Using expression data to detect small QTL in dairy cattle

Irene van den Berg[1] *(irene.vandenberg@unimelb.edu.au)*, *BJ Hayes*[2,3], *ME Goddard*[1,3]
[1]Faculty of Veterinary & Agricultural Science, University of Melbourne, Parkville, Victoria, Australia
[2]Queensland Alliance for Agriculture and Food Innovation, Centre for Animal Science, University of Queensland, St Lucia, Queensland, Australia
[3]Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, Victoria, Australia

GenSAP

THE UNIVERSITY OF MELBOURNE

AGRICULTURE VICTORIA

Introduction

- Sequence data → detect causative mutations, use for genomic prediction
- Most QTL have a small effect → difficult to detect
- Many variants associated with gene expression → eQTL → use to detect QTL?

Introduction

- Sequence data → detect causative mutations, use for genomic prediction
- Most QTL have a small effect → difficult to detect
- Many variants associated with gene expression → eQTL → use to detect QTL?

Objective

Detect variants associated with QTL and eQTL

-Correlate local GEBVs & gene expression

-Colocalisation of QTL and eQTL

Methods: Correlate local GEBVs & gene expression

- Calculate local GEBVs:
    - HD genotypes of 35,775 Holstein, Jersey and crossbred bulls and cows
    - DTD / TD for milk, fat and protein yield, milk and fat percentage, and fertility
    - SNP effects estimated using Bayes R hybrid
    - Sum up SNP effects over 250kb windows
    - Select intervals that explain at least 1/10,000th of the total additive genetic variance

Methods: Correlate local GEBVs & gene expression

- Calculate local GEBVs:
  - HD genotypes of 35,775 Holstein, Jersey and crossbred bulls and cows
  - DTD / TD for milk, fat and protein yield, milk and fat percentage, and fertility
  - SNP effects estimated using Bayes R hybrid
  - Sum up SNP effects over 250kb windows
  - Select intervals that explain at least 1/10,000th of the total additive genetic variance

- Gene expression:
  - 105 Holstein cows (milk + blood samples) and 26 Jersey cows (milk samples)
  - Gene expression of each gene within 1Mb of an interval

Methods: Correlate local GEBVs & gene expression

- Calculate local GEBVs:
    - HD genotypes of 35,775 Holstein, Jersey and crossbred bulls and cows
    - DTD / TD for milk, fat and protein yield, milk and fat percentage, and fertility
    - SNP effects estimated using Bayes R hybrid
    - Sum up SNP effects over 250kb windows
    - Select intervals that explain at least 1/10,000th of the total additive genetic variance

- Gene expression:
    - 105 Holstein cows (milk + blood samples) and 26 Jersey cows (milk samples)
    - Gene expression of each gene within 1Mb of an interval

→ Correlate gene expression & local GEBVs, select correlations with $p \leq 10^{-5}$

Methods: GWAS & eQTL

- GWAS
  - Same individuals as used for local GEBVs, imputed up to full sequence data, ~4.8M variants after filtering & LD pruning
  - Mixed linear model association analysis in GCTA (Yang *et al.*, 2011), separate for bulls & cows
  - Meta-analysis of bulls & cow GWAS using weighted z-scores model in Metal (Willer *et al.*, 2010)

Methods: GWAS & eQTL

- GWAS
  - Same individuals as used for local GEBVs, imputed up to full sequence data, ~4.8M variants after filtering & LD pruning
  - Mixed linear model association analysis in GCTA (Yang *et al.*, 2011), separate for bulls & cows
  - Meta-analysis of bulls & cow GWAS using weighted z-scores model in Metal (Willer *et al.*, 2010)

- eQTL analysis
  - Same cows as gene expression
  - ~10 million sequence variants, 12,772 genes in milk and 11,577 in blood
  - For all genes, association of expression level with all variants on the same chromosome

Methods: Colocalisation of eQTL and QTL

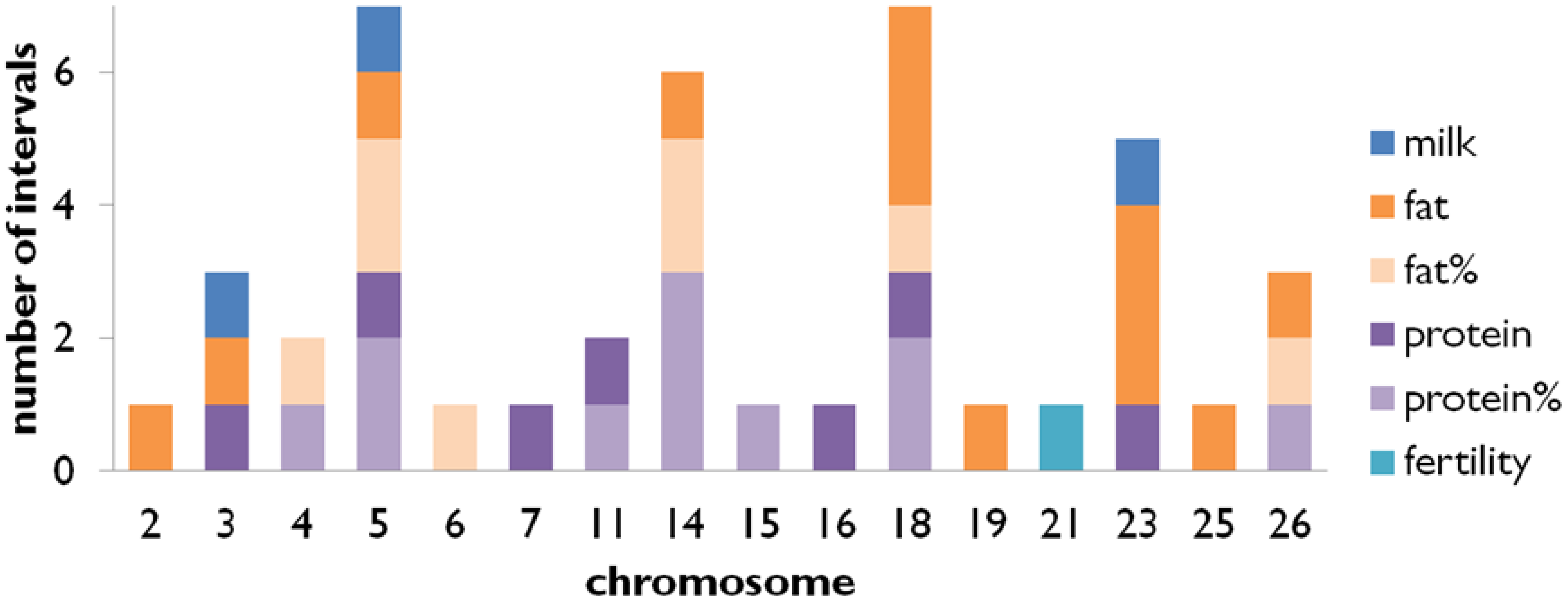
- Colocalisation posterior probability (CLPP) = the probability that the same variant is causal for QTL and eQTL

- eCaviar software (Hormozdiairi *et al.*, 2016)

- CLPP computed based on GWAS & eQTL effects and LD between variants

- CLPP ≥ 0.01 → colocalisation

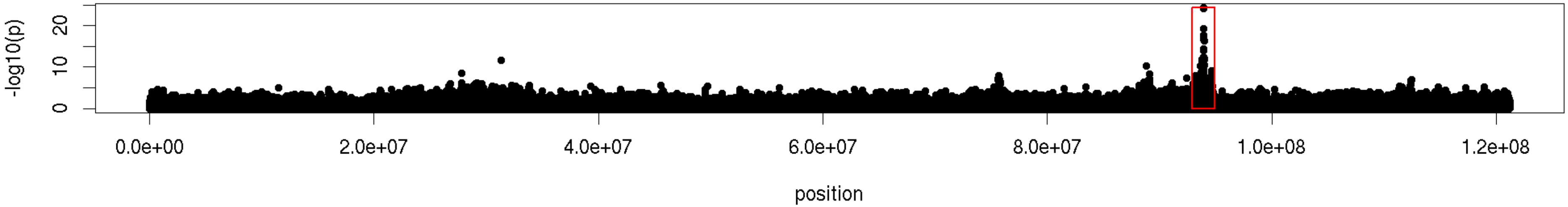
Intervals with significant correlations between local GEBVs & gene expression



Example: MGST1, chr 5, milk yield

Known QTL and eQTL (Littlejohn *et al.*, 2016)

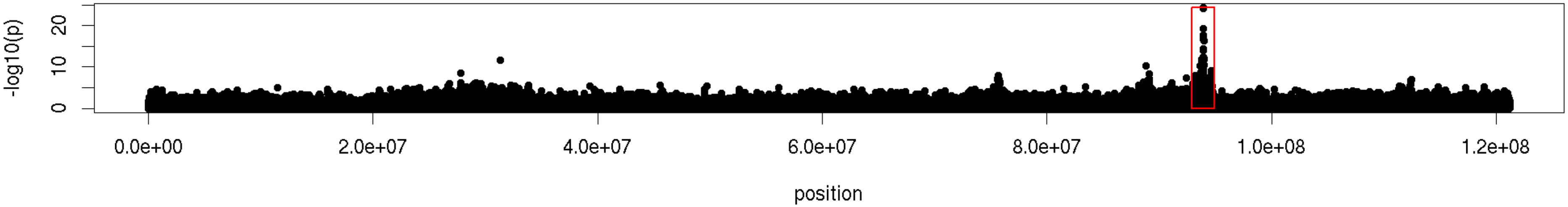
GWAS



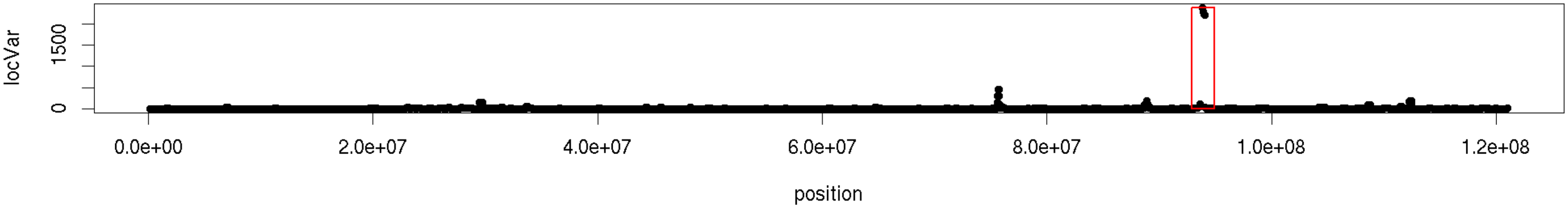
Example: MGST1, chr 5, milk yield

Known QTL and eQTL (Littlejohn *et al*., 2016)

GWAS



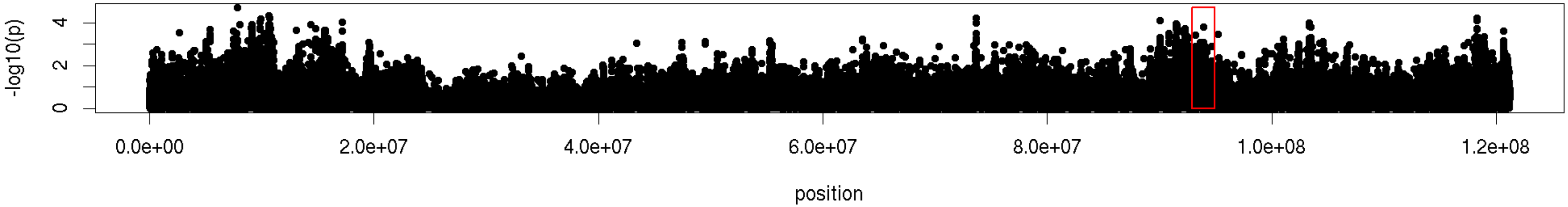
var(locGEBV)



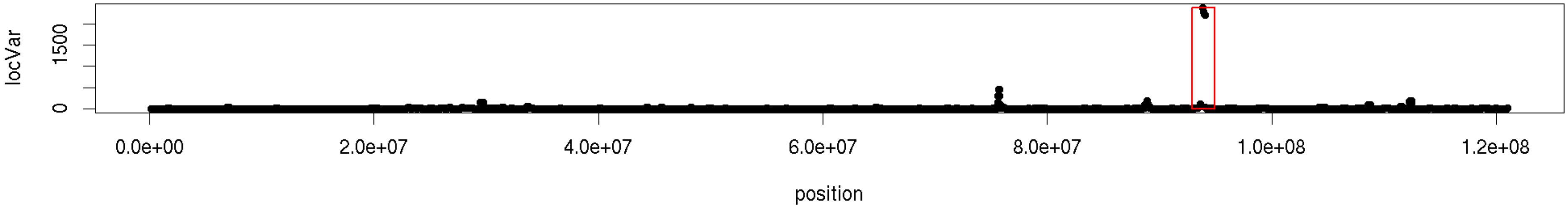
Example: MGST1, chr 5, milk yield

In our study, p=0.008 for eQTL reported by Littlejohn *et al.* → lack of power?

eQTL

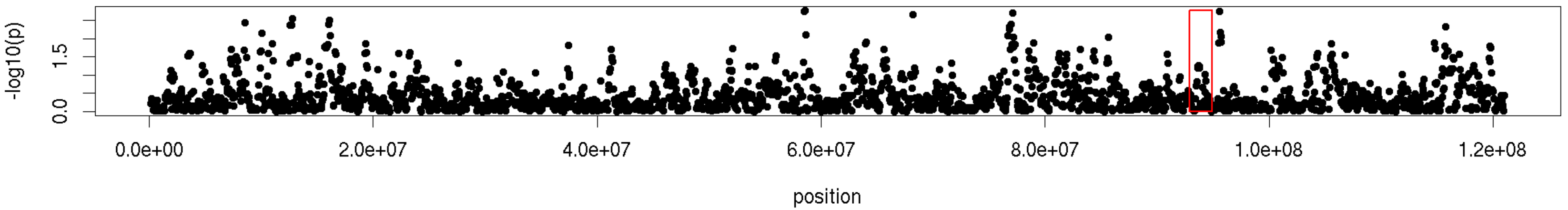


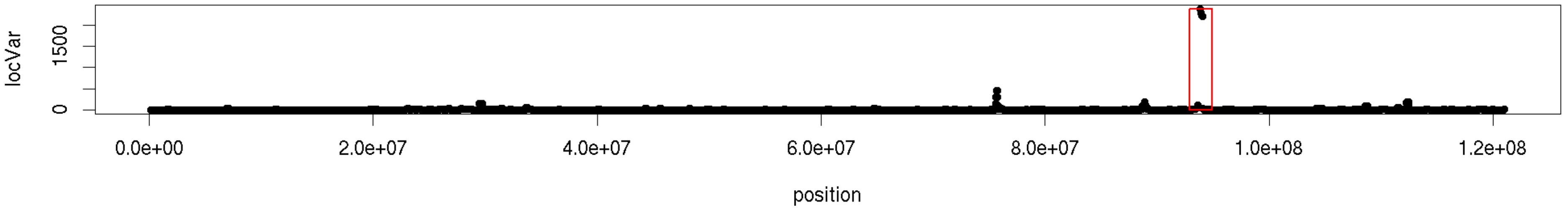
var(locGEBV)



Example: MGST1, chr 5, milk yield

In our study, p=0.008 for eQTL reported by Littlejohn *et al.* → lack of power?

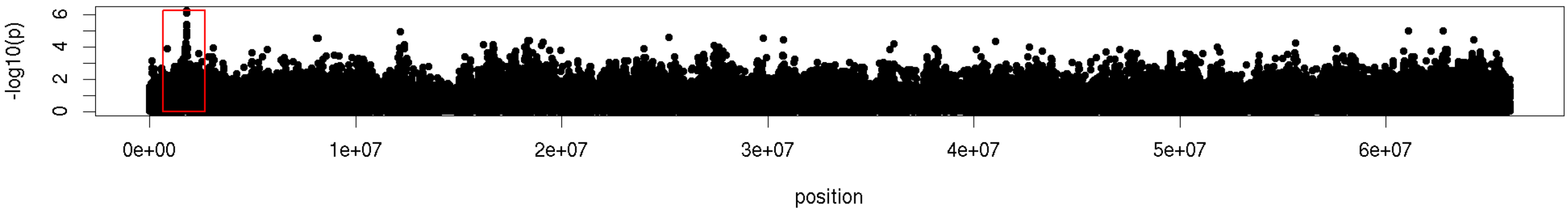

cor(locGebv,expr)


var(locGEBV)

Example: FUK, chr 18, fat yield

Small peaks in GWAS & local GEBV variance

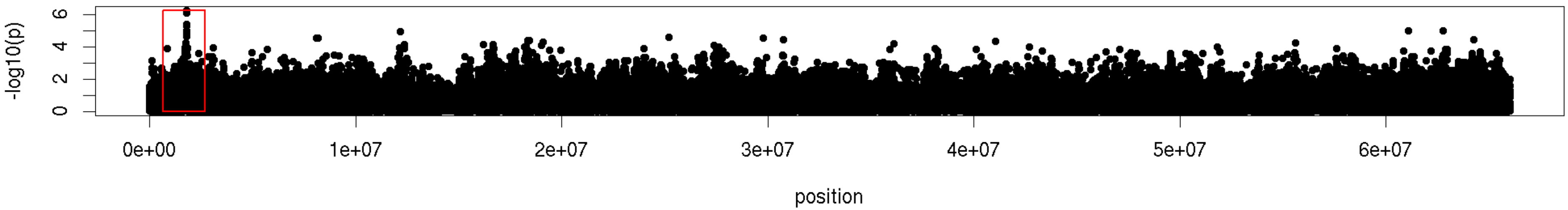
GWAS



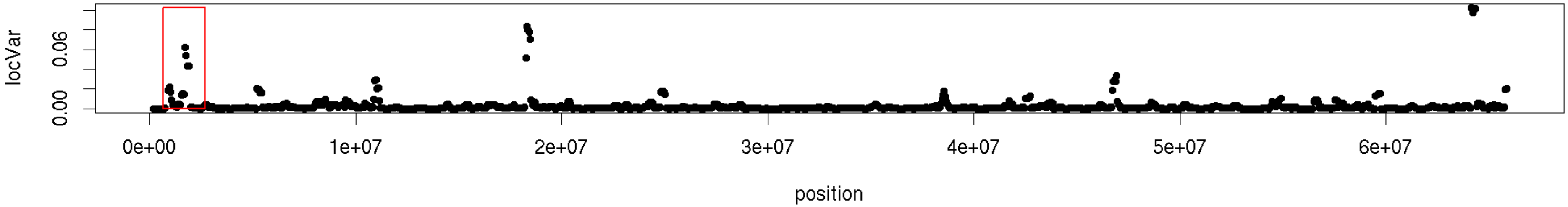
Example: FUK, chr 18, fat yield

Small peaks in GWAS & local GEBV variance

GWAS



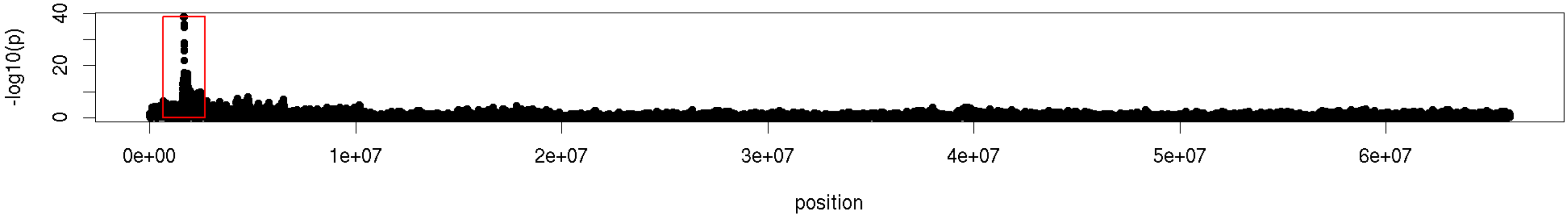
var(locGEBV)



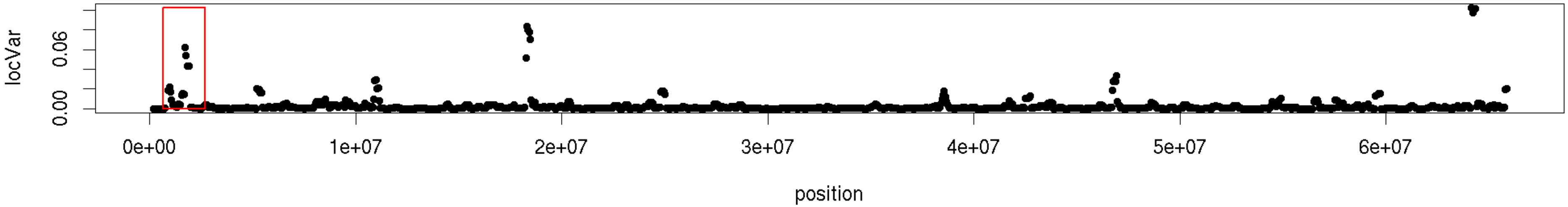
Example: FUK, chr 18, fat yield

Highly significant eQTL

eQTL



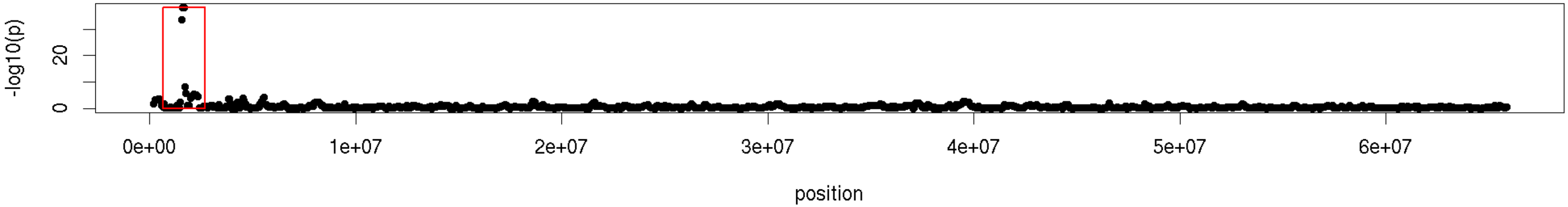
var(locGEBV)



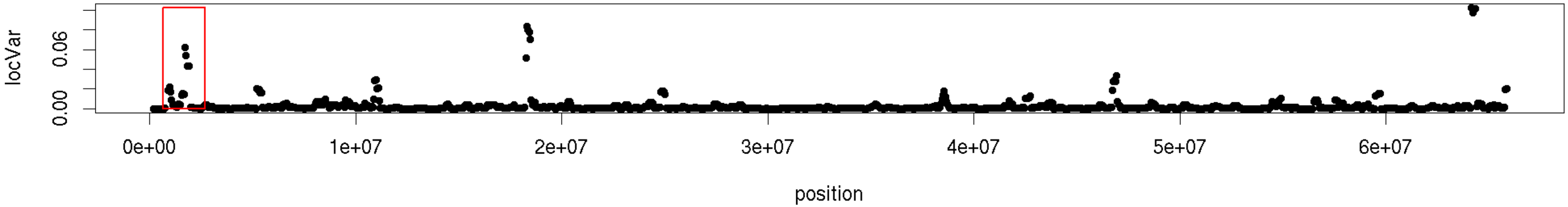
Example: FUK, chr 18, fat yield

Correlation between FUK expression & local GEBVs = 0.90 (p = 5×10^{-39})

cor(locGEBV,expr)

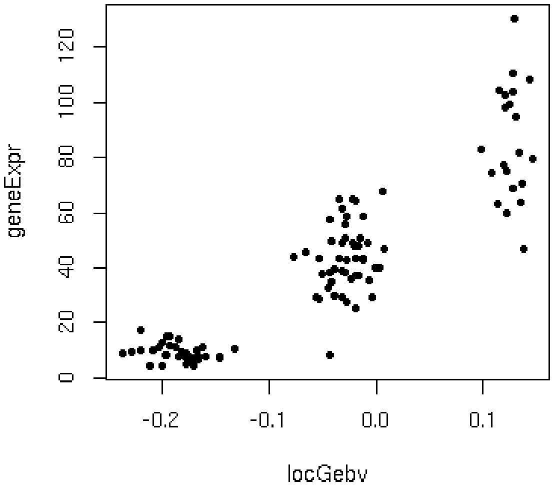


var(locGEBV)



Example: FUK, chr 18, fat yield

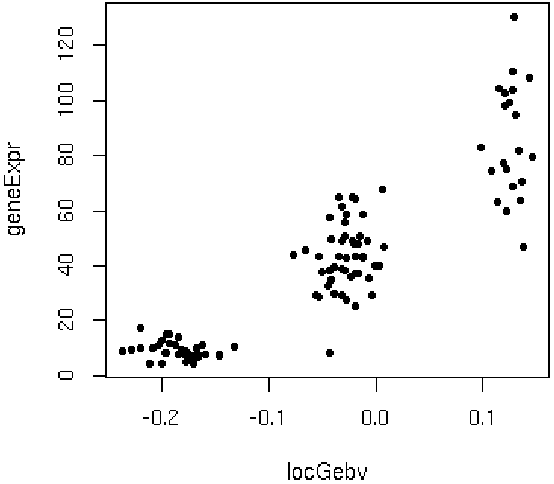
locGEBV vs expression



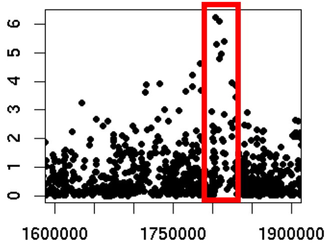
Example: FUK, chr 18, fat yield

Strong correlation, but GWAS, local GEBV variance & eQTL peaks on different locations

locGEBV vs expression



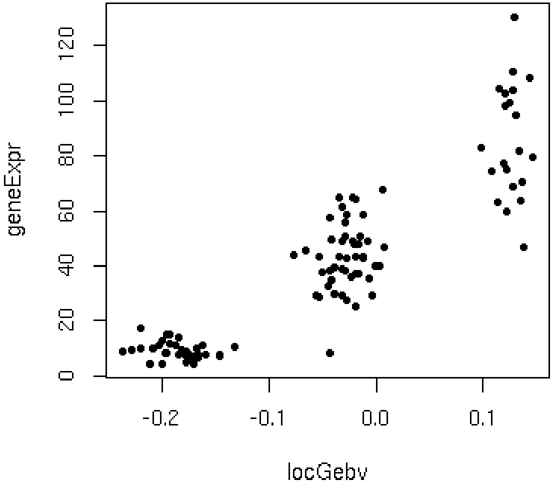
GWAS



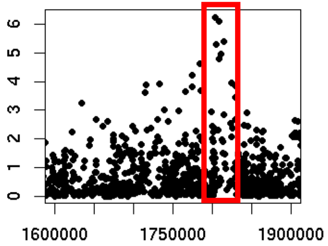
Example: FUK, chr 18, fat yield

Strong correlation, but GWAS, local GEBV variance & eQTL peaks on different locations

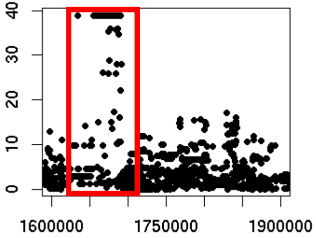
locGEBV vs expression



GWAS



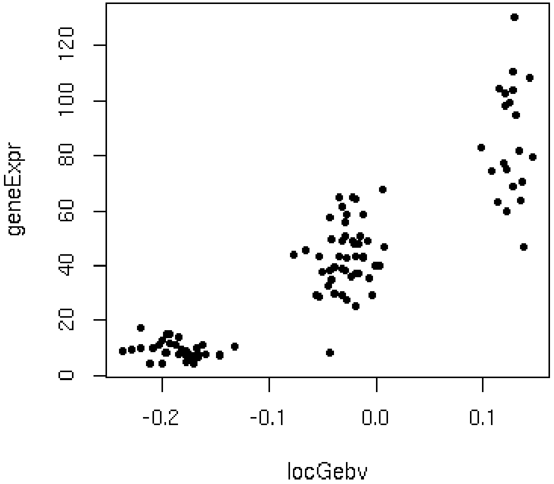
eQTL



Example: FUK, chr 18, fat yield

Strong correlation, but GWAS, local GEBV variance & eQTL peaks on different locations

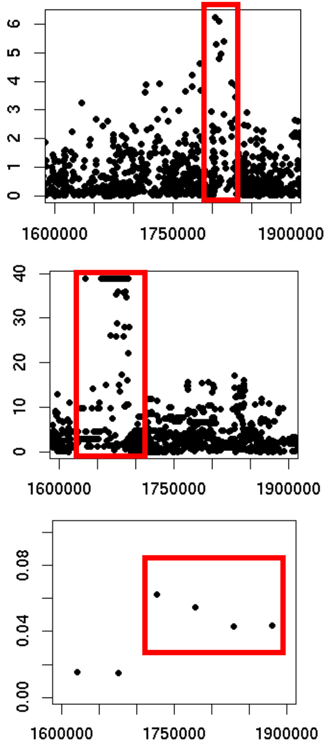
locGEBV vs expression



GWAS

eQTL

var(locGEBV)



Results colocalisation

- Out of 919 eQTL ($p \leq 10^{-5}$, max 1 per gene), 47 colocalise with a GWAS QTL

Results colocalisation

- Out of 919 eQTL ($p \leq 10^{-5}$, max 1 per gene), 47 colocalise with a GWAS QTL

- Out of 178 GWAS QTL ($p \leq 10^{-6}$, max 1 QTL per Mb), 27 colocalise with an eQTL

Results colocalisation

- Out of 919 eQTL ($p \leq 10^{-5}$, max 1 per gene), 47 colocalise with a GWAS QTL
- Out of 178 GWAS QTL ($p \leq 10^{-6}$, max 1 QTL per Mb), 27 colocalise with an eQTL
- QTL that colocalised with eQTL had small effects on the trait

Results colocalisation

- Out of 919 eQTL ($p \leq 10^{-5}$, max 1 per gene), 47 colocalise with a GWAS QTL
- Out of 178 GWAS QTL ($p \leq 10^{-6}$, max 1 QTL per Mb), 27 colocalise with an eQTL
- QTL that colocalised with eQTL had small effects on the trait
- High LD in dairy cattle → GWAS effects influenced by multiple QTL

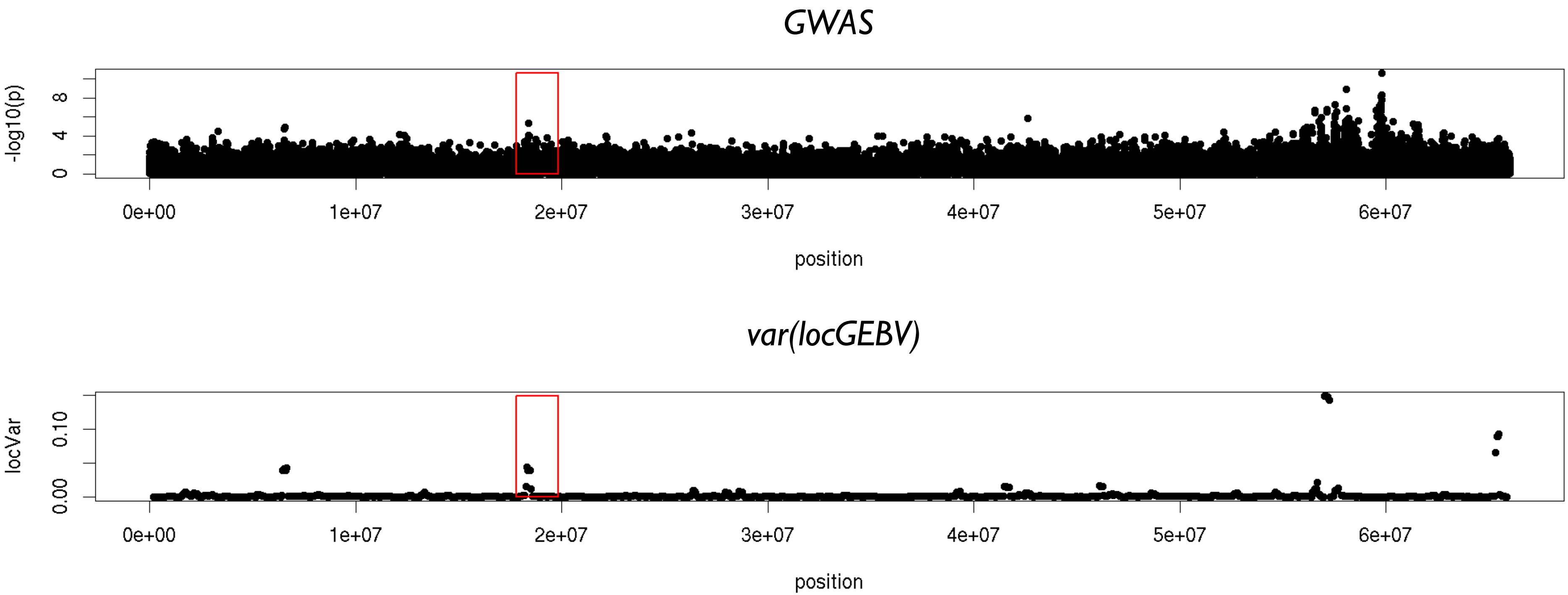
Results colocalisation

- Out of 919 eQTL ($p \leq 10^{-5}$, max 1 per gene), 47 colocalise with a GWAS QTL
- Out of 178 GWAS QTL ($p \leq 10^{-6}$, max 1 QTL per Mb), 27 colocalise with an eQTL
- QTL that colocalised with eQTL had small effects on the trait
- High LD in dairy cattle → GWAS effects influenced by multiple QTL

→ The majority of QTL and eQTL do not seem to colocalise, and the ones that do explain only a minor part of the genetic variance

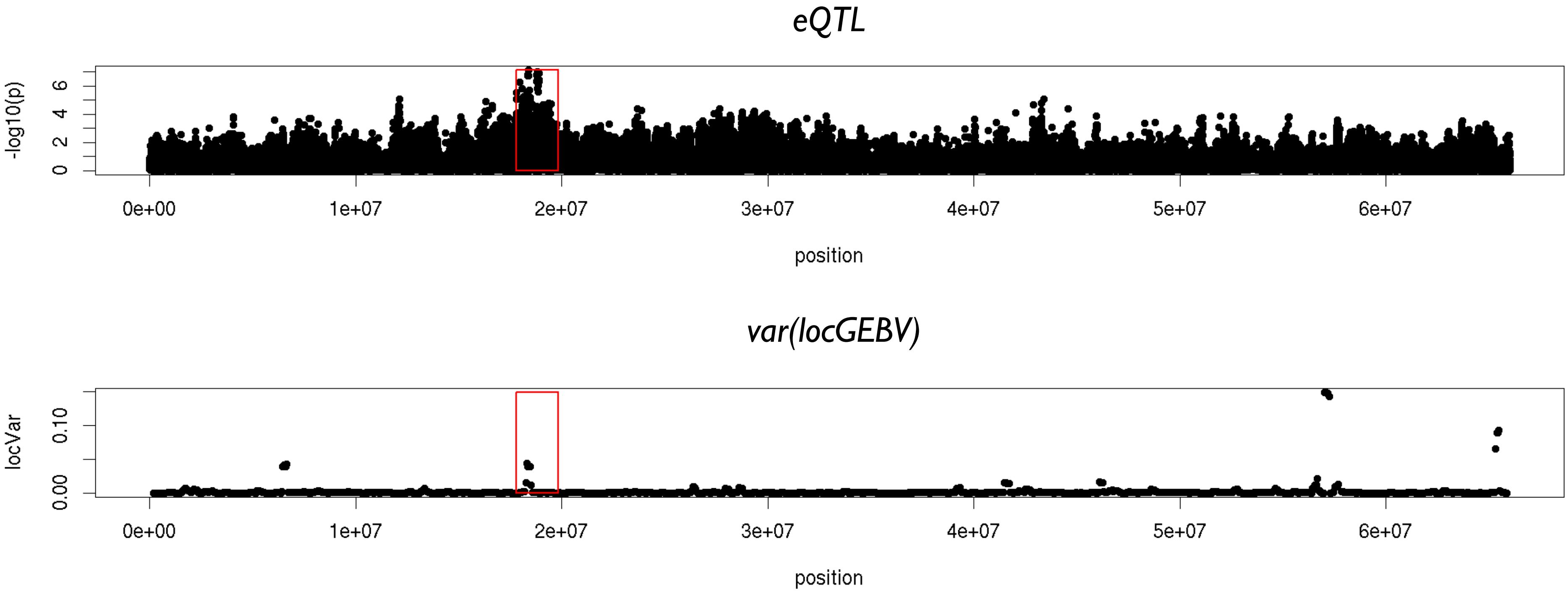
Example: BRD7, chr 18, protein yield

Small peaks in GWAS & local GEBV variance



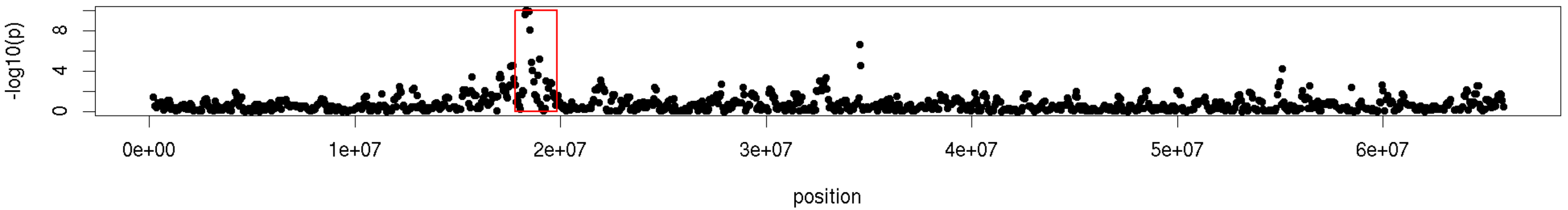
Example: BRD7, chr 18, protein yield

Significant eQTL

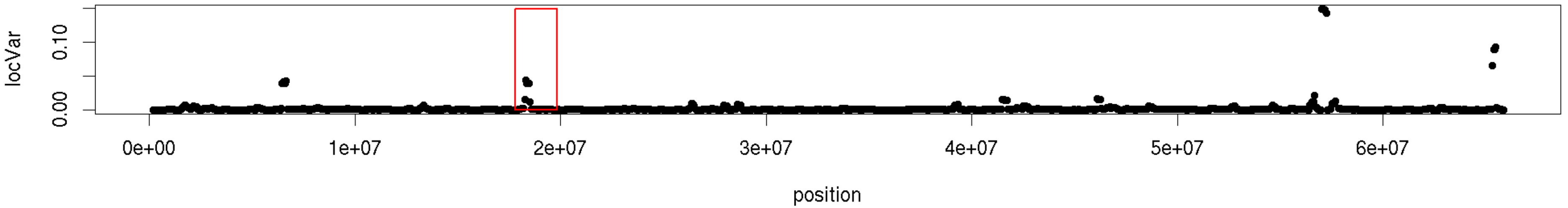


Example: BRD7, chr 18, protein yield

Correlation between BRD7 expression & local GEBVs = 0.58 (p = 9.3×10^{-11})



cor(locGEBV,expr)

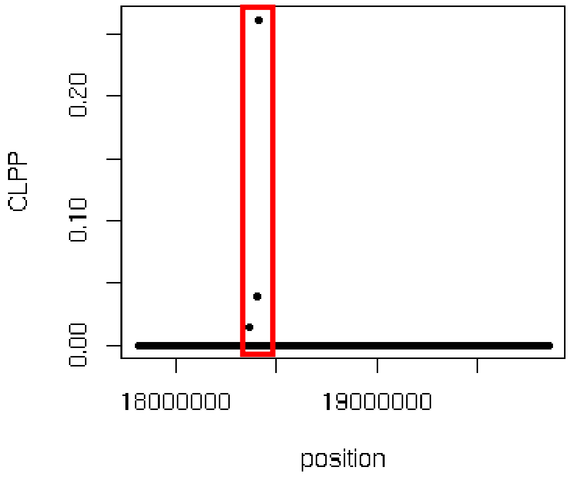


var(locGEBV)

Example: BRD7, chr 18, protein yield

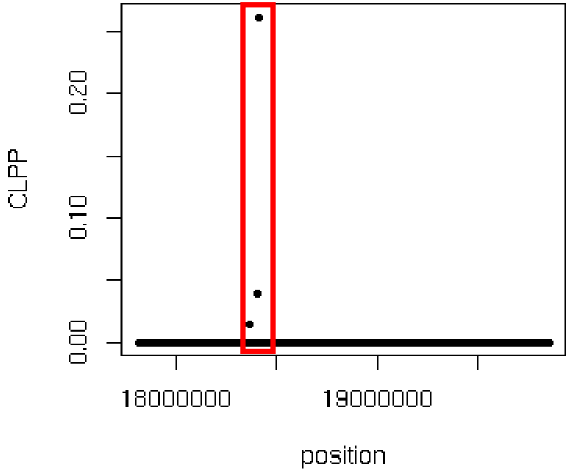
Strong correlation, CLPP = 0.26 for an intergenic variant near *BRD7*

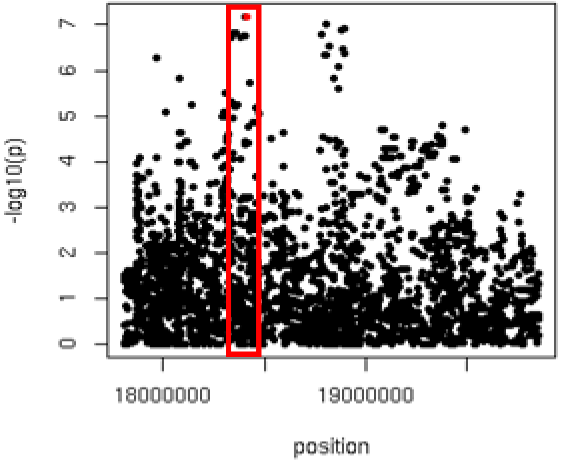
colocalisation



Example: BRD7, chr 18, protein yield

Strong correlation, CLPP = 0.26 for an intergenic variant near *BRD7*, top variant in eQTL

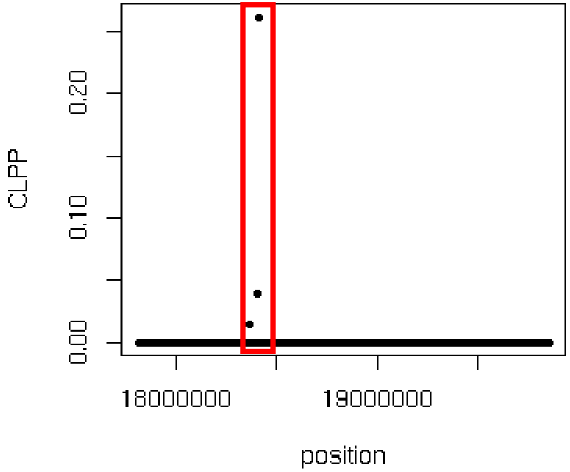

colocalisation


eQTL

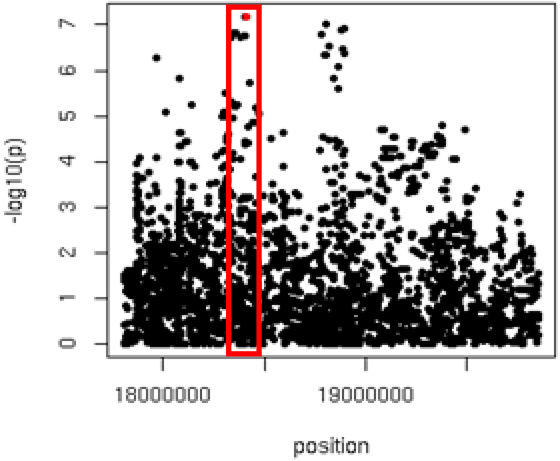
Example: BRD7, chr 18, protein yield

Strong correlation, CLPP = 0.26 for an intergenic variant near *BRD7*,
top variant in eQTL but not in GWAS

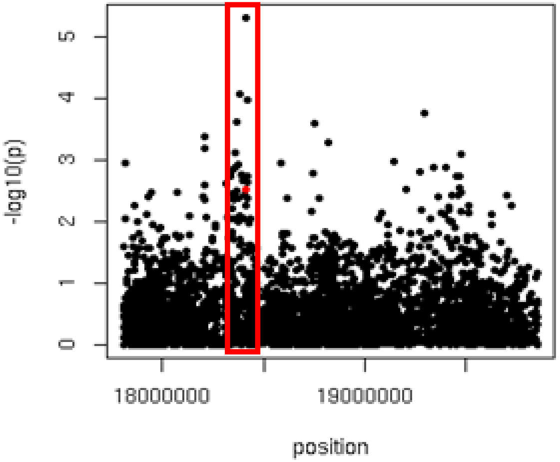
colocalisation



eQTL



GWAS



Conclusions

- We found little overlap between QTL and eQTL

- Lack of power, need more powerful eQTL experiment?

- The QTL that we detected using expression data explained only a very small part of the total genetic variance

→ Expression data can help to identify small QTL, but in our study not enough of them to explain a substantial part of the total genetic variance

Acknowledgements

GenSAP

DataGene
Solutions for Herd Development

AGRICULTURE VICTORIA

DairyBio