Estimation of non-additive variance components and SNP effects in five dairy cattle breeds

Andrew Marete

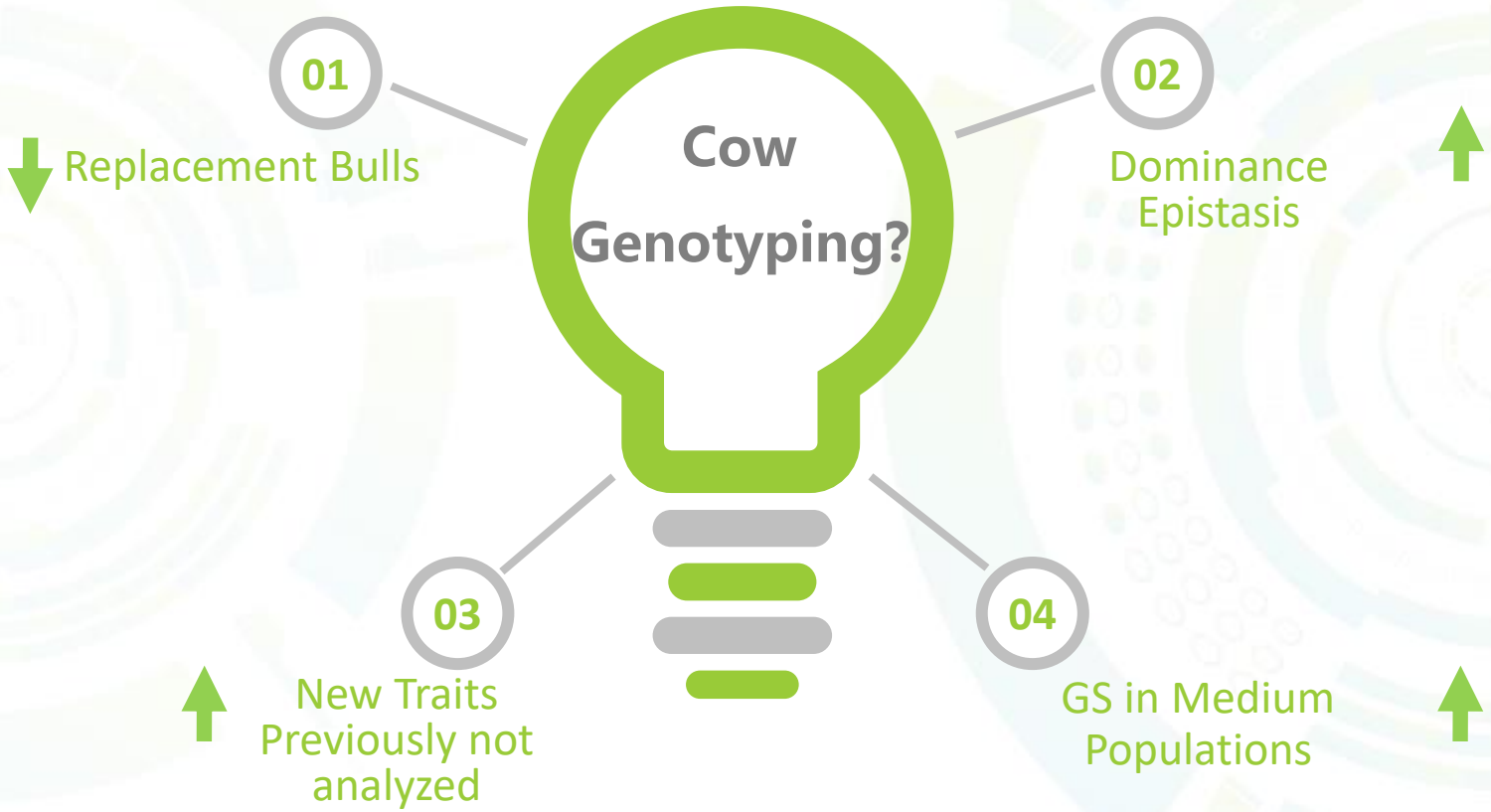
The Premise

- Milk production traits are affected by many genes, many with a small effect and some with a major effect, thus complex traits
- Complex traits are regulated by complex interplay between multiple genes
- Theoretically, these genes can interact
- Interaction can be explained as non-additive genetic variance (NAGV)
    1. same locus interaction (dominance deviation)
    2. different loci interaction (epistatic variance)

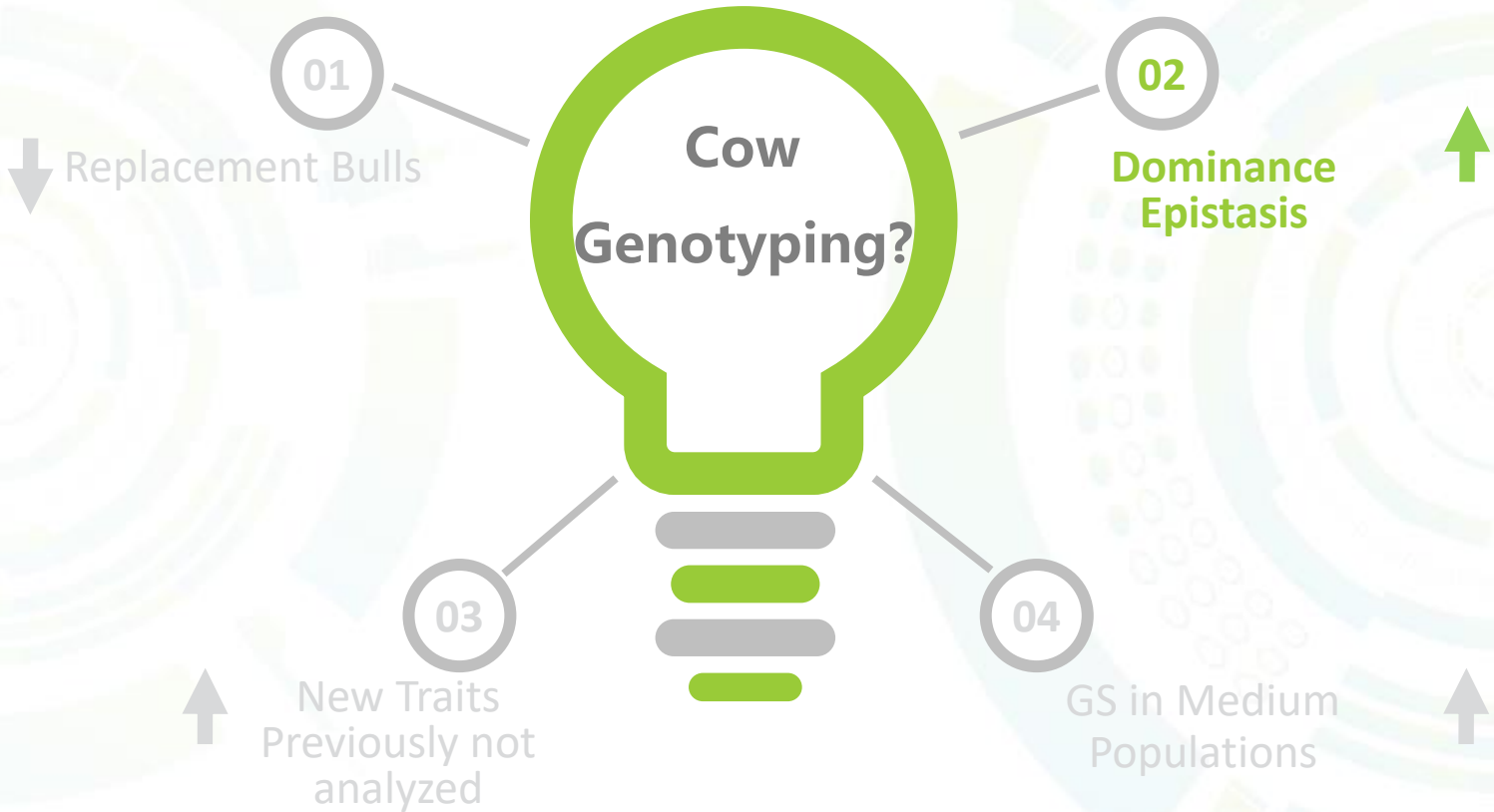
The challenge

- Non-additive genetic variance (NAGV) is usually omitted in Genomic Selection (GS) and when estimating life-time merit of cows
- GS is based on additive effects of sire
- Especially the case when reference population is made of progeny tested bulls
- But cows with own performances are becoming more popular as a resource of choice for GS
- If NAGV is substantial, it can ameliorate the life-time merit of cows if considered in addition to additive variation

The promise


01
Replacement Bulls
Cow
Genotyping?
02
Dominance
Epistasis
03
New Traits
Previously not
analyzed
04
GS in Medium
Populations

For this study



The workflow

INPUTS
Phenotypes
Genotypes

OUTPUTS
Non-Additive
Genetic Variances

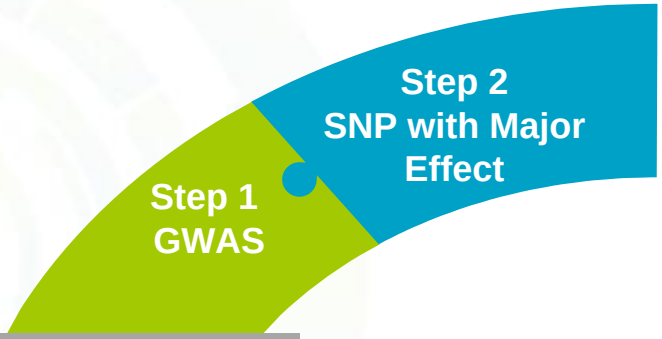
The workflow

Step 1
GWAS

INPUTS
Phenotypes
Genotypes

OUTPUTS
Non-Additive
Genetic Variances

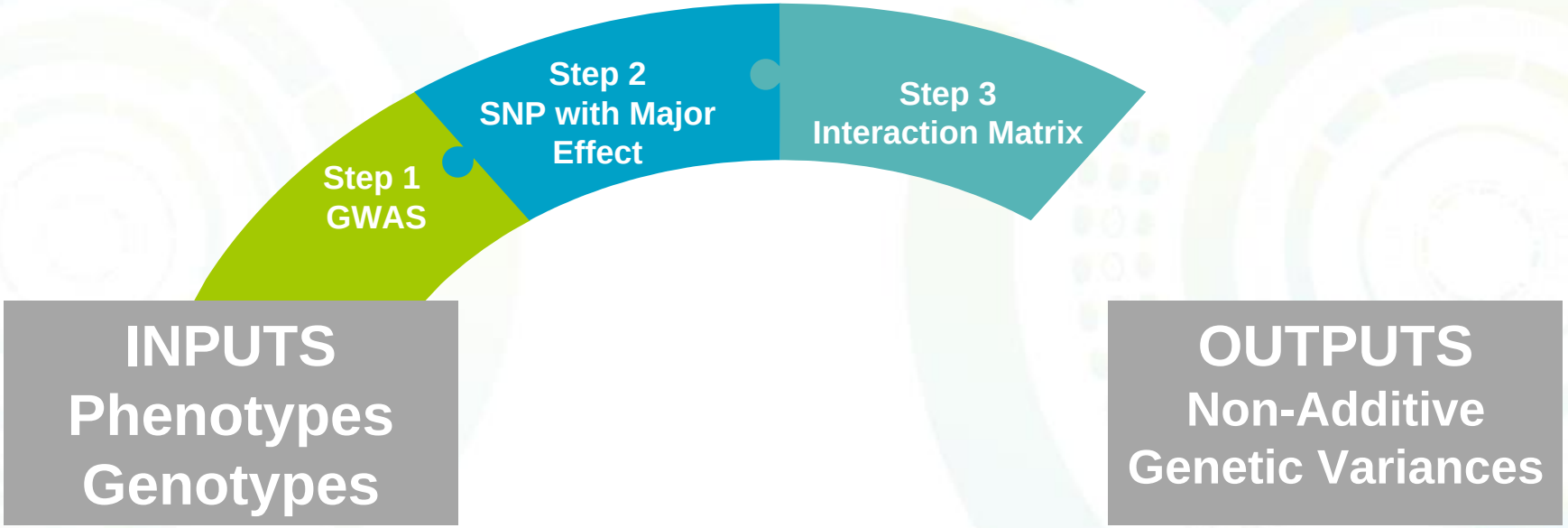
The workflow



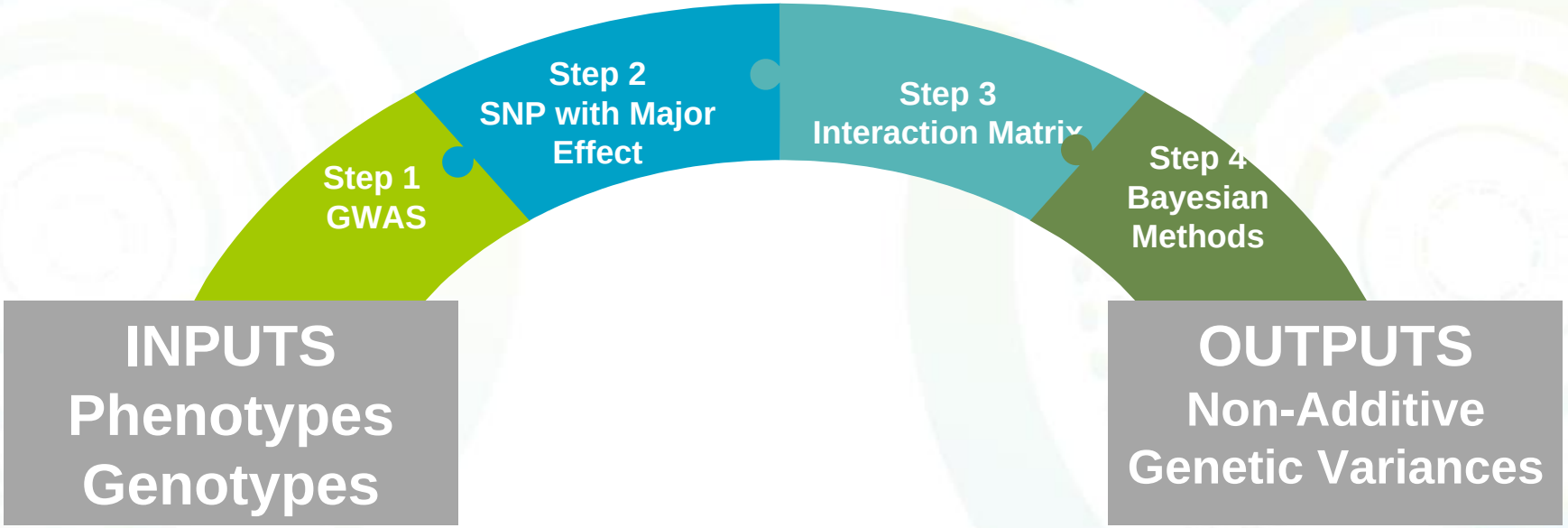
INPUTS
Phenotypes
Genotypes

OUTPUTS
Non-Additive
Genetic Variances

The workflow



The workflow



Data

Breeds	Phenotypes	Genotypes
	Yield Deviation: Milk, Protein, Fat	
	11,000	• Genotyped with 50k SNP chip • 6,132 SNPs selected from whole genome sequence • All imputed to 50k density
	20,000	
	46,000	
	10,000	
	12,000	

Selection of SNP with Major Effect

GWAS

$$\boldsymbol{\varepsilon} \sim N(0,\ \boldsymbol{I}\sigma_{\varepsilon}^{2})$$

phenotypes — fixed effects — Residual effects

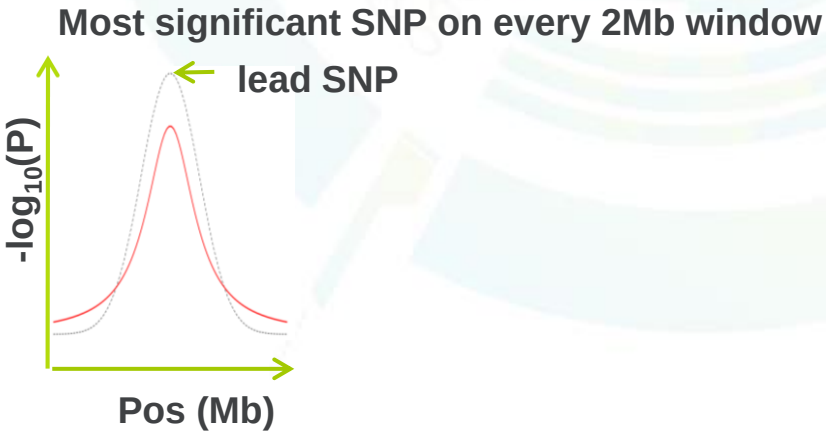
$$\boldsymbol{y} = x\boldsymbol{\beta} + \boldsymbol{g} + \boldsymbol{\varepsilon}$$

genetic effects as explained by all SNP

$$\boldsymbol{g} \sim N(0, \boldsymbol{G}\sigma_{g}^{2})$$

$$\mathrm{var}(\mathbf{y}) = \boldsymbol{G}\sigma_{g}^{2} + \boldsymbol{I}\sigma_{\varepsilon}^{2}$$

- **539 SNP with major Effect**



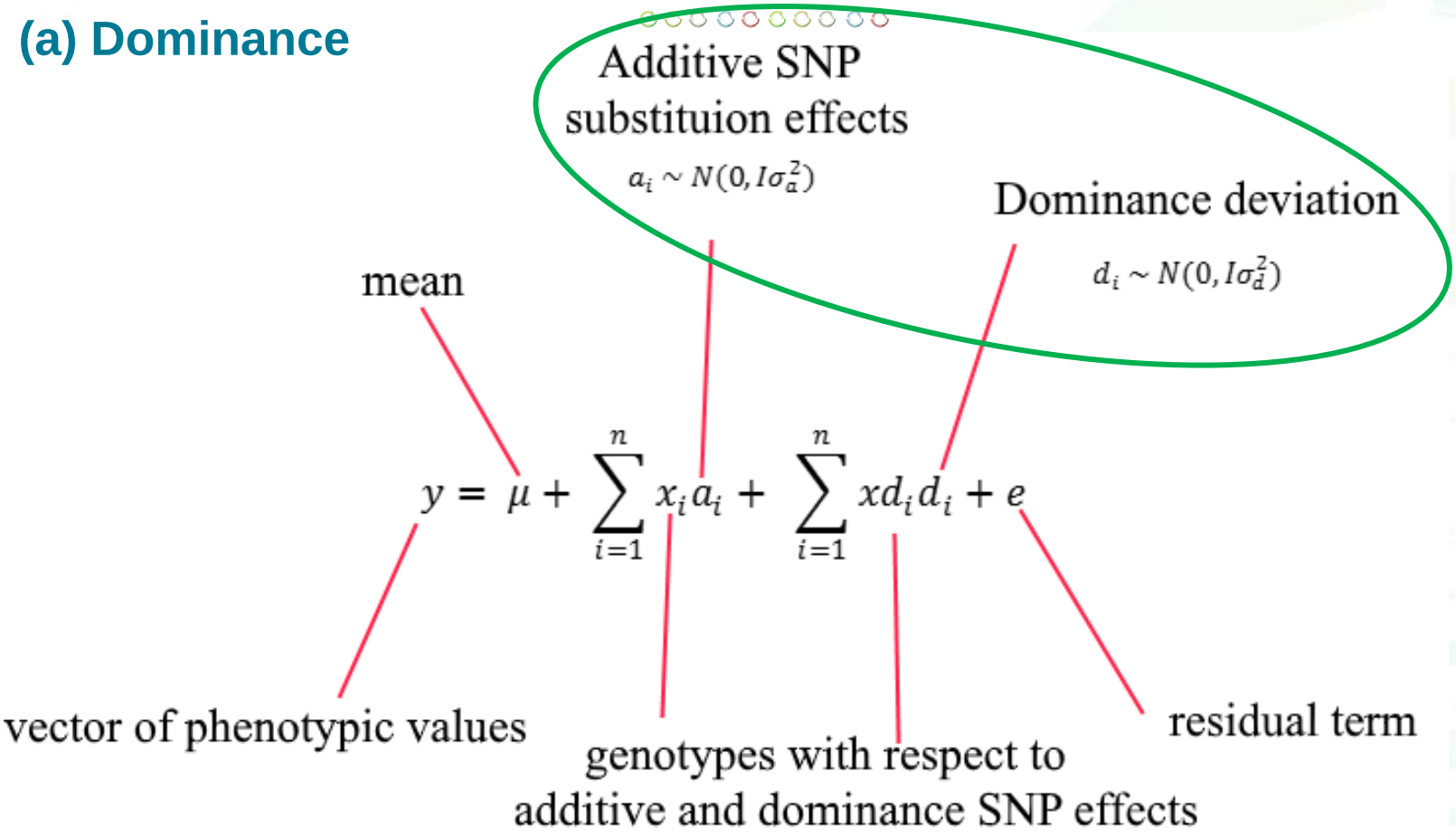
Non-additive models

- **Symmetric marker coding**

W_{ij}	**Locus k**		
Locus j	AA	AB	BB
AA	1	0	-1
AB	0	0	0
BB	-1	0	1

Non-additive models

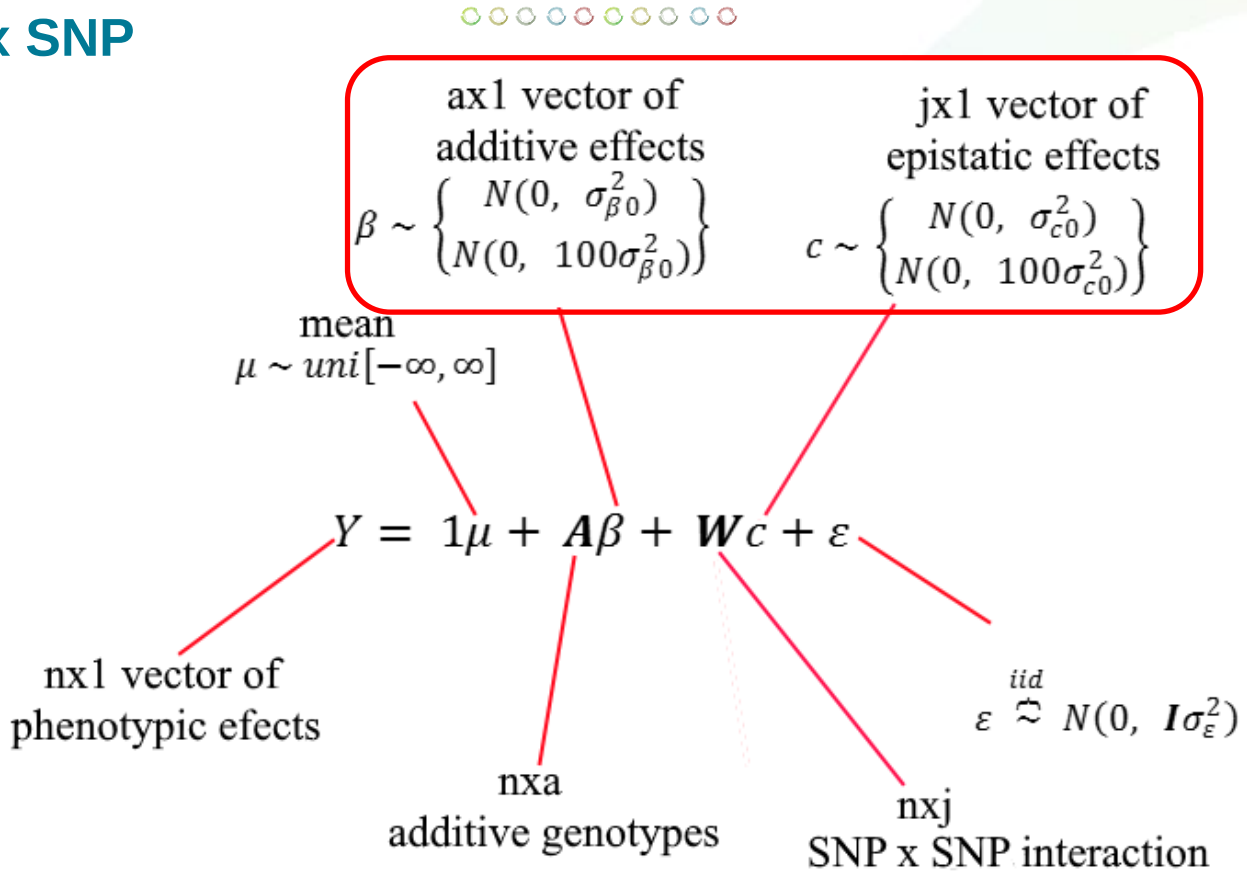
(a) Dominance



- **Chain=200k, burn=50k, skip=0.5k**
- **Metropolis-Hastings sampler – (a)**

Non-additive models

(b) SNP x SNP



$$(\pi\beta_0, \pi 100\beta_0) \sim beta(10,1) \; and \; (\pi c_0, \pi 100 c_0) \sim beta(10,1)$$

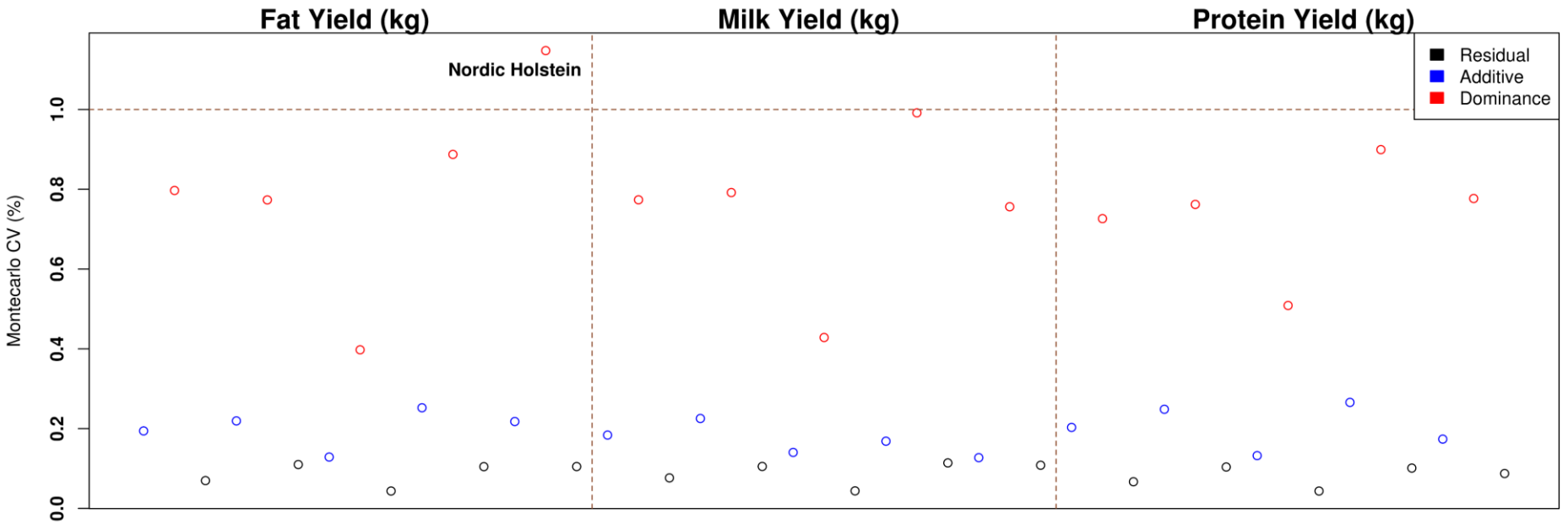
$$\sigma^2_{\beta 0}, \sigma^2_{c0}, \sigma^2_{\varepsilon 0} \sim uni[0, \infty]$$

- **Chain=200k, burn=50k, skip=0.5k**
- **Bernoulli distribution**

Model Convergence

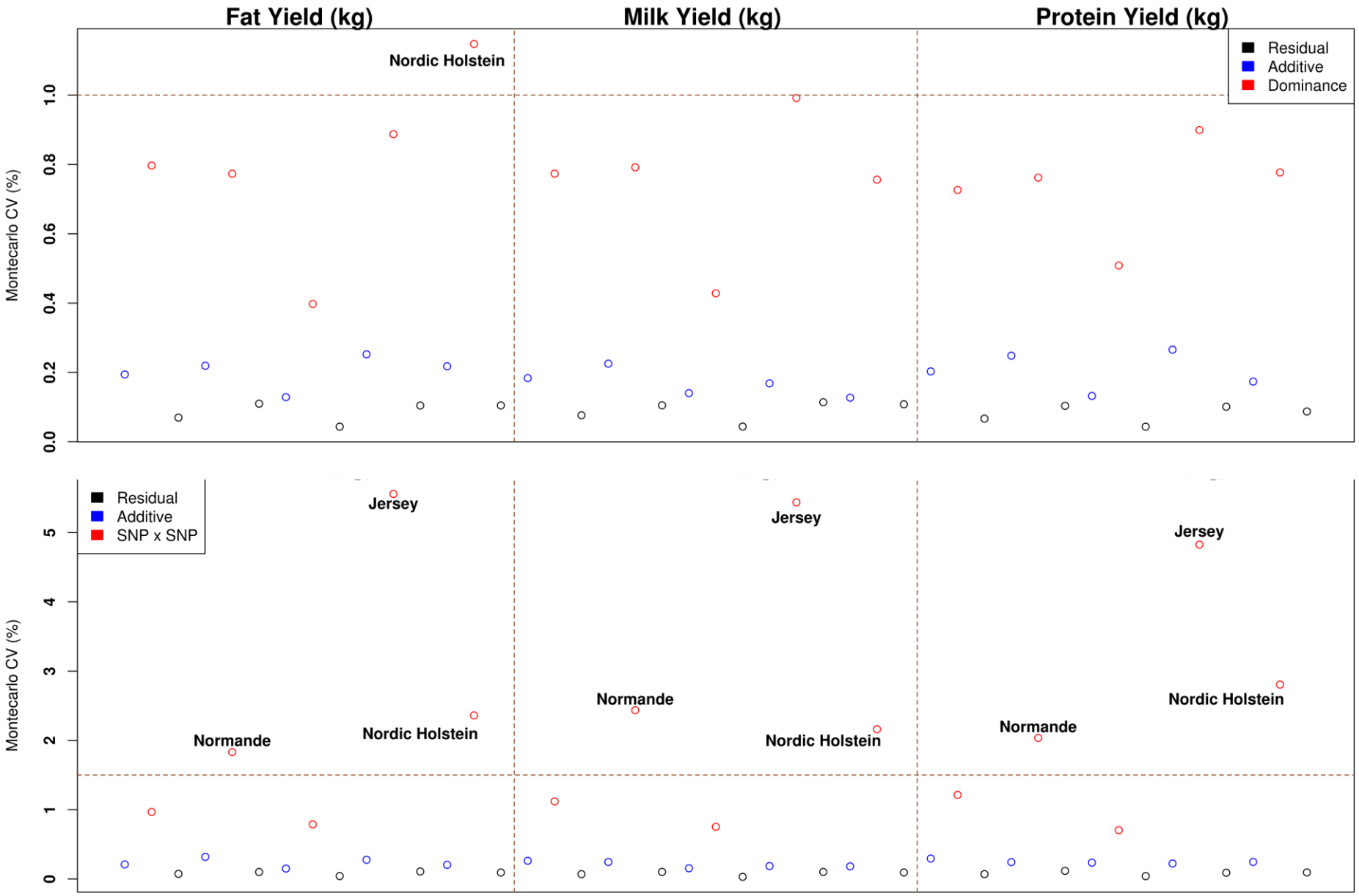
$$\textbf{MonteCarlo Coefficient of variation} = \frac{\textbf{Time Series SE}}{\textbf{Parameter Mean}}$$

Model Convergence



$$\text{MonteCarlo Coefficient of variation} = \frac{\text{Time Series SE}}{\text{Parameter Mean}}$$

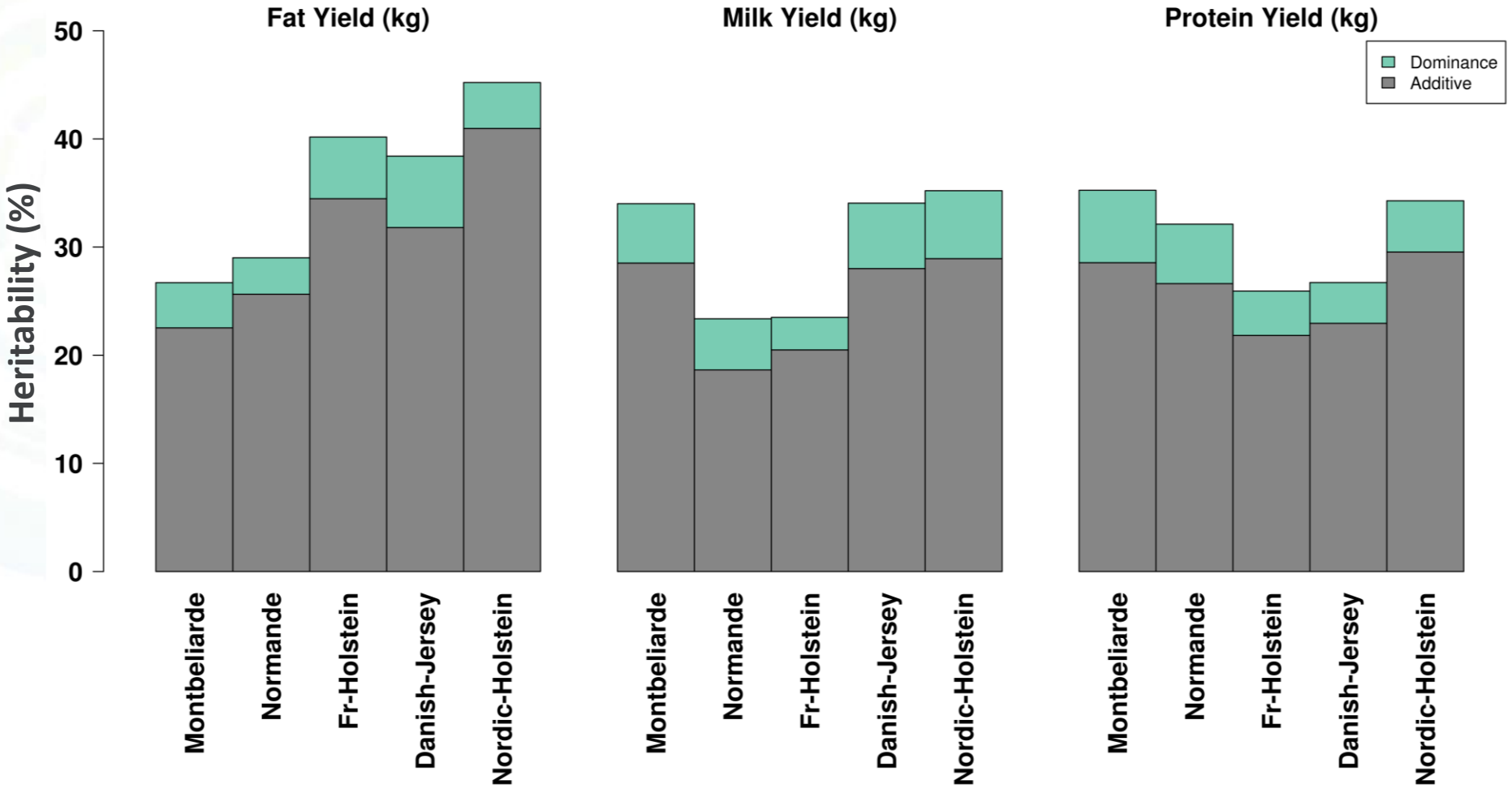
Model Convergence



Heritability

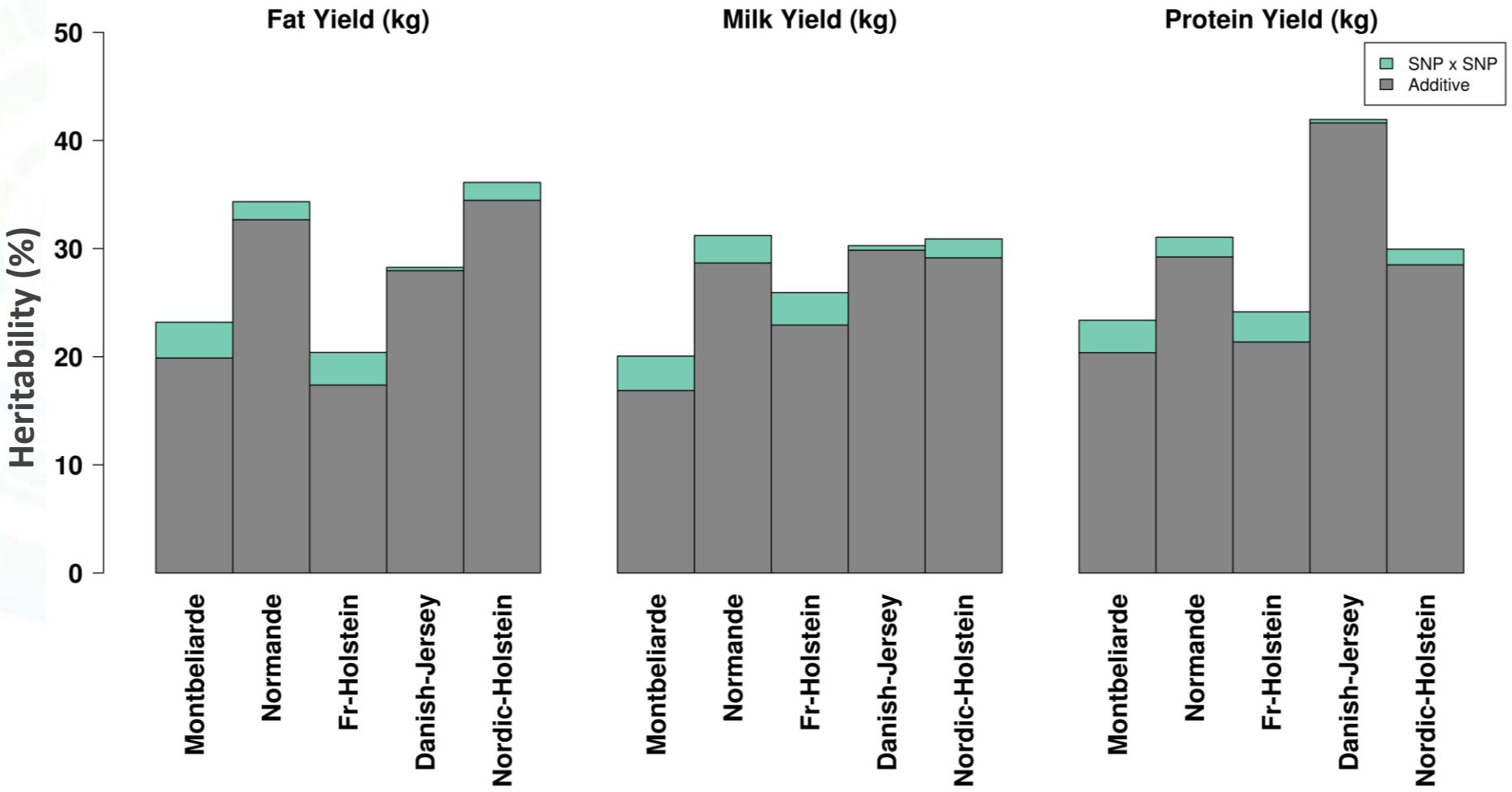
variation in a trait due to variation in genetic factors as explained by SNP

Dominance



Heritability

SNP x SNP



Allele Substitution Effect

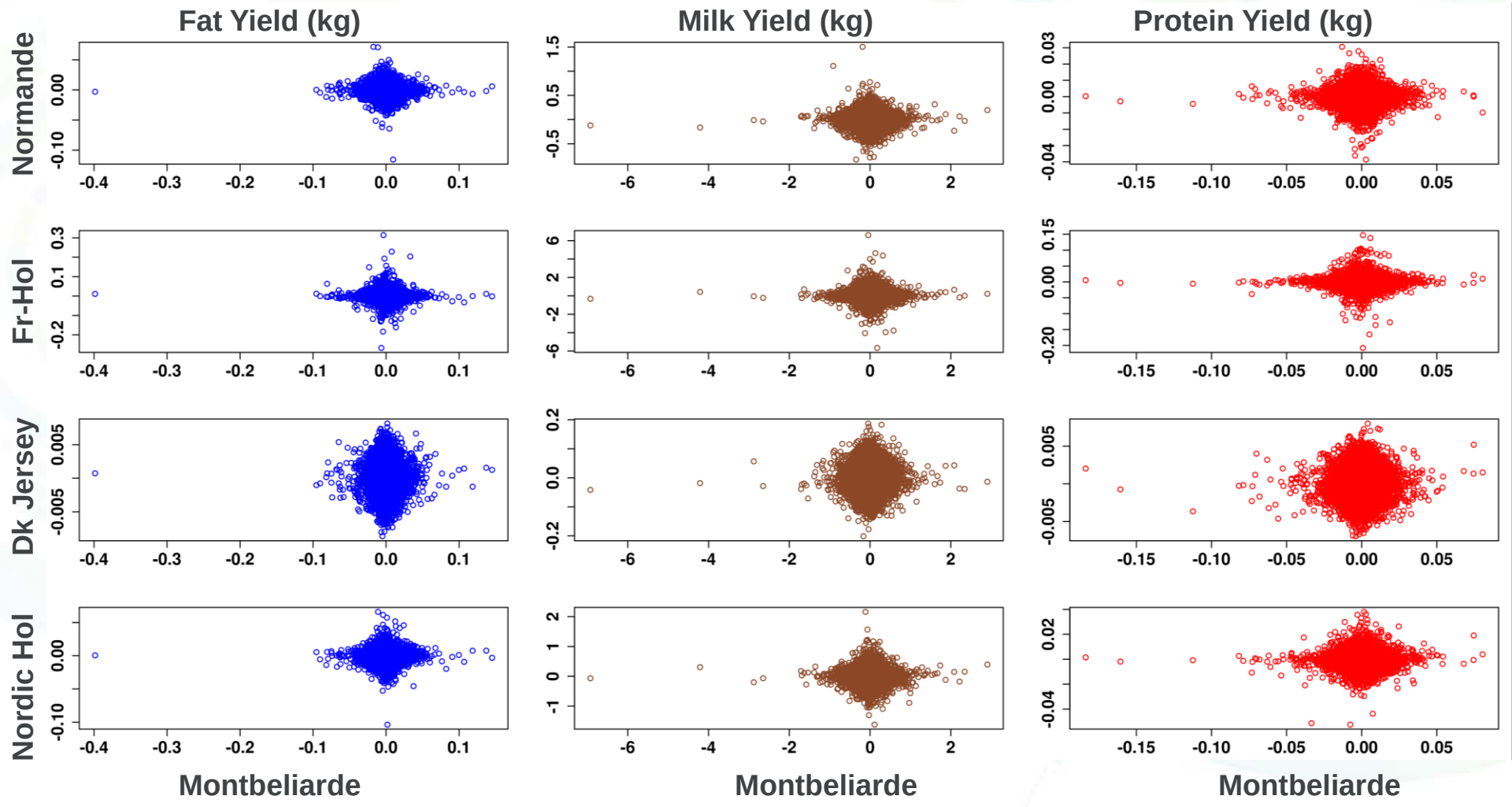
SNP x SNP effects

1. **Across breeds same trait**
2. **Within breed different traits**

Allele Substitution Effect

SNP x SNP effects

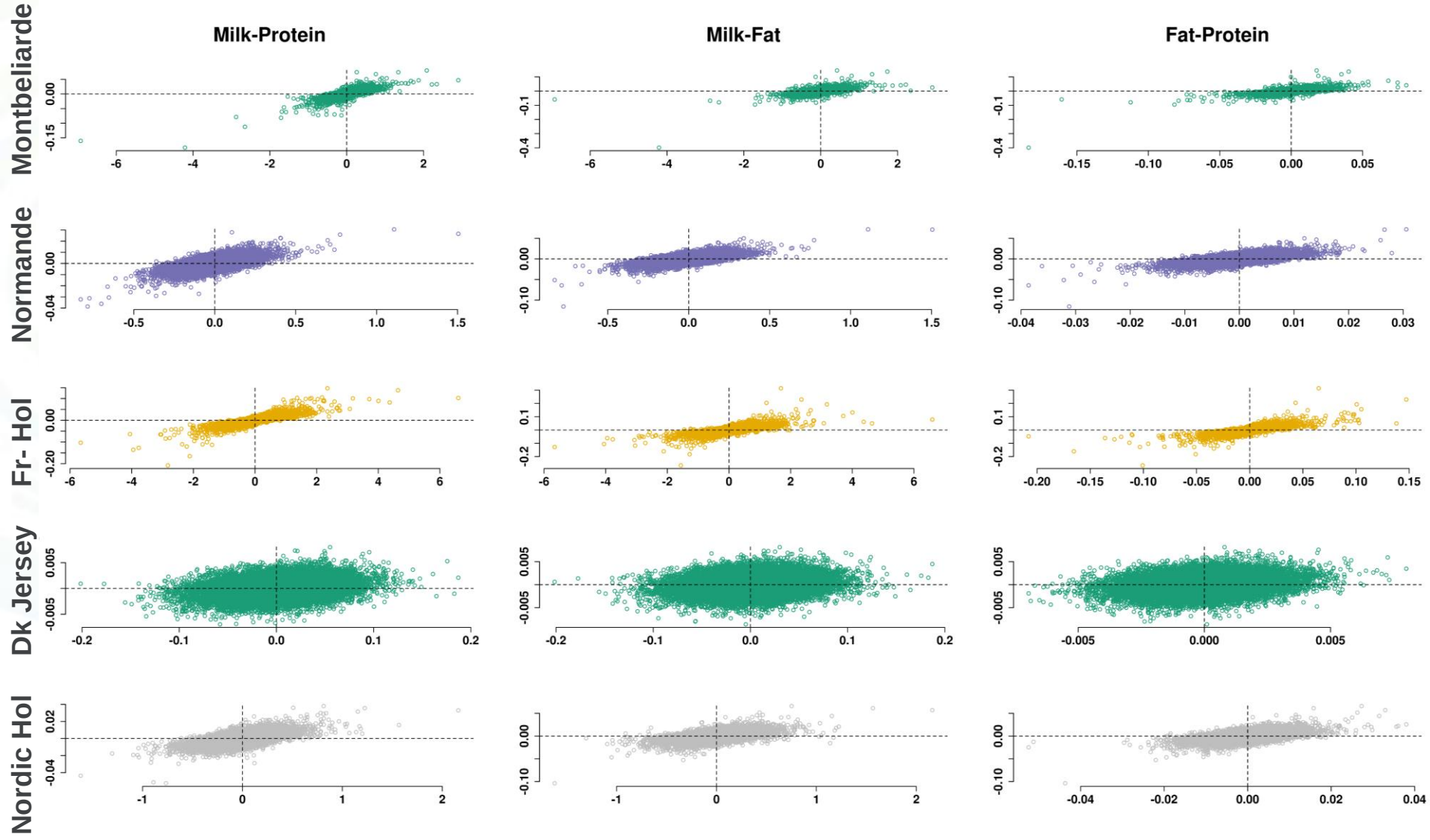
1. Across breeds same trait



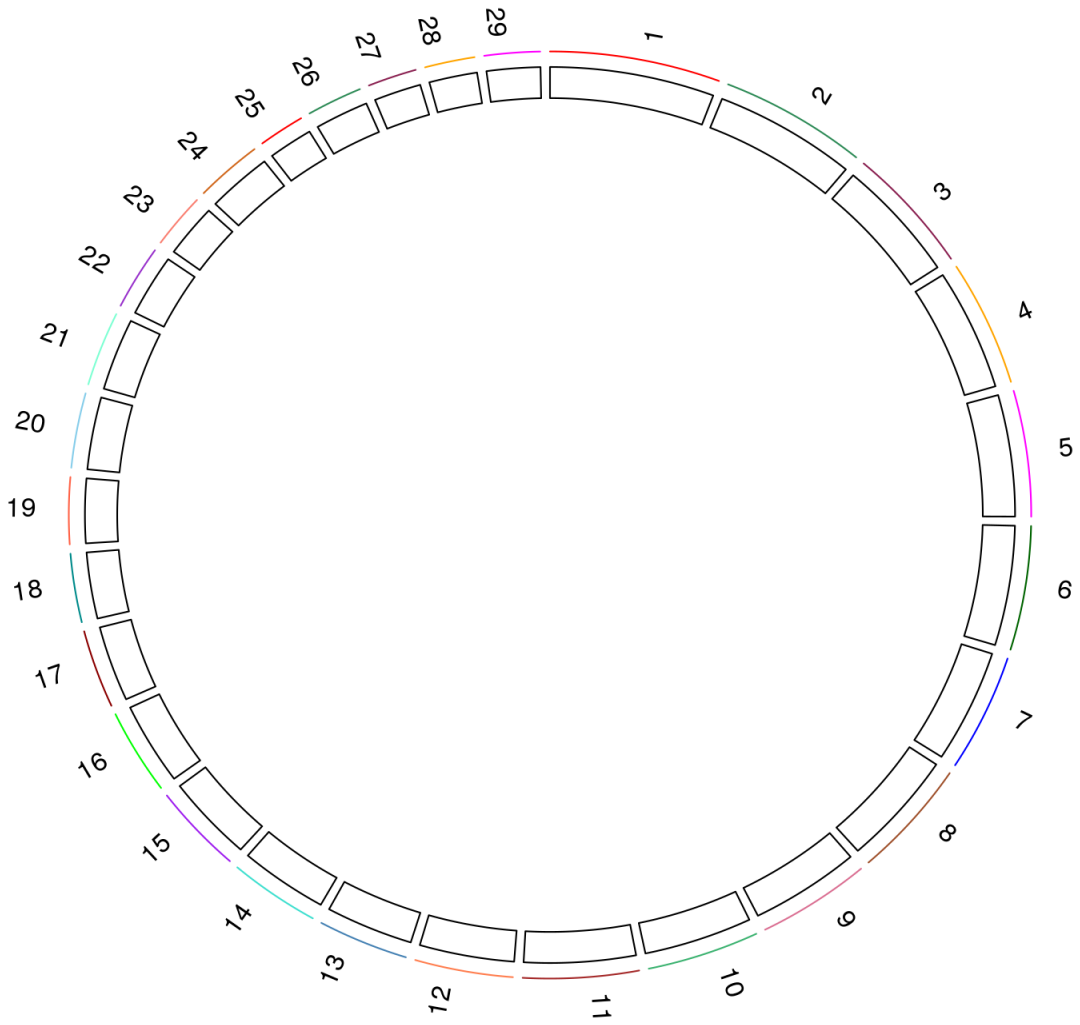
Allele Substitution Effect

SNP x SNP effects

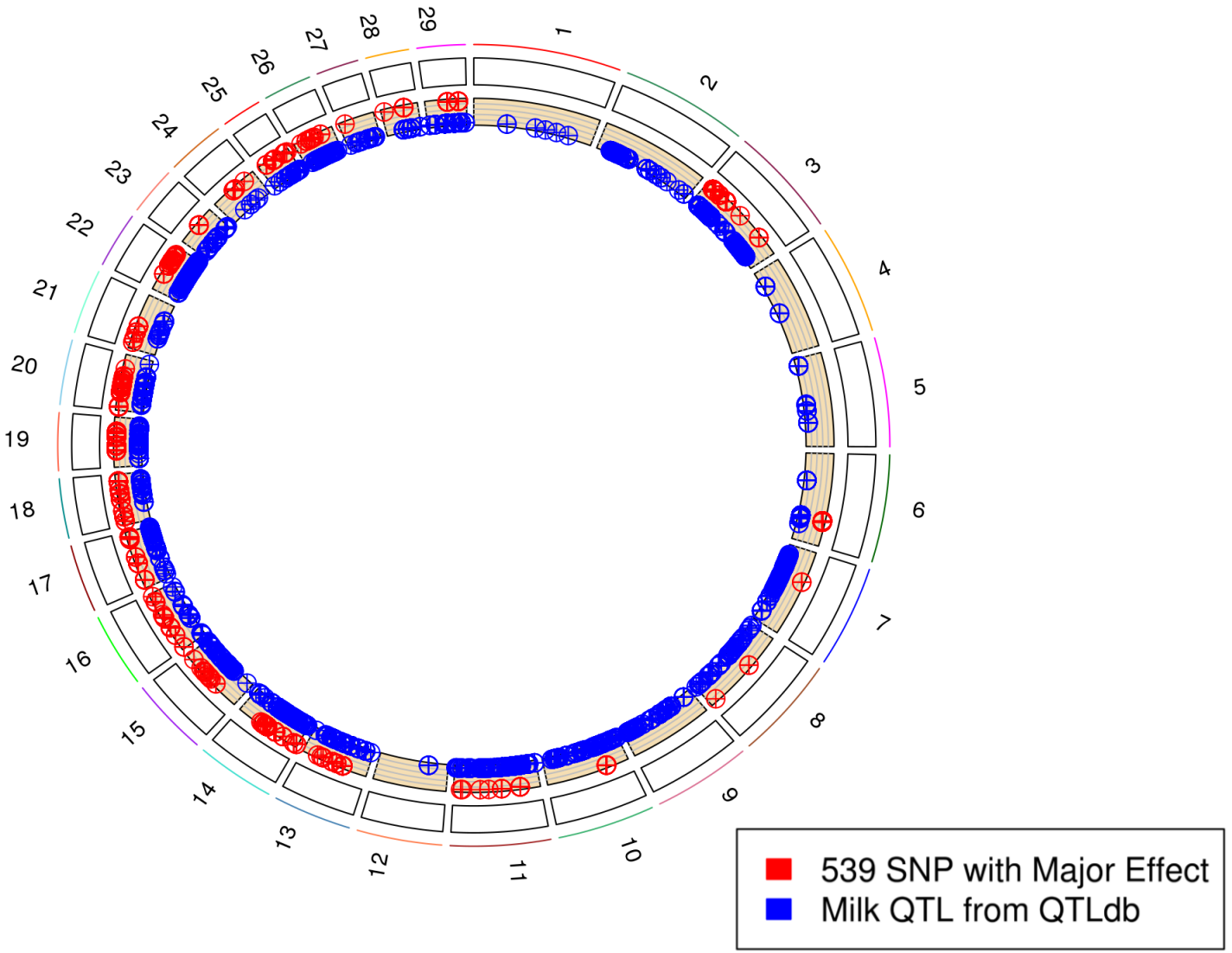
2. Within breed different traits



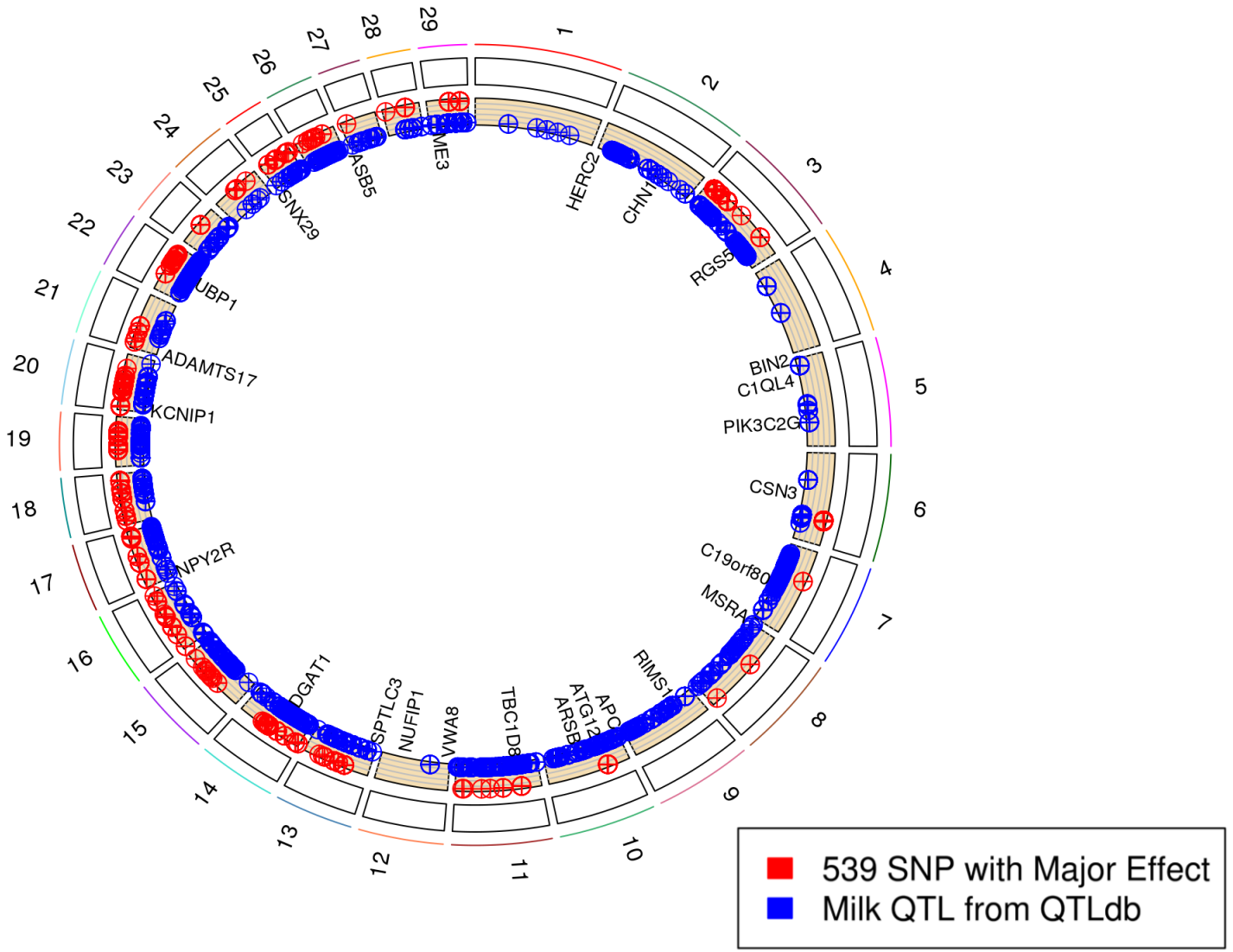
In Summary: Holstein Milk Yield



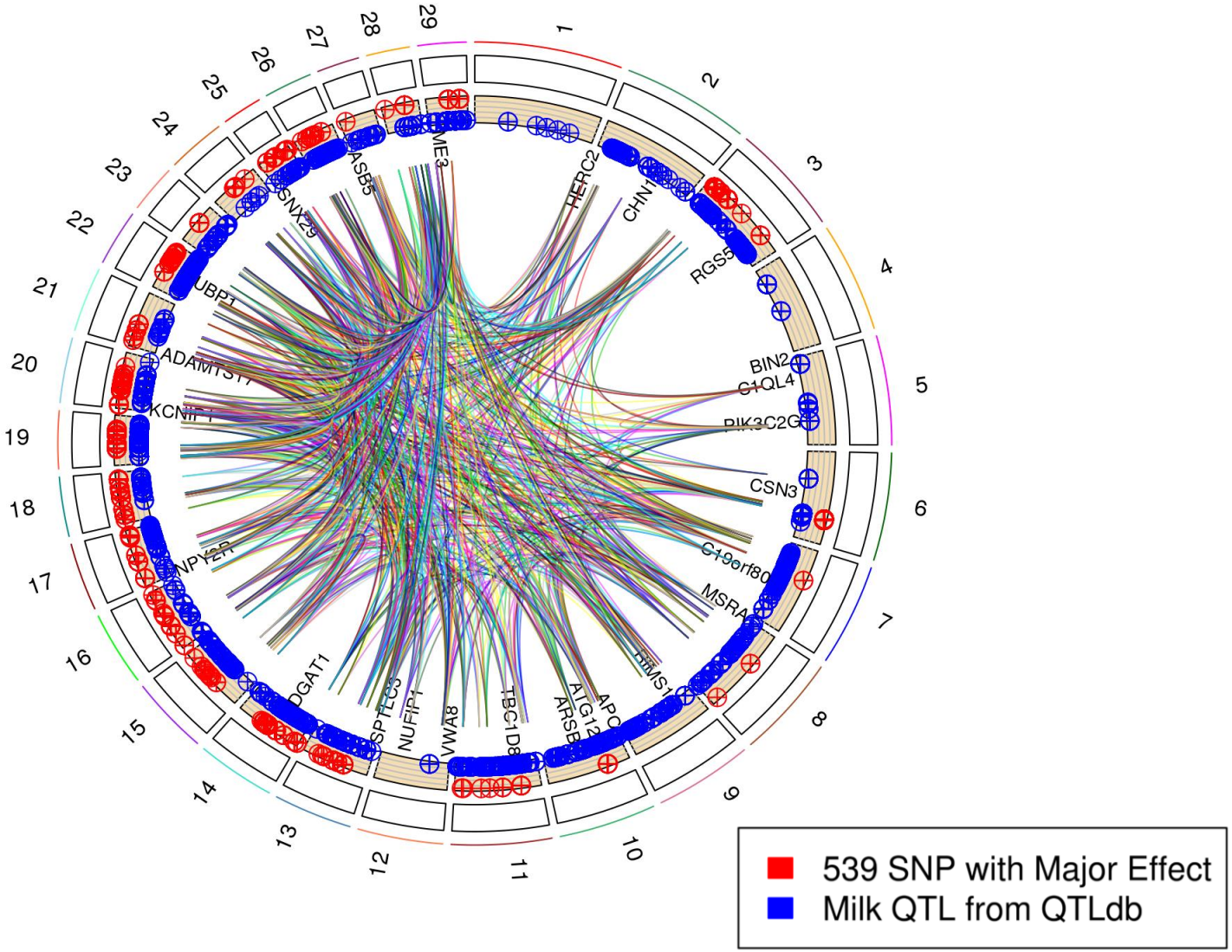
In Summary: Holstein Milk Yield



In Summary: Holstein Milk Yield



In Summary: Holstein Milk Yield



Conclusion

SNP with major effect including those selected from whole genome sequence explain non-additive variation

This interaction matrix can augment the normal GRM when predicting breeding accuracies

Thank you for your attention

Any questions?