Sequence-based genome-wide association study of milk mid-infrared wavenumbers in dairy cattle

Kathryn Tiplady

T. Lopdell[1], E. Reynolds[2], R. Sherlock[1], M. Keehan[1], T. Johnson[1], J. Pryce[3,4], H. Blair[2], S. Davis[1], M. Littlejohn[1,2], D. Garrick[2], R. Spelman[1] and B. Harris[1]

[1]Livestock Improvement Corporation, Hamilton, NZ
[2]Massey University, AL Rae Centre, Hamilton, NZ
[3]Agriculture Victoria, AgriBio, VIC, Australia
[4]La Trobe University, VIC, Australia

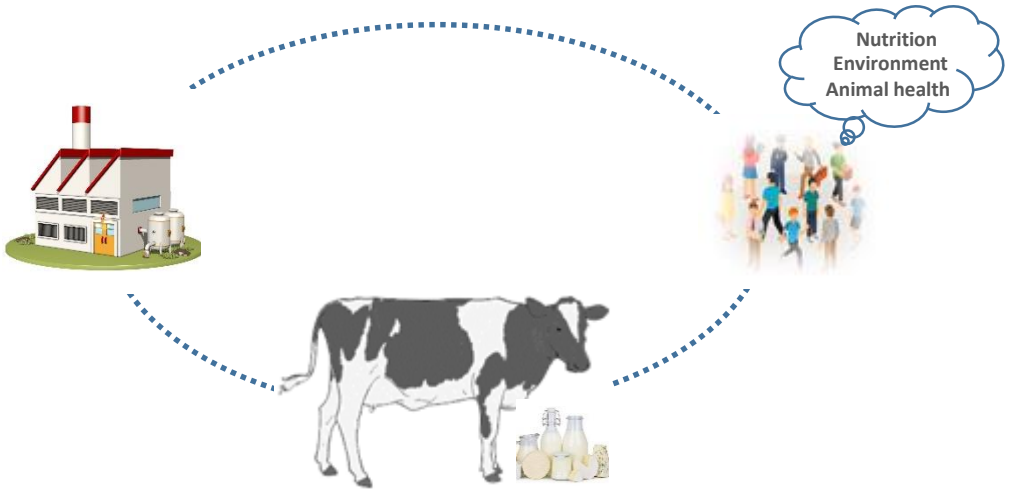
MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

LIC

EAAP · EUROPA 2019 – Ghent – Belgium · 70th · EVT · FEZ · EAAP · EAX

Introduction

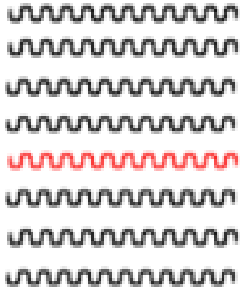
- Milk is highly complex.

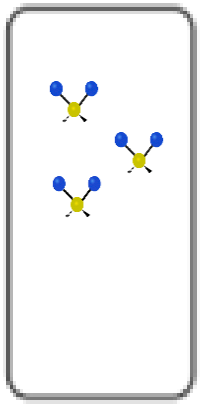


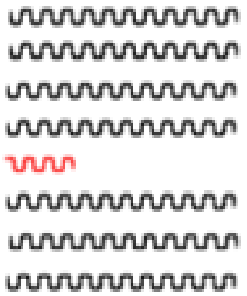
- Fourier-transform infrared (FTIR) spectroscopy is widely used to estimate milk composition and other indicator traits.
- High throughput, inexpensive, readily available from routine milk testing.

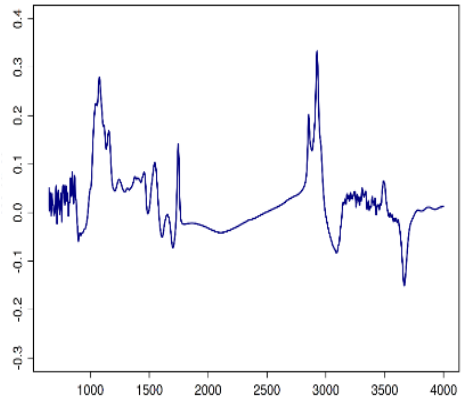
FTIR spectroscopy

Infrared light

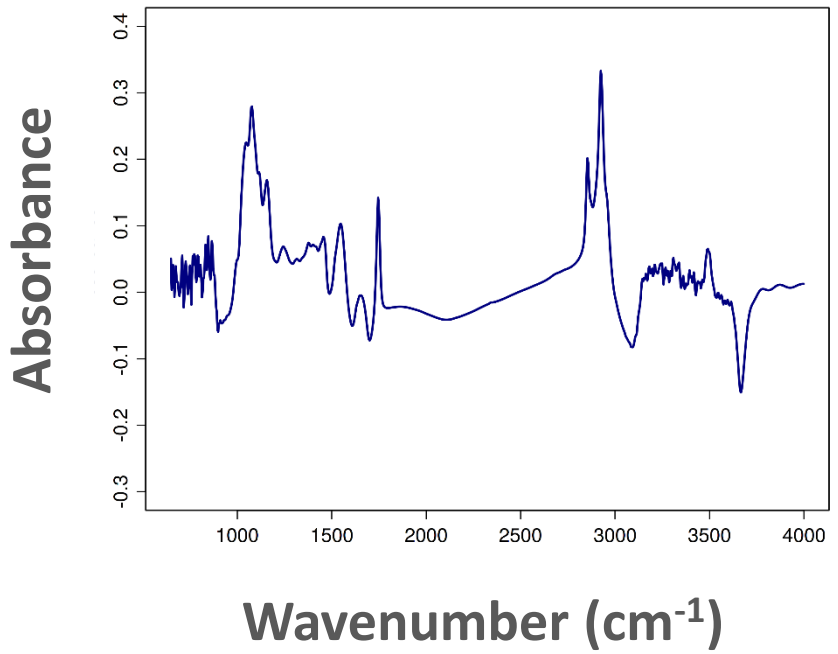








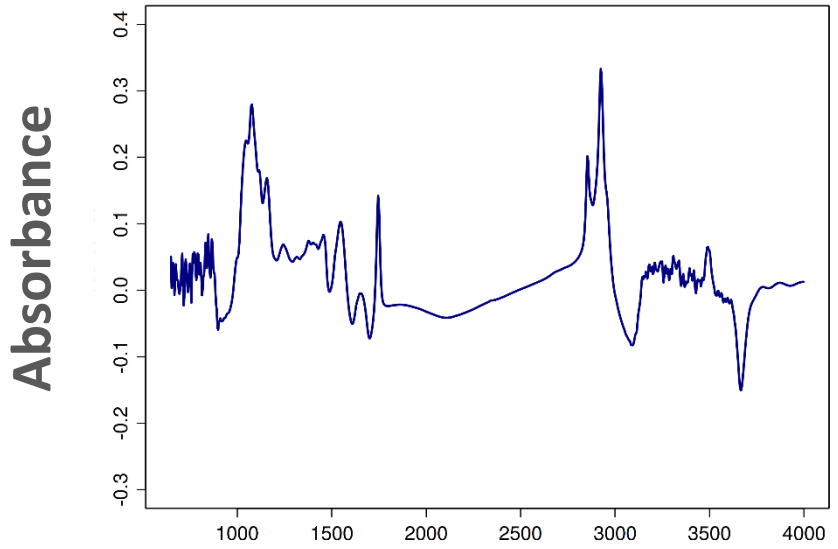
FTIR spectroscopy



$$Fat = F_0 + F_1 w_1 + F_2 w_2 + \cdots + F_k w_k$$

$$Protein = P_0 + P_1 w_1 + P_2 w_2 + \cdots + P_k w_k$$

FTIR spectroscopy



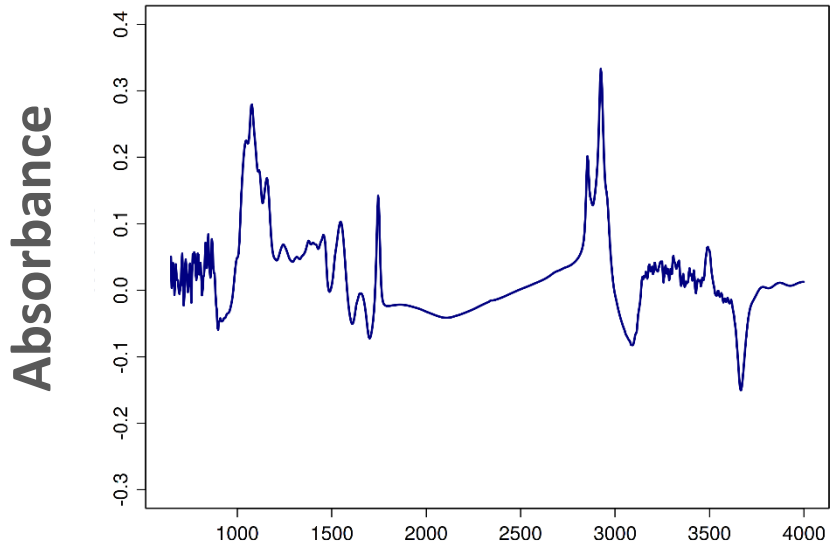
$$Fat = F_0 + F_1 w_1 + F_2 w_2 + \cdots + F_k w_k$$

$$Protein = P_0 + P_1 w_1 + P_2 w_2 + \cdots + P_k w_k$$

Energy balance
Fertility
Fatty acid profiles

Nitrogen
Methane

FTIR spectroscopy



$$Fat = F_0 + F_1 w_1 + F_2 w_2 + \cdots + F_k w_k$$

$$Protein = P_0 + P_1 w_1 + P_2 w_2 + \cdots + P_k w_k$$

Energy balance
Fertility
Fatty acid profiles

Nitrogen
Methane

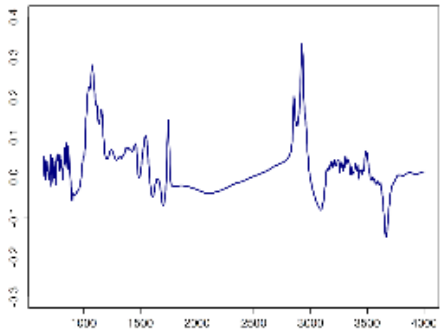
Genetics of FTIR data

- Individual wavenumbers have moderate to high heritability across the mid-infrared region (Bittante & Cecchinato 2013; Soyeurt et al. 2010).
- Some promising association peaks have been identified in other studies (Wang & Bovenhuis 2018; Bendet et al. 2019).

Aim of our study

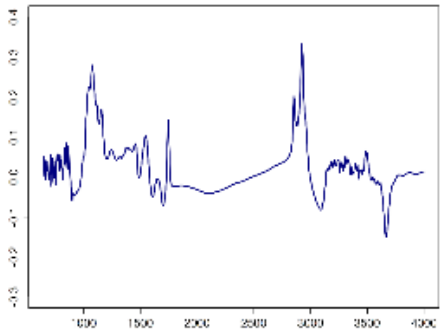
- Identify associations between individual mid-infrared wavenumbers and regions of the genome.

Identify genomic associations



GWAS

Identify genomic associations



GWAS

Dataset

- 100,671 FTIR records: 38,085 mixed-breed NZ dairy cattle.
- Imputed sequence: 17,776,293 variants.
- Individual wavenumbers standardized (Grelet et al., 2015; Tiplady et al., 2019).
- Adjusted for parity, days in milk, breed, heterosis and herd by test date (ASReml mixed linear model).

Identify genomic associations

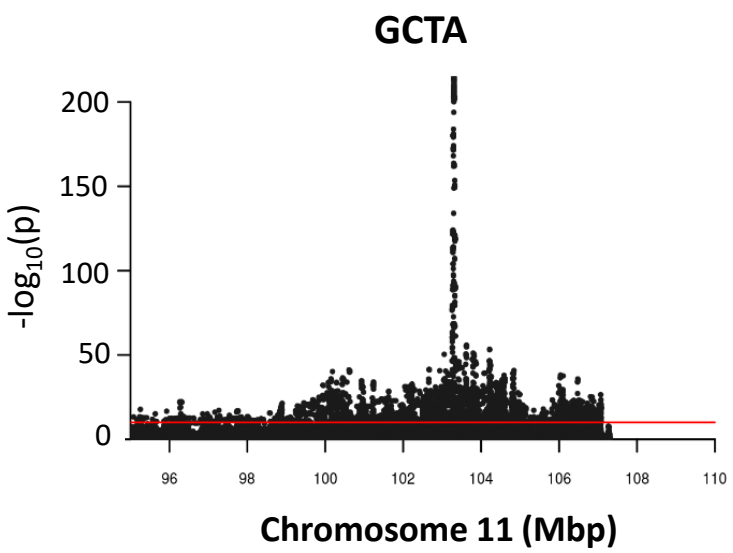
Sequence-based GWAS: ~38k animals, ~17.7m variants, 895 traits

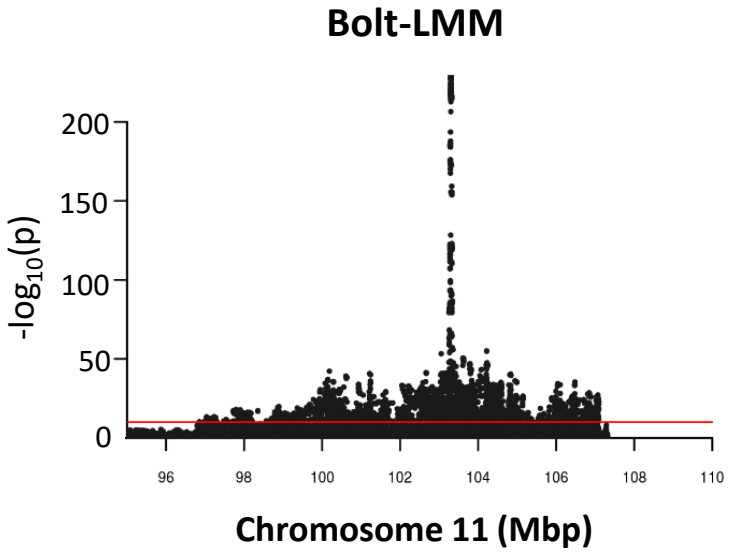
Computation challenges

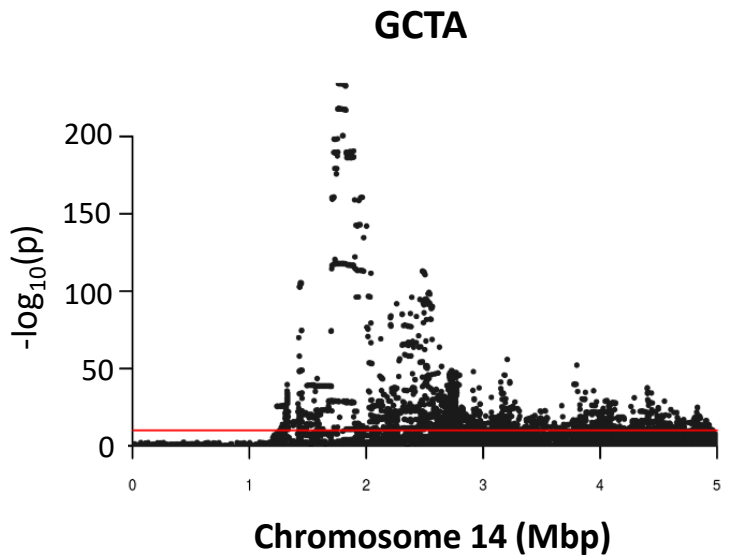
- GCTA (Yang et al., 2011): 1800 hours per trait ⇨ ~ 6 years
- Bolt-LMM (Loh et al., 2015): 6 hours per trait ⇨ ~ 7 days

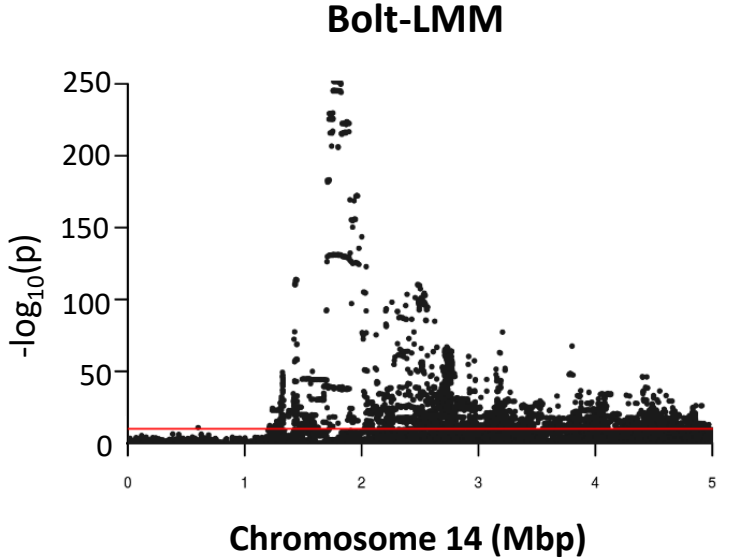
- Comparisons made using 43,851 SNP to account for population stratification.

Software comparisons





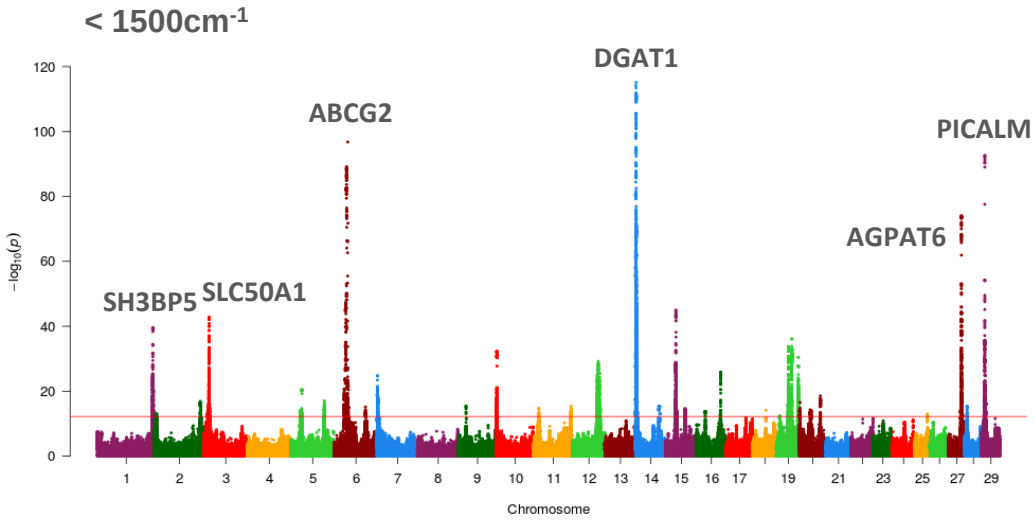


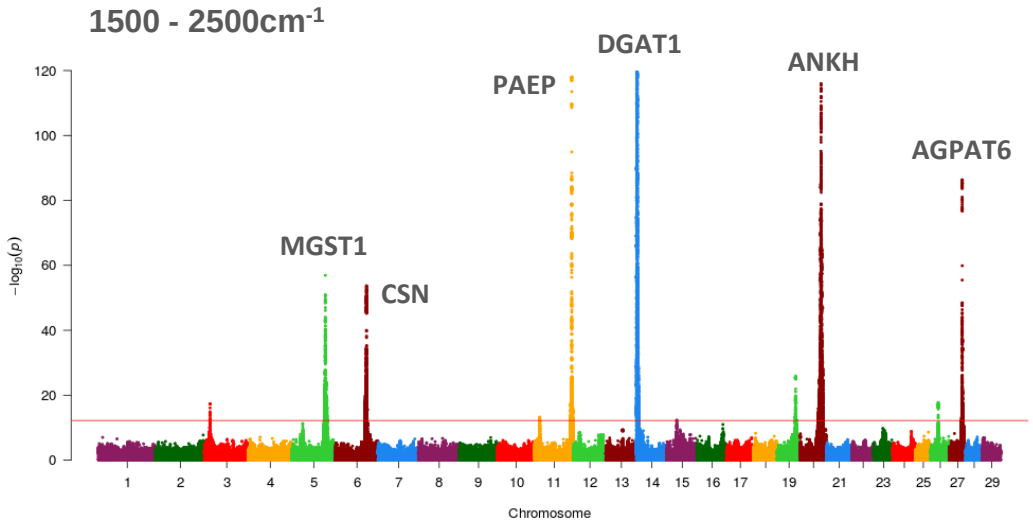


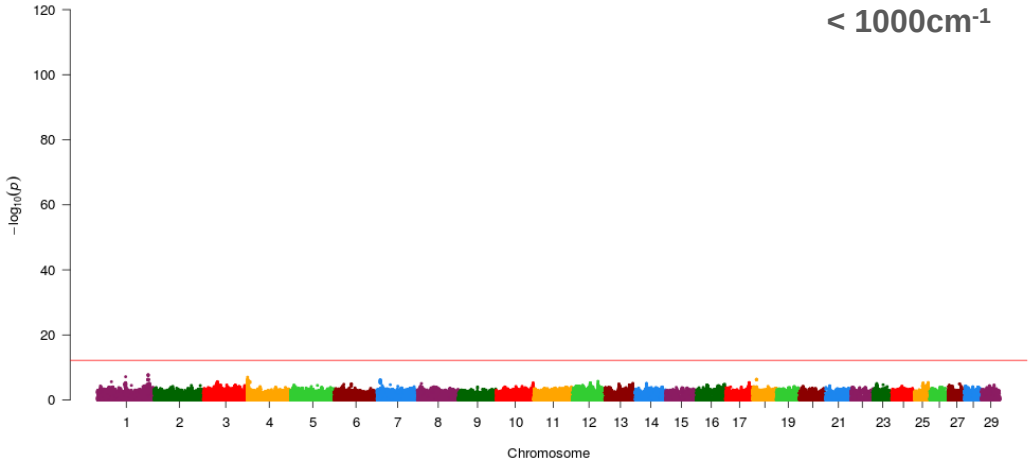
~1800 hours/trait

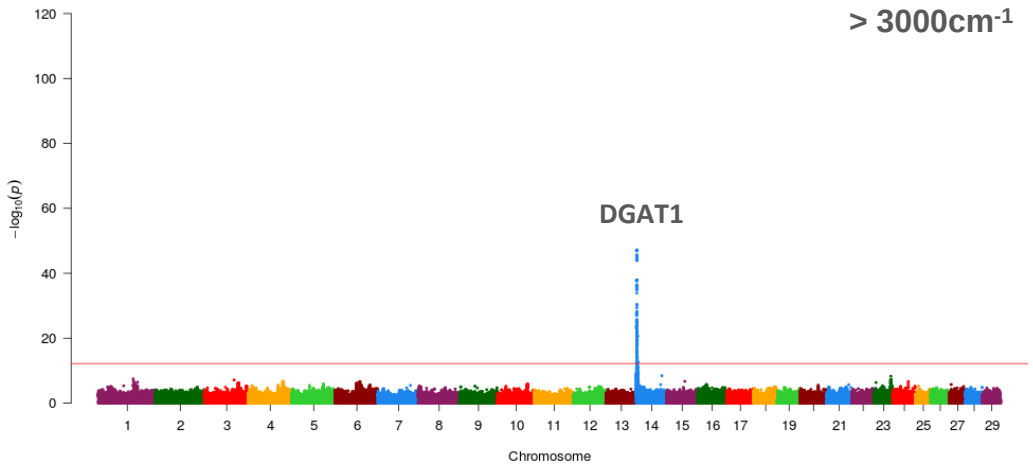
~6 hours/trait

Results







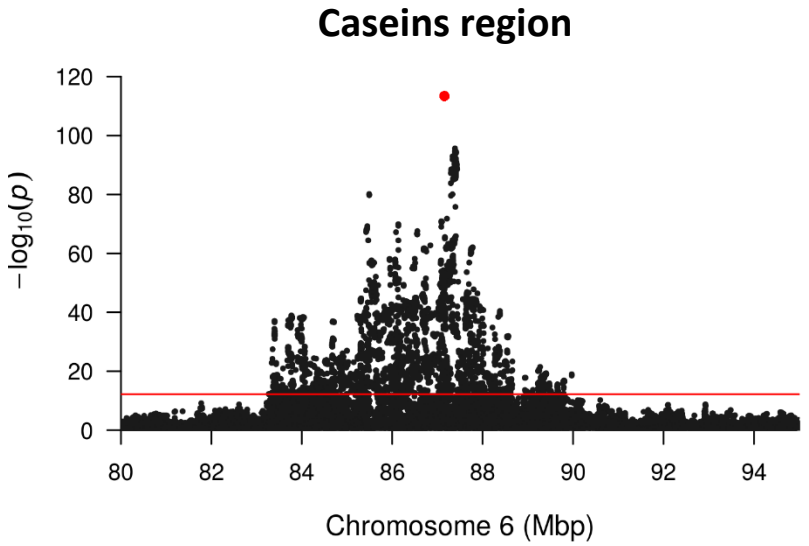


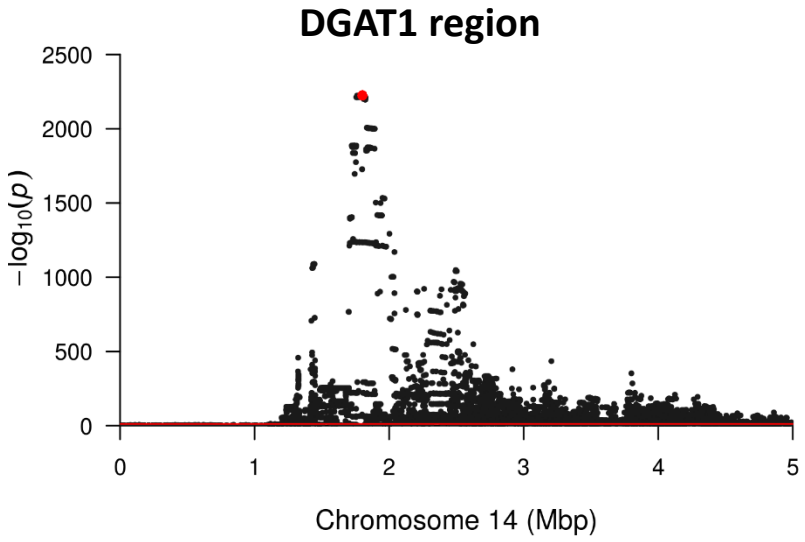
Bonferroni-threshold: $0.01/17.7m/895=6.3 \times 10^{-13}$.

Significant associations conserved across multiple wavenumbers: DGAT1 significant for over 750 wavenumbers!

Peak identification

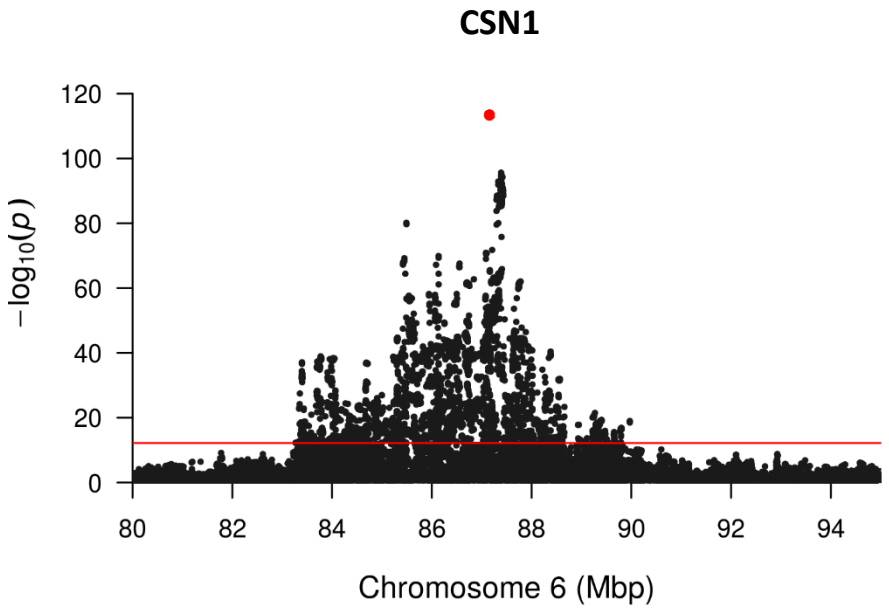
- How to deal with complex regions?
  e.g: Caseins region on BTA6; DGAT1 region at the start of BTA14.

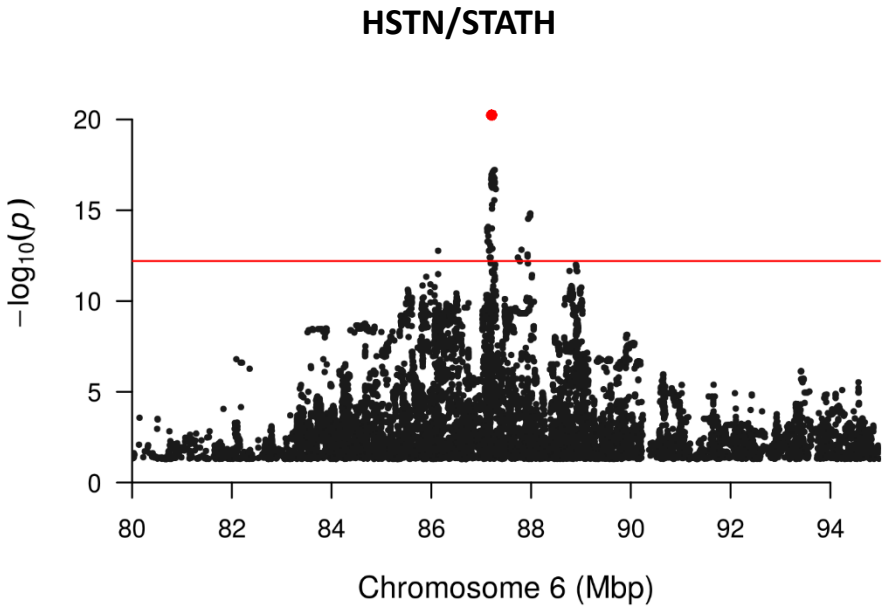


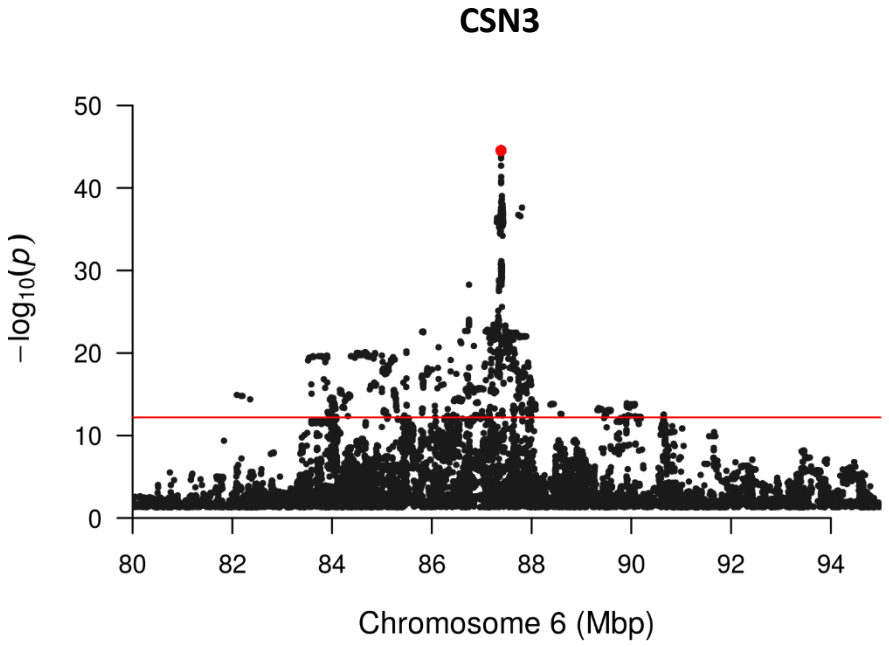


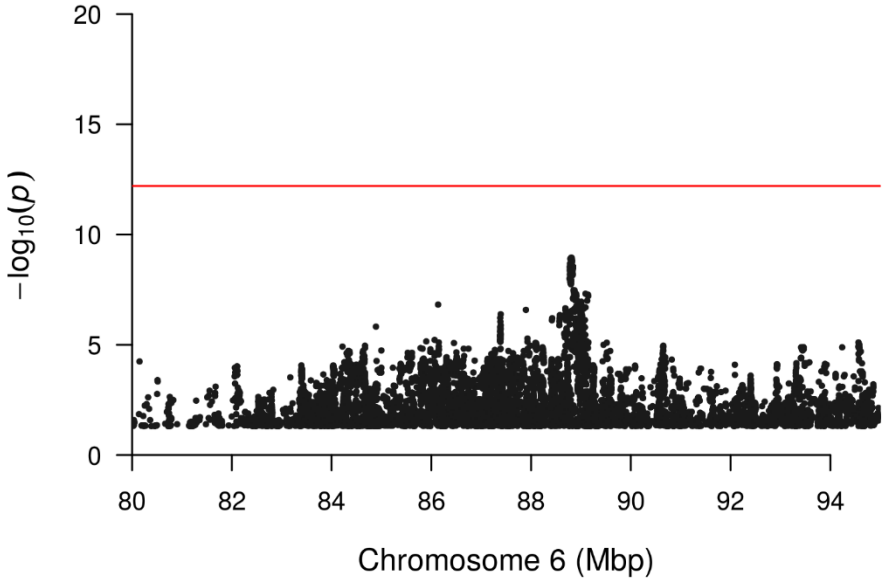
- Conditional iterative GWAS for each wavenumber.
  - Top SNP on each chromosome fitted as a covariate.
  - Iteratively run until there no signal left.

Peak identification

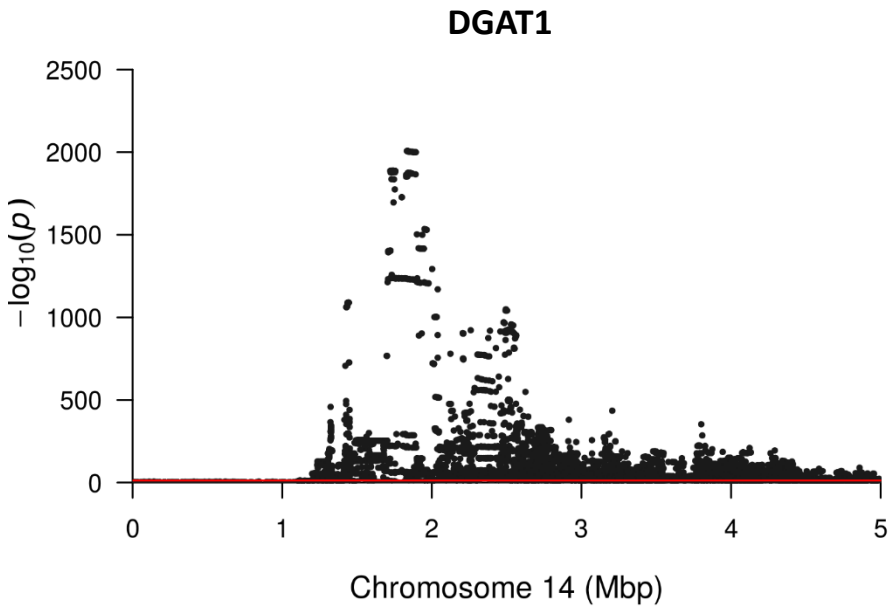


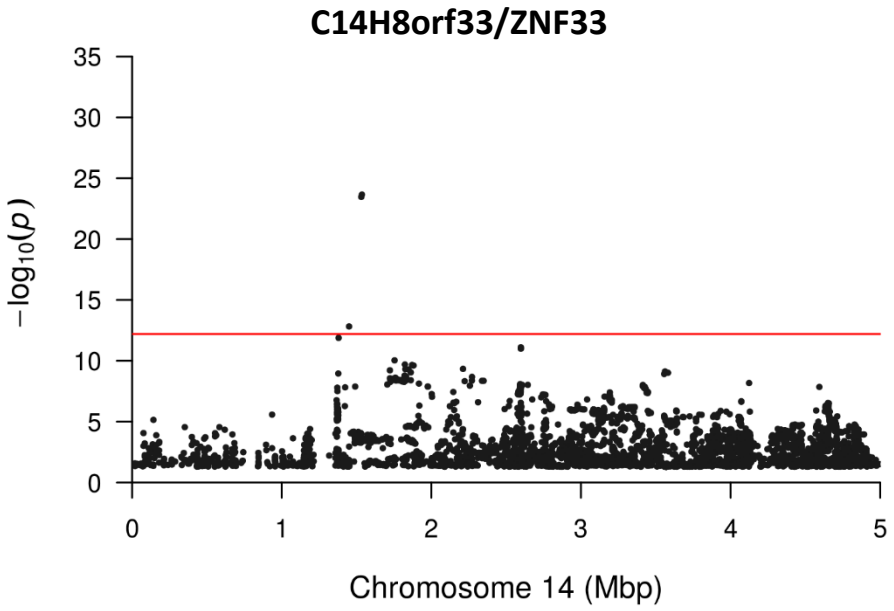


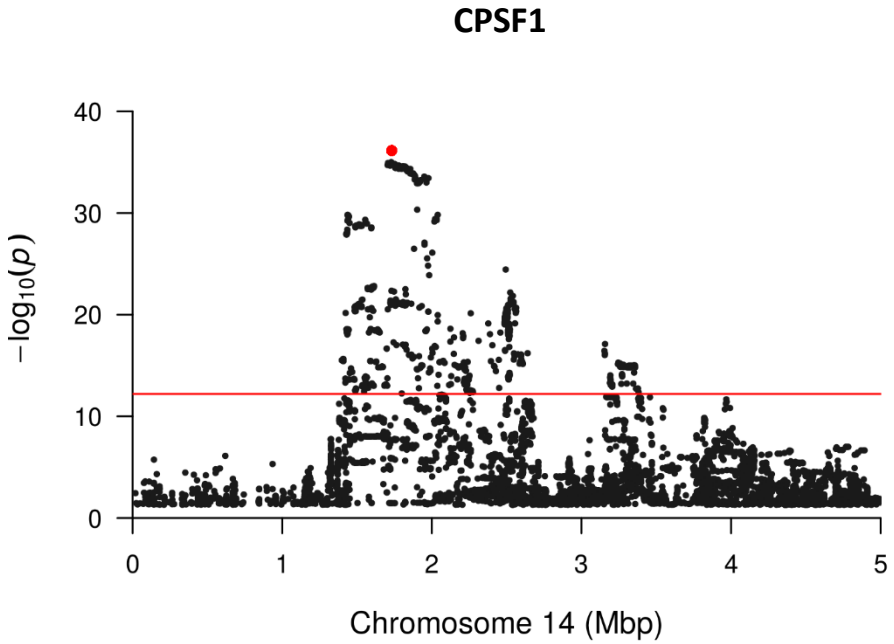


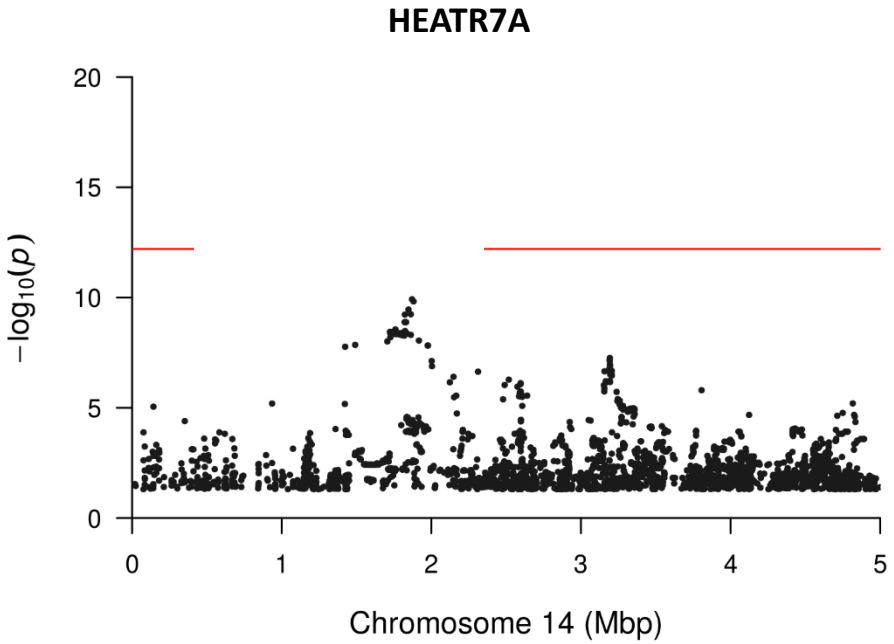


Peak identification



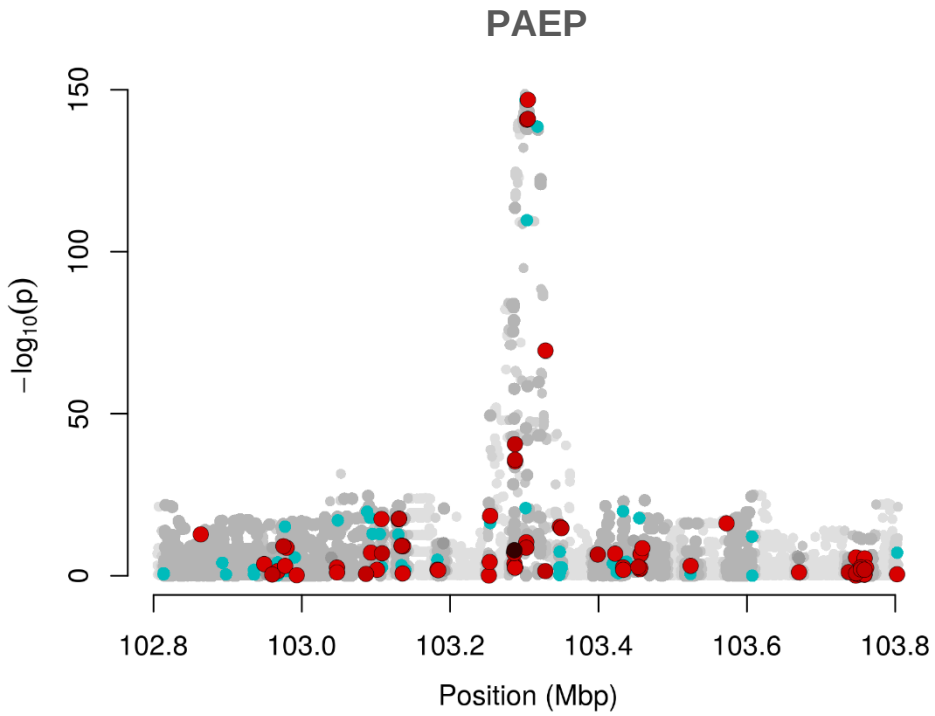


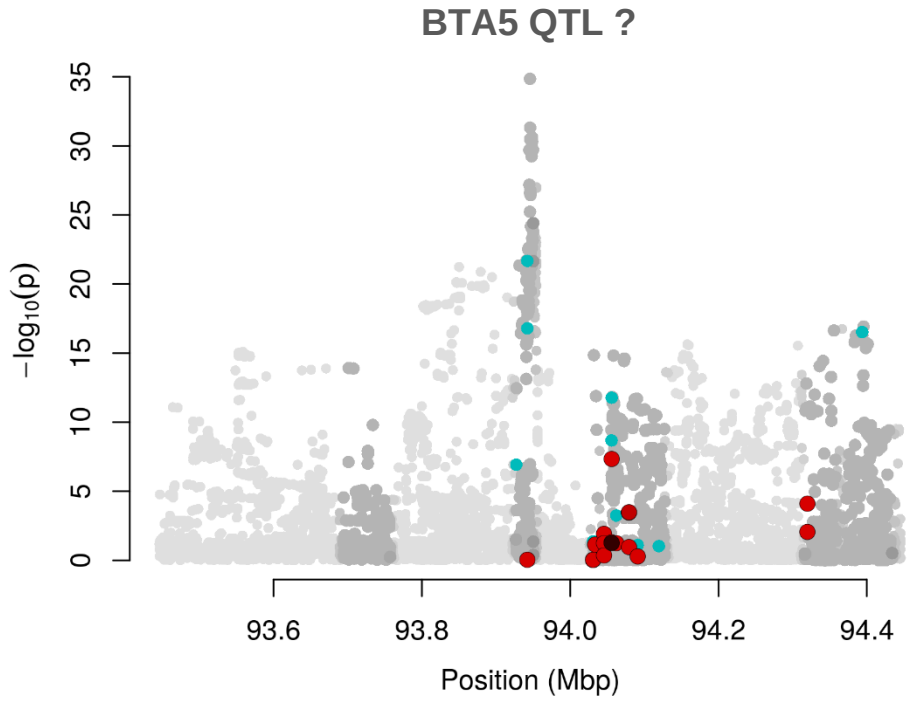




Determine likely candidates

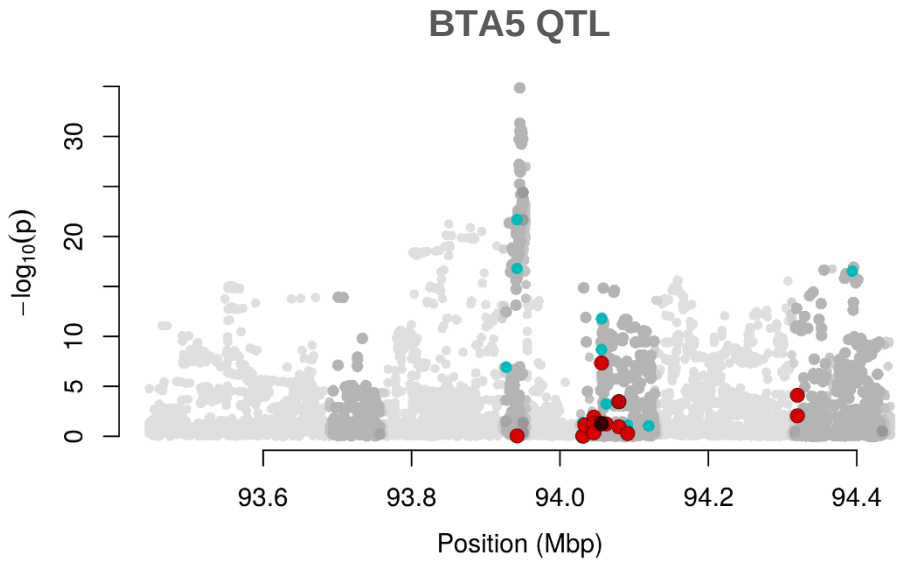
Link associations to SnpEff annotation - assess the candidacy of variants in terms of their predicted impact on protein sequence.





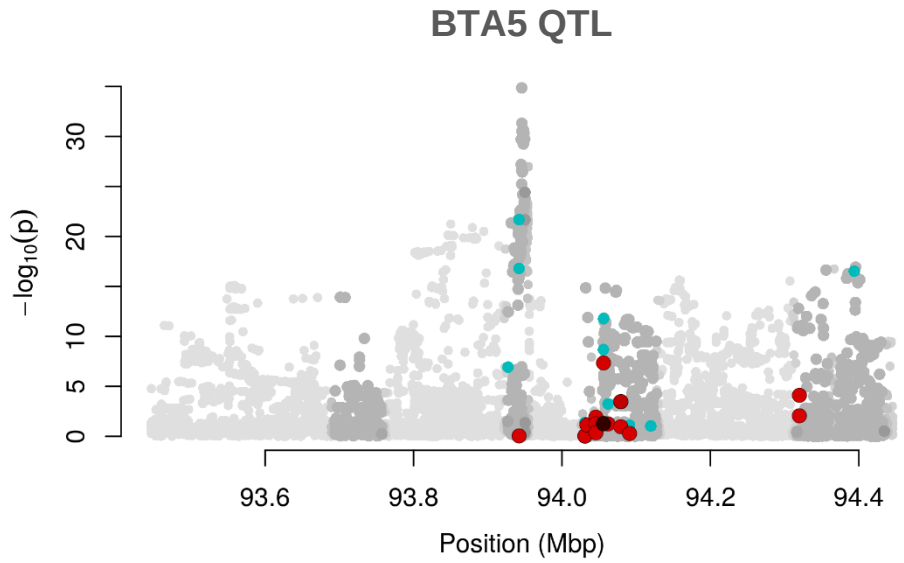
- High impact: Slice donor/acceptor, frameshift variants
- Moderate: Missense, splice region variants
- Low impact: Synonymous variants, non-coding exons
- Modifiers: Intron and up/downstream variants, UTRs.

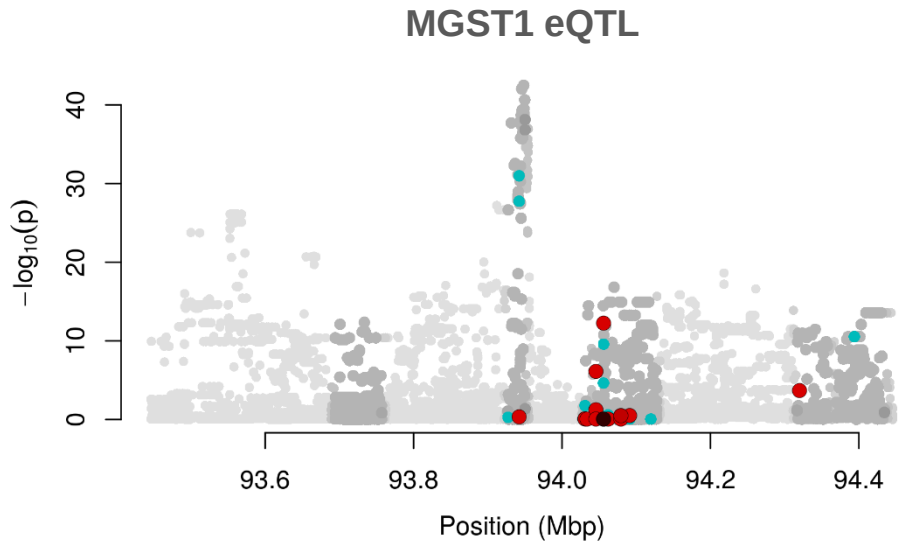
Co-locating eQTLs



- Gene expression data from mammary tissue in 375 animals (Lopdell et al., 2017).
- A co-locating eQTL indicates genetic regulation of the gene in the same region that we see the trait QTL.

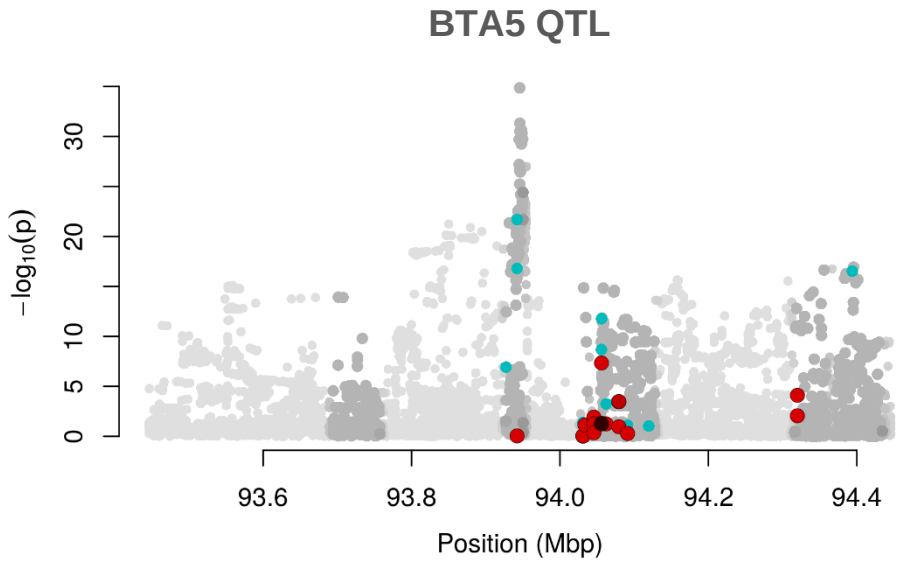
Co-locating eQTLs

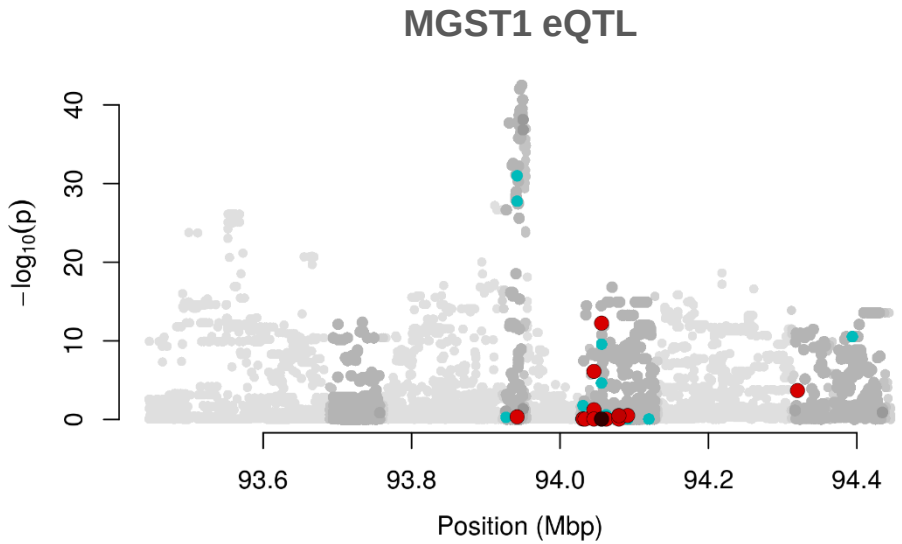


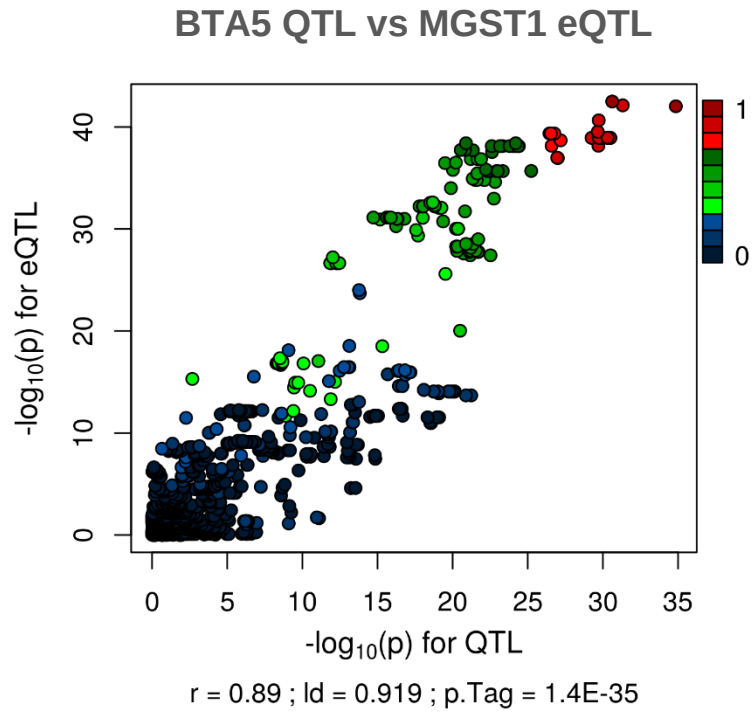


- Gene expression data from mammary tissue in 375 animals (Lopdell et al., 2017).
- A co-locating eQTL indicates genetic regulation of the gene in the same region that we see the trait QTL.

Co-locating eQTLs

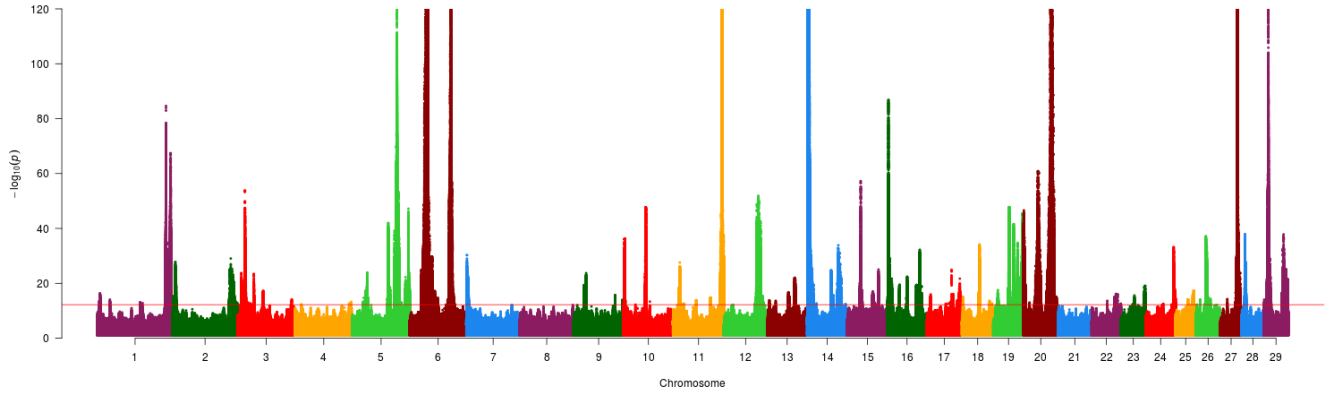


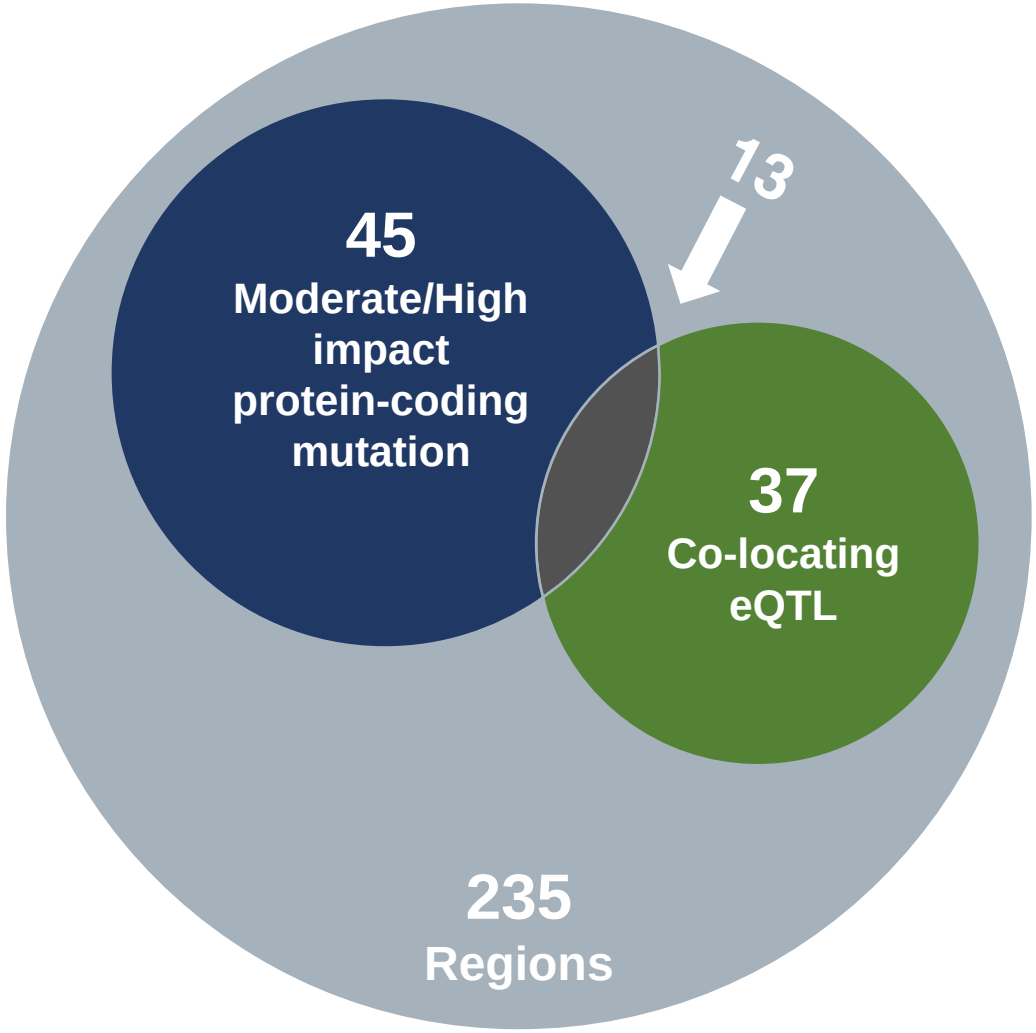




- Gene expression data from mammary tissue in 375 animals (Lopdell et al., 2017).
- A co-locating eQTL indicates genetic regulation of the gene in the same region that we see the trait QTL.

Peak overview





Summary

- GWAS on mid-infrared wavenumbers conducted
  - Bolt-LMM: ~38k animals, ~17.7m variants, 895 traits
  - Multiple QTL in complex regions ⇨ Conditional iterations.

- Promising peaks with candidate variants and co-locating eQTLs identified.
- Ongoing research to consolidate results and determine how we can use this information to improve genomic selection.

Acknowledgements

Sequence-based genome-wide association study of milk mid-infrared wavenumbers in dairy cattle

T. Lopdell, E. Reynolds, R. Sherlock, M. Keehan, T. Johnson, J. Pryce, H. Blair, S. Davis, M. Littlejohn, D. Garrick, R. Spelman and B. Harris

Tracey Monehan (Project manager)

Dorian Garrick (Massey, AL Rae Centre)

Matt Littlejohn (Massey/LIC)

Jennie Pryce (Agriculture Victoria)

Hugh Blair (Massey)

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

LIC