SNP} = h^2_g = \frac{V(A:SNP)}{V(P)}$

$\mathrm{SNP-based\ heritability}$

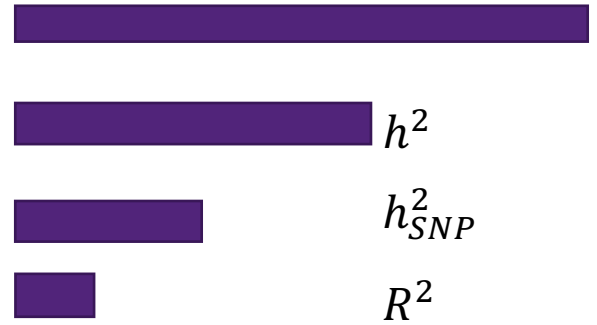
Limitations in prediction accuracy

- PGS have a **theoretical** upper limit dependent on the **heritability of the trait** (how much of the variance of trait values between people is attributed to genetic factors).

- PGS have a **technical** upper limit associated with the proportion of **variance tagged** by the DNA variants measured.

- PGS have a **practical** upper limit dependent on the **sample size of the discovery sample** used to estimate effect sizes of risk alleles, and the **quality** of the discovery sample.

- PGS can be pushed closer to the technical upper limit by the **statistical methodology** used to generate the optimal weighting given to the risk alleles, and new methods integrate new biological data.



Schizophrenia

Max:
25% Liability
AUC 0.84

Current:
11% Liability
AUC 0.74

Polygenic scores cannot be highly accurate predictors of phenotypes

Family history

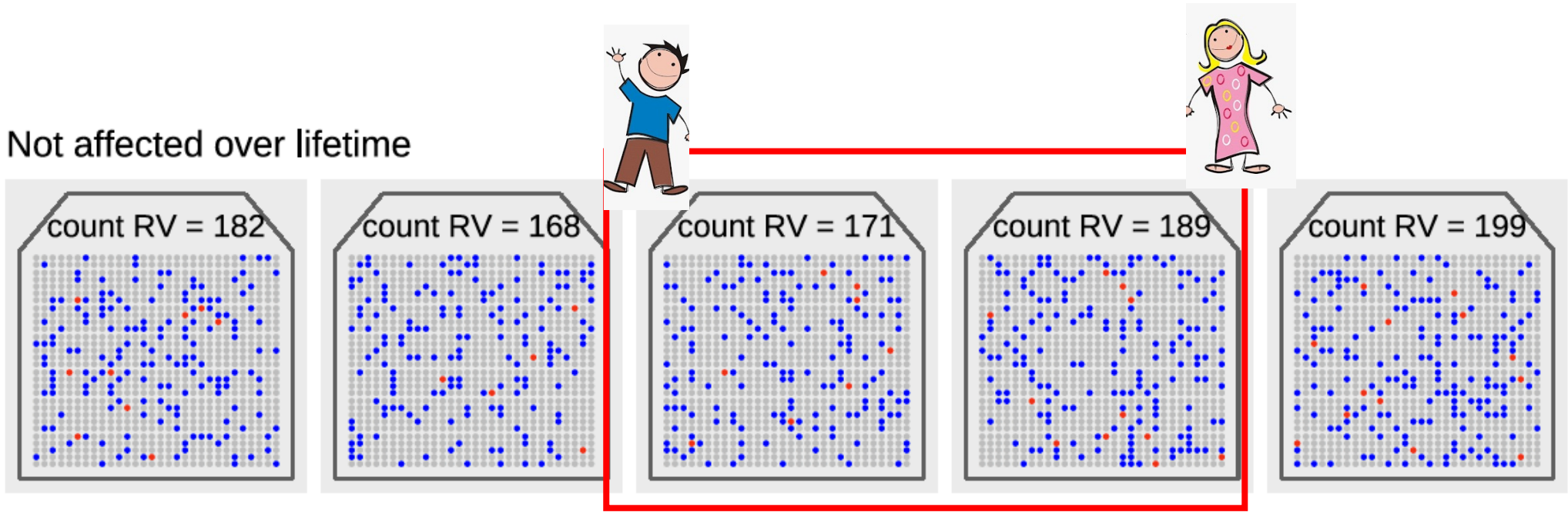
Will people withOUT known family history have high PGS?

Maybe, and that's important!

JAMA Psychiatry | Review

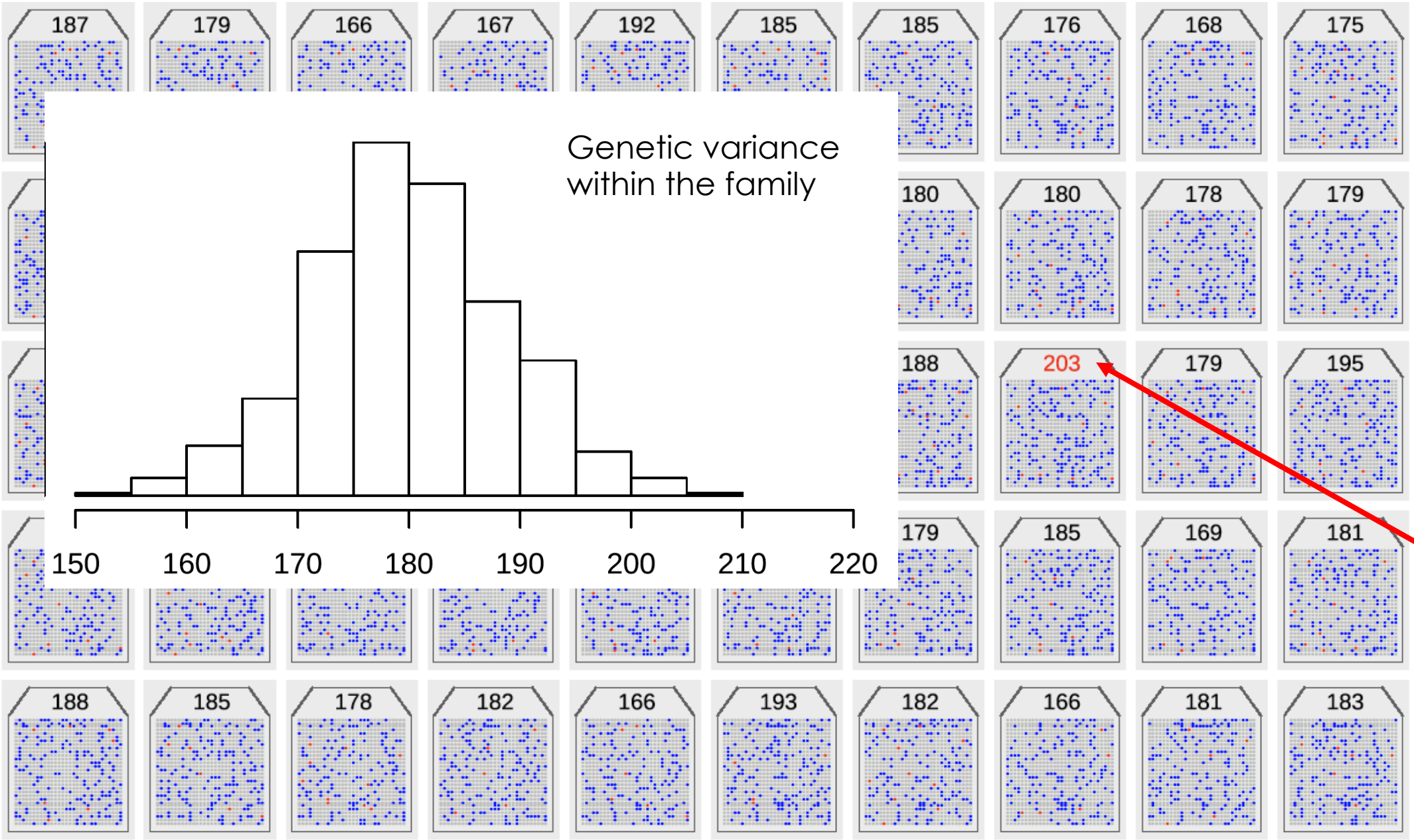
From Basic Science to Clinical Application of Polygenic Risk Scores
A Primer

Naomi R. Wray, PhD; Tian Lin, PhD; Jehannine Austin, PhD; John J. McGrath, MD, PhD; Ian B. Hickie, MD; Graham K. Murray, MD, PhD; Peter M. Visscher, PhD



Grey: Homozygote: Two non-risk/protective alleles – always passes a non-risk allele to child at the locus
Red: Homozygote: Two risk alleles – always passes a risk allele to child at the locus
Blue: Heterozygotes: One risk allele & one non-risk allele –
passes a risk allele 50% of the time & a non-risk allele 50% of the time

Children (Parents: 171 & 189)



Children of these parents
Mean: 180
+/-3SD: 153-207

Population
Mean: 180
+/-3SD: 142-218

No family history, but by chance segregation of alleles has high genetic risk

Family history

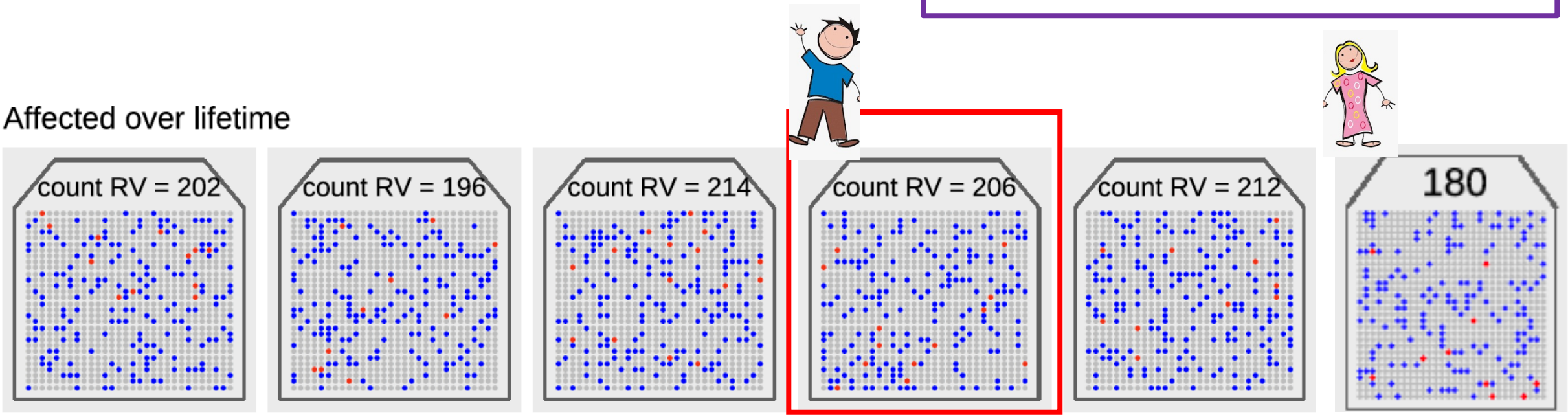
Will people with known family history have high PGS?

Maybe, maybe not!!

JAMA Psychiatry | Review

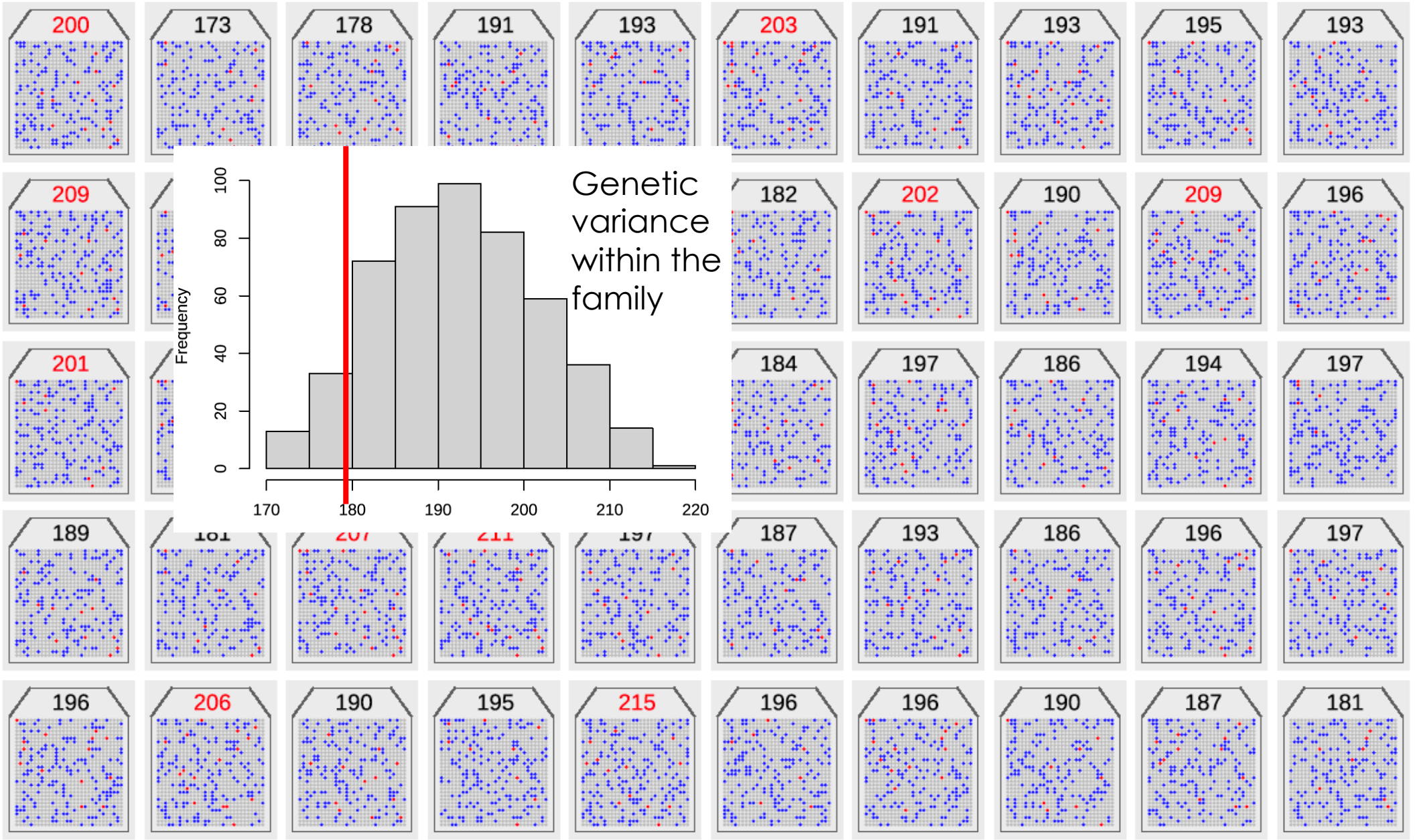
From Basic Science to Clinical Application of Polygenic Risk Scores
A Primer

Naomi R. Wray, PhD; Tian Lin, PhD; Jehannine Austin, PhD; John J. McGrath, MD, PhD; Ian B. Hickie, MD; Graham K. Murray, MD, PhD; Peter M. Visscher, PhD



Grey: Homozygote: Two non-risk/protective alleles – always passes a non-risk allele to child at the locus
Red: Homozygote: Two risk alleles – always passes a risk allele to child at the locus
Blue: Heterozygotes: One risk allele & one non-risk allele –
passes a risk allele 50% of the time & a non-risk allele 50% of the time

Children (Parents: 206 & 180)



Children of these parents
Mean: 193
+/-3SD: 166-220

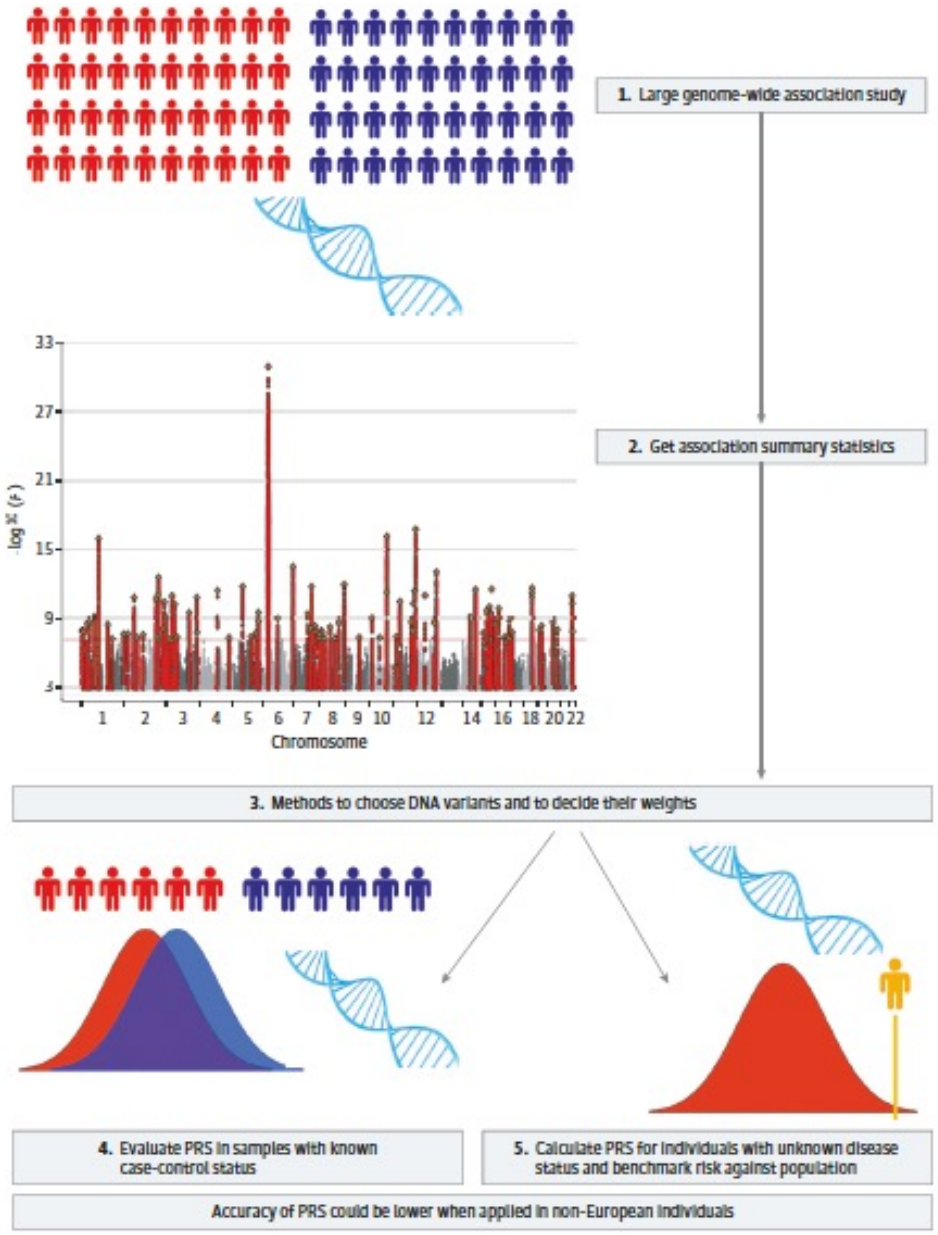
Population
Mean: 180
+/-3SD: 142-218

Summary

- Complex traits are polygenic, with many variants of small effects.
- Prediction accuracy is limited by heritability, SNP set, sample size & statistical method.
- Substantial genetic variation within the family (half of that in the whole population).
- A high PGS is mostly a consequence of genetic sampling.

Evaluations and applications

Polygenic scores



- A weighted count of risk alleles

$$PGS = \widehat{\beta_1} x_{i1} + \widehat{\beta_2} x_{i2} + \widehat{\beta_3} x_{i3} + \cdots = \sum_{j=1}^{n_{SNP}} \widehat{\beta_j} x_{ij}$$

0, 1 or 2
Risk alleles

Which SNPs?

What weights?

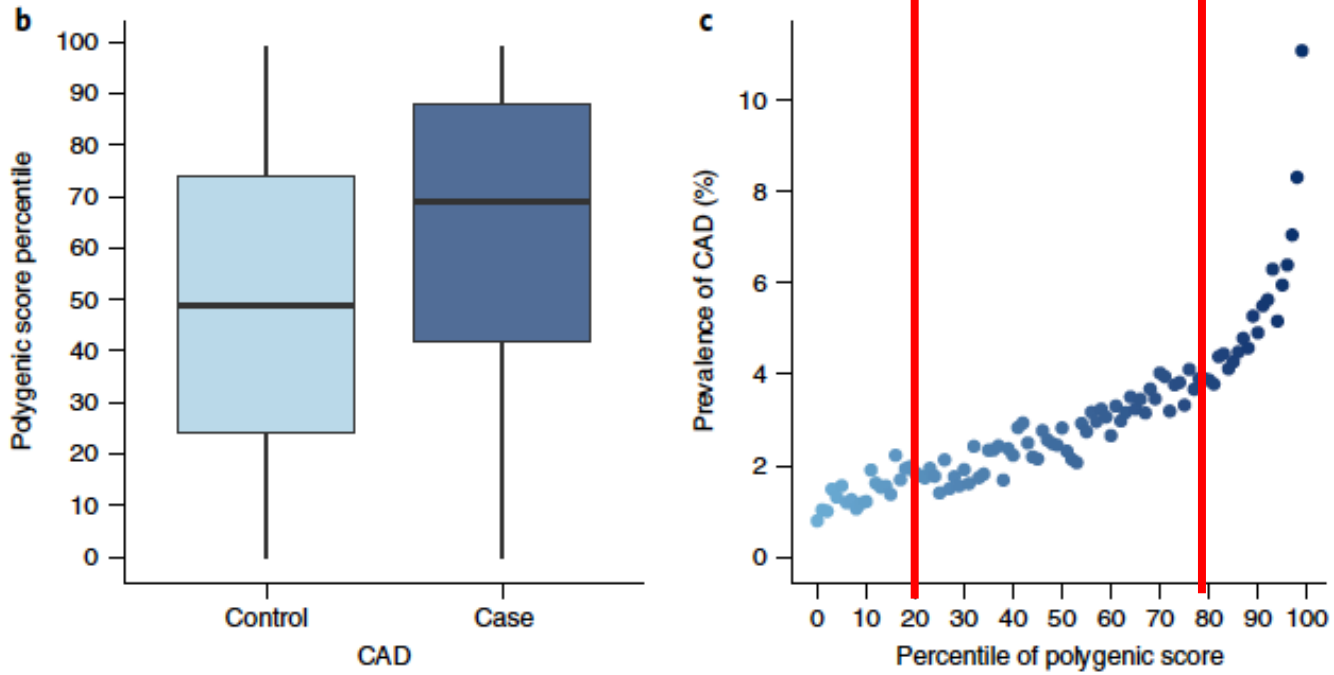
- Don't need to know causal variants for prediction!
- Prediction can be based on correlated variants.

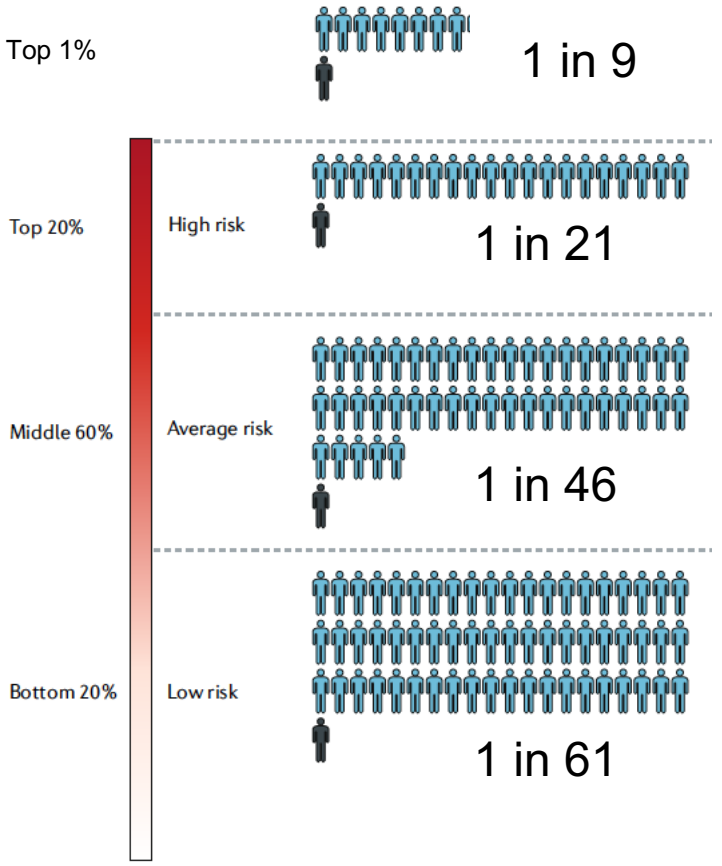
4. Evaluate
Y= b*PGS + e
R^2 = var(b*PGS)/Var(Y)

AUC statistic:
Probability that a case ranks higher than a control

Evaluating PRS

Different views of the same data



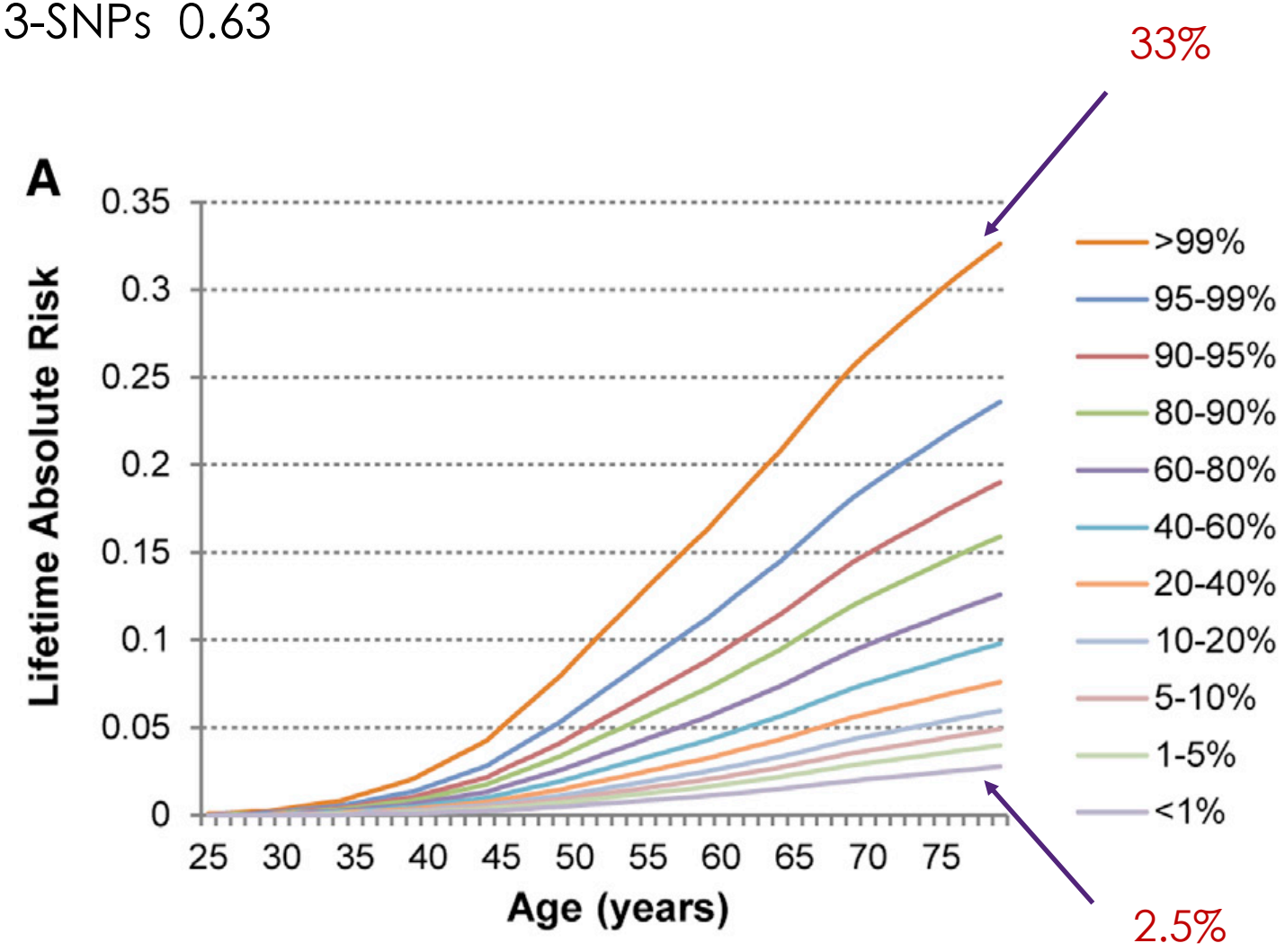


Khera et al (2018) Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nature Genetics

Torkamani et al, Nat Rev Genetics, 2018

Breast Cancer

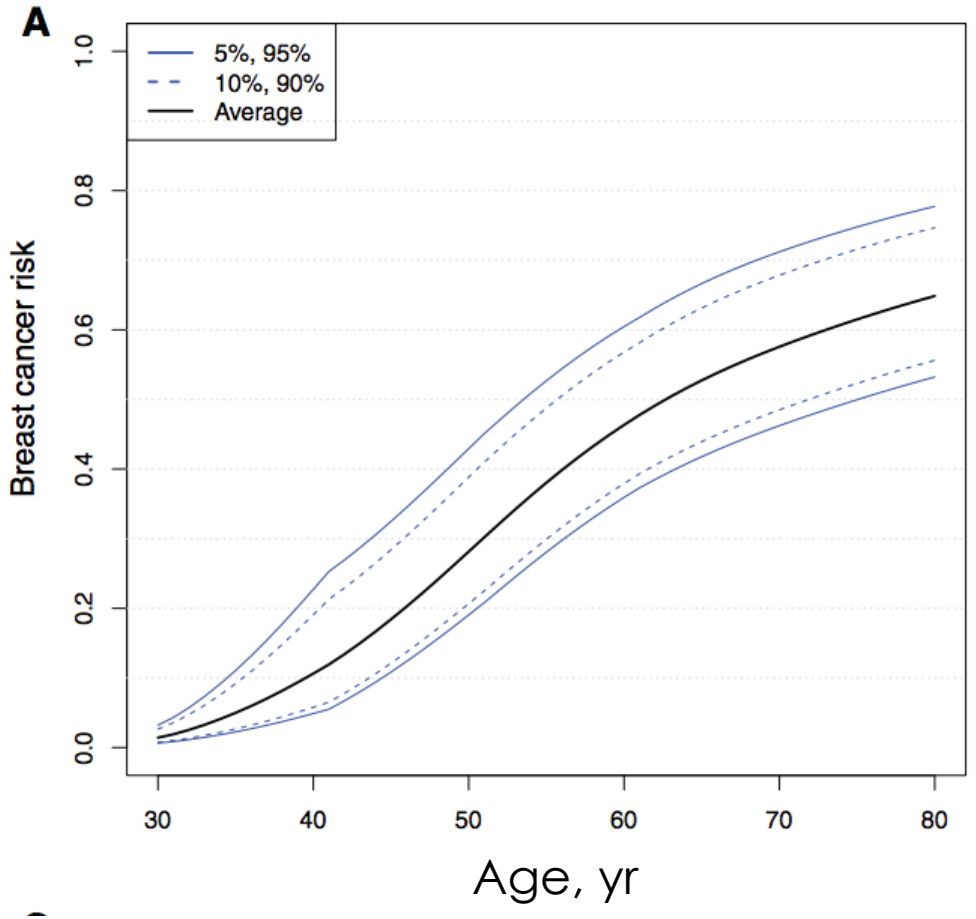
AUC 313-SNPs 0.63



Mavaddat et al (2019) Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. AJHG

Increase prediction accuracy….

Combine PRS with known risk mutations
Breast cancer

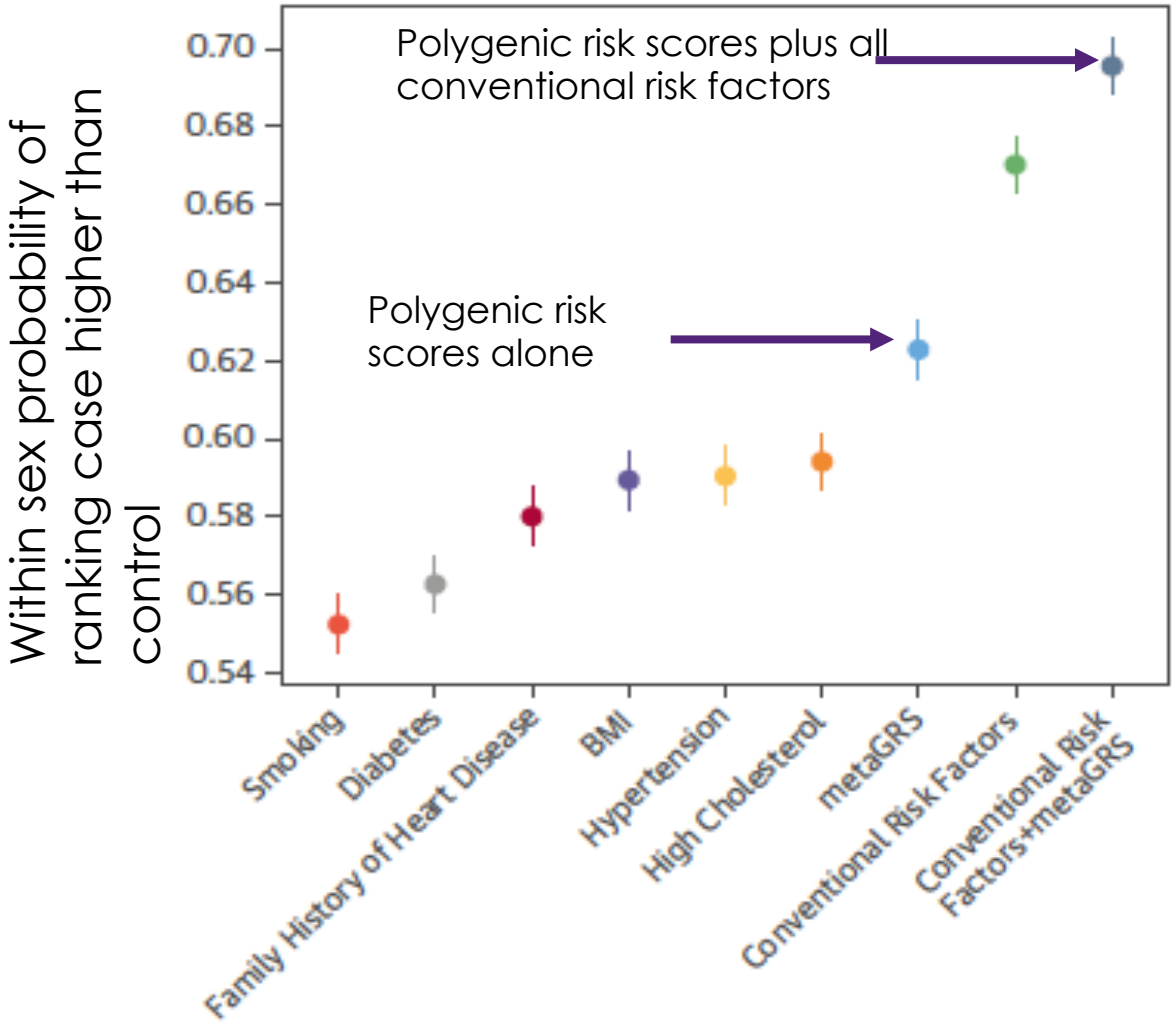


BRCA1 carriers

Kuchenbaecker et al: Evaluation of polygenic risk scores for breast and ovarian cancer risk prediction in BRCA1 and BRCA2 mutation carriers. J Natl Cancer Inst (2017)

Increase prediction accuracy….

Combine PRS with conventional risk predictors Coronary Artery Disease



Inouye et al (2018) Genomic risk prediction of CAD in 480K adults. JACC

Polygenic risk score applications

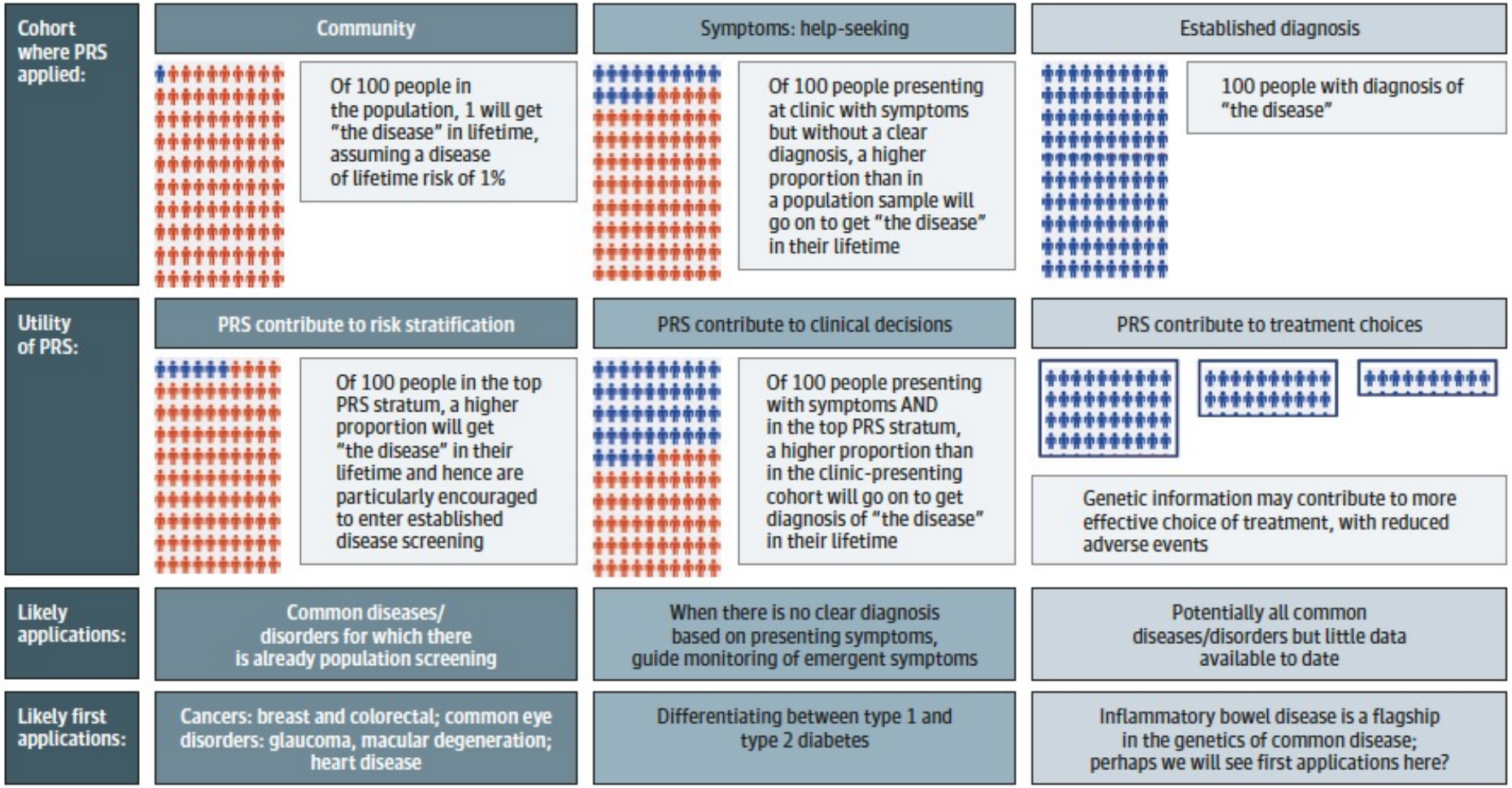
JAMA Psychiatry | Review

From Basic Science to Clinical Application of Polygenic Risk Scores
A Primer

Naomi R. Wray, PhD; Tian Lin, PhD; Jehannine Austin, PhD; John J. McGrath, MD, PhD; Ian B. Hickie, MD; Graham K. Murray, MD, PhD; Peter M. Visscher, PhD

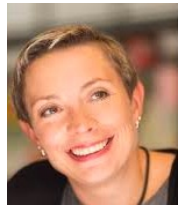
Goal:

- Understandable by interested clinician
- Technically accurate – backed up in Supplement & Rscript



Naomi Wray, UQ & UoOxford

Graham Murray, UoCambridge



Jehannine Austin, UoBritish Columbia

Ian Hickie, UoSydney

John McGrath, UQ

Tian Lin, UQ

Polygenic risk score applications

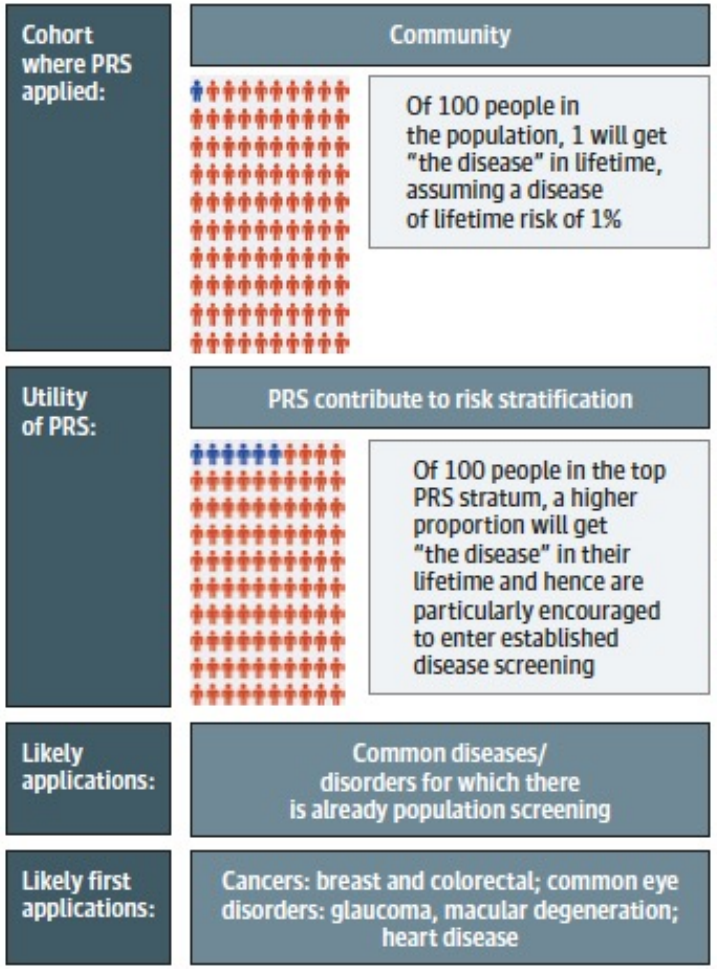
JAMA Psychiatry | Review

From Basic Science to Clinical Application of Polygenic Risk Scores
A Primer

Naomi R. Wray, PhD; Tian Lin, PhD; Jehannine Austin, PhD; John J. McGrath, MD, PhD; Ian B. Hickie, MD; Graham K. Murray, MD, PhD; Peter M. Visscher, PhD

Goal:

- Understandable by interested clinician
- Technically accurate – backed up in Supplement & Rscript

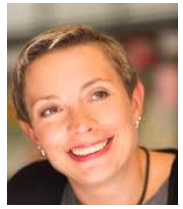


PRS could have utility in community settings (stratification to better triage people into established screening programs)

Naomi Wray, UQ & UoOxford

Graham Murray, UoCambridge

Ian Hickie, UoSydney



Jehannine Austin, UoBritish Columbia

John McGrath, UQ

Tian Lin, UQ

Polygenic risk score applications

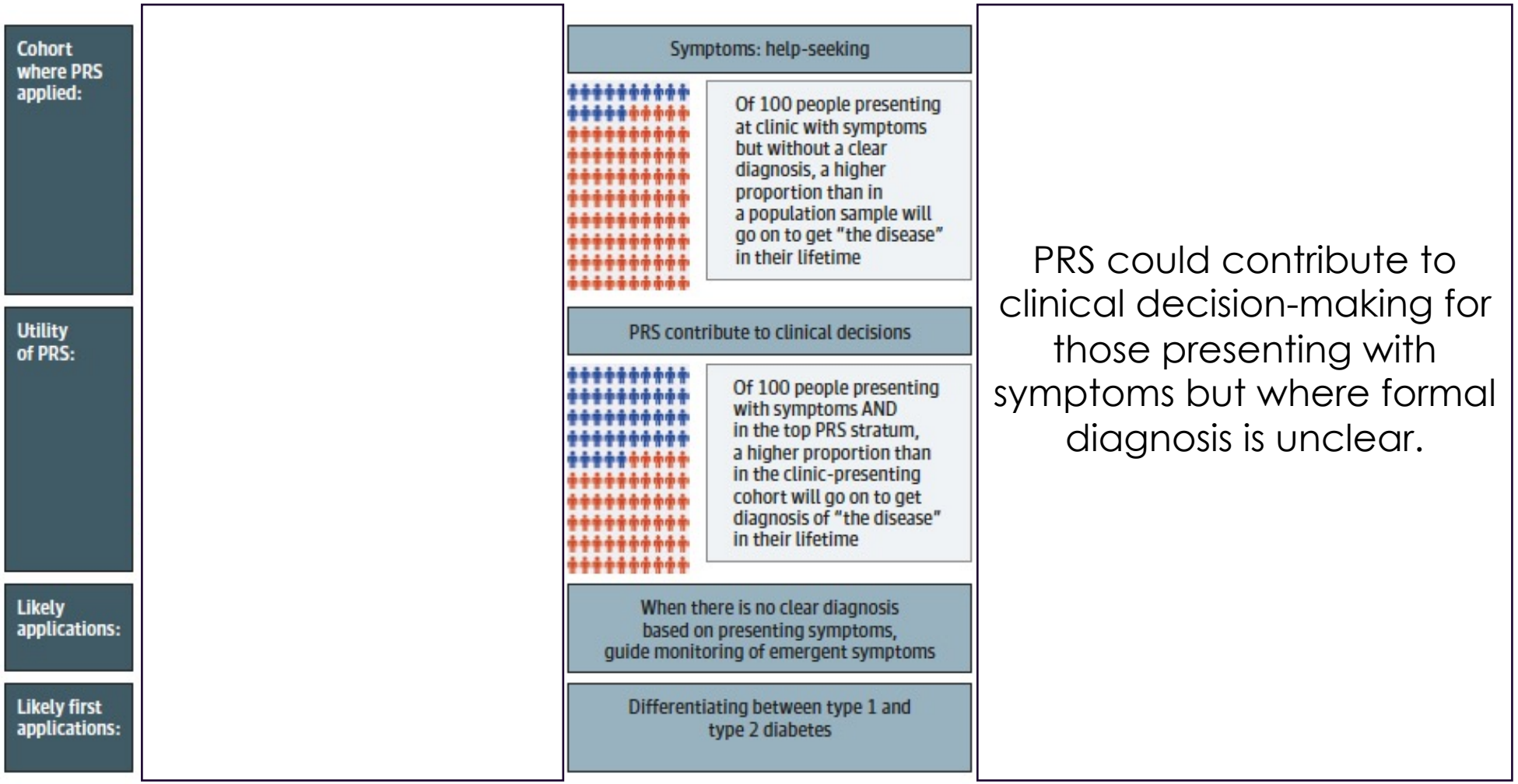
JAMA Psychiatry | Review

From Basic Science to Clinical Application of Polygenic Risk Scores
A Primer

Naomi R. Wray, PhD; Tian Lin, PhD; Jehannine Austin, PhD; John J. McGrath, MD, PhD; Ian B. Hickie, MD; Graham K. Murray, MD, PhD; Peter M. Visscher, PhD

Goal:

- Understandable by interested clinician
- Technically accurate – backed up in Supplement & Rscript

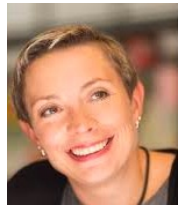


PRS could contribute to clinical decision-making for those presenting with symptoms but where formal diagnosis is unclear.

Naomi Wray, UQ & UoOxford

Graham Murray, UoCambridge

Ian Hickie, UoSydney



Jehannine Austin, UoBritish Columbia

John McGrath, UQ

Tian Lin, UQ

Polygenic risk score applications

JAMA Psychiatry | Review

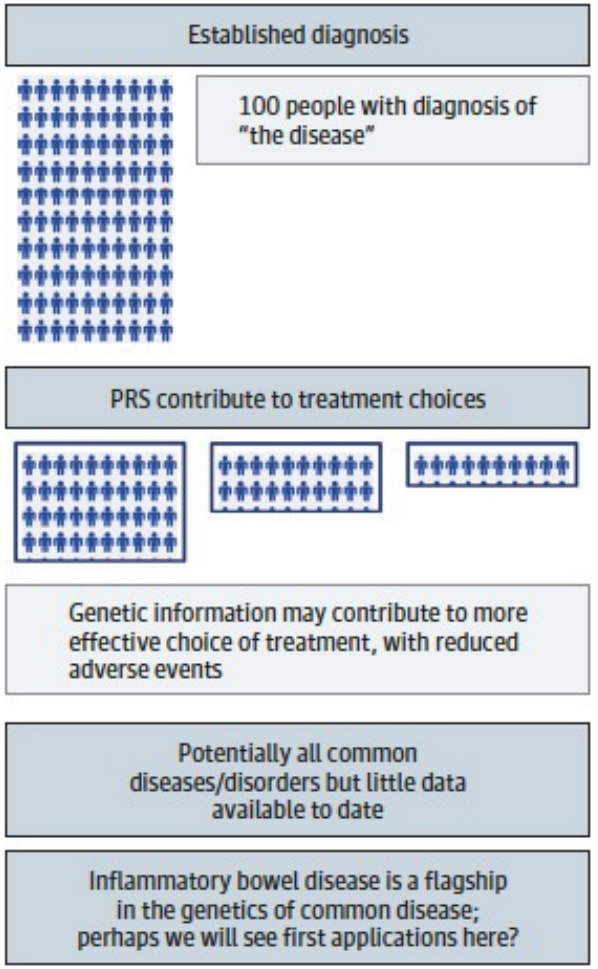
From Basic Science to Clinical Application of Polygenic Risk Scores
A Primer

Naomi R. Wray, PhD; Tian Lin, PhD; Jehannine Austin, PhD; John J. McGrath, MD, PhD; Ian B. Hickie, MD; Graham K. Murray, MD, PhD; Peter M. Visscher, PhD

Goal:

- Understandable by interested clinician
- Technically accurate – backed up in Supplement & Rscript

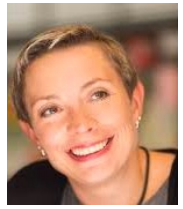
Cohort where PRS applied:	PRS could contribute to treatment choices, but more data are needed to allow development of PRS in this context.
Utility of PRS:	
Likely applications:	
Likely first applications:	



Naomi Wray, UQ & UoOxford

Graham Murray, UoCambridge

Ian Hickie, UoSydney



Jehannine Austin, UoBritish Columbia

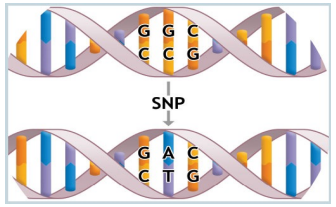
John McGrath, UQ

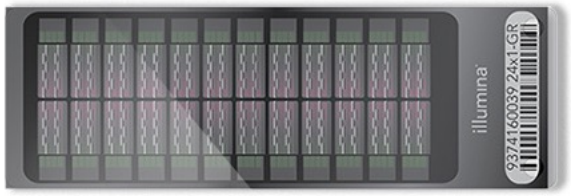
Tian Lin, UQ

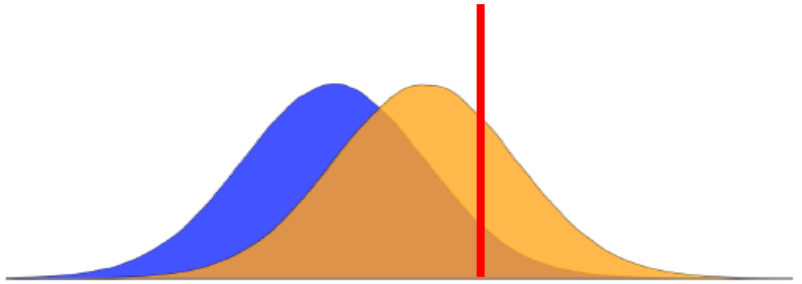
Justify for one disease and the rest come for free!

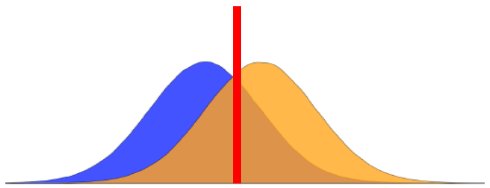
The University Of Queensland AUSTRALIA

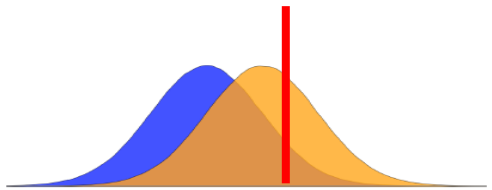
One disease

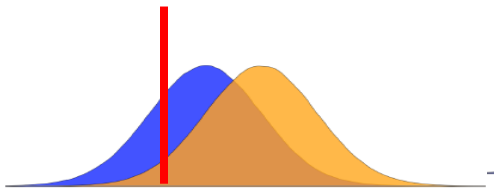


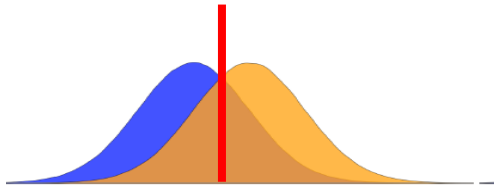


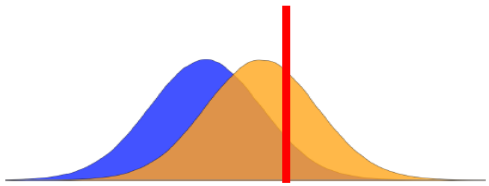


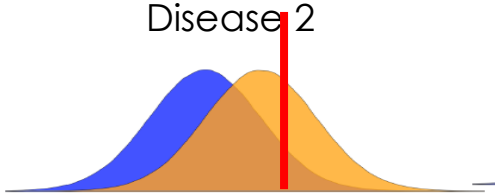

Disease 2

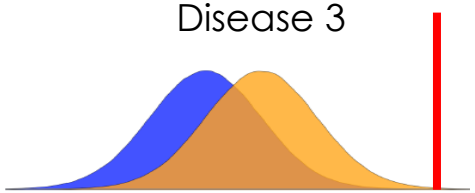

Disease 3

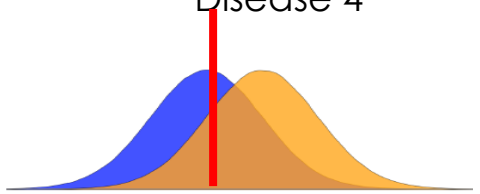

Disease 4

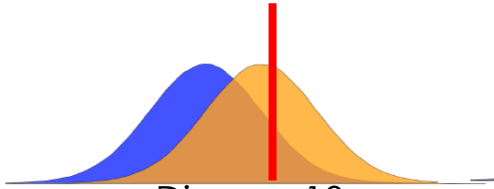

Disease 5

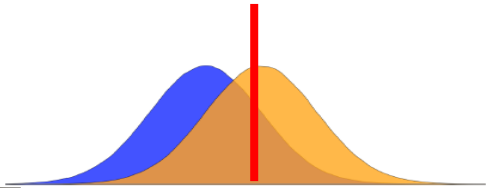

Disease 6

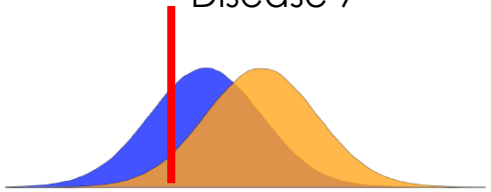

Disease 7

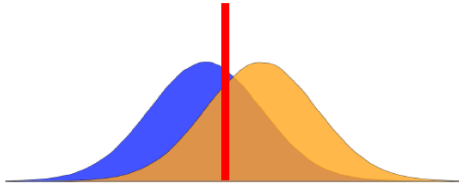

Disease 8

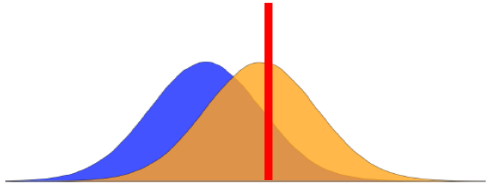

Disease 9

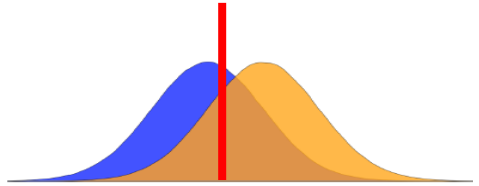

Disease 10

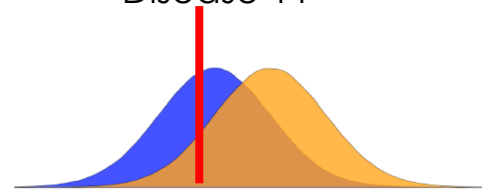

Disease 11


Disease 12


Disease 13


Disease 14


Disease 15


Disease 16

Methodology and challenges

Polygenic risk score methods

A weighted sum of the count of risk alleles

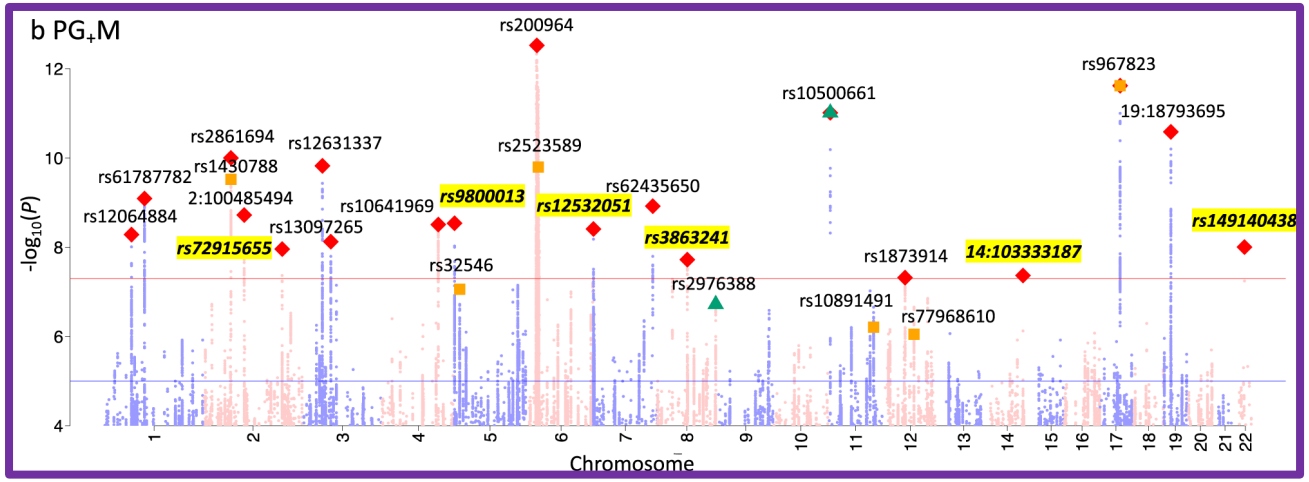
$$\text{PRS} = \widehat{\beta_1} x_{i1} + \widehat{\beta_2} x_{i2} + \widehat{\beta_3} x_{i3} + \cdots = \sum_{j=1}^{n_{SNP}} \widehat{\beta_j} x_{ij}$$

How many SNPs?
Which SNPs?
What weights?

Basic method:

Clumping & P-value thresholding (C+PT):

- Select most associated SNP in tower – LD-based clumping
- Select on a p-value threshold



Polygenic risk score methods

A weighted sum of the count of risk alleles

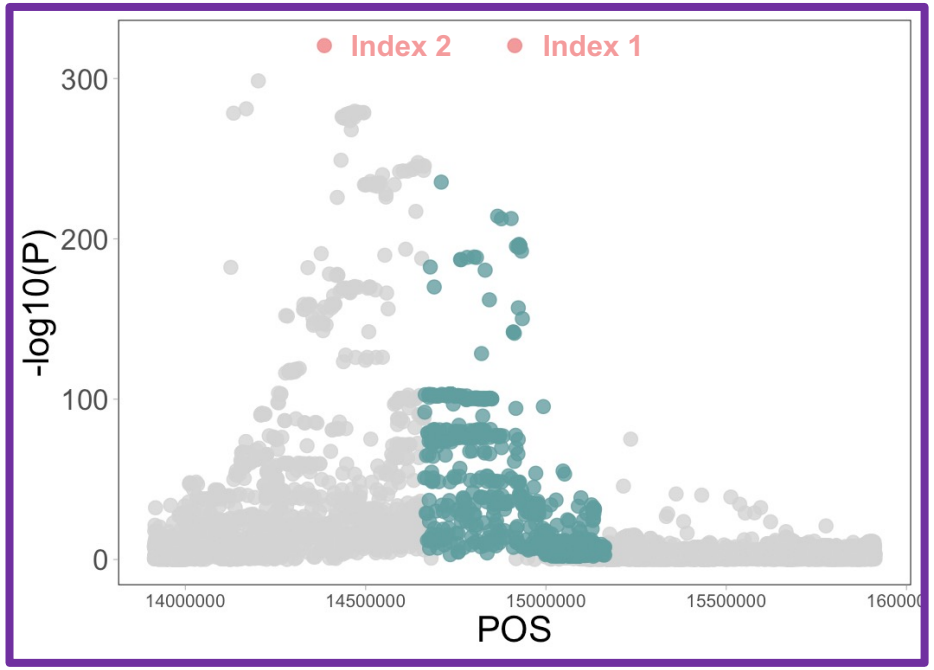
$$\text{PRS} = \widehat{\beta_1} x_{i1} + \widehat{\beta_2} x_{i2} + \widehat{\beta_3} x_{i3} + \cdots = \sum_{j=1}^{n_{SNP}} \widehat{\beta_j} x_{ij}$$

How many SNPs?
Which SNPs?
What weights?

Basic method:

Clumping & P-value thresholding (C+PT):

- Select most associated SNP in tower – LD-based clumping
- Select on a p-value threshold



Polygenic risk score methods

A weighted sum of the count of risk alleles

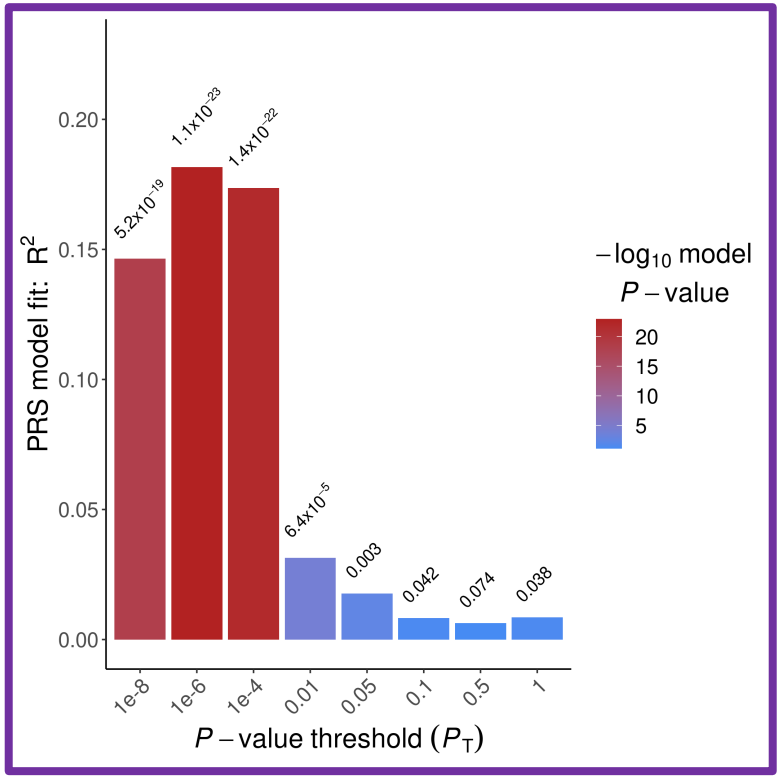
$$\text{PRS} = \widehat{\beta_1} x_{i1} + \widehat{\beta_2} x_{i2} + \widehat{\beta_3} x_{i3} + \cdots = \sum_{j=1}^{n_{SNP}} \widehat{\beta_j} x_{ij}$$

How many SNPs?
Which SNPs?
What weights?

Basic method:

Clumping & P-value thresholding (C+PT):

- Select most associated SNP in tower – LD-based clumping
- Select on a p-value threshold



Polygenic risk score methods

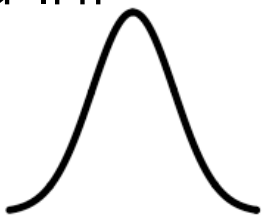
A weighted sum of the count of risk alleles

$$\text{PRS} = \widehat{\beta_1} x_{i1} + \widehat{\beta_2} x_{i2} + \widehat{\beta_3} x_{i3} + \cdots = \sum_{j=1}^{n_{SNP}} \widehat{\beta_j} x_{ij}$$

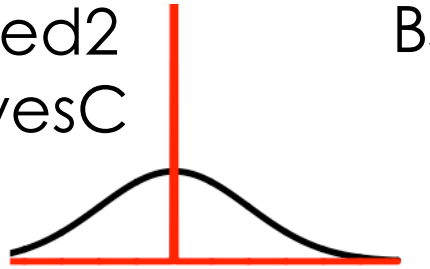
How many SNPs?
Which SNPs?
What weights?

New methods model genetic architecture

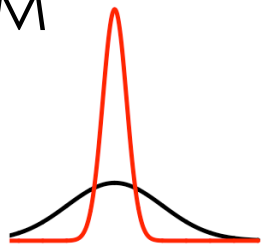
LDpred-Inf
SBLUP



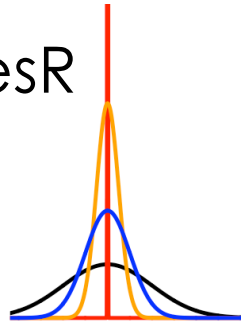
LDPred2
SBayesC



BSLMM



SBayesR



Polygenic risk score methods

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Table 1. Summary of Methods Used to Generate Polygenic Scores

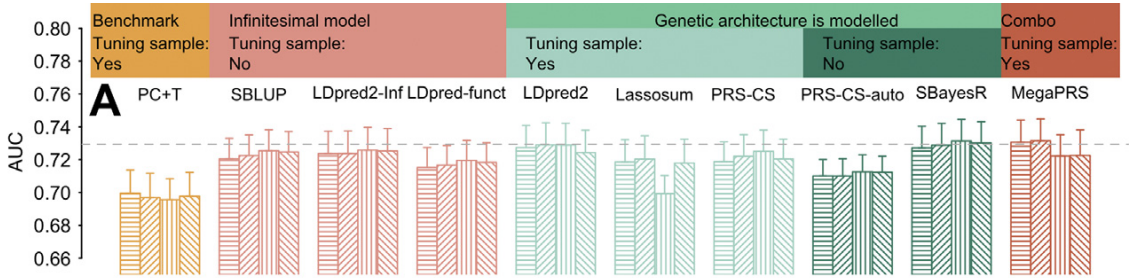
Method	Distribution of SNP Effects (β)	Tuning Sample	Predefined Parameters	Parameters Estimated in Tuning Sample
PC+T	None	Yes	–	*p*-value threshold
SBLUP	$\beta \sim N\left(0, \frac{h_g^2}{m}\right)$ h_g^2: SNP-based heritability, m: number of SNPs; $\lambda = m(1 - h_g^2)/h_g^2$	No	λ LD radius in kb	–
Ldpred2-Inf	Same as SBLUP	No	h_g^2 LD radius in cM or kb	–
LDpred-funct	$\beta_j \sim N(0, c\sigma_j^2)$ $\sum_{j=1}^{M} 1_{\sigma_j^2>0} c\sigma_j^2 = h_g^2$, c is a normalizing constant, σ_j^2 is the expected per SNP heritability under the baseline-LD annotation model estimated by stratified LDSC from the discovery GWAS within LDpred-funct software	No	h_g^2 LD radius in number of SNPs	–
LDpred2	$\beta_j \sim \begin{cases} N\left(0, \frac{h_g^2}{\pi m}\right), & \text{with probability of } \pi \\ 0, & \text{with probability of } 1-\pi \end{cases}$ When sparsity is "true," the β_j for SNPs in the $(1 - \pi)$ partition are all set to zero	Yes	h_g^2 π software default values, LD radius in cM or kb	π, sparsity
Lassosum	$f(\boldsymbol{\beta}) = \mathbf{y}^T\mathbf{y} + (1-s)\boldsymbol{\beta}^T\mathbf{X}_i^T\mathbf{X}_i\boldsymbol{\beta} - 2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y} + s\boldsymbol{\beta}^T\boldsymbol{\beta} + 2\lambda\|\boldsymbol{\beta}\|_1^1$ $\mathbf{X}_i$: $n \times m$ matrix of genotypes of LD reference sample, where n is sample size	Yes	LD blocks	λ, s
PRS-CS	$\beta_j \sim N\left(0, \frac{\sigma^2}{n}\psi_j\right)$ $\psi_j \sim G(a, \delta_j)$ $\delta_j \sim G(b, \phi)$, ϕ is a global scaling parameter	Yes	$a = 1$, $b = 0.5$ n LD blocks	ϕ
PRS-CS-auto	Same as PRS-CS, but estimates ϕ from the discovery GWAS	No	$a = 1$, $b = 0.5$ n LD blocks	–
SBayesR	$\beta_j \mid \boldsymbol{\pi}, \sigma_\beta^2 \sim \begin{cases} 0, & \text{with probability of } \pi_1 \\ N(0, \gamma_2\sigma_\beta^2), & \text{with probability of } \pi_2 \\ \vdots \\ N(0, \gamma_C\sigma_\beta^2), & \text{with probability of } 1 - \sum_{c=1}^{C-1}\pi_c \end{cases}$ $\sigma_\beta^2 \sim Inv - \chi^2(d.f. = 4)$ $\pi_i \sim Dir(\mathbf{1})$, estimated from discovery GWAS in SBayesR software γ_i are scaling parameters	No	LD radius in cM or kb C = 4 γ software default values	–
MegaPRS	Lasso: $\beta_j \sim DE(\lambda/\sigma_j)$ Ridge regression: $\beta_j \sim N(0, v\sigma_j^2)$ BOLT-LMM: $\beta_j \sim \begin{cases} N\left(0, \frac{(1-f_2)\sigma_j^2}{\pi}\right), & \text{with probability of } \pi \\ N\left(0, \frac{f_2\sigma_j^2}{1-\pi}\right), & \text{with probability of } 1-\pi \end{cases}$ f_2 is the proportion of the total mixture variance in the second normal distribution BayesR: similar to SBayesR with C = 4, and π_i and γ_i estimated in the tuning sample σ_j^2 is the expected per SNP-heritability under BLD-LDAK model using SumHer	Yes	LD radius in cM or kb Parameters used in BLD-LDAK Grid search parameter values for each method	The tuning cohort is used to estimate the parameters that maximize prediction for each model, and from these the model that maximizes prediction is selected

Archival Report

Biological Psychiatry

A Comparison of Ten Polygenic Score Methods for Psychiatric Disorders Applied Across Multiple Cohorts

Guiyan Ni, Jian Zeng, Joana A. Revez, Ying Wang, Zhili Zheng, Tian Ge, Restuadi Restuadi, Jacqueline Kiewa, Dale R. Nyholt, Jonathan R.I. Coleman, Jordan W. Smoller, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Jian Yang, Peter M. Visscher, and Naomi R. Wray

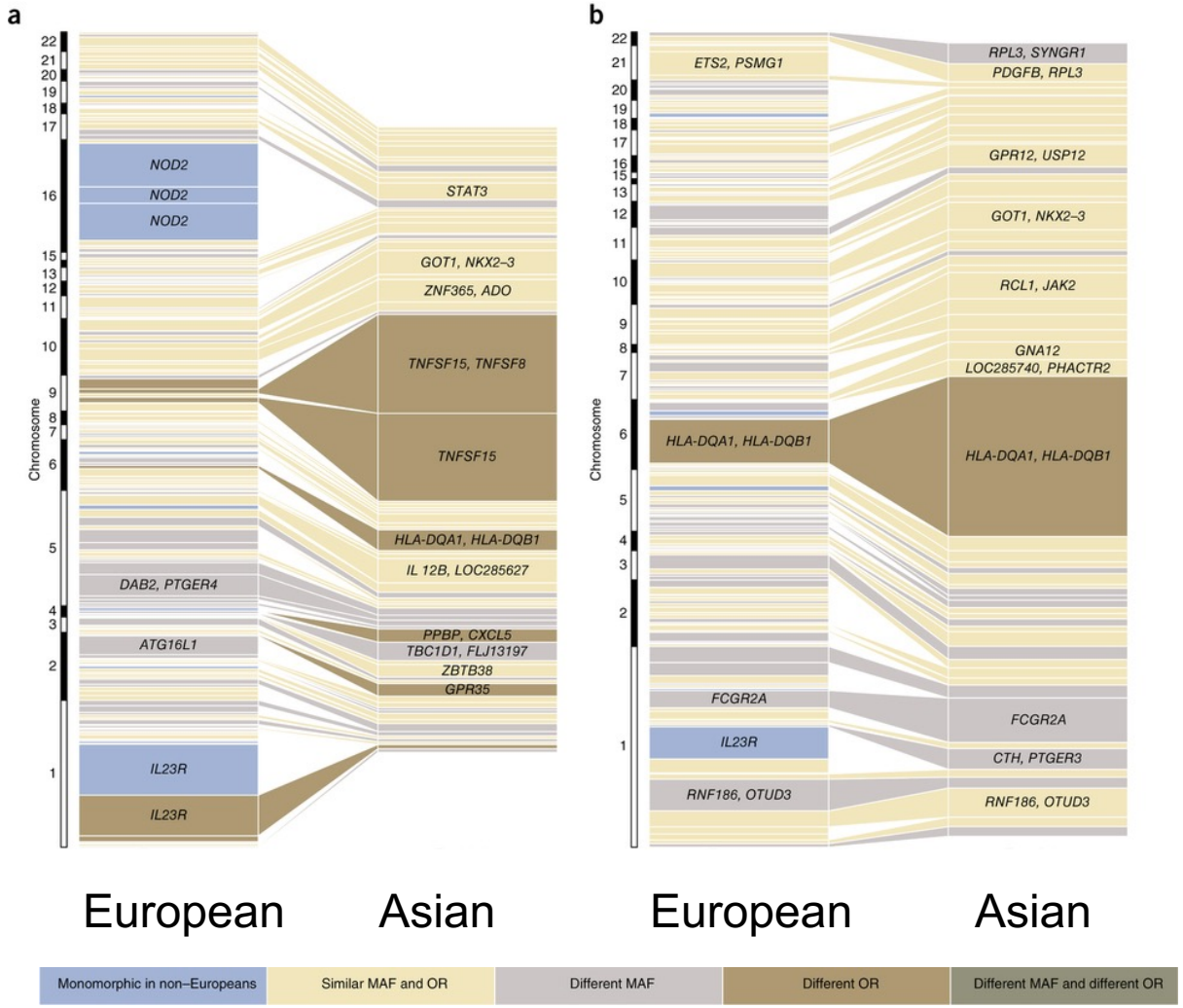


- Random effects models > fixed effects models
- Mixture models > non-mixture (infinitesimal) models

Ancestry

Crohn's Disease
r_g EUR-ASN 0.76

Ulcerative Colitis
r_g EUR-ASN 0.79

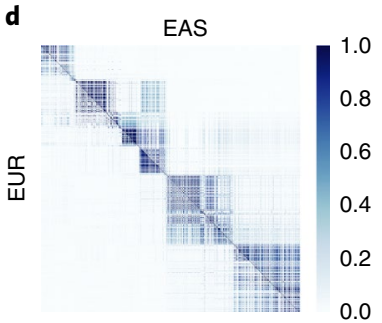


Issues

- Same causal variants
  - Different allele frequencies
  - LD differences
  - Different effect sizes
- Different causal variants
  - GxE
  - Different phenotype

In *general:*
We expect common causal variants to be shared across ancestries

But correlation structure differs



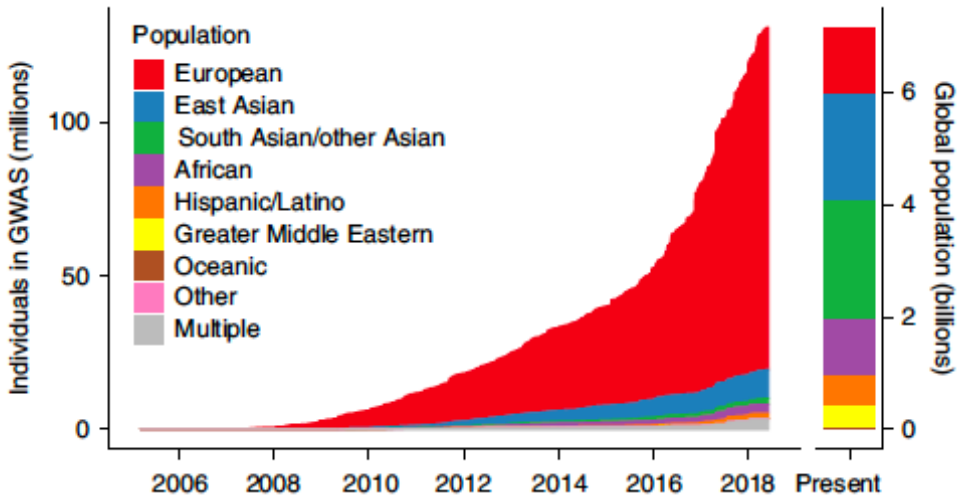
Liu et al (2015) Association analysis identifies 38 susceptibility loci for IBD and highlight shared genetic risk across populations. Nat Gen 2015

Ancestry

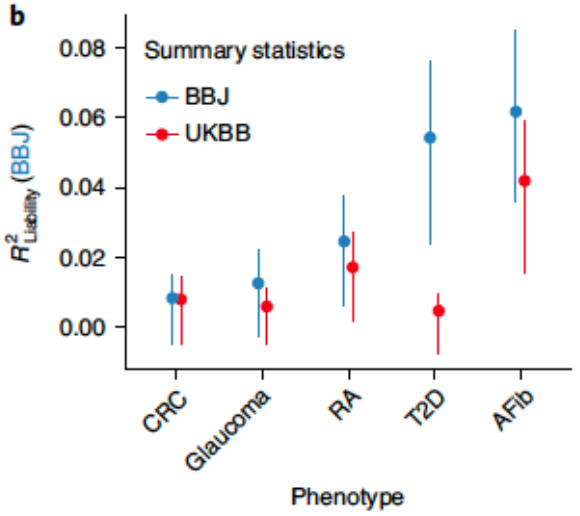
PERSPECTIVE
https://doi.org/10.1038/s41588-019-0379-x
nature genetics

Clinical use of current polygenic risk scores may exacerbate health disparities

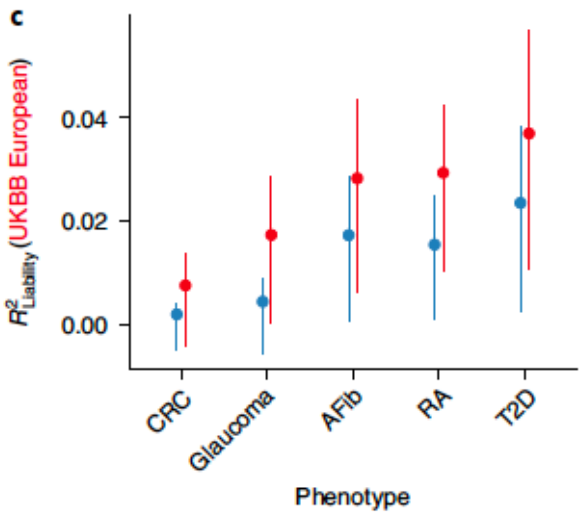
Alicia R. Martin[1,2,3*], Masahiro Kanai[1,2,3,4,5], Yoichiro Kamatani[5,6], Yukinori Okada[5,7,8], Benjamin M. Neale[1,2,3] and Mark J. Daly[1,2,3,9]

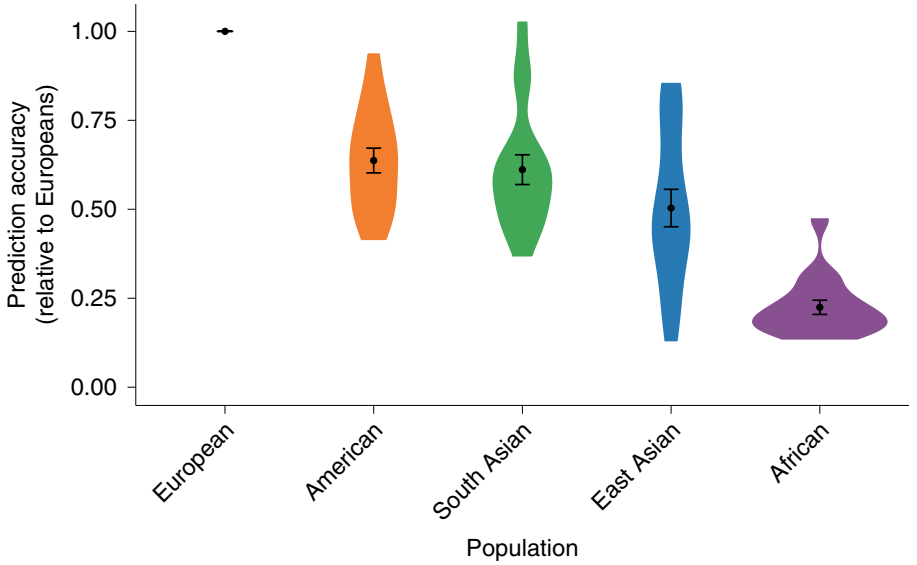


Predicted into Japanese



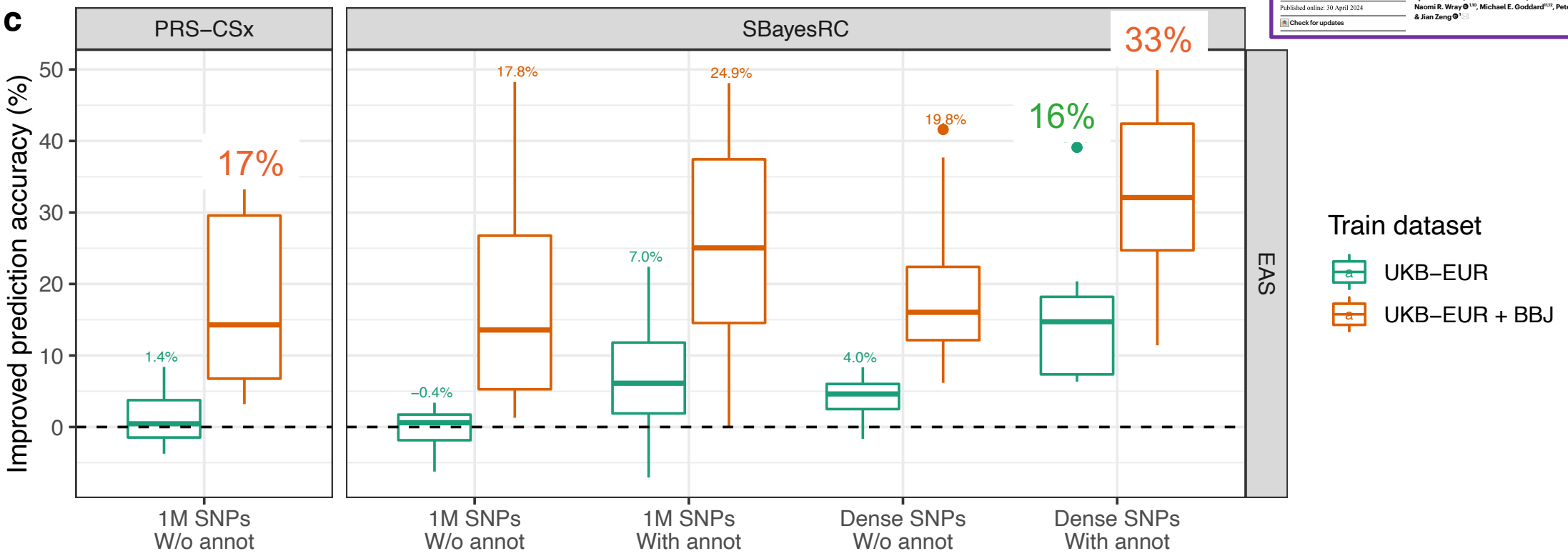
Predicted into European

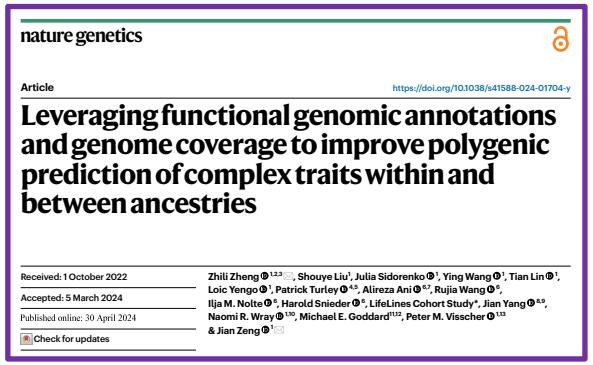




Trans-ancestry prediction

Use GWAS data from UK Biobank (UKB) European samples and Biobank Japan East Asian (EAS) samples to predict UKB EAS




nature genetics

Article

https://doi.org/10.1038/s41588-024-01704-y

Leveraging functional genomic annotations and genome coverage to improve polygenic prediction of complex traits within and between ancestries

Received: 1 October 2022

Accepted: 5 March 2024

Published online: 30 April 2024

Check for updates

Zhili Zheng, Shouye Liu, Julia Sidorenko, Ying Wang, Tian Lin, Loic Yengo, Patrick Turley, Alireza Ani, Rujia Wang, Ilja M. Nolte, Harold Snieder, LifeLines Cohort Study, Jian Yang, Naomi R. Wray, Michael E. Goddard, Peter M. Visscher & Jian Zeng

Summary

PGS are …

- Imperfect genetic predictors with inherently limited accuracy.
- Being evaluated in clinical settings and are often combined with other predictive measures to predict the total disease risk.
- Useful in risk stratification to better triage people into established screening programs.
- Available for an individual for all common diseases from birth.
- being improved with more data and better methods, especially for its utility across ancestries.

PGS are not …

- Not diagnostic.
- Not absolute risk and do not provide a baseline or timeframe for the progression of a disease.
- Not and never will be stand-alone predictors of common diseases.
- PGS accuracy will increase with GWAS sample size but are never going to be able to definitively predict complex conditions.

Practical 1: Computation of PRS using C+PT

https://cnsgenomics.com/data/teaching/GNGWS24/module5/Practical1_PRS.html

To log into your server, type command below in **Terminal** for Mac/Linux users or in **Command Prompt** or **PowerShell** for Windows users.

```
ssh username@hostname
```

And then key in the provided password.