Investigating genotype-phenotype relationships for tuber bruising in autotetraploid potatoes

Tools for Polyploids Training Workshop 2022

O. Angelin-Bonnet, S. Thomson, P. J. Biggs, M. Vignes, S. Baldwin

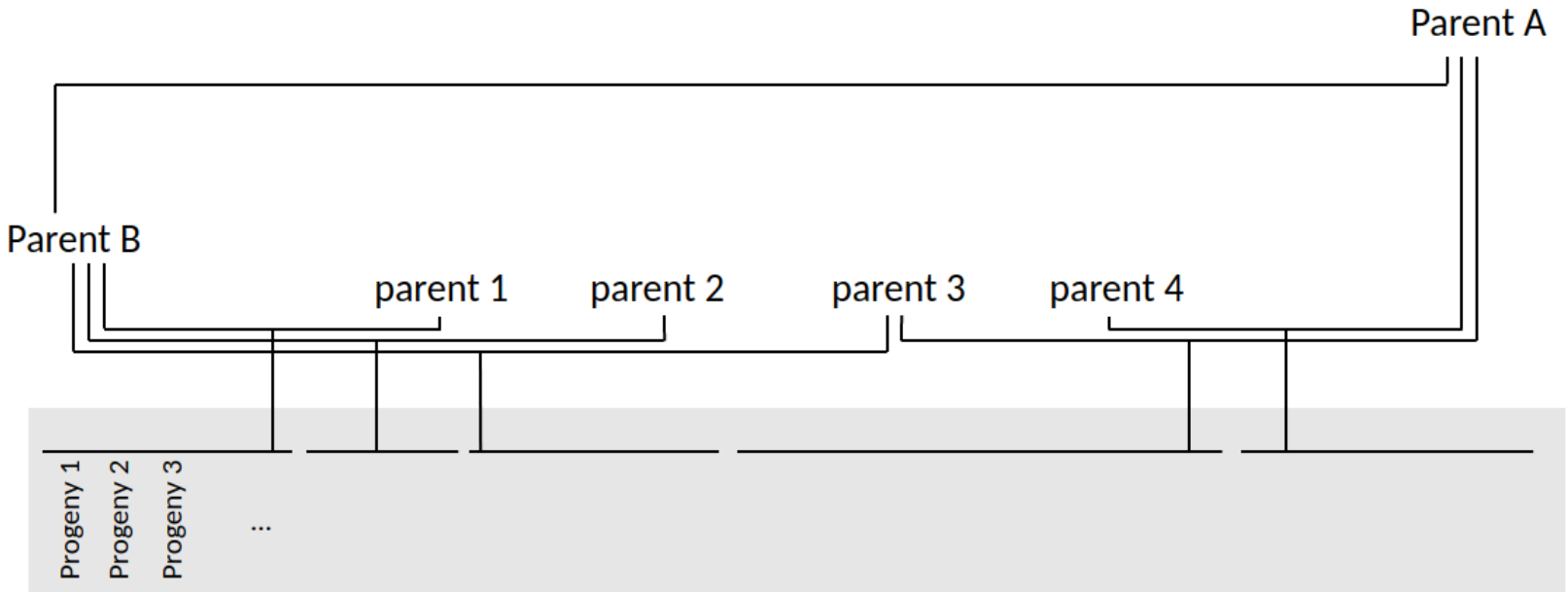
Massey University, New Zealand

14 January 2022

Introduction

Research question: investigate the molecular mechanisms of tetraploid potato bruising

Multi-factorial design: half-sibling progeny samples, from a breeding programme



Genomics data

Capture by sequencing using exon capture:

- baits designed to target exons, reflect exon density
- 178 samples (160 progeny genotypes and 13 parent genotypes)
- Genotype calling with `polyRAD`: 454,246 biallelic SNPs (after filtering)

Phenotypic data

Bruising experiment performed on a subset of the progeny data:

- Tubers bruised by throwing weight from controlled height
- Image of the bruised site taken after 24h
- Bruising scored visually on a scale from 0 (no bruising) to 5 (extensive bruising)

Other phenotypes recorded: dry matter, sugar content, maturity, vigour, etc.

Molecular data

Transcriptomics data

- RNA sequencing for 100 progeny samples
- 25,163 transcripts measured (after filtering)

Metabolomics data

- LC-MS for 122 progeny samples, each with 2 biological replicates
- 4,604 compounds measured
- Compounds identification main bottleneck for this dataset

Genome-Wide Association Study

With GWASpoly

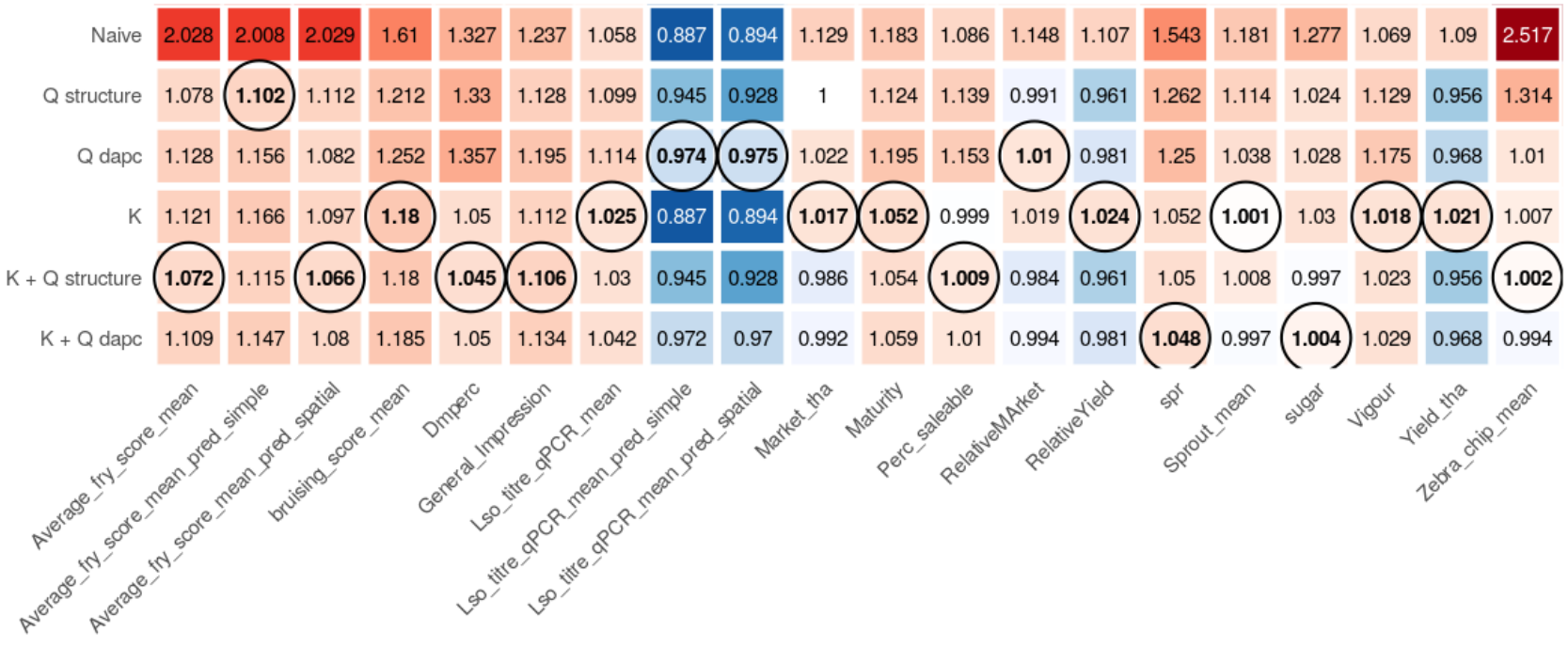
GWAS analysis

- Used the `GWASpoly` package
- Population structure explored with `STRUCTURE` and `DAPC` (`adegenet` R package)
- For each trait of interest, tested 8 genetic models $\times$ 6 population models:

Population model	**Kinship matrix**	**Subopulations membership probabilities as covariates**
Naive model	-	-
K model	LOCO method	-
$Q_{STRUCTURE}$ model	-	STRUCTURE
Q_{DAPC} model	-	DAPC
$K + Q_{STRUCTURE}$ model	LOCO method	STRUCTURE
$K + Q_{DAPC}$ model	LOCO method	DAPC

Correcting for population structure

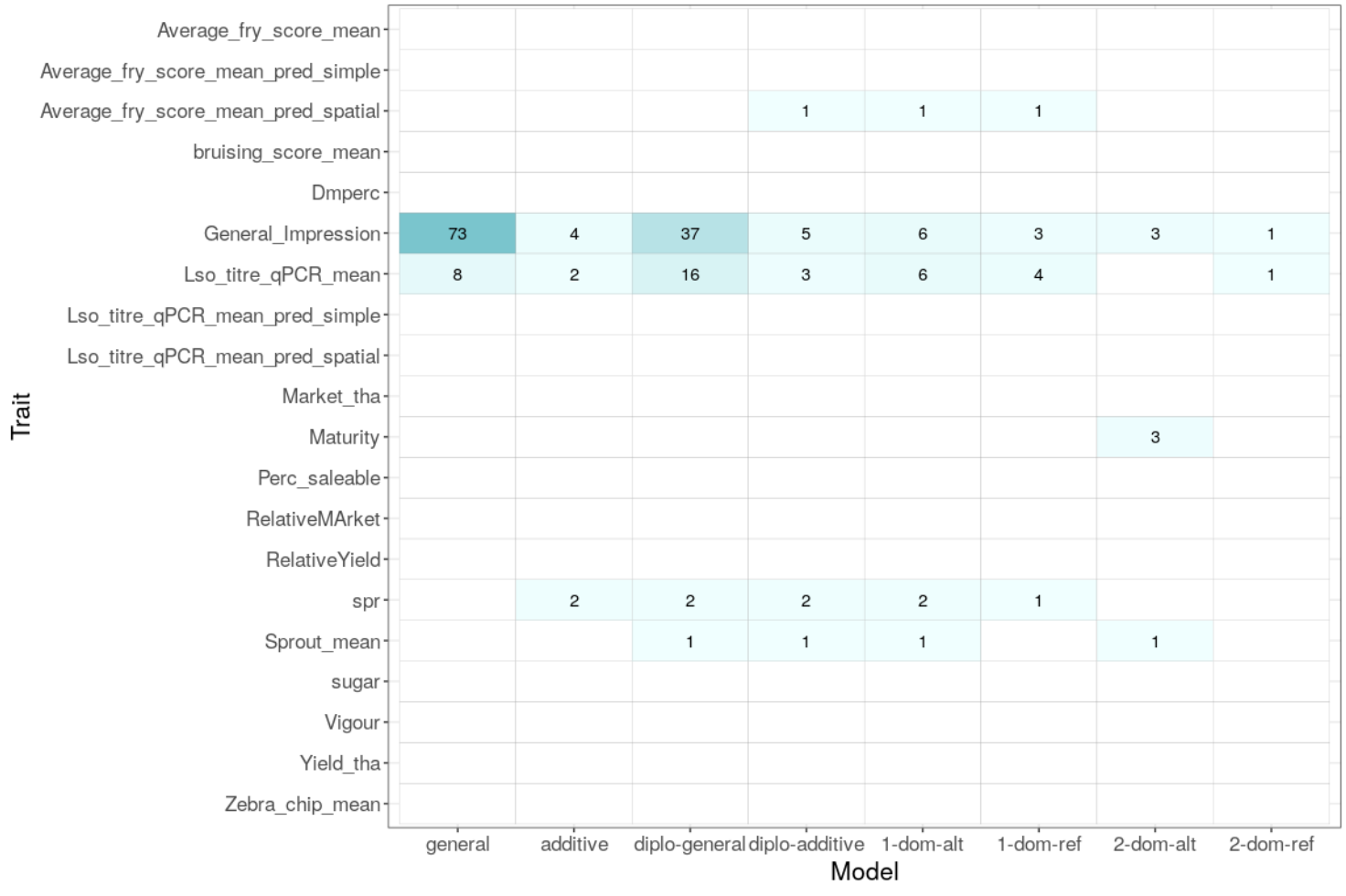
For each trait, average inflation factor over all genetic models used to select the best population model:



Ideal average inflation factor is closest to 1 and > 1.

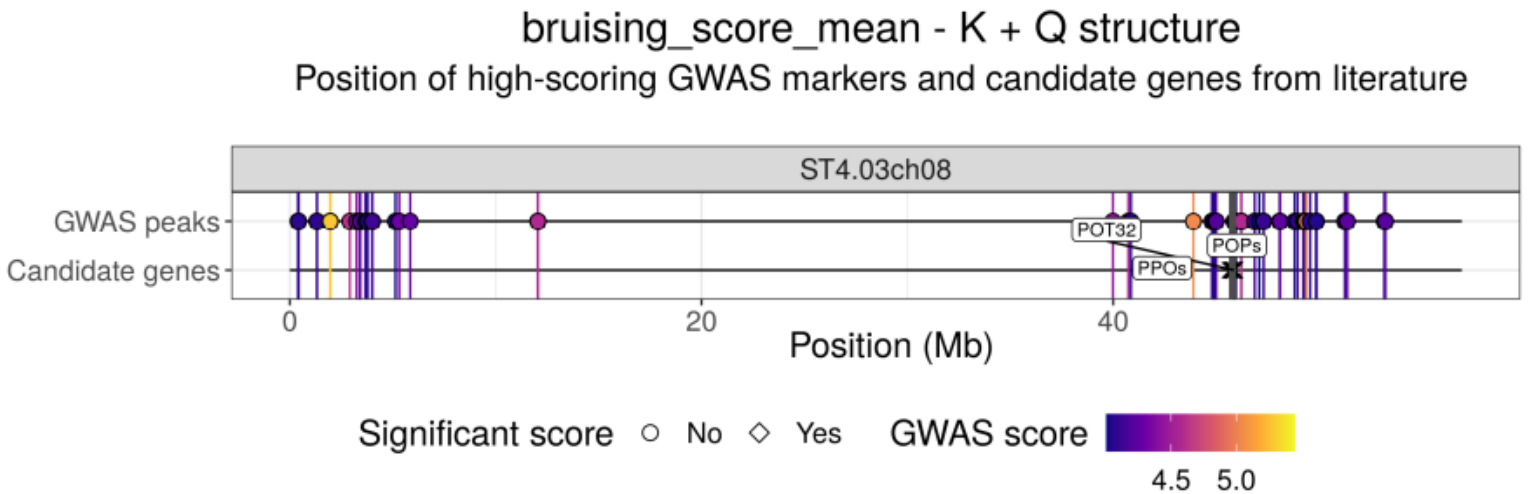
Significant QTLS

- Used `Meff` method to compute significance threshold
- Had to adapt the function since cannot handle that many variants



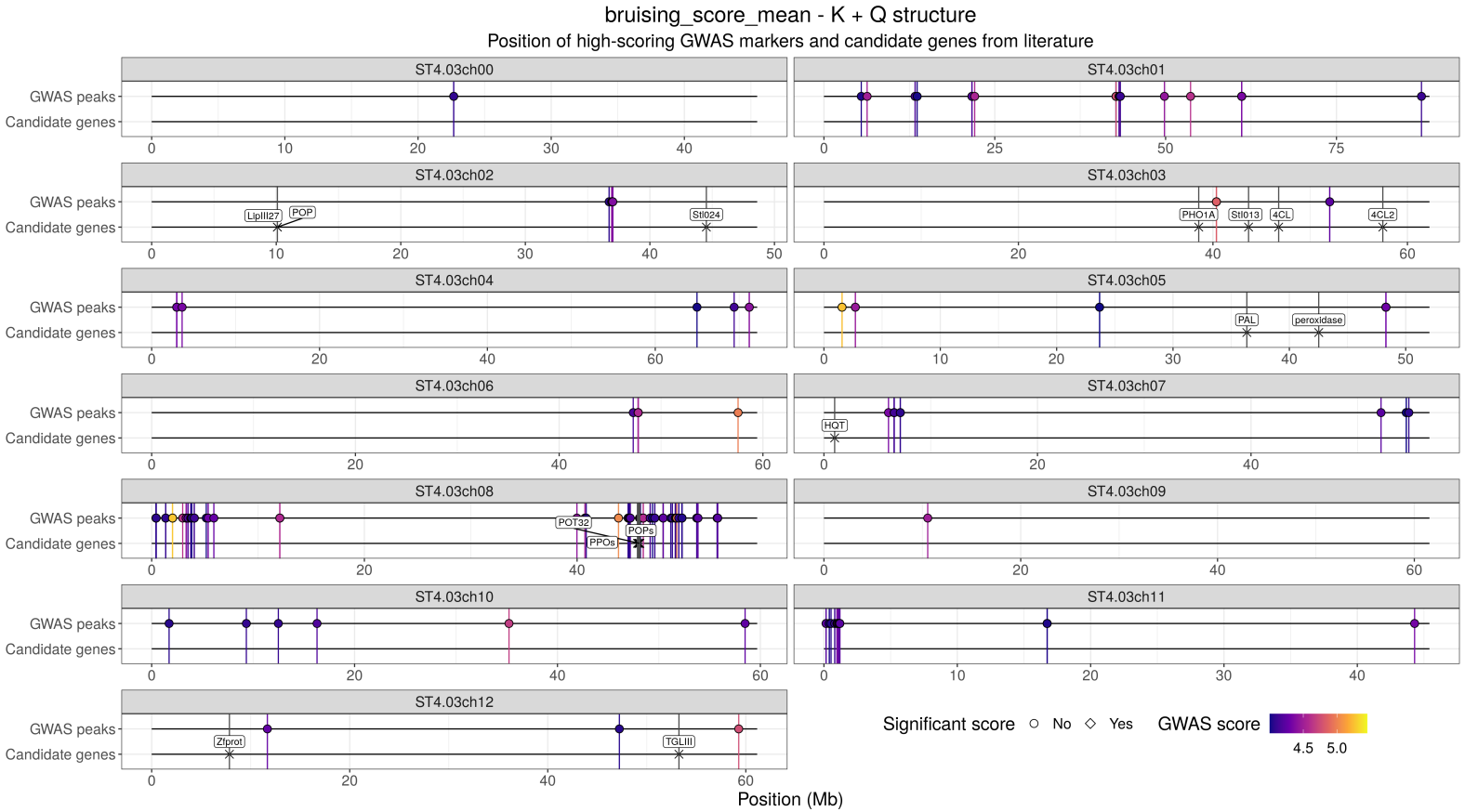
GWAS score peaks vs known QTLS and candidate genes

Comparing genomic position of variants with unadjusted p-value $< 10^{-4}$ with position of QTL regions identified by previous studies:



GWAS score peaks vs known QTLS and candidate genes

Comparing genomic position of variants with unadjusted p-value $< 10^{-4}$ with position of QTL regions identified by previous studies:



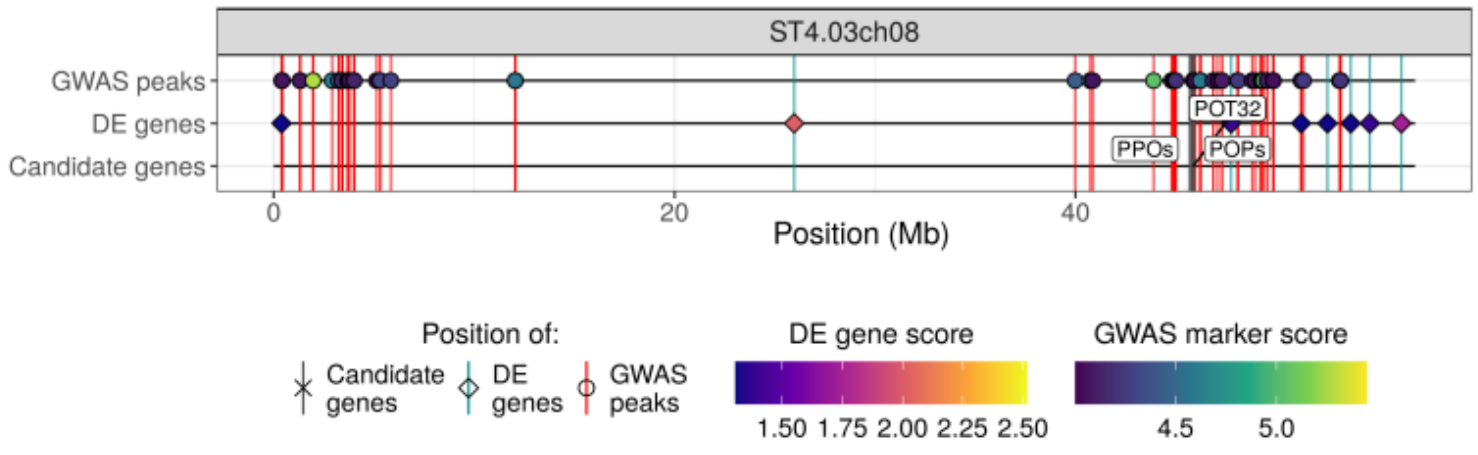
Transcriptomics differential expression

Transcriptomics differential expression

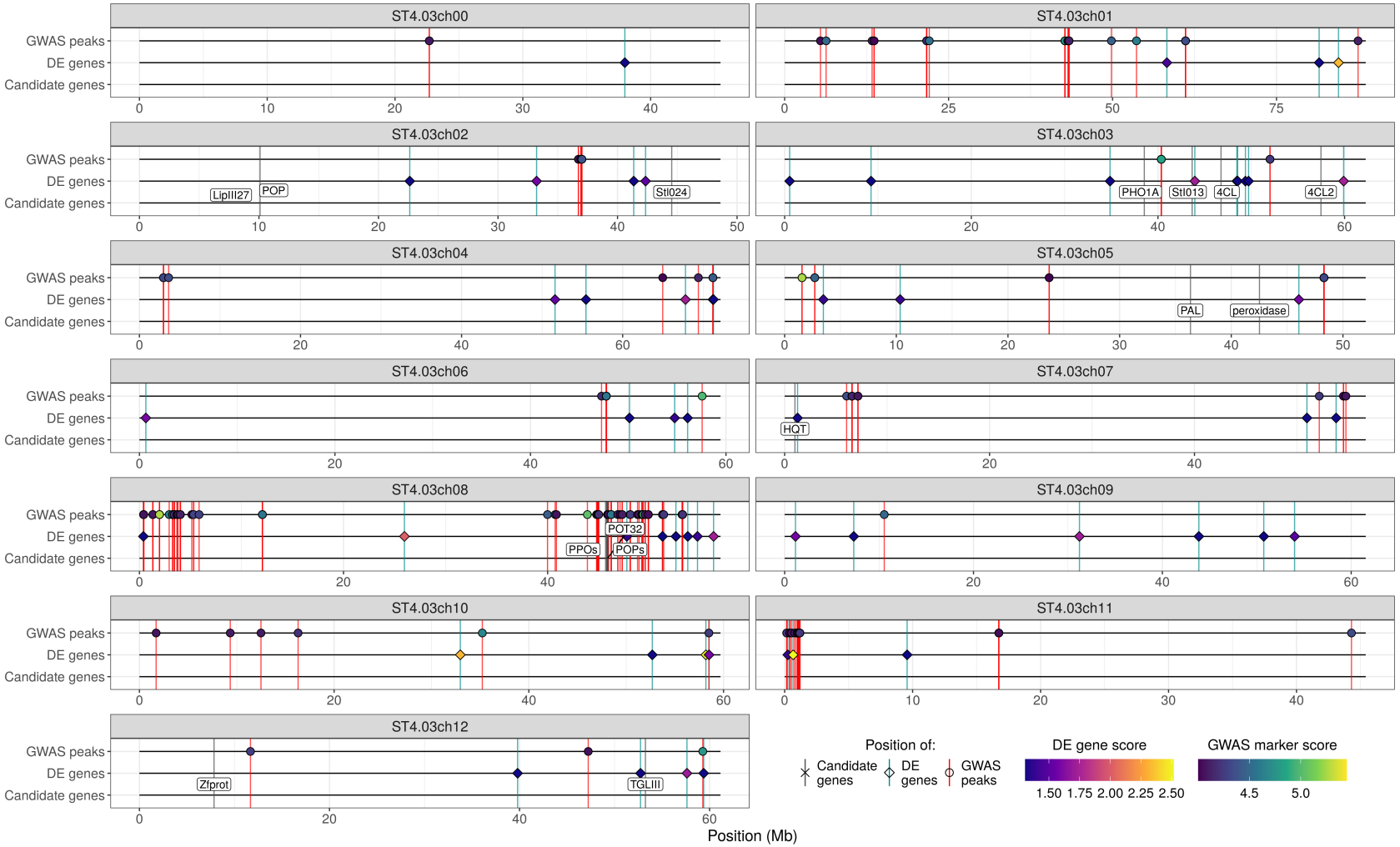
Used the `DESeq2` package for transcriptomics differential expression analysis:

- 30 up-regulated genes
- 27 down-regulated genes

Can compare the results from the GWAS and differential expression analyses:



Transcriptomics differential expression to interpret GWAS results



Multi-omics data integration

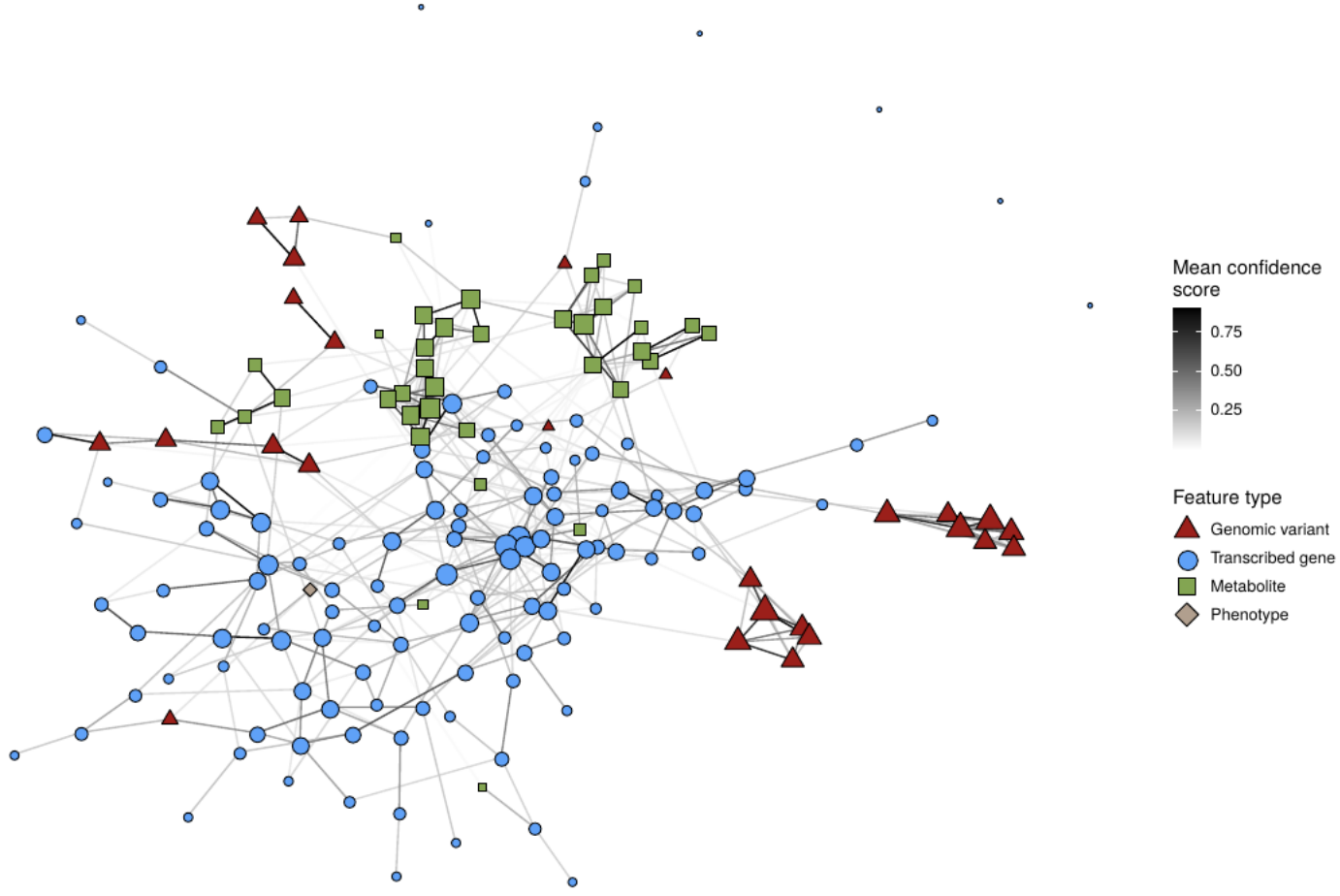
Feature selection using DIABLO

Used `DIABLO` from the `mixOmics` package to select variants, transcripts and compounds associated with the tuber bruising score.

- "Low bruising score" samples vs "high bruising score" samples
- `DIABLO` selected transcripts and compounds not found differentially expressed (e.g. glutathione-S-transferase transcript)
- Caveat: population structure influences which variants are selected $\Rightarrow$ restricted the variants to only those with a high GWAS score

Reconstruction of a multi-omics network

Used several causal inference methods to assess the causal relationships between the selected features.



Conclusion

Take-home message

- Genomics data from breeding programme can be useful to locate genomic regions associated with a trait of interest
- Addition of other omics data (e.g. transcriptomics and metabolomics) can provide an alternative way to detect potential causal genes or biological pathways associated with a trait of interest

Acknowledgements

- Samantha Baldwin (PFR)
- Susan Thomson (PFR)
- Matthieu Vignes (Massey)
- Patrick Biggs (Massey)
- Rebecca Bloomer (PFR)
- Katrina Monaghan (PFR)
- Nigel Joyce (PFR)
- Martin Shaw (PFR)

Thank you for your attention!

Any questions?