Towards an Integrated Functional Annotation of the Bovine Genome

Christa Kühn on behalf of the BovReg consortium

PAG 2023, Functional Annotation of Animal Genomes (FAANG) workshop
January, 13rd, 2023

BovReg features

- EU H2020 – funded consortium as contribution to global FAANG with focus on biology-informed breeding (Sep 2019 – Feb 2024)

Objectives:

▶ Establish **new laboratory and bioinformatics tools**

▶ **Annotate functionally active genomic regions** for tissues that are highly relevant for the BovReg target traits across ontogenetic stages, gender, metabolic breed types and environments

▶ **Map molecular and whole-animal phenotypes** (ranging from transcriptome to intermediate phenome to performance/health) to newly annotated genomic features

▶ Develop **prototype models for integrating biological knowledge** on regulatory variation in genomic selection

▶ Provide **targeted training, dissemination and communication** activities and assess **social perception**

Aims:

- Annotation of the *Bos taurus* genome

- Key traits: robustness (fertility), health (mastitis susceptibility), biological efficiency (feed efficiency)

BovReg: Towards an Integrated Functional Annotation of the Bovine Genome
Ch. Kühn on behalf of the BovReg Consortium, PAG 2023, San Diego, Jan 13rd 2023

2

BovReg Consortium

- 20 (19) partners
- 13 countries
  & 2 international organisations
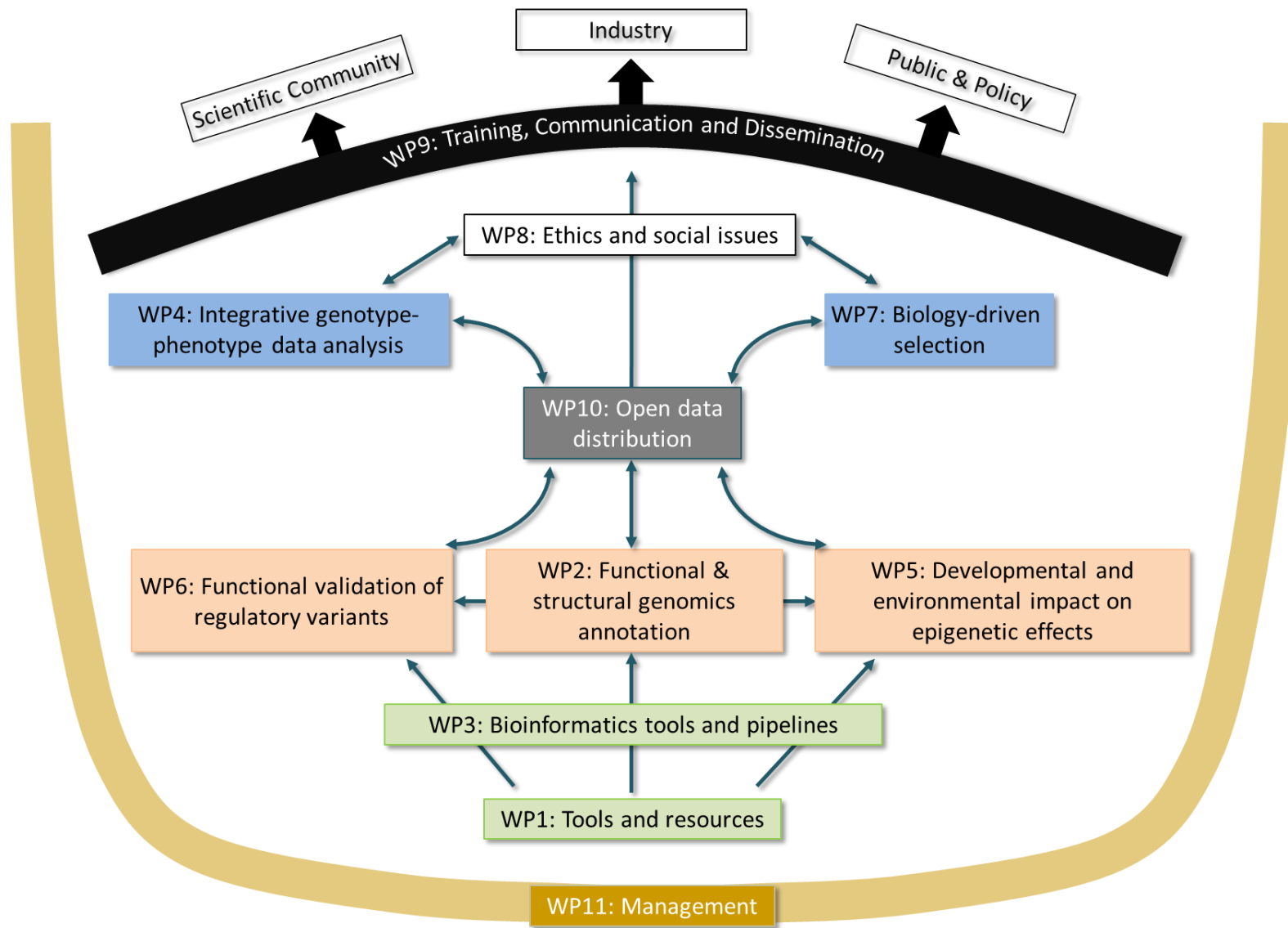
Belgium (2)
Denmark
Finland
France (2)
Germany (2)
Netherlands (2)
Poland
Portugal
Spain
Switzerland
United Kingdom (2)
Canada
(Australia)

EAAP
EMBL-EBI

Map labels: UAL, AgVic, UU, WR, LUKE, DIAG, GIGA, Edinethics, UEDIN, AU, EMBL-EBI, PAS, FBN, INRA, TiHo, ETH, ULIM, CRG, EAAP, FMV-LIS

https://mapswire.com/download/europe/europe-political-map-miller.jpg

BovReg: Towards an Integrated Functional Annotation of the Bovine Genome
Ch. Kühn on behalf of the BovReg Consortium, PAG 2023, San Diego, Jan 13rd 2023
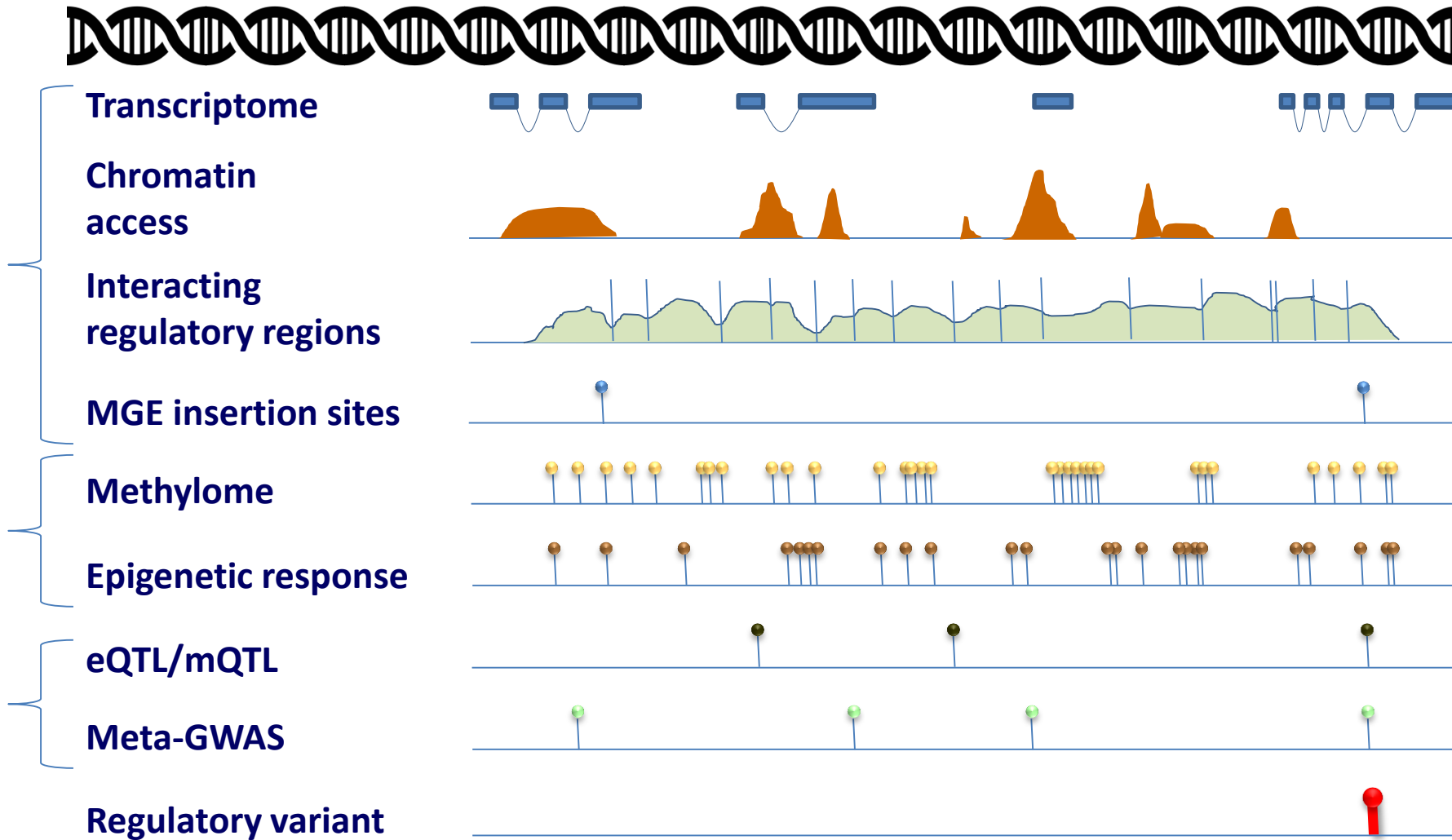
3

BovReg: Towards an Integrated Functional Annotation of the Bovine Genome
Ch. Kühn on behalf of the BovReg Consortium, PAG 2023, San Diego, Jan 13rd 2023
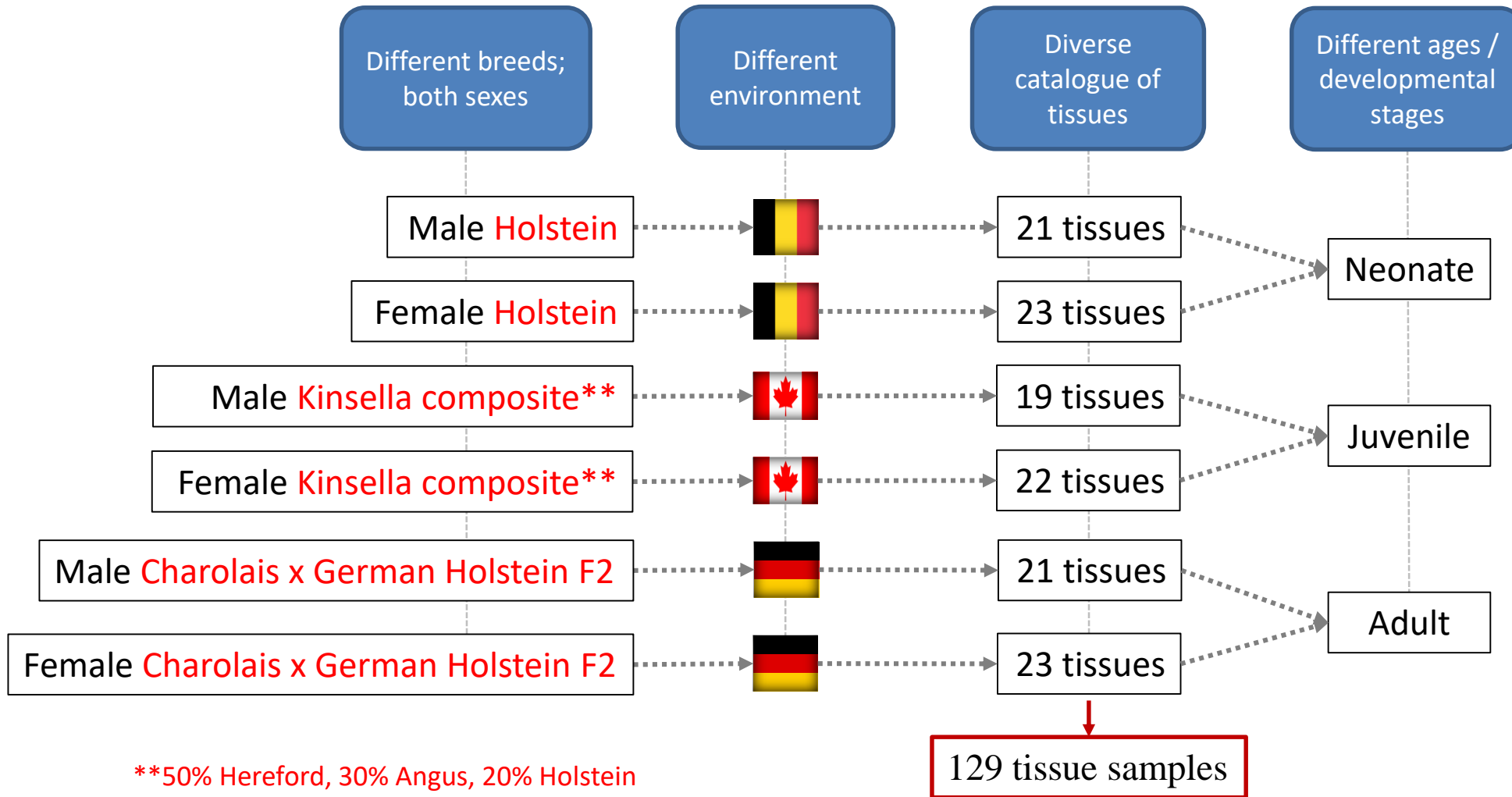
4

Integrated functional genome annotation in BovReg

Transcriptome

Chromatin access

Interacting regulatory regions

MGE insertion sites

Methylome

Epigenetic response

eQTL/mQTL

Meta-GWAS

Regulatory variant

Diversity in sample selection for a comprehensive annotation

Different breeds; both sexes	Different environment	Diverse catalogue of tissues	Different ages / developmental stages
Male Holstein	🇩🇪	21 tissues	Neonate
Female Holstein	🇩🇪	23 tissues	
Male Kinsella composite**	🇨🇦	19 tissues	Juvenile
Female Kinsella composite**	🇨🇦	22 tissues	
Male Charolais x German Holstein F2	🇩🇪	21 tissues	Adult
Female Charolais x German Holstein F2	🇩🇪	23 tissues	

**50% Hereford, 30% Angus, 20% Holstein

129 tissue samples

PLANT & ANIMAL GENOME CONFERENCE / PAG 30
THE PREMIER AND LARGEST AG-GENOMICS MEETING IN THE WORLD
JANUARY 13-18, 2023
TOWN & COUNTRY HOTEL
SAN DIEGO, CA

Transcriptome annotation

6 animals x 24 tissue types

+

**EBL, MAC-T, MDBK, F3 and Luteal**

mRNA

Total RNA

- 47,914 genes / transcriptional units assembled (27,607 Ensembl; 35,143 NCBI)
- 290,706 transcripts assembled (14.75% & 12.61% perfect match, 10.60% & 7.67% are novel; comparing with Ensembl and NCBI annotations)

Small RNA

- 1,265 miRNAs detected (638 bovine, 35 other species, 627 novel – 249 overlap repeats)

CAGE

- 51,295 TSS and 2,328 TSS-Enhancers detected (each present in at least 2/3rds of the tissues)
- cattle-specific set of TSSs detected (via comparative analysis with one avian and seven mammalian species using public CAGE derived promoter datasets)

BovReg tissue samples, primary cells, milk

scRNAseq

- 17,600 cells sequenced, FAANG DCC scRNAseq ruleset established
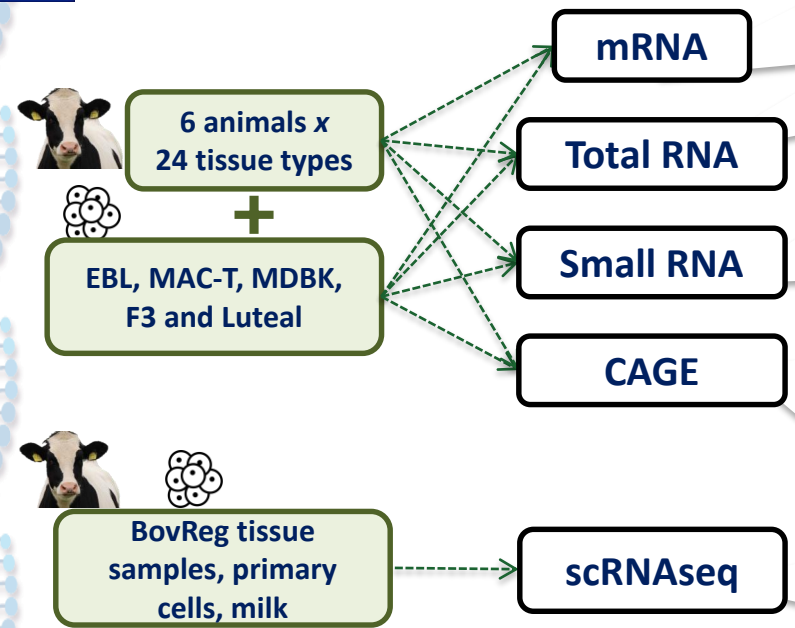
Data generated	Available at	Accession numbers
121 BAM files from mRNA tissue samples	FAANG DCC BovReg portal	PRJEB48309
5 BAM files from mRNA cell lines	FAANG DCC BovReg portal	PRJEB51474
126 BAM files from total RNA (tissues & cells)	FAANG DCC BovReg portal	PRJEB48026
102 BAM files from CAGE experiment	FAANG DCC BovReg portal	PRJEB43235
BovReg *de novo* Transcriptome Atlas (GTF)	MinIO server (EBI)	
fastq files + Seurat objects from scRNA-Seq assays	FAANG DCC BovReg portal, Figshare	PRJEB41576
FAANG meta-data standards for scRNAseq	FAANG DCC rulesets	

BovReg: Towards an Integrated Functional Annotation of the Bovine Genome
Ch. Kühn on behalf of the BovReg Consortium, PAG 2023, San Diego, Jan 13rd 2023

7

Transcriptome annotation

Further details on CAGE and miRNA results see:

PE0760 Expanding the Repertoire of miRNAs Annotated in Bovine for the BovReg Project

Gabriel Costa Monteiro Moreira et al.

LIÈGE université
GIGA

Mapping transcription start sites and enhancers unique to cattle using CAGE-Seq:
A diverse transcriptome approach

Mazdak Salavati et al.

ROSLIN   THE UNIVERSITY *of* EDINBURGH

>>https://www.bovreg.eu/communication-dissemination/events-with-bovreg-participation/

BovReg: Towards an Integrated Functional Annotation of the Bovine Genome
Ch. Kühn on behalf of the BovReg Consortium, PAG 2023, San Diego, Jan 13rd 2023

8

Chromatin access

6 animals x 24 tissue types

+

EBL, MAC-T, MDBK, F3 and Luteal

ATAC-Seq

ChIP-Seq

H3K4me3, H3K4me1, H3K27me3, H3K27ac and CTCF

Whole-genome sequenicng

Consensus...

Mark	Tissue samples	Cell lines
ATAC	720,770	153,625
H3K4me3	243,759	38,698
H3K4me1	423,807	271,978
H3K27me3	359,005	207,593
H3K27ac	439,835	118,298
CTCF	494,031	116,955

6 BAM files from the BovReg animals (WP2) and 5 BAM files from the cell lines (WP1) with > 30X

Example...
Neonate specific active enhancer
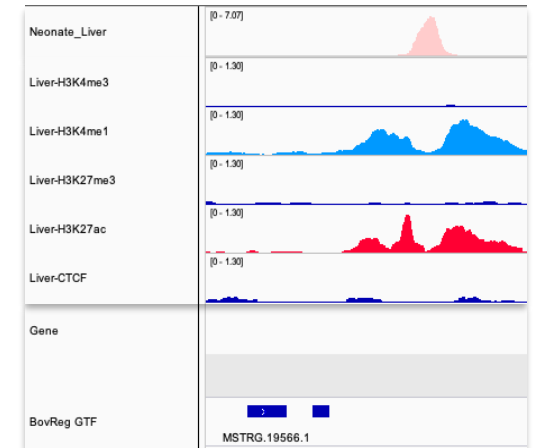
Data generated	Available at	Accession numbers
119 BAM files from ATAC-Seq assay	FAANG DCC BovReg portal	PRJEB51163
655 BAM files from ChIP-Seq assay	FAANG DCC BovReg portal	PRJEB51167
Narrow peak annotations in the tissue samples from ATAC- and ChIP-seq assays	MinIO server (EBI)	--
Narrow peak annotations in the cell lines from ATAC- and ChIP-seq assays	MinIO server (EBI)	--

BovReg: Towards an Integrated Functional Annotation of the Bovine Genome
Ch. Kühn on behalf of the BovReg Consortium, PAG 2023, San Diego, Jan 13rd 2023

9

Methylome and epigenetic response

6 animals *x* 24 tissue types

+

EBL, MAC-T, MDBK, F3 and Luteal

+

Embryos after (intergenerational) environmental challenge

WGMS

RRBS

- application of enzymatic methylation sequencing (EM-Seq) instead of WGBS

- clustering of tissues showed a similar pattern for RRBS and WGMS
- % mCpGs reflects the anticipated cellular functional activity of the respective tissue type x ontogenetic stage combinations

- between 15,509 and 270,998 differentially methylated CpGs in pairwise tissue RRBS comparisons

- proportion of DMCs from CpG island in the embryo comparison twice compared to DMC from comparisons between tissues

Data generated	Available at	Accession numbers
162 fastq/BAM files from RRBS tissue samples	FAANG DCC BovReg portal	PRJEB57838
5 fastq/BAM files from WGMS cell lines	FAANG DCC BovReg portal	in upload
22 fastq/BAM files from WGMS tissue samples	FAANG DCC BovReg portal	in upload
Pairwise differential methylation sites (BED)	MinIO server (EBI)	--
26 fastq/BAM from mRNA, RRBS and ATACseq from embryos after intergenerational heat stress		
37 fastq/BAM from mRNA, RRBS and ATACseq from embryos after metabolic stress		

BovReg: Towards an Integrated Functional Annotation of the Bovine Genome
Ch. Kühn on behalf of the BovReg Consortium, PAG 2023, San Diego, Jan 13rd 2023

10

Integration and standardisation of available data analysis pipelines

BovReg: Towards an Integrated Functional Annotation of the Bovine Genome
Ch. Kühn on behalf of the BovReg Consortium, PAG 2023, San Diego, Jan 13rd 2023

11
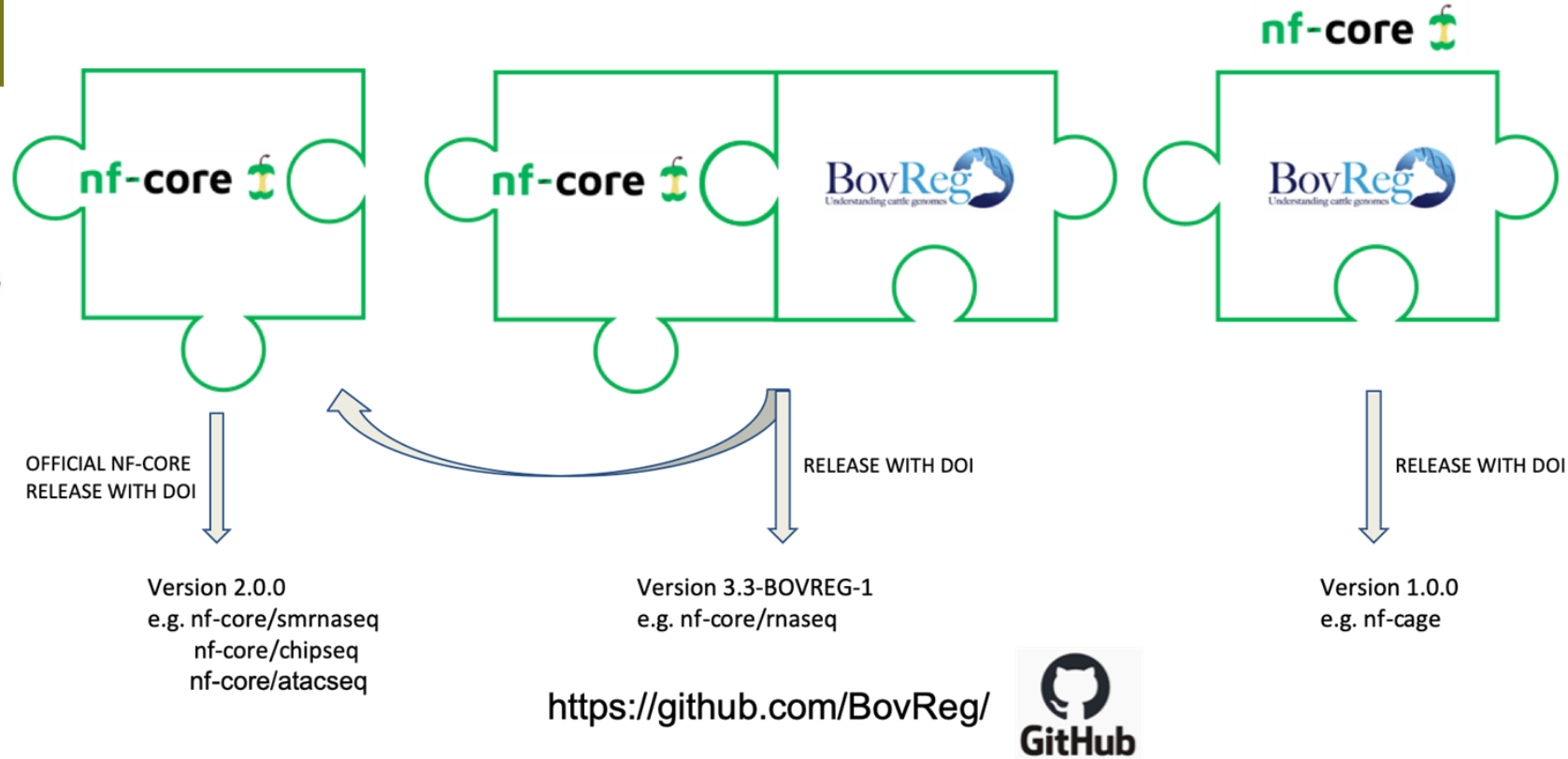
Overview of new or adapted Nextflow-based pipelines applied in BovReg

Group	Main developer	Analysis Type	Language	Standard	Github / DOI
GIGA	CRG	ATAC-seq	Nextflow	nf-core	nf-core/atacseq
UEDIN	UEDIN	CAGE-seq	Nextflow	nf-core	BovReg/nf-cage
GIGA	CRG	ChIP-seq	Nextflow	nf-core	nf-core/chipseq
FBN	FBN	circRNAs	bash (Nextflow planned)	-	
FBN	FBN	eQTLs	Nextflow	-	
INRAE	CRG	Hi-C	Nextflow	nf-core	nf-core/hic
GIGA	CRG	miRNAs expression	Nextflow	nf-core	nf-core/smrnaseq
FMV	FMV	miRNAs expression	Perl, R (Nextflow planned)	-	andreiaamaral/IsomiR-Window
FBN	FBN	RRBS	Nextflow + specific field tools	nf-core	nf-core/methyseq
UEDIN	UEDIN	SNP heritability	Nextflow	-	
INRAE	INRAE	SNPs/indels/SVs	Nextflow	nf-core	nf-core/sarek
GIGA	CRG	Transcript annotation and quantification (mRNA and lncRNA)	Nextflow	nf-core	BovReg/rnaseq
UEDIN	UEDIN	TWAS	Nextflow	nf-core	
INRAE	INRAE	WGMS	Nextflow + specific field tools	nf-core	nf-core/methyseq
CRG	CRG	Protein structure prediction	Nextflow	nf-core	nf-core/proteinfold

BovReg: Towards an Integrated Functional Annotation of the Bovine Genome
Ch. Kühn on behalf of the BovReg Consortium, PAG 2023, San Diego, Jan 13rd 2023

12

Example:
Development of a container- and workflow manager-based eQTL detection pipeline

partners

Genotype
WGS

Phenotype
RNAseq from various tissues

docker + nextflow

DSL2

Workflow
- RNA-seq alignment
- Quantification
- Count matrices
- Extract WGS data
- Covariates estimation
- QTL mapping
- QTL permutation

Example:
Development of a container- and workflow manager-based eQTL detection pipeline

BovReg: Towards an Integrated Functional Annotation of the Bovine Genome
Ch. Kühn on behalf of the BovReg Consortium, PAG 2023, San Diego, Jan 13rd 2023
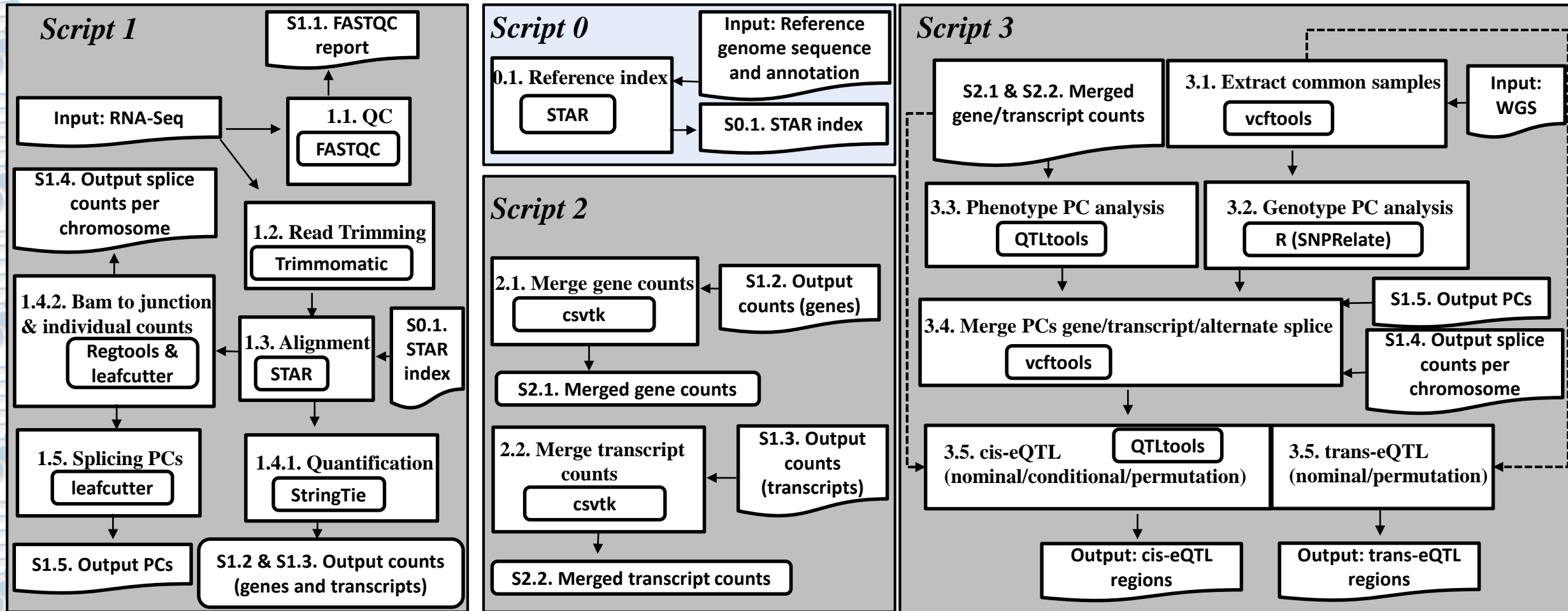
Genomic regions contributing to diversity in zootechnical and molecular phenotypes

Meta-GWAS
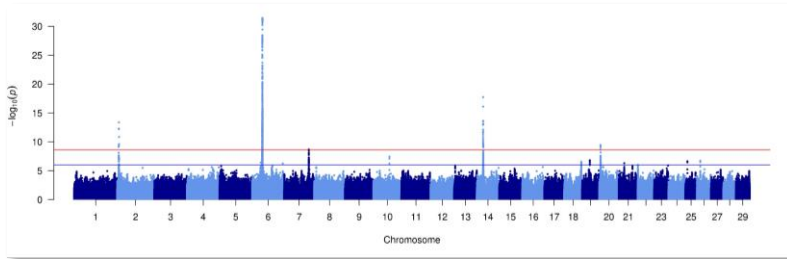
M. Sanchez et al. 2022

metabolite GWAS

unpublished

eQTL

P. Chitneedi et al. 2022

- 13 populations
- 2266 to 125,204 individuals
- 48 meta-analyses
- up to 60,364 trait-associated variants $(p < 10^{-8})$

metabolite QTL
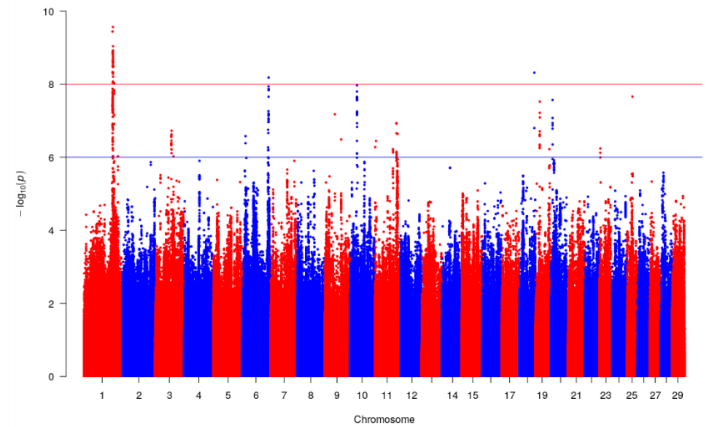
- 3 populations
- 797 different compounds
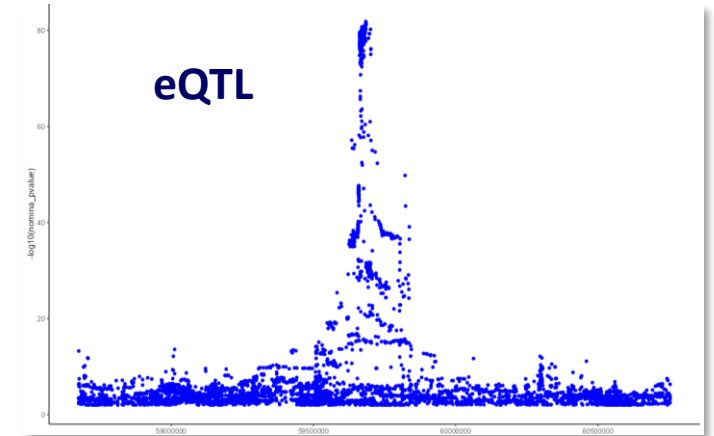
- 7 populations
- cis, trans, splicing eQTL
- 44,638 genes in analysis
- up to 7,598 cis eQTL $(p_{adj} < 0.05)$

BovReg: Towards an Integrated Functional Annotation of the Bovine Genome
Ch. Kühn on behalf of the BovReg Consortium, PAG 2023, San Diego, Jan 13rd 2023

15

TWAS Workflow for highlighting candidate genes

TWAS cohort

- Genotypes
- RNA-seq
- Or
- Genotypes imputed from RNA-seq

Train gene prediction models → Imputed expression values (across 1000s of cattle) → **Genes associated with phenotypes**

Gene / Transcript annotations

SNP locations

High-resolution cow transcriptome

GWAS cohort

- Summary GWAS
- Or
- Genotypes
- Phonotypes

BovReg: Towards an Integrated Functional Annotation of the Bovine Genome
Ch. Kühn on behalf of the BovReg Consortium, PAG 2023, San Diego, Jan 13rd 2023

Integrated functional genome annotation in BovReg

Transcriptome

Chromatin access

Interacting regulatory regions

MGE insertion sites

Methylome

Epigenetic response

eQTL/mQTL

Meta-GWAS

Regulatory variant

Functional region/variant validation

For details on validation of selected functional regions/variants see

PE0772 Functional Validation of Regulatory Sequences and Variants in Bovine Cell Lines Using CRISPR/Cas

Caroline Mörke et al.

BovReg: Towards an Integrated Functional Annotation of the Bovine Genome
Ch. Kühn on behalf of the BovReg Consortium, PAG 2023, San Diego, Jan 13rd 2023

18

Application of functional annotation data for improving genomic predictions

- INRAE tested the BayesRC model allowing for overlapping annotations in genomic prediction models

- WR conducted first validation studies for feed intake using meta-GWAS QTL and mQTL information for genomic prediction of feed efficiency

- AU: genomic prediction integrating multi-omics data for within and across breed prediction using simulations

BovReg: Towards an Integrated Functional Annotation of the Bovine Genome
Ch. Kühn on behalf of the BovReg Consortium, PAG 2023, San Diego, Jan 13rd 2023

19

Clustering Work in Global and European FAANG

↳ Joint contribution to the new global FAANG 2.0 task forces, co-organisers for
- FAANG single cell
- FAANG Compgen
- FAANG high-throughput phenotyping and data storage
- FAANG metaFAIR

↳ Collaboration with AG2PI and AgBio initiatives
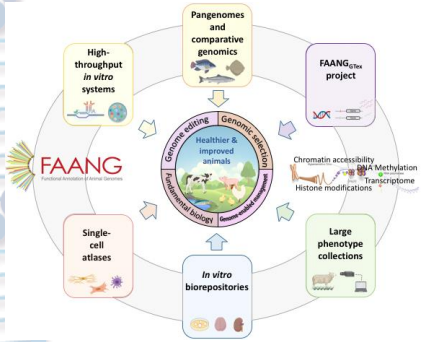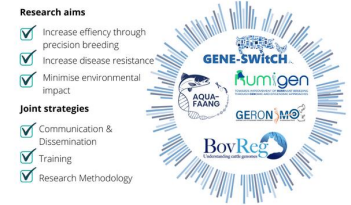
↳ EuroFAANG extended: new members **humigen** **GERONIMO** **HoloRuminant** Understanding microbiomes of the ruminant holobiont

EuroFAANG
Leveraging functional genomic information to make animal production more robust and sustainable in Europe and beyond

Research aims
- ☑ Increase effiency through precision breeding
- ☑ Increase disease resistance
- ☑ Minimise environmental impact

Joint strategies
- ☑ Communication & Dissemination
- ☑ Training
- ☑ Research Methodology

BovReg: Towards an Integrated Functional Annotation of the Bovine Genome
Ch. Kühn on behalf of the BovReg Consortium, PAG 2023, San Diego, Jan 13rd 2023

20

EuroFAANG

(Clark et al., 2020)

- New Horizon Europe infrastructure project

- Run time: 2023 - 2025

- Coordinator: FBN

- Aim: Establish concepts for infrastructures to facilitate research and innovation for genotype to phenotype (G2P) prediction in farmed animals (terrestrial and aquatic) to achieve sustainable, efficient and socially accepted farmed animal production in Europe

Partners:

University Edinburgh

INRAE

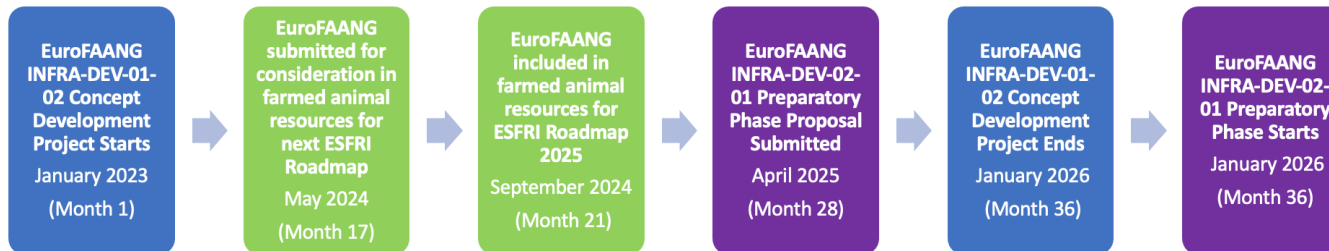Wageningen University

FBN

Norwegian University of Life Sciences

EMBL-EBI

European Forum for Farm Animal Breeders

Perspective: paving the way to embed farmed animal research in the ESFRI-Roadmap

| EuroFAANG INFRA-DEV-01-02 Concept Development Project Starts January 2023 (Month 1) | → | EuroFAANG submitted for consideration in farmed animal resources for next ESFRI Roadmap May 2024 (Month 17) | → | EuroFAANG included in farmed animal resources for ESFRI Roadmap 2025 September 2024 (Month 21) | → | EuroFAANG INFRA-DEV-02-01 Preparatory Phase Proposal Submitted April 2025 (Month 28) | → | EuroFAANG INFRA-DEV-01-02 Concept Development Project Ends January 2026 (Month 36) | → | EuroFAANG INFRA-DEV-02-01 Preparatory Phase Starts January 2026 (Month 36) |

BovReg: Towards an Integrated Functional Annotation of the Bovine Genome
Ch. Kühn on behalf of the BovReg Consortium, PAG 2023, San Diego, Jan 13ʳᵈ 2023

BovReg team at work

BovReg: Towards an Integrated Functional Annotation of the Bovine Genome
Ch. Kühn on behalf of the BovReg Consortium, PAG 2023, San Diego, Jan 13rd 2023

22

The BovReg functional genome annotation team

FBN
- D. Becker
- P. Chitneedi
- C. Mörke
- R. Weikard
- F. Hadlich
- N. Trakooljul
- F. Becker
- J. Vanselow
- Ch. Kühn

INRAE
- M. Sanchez
- D. Boichard
- M. Charles
- M. Bernard
- P. Croiseau
- D. Rocha

GIGA
- G. Costa Moreira
- S. Dupont
- L. Tang
- H. Takeda
- L. Karim
- C. Yuan
- C. Charlier

UEDIN
- M. Salavati
- J. Prendergast
- S. Jayaraman
- E. Clark

CRG
- B. Langer
- J. Espinosa-Carrasco
- C. Notredame

EMBL-EBI
- A. Sokolov
- P. Harrison

AU
- G. Sahana
- L. Janss
- E. Karaman

WR
- A. Bouwman
- B. Gredler
- R. Veerkamp

LUKE
- T. Iso-Touru
- D. Fischer
- J. Vilkki

ETH
- N. Kadri
- H. Pausch

FMV-Lisboa
- A. Amaral

IGHZ PAN
- N. Stojak
- H. Taniguchi

ULIM
- N. Duprat
- V. Blanchet

UAL
- C. Li
- J. Li
- G. Plastow

AgVic
- A. Chamberlain

BovReg: Towards an Integrated Functional Annotation of the Bovine Genome
Ch. Kühn on behalf of the BovReg Consortium, PAG 2023, San Diego, Jan 13rd 2023

23

BovReg partners

www.bovreg.eu

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

BovReg: Towards an Integrated Functional Annotation of the Bovine Genome
Ch. Kühn on behalf of the BovReg Consortium, PAG 2023, San Diego, Jan 13rd 2023

24

Thank you for your attention

Hiring project manager within EuroFAANG:
great opportunity to join the FBN team

https://www.fbn-dummerstorf.de/en/news/career/vacancies/
Contact: kuehn@fbn-dummerstorf.de or here at PAG

BovReg: Towards an Integrated Functional Annotation of the Bovine Genome
Ch. Kühn on behalf of the BovReg Consortium, PAG 2023, San Diego, Jan 13rd 2023

25

Location: https://www6.jouy.inrae.fr/gabi
Contacts: Elisabetta.Giuffra@inrae.fr, Michele.Boichard@inrae.fr

Location: www.fbn-dummerstorf.de
Contact: kuehn@fbn-dummerstorf.de