Factors Influencing accuracy of genomic selection with sequence Information

Ignacy Misztal, Ivan Pocrnic*, Daniela Lourenco

University of Georgia

*now Roslin Institute

Genomic selection, effective population size and sequence data

- Efficient genomic selection with sequence data - need QTN location and QTN variance (Fragomeni et al., 2017)
- GWAS developed mostly for human data – large effective population size (Ne=3000-10000)
- Farm populations have small Ne: 40 to 150
  - 5k to 15k independent chromosome segments (Stam, 1980; Pocrnic et al., 2016)
  - Each segment size 200k to 600k

- Questions
  - Does small effective population size affect GWAS?
  - What is Manhattan plot composed of?
  - Why small benefits of using sequence data for prediction?

Simulation study

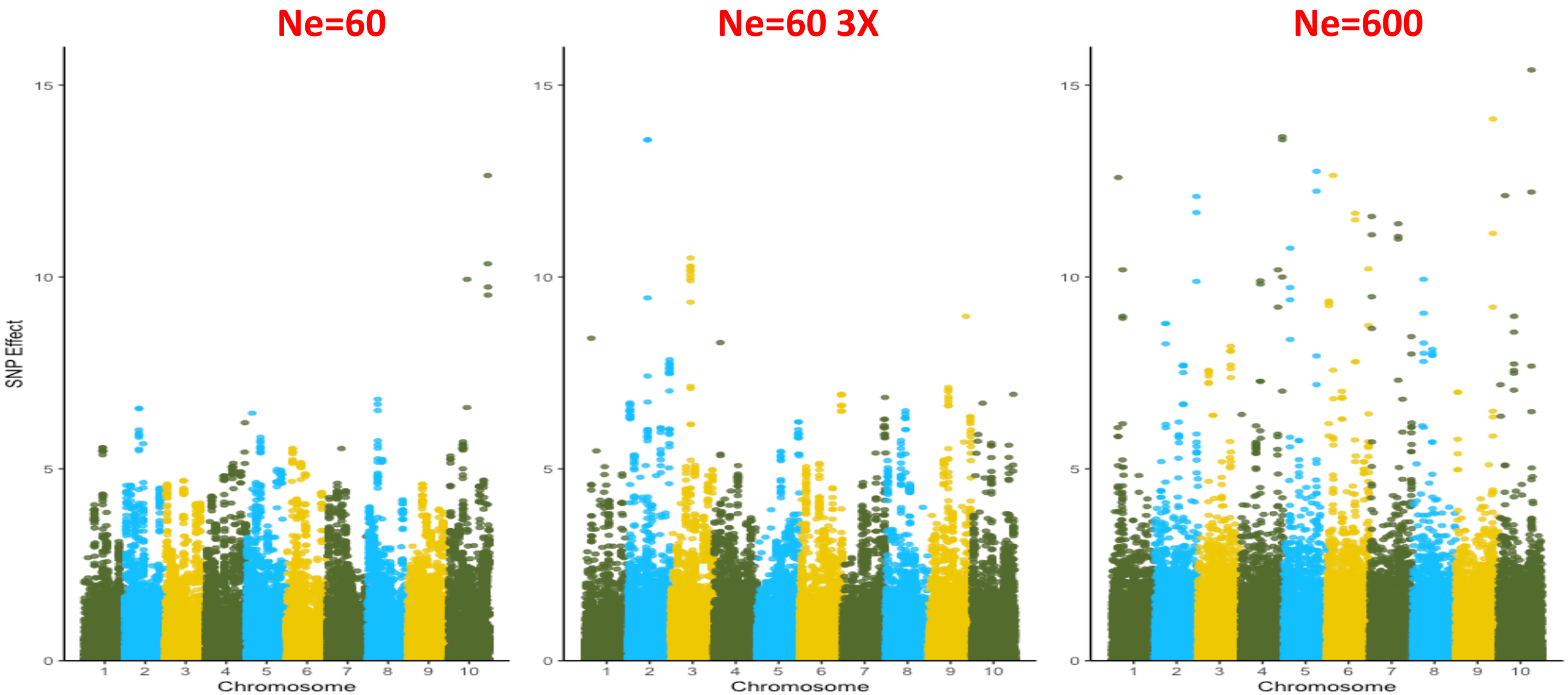
- AlphaSimR (Faux et al., 2016)

- 100 QTN with same effect and equally spaced
- 10 generations with 2000 animals in each
- Last 3 generations genotyped for 50k SNP – causative SNP included

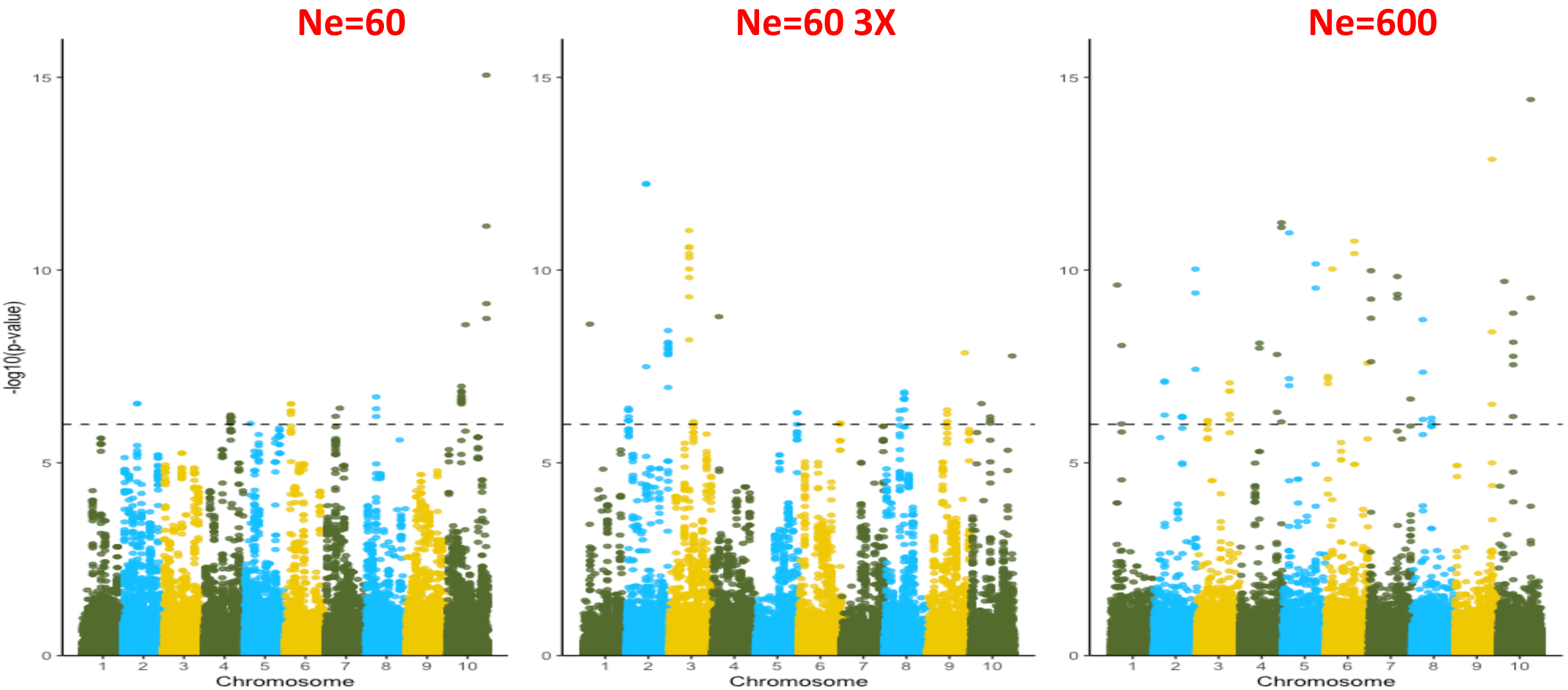
- 3 data sets
  - Ne=60
  - Ne=600
  - Ne=60_3x --3 times more data (6000 per generation)

- Analyzes by ssGBLUP with p-value option

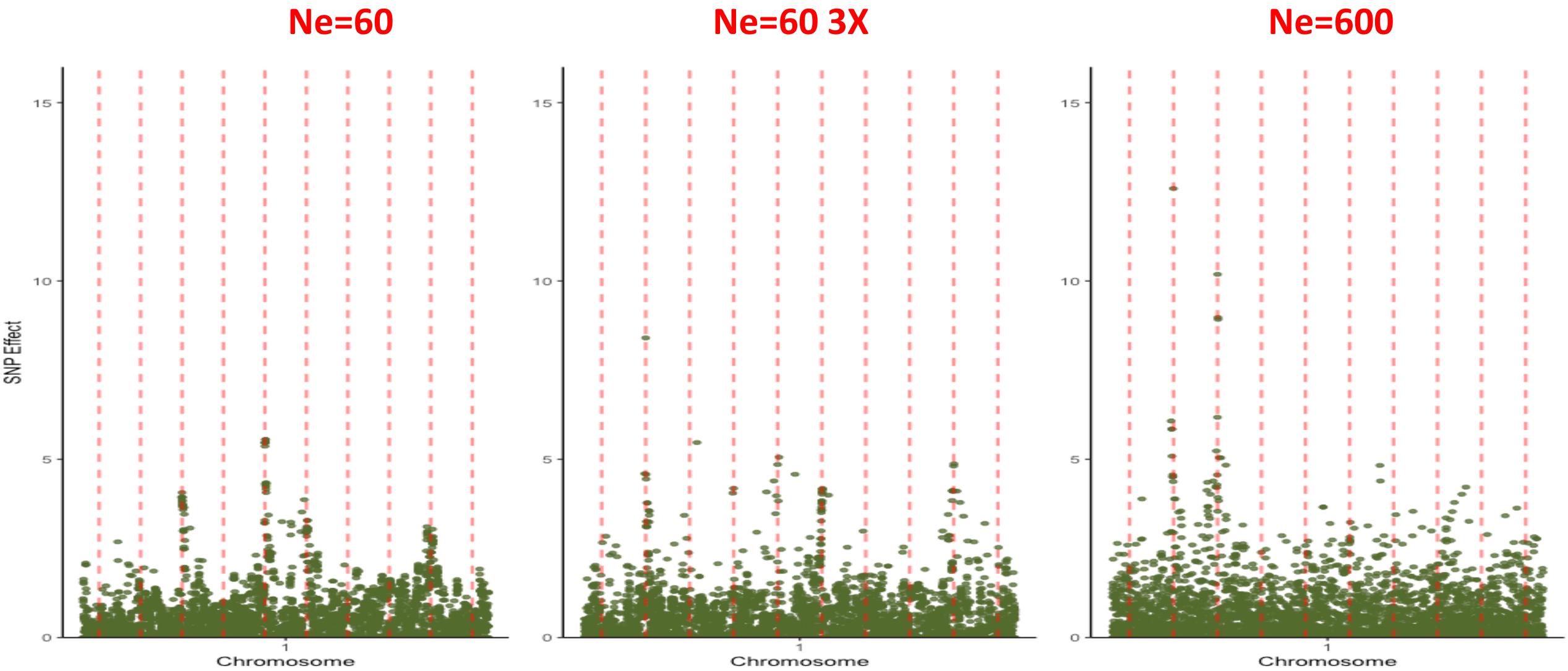
Manhattan plots for SNP effects



Manhattan plots for P-values

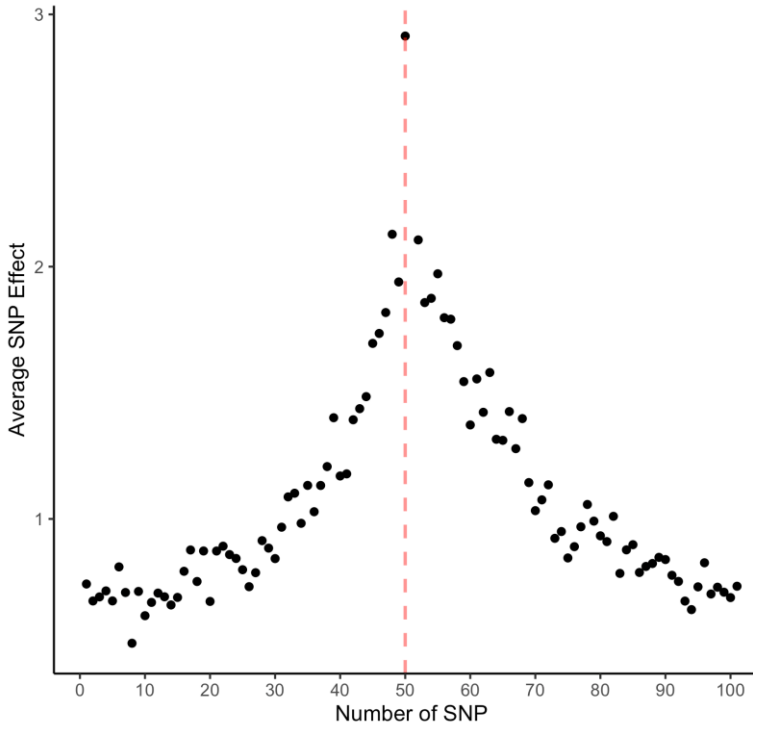


Manhattan plots for first chromosome

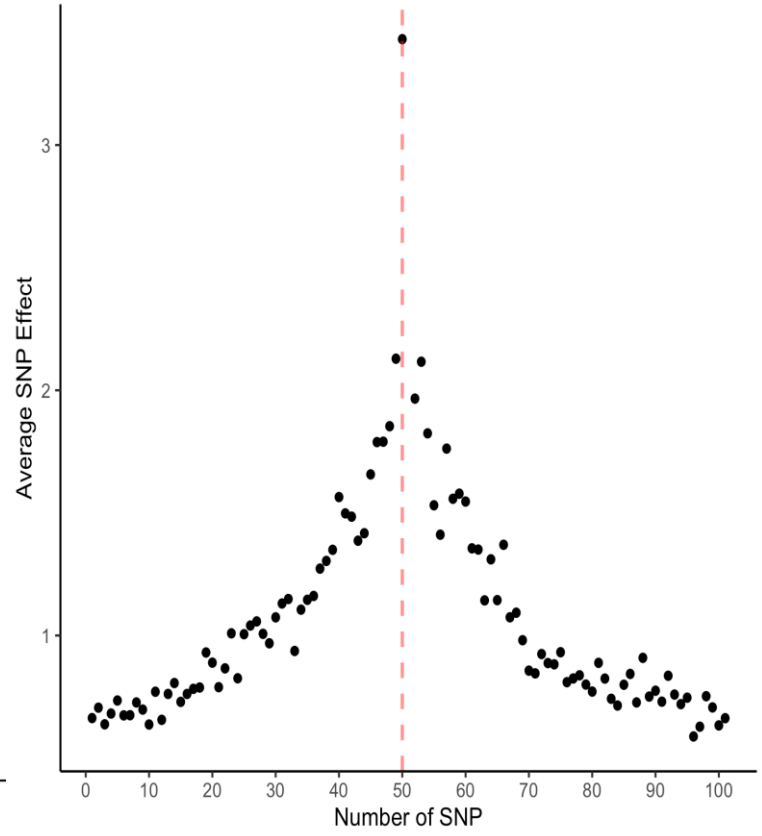


Plots averaged for all QTNs

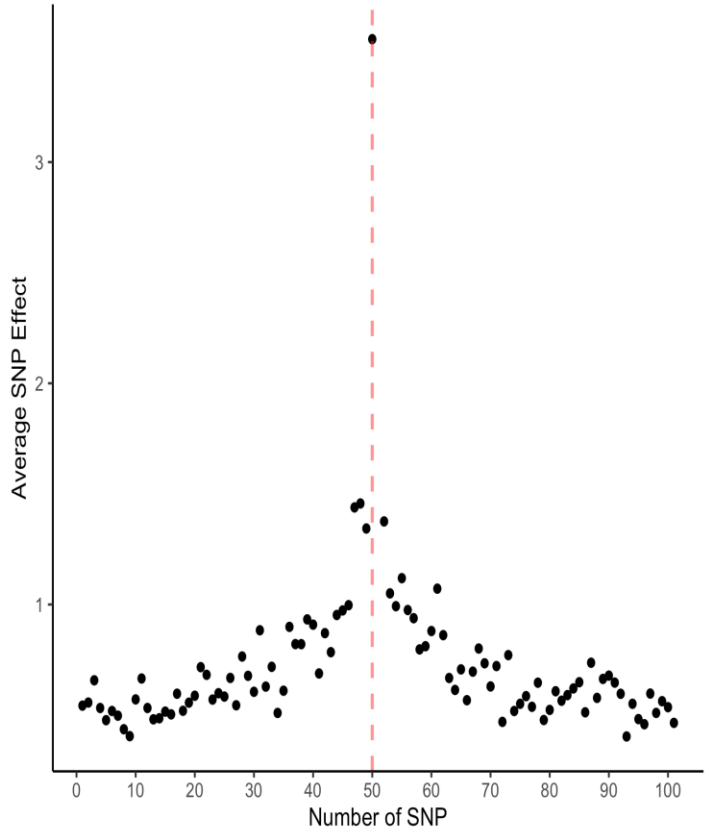
Ne=60



Ne=60 3X



Ne=600

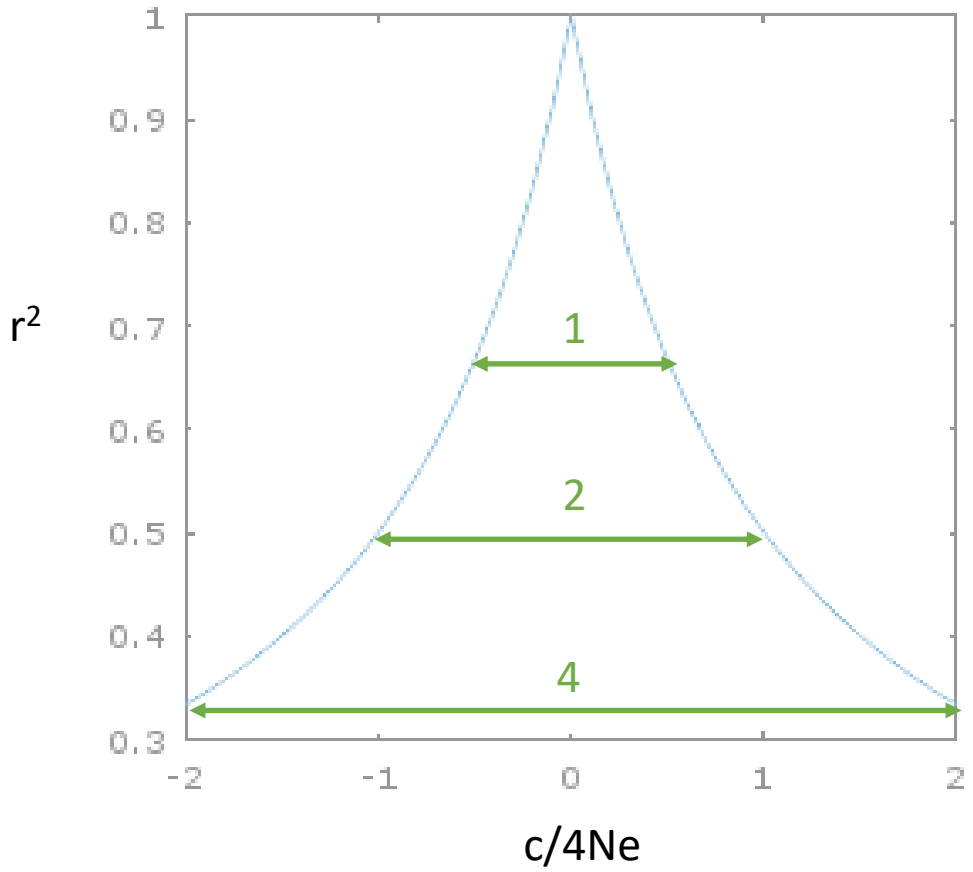


Pairwise linkage disequilibrium (r^2)

$E(r^2)=1/(4\ c\ Ne\ +1)$ Sved (1971)

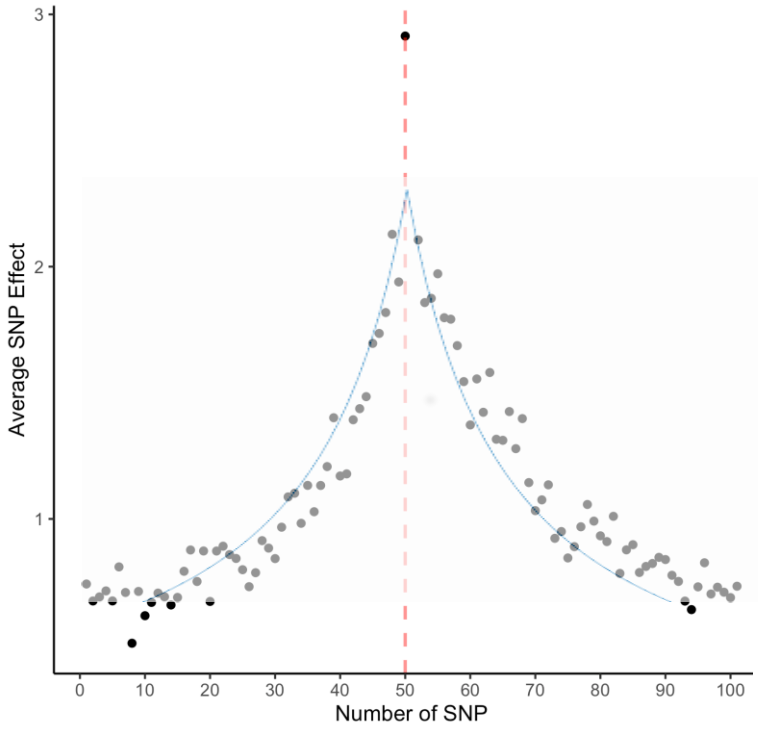
c – distance between genes
Ne – effective population size

One Stam segment - (1/4Ne Morgan)

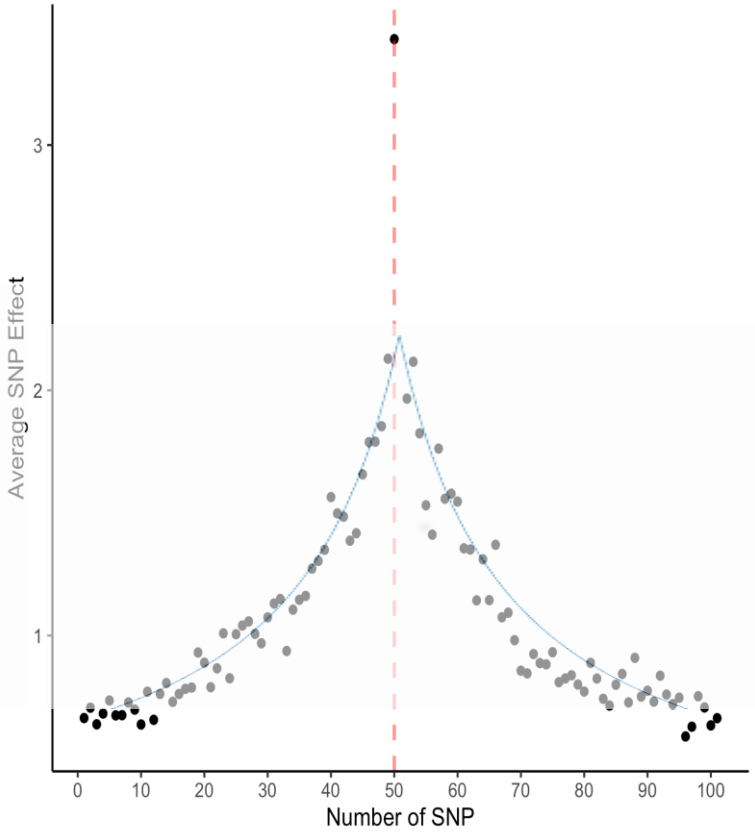


LD fitting….

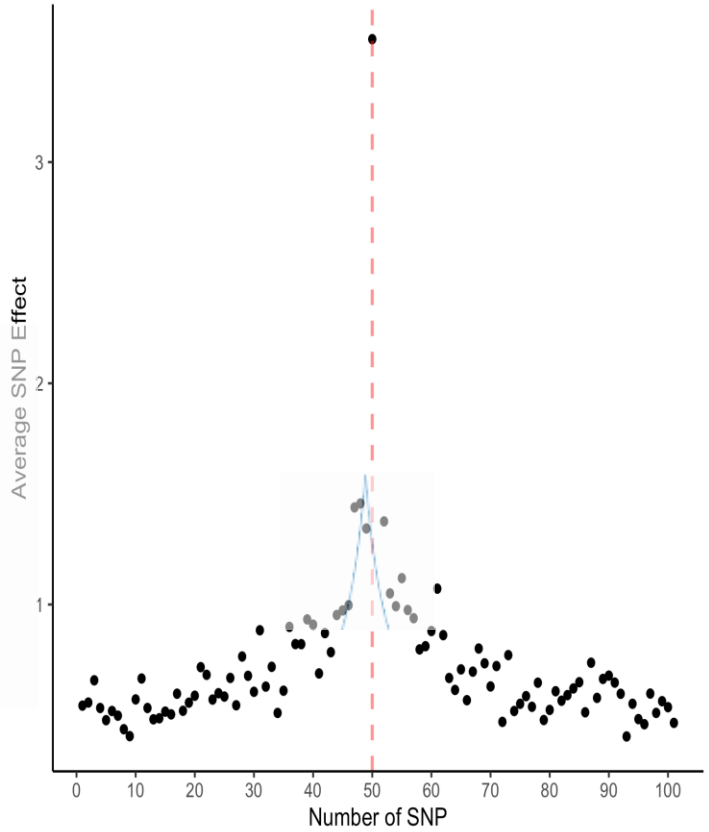
Ne=60



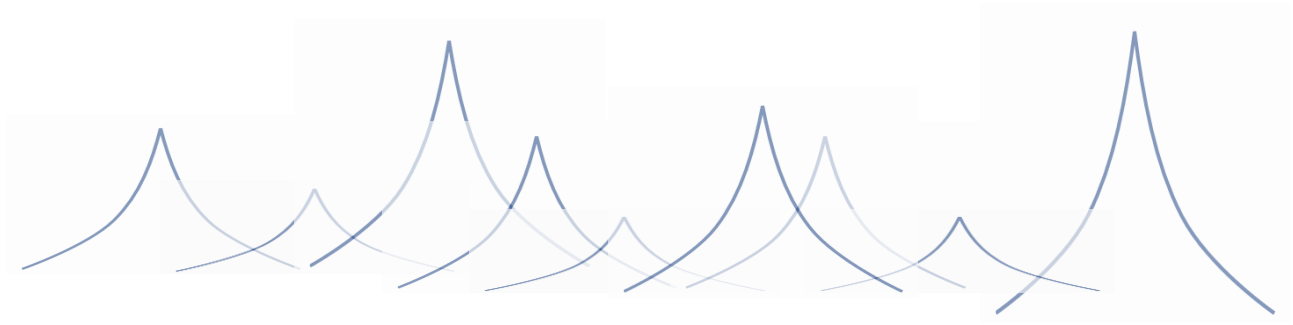
Ne=60 3X



Ne=600



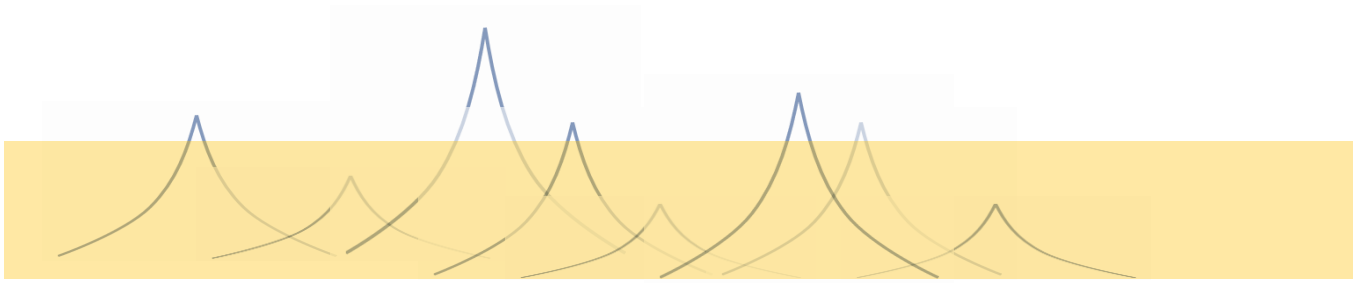
What is Manhattan plot composed of?



QTNs

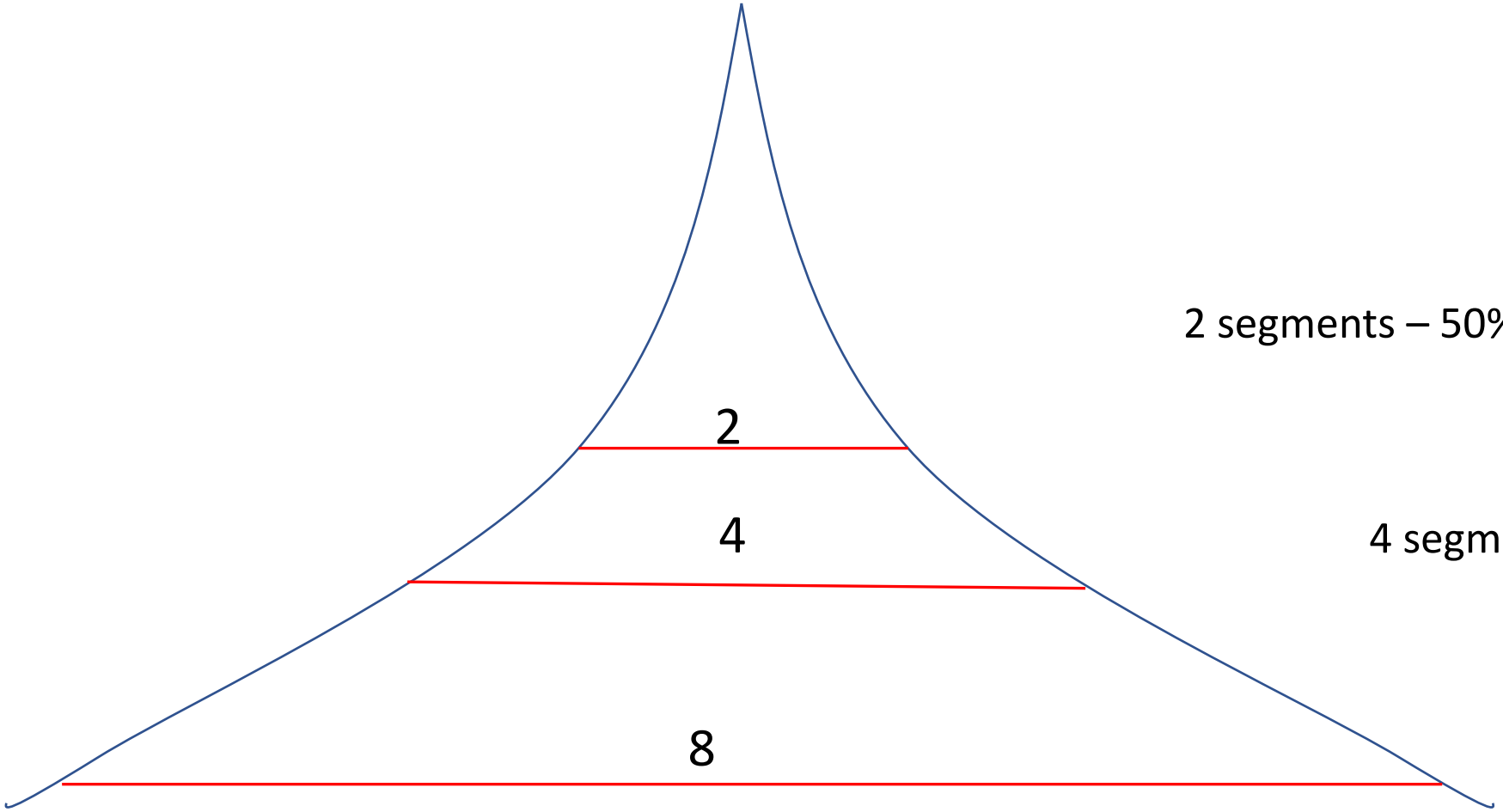
Relationships

Noise



Composite plot

Proportion of QTN variance explained by different number of segments

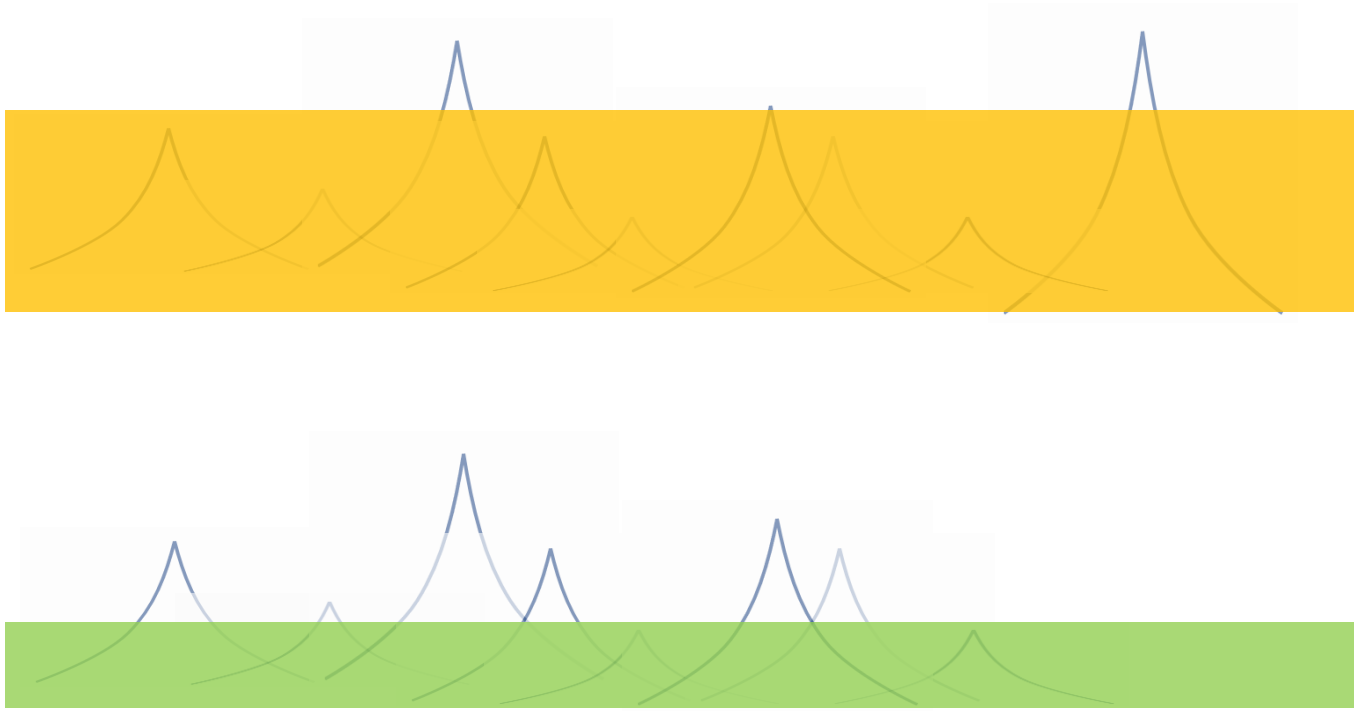


2 segments – 50% QTN variance

4 segments – 66% QTN variance

8 segments – 80% QTN variance

Size of data

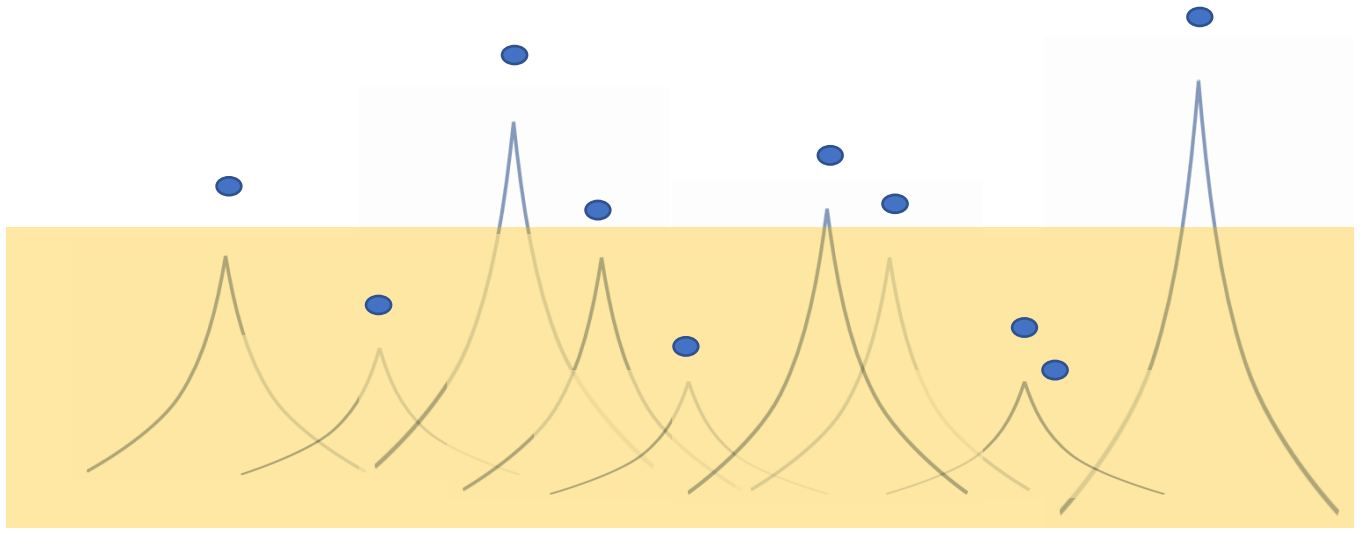


Small data

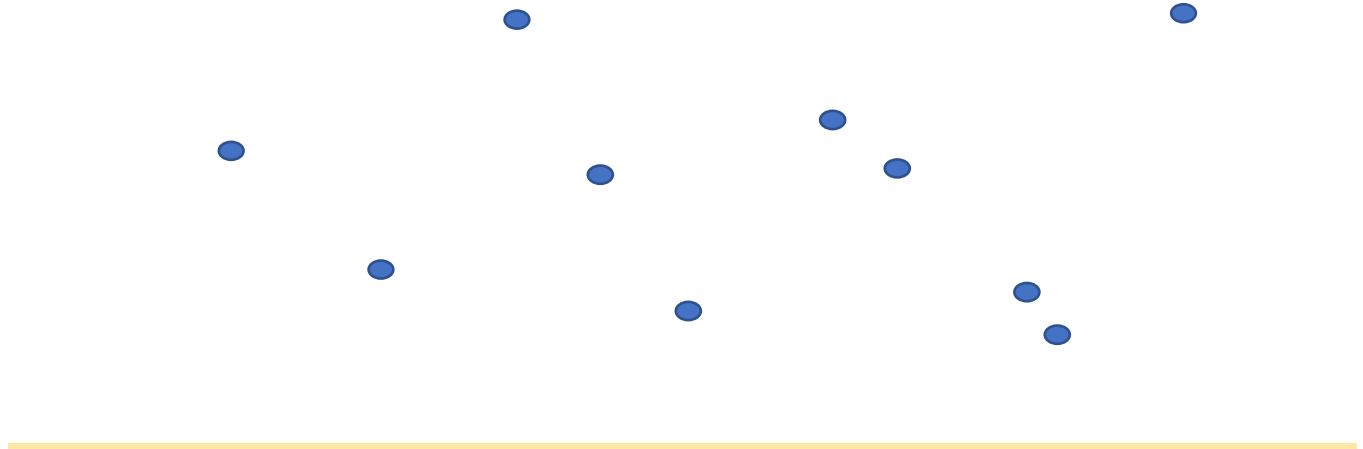
Large data

If GWAS by sliding window, optimal size 4-8 Stam segments
1-2 Mb cattle, 3-6 Mb pigs and chicken

+ QTN in the data

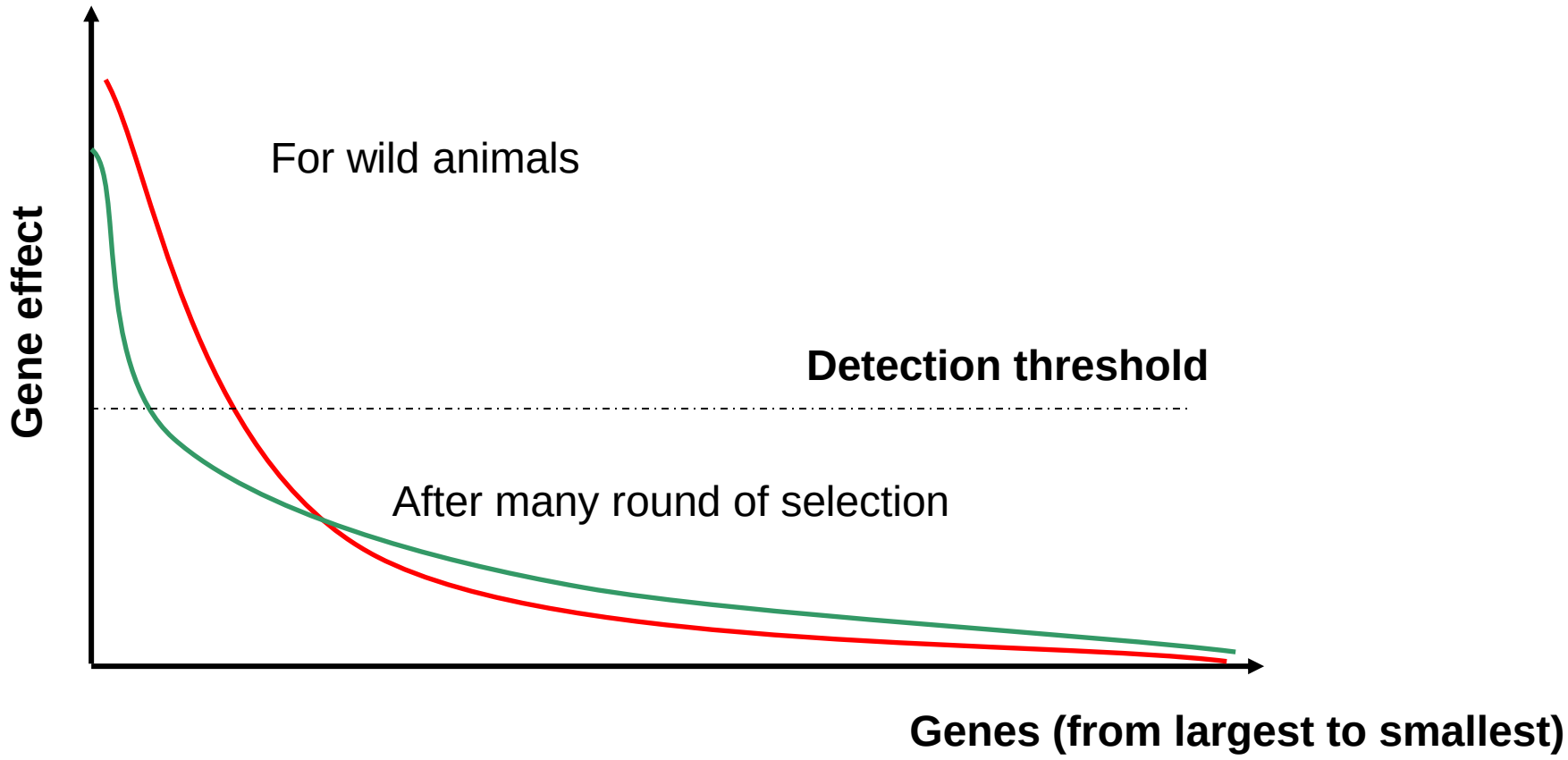


All SNP including QTN unweighted



QTN properly weighted

Distribution of QTL effects



Other factors

- Bayesian methods pick up large signals due to QTL, relationships and noise
  - If small signals undetected, probably large signals inflated

- Large signals without LD curves due to imputation errors (Li Ma, personal communication)

Conclusions

- QTN profile wide with small effective population size
  - Optimal window size 4 to 8 segments
  - About 1-2 Mb at Ne=100
- Sharp peak for QTN
- Large signals in GWAS due to QTN, relationships and noise and imputation
- Most QTNs probably below detection limit

UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences

Acknowledgements

USDA NIFA
ANGUS THE BUSINESS BREED
AMERICAN ANGUS ASSOCIATION
3201 Frederick Avenue • St. Joseph, MO 64506 • (816) 383-5100 • Fax (816) 233-9703 • E-mail: angus@angus.org
AGI
zoetis
WAGYU
AMERICAN WAGYU ASSOCIATION
PIC
U.S. Registered Holsteins
FOR MAXIMUM PROFIT
HOLSTEIN ASSOCIATION USA, INC.
Cobb
Smithfield
THE MASCHHOFFS
Progressive Farming. Family Style.
DNA GENETICS
Greater gains. Improved performance.
methodgenetics
USDA United States Department of Agriculture
Agricultural Research Service
Warmwater Aquaculture Research Unit
PLEMDAT
ABCZ
SEMEX
Genetics for Life®
PFHBiPM