

# ➤ CAPITALIZING ON COMPLEX ANNOTATIONS IN BAYESIAN GENOMIC PREDICTION FOR A BACKCROSS POPULATION OF GROWING PIGS

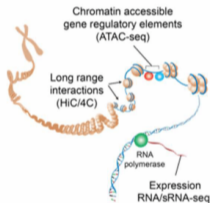
F. Mollandin, P. Croiseau, H. Gilbert, A. Rau

## >CONTEXT

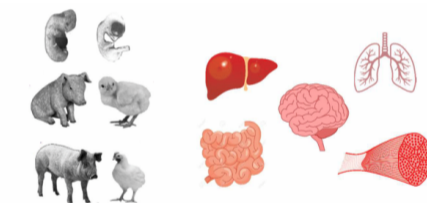
- ▶ Genomic selection is widely used in the breeding community, focusing on efficient evaluation methods

## CONTEXT

- ▶ Genomic selection is widely used in the breeding community, focusing on efficient evaluation methods
- ▶ Biotechnology is improving, providing more and more **heterogeneous data**, from different independent studies, different type of OMICS, etc...

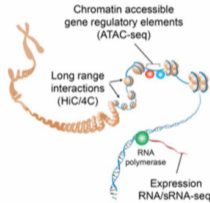


(Image: <http://www.fragencode.org>)



## >CONTEXT

- ▶ Genomic selection is widely used in the breeding community, focusing on efficient evaluation methods
- ▶ Biotechnology is improving, providing more and more **heterogeneous data**, from different independent studies, different type of OMICS, etc...



(Image: <http://www.fragencode.org>)



- ▶ We would like to more fully exploit the available information to **predict complex traits**, and better represent the underlying **genomic architecture**.

## ➤ GENOMIC PREDICTION : EXISTING MODELS

$$\mathbf{y} = \mu \mathbf{1}_n + \mathbf{X}\boldsymbol{\beta} + \mathbf{e},$$
$$\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}_n \sigma_e^2)$$

Method	SNP effect prior distribution	Prior Annotations classes	# A(i)
BayesR <sup>1</sup>	$\beta_i \sim \sum_{\ell=1}^4 \pi_{\ell} \mathcal{N}(0, k\sigma_g^2)$	No	-
BayesRC <sup>2</sup>	$\beta_i   A(i) \sim \sum_{\ell=1}^4 \pi_{\ell, \alpha} \mathcal{N}(0, k\sigma_g^2)$	Yes, disjointed	=1

<sup>1</sup> Erbe et al. 2012; <sup>2</sup> MacLeod et al. 2016

## ➤ GENOMIC PREDICTION : EXISTING MODELS

$$y = \mu \mathbf{1}_n + X\beta + e,$$

$$e \sim N(0, I_n \sigma_e^2)$$

Method	SNP effect prior distribution	Prior Annotations classes	# A(i)
BayesR <sup>1</sup>	$\beta_i \sim \sum_{\ell=1}^4 \pi_{\ell} \mathcal{N}(0, k\sigma_g^2)$	No	-
BayesRC <sup>2</sup>	$\beta_i   A(i) \sim \sum_{\ell=1}^4 \pi_{\ell, a} \mathcal{N}(0, k\sigma_g^2)$	Yes, disjointed	=1

<sup>1</sup> Erbe et al. 2012; <sup>2</sup> MacLeod et al. 2016



Reference genome ...ACTCCGTAAGTACTAGCCTACAAAGGCTAACTTACAAAAGATTTA...

Capitalizing on complex annotations in Bayesian genomic prediction for a backcross population of growing pigs

F. Mollandin, P. Croiseau, H. Gilbert, A. Rau

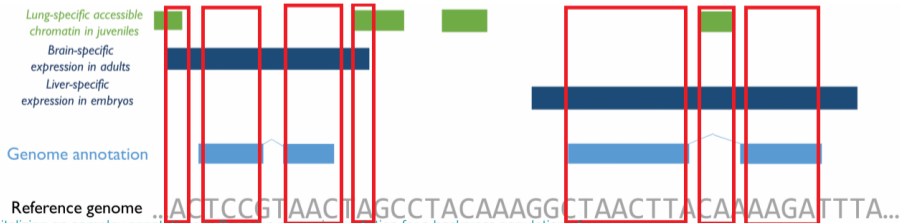
## ➤ GENOMIC PREDICTION : EXISTING MODELS

$$y = \mu \mathbf{1}_n + \mathbf{X}\beta + \mathbf{e},$$

$$\mathbf{e} \sim N(0, \mathbf{I}_n \sigma_e^2)$$

Method	SNP effect prior distribution	Prior Annotations classes	# A(i)
BayesR <sup>1</sup>	$\beta_i \sim \sum_{\ell=1}^4 \pi_{\ell} \mathcal{N}(0, k\sigma_g^2)$	No	-
BayesRC <sup>2</sup>	$\beta_i   A(i) \sim \sum_{\ell=1}^4 \pi_{\ell, \alpha} \mathcal{N}(0, k\sigma_g^2)$	Yes, disjointed	=1

<sup>1</sup> Erbe et al. 2012; <sup>2</sup> MacLeod et al. 2016



Capitalizing on complex annotations in Bayesian genomic prediction for a backcross population of growing pigs

F. Mollandin, P. Croiseau, H. Gilbert, A. Rau

## ➤ EXTENSIONS TO OVERLAPPING DATA

Method	SNP effect prior distribution	Prior Annotations classes	# A(i)
BayesR	$\beta_i \sim \sum_{\ell=1}^4 \pi_{\ell} \mathcal{N}(0, k\sigma_g^2)$	No	-
BayesRC	$\beta_i   A(i) \sim \sum_{\ell=1}^4 \pi_{\ell, a} \mathcal{N}(0, k\sigma_g^2)$	Yes, disjointed	=1
BayesRC+	$\beta_i   A(i) \sim \sum_{a \in A(i)} \sum_{\ell=1}^4 \pi_{\ell, a} \mathcal{N}(0, k\sigma_g^2)$	Yes, overlapping	$\geq 1$
BayesRC $\pi$	$\beta_i   A(i) \sim \sum_{a \in A(i)} p_{i, a} \sum_{\ell=1}^4 \pi_{\ell, a} \mathcal{N}(0, k\sigma_g^2)$	Yes, overlapping	$\geq 1$

Two proposed models for multi-annotated markers :

1. Multi-annotations should drive inclusion of markers in the model → **cumulative** additive model : **BayesRC+**
2. Multi-annotated markers should be assigned to the annotation category most representative of their effect → **mixture** of mixtures model : **BayesRC $\pi$**



## > ILLUSTRATION ON PIG DATA

### Genotypes and phenotypes data

- ▶  $n = 1297$  backcross pig population 3/4 Large-White, 1/4 Creole
- ▶ 2 traits : **backfat thickness** (BFT) and **average daily weight gain** (ADG), **pre-corrected** for age, sex, and farm
- ▶ Illumina Porcine 60k BeadChip array, after filtering we have **46,908** and **46,881** markers for ADG and BFT
- ▶ **10-fold** cross validation procedure

## > ILLUSTRATION ON PIG DATA

### Genotypes and phenotypes data

- ▶  $n = 1297$  backcross pig population 3/4 Large-White, 1/4 Creole
- ▶ 2 traits : **backfat thickness** (BFT) and **average daily weight gain** (ADG), **pre-corrected** for age, sex, and farm
- ▶ Illumina Porcine 60k BeadChip array, after filtering we have **46,908** and **46,881** markers for ADG and BFT
- ▶ **10-fold** cross validation procedure

### Annotations data

- ▶ **pigQTLdb**= public database of genotype-phenotype association studies
- ▶ **11 annotations** relevant to **production** (anatomy, behavioral, blood parameters, conformation, fatness, fatty acid content, feed conversion, fowth, immune capacity, litter traits, reproductive organs)

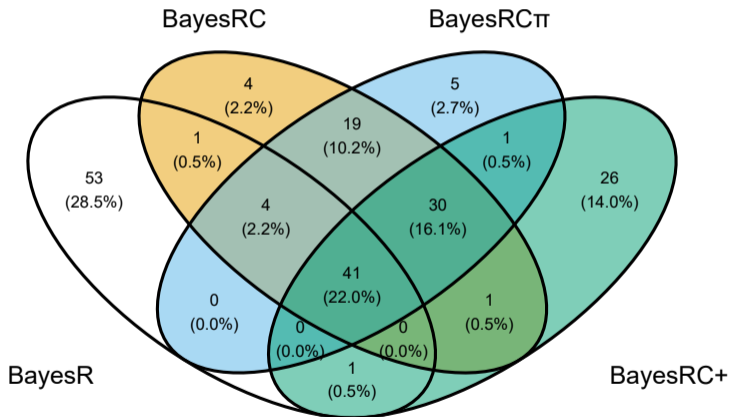
## > PREDICTION ACCURACY

	BayesR	BayesRC	BayesRC $\pi$	BayesRC+
ADG	0.21 ( $\pm 0.08$ )	+1.2 pts	+1.7 pts	+1.4 pts
BFT	0.26 ( $\pm 0.16$ )	-0.6 pts	-1 pts	+0.6 pts

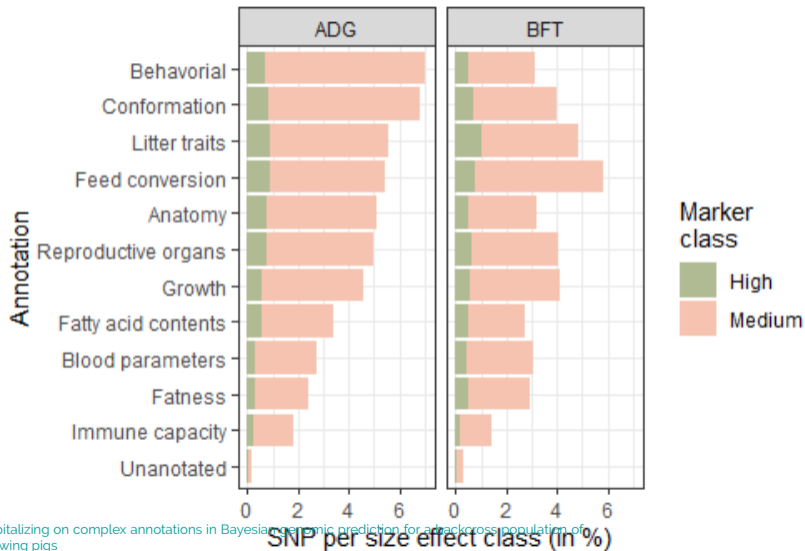
- ▶ Pearson correlation on validation data
- ▶ Lot of variability for BFT

# ➤ ANNOTATED SNPs ARE HIGHLIGHTED

Top 100 SNPs, ranked by *posterior* variance



## ➤ ANNOTATION EXPLOITATION, BAYESRC $\pi$



## > CONCLUSION

- ▶ Prediction accuracy can **increase** with the integration of functional annotations, with **suitable annotations**.

## > CONCLUSION

- ▶ Prediction accuracy can **increase** with the integration of functional annotations, with **suitable annotations**.
- ▶ (Multi-)annotated SNPs are **highlighted**, especially with BayesRC+.

## > CONCLUSION

- ▶ Prediction accuracy can **increase** with the integration of functional annotations, with **suitable annotations**.
- ▶ (Multi-)annotated SNPs are **highlighted**, especially with BayesRC+.
- ▶ New outputs to **explore annotations** usefulness.



## > CONCLUSION

- ▶ Prediction accuracy can **increase** with the integration of functional annotations, with **suitable annotations**.
- ▶ (Multi-)annotated SNPs are **highlighted**, especially with BayesRC+.
- ▶ New outputs to **explore annotations** usefulness.
- ▶ The **construction** of annotations have an **impact** on correlation and QTL mapping.

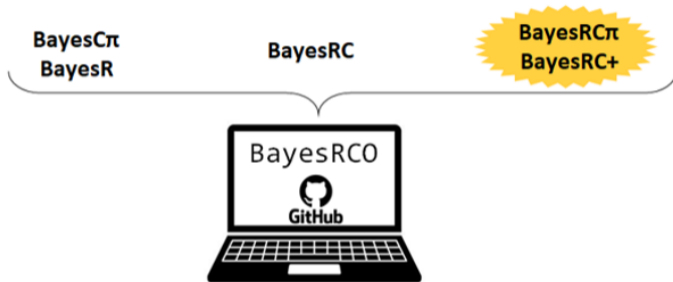
## > CONCLUSION

- ▶ Prediction accuracy can **increase** with the integration of functional annotations, with **suitable annotations**.
- ▶ (Multi-)annotated SNPs are **highlighted**, especially with BayesRC+.
- ▶ New outputs to **explore annotations** usefulness.
- ▶ The **construction** of annotations have an **impact** on correlation and QTL mapping.

→ BayesRC $\pi$ /+ flexibility offers new perspectives for the integration of functional annotations!

# ➤ THANK YOU!

Open-source software available here, with a full user's guide :  
<https://github.com/fmollandin/BayesRCO>



*This work is part of the **GENE-SWitCH** project that has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the grant agreement n°817998. This work also benefited from the clustering activities organized with the **BovReg** project, part of the European Union's Horizon 2020 Research and Innovation Programme under the grant agreement 37 n°815668.*



**INRAE**

Capitalizing on complex annotations in Bayesian genomic prediction for a backcross population of growing pigs

F. Mollandin, P. Croiseau, H. Gilbert, A. Rau