Accounting for predicted variant effects in genomic prediction in poultry

Andrea Rau, Eirini Tarsani, Andreas Kranis

GENE-SWitCH

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

Accounting for predicted variant effects in genomic prediction in poultry

Andrea Rau, Eirini Tarsani, and Andreas Kranis

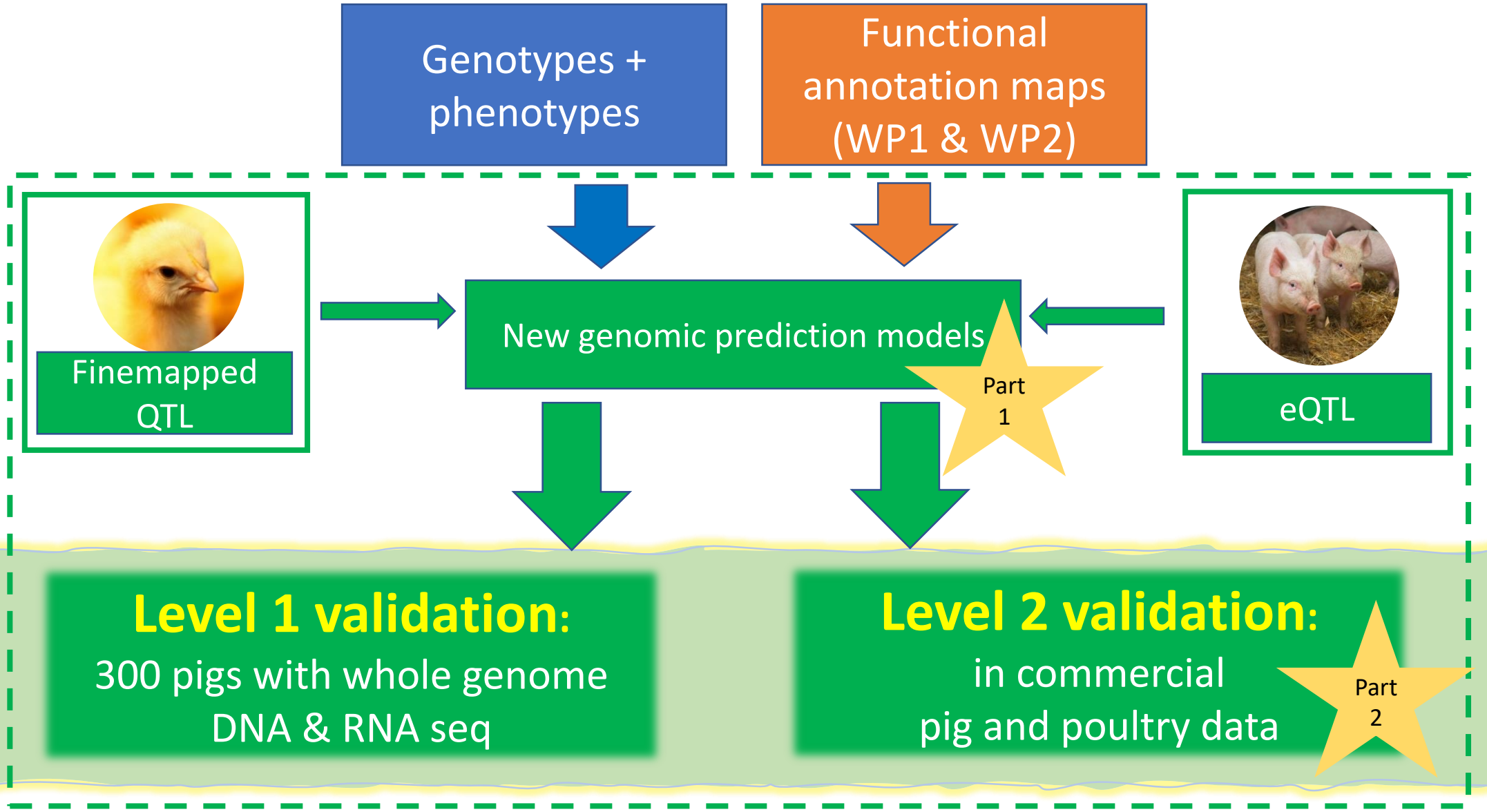
GENE-SWitCH 2nd Webinar for EM-ABG, Genomic Prediction

4 May 2023

IRTA · The University of Edinburgh · Uppsala Universitet · INRAE · Wageningen University & Research · Hendrix Genetics

*This project has received funding from the European Union's **Horizon 2020** **Research and Innovation** Programme under the grant agreement **n° 817998***

GENE-SWitCH WP4 overview



Part 1: Genomic prediction in a nutshell

Objective: select the best animals for reproduction to obtain **genetic improvement** of the population on **traits of interest**

$$Y = \mathbf{Z}\theta + \mathbf{X}\beta + \varepsilon$$

- Y = n-vector of traits
- $\mathbf{Z}$ = $n \times m$ matrix of covariates
- θ = m-vector of covariate effect parameters
- $\mathbf{X}$ = $n \times p$ matrix of (suitably coded) genotypes
- β = p-vector of genetic effect parameters
- ε = n-vector of errors representing noise, assumed to be iid and (usually) normally distributed

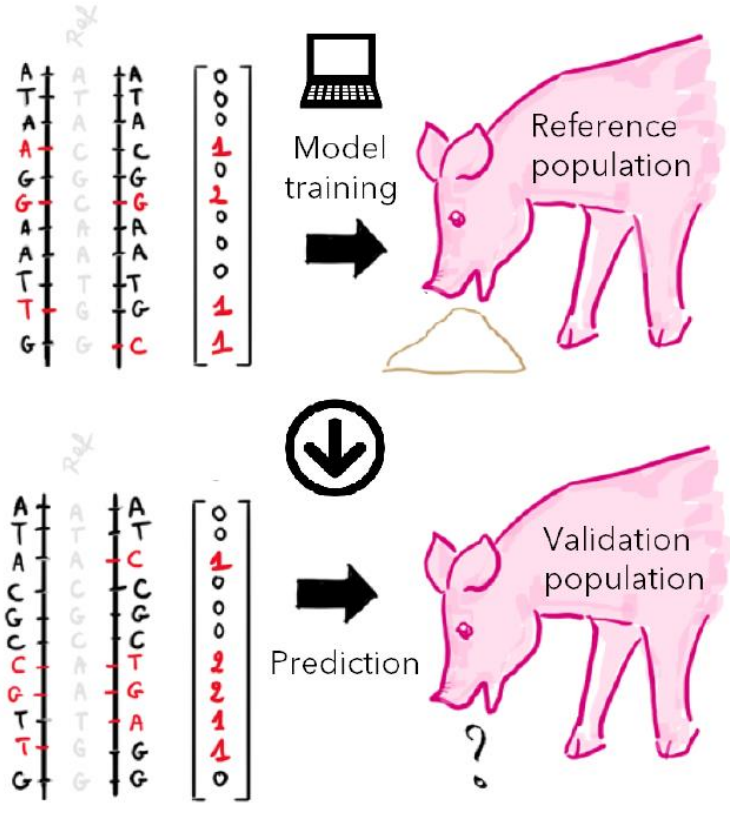


Image: F. Mollandin

Bayesian methods for genomic prediction

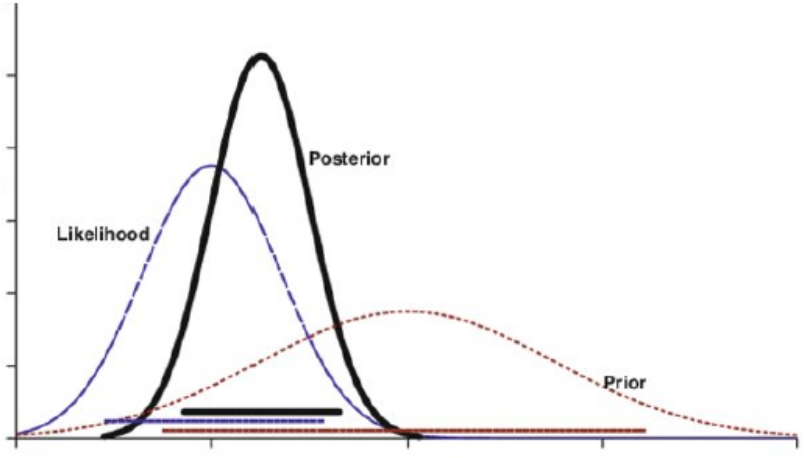


Image: 10.1007/s10681-007-9516-1

likelihood × prior

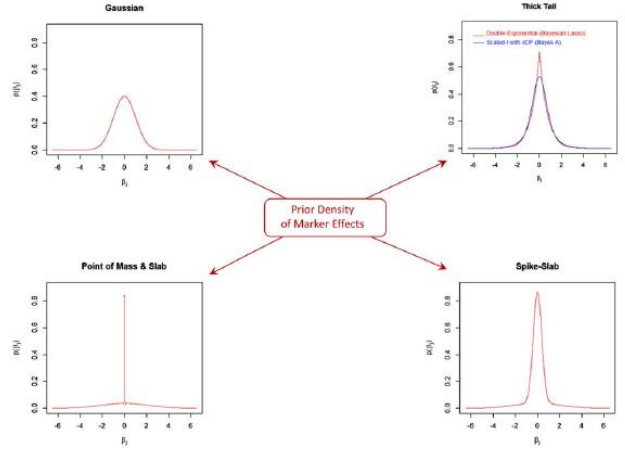
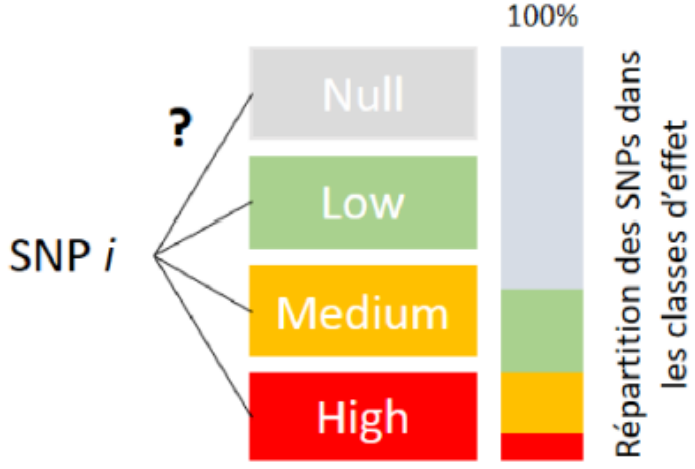


Image: 10.1543/genetics.112.143313

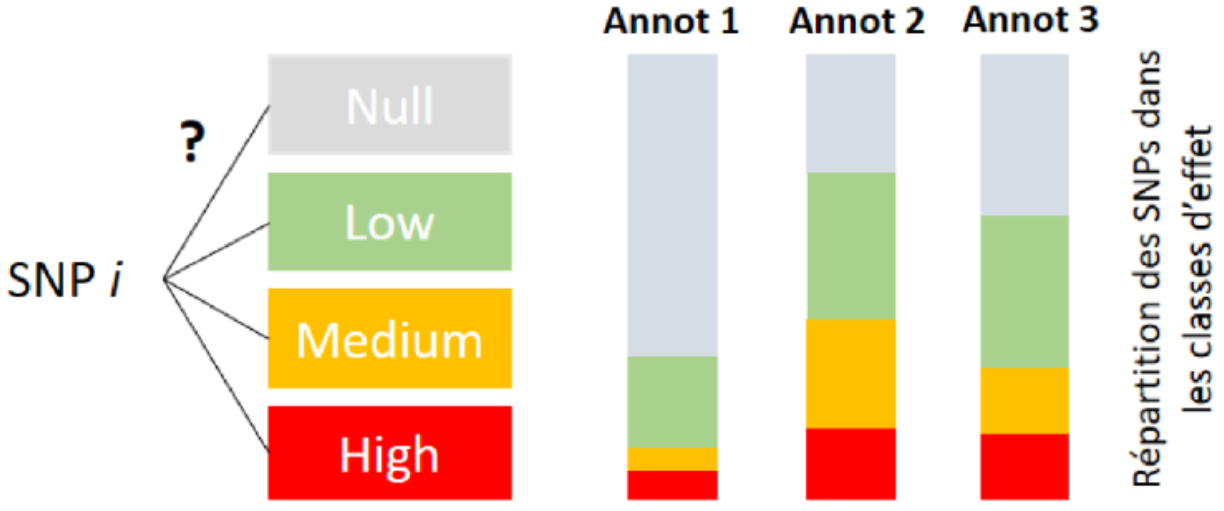
Should reflect a trait's genetic architecture (and be computationally feasible…)

Flexible but efficient: **BayesR**

(Erbe *et al*., 2012)

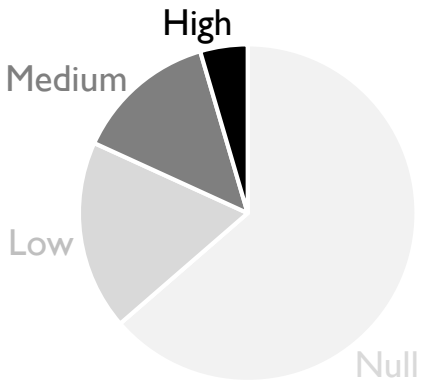


BayesR for SNP categories: **BayesRC**
(MacLeod *et al.*, 2016)

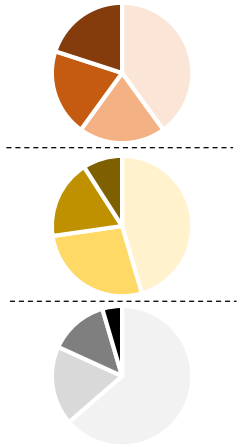


BayesR for overlapping SNP categories: **BayesRCO**

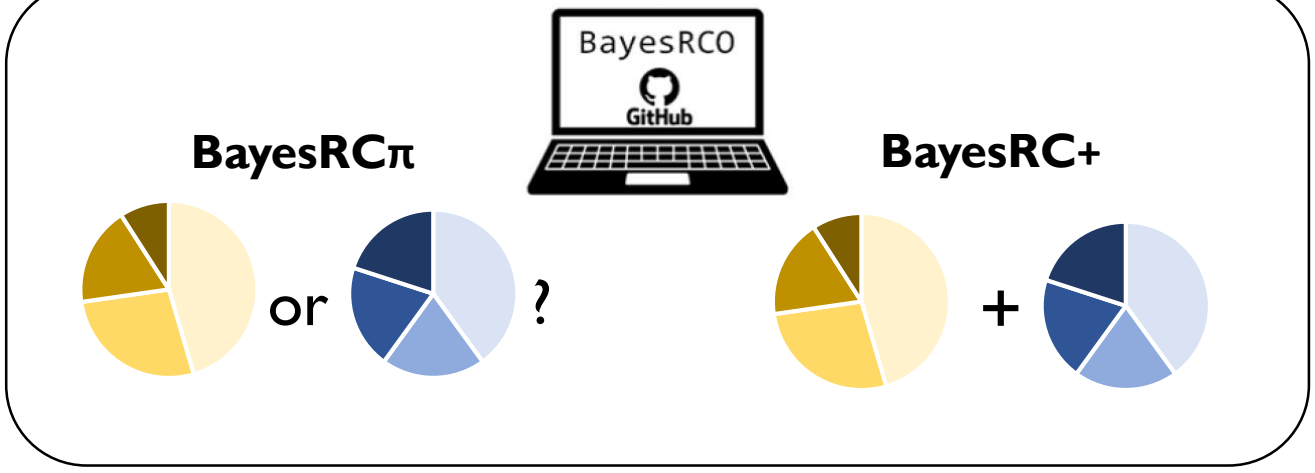
BayesR



BayesRC



(Single-annotated SNPs)



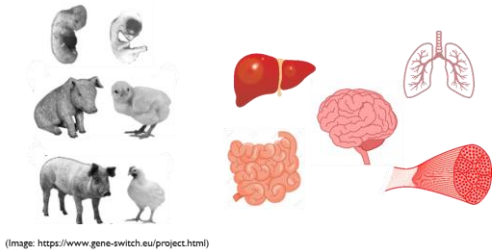
Multi-annotated SNP

Accessible chromatin
(embryo liver)

Unmethylated
(piglet liver)

GWAS hits

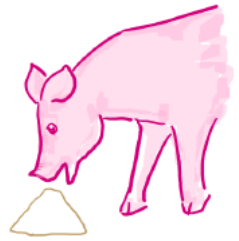
AnimalQTLdb



Genotype ...0000010012010021002000101000010110010111110...

...ACTCCGTAACTAGCCTACAAAGGCTAACTTACAAAAGATTTA...

Predict



GBV

Mollandin *et al.* (2022), *BMC Bioinformatics*; https://github.com/FAANG/BayesRCO

FAANG
Functional Annotation of Animal Genomes

Part 2: validation in commercial poultry data

Genotype and phenotype data

- p = 47,447 autosomal SNPs and n = 60,558 samples with records for BW
- BayesRCπ × 10 random splits into training/test data
  - Training: n = 48,446 samples
  - Test: n = 12,112 samples

SNP categories (“annotations”)

- Variant annotation using Ensembl Variant Effect Predictor (VEP) tool for *Gallus gallus*

Does the granularity of SNP categories affect results?

Predicted variant effects using VEP

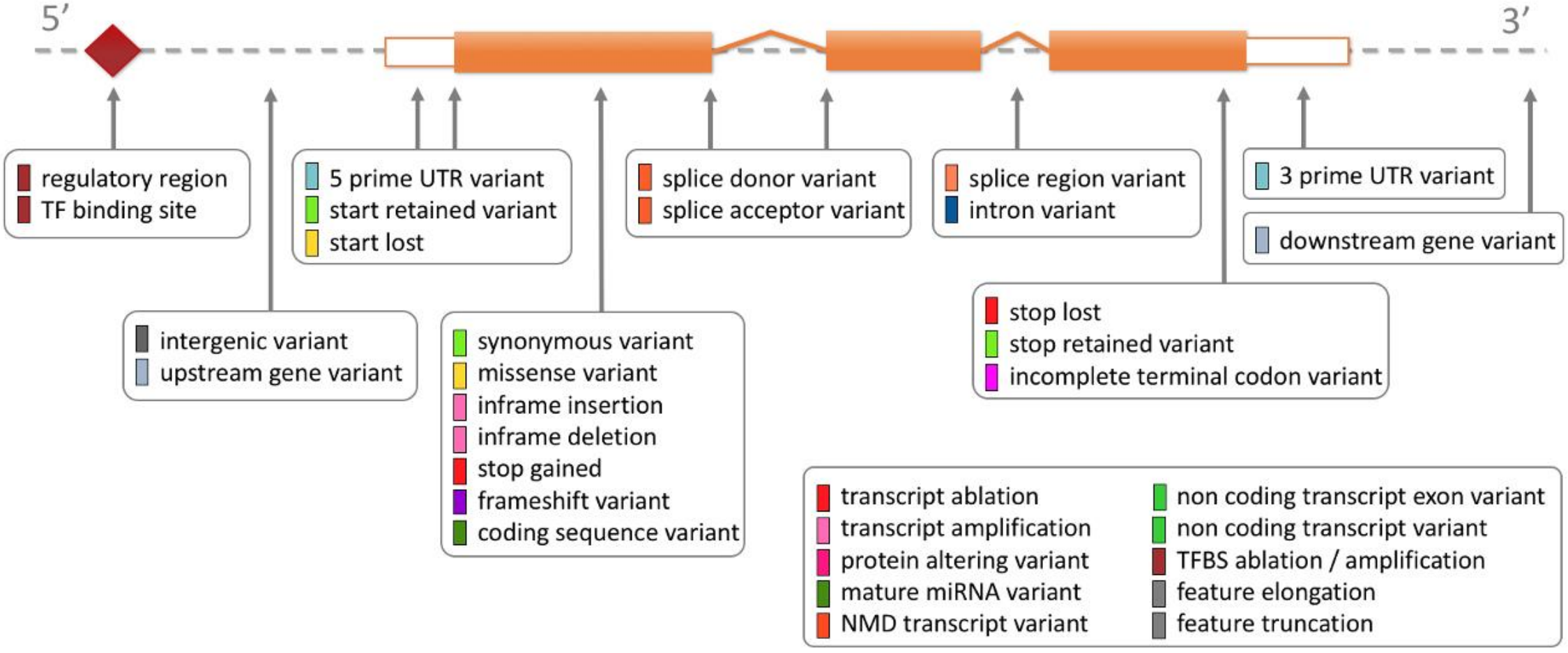


Image: https://grch37.ensembl.org/info/genome/variation/prediction/predicted_data.html

Results: Predicted variant effects by category

21 annotation categories

Table 1. Number of SNPs assigned to predicted variant annotations

Predicted consequence type	Number of variants
Intron variant	33351
Missense variant	940
Upstream gene variant	9556
Downstream gene variant	8822
3_prime UTR variant	1310
Synonymous variant	226
Intergenic variant	6409
Non-coding transcript exon variant	9270
5_prime UTR variant	317
Splice polypyrimidine tract variant	194
Splice region variant	130
Splice donor variant	45
Stop lost	57
Stop gained	87
Splice donor 5th base variant	9
Splice donor region variant	25
Stop retained variant	15
Splice acceptor variant	10
Start lost	5
Non-coding transcript variant	8443
Coding sequence variant	3

5 annotation categories

Table 2. Number of variants assigned to annotation categories by Sequence Ontology (SO)

Variant annotation category according to SO	Number of variants
intron variant	33351
exon variant	12157
intergenic variant	21625
splicing variant	303
Noncoding variant	8443

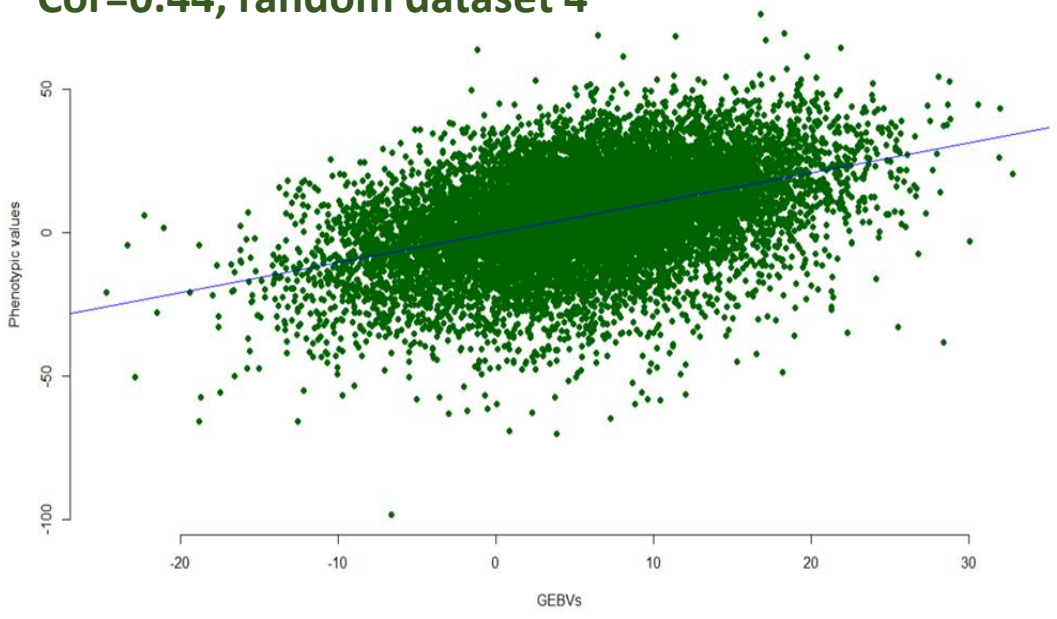
Note: SNPs may be assigned to multiple categories!

Results: BayesRCπ test correlation, **fine annotations** (21)

Test dataset (*n*=12,112 samples)	Correlation of GEBVs with phenotypic values
1	0.425
2	0.432
3	0.434
4	0.442
5	0.431
6	0.424
7	0.438
8	0.428
9	0.423
10	0.428

Average r= 0.43

Cor=0.44, random dataset 4

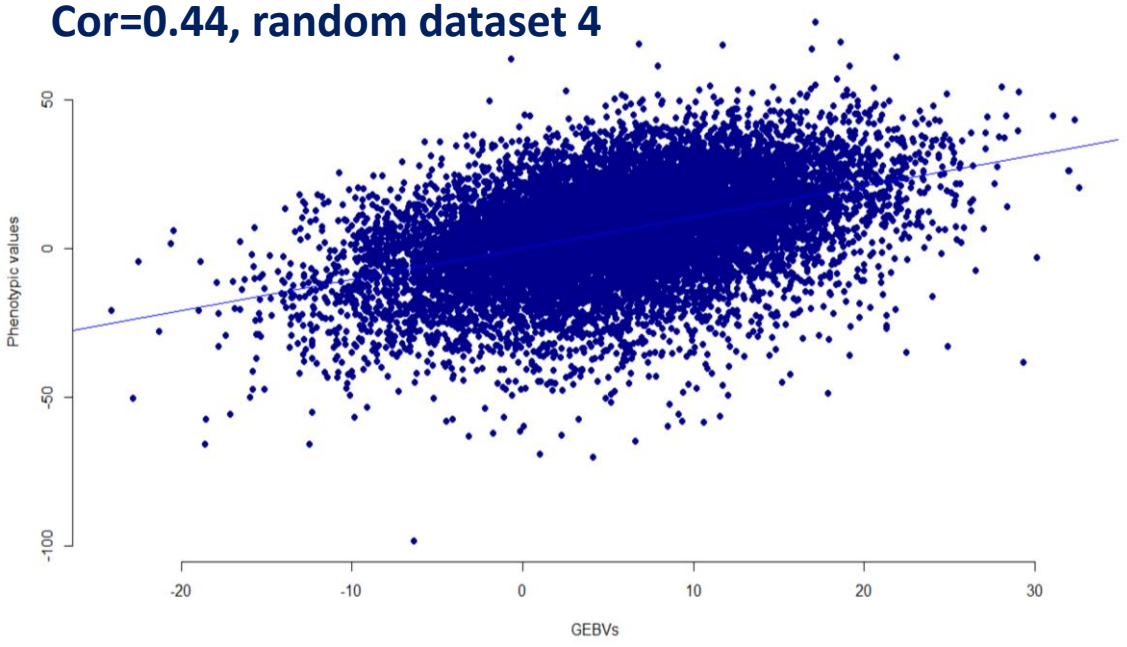


Results: BayesRCπ test correlation, **broad annotations** (5)

Test dataset (*n*=12,112 samples)	Correlation of GEBVs with phenotypic values
1	0.425
2	0.432
3	0.435
4	0.443
5	0.431
6	0.424
7	0.438
8	0.429
9	0.424
10	0.429

Average r= 0.43

Cor=0.44, random dataset 4



Take home messages

BayesRCO = flexible Bayesian genomic prediction model
to account for overlapping annotations

Our findings showed:

- BayesRCO can be used on commercial-scale data (60k animals), at least for medium-density genotypes
- Using a finer granularity of predicted variant effects does not appear to affect results (21 vs 5)

Limitations:

- **Genotyping resolution**: 47k SNPs used here, but greater added value from annotations expected when using WGS resolution
- **Choice of annotation categories**: different granularities of predicted variant effects here, but other (functional) annotation maps may be more informative

*This project has received funding from the European Union's **Horizon 2020** **Research and Innovation** Programme under the grant agreement **n° 817998***