

# Going Further: Introduction to Polygenic Risk Scores

 [jodie.lord@kcl.ac.uk](mailto:jodie.lord@kcl.ac.uk)

 @JodieLord5

# Learning Objectives

- Understand the multiple testing challenge facing GWAS approaches.
- Grasp the concept of “Polygenic Risk Scores” and understand how they can help to provide additional insights to standard GWAS outputs.
- Gain knowledge in two approaches to polygenic scoring:
  1. Within-trait genetic signal
  2. Cross-trait genetic overlap
- Understand how polygenic scores can be utilized to integrate knowledge from genetics with other omic modalities.

# Learning Objectives

- Understand the multiple testing challenge facing GWAS approaches.
- Grasp the concept of “Polygenic Risk Scores” and understand how they can help to provide additional insights to standard GWAS outputs.
- Gain knowledge in two approaches to polygenic scoring:
  1. Within-trait genetic signal
  2. Cross-trait genetic overlap
- Understand how polygenic scores can be utilized to integrate knowledge from genetics with other omic modalities.

# Multiple Testing Challenge facing GWAS

---

- A reminder from the previous presentation...

# Multiple Testing Challenge facing GWAS

---

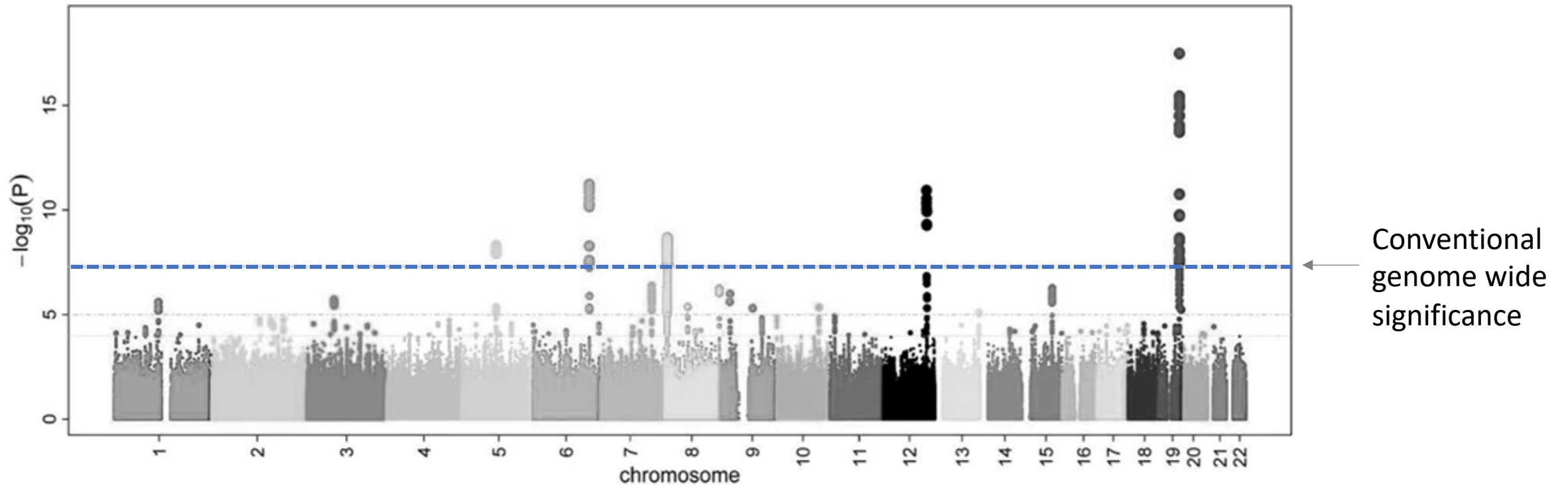
- A reminder from the previous presentation...

## Genome Wide Association Studies

- Because you perform pairwise associations between each SNP for a single sample → have to perform statistical corrections to correct for multiple-testing.
- This is typically a corrected threshold of  $5 \times 10^{-8}$ 
  - Conventional significance threshold = 0.05 (allowing for a 5% chance of a false association being identified).
  - As you perform more and more tests, the chances of you detecting an association simply by chance increases.
  - To correct for this increased likelihood of identifying an incorrect association, we divide the conventional significance threshold by the number of tests performed.
  - As there are typically ~1million SNP-phenotype associations tested within standard GWAS – we divide 0.05 by 1million, which =  $5 \times 10^{-8}$ .
  - The consequence of this = that in order to claim a “significant association”, the effect size of the association between the SNP and phenotype needs to be very large (which we know, from previous knowledge it is unlikely to be) or sample sizes need to be HUGE, so that we have enough power to see an effect if it is there.

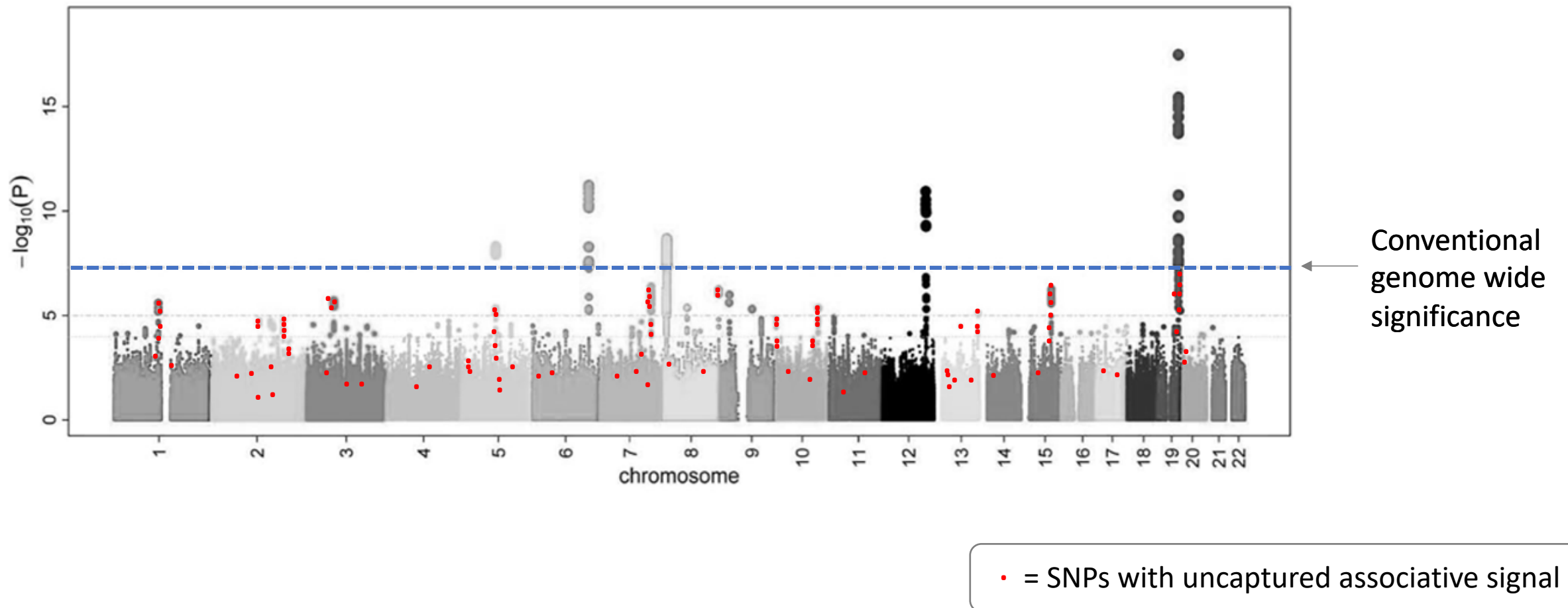
# Multiple Testing Challenge facing GWAS

---



# Multiple Testing Challenge facing GWAS

---

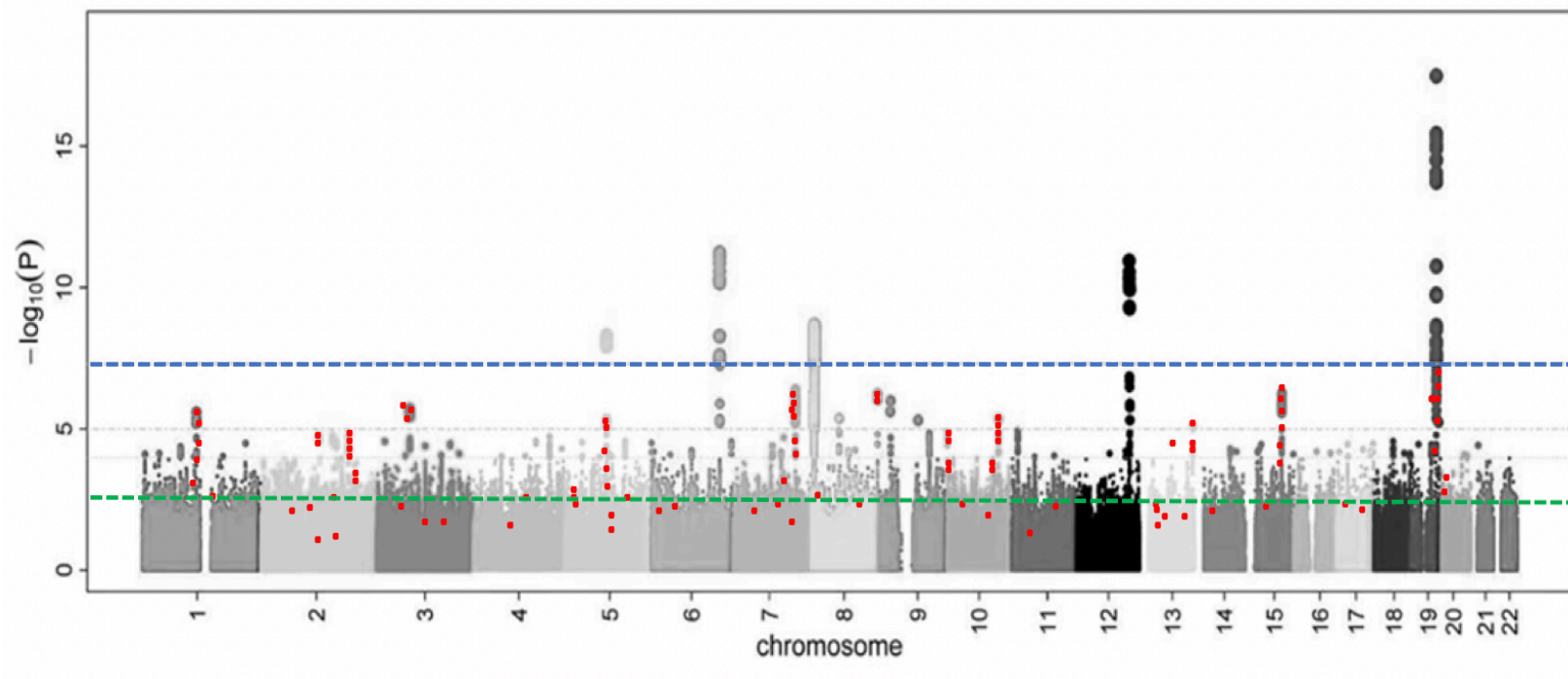


# Learning Objectives

- Understand the multiple testing challenge facing GWAS approaches.
- Grasp the concept of “Polygenic Risk Scores” and understand how they can help to provide additional insights to standard GWAS outputs.
- Gain knowledge in two approaches to polygenic scoring:
  1. Within-trait genetic signal
  2. Cross-trait genetic overlap
- Understand how polygenic scores can be utilized to integrate knowledge from genetics with other omic modalities.



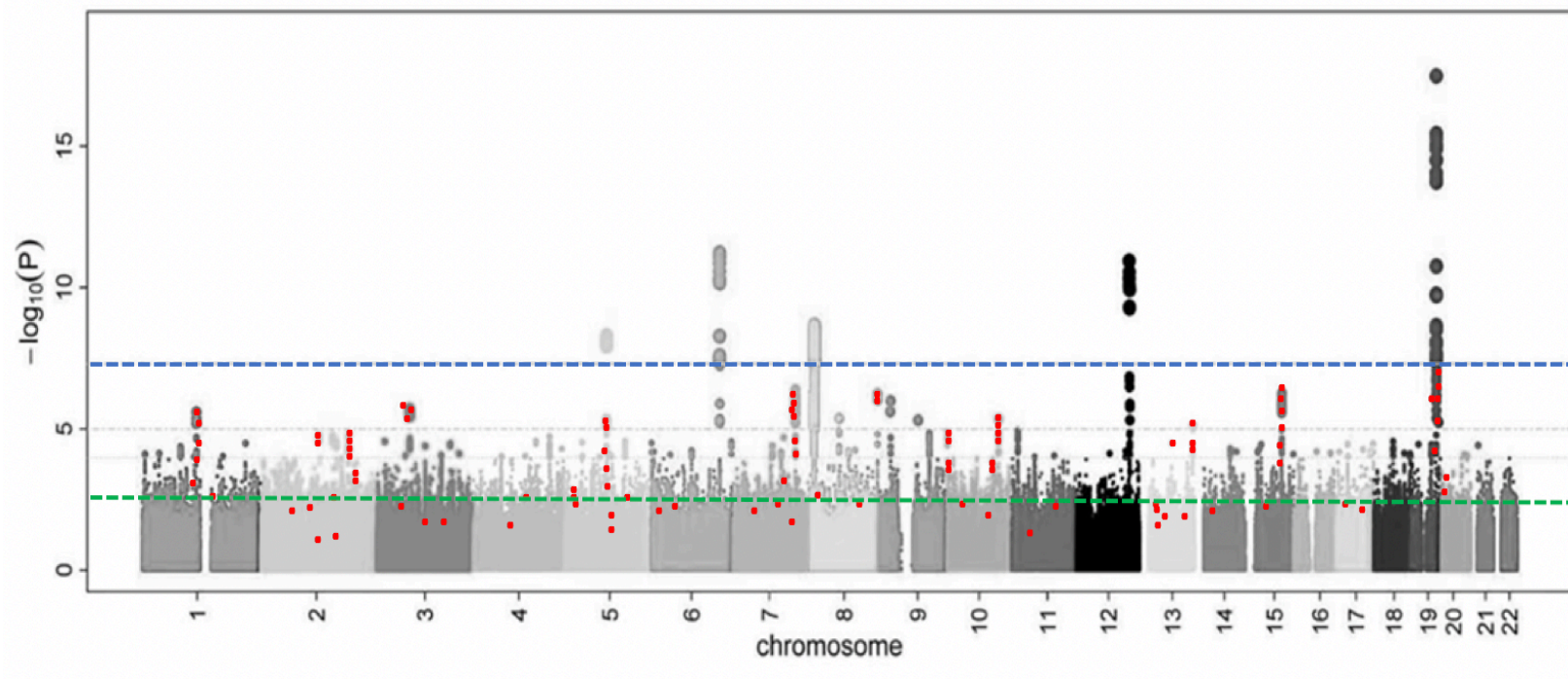
# What is a Polygenic Score?



SNP	Weight
rs3115860	1
rs3131970	0
rs3115857	1
rs2071111	2
rs528746615	4
Rs551793372	1

**Scoring  
model**

# What is a Polygenic Score?

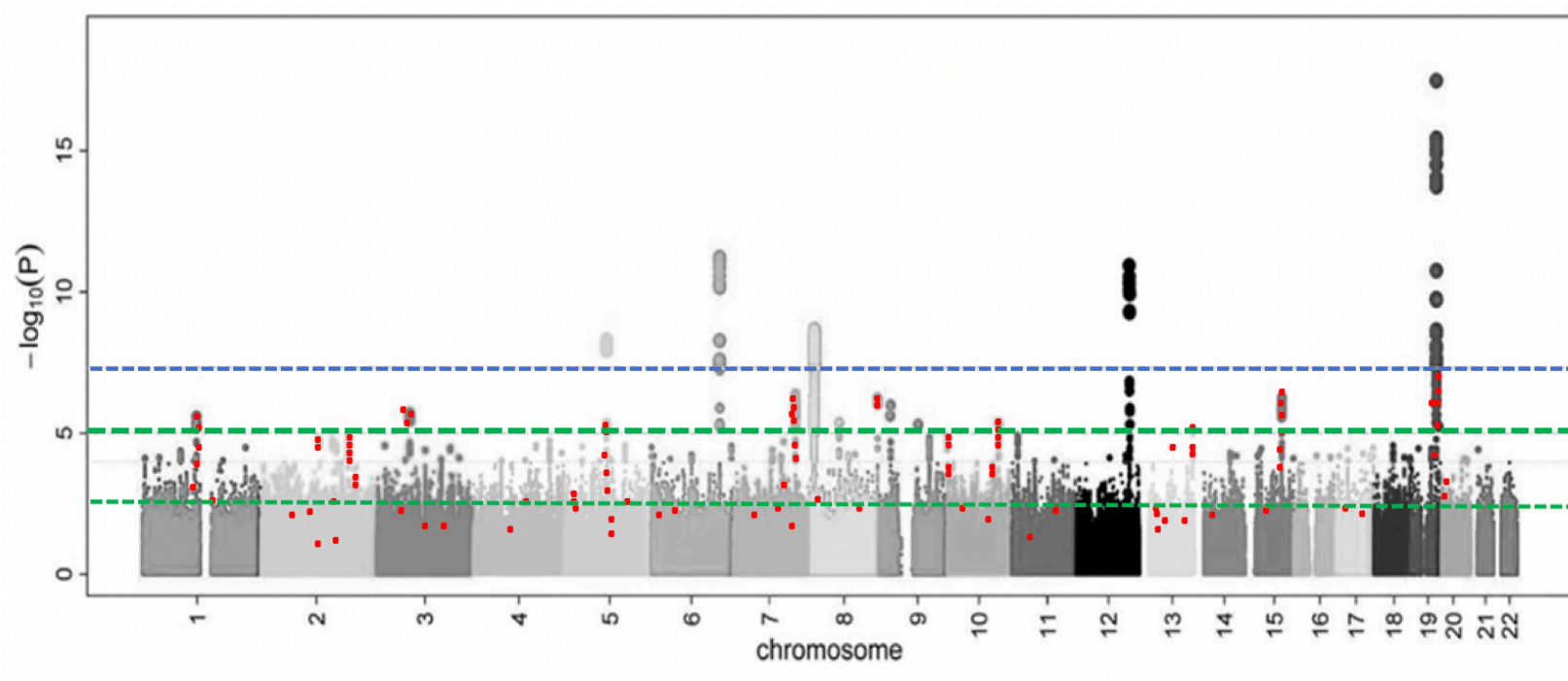


SNP	Weight
rs3115860	1
rs3131970	0
rs3115857	1
rs2071115	3
rs528746615	4
Rs551793372	1



Target  
phenotype

# What is a Polygenic Score?

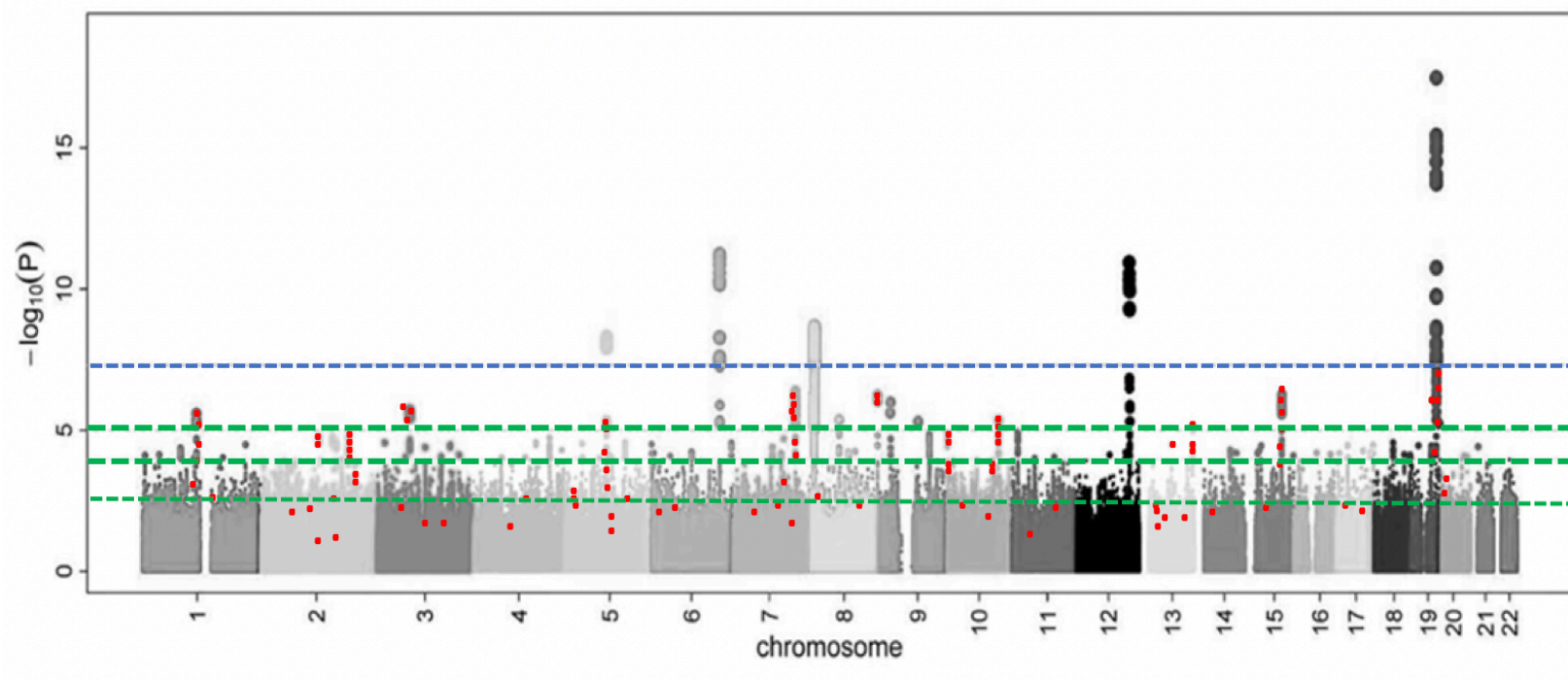


SNP	Weight
rs3115860	1
rs3131970	0
rs3115857	1
rs20746615	2
rs528746615	4
Rs551793372	1



**Target  
phenotype**

# What is a Polygenic Score?



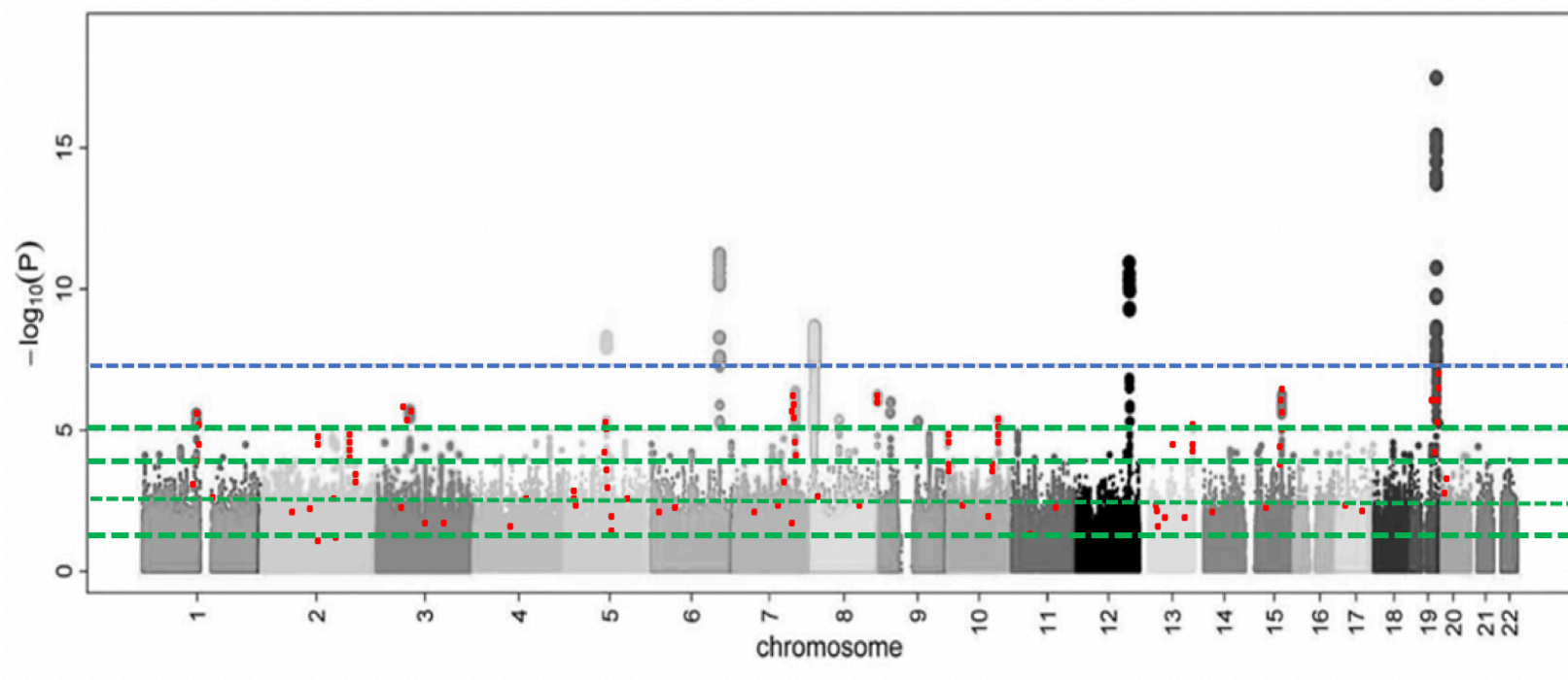
SNP	Weight
rs3115860	1
rs3131970	0
rs3115857	1
rs20746615	2
rs528746615	4
Rs551793372	1

**Scoring  
model**



**Target  
phenotype**

# What is a Polygenic Score?



SNP	Weight
rs3115860	1
rs3131970	0
rs3115857	1
rs2061155	2
rs528746615	4
Rs551793372	1



**Target  
phenotype**

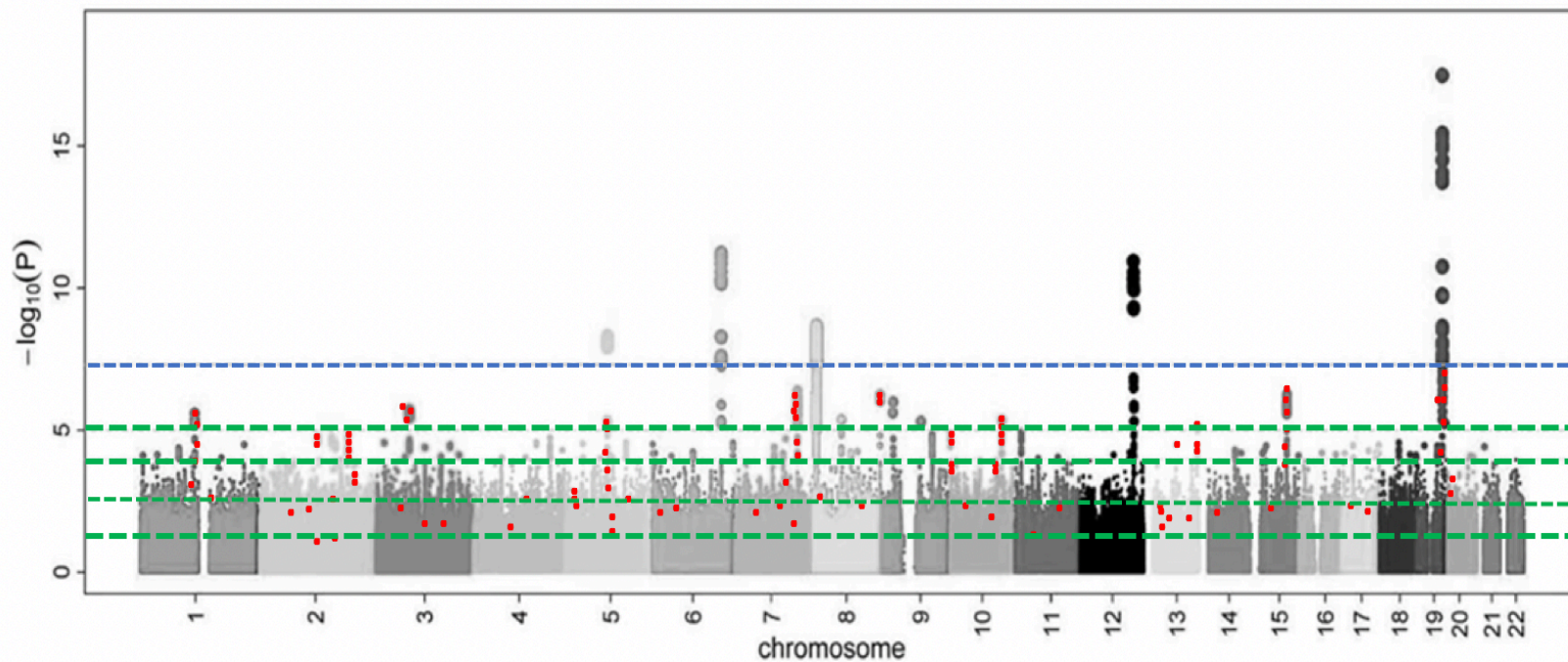
# Learning Objectives

- Understand the multiple testing challenge facing GWAS approaches.
- Grasp the concept of “Polygenic Risk Scores” and understand how they can help to provide additional insights to standard GWAS outputs.
- Gain knowledge in two approaches to polygenic scoring:
  1. Within-trait genetic signal
  2. Cross-trait genetic overlap
- Understand how polygenic scores can be utilized to integrate knowledge from genetics with other omic modalities.



# What is a Polygenic Score?

Alzheimer's Disease



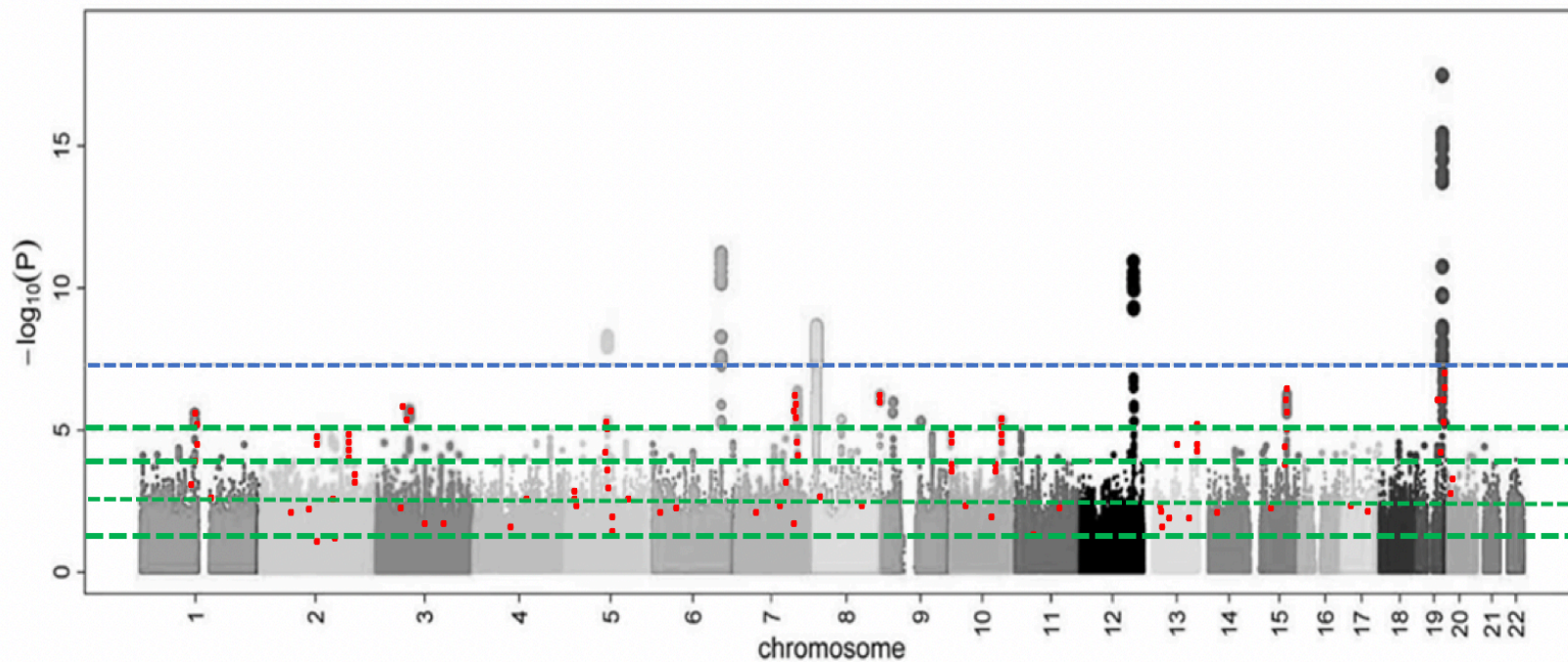
SNP	Weight
rs3115860	1
rs3131970	0
rs3115857	1
rs2070762	2
rs528746615	4
Rs551793372	1



Alzheimer's  
Disease

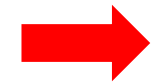
# What is a Polygenic Score?

Alzheimer's Disease



SNP	Weight
rs3115860	1
rs3131970	0
rs3115857	1
rs20	2
rs528746615	4
Rs551793372	1

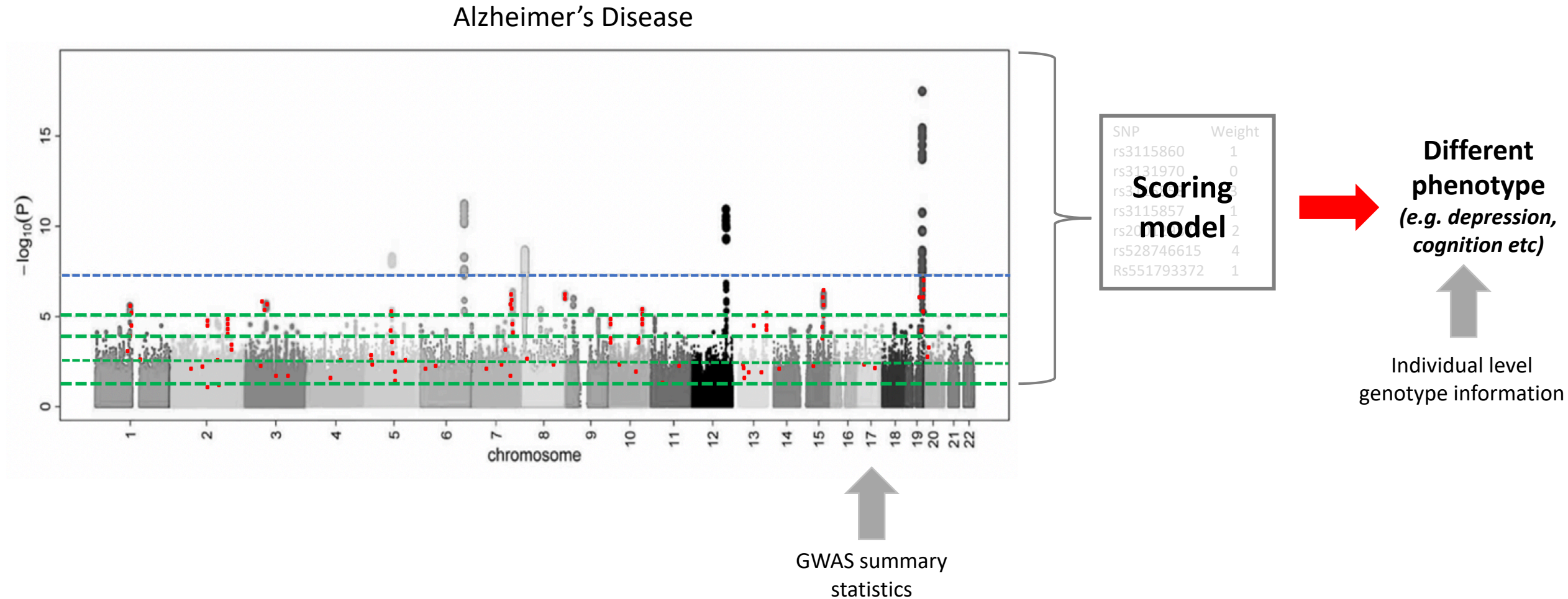
**Scoring  
model**



**Different  
phenotype**  
(e.g. depression,  
cognition etc)



# What is a Polygenic Score?



# Learning Objectives

- Understand the multiple testing challenge facing GWAS approaches.
- Grasp the concept of “Polygenic Risk Scores” and understand how they can help to provide additional insights to standard GWAS outputs.
- Gain knowledge in two approaches to polygenic scoring:
  1. Within-trait genetic signal
  2. Cross-trait genetic overlap
- Understand how polygenic scores can be utilized to integrate knowledge from genetics with other omic modalities.

# Working example:

---

- Using PRS to identify metabolites which show genetic overlap with Alzheimer's Disease
  - integrating metabolomic and genomic information to expand understanding of Alzheimer's.

# Working example: Cross-trait PRS: Metabolites & Alzheimer's

- Base Datasets:  
106 metabolite GWAS datasets from Kettunen et al., 2016.
- 10 p-value thresholds:  
5e-08 | 1e-05 | 1e-04 | 0.0001 | 0.001 | 0.01 | 0.05 | 0.1 | 0.2 | 0.5 | 1
- Target Datasets:  
**3 raw AD datasets (106\*3):**
  1. AddNeuroMed (N=648)
  2. ADNI (N=886)
  3. GERAD (N=3191).
- Covariates:  
Top 7 PCs, age, sex | ApoE genomic region removed.
- Random-effects **meta-analysis**

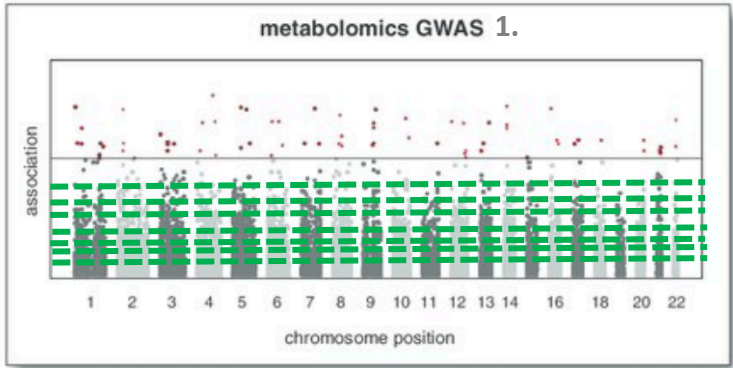


Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA

Johannes Kettunen , Ayşe Demirkan [...] Mika Ala-Korpela

*Nature Communications* 7, Article number: 11122 (2016) | [Download Citation](#)

- 123 GWAS datasets available.
- 17 removed due to low  $h^2$ .
- Blood metabolites extracted via Nuclear Magnetic Resonance.
- Total N = 24,925



**Scoring model**

SNP	Weight
rs3115860	1
rs3131970	0
rs3115857	1
rs20	2
rs528746615	4
Rs551793372	1



**AD dataset 1**



**AD dataset 2**



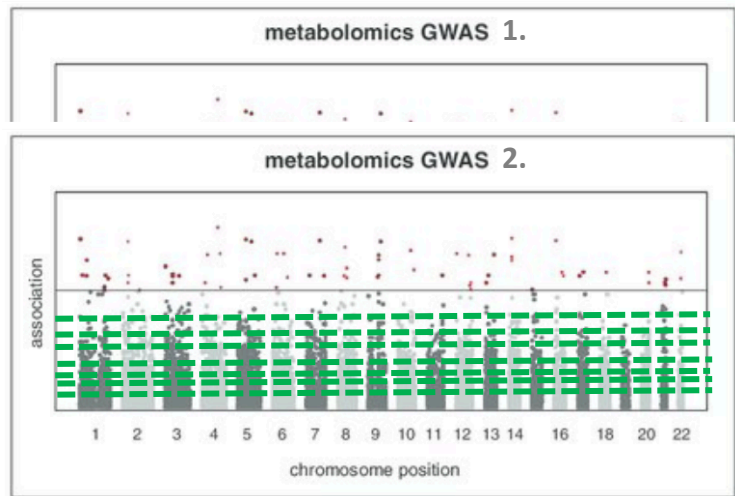
**AD dataset 3**



**Meta-analysed  
metabolite1 –  
AD polygenic  
score**  
**\*10**



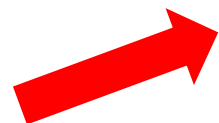
- 1.
- 2.
- 3.
- .....
- 10.



SNP	Weight
rs3115860	1
rs3131970	0

SNP	Weight
rs3115860	1
rs3131970	0
rs3115857	1
rs20	2
rs528746615	4
Rs551793372	1

**Scoring  
model**



**AD dataset 1**



**AD dataset 1**



**AD dataset 2**

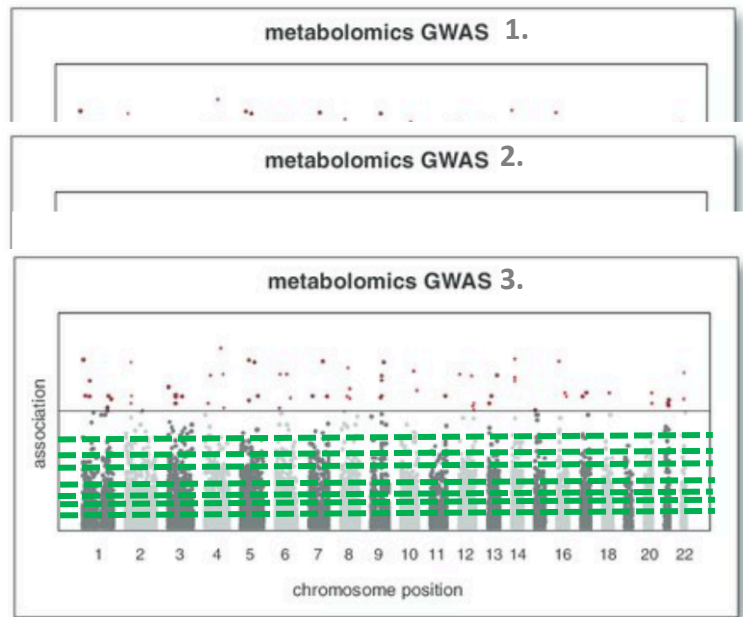


**AD dataset 3**

<b>Meta-analysed metabolite1 –</b>
<b>Meta-analysed metabolite2 – AD polygenic score *10</b>



<b>1.</b>
<b>2.</b>
<b>1.</b>
<b>2.</b>
<b>3.</b>
<b>.....</b>
<b>10.</b>



SNP	Weight
rs3115860	1
rs3131970	0

SNP	Weight
rs3115860	1

SNP	Weight
rs3115860	1
rs3131970	0
rs3115857	1
rs20	2
rs528746615	4
Rs551793372	1

**Scoring  
model**



**AD dataset 1**



**AD dataset 1**



**AD dataset 1**



**AD dataset 2**



**AD dataset 3**



**Meta-analysed  
metabolite1 –**

**Meta-analysed**

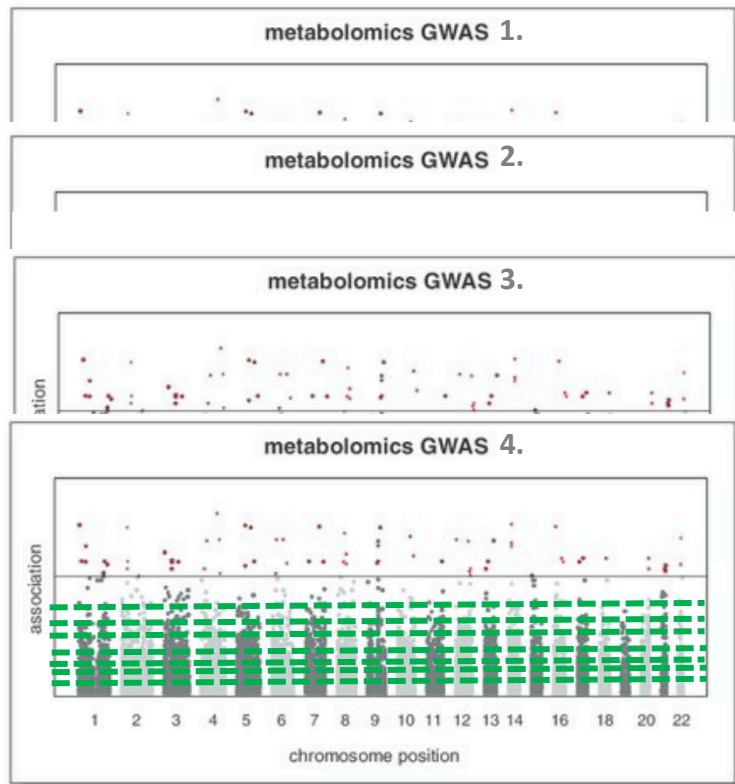
**Meta-analysed  
metabolite3 –  
AD polygenic  
score  
\*10**



**1.  
2.**

**1.**

**1.  
2.  
3.  
.....  
10.**



SNP	Weight
rs3115860	1
rs3131970	0

SNP	Weight
rs3115860	1

SNP	Weight
rs3115860	1
rs3131970	0
rs3115857	1
rs2000000	2
rs528746615	4
Rs551793372	1

**Scoring  
model**



**AD dataset 1**



**AD dataset 1**



**AD dataset 1**



**AD dataset 2**



**AD dataset 3**

**Meta-analysed  
metabolite1 –**

**Meta-analysed**

**Meta-analysed  
metabolite 4 –  
AD polygenic  
score  
\*10**

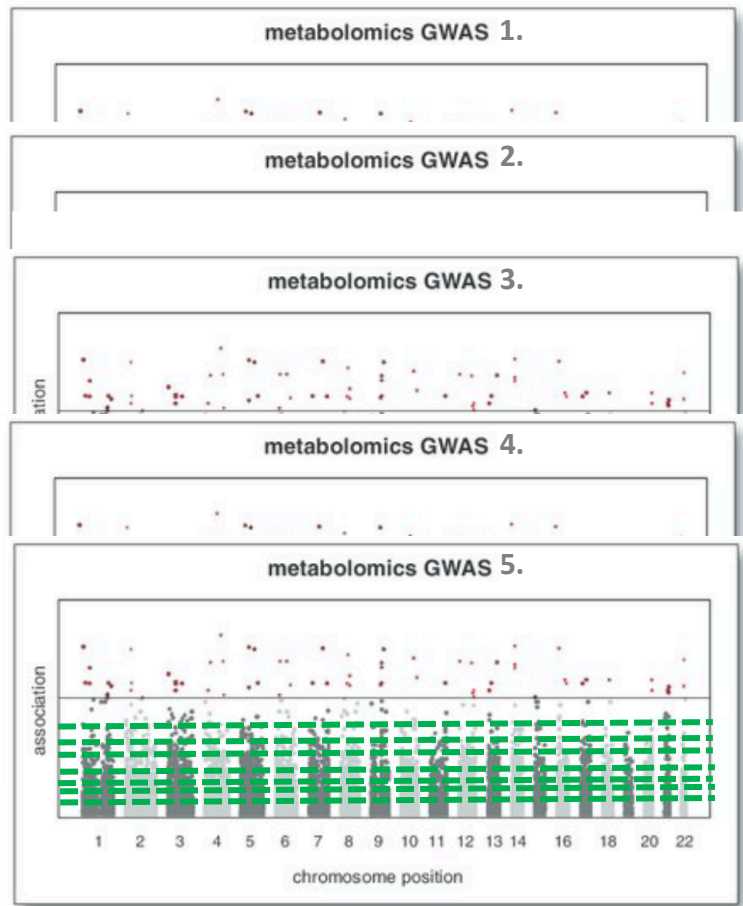
*Rank*

**1.  
2.**

**1.**

**1.  
2.  
3.  
.....  
10.**





SNP	Weight
rs3115860	1
rs3131970	0

SNP	Weight
rs3115860	1

SNP	Weight
rs3115860	1
rs3131970	0
rs3115857	1
rs2000000	2
rs528746615	4
Rs551793372	1

**Scoring  
model**



**AD dataset 1**



**AD dataset 1**



**AD dataset 1**



**AD dataset 2**



**AD dataset 3**

**Meta-analysed  
metabolite1 –**

**Meta-analysed**

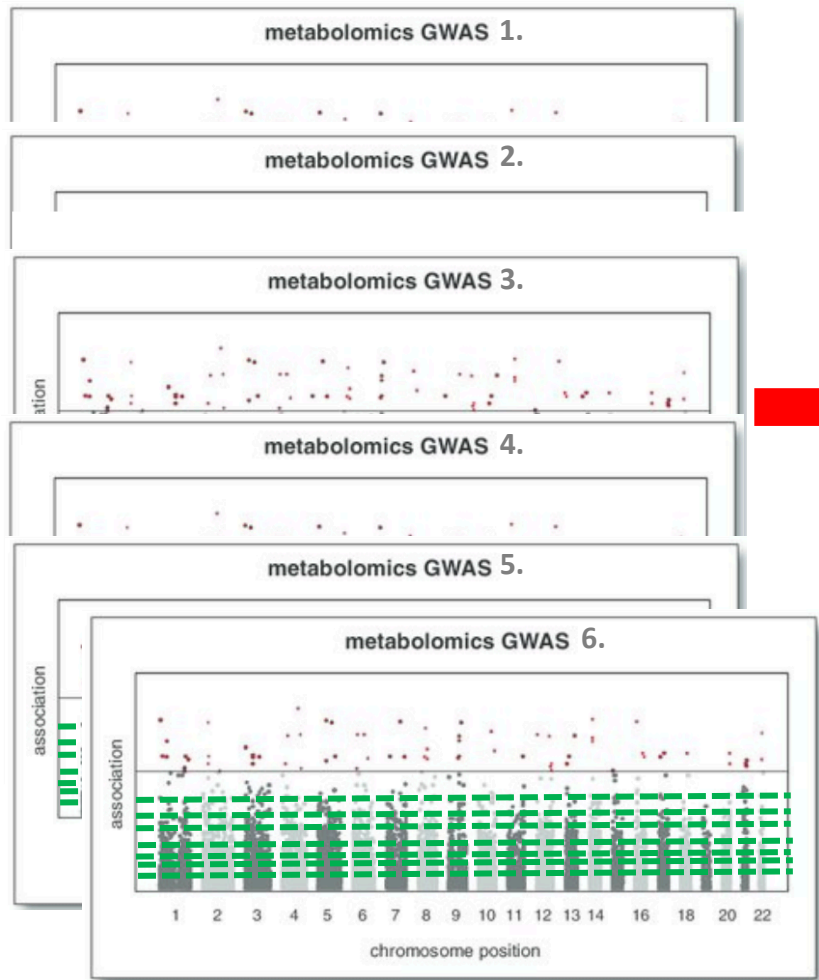
**Meta-analysed  
metabolite 4 –  
AD polygenic  
score  
\*10**

*Rank*

**1.  
2.**

**1.**

**1.  
2.  
3.  
.....  
10.**

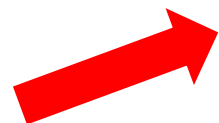


SNP	Weight
rs3115860	1
rs3131970	0

SNP	Weight
rs3115860	1

SNP	Weight
rs3115860	1
rs3131970	0
rs3115857	1
rs2000000	2
rs528746615	4
Rs551793372	1

**Scoring  
model**



**AD dataset 1**



**AD dataset 1**



**AD dataset 1**



**AD dataset 2**



**AD dataset 3**

**Meta-analysed  
metabolite1 –**

**Meta-analysed**

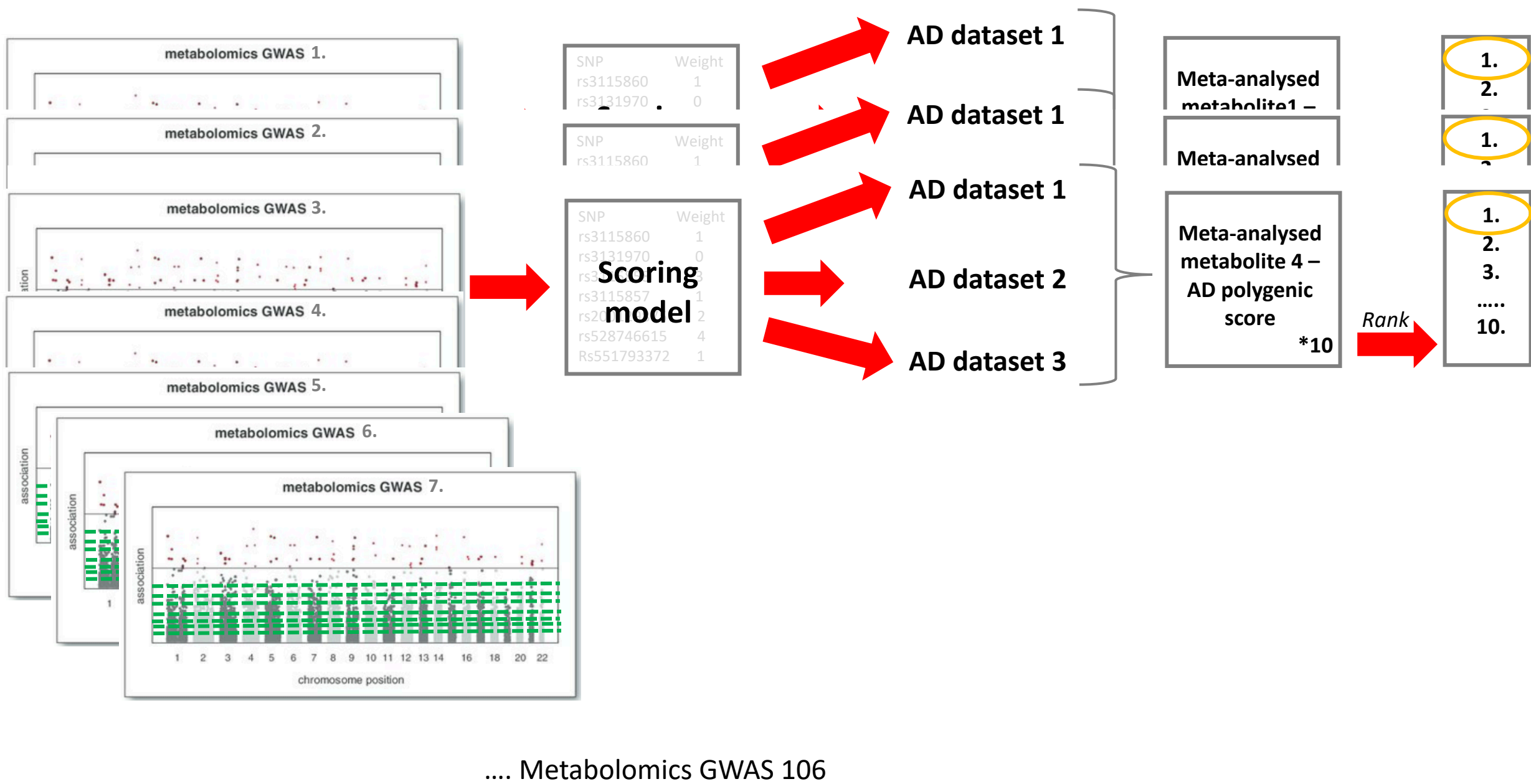
**Meta-analysed  
metabolite 4 –  
AD polygenic  
score  
\*10**

*Rank*

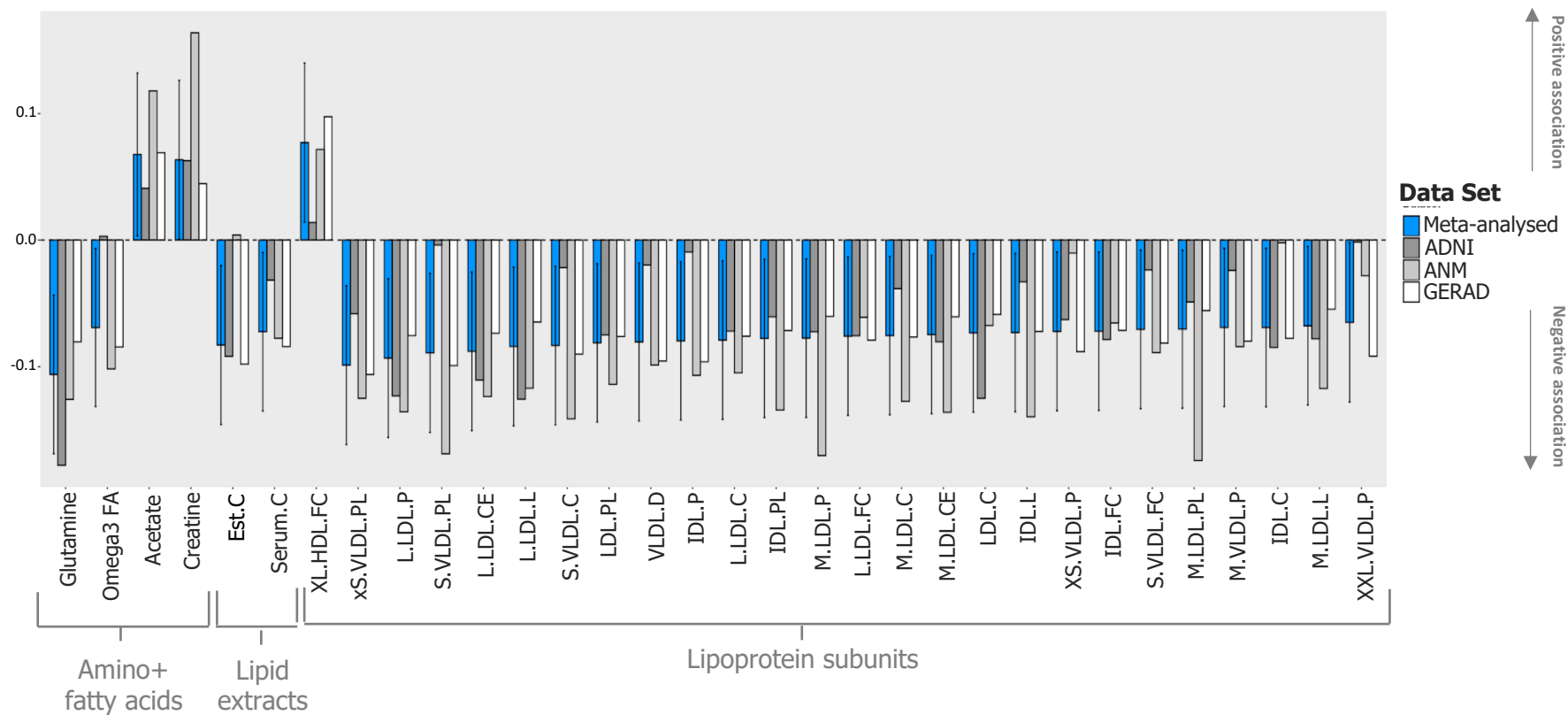
**1.  
2.**

**1.**

**1.  
2.  
3.  
.....  
10.**



# Results



34 metabolite-AD  
genetic associations at  
 $p < 0.05$

Majority nominal  
associations =  
lipoprotein subunits.

Glutamine = most sig  
genetic association with  
AD ( $\beta = -0.11$ ,  $se = 0.03$ ,  
 $p = 0.0009$ )

# Takeaways...

---



- Polygenic scores can be a useful way of uncovering signal below the hood of conventional genome-wide significance
- They can be used both within trait to get an idea of within-trait genetic signal, or across two different traits to get an ideas of potential genetic overlap across those traits.
- They can be a useful starting block in integrating information across omic modalities to look for clues of shared relationships which could indicate shared pathways (though subsequent steps required to take further).