Identifying regulatory variants to enhance mapping and prediction of cattle traits

Ruidong Xiang et al

Computational Biology | Genomics and Cellular Science

Agriculture Victoria Research,

AgriBio Research centre,

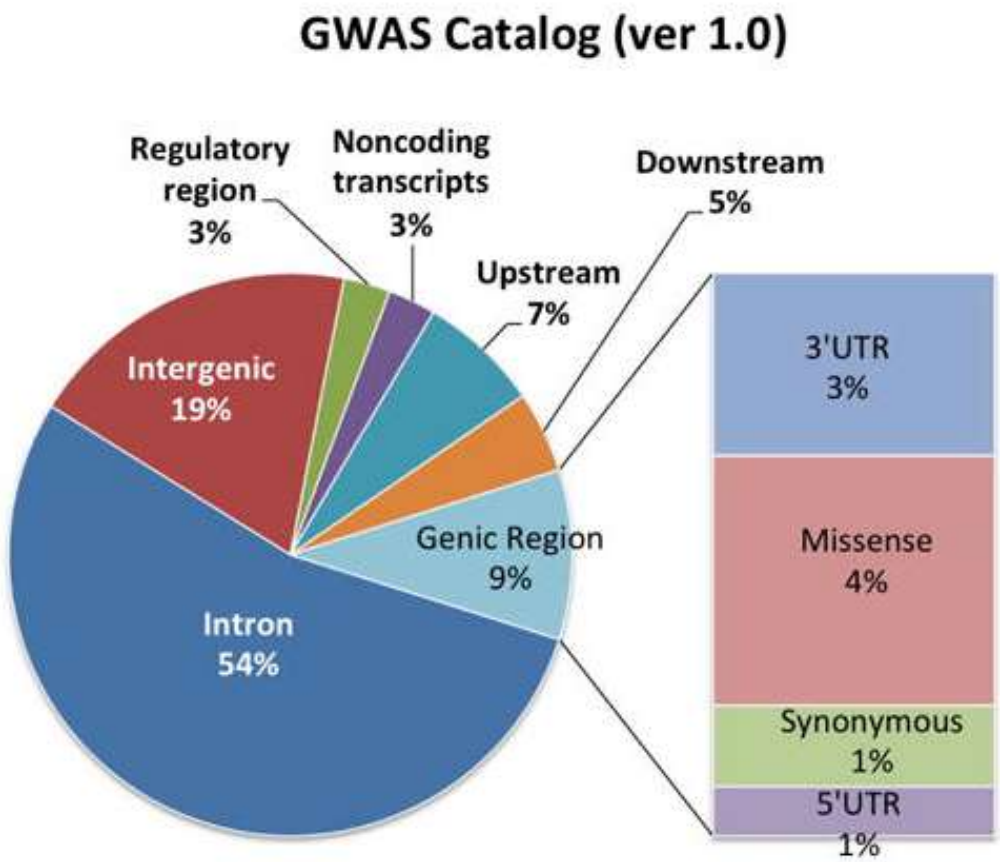
5 Ring road, Bundoora, 3083 Victoria, Australia

DairyBio Dairy Australia GARDINER FOUNDATION THE UNIVERSITY OF MELBOURNE

AGRICULTURE VICTORIA

Most complex traits have polygenic architecture

- Most GWAS hits fall into non-coding regions



➢ **Quantitative trait loci (QTL) should impact gene regulation**

Controversies on the importance of regulatory variants

- Previously, eQTL are reported to only contribute a small fraction of trait h^2

❖11% of h^2 for human disease (Yao et al 2020)

❖Term “missing regulation” in human genetics (Connally et al 2021, Mostafavi et al 2022)

❖12% of h^2 for cattle traits (Xiang et al 2019)

❖< 5% of cattle QTL of fat% overlapped with eQTL (van den Berg et al. 2020)

e/sQTLs explained ~70% of h^2 of dairy cattle traits

Cell Genomics

CellPress
OPEN ACCESS

Article

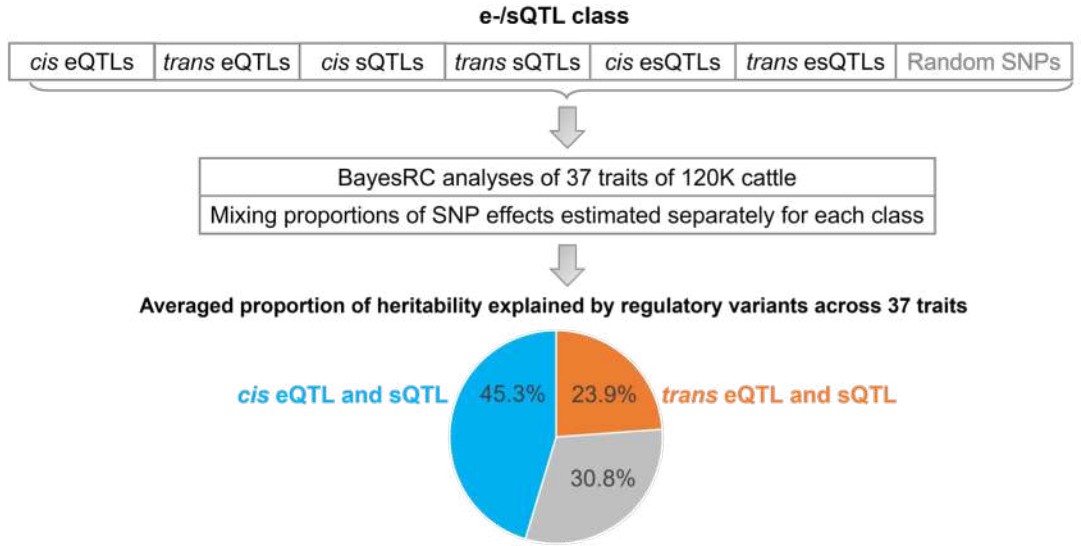
Gene expression and RNA splicing explain large proportions of the heritability for complex traits in cattle

Ruidong Xiang,[1,2,3,12,*] Lingzhao Fang,[4,5] Shuli Liu,[6] Iona M. Macleod,[2] Zhiqian Liu,[2] Edmond J. Breen,[2] Yahui Gao,[7] George E. Liu,[7] Albert Tenesa,[4,8] CattleGTEx Consortium, Brett A. Mason,[2] Amanda J. Chamberlain,[2,11] Naomi R. Wray,[9,10] and Michael E. Goddard[1,2]

[1]Faculty of Veterinary & Agricultural Science, the University of Melbourne, Parkville, VIC 3052, Australia
[2]Agriculture Victoria, AgriBio, Centre for AgriBiosciences, Bundoora, VIC 3083, Australia
[3]Cambridge-Baker Systems Genomics Initiative, Baker Heart and Diabetes Institute, Melbourne, VIC 3004, Australia
[4]MRC Human Genetics Unit at the Institute of Genetics and Cancer, the University of Edinburgh, Edinburgh, UK
[5]Center for Quantitative Genetics and Genomics, Aarhus University, Aarhus, Denmark
[6]Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang 310024, China
[7]Animal Genomics and Improvement Laboratory, Henry A. Wallace Beltsville Agricultural Research Center, Agricultural Research Service, USDA, Beltsville, MD 20705, USA
[8]The Roslin Institute, Royal (Dick) School of Veterinary Studies, the University of Edinburgh, Midlothian EH25 9RG, UK
[9]Institute for Molecular Bioscience, the University of Queensland, Brisbane, QLD 4072, Australia
[10]Queensland Brain Institute, the University of Queensland, Brisbane, QLD 4072, Australia
[11]School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3083, Australia
[12]Lead contact
*Correspondence: ruidong.xiang@unimelb.edu.au
https://doi.org/10.1016/j.xgen.2023.100385

SUMMARY

Many quantitative trait loci (QTLs) are in non-coding regions. Therefore, QTLs are assumed to affect gene regulation. Gene expression and RNA splicing are primary steps of transcription, so DNA variants changing gene expression (eVariants) or RNA splicing (sVariants) are expected to significantly affect phenotypes. We quantify the contribution of eVariants and sVariants detected from 16 tissues (n = 4,725) to 37 traits of ~120,000 cattle (average magnitude of genetic correlation between traits = 0.13). Analyzed in Bayesian mixture models, averaged across 37 traits, *cis* and *trans* eVariants and sVariants detected from 16 tissues jointly explain 69.2% (SE = 0.5%) of heritability, 44% more than expected from the same number of random variants. This 69.2% includes an average of 24% from *trans* e-/sVariants (14% more than expected). Averaged across 56 lipidomic traits, multi-tissue *cis* and *trans* e-/sVariants also explain 71.5% (SE = 0.3%) of heritability, demonstrating the essential role of proximal and distal regulatory variants in shaping mammalian phenotypes.



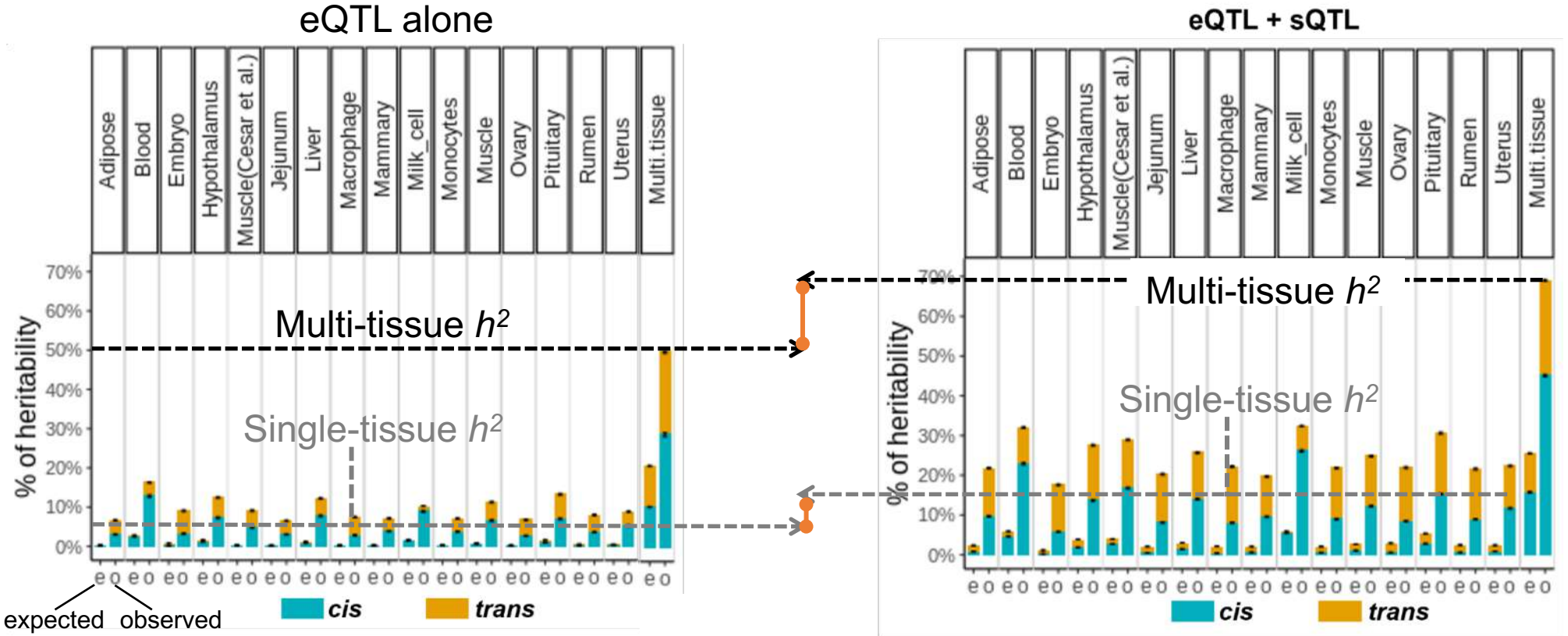
Highlights

- Map *cis* and *trans* eQTLs and RNA splicing sQTLs in 16 tissues of 4,725 cattle
- Use *cis* and *trans* e/sQTLs to partition heritability (h^2) of 37 traits of 120,000 cattle
- *cis* and *trans* e/sQTLs explained an average of 69.2% of h^2 across phenotypic traits
- *cis* and *trans* e/sQTLs are essential for mammalian phenotypes

DairyBio Dairy Australia GARDINER FOUNDATION THE UNIVERSITY OF MELBOURNE

AGRICULTURE VICTORIA

h^2 due to regulatory variants

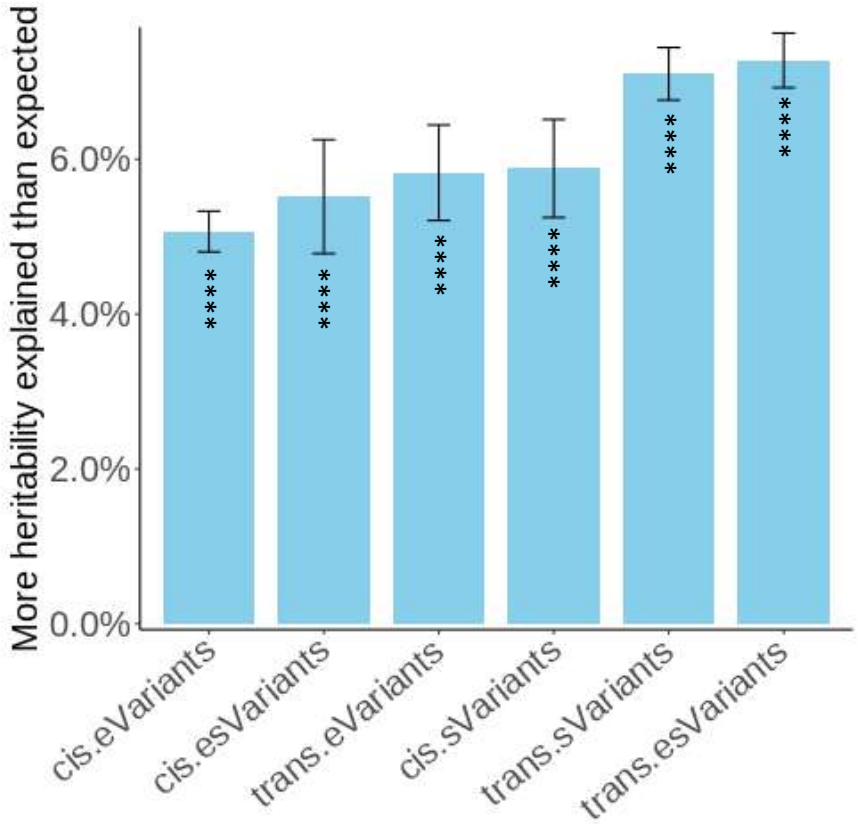


16 single-tissue + 1 multi-tissue BayesRC analysis

➢ **Including more regulatory variants explains higher h^2**

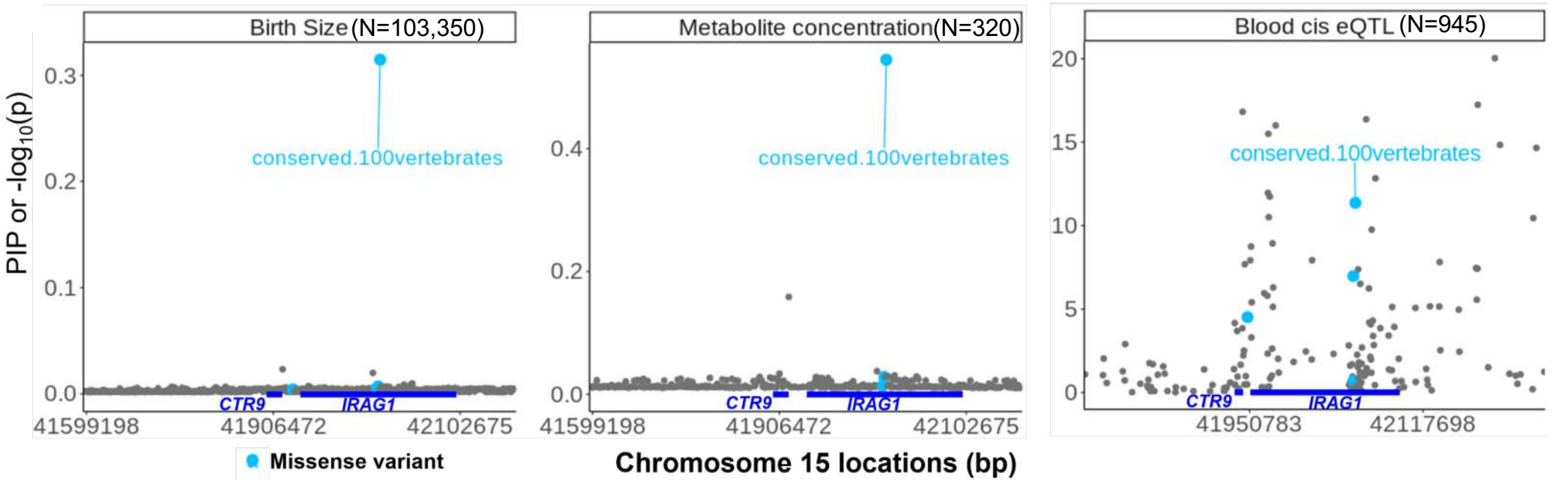
h^2 of lipidomic traits due to regulatory variants

- The same methods and e/sQTL prior
- Applied to 56 lipidomic traits measured by liquid chromatography-mass spectrometry (LC-MS)



➢ 71% of lipidomic traits h^2 due to cis and trans e/sQTL

Causal variant for fertility and multi-omics

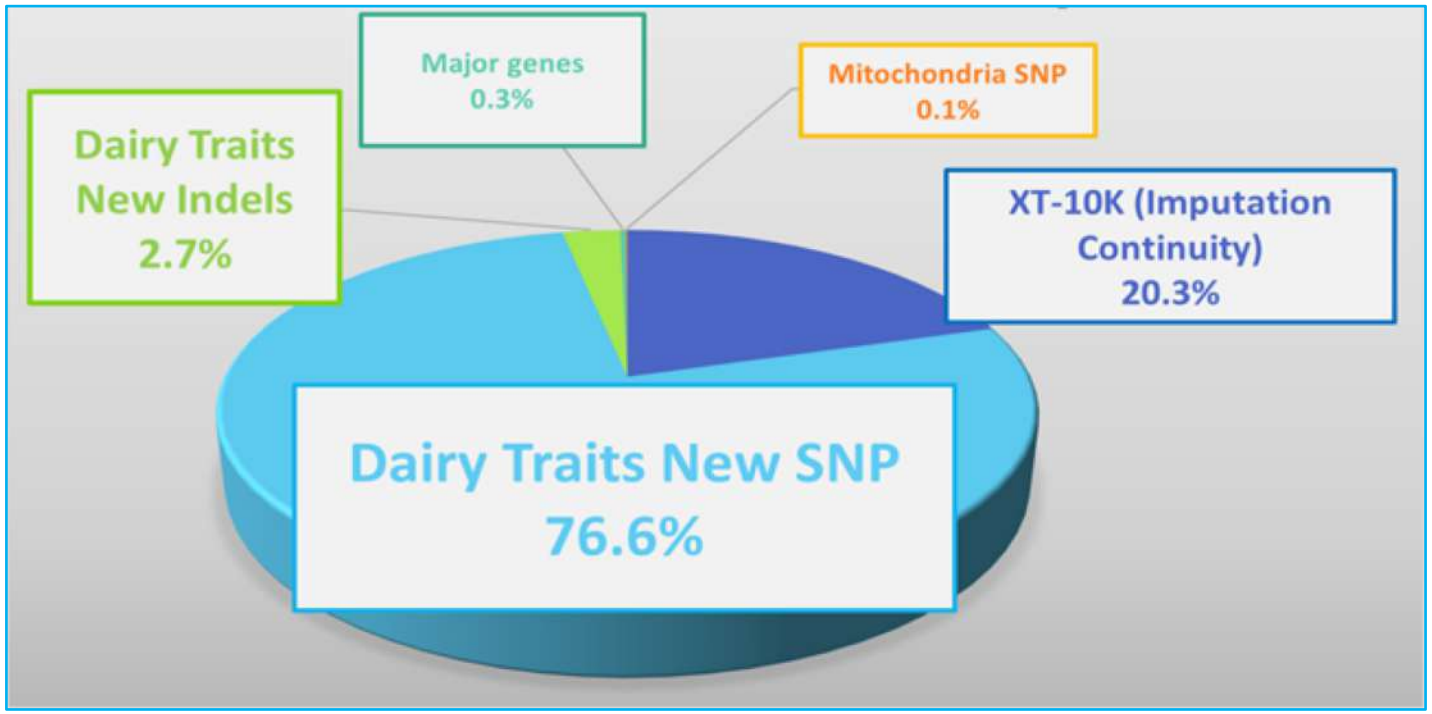


- chr15:42,044,576 (rs137255300) within *IRAG1* affects birth size and the concentration of lactosylceramide and is a *cis* eQTL for *CTR9* in blood

- chr15:42,044,576 (rs137255300) is also a missense mutation and conserved across 100 vertebrates

Customising SNP chip using regulatory variants

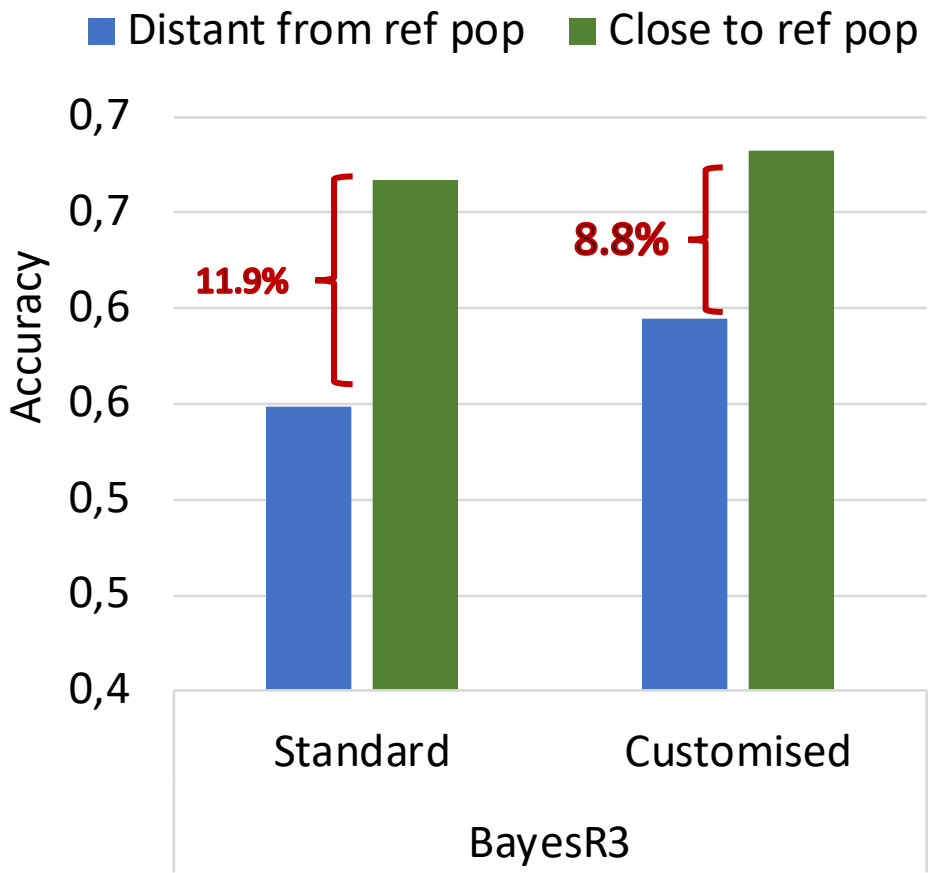
Research Discovery & Validation → Chip Design & Manufacture → Genotype Imputation Reference → Impute historical & new animals → Re-evaluate & use in industry → Uptake on Commercial Chips



Credit: Iona M. MacLeod

Industry adoption of customised SNP Chip

- Used to genotype and genetically evaluate ~20,000 cows
- Implemented for Holstein and Australian red
- Validated genomic prediction accuracy in industry settings

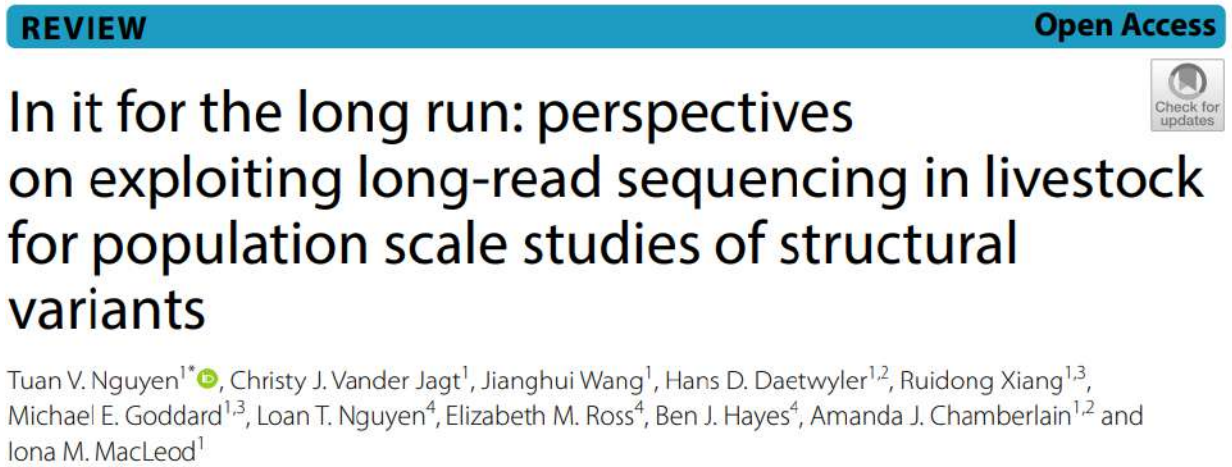


Credit: Iona M. MacLeod and Irene van den Berg

Future steps

- Some markers on the SNP chip didn't work well – chip manufacturing process, genotyping QCs, etc
- New research, new traits, new variants ….
- low-pass (0.5-2X) sequencing: costs have been moving towards SNP chip
- Long-read sequencing consortium:

iona.macleod@agriculture.vic.gov.au;
tuan.nguyen@agriculture.vic.gov.au
amanda.chamberlain@agriculture.vic.gov.au


REVIEW Open Access

Check for updates

In it for the long run: perspectives on exploiting long-read sequencing in livestock for population scale studies of structural variants

Tuan V. Nguyen[1*], Christy J. Vander Jagt[1], Jianghui Wang[1], Hans D. Daetwyler[1,2], Ruidong Xiang[1,3], Michael E. Goddard[1,3], Loan T. Nguyen[4], Elizabeth M. Ross[4], Ben J. Hayes[4], Amanda J. Chamberlain[1,2] and Iona M. MacLeod[1]

Acknowledgements

DATA ACCESS

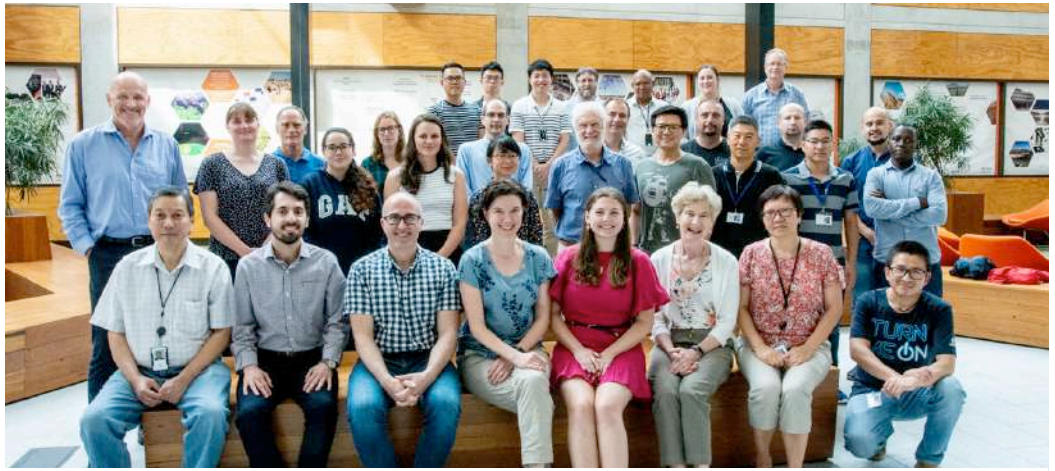
DataGene: Gert Nieuwhof - preparation & provision of data

Cattle GTEx | 1000 Bull Genomes

TEAM MEMBERS AND COLLABORATORS :

Mike Goddard, Iona MacLeod, Lingzhao Fang, Shuli Liu, Edmond Breen, Zhiqian Liu, Simone Rochfort, Yahui Gao, George Liu, Albert Tenesa, Brett Mason, Amanda Chamberlain, Naomi Wray

BASC support team (past and present)



Computational Biology Group