Polygenic Risk Score (PRS)

X-chromosome aware PRS - Advances and Challenges

Drs. Wei Deng, Lei Sun, Andrew Paterson

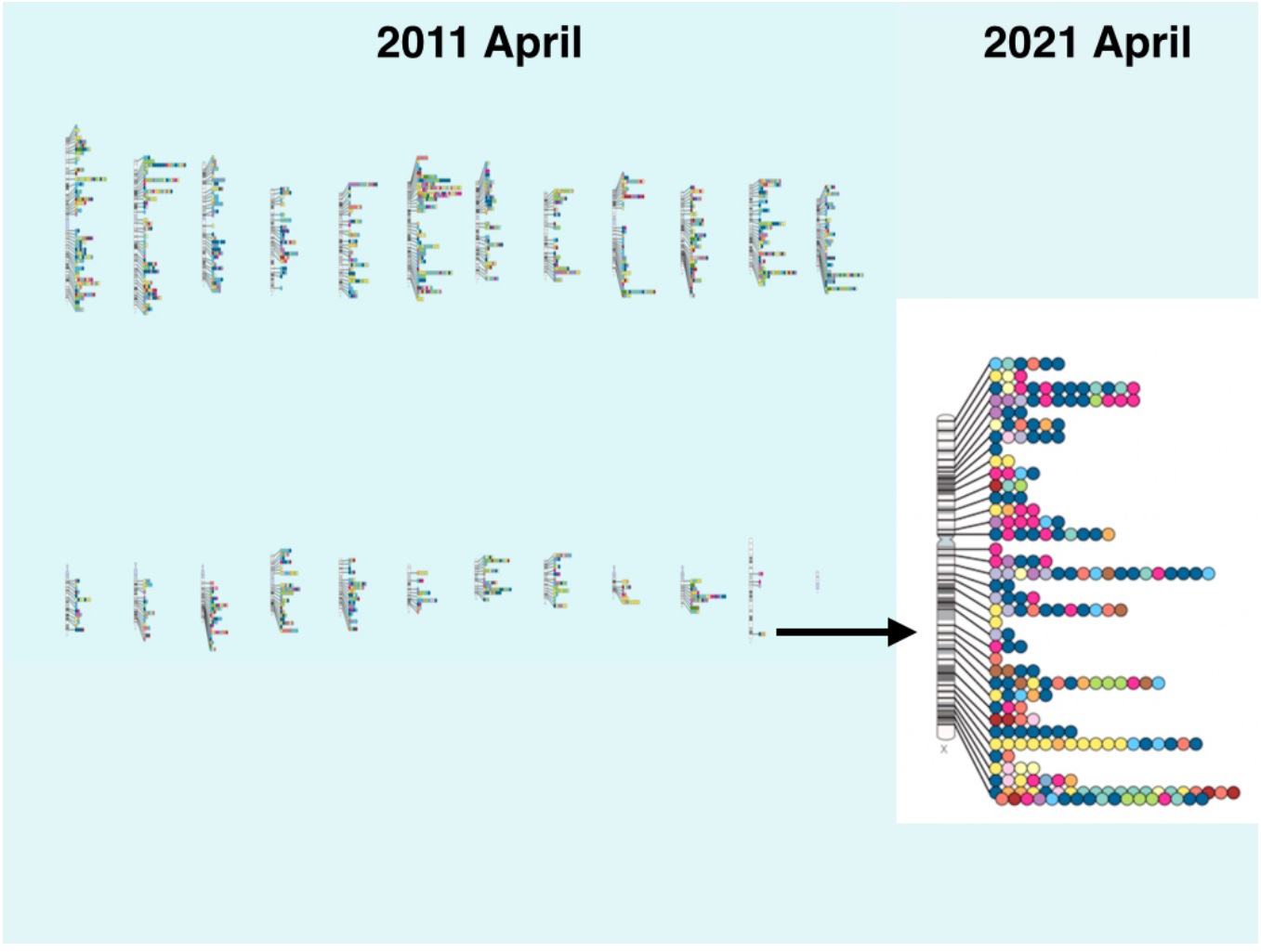
Department of Statistical Sciences, FAS

Division of Biostatistics, DLSPH

University of Toronto

12 October, 2021

An X-inclusive trend towards GWAS results reporting



Wise et al. (2013). *American Journal of Human Genetics*. eXclusion: toward integrating the X chromosome in genome-wide association analyses.

From autosomes to the X-chromosome (Xchr)

Quick facts:

- Xchr accounts for 5% of the nuclear genome (155Mb)
- 800 protein-coding genes
- Many X-linked (recessive) diseases

Characteristics:

- Female (two copies) vs. Male (one copy from mother);
- Dosage compensation via random Xchr-inactivation (XCI);
- Two pseudo-autosomal regions (PAR1 & PAR2) at termini of Xchr.

Dosage compensation and Xchr-inactivation (XCI)

Dosage compensation: a process to equalize expression of genes between sexes through **XCI** in females.

- Random XCI: randomly silence one copy of XX in females
- No XCI: some genes escape silencing
- Skewed XCI: preferentially silence one X over the other

Carrel et al (2005). *Nature*. X-inactivation profile reveals extensive variability in X-linked gene expression in females.

Tukiainen et al. (2017). *Nature*. Landscape of X chromosome inactivation across human tissues.

The same 'weighted sum' approach in principle, BUT

$$\text{X-}PRS_i \stackrel{?}{=} \sum_{j=1}^{J} \hat{\beta}_j G_{ij}$$

- X-PRS_i: Xchr PRS for individual i

  (sex-specific? combining males and females)

- J: the total number of "relevant" bi-allelic (common) SNPs

  (which **Xchr association test**; under no, random, or skewed XCI?)

- G_{ij}: the number of copies of the risk allele

  (counting a or A: **coding choices**?)

- $\hat{\beta}_j$: **estimated effect size of SNP** j

  (sex-specific or combined? **effect size interpretation** in the context of XCI uncertainty and G×Sex interaction effect!)

- $\sum$: (how to measure LD for Xchr?)

Coding choices: what does it mean analytically for G_{ij}?

G_j^M: Genotype of a (single allele, male Xchr) SNP j

- a = the reference allele
- A = the alternative allele
  (often the minor allele with MAF of p)
- coded (0, 1) for (a, A)

G_j^F: Genotype of a (bi-allelic, female Xchr) SNP j

- **no XCI**: (0, 1, 2) for (aa, Aa, AA)
- **random XCI**: (0, 0.5, 1) for (aa, Aa, AA)

  male G=(0, 2) and female G=(0, 1, 2), the PLINK default and
  male G=(0, 1) and female G=(0, 0.5, 1) are statistically the same.
- **skewed XCI**?

Modelling considerations in Xchr-inclusive GWAS

- status of XCI
- G_A: baseline allele a vs A
- S: sex as a confounder
- $G_A \times S$: interaction?
- G_D: dominance effect

Chen et al. (2021). *Genetic Epidemiology*. The X factor: a robust and powerful approach to X-chromosome-inclusive whole-genome association studies.

Coding and Modelling choice **critical** for Xchr!

Not including Sex as a covariate leads to

- Type I error issue; sex is an inherent confounder for Xchr
- Biased β estimation

Misspecified XCI leads to

- Inaccurate β testing
- Biased β estimation

This 3 d.f. test resolves above issues simultanuously

$$g(E(Y)) = \beta_0 + \beta_S Sex + \beta_A G_{Additive} + \beta_D G_{Dominance} + \beta_{GS} G \times Sex,$$

$$H_0 : \beta_A = \beta_D = \beta_{GS} = 0.$$

- Code the G_A whichever way you want, and the association testing results will be the same!
- XCI uncertainty is statistically equivalent to $G \times Sex$.
- XCI skewness is statistically equivalent to a dominance effect.

Model, $g(E(Y)) =$	Testing H_0 :	a vs A & S	XCI and $G \times S$	skewed XCI and G_D
$M_0 : \beta_0 + \beta_A G_A$	$\beta_A = 0$	$\times$	$\times$	$\times$
$M_1 : \beta_0 + \beta_S S + \beta_A G_A$	$\beta_A = 0$	$\surd$	$\times$	$\times$
$M_2 : \beta_0 + \beta_S S + \beta_A G_A + \beta_D G_D$	$\beta_A = \beta_D = 0$	$\surd$	$\times$	$\surd$
$M_3 : \beta_0 + \beta_S S + \beta_A G_A + \beta_{GS} GS$	$\beta_A = \beta_{GS} = 0$	$\surd$	$\surd$	$\times$
$M_4 : \beta_0 + \beta_S S + \beta_A G_A + \beta_D G_D + \beta_{GS} GS$	$\beta_A = \beta_D = \beta_{GS} = 0$	$\surd$	$\surd$	$\surd$

Chen et al. (2021). *Genetic Epidemiology*. The X factor: a robust and powerful approach to X-chromosome-inclusive whole-genome association studies.

Trouble in paradise

Chen et al. (2021). *Genetic Epidemiology*. The X factor: a robust and powerful approach to X-chromosome-inclusive whole-genome association studies.

"**GWAS alone cannot identify the underlying true genetic model**, because we have shown, for example, XCI uncertainty is analytically equivalent to a gene-sex interaction effect, while XCI skewness is analytically equivalent to dominance effect."

Song et al. (2021) *Genetic Epidemiology*. Testing and estimation of X-chromosome SNP effects: Impact of model assumptions.

"We demonstrated sex and **SNP coefficient biases in several situations**, particularly if the assumptions about XCI made by the coding scheme used and the assumptions made about sex differences in SNP effect of the fitted model were incorrect."

Numerical demonstration of the biased β estimates

Consider the commonly used model *without* G_D or $G \times Sex$,

$$g(E(Y)) = \beta_0 + \beta_S \text{Sex} + \beta_A G_{Additive}$$

			Estimated β_A			
			Bias		SD	
	β_S	β_A	XCI fit	no XCI fit	XCI fit	no XCI fit
XCI	0	0.75	0.006	0.012	0.074	0.110
	0.1	0.5	0.006	0.040	0.079	0.109
	0.2	0.2	0.012	0.081	0.083	0.120
	0.5	0.1	0.005	0.179	0.073	0.106
	0.75	0	0.009	0.278	0.090	0.128
no XCI	0	0.75	-0.002	-0.008	0.076	0.104
	0.1	0.5	-0.026	0.011	0.078	0.110
	0.2	0.2	-0.077	-0.013	0.075	0.106
	0.5	0.1	-0.175	-0.012	0.078	0.115
	0.75	0	-0.251	0.009	0.076	0.108

Song et al. (2021) *Genetic Epidemiology*. Testing and estimation of X-chromosome SNP effects: Impact of model assumptions.

Xchr association analysis: only partially solved

Recommendation for association **testing**:

$$g(E(Y)) = \beta_0 + \beta_S Sex + \beta_A G_{Additive} + \beta_D G_{Dominance} + \beta_{GS} G \times Sex,$$

$$3 \text{ d.f. test of } H_0 : \beta_A = \beta_D = \beta_{GS} = 0$$

Recommendation for association **reporting**:

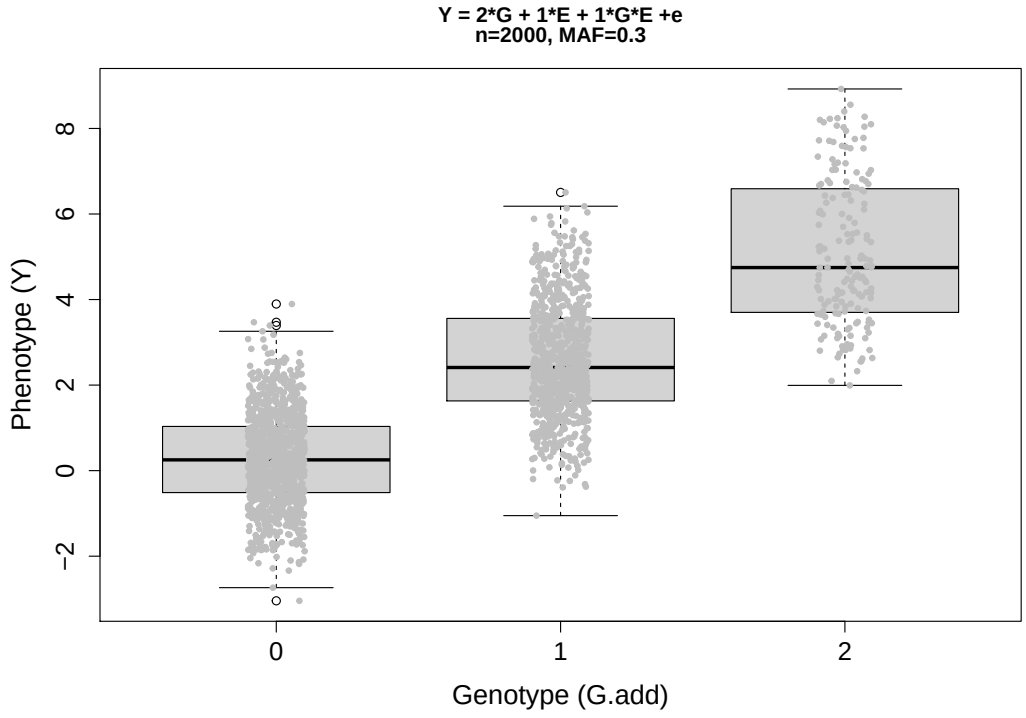
Wang, Sun, Paterson (working paper). Challenges with X-chromosome analyses and reporting in genome-wide association studies.

- specify the testing model, and if not the 3 d.f. model above
- specify the baseline allele
- specify the coding choice and the XCI assumption if invoked
- provide sex-specific estimates

Recommendation for effect size **estimation: An Open Problem!**

Back to association testing (not parameter estimation)

$G \times E$ interaction leads to variance heterogeneity in Y:

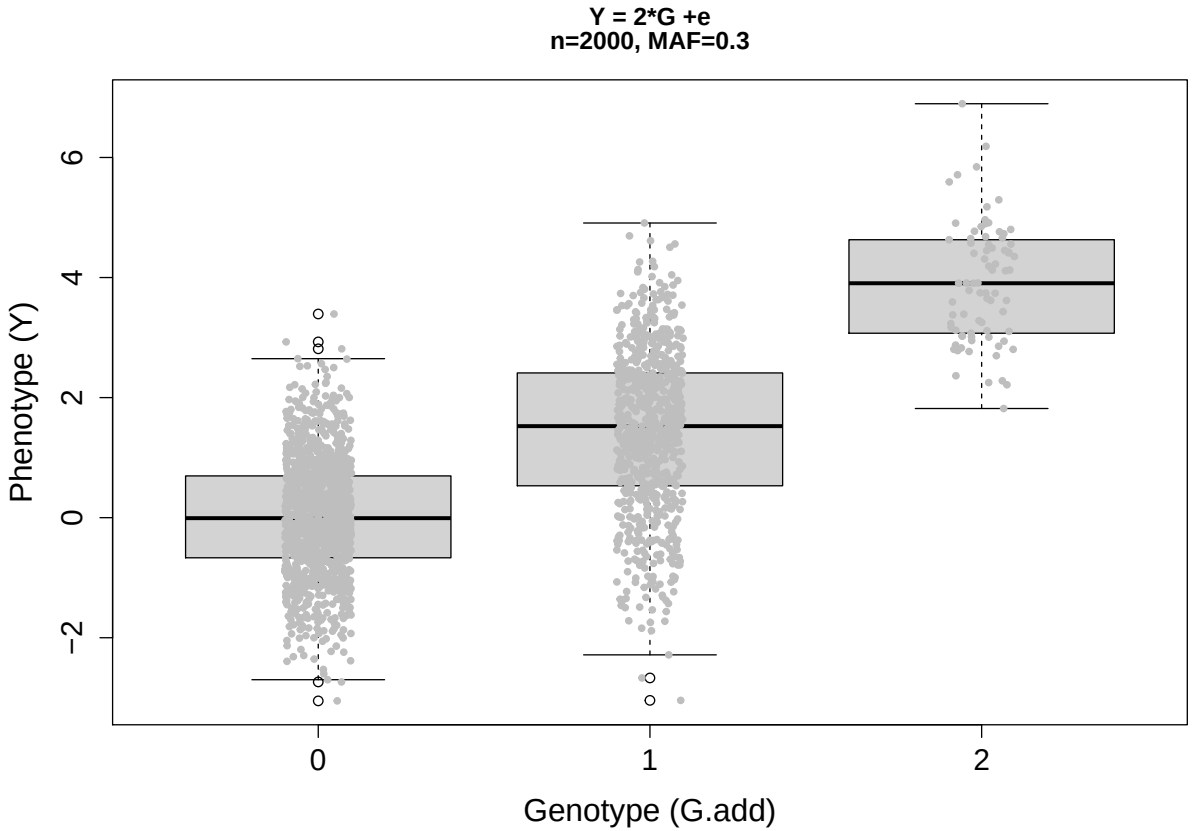


When E is not available, can we leverage the variance heterogeneity observation to **indirectly detect the $G \times E$ interaction**?

Soave and Sun (2017). *Biometrics*. A generalized Levene's scale test for variance heterogeneity in the presence of sample correlation and group uncertainty

Another source of variance heterogeneity specific to Xchr

A higher phenotypic variance in **heterozygous** females when true model is under XCI:



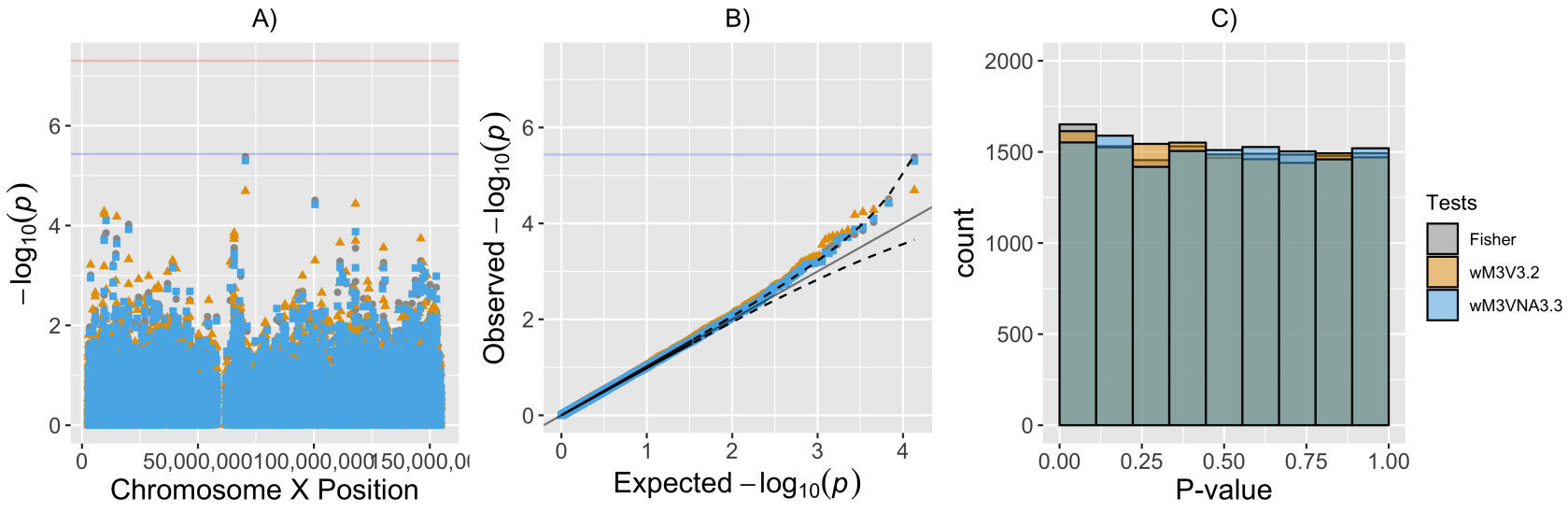

```
## Var(Y|G=0) Var(Y|G=1) Var(Y|G=2)
##  0.9603048  1.8152005  0.9969926
```

Ma et al. (2015). *BMC Genomics*. X-inactivation informs variance-based testing for X-linked association of a quantitative trait.

UK Biobank and MESA Xchr application

Variance heterogeneity of *height* in UKB (n>300k)

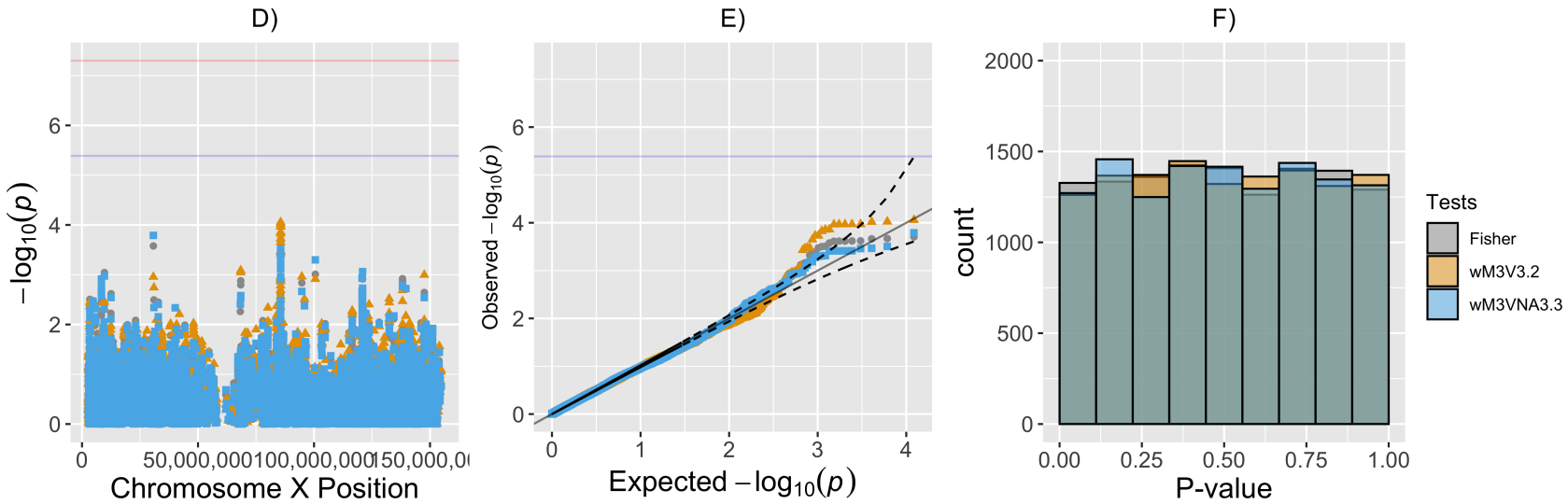
No Xchr-wide signals at $0.05/13,621$



Deng et al. (2019). *Genetic Epidemiology*. Analytical strategies to include the X-chromosome in variance heterogeneity analyses: Evidence for trait-specific polygenic variance structure

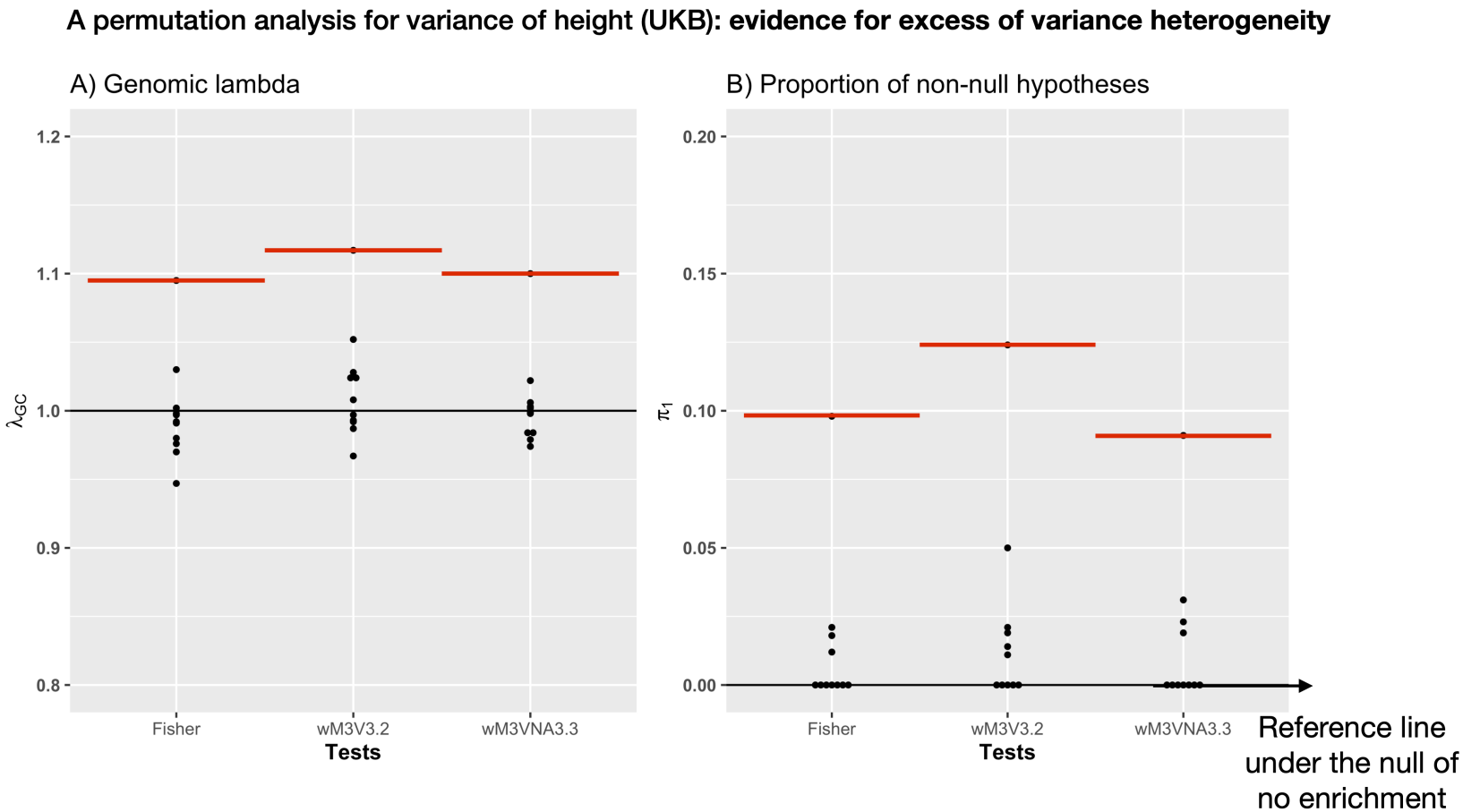
UK Biobank and MESA Xchr application (cont'd)

Variance heterogeneity of *height* in MESA (n = 2k)



Deng et al. (2019). *Genetic Epidemiology*. Analytical strategies to include the X-chromosome in variance heterogeneity analyses: Evidence for trait-specific polygenic variance structure

Height potentially enriched for $G \times E$ interactions (UKB)



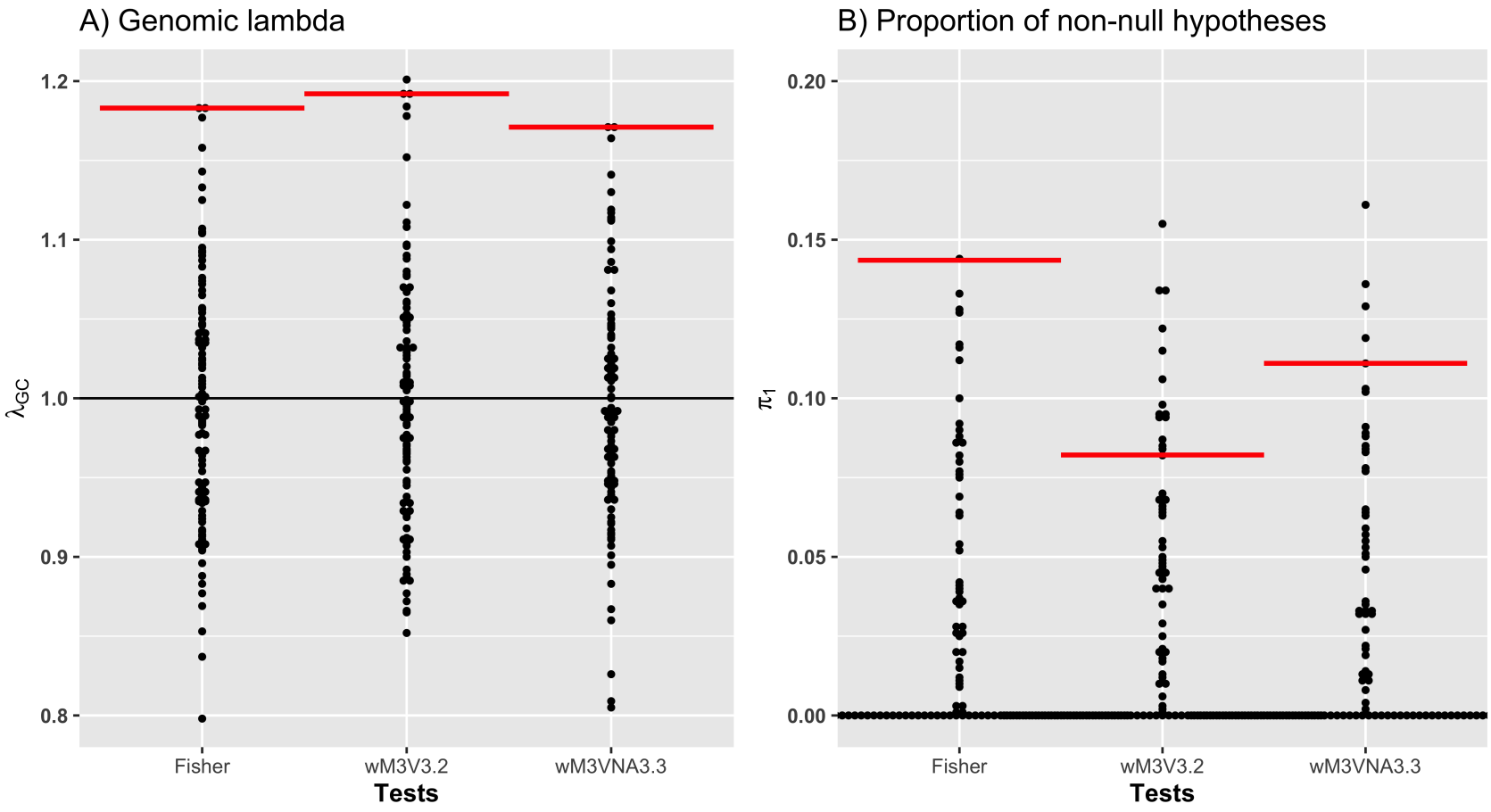
Permuted statistics

Observed statistics

- $\lambda_{GC} = \dfrac{\text{observed median}}{\text{theoretical median}}$
- π_1; proportion of non-null hypothesis

Deng et al. (2019). *Genetic Epidemiology*. Analytical strategies to include the X-chromosome in variance heterogeneity analyses: Evidence for trait-specific polygenic variance structure

Height potentially enriched for $G \times E$ interactions (MESA)



Deng et al. (2019). *Genetic Epidemiology*. Analytical strategies to include the X-chromosome in variance heterogeneity analyses: Evidence for trait-specific polygenic variance structure

Recap the goal of this talk: a better understanding of

- Why association testing of an Xchr variant is actually 'easy'.
  advances in both location (mean) and scale (variances) methods
  "=" boosted power for discovery

- Why **genetic effect estimation is hard for Xchr**!
  GWAS data alone cannot identify the true Xchr genetic model.
- Why $\hat{\beta}$ is not meaningful for an Xchr variant.
  Good Xchr-inclusive PRS analysis is still an open question!

Xchr open but **consequential** questions

Improving data quality

- Xchr-aware calling of sequencing data
- Xchr-aware imputation
- Xchr-aware QC steps

Inference issues

- Xchr association **testing** problem solved, BUT
- **Which effect estimates to use for Xchr-PRS?**
- Sex-stratified, sex-combined effects, $G \times S$ interaction effects?

Other issues

- How to measure **LD for the Xchr**?
- Is it okay to use autosomes-derived PCA for the Xchr?
- How to perform **PCA for the Xchr**? How to do LD prunning?
- How to estimate h^2 **for the Xchr**? Particularly in the context of the β estimation issue!

References

- Carrel et al (2005). *Nature*. X-inactivation profile reveals extensive variability in X-linked gene expression in females.
- Tukiainen et al. (2017). *Nature*. Landscape of X chromosome inactivation across human tissues.

- Clayton (2008). *Biostatistics*. Testing for association on the X chromosome.
- Wise et al. (2013). *American Journal of Human Genetics*. eXclusion: toward integrating the X chromosome in genome-wide association analyses.
- Konig et al. (2014). *Genetic Epidemiology*. How to include chromosome X in your genome-wide association study.
- Wang et al. (2014). *Genetic Epidemiology*. X-chromosome genetic association test accounting for X-inactivation, skewed X-inactivation, and escape from X-inactivation.
- Chen et al. (2020). *Biostatistics*. Bayesian model averaging for the X-chromosome inactivation dilemma in genetic association study.

- Ma et al. (2015). *BMC Genomics*. X-inactivation informs variance-based testing for X-linked association of a quantitative trait.
- Soave and Sun (2017). *Biometrics*. A generalized Levene's scale test for variance heterogeneity in the presence of sample correlation and group uncertainty
- Deng et al. (2019). *Genetic Epidemiology*. Analytical strategies to include the X-chromosome in variance heterogeneity analyses: Evidence for trait-specific polygenic variance structure.

- Chen et al. (2021). *Genetic Epidemiology*. The X factor: a robust and powerful approach to X-chromosome-inclusive whole-genome association studies.
- Song et al. (2021) *Genetic Epidemiology*. Testing and estimation of X-chromosome SNP effects: Impact of model assumptions.