GENE-SWitCH

The regulatory GENomE of SWine and CHicken:
functional annotation during development

Dynamic inclusion of functional genome annotations to improve accuracy of genomic prediction in pigs.

Derek Bickhart, Bruno Perez ... Mario Calus

@GeneSwitch - www.gene-switch.eu

WAGENINGEN
UNIVERSITY & RESEARCH

HENDRIX GENETICS

Biological knowledge in genomic predictions

Objective: Leverage functional annotation (FA) information to augment genomic prediction models.

- There are **numerous sources** of FA information…
  - Results using only **one source** of FA are often highly **trait-dependent**.

- **Integrating** the **many layers** of functional information available seem to be the way forward.

GENE-SWitCH

How can we do that?

FAETH score

- **F**unctional and **E**volutionary **T**rait **H**eritability

Ranks **SNP variants** according to how much **genetic variance** it explains across **phenotypes** and **functional annotation** layers
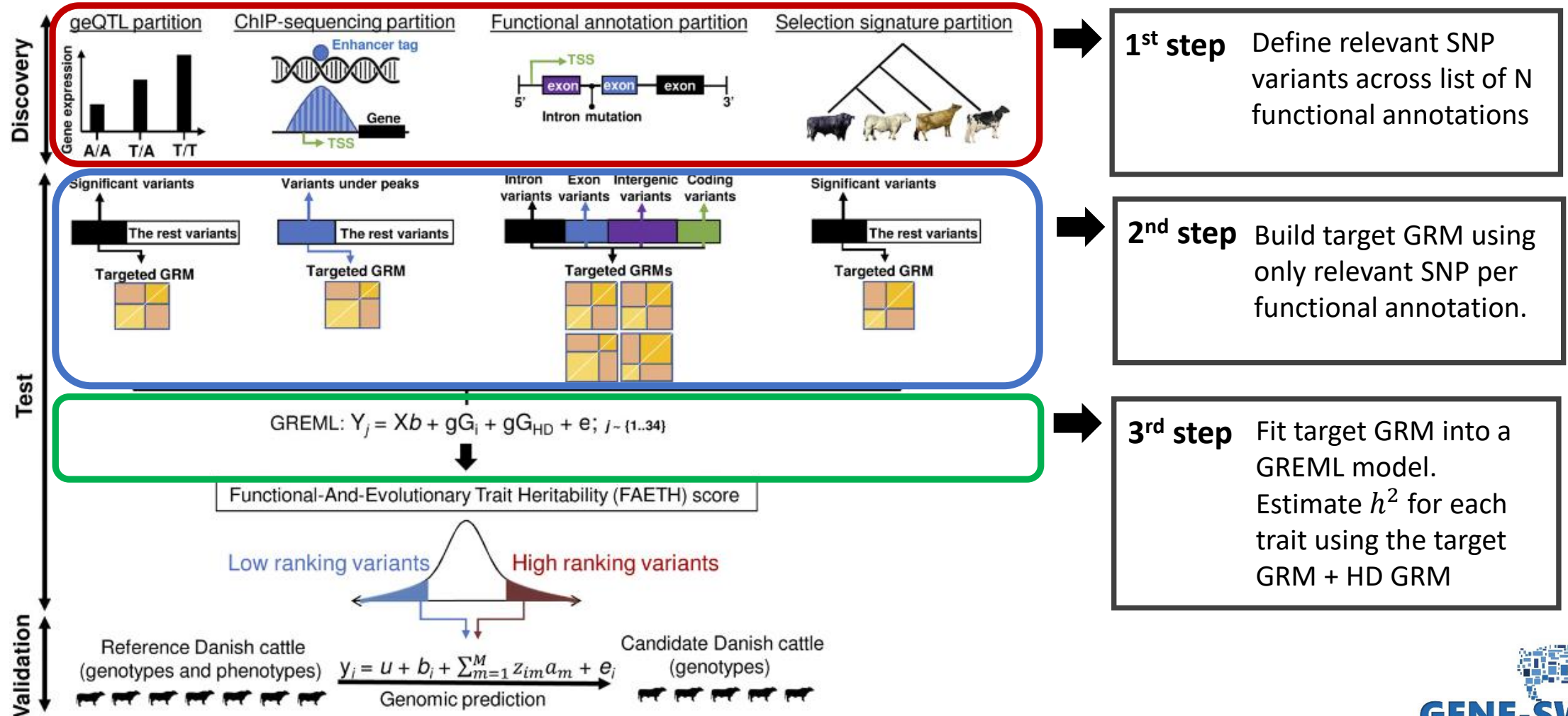
Calculating FAETH scores

1ˢᵗ step   Define relevant SNP variants across list of N functional annotations

2ⁿᵈ step   Build target GRM using only relevant SNP per functional annotation.

3ʳᵈ step   Fit target GRM into a GREML model. Estimate h^2 for each trait using the target GRM + HD GRM

FAETH score pipeline

Data description

- **100 Landrace with WGS data available**
  - Approximately 25K imputed WGS = 15M SNP

- **6 traits recorded on our breeding population**
  - Landrace
  - Ranging from 6K to 22K records per trait

Trait	Number of records		h^2
	Reference	Validation	
TRT_1	19,900	1,161	0.38
TRT_2	20,302	1,207	0.24
TRT_3	6,016	780	0.14
TRT_4	22,442	1,207	0.10
TRT_5	19,809	1,162	0.25
TRT_6	15,612	773	0.21

- **Functional annotation data**
  - GENESWitCH
  - pigGTEx (publicly available)
  - Other publicly available data sources…

  Currently in total = **32 functional annotation layers**

GENE-SWitCH

Functional annotation layers
Maps from consortium data

PigGTEx

GENE-SWitCH

| Functional Annotation	Description	Source	Targeted variants sets (no. of variants)	No. of variants
eQTL x3brd	eQTL analysis (3 breeds)	Gene-Switch (IRTA)	eQTL with an adjusted p-value < 0.05	43,059
DMR_GNSW	Differentially methylated regions	Gene-Switch (WUR)	Variants with LogFoldChange >= 1 or LogFoldChange <= -1	10,337
AtacSeq	Chromatin accessibility	Gene-Switch (IRAE)	Peaks of <500Kb and with variance across development	16,326
ChipSeq	H3K4me1, H3K4me3, H3K27ac, H3K27me3	Gene-Switch (WUR)	Peaks with adjusted p-value < 0.05 (Muscle, Kidney, Liver)	-
lncQTL	Long non-coding RNA based associations	Pig GTEx	lncQTLs with an adjusted p-value < 0.05 across all tissues	12,166
sQTL	Splicing gene-based associations	Pig GTEx	sQTLs with an adjusted p-value < 0.05 across all tissues	20,077
eQTL	Protein-coding gene associations	Pig GTEx	eQTLs with an adjusted p-value < 0.05 across all tissues	48,571
enQTL	Enhancer-based associations	Pig GTEx	enQTLs with an adjusted p-value < 0.05 across all tissues	48,652
bieQTL	Breed-interaction gene associations (LN)	Pig GTEx	bieQTLs with an adjusted p-value < 0.05 across all tissues	10,940
eeQTL	Exon-based associations	Pig GTEx	eeQTLs with an adjusted p-value < 0.05 across all tissues	56,535

GENE-SWitCH

Functional annotation layers
Maps from other sources

Functional Annotation	Description	Source	Targeted variants sets (no. of variants)	No. of variants
UpDownRegulated	Differentially expressed genes: Inter Muscular fat experiment	Xu *et al.* 2022	Regions with logFold change > 1.2	3,227
DiffMethy_CpG_MD	Differentially methylated sites Muscle depth experiment	Yang *et al.* (2021)	Variants placed up to 100bp around differentially methylated sites with an adjusted p-value < 0,001	12,902
ActiveTSS	Annotated chromatin states	Pan *et al.* (2021)	Variant present in 5 tissues or more	139,392
Bivalent/Poised TSS	Annotated chromatin states	Pan *et al.* (2021)	Variant present in 5 tissues or more	10,310
Zinc fingers	Annotated chromatin states	Pan *et al.* (2021)	Variant present in 5 tissues or more	1,359,251
Strong transcription	Annotated chromatin states	Pan *et al.* (2021)	Variant present in 5 tissues or more	211,077
Genic enhancers	Annotated chromatin states	Pan *et al.* (2021)	Variant present in 5 tissues or more	729,651
Enhancers	Annotated chromatin states	Pan *et al.* (2021)	Variant present in 5 tissues or more	1,035,468
Histone Modifications	Histone modifications across embryo development (50 to 95 days)	Han *et al.* (2019)	Regions (+/- 2Kb) with a logFold change >=2	148,319
Conserved Sites 100	Across 100 species	NCBI	Sites conserved across 100 species (lifted from the Human genome)	227,447
Ataq-Seq D90	Open chromatin profiles in muscle	Salavati *et al.* (2021)	Peaks up to 1000Kb found in small, normal and big piglets	66,808
SE and BD	Super Enhancers (SE) and Broad Domains (BD)	Peng *et al.* (2021)	Pig-specific H3K27ac (SE) and H3K4me3 (BD) enriched peaks	189,136

8

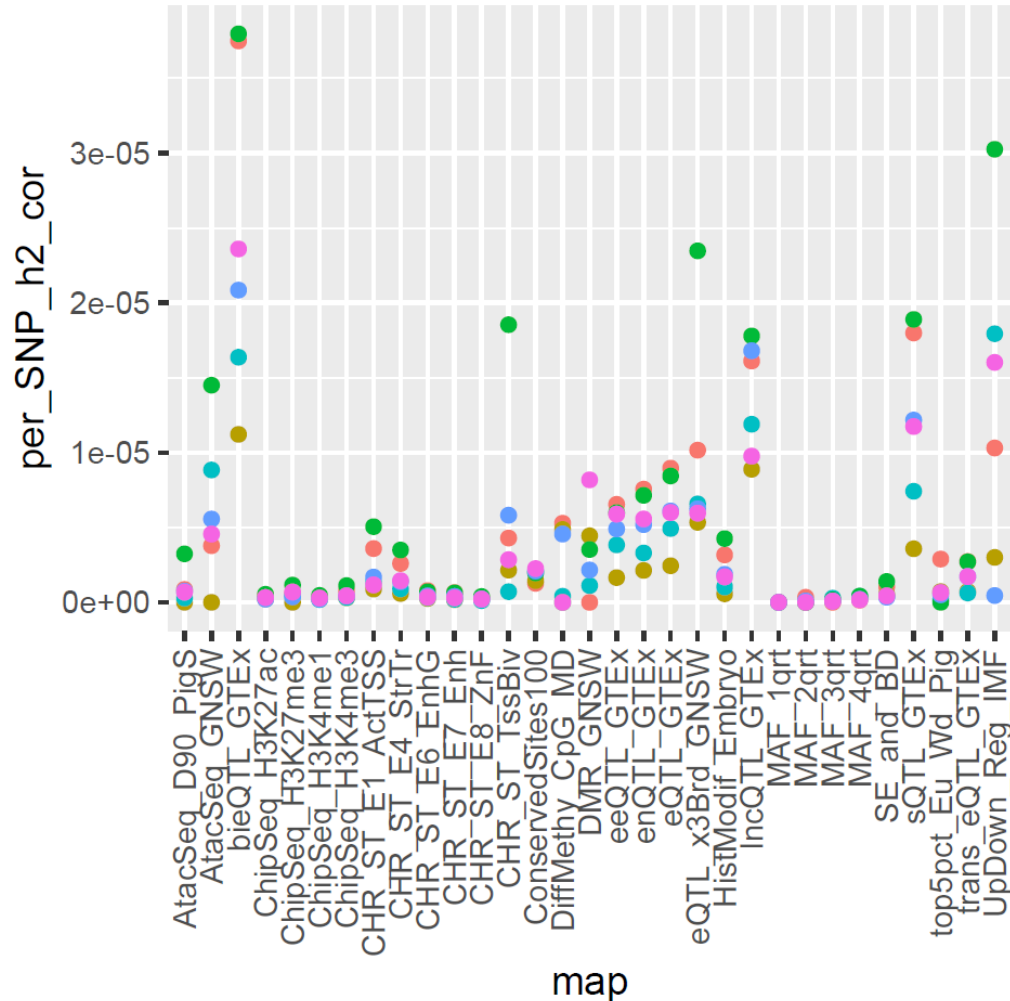GENE-SWiTCH

SNP-based h^2

1. Molecular QTL score highly
   a) bieQTL, lncQTL, sQTL…

2. Big SD indicate trait specificity

3. Similar FA maps grouped together

4. Conserved sites are trait independent (low SD)
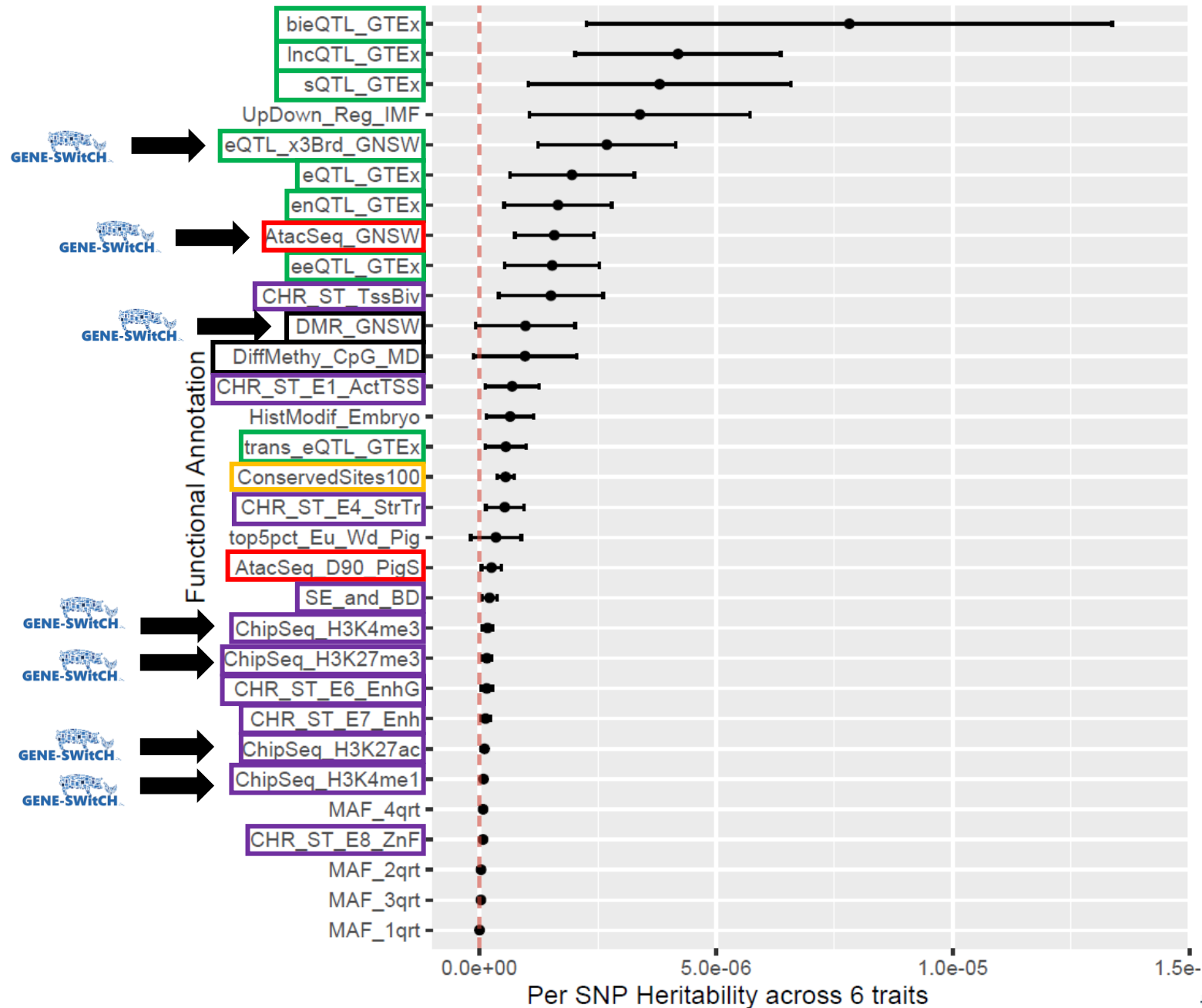
32 functional annotation maps

Per trait SNP-based h^2 (within FA layers)
Corrected for trait heritability

Correction: $\dfrac{SNP\_based\_h^2}{trait\_h^2}$

- Functional/regulatory variants seem to capture higher % of genetic variance for **Trait 3**

Both functional **annotation map** and **phenotype** sets will influence results.

GENE-SWitCH

Validation of FAETH score variants for genomic prediction

Compare genetic variance captured and predictive accuracy between...

- **High** vs **Low** FAETH score variants
- **High** FAETH scores vs **randomly** sampled variants

- Is the FAETH score able to indicate informative SNP?

- Does it outperform a random placement of SNP in GP?

GENE-SWitCH

Heritability

At 1 million SNP level:

- **Low FAETH** always captures less genetic variance
  - From 10-30% less

- **Random** and **High FAETH** seem to capture the same amount of variance.

Trait	Heritability			Difference
	LOW_1M	**RANDOM_1M**	**HIGH_1M**	**LOW x HIGH**
TRT_1	0.32	0.38	0.37	14%
TRT_2	0.18	0.24	0.24	24%
TRT_3	0.10	0.14	0.14	30%
TRT_4	0.07	0.10	0.10	31%
TRT_5	0.23	0.25	0.25	9%
TRT_6	0.17	0.21	0.21	18%

[1] standard error = ±0.01

GENE-SWitCH

Predictive accuracies

At 1 million SNP level:

- **Low FAETH** yield the lowest predictive accuracies
  - Approx. -15% relative to Random

- **High FAETH** yield similar or higher predictive accuracies than Random.
  - From 0.0 – 4.0%

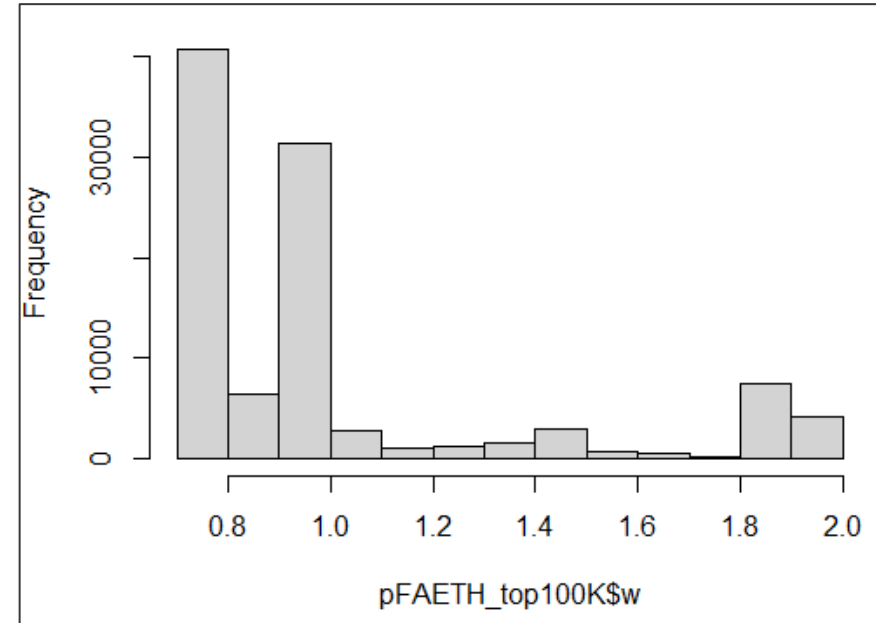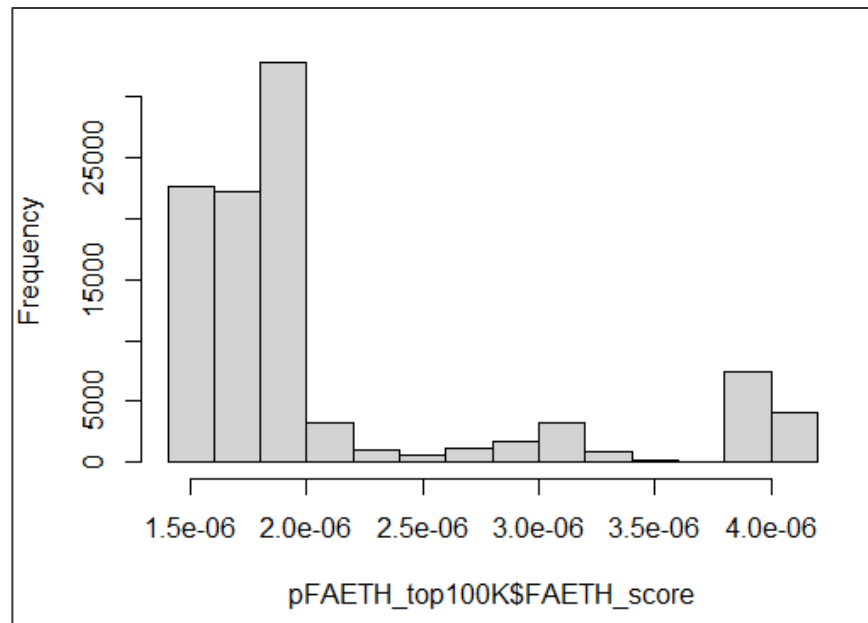- **High FAETH** never seem to deteriorate predictive accuracies

* standard errors calculated by bootstrapping

GENE-SWitCH
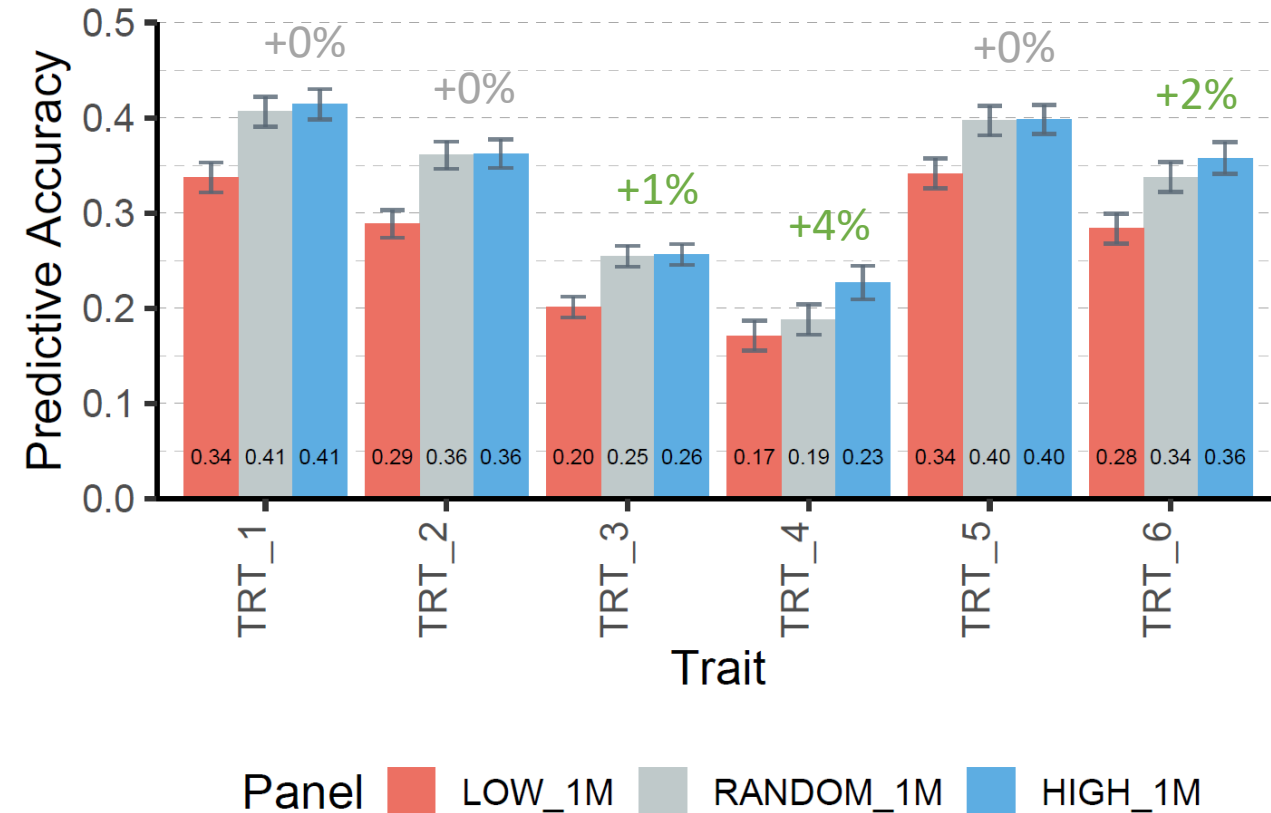
"Informing" the model a bit more

From FAETH score to SNP weights:

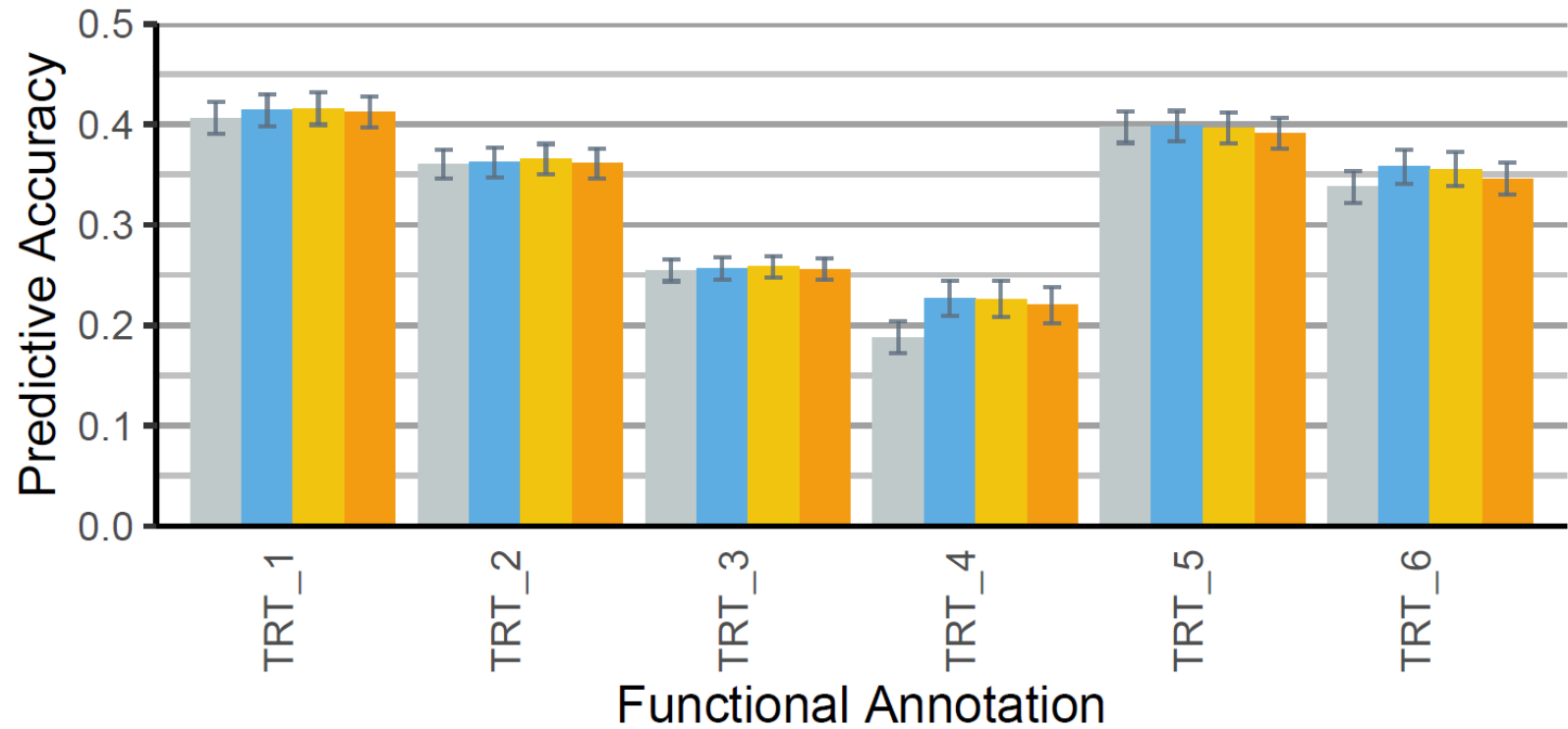- $wFAETH_{SNP_i} = {FAETH_{SNP_i}}\Big/{\sum FAETH} * nSNP$
- Mean = 1

Top 100K   **FAETH score** ⟶ **wFAETH**

Predictive accuracy - wGBLUP
using FAETH scores as SNP weights

* standard errors calculated by bootstrapping

Take home messages

- **Pipeline developed for calculating FAETH scores**
  - General pipeline can be applied to any species

- **Publicly available** and/or **proprietary data** can be used

- **Molecular QTL** maps seem to be the **most informative**

  - Also, **more trait-specific** than other maps

GENE-SWitCH

Take home messages

- **FAETH scores can help to improve predictive accuracies in pig breeding**

- **Using FAETH as SNP weights did not result on increase in accuracies**
  - Possible limitations? (imputation accuracy for example)
  - Imputing big blocks may cause loss in resolution for Discovery / Prediction

GENE-SWitCH

Next steps

- **Remove commercial 50K SNP from "*Discovery*"**
  - Avoid inflation of predictive accuracy due to overlap between target_SNP and the commercial 50K SNP
  - SNP in the commercial panel are **not imputed** (actual genotypes)

- **Manuscript**
  - Publication is planned.
  - Start draft: report on FAETH scores results (*presented here*)

GENE-SWitCH

Acknowledgements: