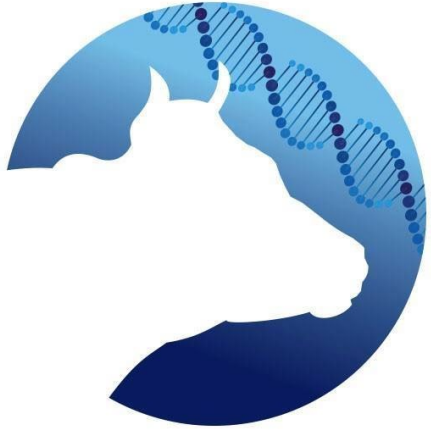
Biology-driven genomic predictions for dry matter intake within and across-breeds using WGS data

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



WAGENINGEN UNIVERSITY & RESEARCH

UNIVERSITY OF ALBERTA

BovReg Understanding cattle genomes

EuroFAANG

Outline

1. Background and project fit within BovReg
2. Biology-driven genomic predictions
3. Aim of the study
4. Data available, validation, BovReg genomic features
5. Selection and prioritization of genomic features
6. Scenarios and software
7. Within-breed results: NLD and CAN
8. Next steps and implications

1. Genomic predictions using WGS data

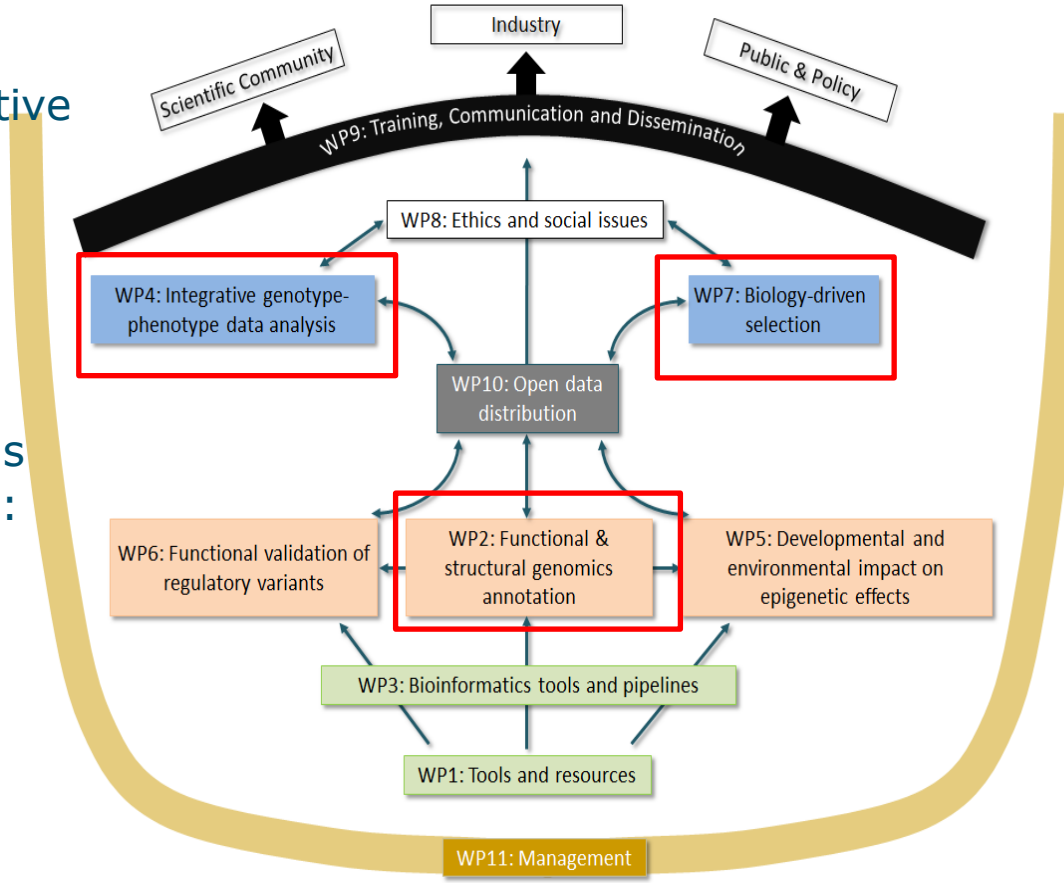
- Accuracy of genomic predictions (GP) → selection response (genomic selection efficiency)
- Large and representative reference populations (small breeds?)
- Whole-genome-sequence data (WGS) include causal variants with direct effects on the phenotypes
- GP using 50k → GP using WGS: no large benefits
- Pre-selecting SNPs from WGS → more accurate GP
- Using **genomic annotation** information could improve predictions across genetically lowly related individuals

1. The BovReg project

Catalogue of genomic annotation → BovReg: catalogue of functionally active **genomic features (GF)** in cattle

Work Package 7, biology-driven selection:

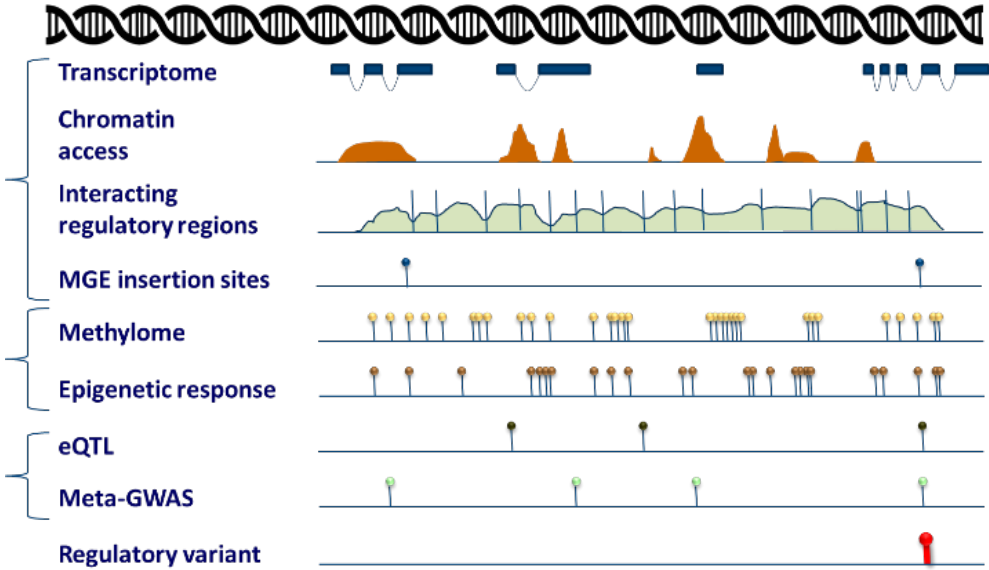
1. Development of novel GP methods integrating functional information: Genomic Features models
2. **Validate models in real data**, using results from WP2 and WP4



WAGENINGEN
UNIVERSITY & RESEARCH

2. Biology-driven genomic predictions

- WGS: millions of variants → pinpoint **causal variants** affecting traits of interest
- Key traits: **biological efficiency**, e.g. **feed efficiency**, small ref. populations
- **GF catalogue** of **"multi-omics"** data: genomics (QTL), transcriptomics (eQTL), epigenomics (mQTL), chromatin accessibility (ATAC-seq/ChIP-seq), …
- **Functional GF** → SNP prioritization → **Biology-driven genomic predictions**



WAGENINGEN
UNIVERSITY & RESEARCH

3. Aim of the study

Validation of within- and across-breed

biology-driven genomic predictions using **genomic features**

for **dry matter intake (feed efficiency)**

WAGENINGEN
UNIVERSITY & RESEARCH

UNIVERSITY OF
ALBERTA

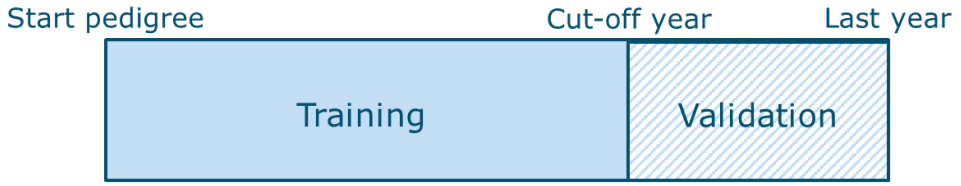
4. Data available for genomic predictions

- **Netherlands (NL)**
- Dairy cattle: Holstein-Friesian
- 6 research farms
- EBVs from routine DMI evaluations for lactation 1, 2, 3+
- Average de-regressed proofs (DRP) as pseudo-phenotypes
- 3,628 cows with genotypes (50K) and DRPs
- Genotypes at 50K SNPs → imputed to HD → imputed to WGS (*Beagle*)

4. Data available for genomic predictions

- **Canada (CAN)**
- Composite crossbred beef cattle
- 14 research farms
- Individual DMI (one observation) → pre-corrected phenotypes (Y^*) for fixed effects & breed composition
- 5,584 individuals (males and females) with genotypes and Y^*
- Genotypes at 50K SNPs → imputed to HD → imputed to WGS (*Beagle*)

4. Validation approach



- Different possible approaches
- Mimic breeding program application → **forward-in-time** cross-validation
- Exclude pedigree relationships >0.5 in validation set (e.g. parent-offspring)
- **Accuracy**: correlation between dependent variable and GEBVs
- **Dispersion**: regression coefficient of dependent variable on GEBVs (expectation in absence of dispersion = 1)
- Standard errors via bootstrapping with replacement (R)

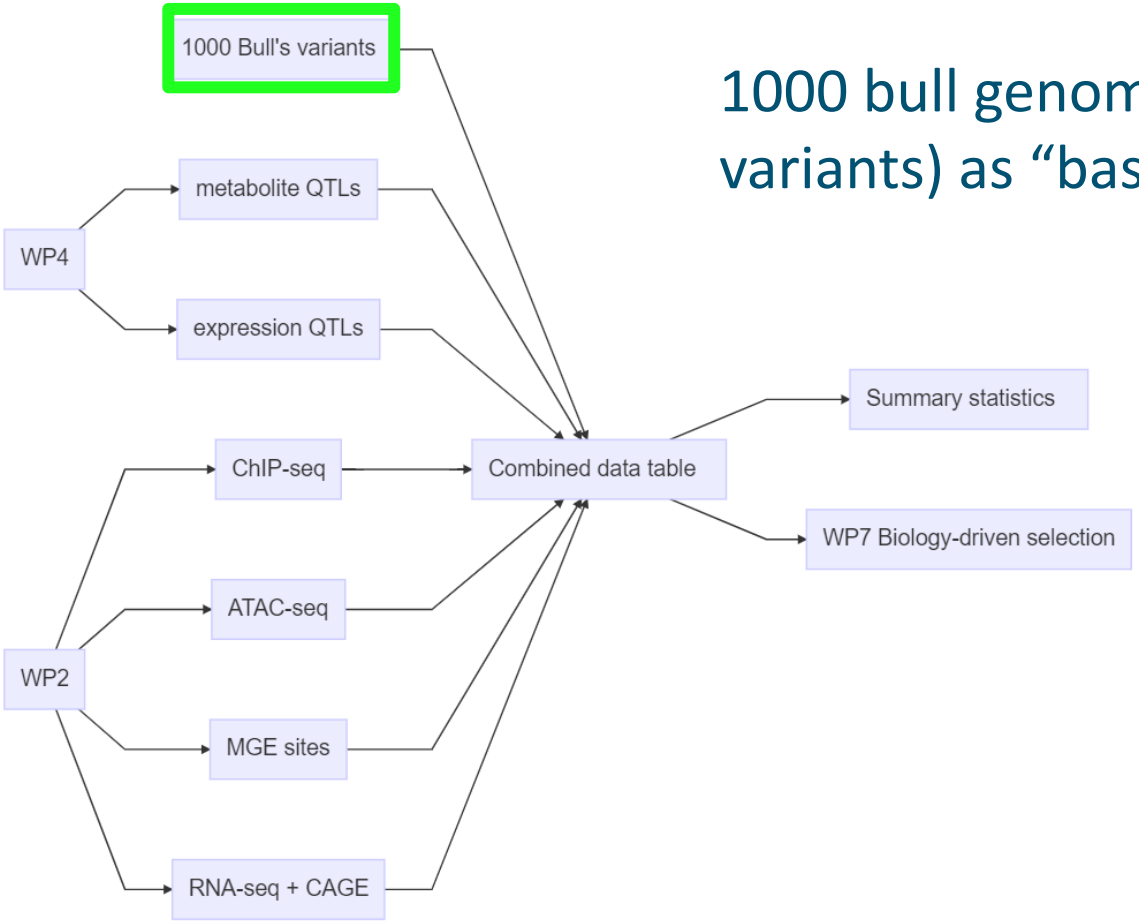
	NLD	CAN
Dependent variable	DRP	Pre-corrected DMI
Cut-off year	2018	2013
# animals training	2,769	4,091
# animals validation	859	1,493

5. Genomic features available

GF	Work Package
Meta-GWAS QTL (QTL)	WP4
Expression QTL (eQTL)	WP4
Metabolic QTL (mQTL)	WP4
ATAC-seq	WP2
ChIP-seq	WP2
Mobile Genetic Elements (MGE)	WP2

- Central database
- Different tissues/samples → Different formats for different GF
- Many files (~1,000) with large size
- How to quick access/query GF information ?

5. Selection of genomic features



1000 bull genome project (~74.9 M variants) as “baseline” to map GF

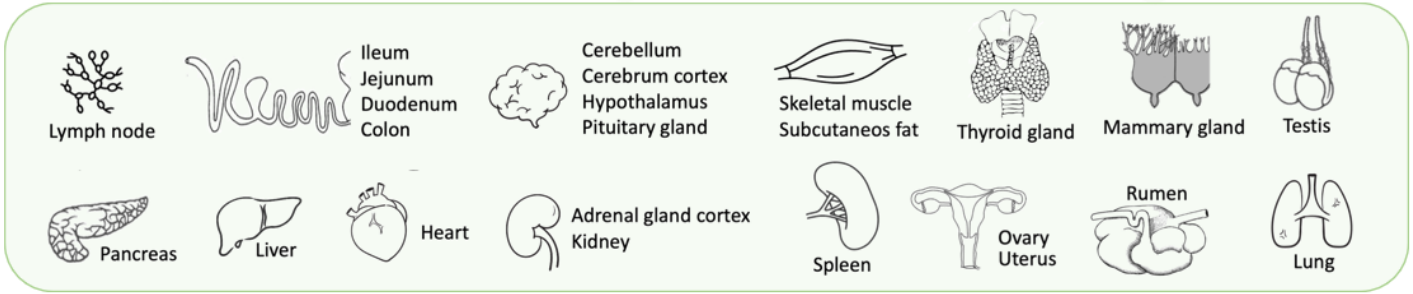
5. Selection of genomic features

Variant ID	CHR	POS	GF1 (DMI QTL)	GF2 (liver eQTL)	GF3 (ATAC tissue1)	GF .	GF*n*
Variant 1	1	659	0	1	0	.	$1.1x10^{-13}$
Variant 2	1	1235	1	0	1	.	NA
Variant 3	1	1578	0	0	1	.	NA
Variant 4	1	1984	0	0	1	.	NA
Variant 5	1	2689	0	0	1	.	$3.9x10^{-5}$
Variant 6	1	3458	0	1	0	.	NA
Variant …	.	.	.	.	.	.	.
Variant 75M	29	49805987	0	0	1	.	NA

For each GF, information store:

- Position/Intervals as (0/1 coding)
- *p*-value (if available) as (NA/value)

5. Genomic features used

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

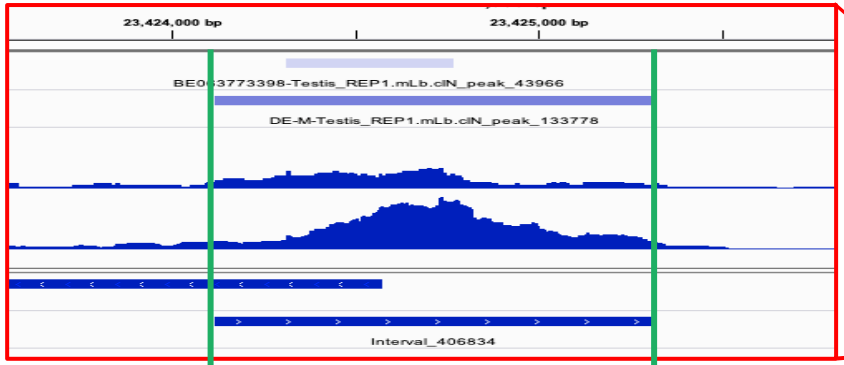
From GC Moreira et al., 2023 EAAP

(intermezzo) ATAC-seq

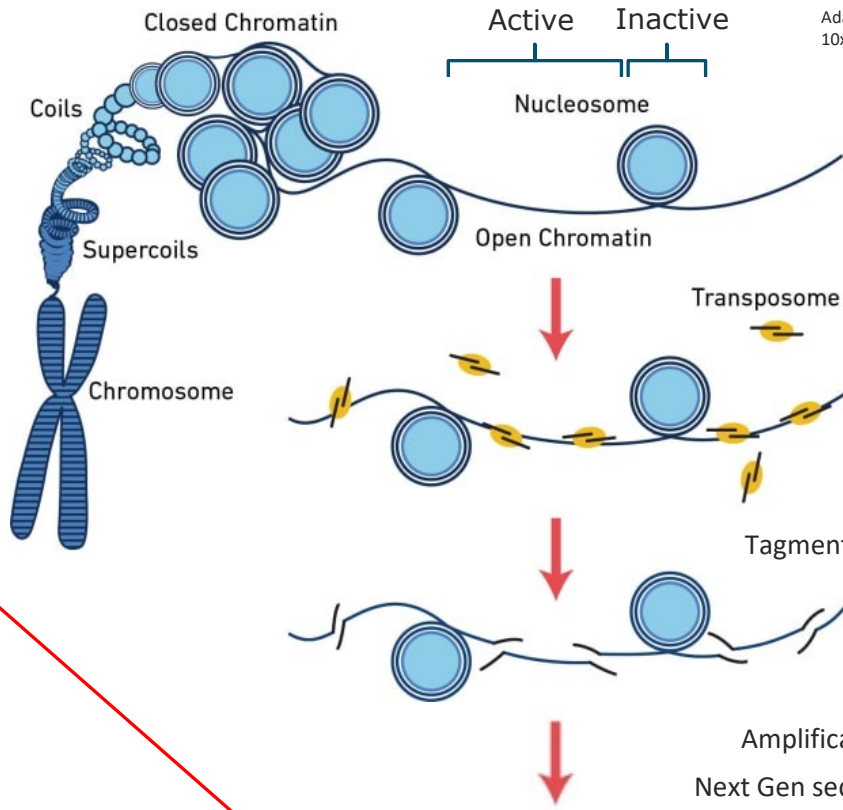
Genome-wide mapping of **chromatin accessibility → accessible DNA**

⇩

Promotor regions
Enhancers
Regulatory elements



Consensus peaks across samples



Adapted from 10xgenomics.com

GC Moreira et al. 2023, EAAP

nf-core/atacseq

5. Selection of genomic features

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

Select GF group
(across traits/tissues)

Awk/R

List of variants per GF "layer"
(overlapping with 1000G)

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

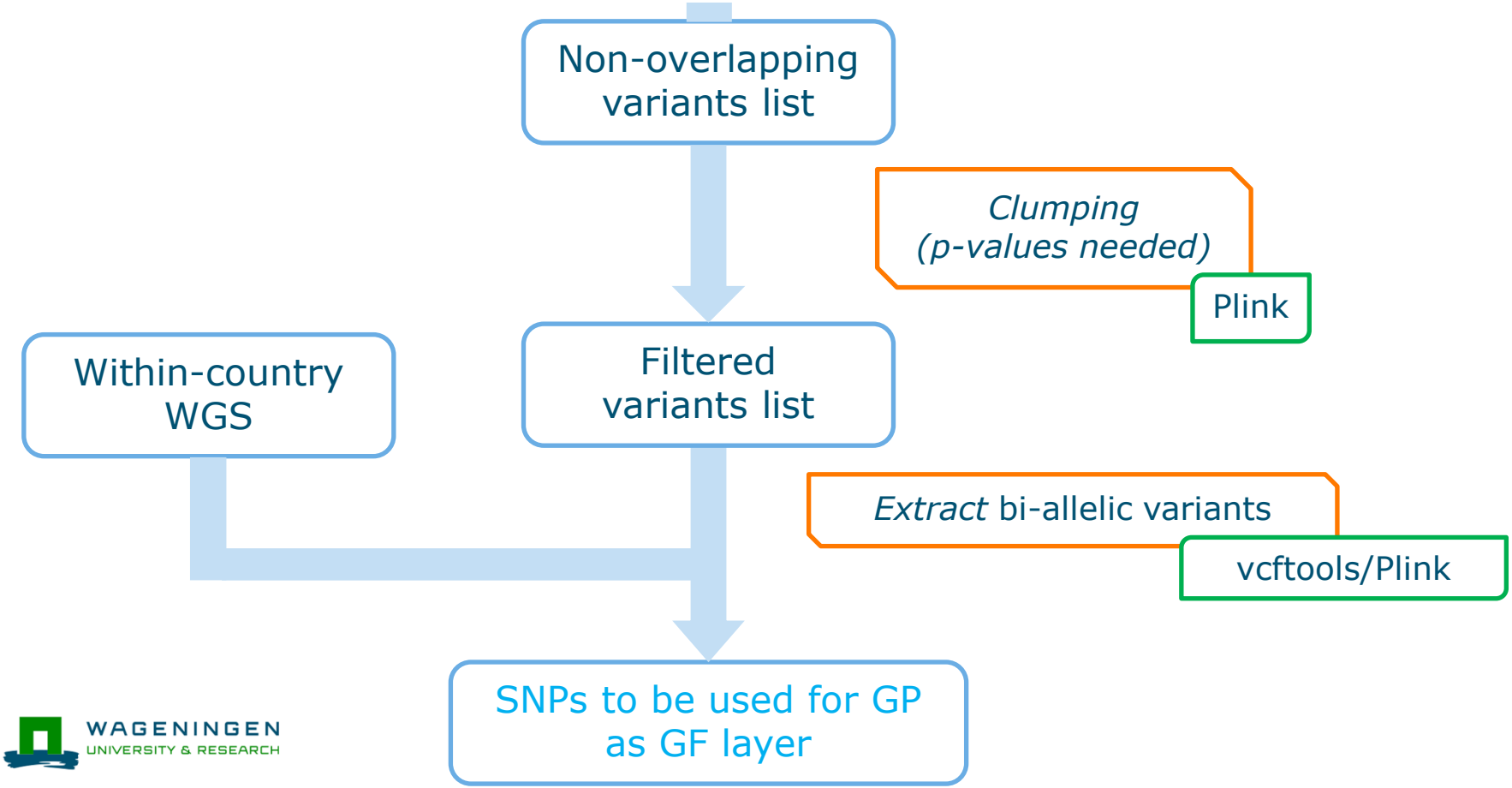
Awk/R

Non-overlapping
variants list

Variant ID	CHR	POS	GF1 (DMI QTL)	GF2 (milk QTL)	GF3 (ATAC tissue1)	GF .	GF*n*
Variant 1	1	659	0	1	0	.	$1.1x10^{-13}$
Variant 2	1	1235	1	0	1	.	NA
Variant 3	1	1578	0	0	1	.	NA
Variant 4	1	1984	0	0	1	.	NA
Variant 5	1	2689	0	0	1	.	$3.9x10^{-5}$
Variant 6	1	3458	0	1	0	.	NA
Variant …	.	.	.	.	.	.	.
Variant 75M	29	49805987	0	0	1	.	NA

WAGENINGEN
UNIVERSITY & RESEARCH

5. Selection of genomic features

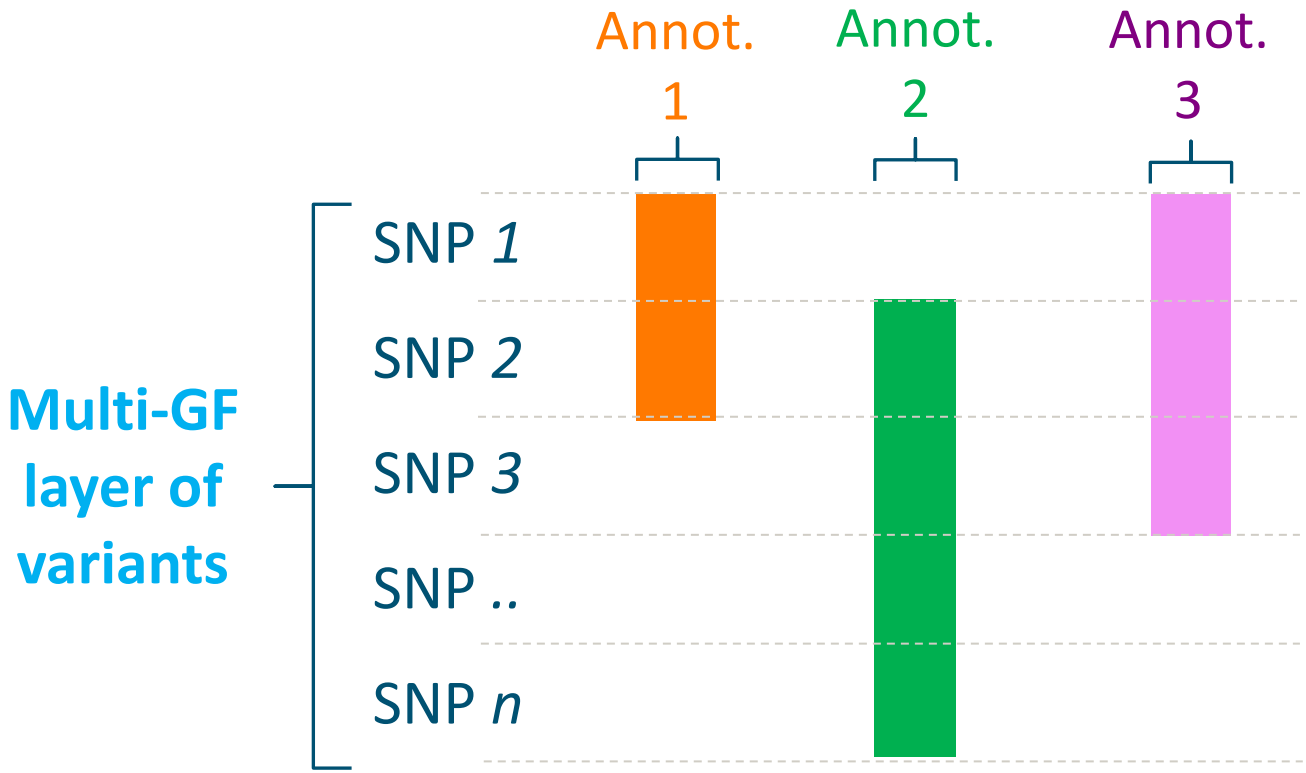


6. Scenarios

Scenario	NLD 	CAN 
50K	48K	46K

- GF as additional layer
- QTL, eQTL, ATAC → Multi-GF → overlapping GF

6. Overlapping genomic features



6. Overlapping genomic features

Multi-GF

QTL	eQTL	ATAC	Variants	%
1	1	0	5	0
1	1	1	16	0
1	0	0	2,576	14
1	0	1	2,819	16
0	1	0	5,051	28
0	1	1	7,329	41
			17,796	100

BayesRCπ (2 classes)

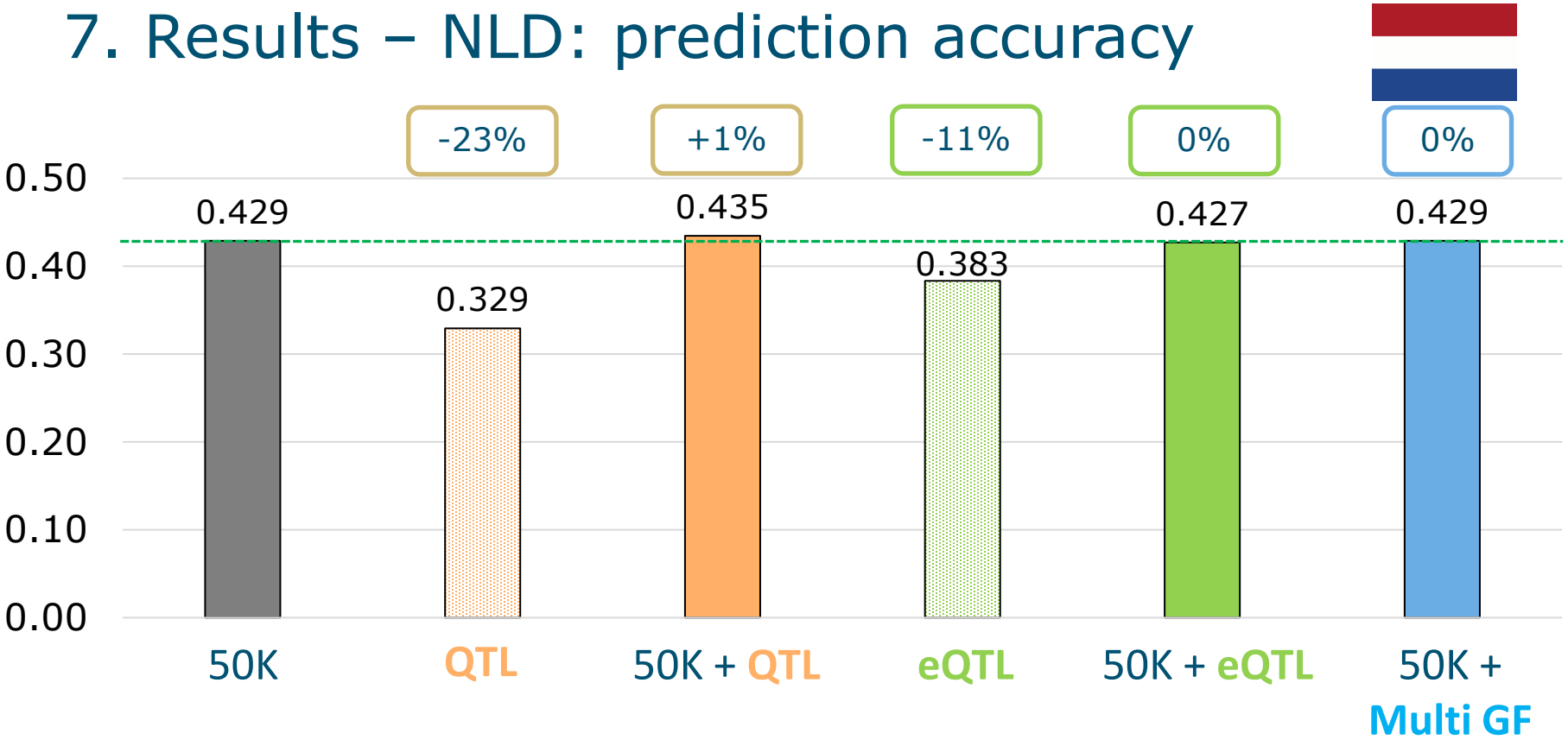
Mollandin *et al.* 2022, BMC bioinformatics

6. Software and settings

- *NextGP.jl* (https://github.com/datasciencetoolkit/NextGP.jl)
- Base 50K: SNPBLUP (BayesC_0), common variance across SNPs
- Additional GF layer (QTL, eQTL, ATAC): SNPBLUP or Bayesian (2 mixture model)
- SNP variance prior: sum(2*pq*) from training genotypes
- No advantage of Bayesian 2 mixture model

Scenario (NLD)	**SNPs**	**Running time**
50k panel, as SNPBLUP	48K	3.98 hours
50k panel, as 2 mix model (pi estim.)	48K	4.96 hours
50k panel, as 2 mix model (pi fixed)	48K	6.41 hours
50k panel + eQTL, both as SNPBLUP	48K + 12K	6.38 hours

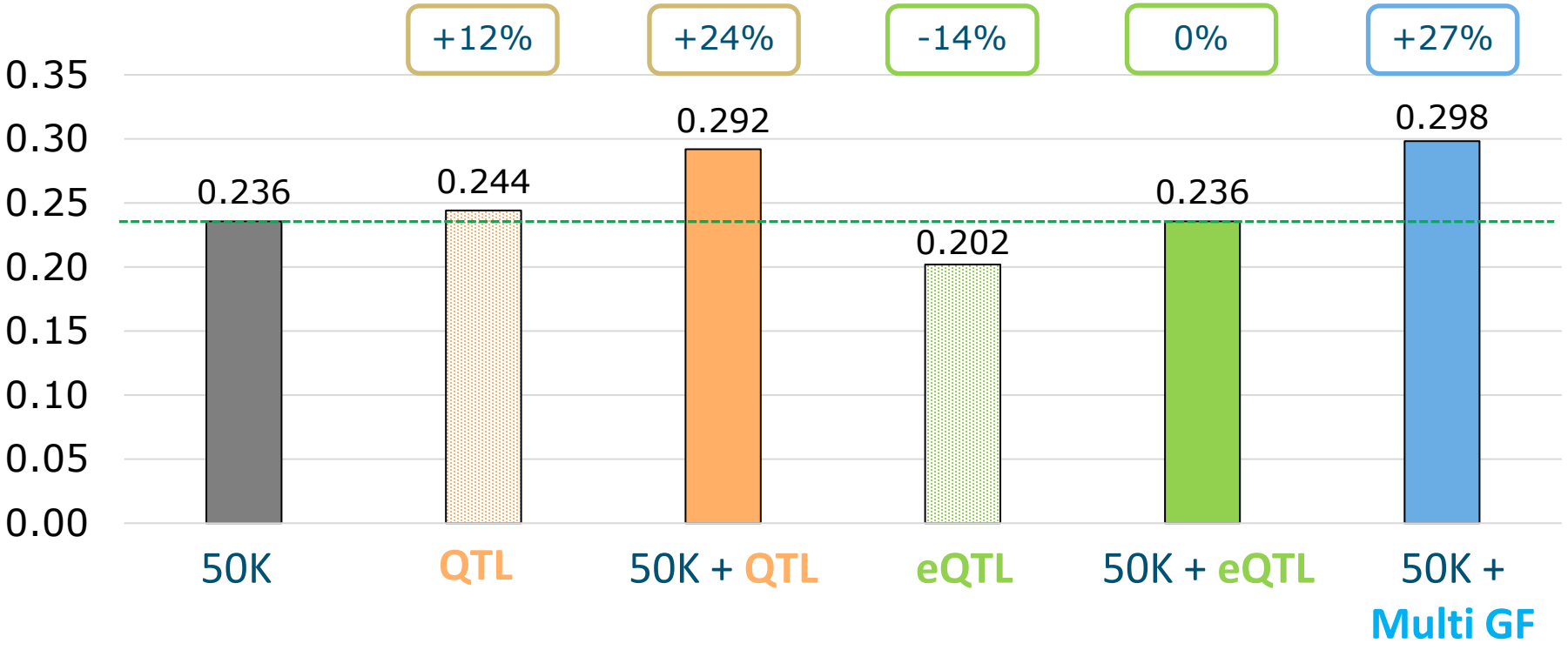
7. Results – NLD: prediction accuracy



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

WAGENINGEN UNIVERSITY & RESEARCH

7. Results – CAN: prediction accuracy



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

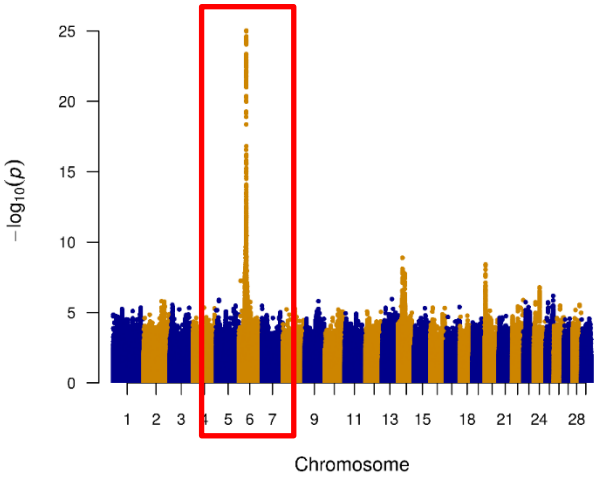
WAGENINGEN UNIVERSITY & RESEARCH

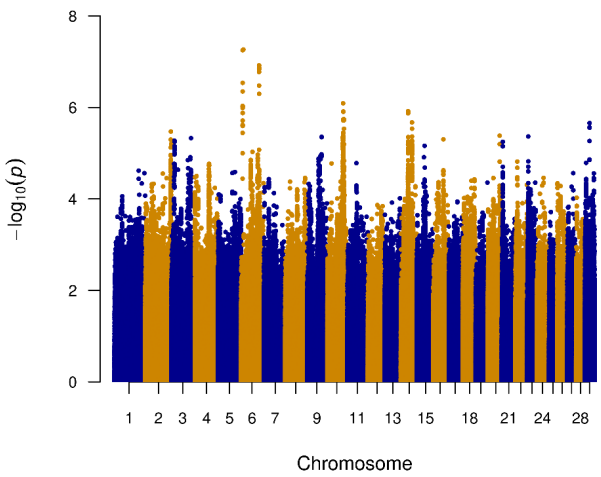
7. Manhattan plots meta-GWAS QTL for DMI

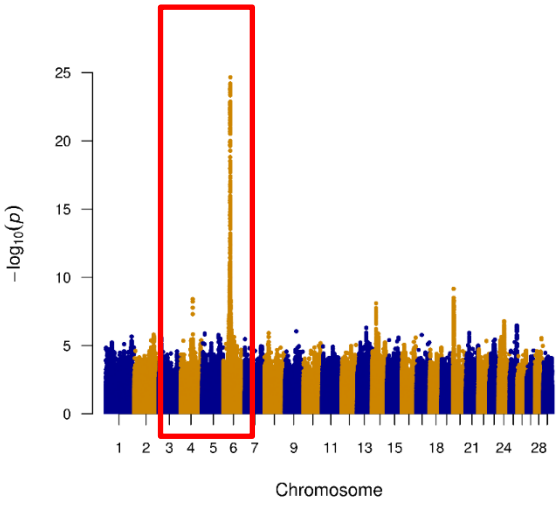
ALL
10,539 animals
Beef, Holstein, Finnish Red
MAF >0.005
30,216,688 variants

HOL
2,368 animals
Only Holstein
MAF >0.005
19,647,876 variants

BEEF
7,805 animals
Only Beef
MAF >0.005
27,839,929 variants







Zhang *et a*l. 2020, BMC genomics

WAGENINGEN
UNIVERSITY & RESEARCH

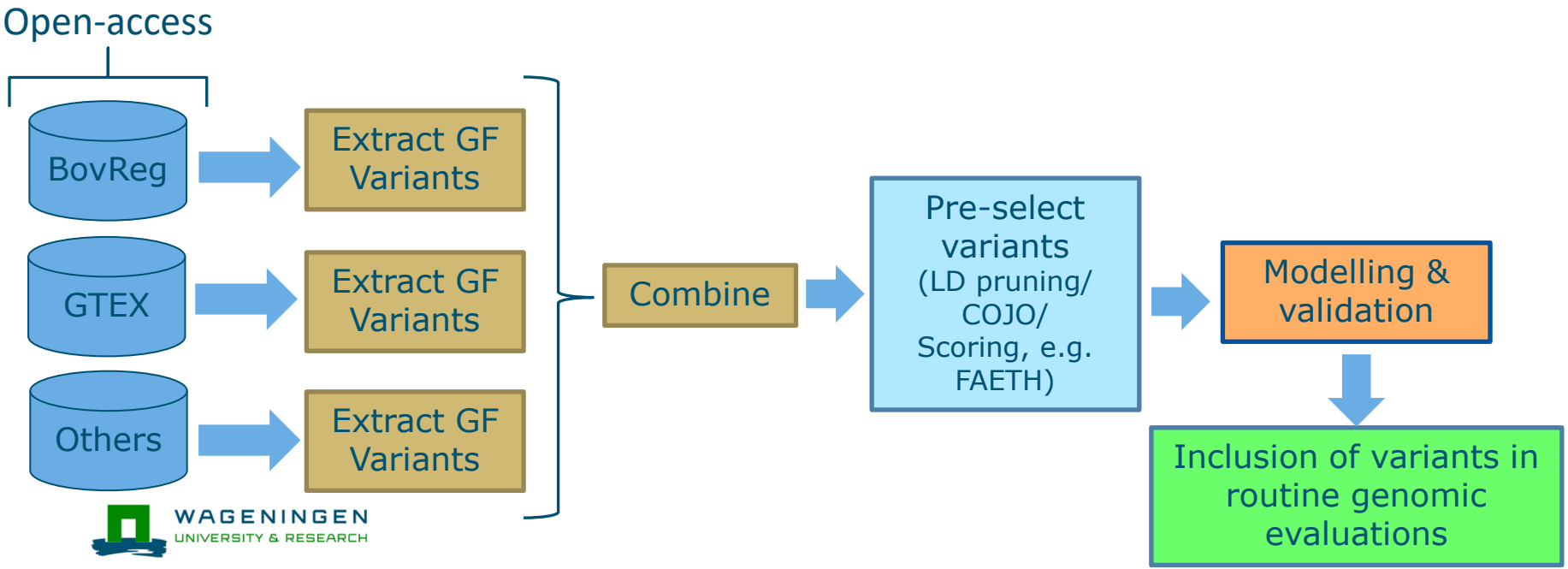
From Gredler-Grandl *et al.* 2022, WCGALP

8. Next steps

- Checking on variants with larger effects on CAN DMI predictions → match with meta-GWAS QTLs?
- GP focussing on trait/tissues-specific variants (now across tissues/traits)
- More detailed annotation modelling using BayesLV (considering *p*-values across annotation layers)
- Across-/multi-breed NLD-CAN

8. Implications

- How can this be used in (routine) genomic predictions?
- CAN case → include QTL variants as markers in 50k chips



Conclusions

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

Acknowledgments

Dr Emre Karaman and Dr Luc Janss

Dr Gabriel Costa M Moreira

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.

EuroFAANG

www.eurofaang.eu

EuroFAANG
RESEARCH INFRASTRUCTURE

AQUA-FAANG

BovReg
Understanding cattle genomes

GENE-SWitCH

GERONIMO
GENOME AND EPIGENOME ENABLED BREEDING IN MONOGASTRICS

Rumigen
TOWARDS IMPROVEMENT OF RUMINANT BREEDING THROUGH GENOMIC AND EPIGENOMIC APPROACHES

HoloRuminant
Understanding microbiomes of the ruminant holobiont

WAGENINGEN
UNIVERSITY & RESEARCH

UNIVERSITY OF ALBERTA

Thanks for your attention

renzo.bonifazi@wur.nl @RenzoBonifazi

FAANG
Functional Annotation of Animal Genomes
10 Years

Disclaimer: Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Take-home messages

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

renzo.bonifazi@wur.nl @RenzoBonifazi

Disclaimer: Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.