

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

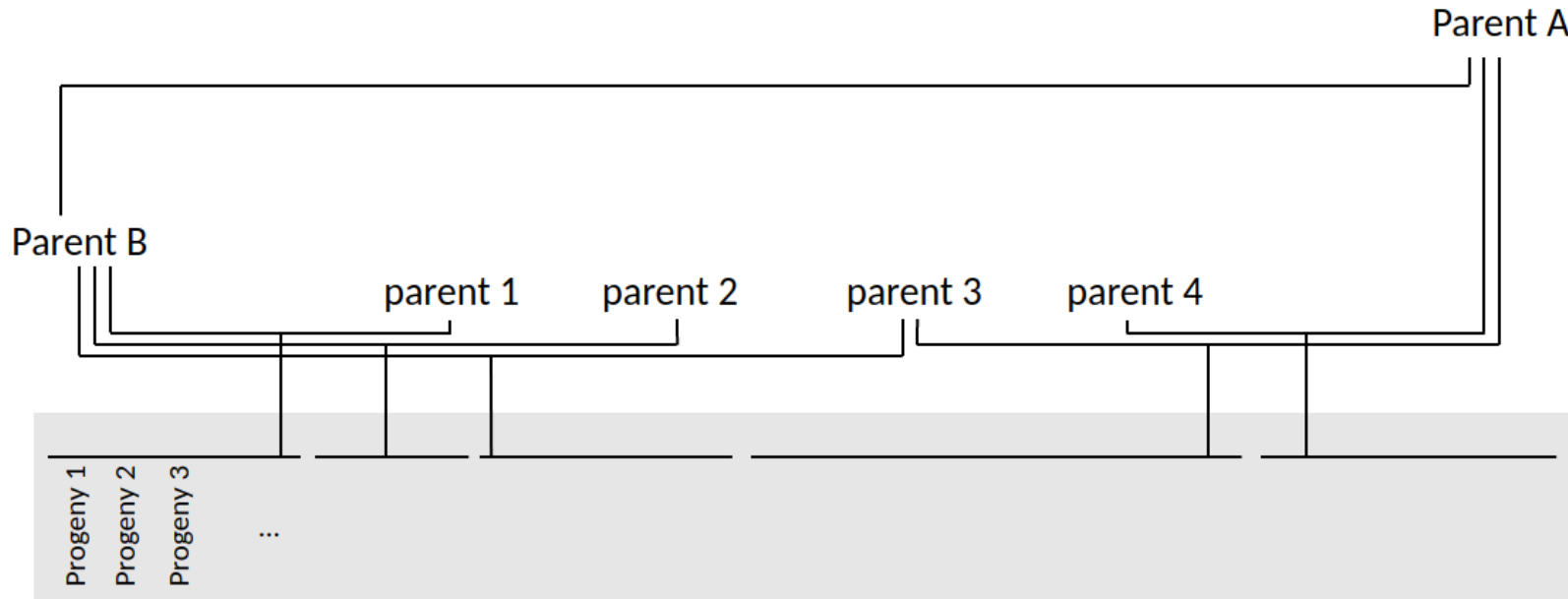
Massey University, New Zealand

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	Subpopulations 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:

Naive	2.028	2.008	2.029	1.61	1.327	1.237	1.058	0.887	0.894	1.129	1.183	1.086	1.148	1.107	1.543	1.181	1.277	1.069	1.09	2.517
Q structure	1.078	1.102	1.112	1.212	1.33	1.128	1.099	0.945	0.928	1	1.124	1.139	0.991	0.961	1.262	1.114	1.024	1.129	0.956	1.314
Q dapc	1.128	1.156	1.082	1.252	1.357	1.195	1.114	0.974	0.975	1.022	1.195	1.153	1.01	0.981	1.25	1.038	1.028	1.175	0.968	1.01
K	1.121	1.166	1.097	1.18	1.05	1.112	1.025	0.887	0.894	1.017	1.052	0.999	1.019	1.024	1.052	1.001	1.03	1.018	1.021	1.007
K + Q structure	1.072	1.115	1.066	1.18	1.045	1.106	1.03	0.945	0.928	0.986	1.054	1.009	0.984	0.961	1.05	1.008	0.997	1.023	0.956	1.002
K + Q dapc	1.109	1.147	1.08	1.185	1.05	1.134	1.042	0.972	0.97	0.992	1.059	1.01	0.994	0.981	1.048	0.997	1.004	1.029	0.968	0.994
Average_fry_score_mean																				
Average_fry_score_mean_pred_simple																				
Average_fry_score_mean_pred_spatial																				
bruising_score_mean																				
Dmperc																				
General_Impression																				
Lso_titre_qPCR_mean																				
Lso_titre_qPCR_mean_pred_simple																				
Lso_titre_qPCR_mean_pred_spatial																				
Market_tha																				
Maturity																				
Perc_saleable																				
RelativeMarket																				
RelativeYield																				
spr																				
Sprout_mean																				
sugar																				
Vigour																				
Yield_tha																				
Zebra_chip_mean																				

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

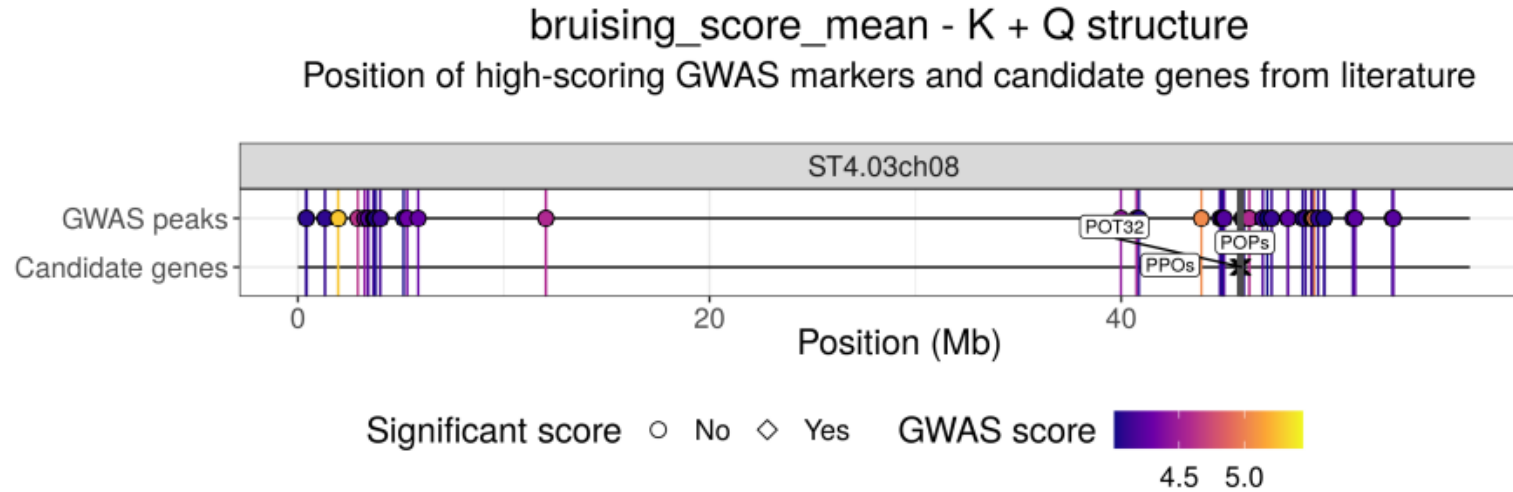
Significant QTLS

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

Trait	Average_fry_score_mean							
	Average_fry_score_mean_pred_simple							
	Average_fry_score_mean_pred_spatial			1	1	1		
	bruising_score_mean							
	Dmperc							
	General_Impression	73	4	37	5	6	3	3
	Lso_titre_qPCR_mean	8	2	16	3	6	4	1
	Lso_titre_qPCR_mean_pred_simple							
	Lso_titre_qPCR_mean_pred_spatial							
	Market_tha							
	Maturity						3	
	Perc_saleable							
	RelativeMarket							
	RelativeYield							
	spr		2	2	2	2	1	
	Sprout_mean			1	1	1		1
	sugar							
	Vigour							
	Yield_tha							
	Zebra_chip_mean							
		general	additive	diplo-general	diplo-additive	1-dom-alt	1-dom-ref	2-dom-alt
								2-dom-ref
		Model						

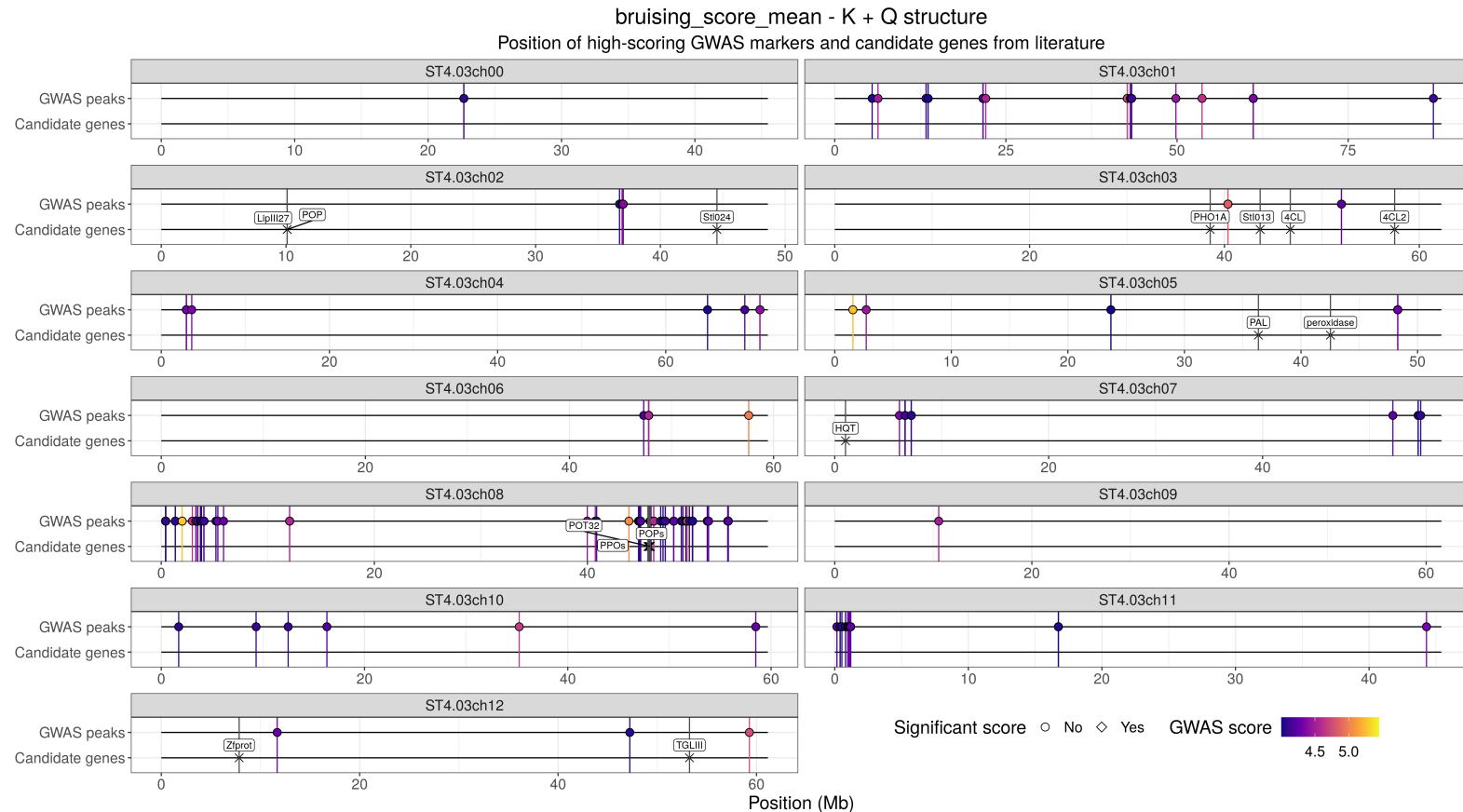
GWAS score peaks vs known QTLs and candidate genes

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:



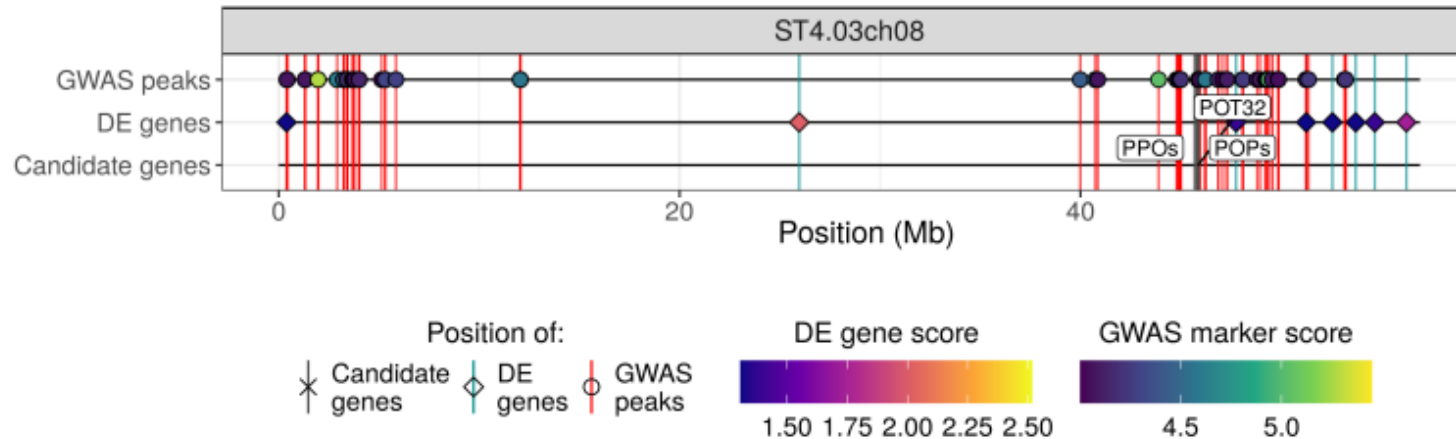
Transcriptomics differential expression

Transcriptomics differential expression

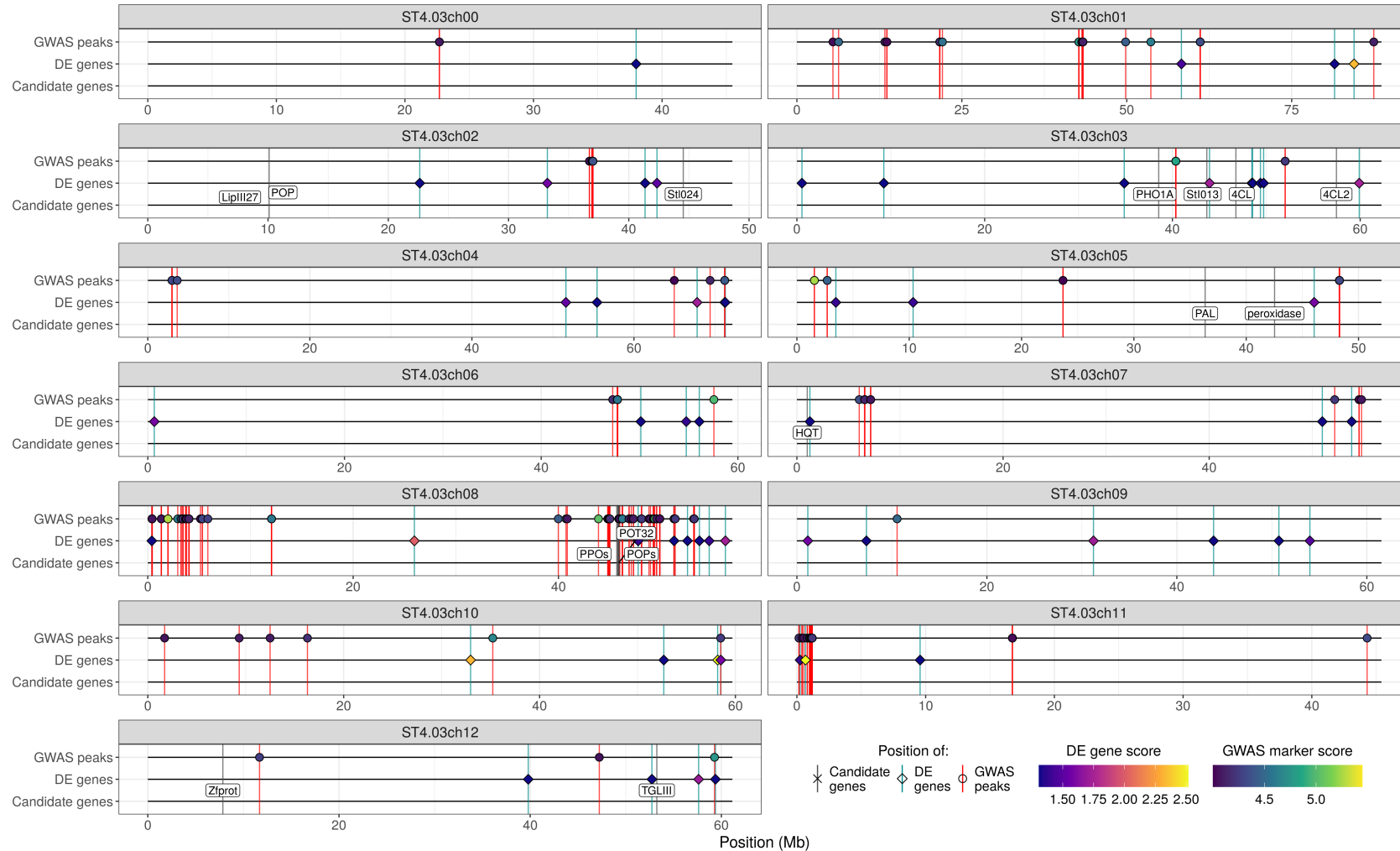
Used the DESeq2 package for transcriptomics differential expression analysis:

- 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

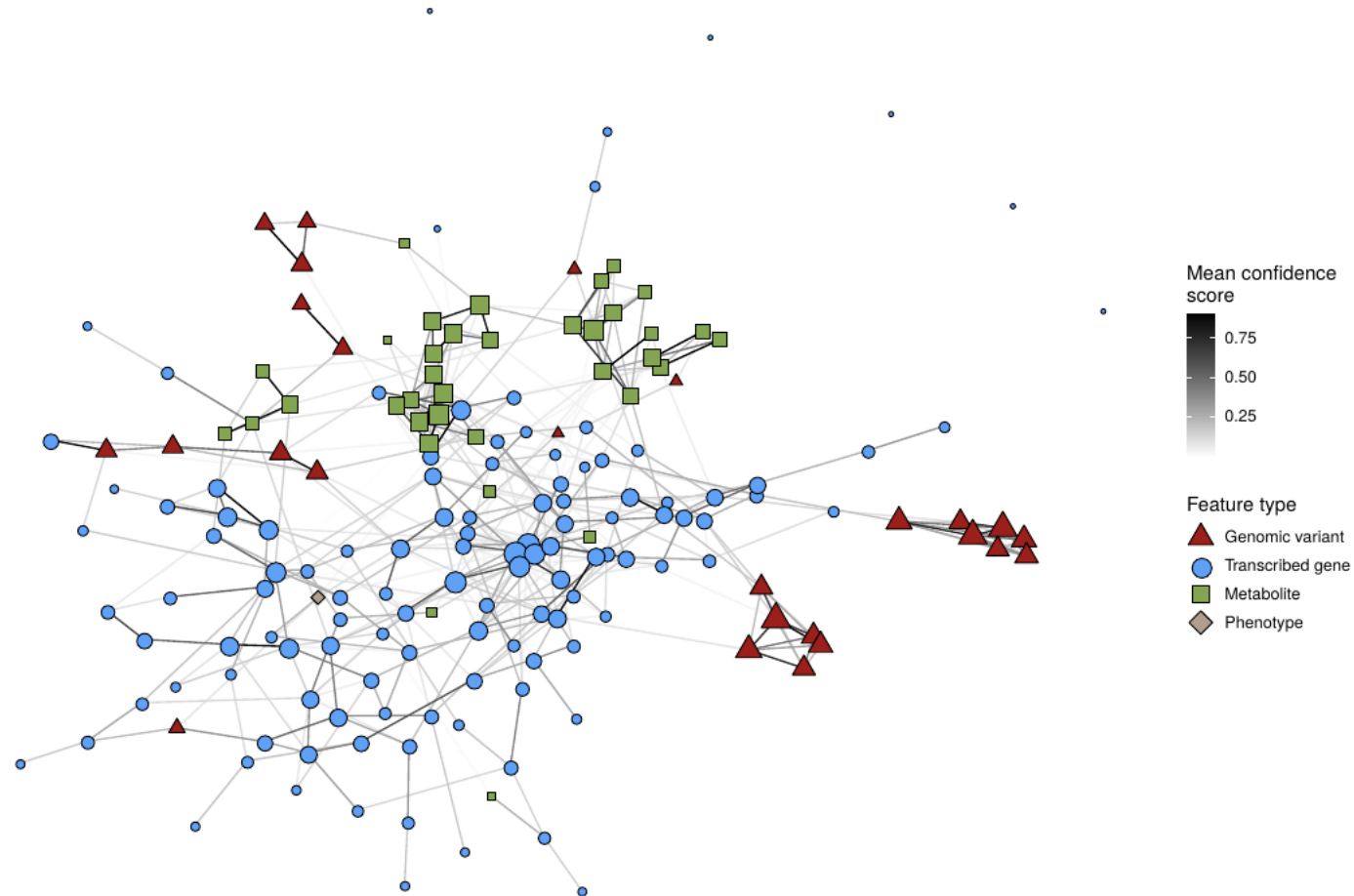
Feature selection using DIABLO

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?