Biology-driven WGS genomic predictions for feed efficiency within and across-breeds

Renzo Bonifazi, **Marzieh Heidaritabar**, AC Bouwman, LD Barlow, L Chen, P Stothard, J Basarab, C Li, G Plastow, E. Karaman, G.C.M. Moreira, the BovReg consortium, B Gredler-Grandl

WAGENINGEN UNIVERSITY & RESEARCH

UNIVERSITY OF ALBERTA

BovReg
Understanding cattle genomes

EuroFAANG

Outline

1. Background and aim
2. Data available and usage of BovReg genomic features
3. Scenarios
4. Within-breed results: NLD and CAN
5. Investigating QTLs
6. Across-breed results
7. Usage of ATAC-seq narrow peaks
8. BayesLV models
9. Conclusions

1. Biology-driven genomic predictions

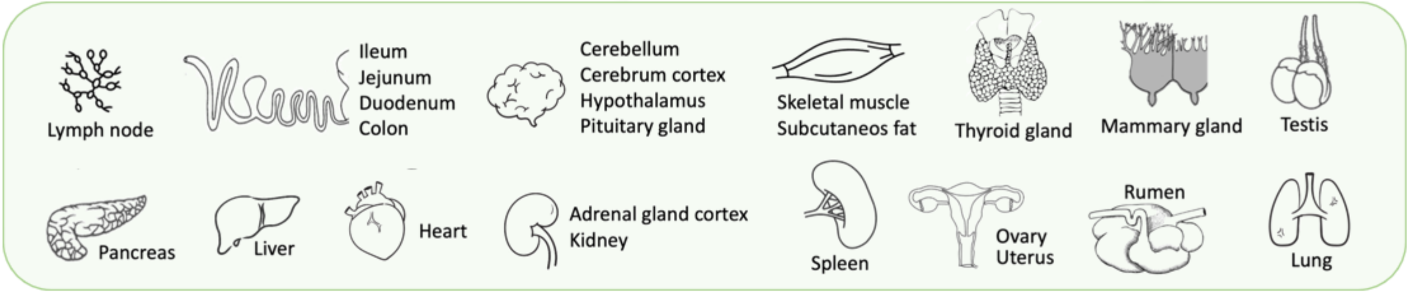
- WGS: millions of variants → pinpoint **causal variants** affecting traits of interest
- Key traits: **biological efficiency**, e.g. **feed efficiency**
- BovReg: catalogue of functionally active **genomic features (GF)** in cattle
- **Functional GF** → SNP prioritization → **Biology-driven genomic predictions**
- **AIM**: Validation of within- and across-breed **biology-driven genomic predictions** using **genomic features** for **dry matter intake (feed efficiency)**

2. Data available for genomic predictions

	NLD	**CAN**
Breed	Holstein	Beef crosses
Dependent variable	DRP on DMI	Pre-corrected DMI
n. animals DMI & geno (training - validation)	~3k (2.2k and 850)	~5.5k (4k and 1.5k)

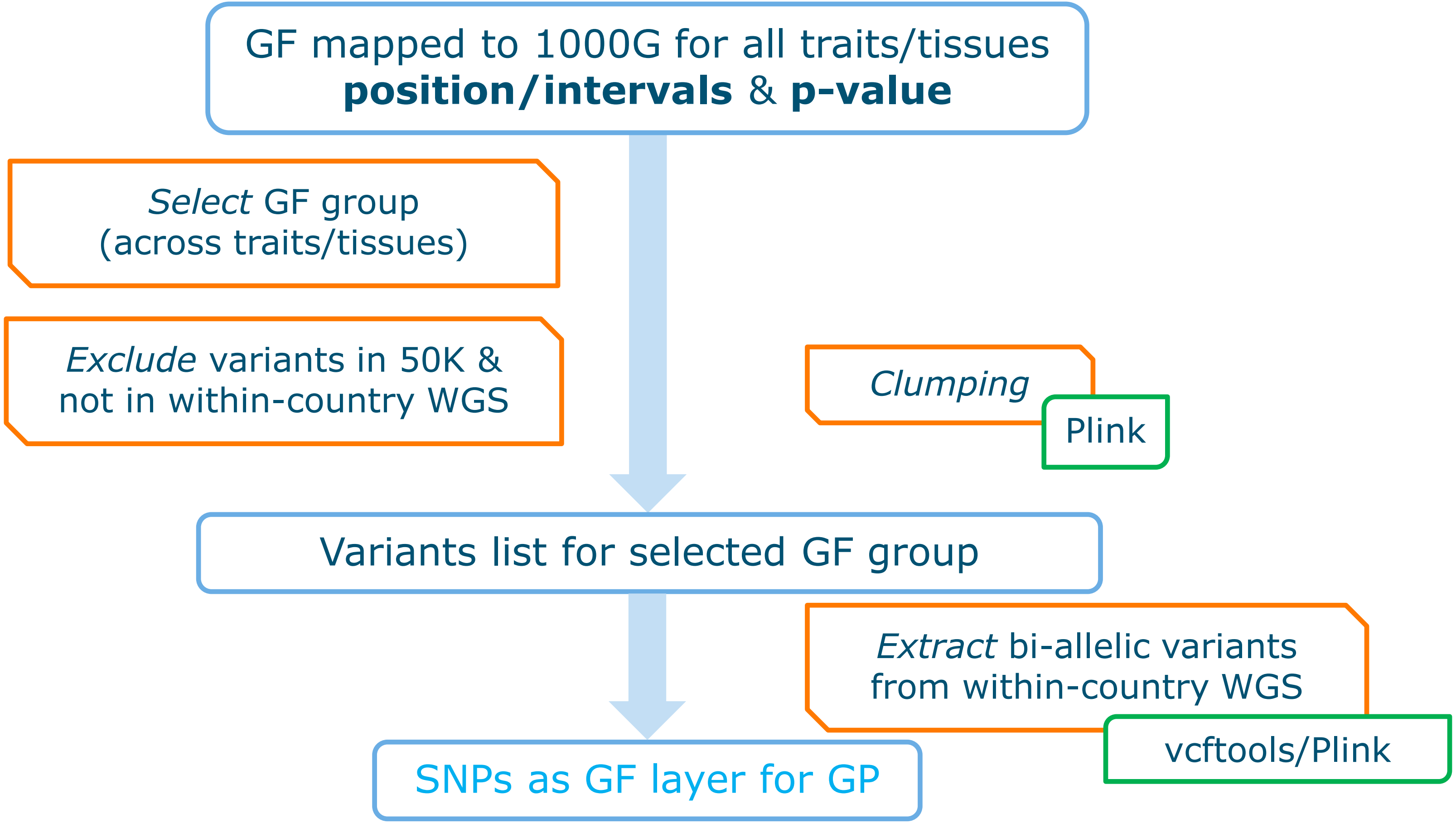
- **forward-in-time validation** (SE via bootstrapping): accuracy and bias
- 50k to imputed WGS (*Beagle*) → prioritize variants based on **GF**

2. Genomic features used

GF	Traits / Tissues
QTL (WP4)	Meat quality, Growth, Milk production, Morphology, Fertility, Health, Feed efficiency, Methane
eQTL (WP4)	(Gene, Transcript, Splice) Jejunum, Blood, Liver, Mammary Gland, Adipose, Muscle, Milk, Rumen
ATAC-seq (WP2)	

From GC Moreira et al., EAAP #939

2. Selection of genomic features



GF mapped to 1000G for all traits/tissues
position/intervals & **p-value**

Select GF group
(across traits/tissues)

Exclude variants in 50K &
not in within-country WGS

Clumping

Plink

Variants list for selected GF group

Extract bi-allelic variants
from within-country WGS

vcftools/Plink

SNPs as GF layer for GP

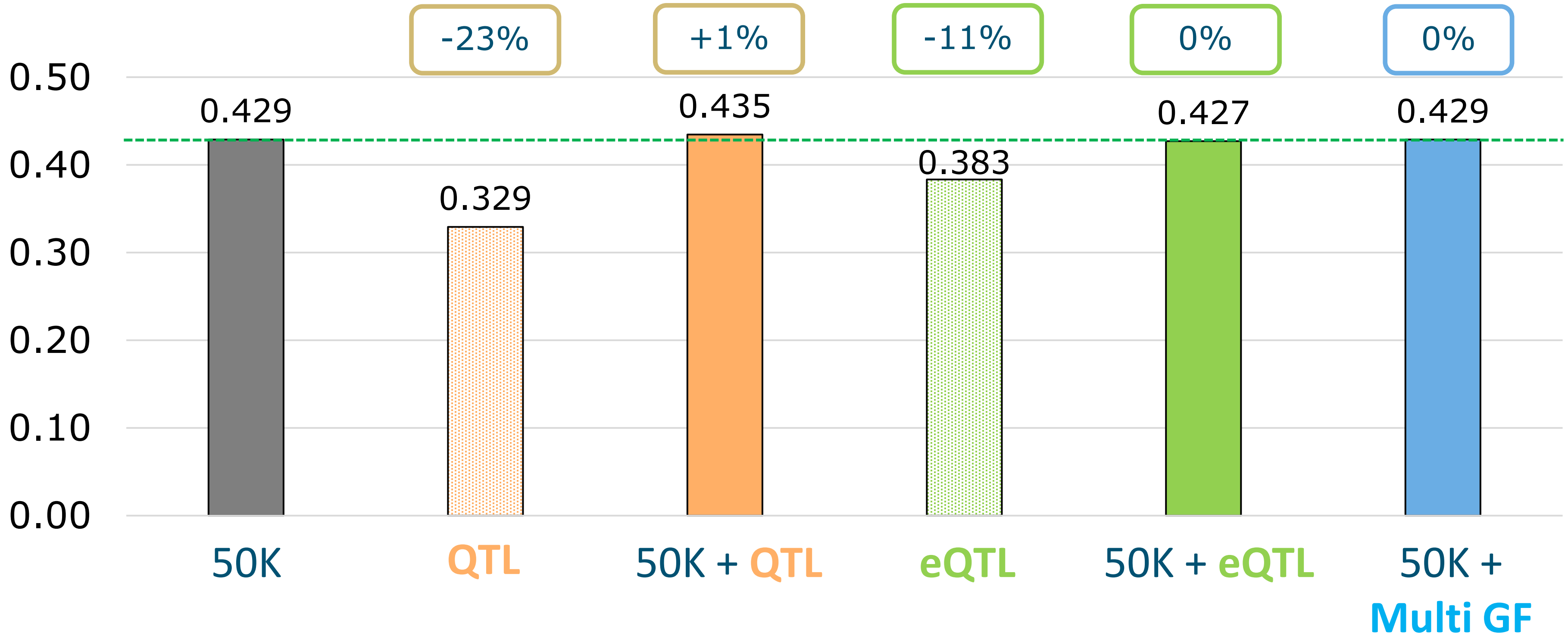
3. Scenarios and software

Scenario	**NLD**	**CAN**
50K	48K	46K
50K + **QTL**	48K + 5,416	46K + 4,222
50K + **eQTL**	48K + 12,401	46K + 11,884
50K + **(QTL, eQTL, ATAC)**	48K + 17,796	46K + 16,089

- *NextGP.jl* [1]
- Base 50K: SNPBLUP (BayesC_0) common variance across SNPs
- Additional GF layer: SNPBLUP or Bayesian (2 mixture model – no advantage)
- QTL, eQTL, ATAC → Multi-GF: overlapping GF (in NLD: 57% within ATAC), BayesRCπ (2 classes)

[1] https://github.com/datasciencetoolkit/NextGP.jl

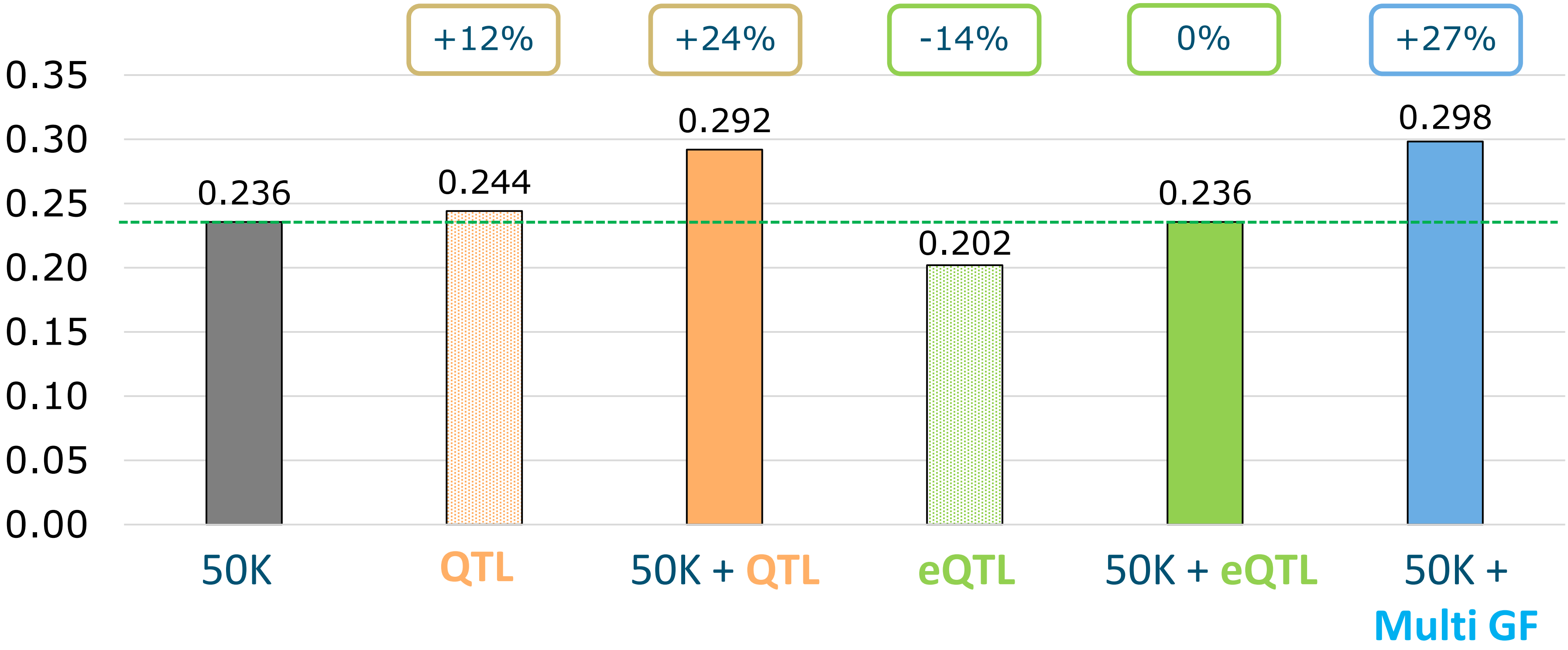
4. Results – NLD: prediction accuracy



- SE: ±0.03
- No impact on dispersion (slope = 0.68 for 50k)

WAGENINGEN
UNIVERSITY & RESEARCH

4. Results – CAN: prediction accuracy

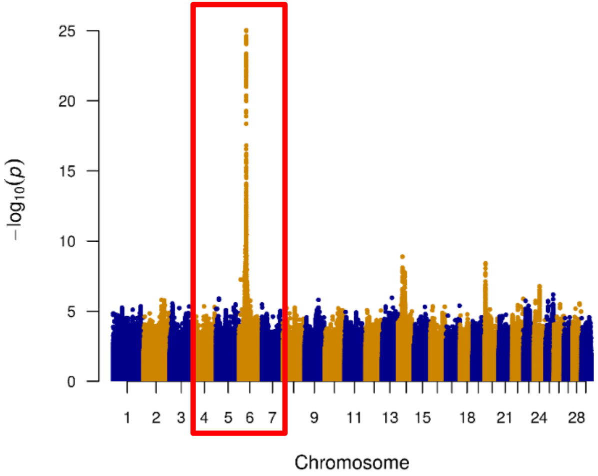


- SE between ±0.02 and ±0.03
- Similar pattern for dispersion (slope = 0.56 for 50k)

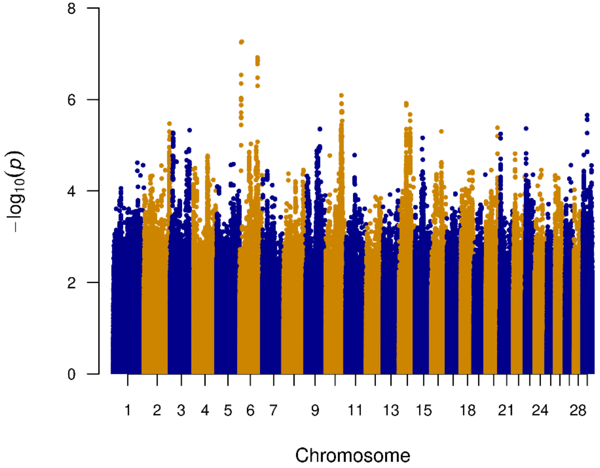
5. Investigating QTL effects in CAN

Gredler-Grandl *et al.* 2022, WCGALP

ALL
10,539 animals
Beef, Holstein, Finnish Red
~30M variants

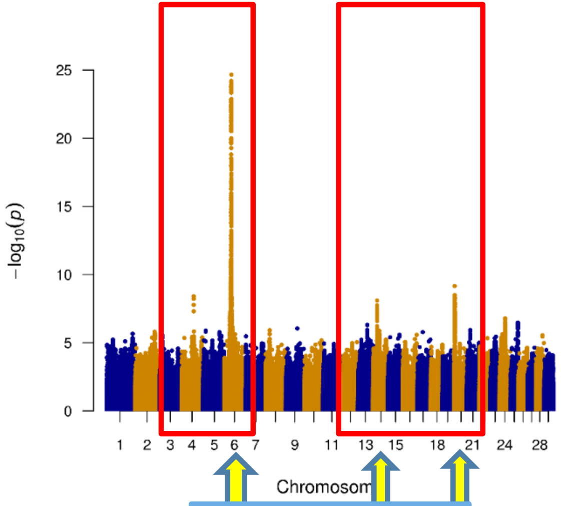


HOL
2,368 animals
Only Holstein
~20M variants



BEEF
7,805 animals
Only Beef
~28M variants

Zhang *et al*. 2020, BMC genomics

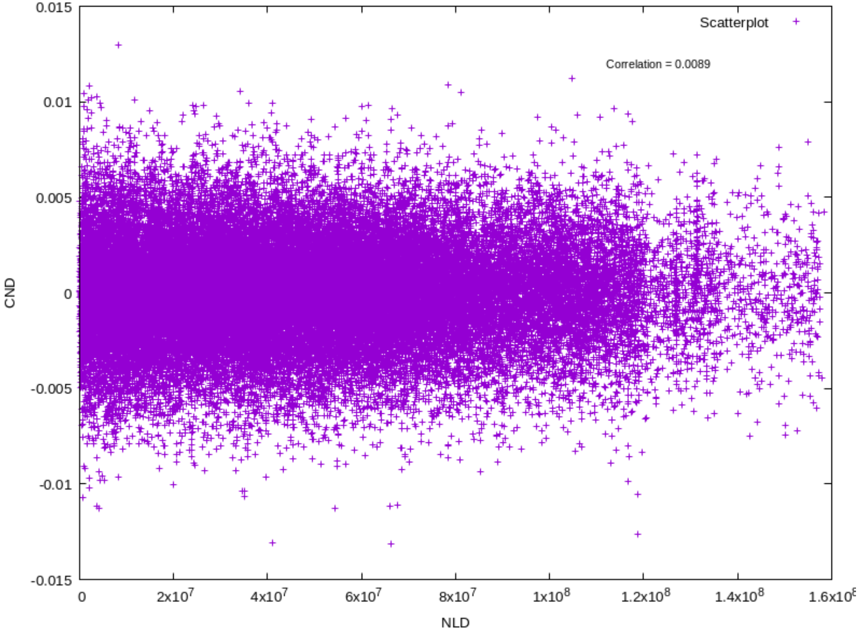


chr 6, 14, 20

🇨🇦 data → filter SNPs with $2pq\alpha^2 \geq 0.0001$

chr	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
count	1	16	6	0	19	21	0	0	1	1	2	0	0	30	6	3	2	1	2	13	0	0	0	2	1	0	1	0	1

6. Across-breed genomic predictions

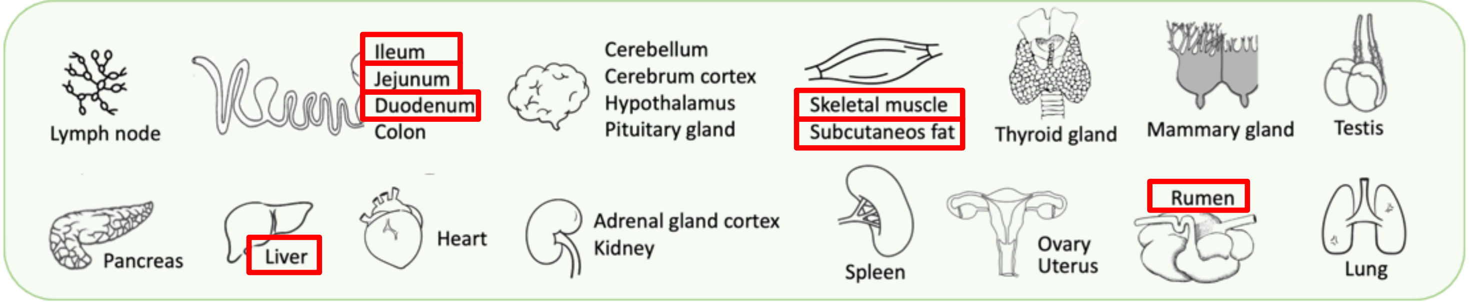


SNP effects from NLD data, prediction on CAN data

Scenario	Number of SNPs overlapped	Prediction accuracy
50K only	46K	0.00
QTL only	2,431	-0.02
eQTL only	11,505	0.00
50K + QTL	46K + 2,431	-0.02
50K + eQTL	46K + 11,505	0.00

Next: multi-breed NLD-CAN

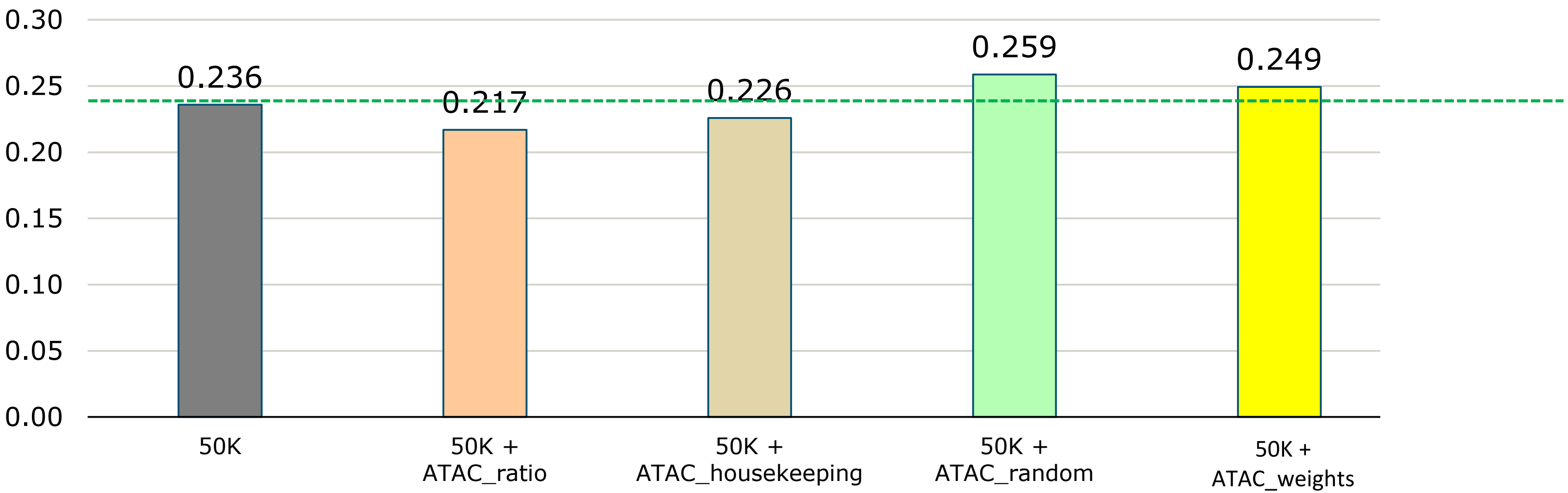
7. Use of narrow-peaks ATAC-seq



Different scenarios implemented using *ATAC-seq* narrow peaks for selected tissues

Scenario	Description	SNPs
ATAC_ratio	Top 10K variants from narrow peaks **ratio** (overlaps selected/ total tissues) **+ adjacent SNPs (200Kb)**	210,919 → 19,523 (LD pruning)
ATAC_housekeep	Top 10K variants that overlaps across all narrow peaks ("**housekeeping" set**)	10,000
ATAC_random	**Random** selection	3,318
ATAC_weights	**Higher weights on less frequent variants** (e.g., QTLs)	390

7. Narrow-peaks ATAC-seq results

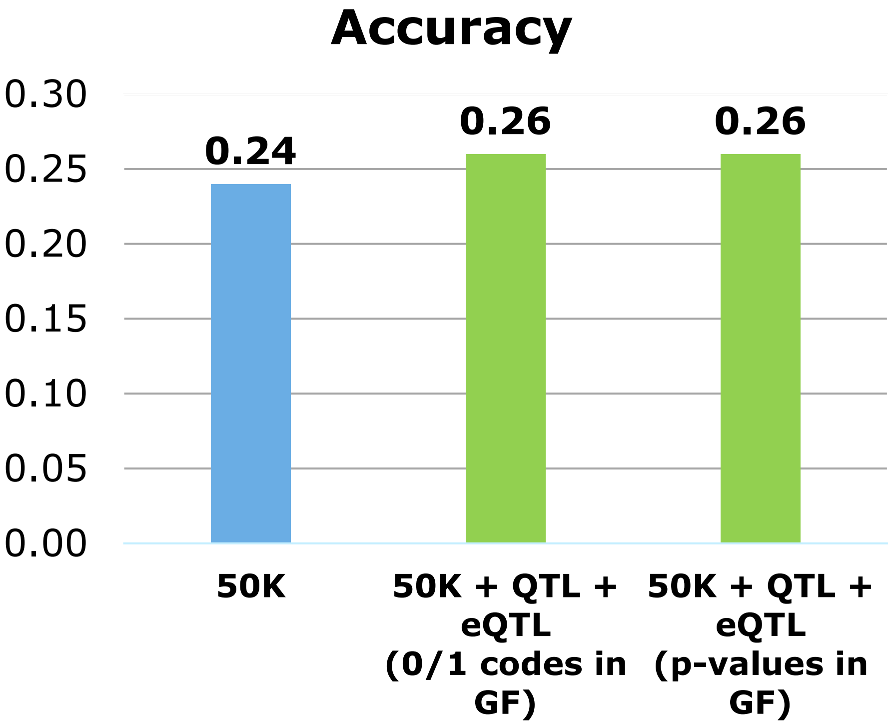


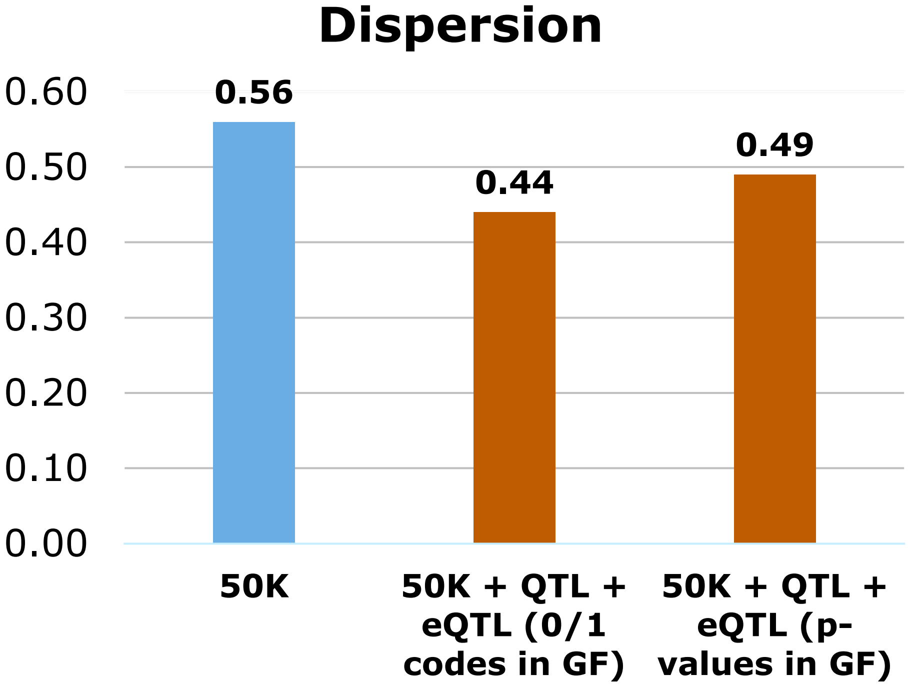
- SNPBLUP models
- Randomly selected variants gave same or higher accuracy than ATAC-seq scenarios (# SNPs)
- ATAC-seq modelled as additional SNP layer → model narrow-peaks as different layers into NextGP annotation matrix

WAGENINGEN UNIVERSITY & RESEARCH

UNIVERSITY OF ALBERTA

8. BayesLV (using p-value in GF layer)





WAGENINGEN UNIVERSITY & RESEARCH

UNIVERSITY OF ALBERTA

9. Conclusions

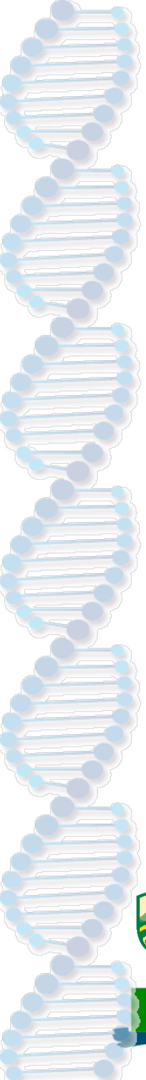
- Inclusion of **Genomic Features** could increase **genomic prediction accuracies** for **Dry Matter Intake**
- Results may vary across datasets
- No advantage using Bayesian (2 mixture model) over SNPBLUP for GF

Acknowledgments

The BovReg consortium

This work is carried out within the BovReg project, which has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 815668.

www.eurofaang.eu



BovReg *PARTNERS*

AARHUS UNIVERSITY
CRG Centre for Genomic Regulation
diagenode Innovating Epigenetics Solutions
edinethics
EMBL-EBI
ETH zürich
FBN
Faculdade de Medicina Veterinária Universidade de Lisboa
Instytut Genetyki i Hodowli Zwierząt PAN
INRAE
LIÈGE université GIGA Medical Genomics
Université de Limoges
Luke Luonnonvarakeskus
Martin-Luther-Universität Halle-Wittenberg
ROSLIN
University of Alberta
The University of Edinburgh
Universiteit Utrecht
Wageningen University & Research

Thank you for your attention

www.bovregproject.eu

University of Alberta
Wageningen University & Research

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 815668