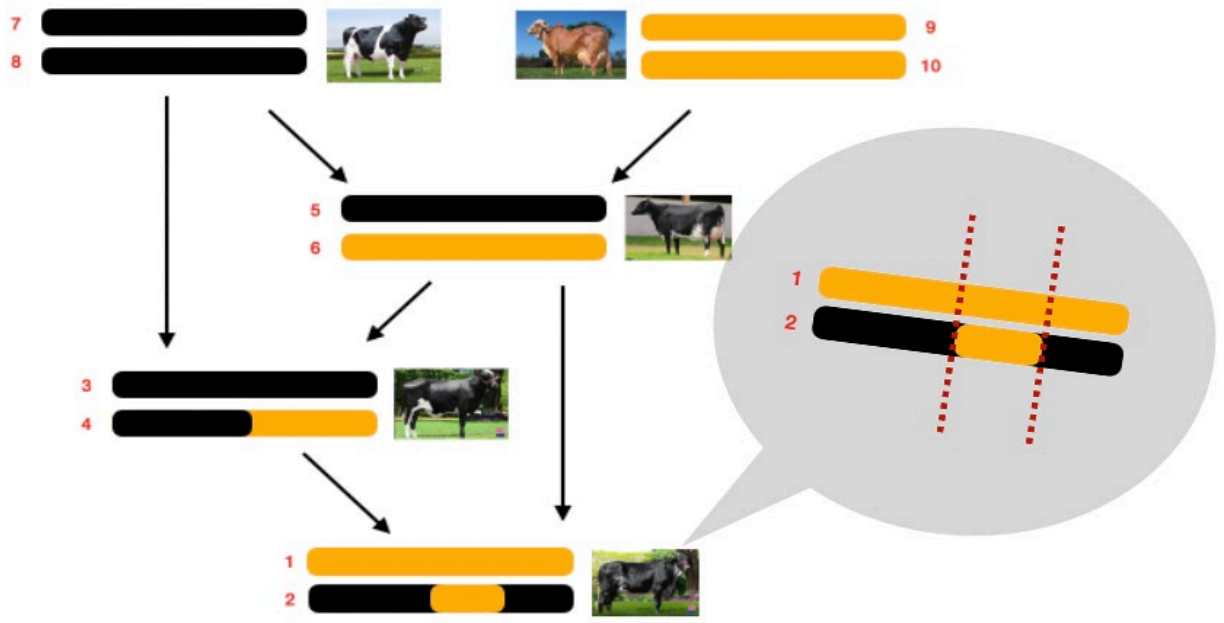
The effective number and profile of causative markers in GWAS

Ivan Pocrnić, D. A. L. Lourenco & I. Misztal

EAAP, Florence IT, 2024/09/04

Background

- Inheritance is chunkular



Fortuna (2024)

Background

- Inheritance is chunkular
- Independent chromosome segments (Me)

• Theory of junctions	Fisher (1949)
• E(Me) = 4NeL	Stam (1980)
• 1 / Var(**G**-**A**)	Wientjes et al. (2016)
• Spectral decomposition **G**	Misztal (2016); Pocrnic et al. (2016)

  - Rank ≤ min(#SNP, #ID, #Me)
  - The number that explains 98% of variance in **G**

- … many more exist …

Background

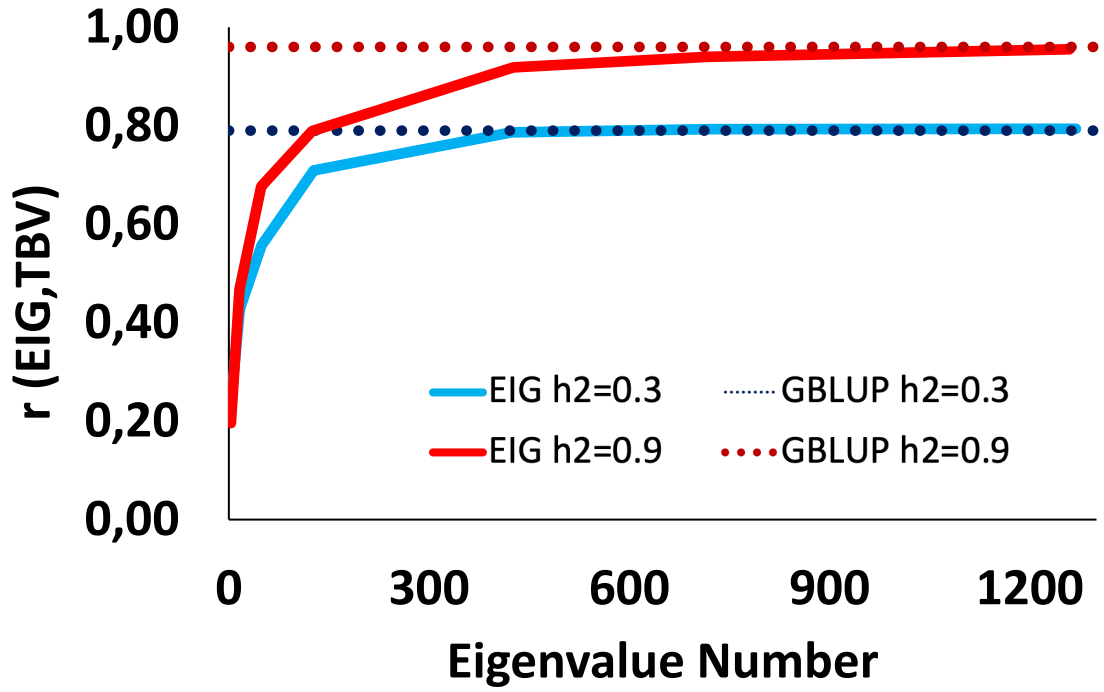
- Inheritance is chunkular
- Independent chromosome segments (Me)
- Genomic dimensionality is limited

Species	Genotyped (x1000)	SNP (x1000)	Dimensionality (x1000)	Ne
Pig	23	37	**4.1**	48
Chicken	16	39	**4.2**	44
Jersey	75	61	**11.5**	101
Angus	81	38	**10.6**	113
Holstein	77	61	**14.0**	149

Pocrnic *et al.* (2016)

Background

- Inheritance is chunkular
- Independent chromosome segments (Me)
- Genomic dimensionality is limited
- Genomic prediction works by estimating chromosome segments (eigenvalues)



Pocrnic *et al.* (2019)

Motivation

- GP works by estimating chromosome segments (eigenvalues)
- How does the effective population size affect GWAS?
- How does the data size affect GWAS?
- What is the profile of a single QTL in Manhattan plots?
- Can the signals on the Manhattan plot be decomposed?

Simulation

- AlphaSimR (Gaynor et al., 2021)

- 10 CHR (100cM / CHR)
- 50,000 equidistant SNP (50/cM)
- 100 equidistant QTL (10/CHR) with equal additive effect
  - 500 SNP (≈10cM)
- Single trait with $h^2 = 0.5$
- 10 generations of random mating

- Populations with Ne 60 (small), Ne 60 (large) and Ne 600 (small)
  - Small; 6k genotyped animals; 24k total (ped + pheno)
  - Large; 18k genotyped animals; 60k total (ped + pheno)

Single-step GWAS

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Wu} + \mathbf{e}$$
$$\mathrm{Var}(\mathbf{e}) = \mathbf{I}\sigma_e^2$$
$$\mathrm{Var}(\mathbf{u}) = \mathbf{H}\sigma_u^2$$

$$\hat{\mathbf{a}}|\hat{\mathbf{u}} = \lambda\alpha\delta\mathbf{DZ}'\mathbf{G}^{-1}\hat{\mathbf{u}}$$

$$Pvalue_i = 2\left(1 - \Phi\left(\left|\frac{\hat{a}_i}{sd(\hat{a}_i)}\right|\right)\right)$$

BLUPF90 family

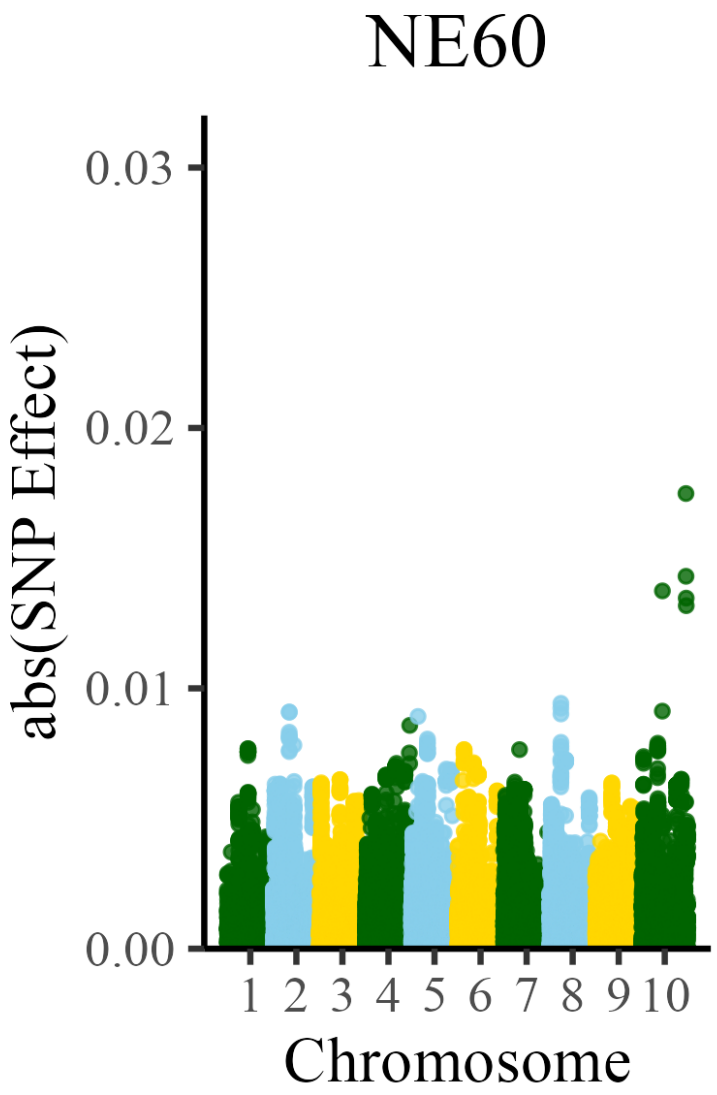
$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_b^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

$$\mathbf{G}_b = \alpha\mathbf{G} + (1-\alpha)\mathbf{A}_{22}$$

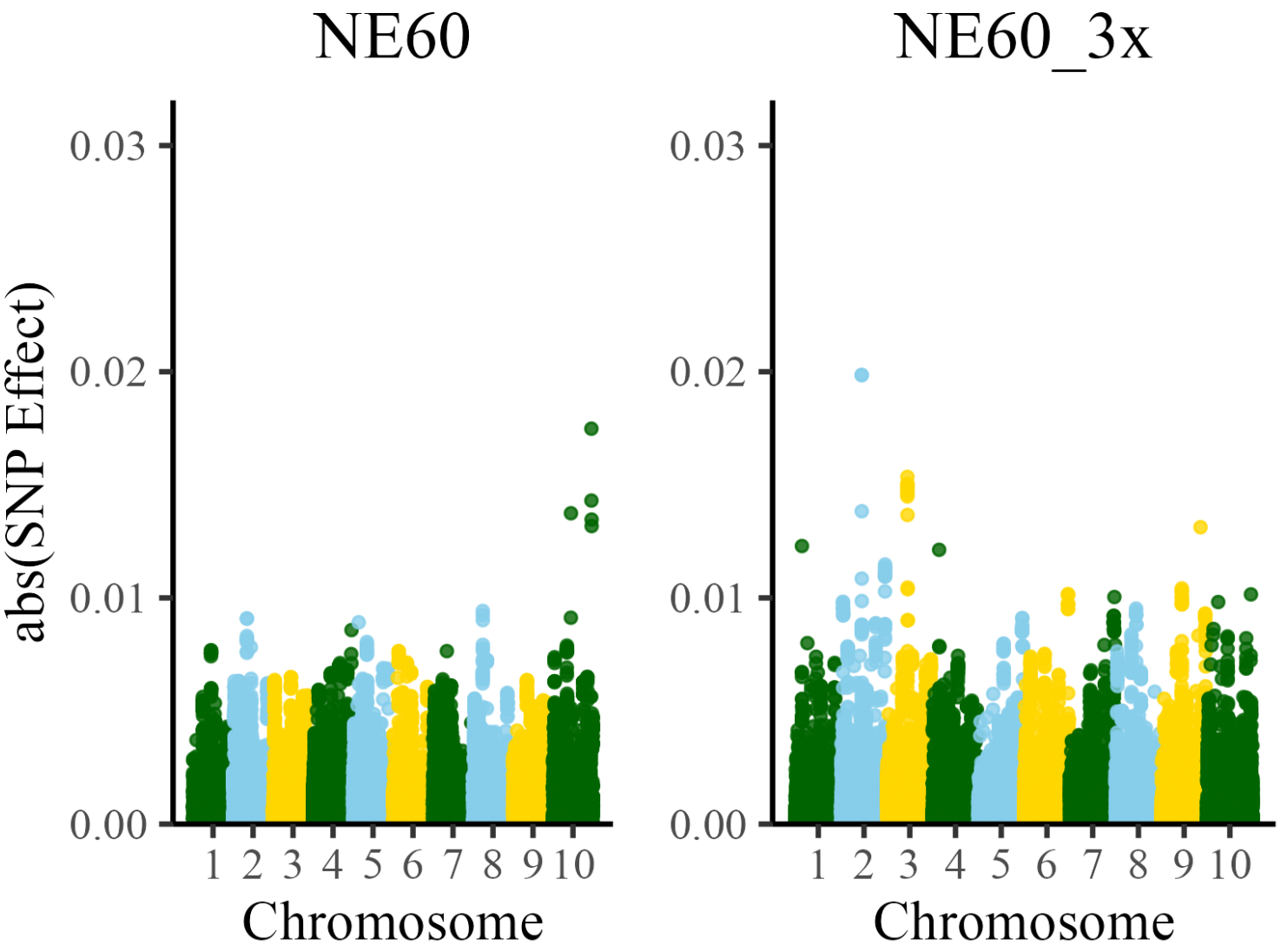
$$\mathbf{G} = \frac{\mathbf{ZDZ}'}{2\sum p_j(1-p_j)}$$

Wang *et al.* (2012); Aguilar *et al.* (2019)

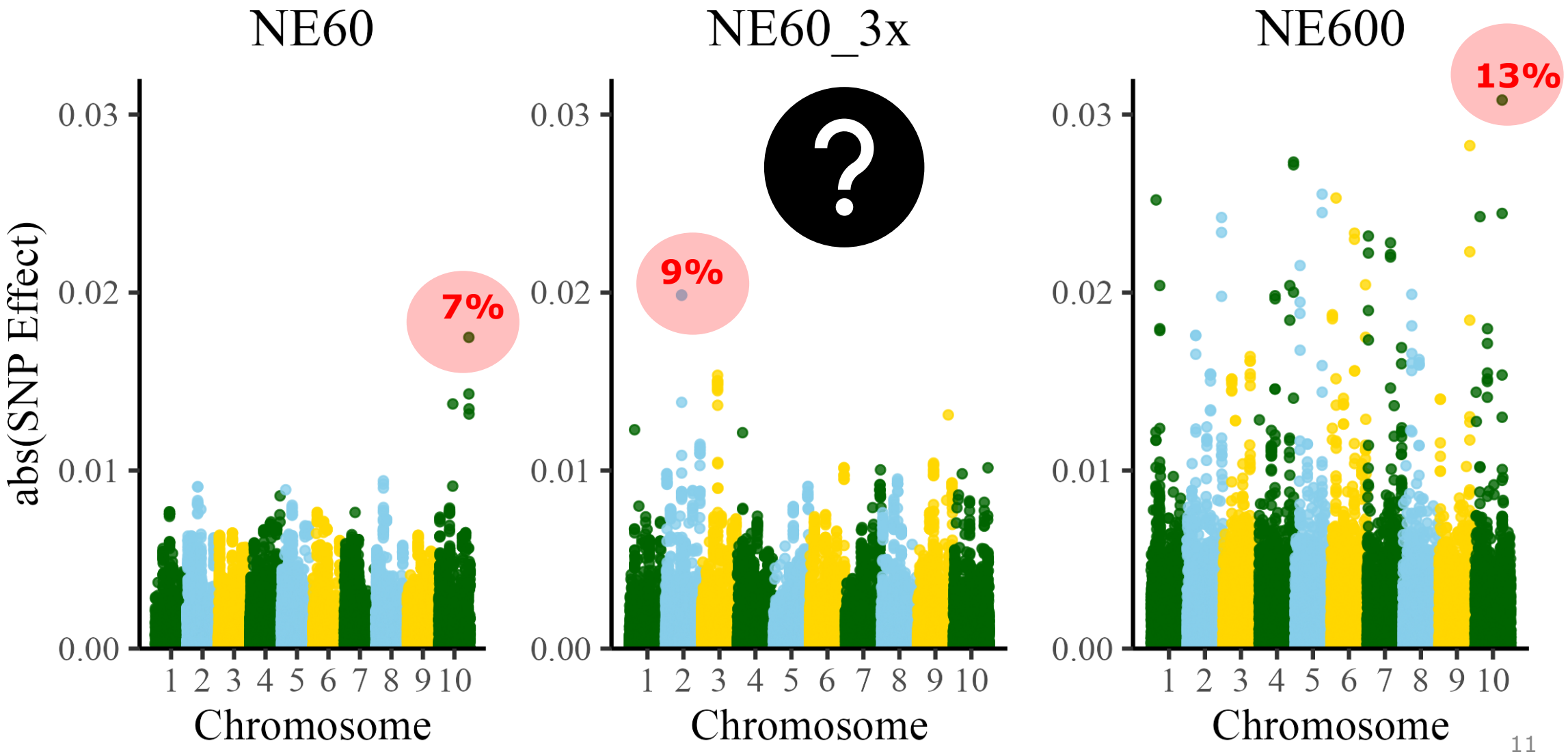
Manhattan plots (SNP-effects)



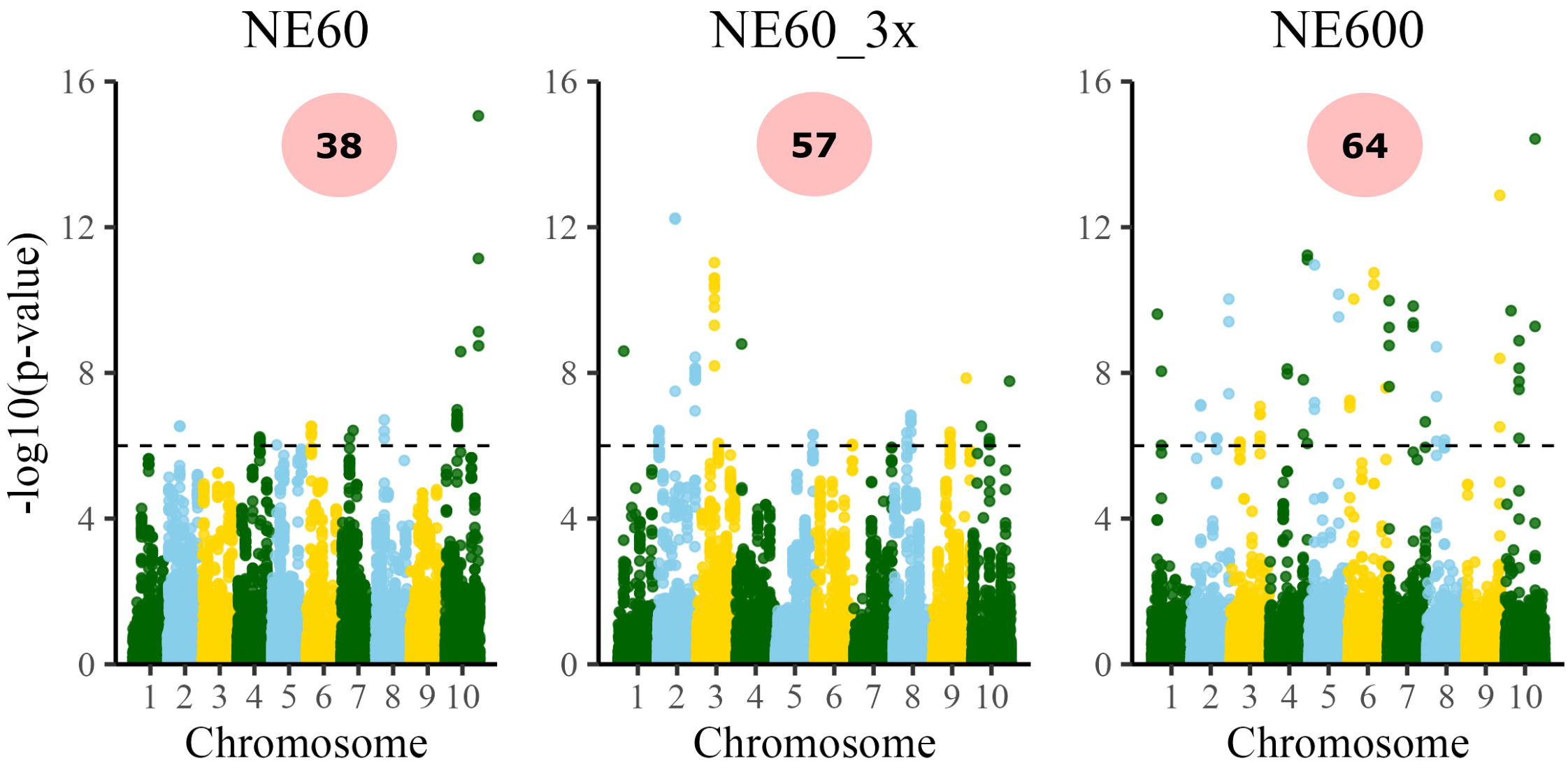
Manhattan plots (SNP-effects)



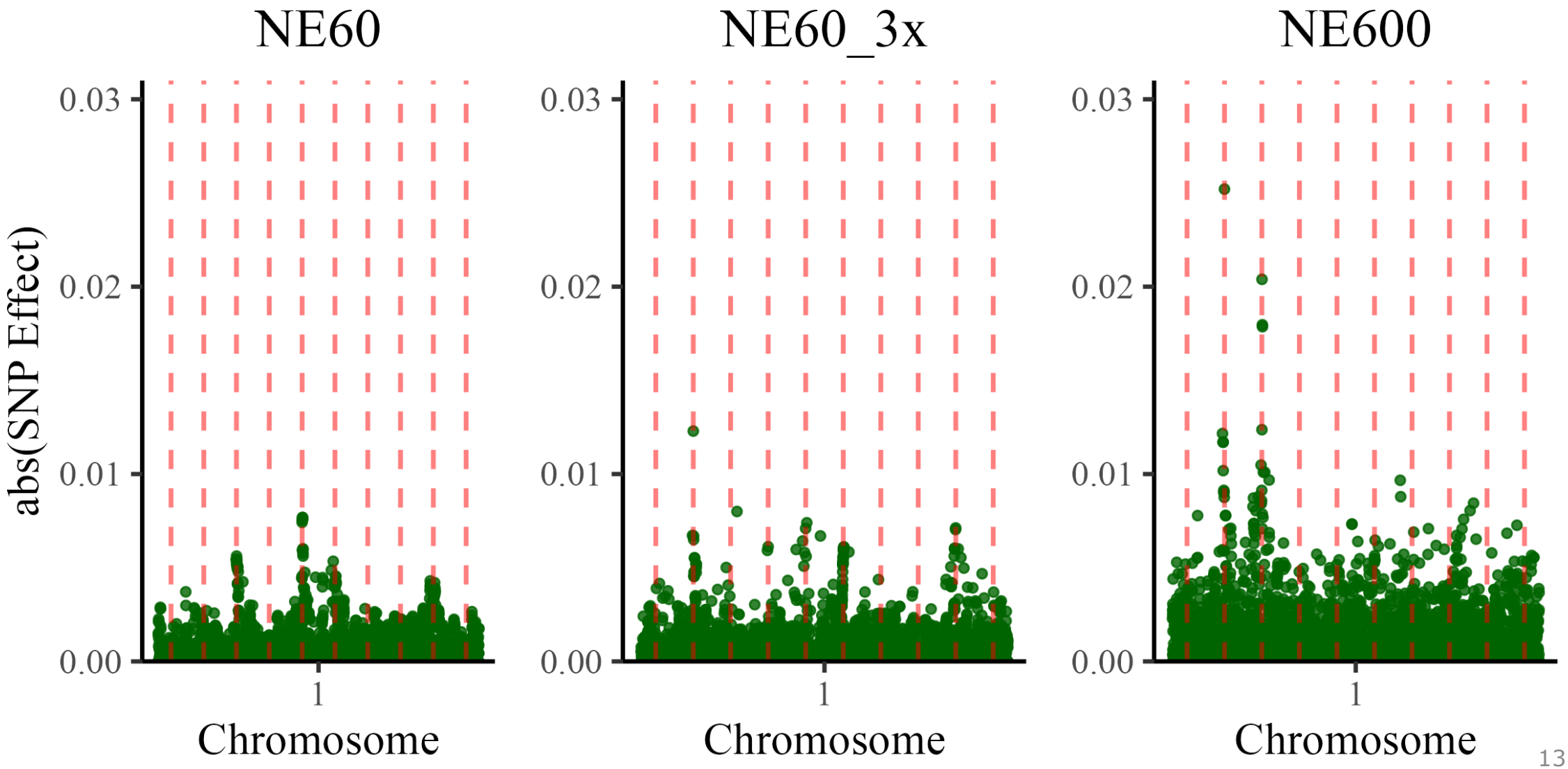
Manhattan plots (SNP-effects)



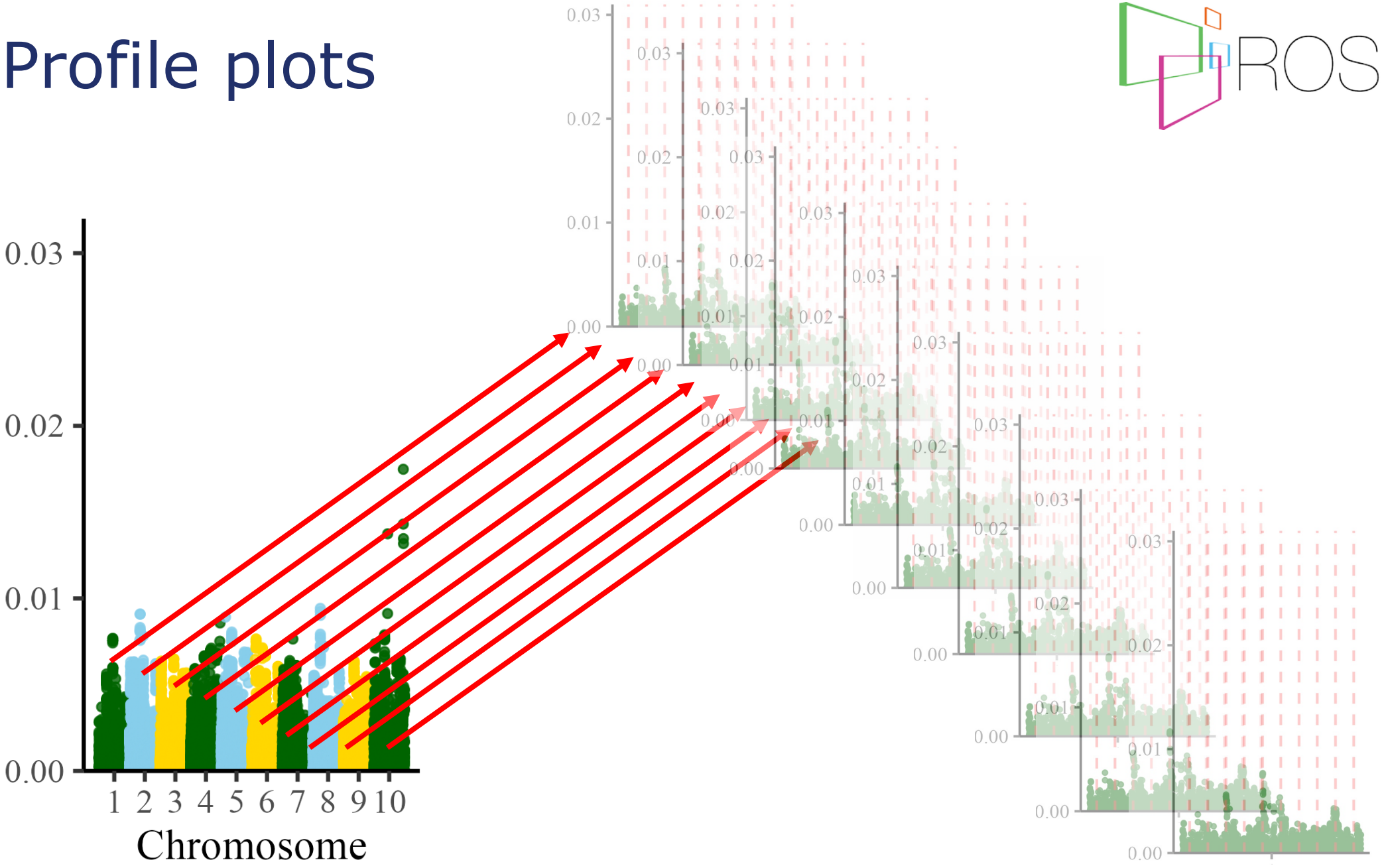
Manhattan plots (p-values)



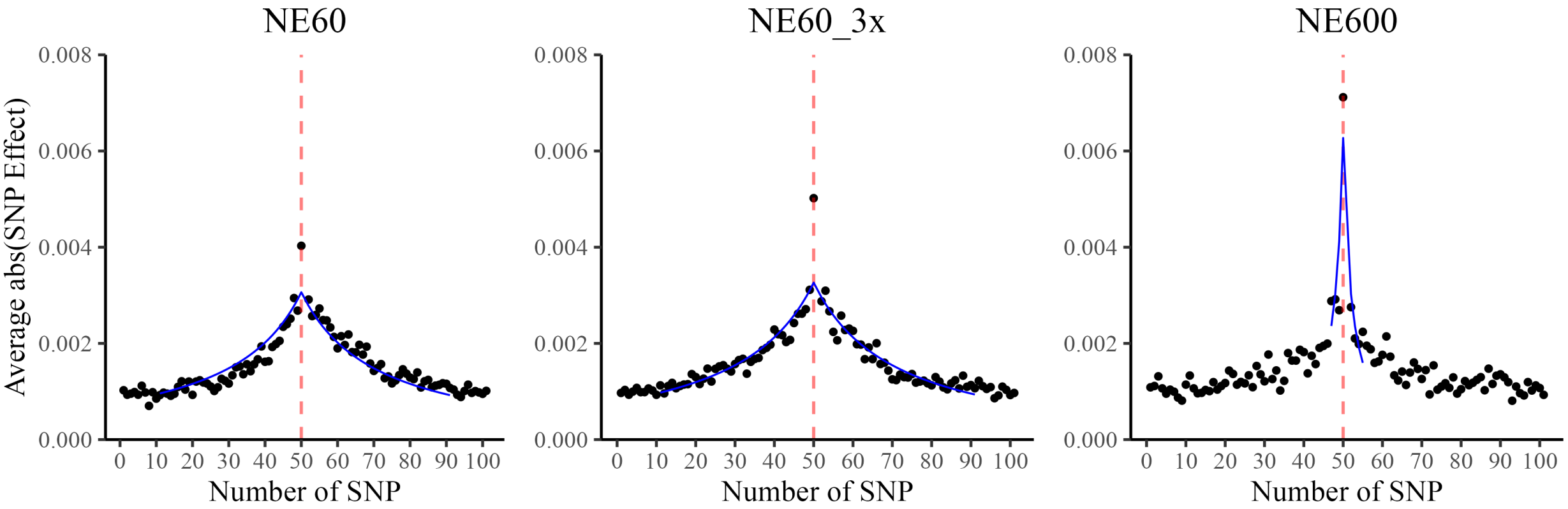
Zoom-in single chromosome



Profile plots



Profile plots



E(Me) = 2,400

E(Me) = 2,400

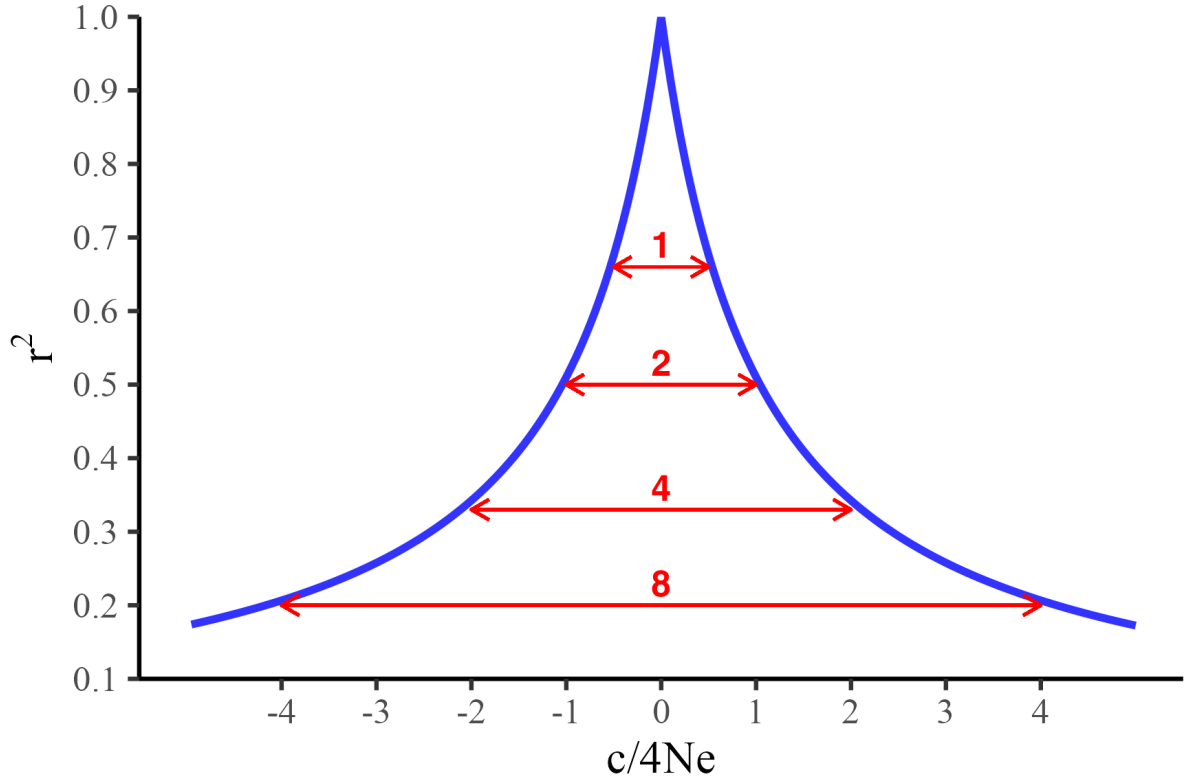
E(Me) = 24,000

What is that profile?

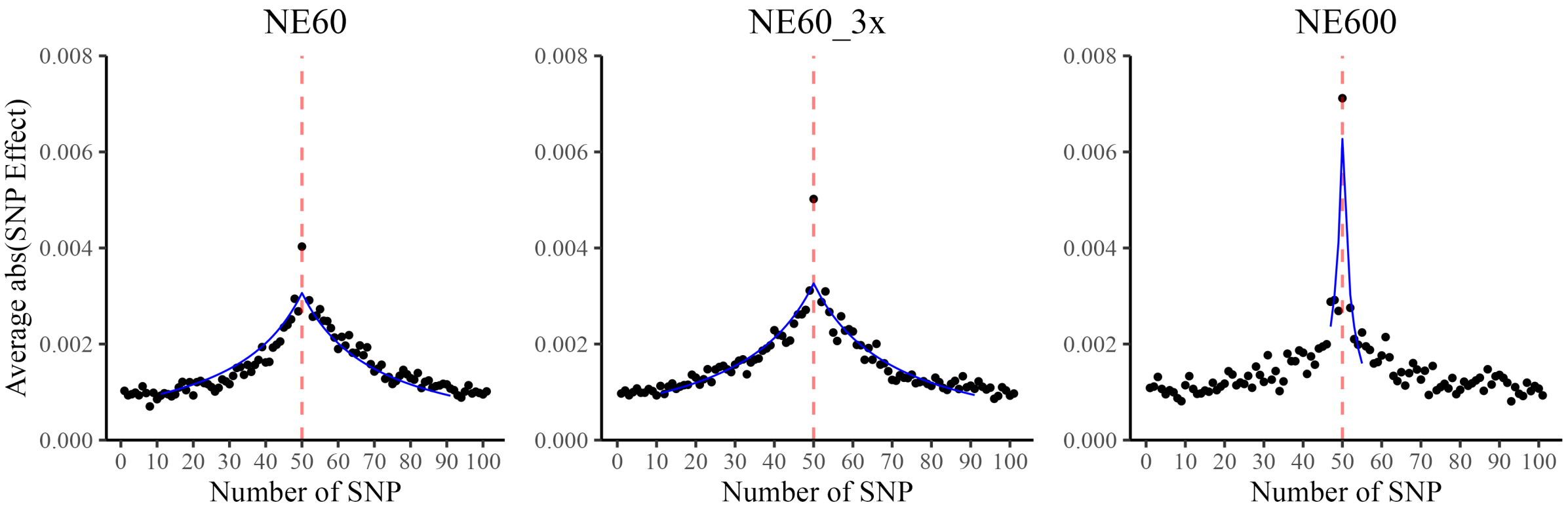
- $E(r^2) = 1/(4cNe + 1)$ Sved(1971)

- Stam segment is 1/4 Ne Morgan

- Optimal size
  - 2Mb (Ne 100)
  - 5Mb (Ne 40)
  - 20kb (Ne 10,000)



Profile plots (fit ± 2 Stam segments)



R^2 = 0.87

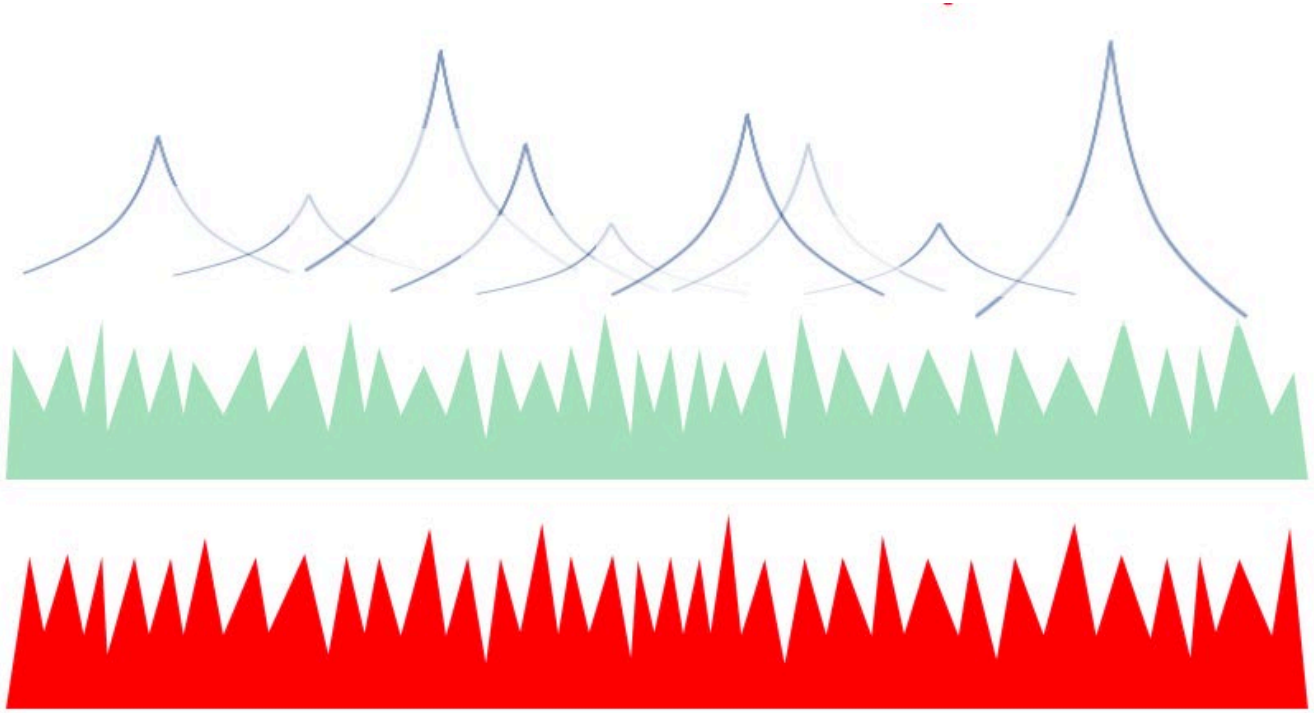
R^2 = 0.85

R^2 = 0.78

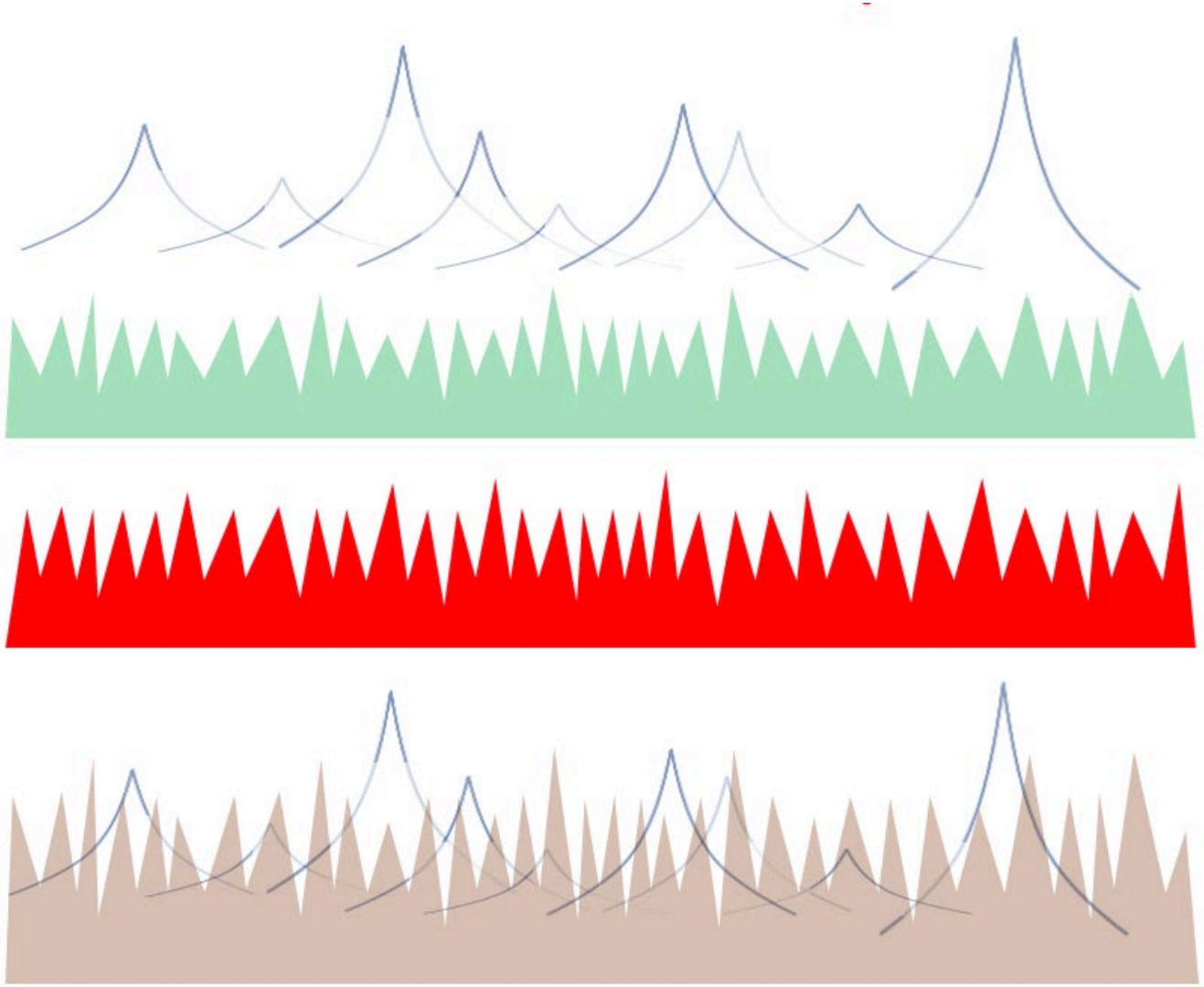
Components of Manhattan plot

QTNs

Components of Manhattan plot



Components of Manhattan plot



QTNs

Relationships

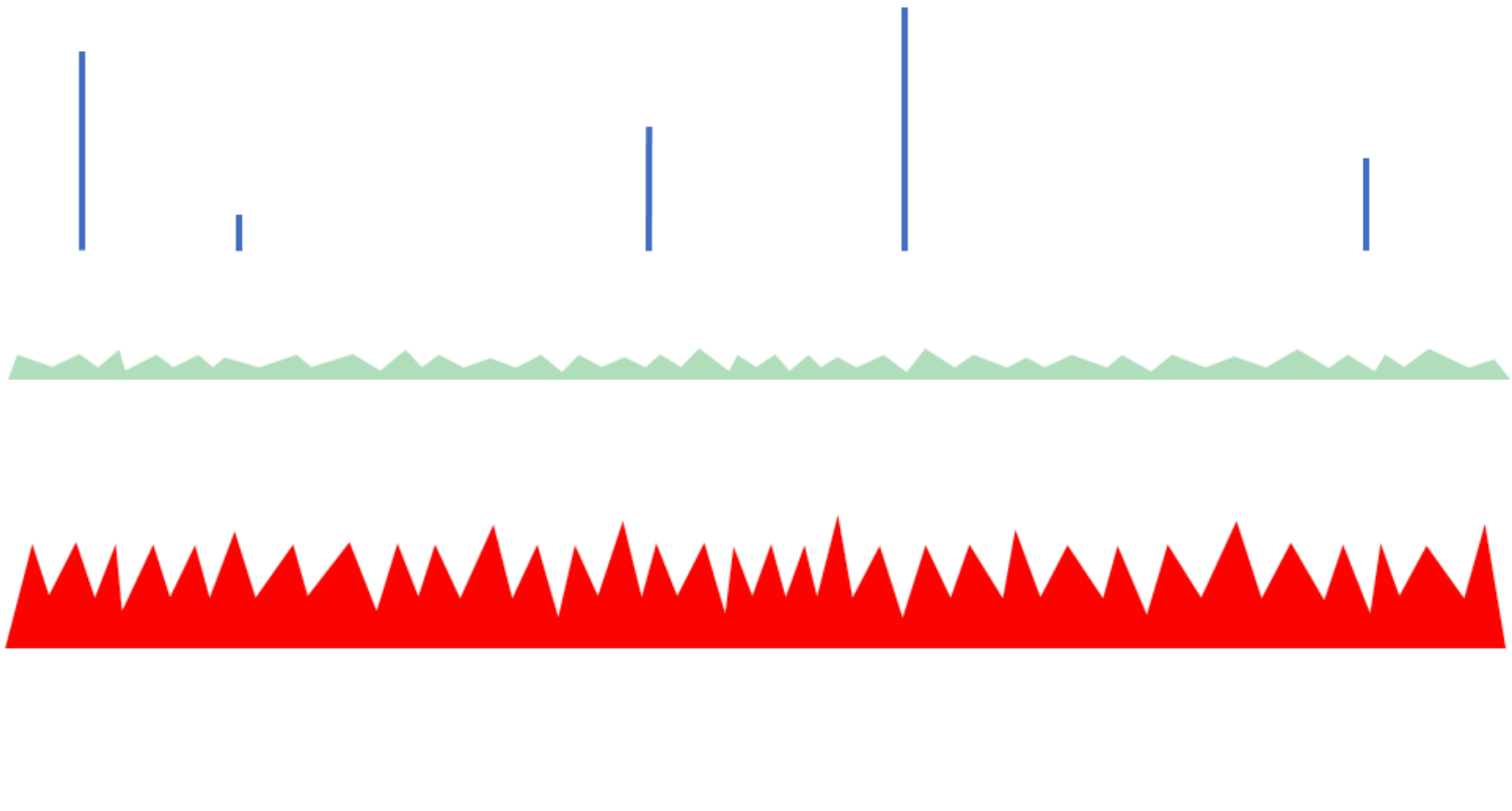
Noise

Combined

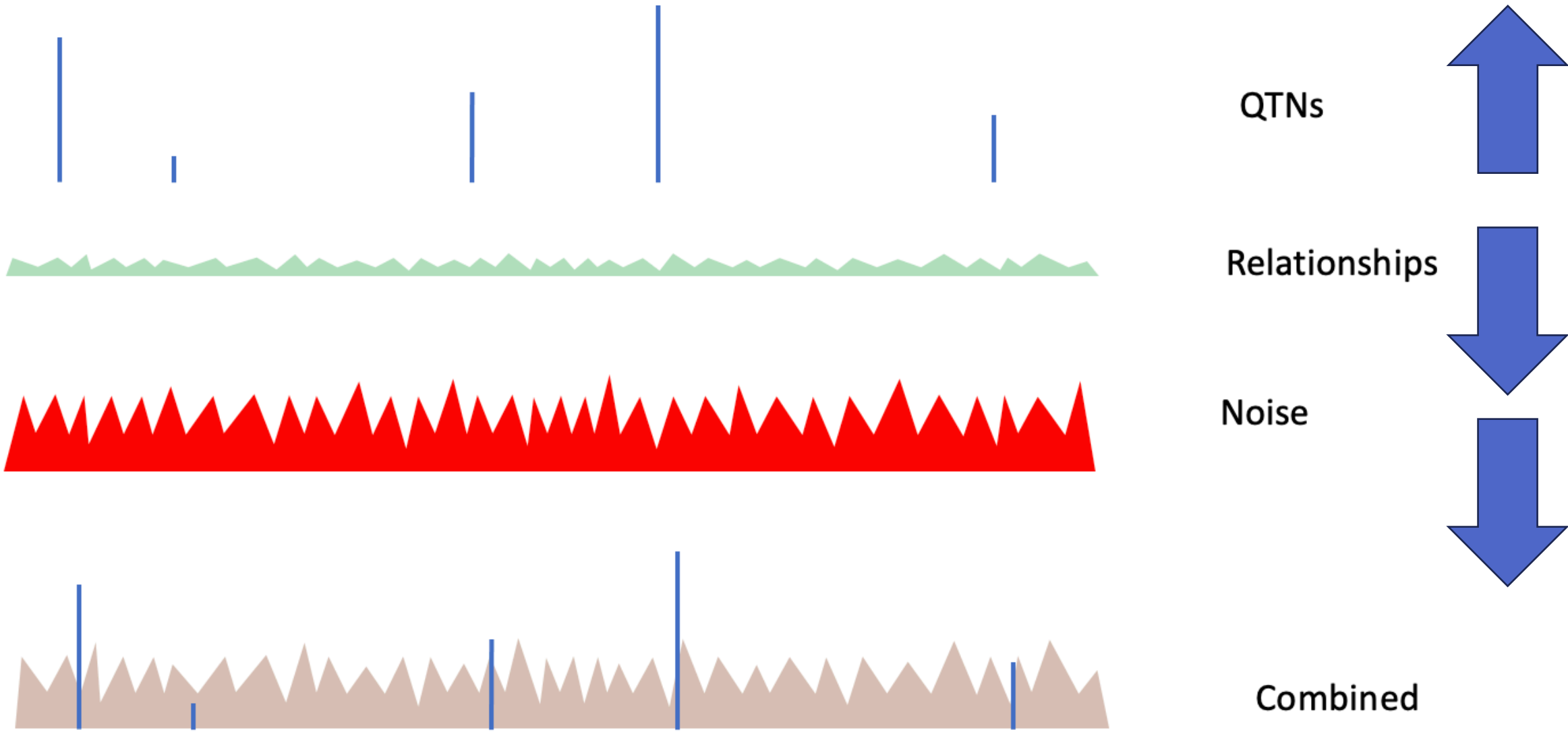
Components of Manhattan plot

QTNs

Components of Manhattan plot

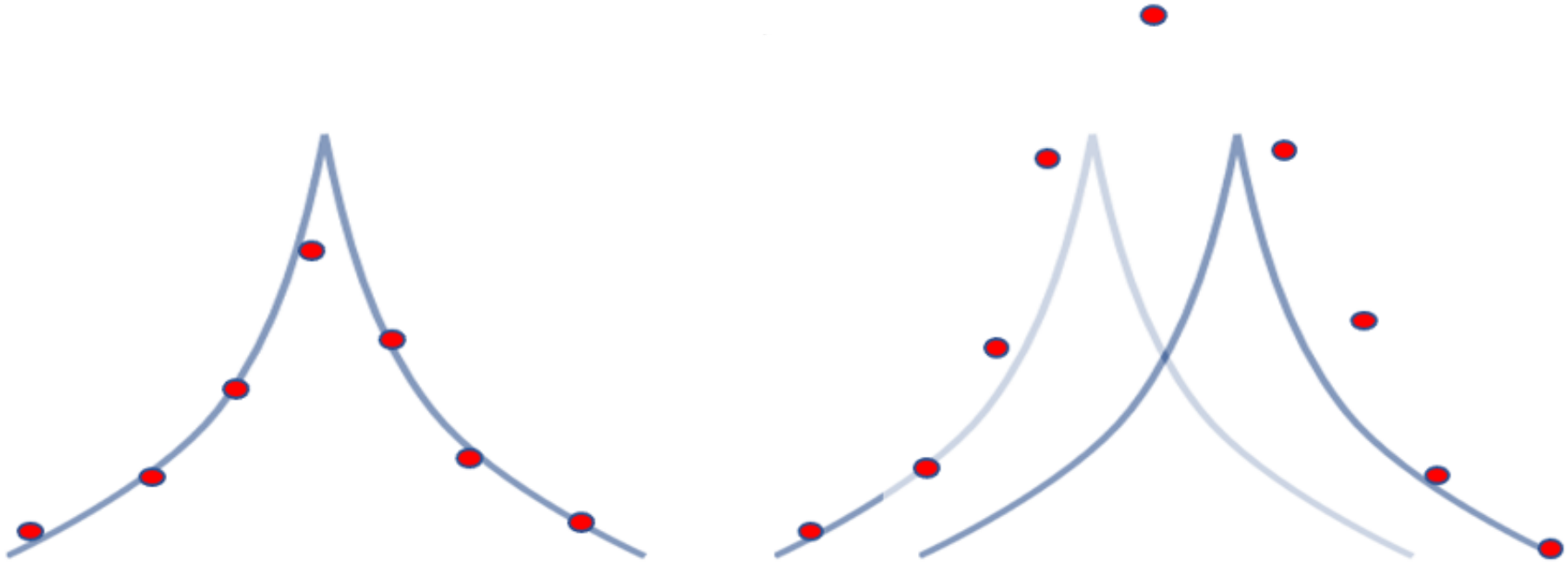


Components of Manhattan plot



Segments overlap

- Why GP captures QTL?
- Optimal SNP-chip size?



Implications

- Signals due to QTN, QTN profile (LD), relationships and noise
  - Small Ne: QTN profile is wide & confounded with relationships
  - Large Ne: QTN profile is narrow & relationships weaker
- Large data needed
- The shape of pairwise LD curve and function of Ne
- GWAS with windows (segments) is useful
  - ~ 1-2 Mb (Ne 100)
  - ~ 0.1-0.2 Mb (Ne 1000)

Thanks

Acknowledgements

BBSRC Institute Strategic Programme

USDA NIFA

Edinburgh Compute and Data Facility

ROSLIN

THE UNIVERSITY of EDINBURGH
The Royal (Dick) School of Veterinary Studies

UKRI Biotechnology and Biological Sciences Research Council

UNIVERSITY OF GEORGIA

OXFORD GENETICS

GENETICS, 2024, 227(4), iyae103
https://doi.org/10.1093/genetics/iyae103
Advance Access Publication Date: 24 June 2024
Genomic Prediction

Single nucleotide polymorphism profile for quantitative trait nucleotide in populations with small effective size and its impact on mapping and genomic predictions

Ivan Pocrnic [1,*,†] Daniela Lourenco [1] Ignacy Misztal [1,*]