

Presented by Peter W. Harrison (EMBL-EBI)

➤ Progress of H2020 GENE-SWitCH: Functional Annotation of Pig and Chicken Genomes during Development

E. Giuffra*, H. Acloque* and GENE-SWitCH Consortium

(*) GABI, AgroParisTech, Université Paris Saclay,

Centre de Recherche Ile-de-France-Jouy-en-Josas – Antony, France



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

Started in July 2019 (4 years)
Extended to: December 2023



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

INRAE

PAG 2023 _ FAANG workshop
Elisabetta.giuffra@inrae.fr



➤ Overall aim



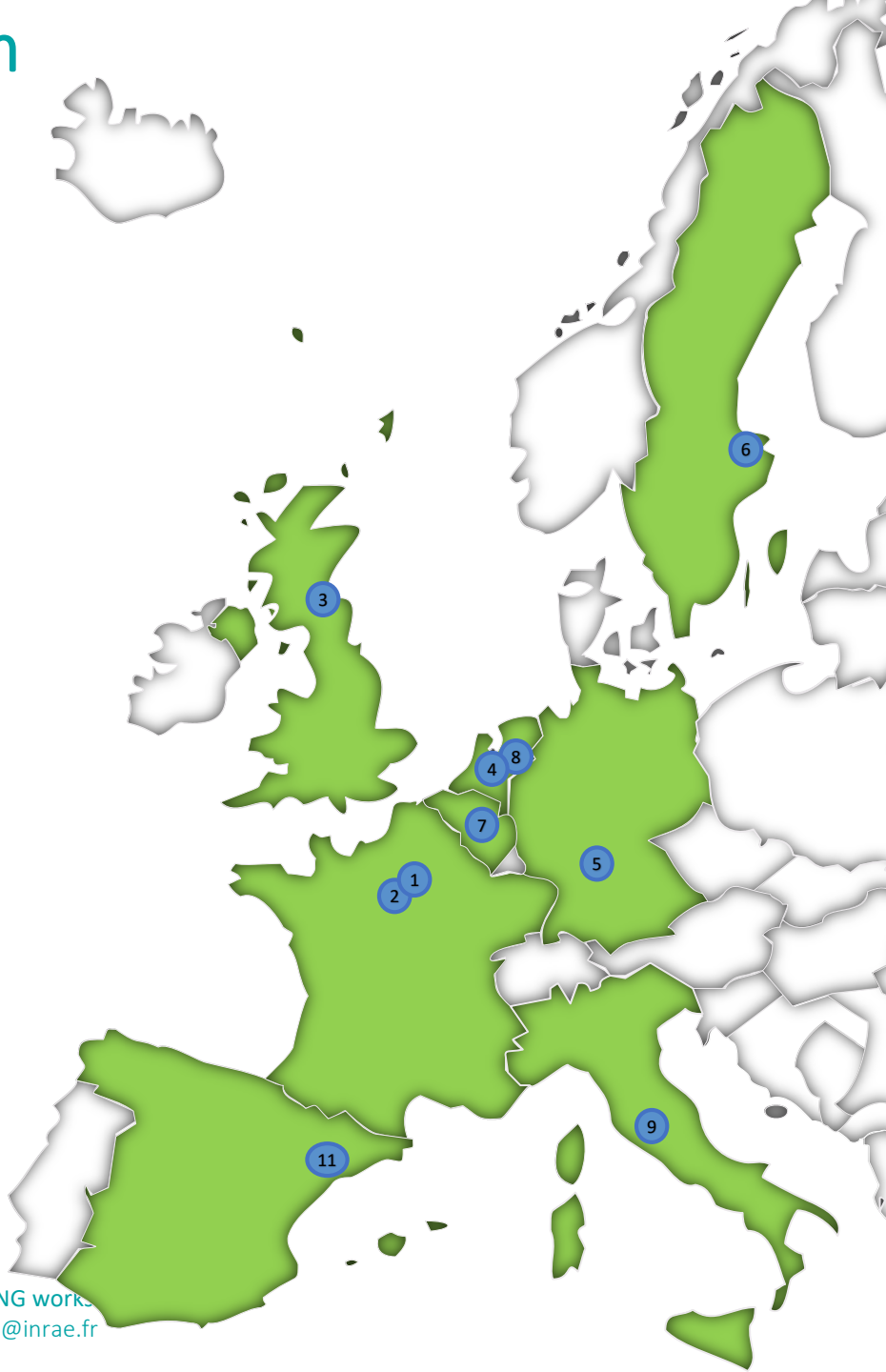
- To deliver underpinning knowledge on the pig and chicken genomes
and,
- to enable its translation to the pig and poultry sectors.













The underlying questions:

Can we identify and characterize the role of functional genomic elements – and in particular those that are active/poised/repressed during development - in the determination of the phenotypes of the adult animal?

What's the relative impact of their genetic variation on main production traits?

> Consortium



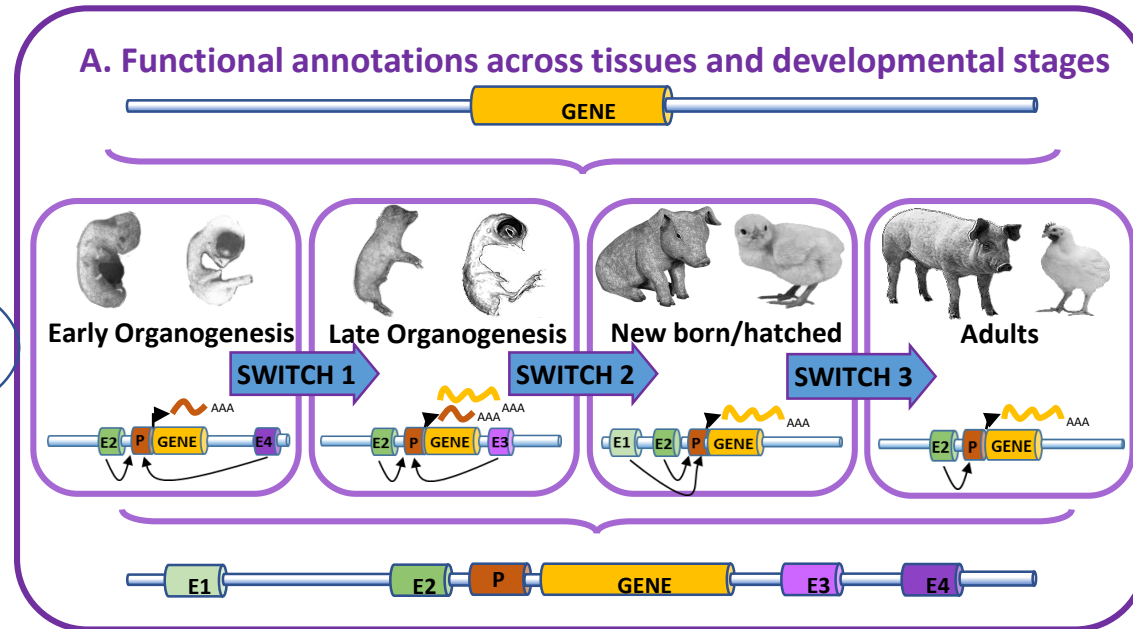
1.   Inserm
2. 
3.  THE UNIVERSITY of EDINBURGH
4.  WAGENINGEN UNIVERSITY & RESEARCH
5.  EMBL-EBI
6.  UPPSALA UNIVERSITET
7.  diagenode
Innovating Epigenetic Solutions
8.  EFFAB
European Forum of Farm Animal Breeders
9.  EAAP
10.  HENDRIX GENETICS
11.  IRTA

Collaboration with 3 breeding companies/ associations (Aviagen, Hypor B.V. and IFIP)

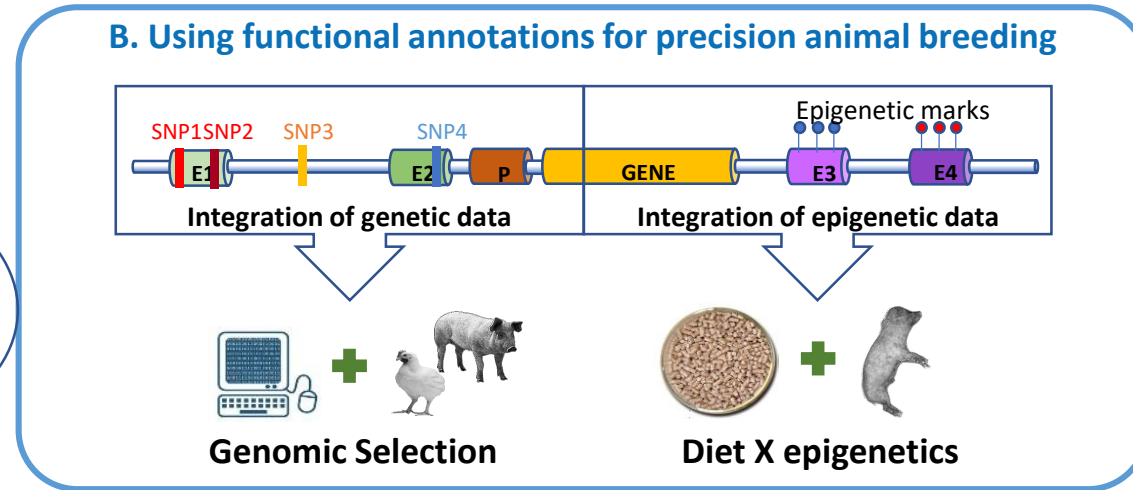
➤ Three specific & interconnected aims



WP1, 2 and 3



WP4 and WP5*
 (*) WP5 not presented



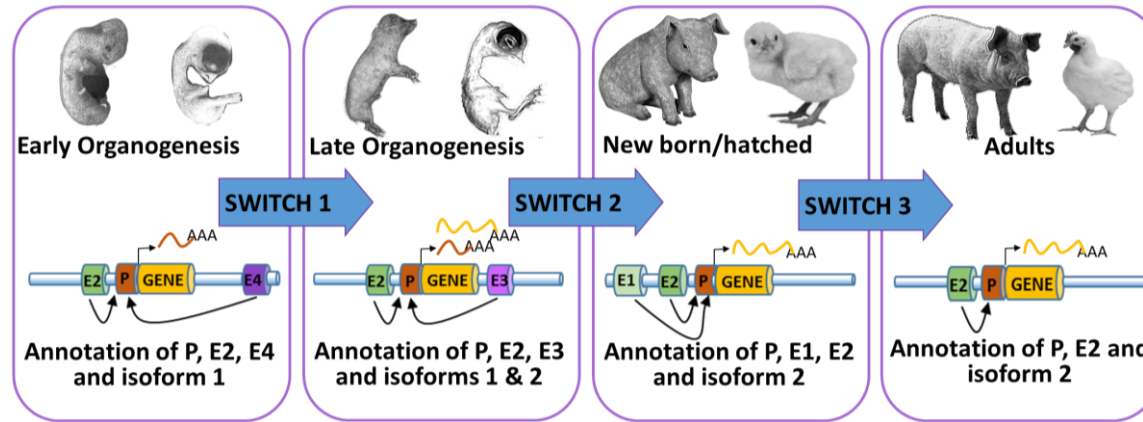
- www.gene-switch.eu
- <https://eurofaang.eu/>
- <https://data.faang.org/projects/GENE-SWitCH>



WP6
INRAE
 PAG 2023 _ FAANG workshop
 Elisabetta.giuffra@inrae.fr



➤ Aim A: Functional annotations across tissues and developmental stages



Led by



Co-led by EMBL-EBI



- Identify the functional elements of chicken and pig genomes across three developmental stages.
- Characterize their temporal dynamics (“switches”) and tissue-specificity, and their patterns of conservation and variation (mammals vs. birds).
- Deliver high-quality, richly annotated genome annotation maps.

INRAE

Inserm

WAGENINGEN
UNIVERSITY & RESEARCH

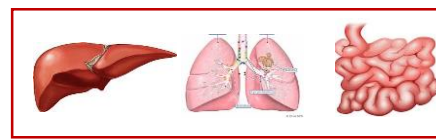
diagenode

INRAE

PAG 2023 _ FAANG workshop
Elisabetta.giuffra@inrae.fr

GENE-SWITCH

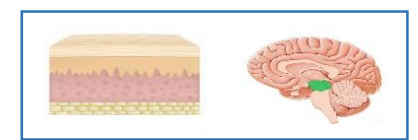
> Samples



Endoderm



Mesoderm



Ectoderm



	<u>Fetus 30D</u>	<u>Fetus 70D</u>	<u>Piglet</u>	<u>Embryo 8D</u>	<u>Embryo 15D</u>	<u>Chick</u>
<u>Core tissues</u>	hindbrain, lungs, kidney, skin, small intestine, liver, hindlimb muscle	cerebellum, lungs, kidney, dorsal skin, small intestine, liver, hindlimb muscle	cerebellum, lungs, kidney, dorsal skin, ileum, liver, gluteus medius	hindbrain, lungs, mesonephros, skin, small intestine, liver, hindlimb muscle	cerebellum, lungs, kidney, dorsal skin, small intestine, liver, hindlimb muscle	cerebellum, lungs, kidney, dorsal skin, small intestine, liver, gluteus medius
<u>Additional tissues</u>	brain cortex, heart, stomach, spleen, large intestine, gonads			brain cortex, heart, gizzard, spleen, large intestine, gonads		
<u>Contributors</u>	INRAE			University of Edinburgh (Roslin Institute)		
<u>BioSample IDs</u>	SAMEA7628703 to SAMEA7629286			SAMEA7679548 to SAMEA7680211		

New and improved standardised sampling and analysis protocols

(<https://data.faang.org>)



> Datasets

Contributors	Assay	Number of libraries		Raw reads number per library (mean)		Accession Number		
		SSC	GGA	SSC	GGA	SSC	GGA	
Diagenode libraries INRAE GeT-PlaGE sequencing	ATAC-seq	84	84	100M	105M	PRJEB44468	PRJEB45945	
	ChIP-seq	IgG	78	77	48M	58M	local server	
		CTCF	82	78	51M	52M	local server	
		H3K4me1	80	76	95M	115M	local server	
		H3K4me3	83	76	50M	57M	local server	
		H3K27me3	83	81	103M	126M	local server	
		H3K27Ac	84	77	53M	61M	local server	
Roslin Institute and INRAE RNA extraction Roslin Institute Iseq libraries Genewiz and Earlham Institute	RNA-seq	mRNA-seq	84	84	150M	PRJEB41970	PRJEB42025	
		smallRNA-seq	84	84	65M	PRJEB42001	PRJEB42041	
		Iso-seq	21	21	3.5M	PRJEB50963	PRJEB48060	
Wageningen University outsourcing to Novogen INRAE and Roslin Institute DNA extraction	DNA methylation	RRBS	63	63	59M	55M	PRJEB41822	PRJEB41829
		WGBS	21	21	36X*	31X*	PRJEB42772	PRJEB42775
INRAE libraries production and sequencing	Capture Hi-C	12	12	180M	200M	PRJEB44486	local server	

Raw data on <https://data.faang.org> and ENA under the terms of the Fort Lauderdale agreement and Toronto Statement.

➤ Analysis pipelines for primary analyses (<https://github.com/FAANG/>)

New development/extensions (manuscripts in preparation):

- RNA-seq data TAGADA (Nextflow DSL2), a highly parallelised annotation and quantification tool.
- GSM pipeline (bisulfite sequencing data) from extension of nf-core methyl-seq pipeline.

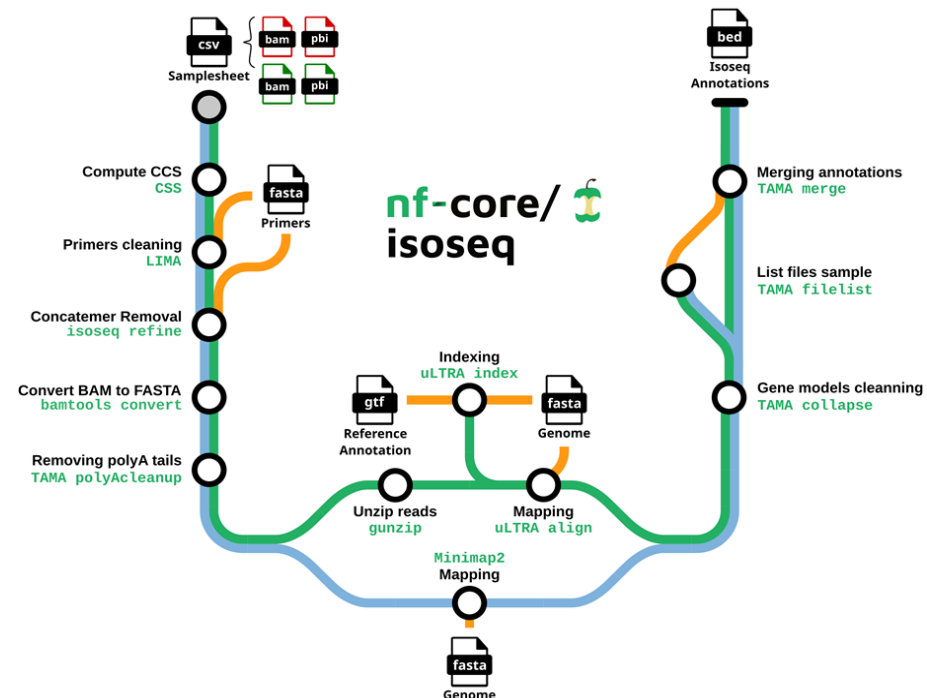
INRAE

Inserm

WAGENINGEN
UNIVERSITY & RESEARCH

Use/refinement of other nf-core community pipelines for:

- sRNA-Seq
- ATAC-Seq
- Iso-seq (also released on <https://github.com/nf-core/iseq>).
- Capture Hi-C
- CHIP-seq



Sebastien
Guizard



➤ Primary analyses

Analyses revealed an excellent clustering of samples by tissue and stages in both species.

- RNA-seq, sRNA-seq and Iso-seq data:
 - Several 'new' genes and transcripts identified.



- ATAC-seq data

- Confirmed the expected enrichment of open chromatin around promoters and TSS.



- Analysis of WGBS and RRBS:

- Analysis of methylation patterns revealed dynamic changes in the methylome during development



WAGENINGEN
UNIVERSITY & RESEARCH



INRAE

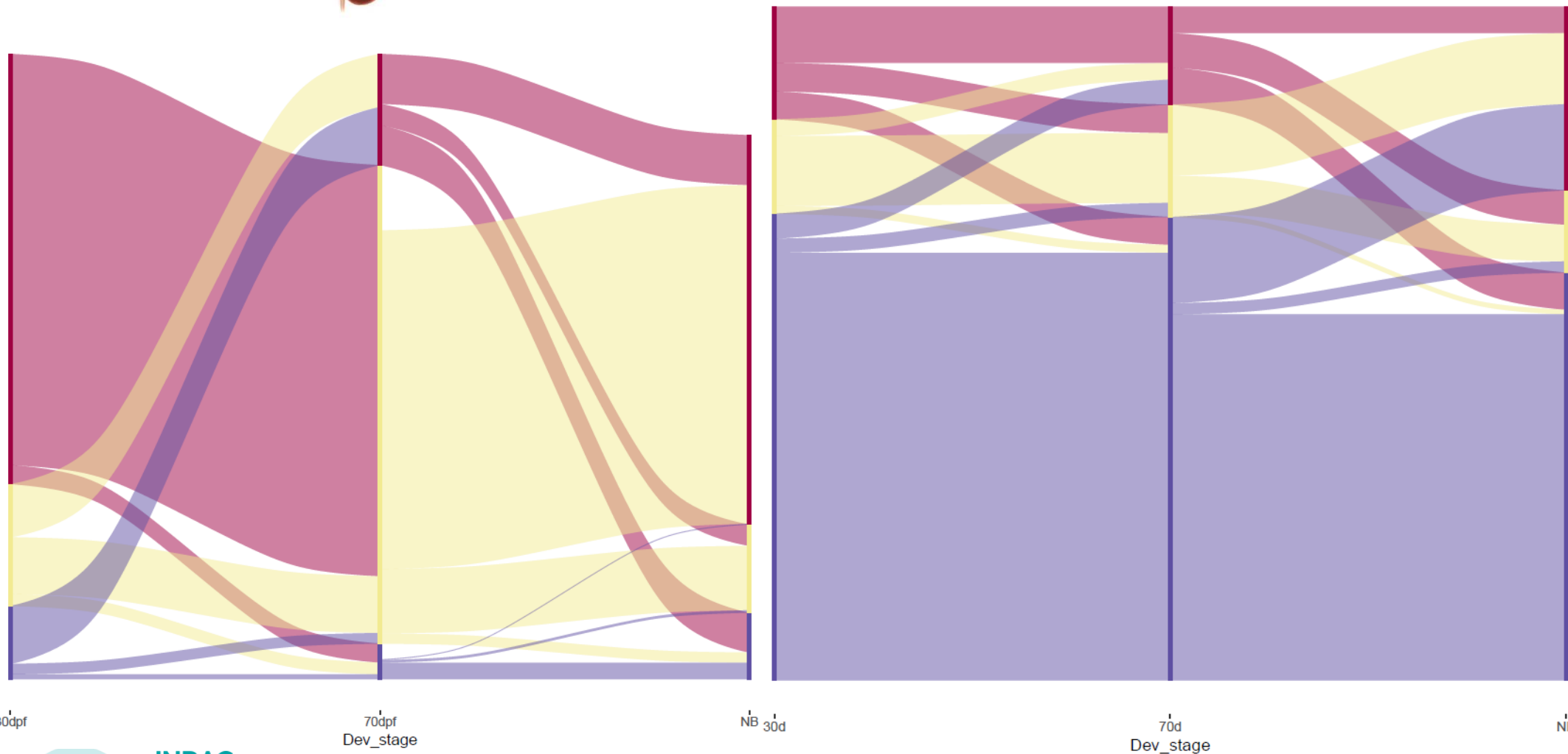
PAG 2023 _ FAANG workshop
Elisabetta.giuffra@inrae.fr



➤ Dynamic changes in methylome during development



Meth_state FMR LMR UMR

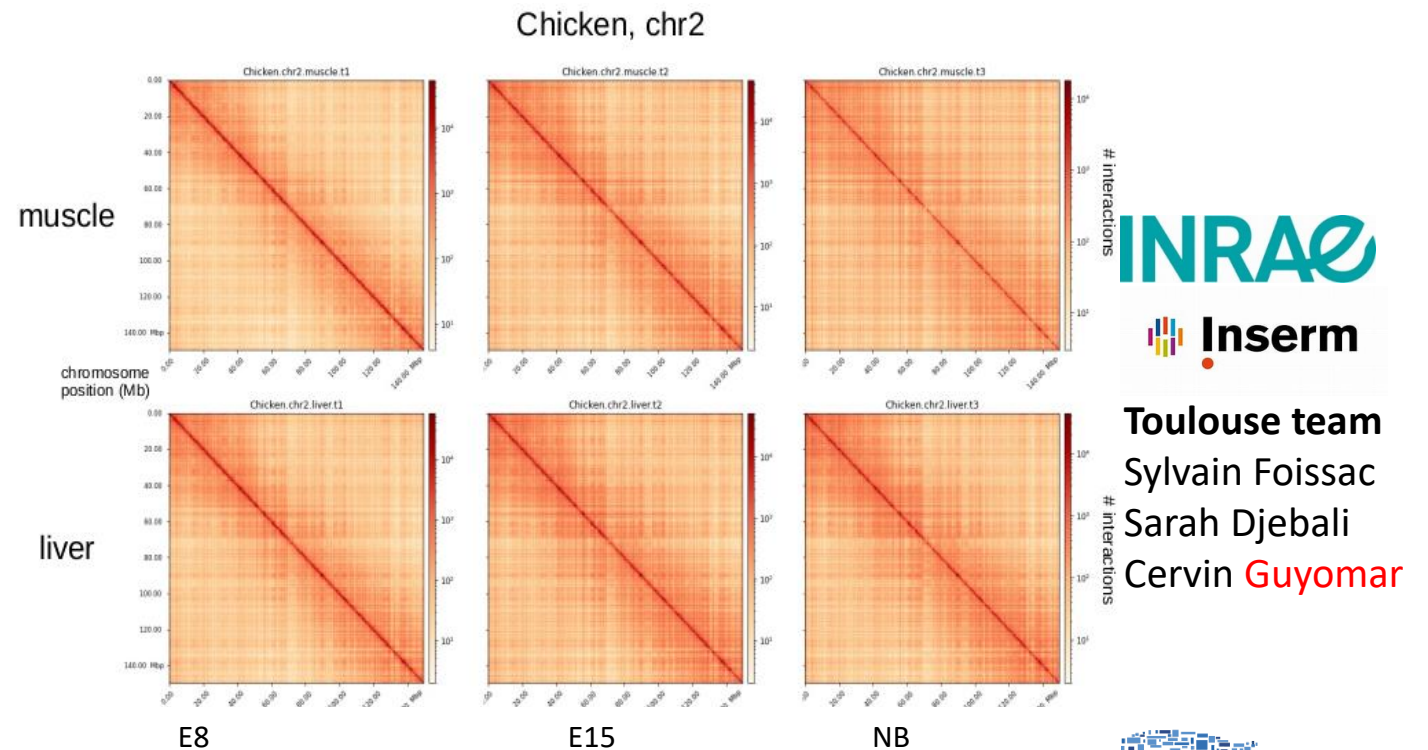


Jani de Vos
Martijn Derks
Ole Madsen



➤ Primary analyses: work in progress

- Joint analysis of RNA-seq, sRNA-seq and Iso-seq data to characterise pig and chicken transcriptome and its dynamics through development (manuscript in preparation by INRAE, Inserm and UEDIN).
- Complete the analysis of capture Hi-C data (muscle and liver tissues) to identify interactions between regulatory elements (e.g. promoters and enhancers)
- Ready to start the analysis of the new ChIP-seq data, with priority given to muscle and liver tissues (spring 2023).
- *Annotation will be updated incrementally as analyses are completed*



Capture Hi-C. Raw interaction matrices after merging replicates at 500Kb resolution (chicken) at three developmental stages.



INRAE

PAG 2023 _ FAANG workshop
Elisabetta.giuffra@inrae.fr



➤ Genome annotation

- The pig reference genome (Sscrofa11.)


&

- The original surrogate chicken (Red Jungle Fowl) + a Broiler and the White Leghorn genomes

have been annotated with gene / transcript models and chromatin accessibility (ATAC-seq) with GENE-SWitCH data.

JOURNAL ARTICLE

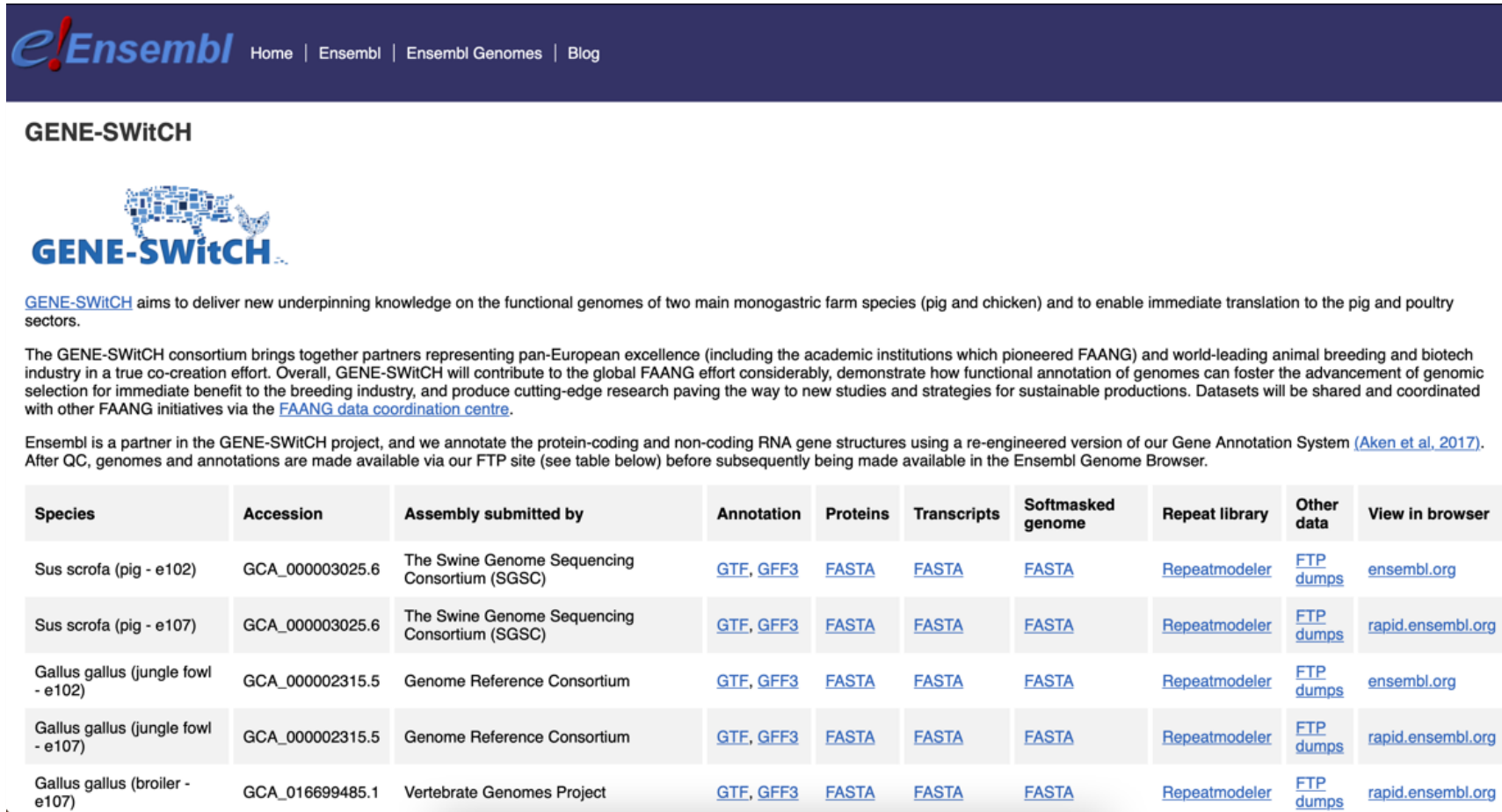
Ensembl 2023

Fergal J Martin , M Ridwan Amode, Alisha Aneja, Olanrewaju Austine-Orimoloye, Andrey G Azov, If Barnes, Arne Becker, Ruth Bennett, Andrew Berry, Jyothish Bhai ...
Show more

Nucleic Acids Research, Volume 51, Issue D1, 6 January 2023, Pages D933–D941,
<https://doi.org/10.1093/nar/gkac958>


➤ GENE-SWitCH Ensembl Collection

- GENE-SWitCH data from multiple releases tracked in a dedicated page.



e!Ensembl Home | Ensembl | Ensembl Genomes | Blog

GENE-SWitCH



GENE-SWitCH aims to deliver new underpinning knowledge on the functional genomes of two main monogastric farm species (pig and chicken) and to enable immediate translation to the pig and poultry sectors.

The GENE-SWitCH consortium brings together partners representing pan-European excellence (including the academic institutions which pioneered FAANG) and world-leading animal breeding and biotech industry in a true co-creation effort. Overall, GENE-SWitCH will contribute to the global FAANG effort considerably, demonstrate how functional annotation of genomes can foster the advancement of genomic selection for immediate benefit to the breeding industry, and produce cutting-edge research paving the way to new studies and strategies for sustainable productions. Datasets will be shared and coordinated with other FAANG initiatives via the [FAANG data coordination centre](#).

Ensembl is a partner in the GENE-SWitCH project, and we annotate the protein-coding and non-coding RNA gene structures using a re-engineered version of our Gene Annotation System ([Aken et al, 2017](#)). After QC, genomes and annotations are made available via our FTP site (see table below) before subsequently being made available in the Ensembl Genome Browser.

Species	Accession	Assembly submitted by	Annotation	Proteins	Transcripts	Softmasked genome	Repeat library	Other data	View in browser
Sus scrofa (pig - e102)	GCA_000003025.6	The Swine Genome Sequencing Consortium (SGSC)	GTF , GFF3	FASTA	FASTA	FASTA	Repeatmodeler	FTP dumps	ensembl.org
Sus scrofa (pig - e107)	GCA_000003025.6	The Swine Genome Sequencing Consortium (SGSC)	GTF , GFF3	FASTA	FASTA	FASTA	Repeatmodeler	FTP dumps	rapid.ensembl.org
Gallus gallus (jungle fowl - e102)	GCA_000002315.5	Genome Reference Consortium	GTF , GFF3	FASTA	FASTA	FASTA	Repeatmodeler	FTP dumps	ensembl.org
Gallus gallus (jungle fowl - e107)	GCA_000002315.5	Genome Reference Consortium	GTF , GFF3	FASTA	FASTA	FASTA	Repeatmodeler	FTP dumps	rapid.ensembl.org
Gallus gallus (broiler - e107)	GCA_016699485.1	Vertebrate Genomes Project	GTF , GFF3	FASTA	FASTA	FASTA	Repeatmodeler	FTP dumps	rapid.ensembl.org



<https://projects.ensembl.org/gene-switch/>



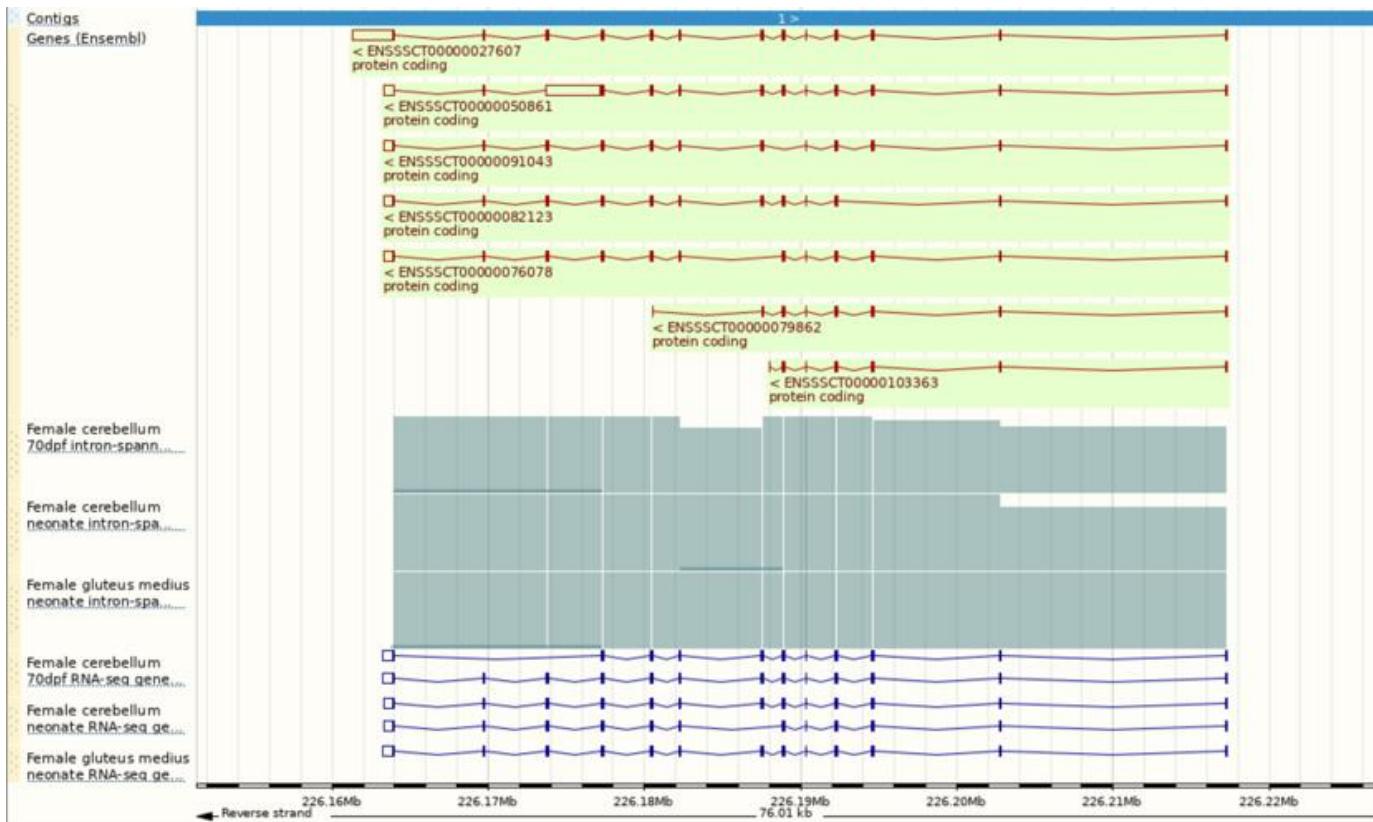
INRAE

PAG 2023 _ FAANG workshop
Elisabetta.giuffra@inrae.fr



➤ GENE-SWitCH Ensembl Collection

- Highly detailed window into the transcriptomes of these species.
- Additional specific GENE-SWitCH analyses performed by consortia's scientists (e.g. ATAC-seq) also made available as loadable tracks to be viewed in their genomic context..

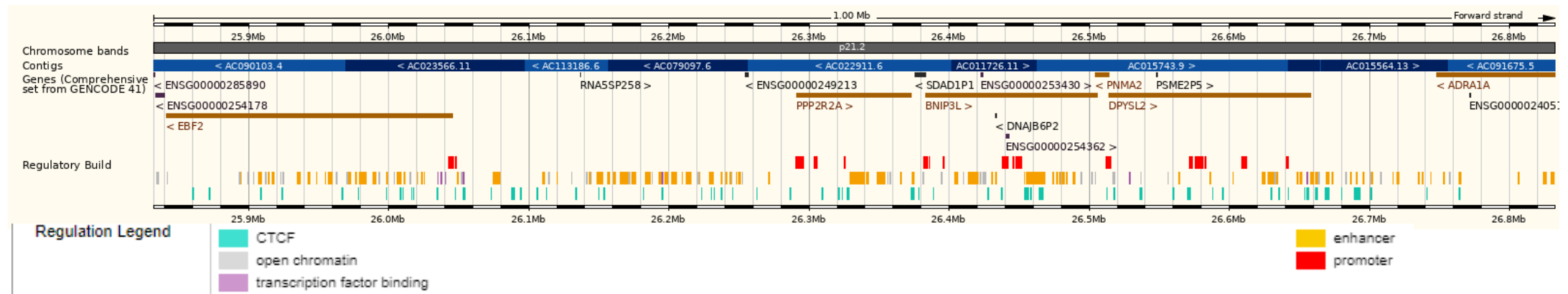


GENE-SWitCH			
Female cerebellum 70dpf	0	0	0
	1	1	1
Female cerebellum neonate	0	0	0
	1	1	1
Female gluteus medius neonate	0	0	0
	1	1	1
Female hindbrain 30dpf	0	0	0
	1	1	1
Female hindlimb muscle 30dpf	0	0	0
	1	1	1
Female hindlimb muscle 70dpf	0	0	0
	1	1	1
Female ileum neonate	0	0	0
	1	1	1
Female kidney 70dpf	0	0	0
	1	1	1
Female kidney neonate	0	0	0
	1	1	1
Female liver 30dpf	0	0	0
	1	1	1
Female liver 70dpf	0	0	0
	1	1	1
Female liver neonate	0	0	0
	1	1	1
Female lung 30dpf	0	0	0
	1	1	1
Female lung 70dpf	0	0	0
	1	1	1
Female lung neonate	0	0	0
	1	1	1



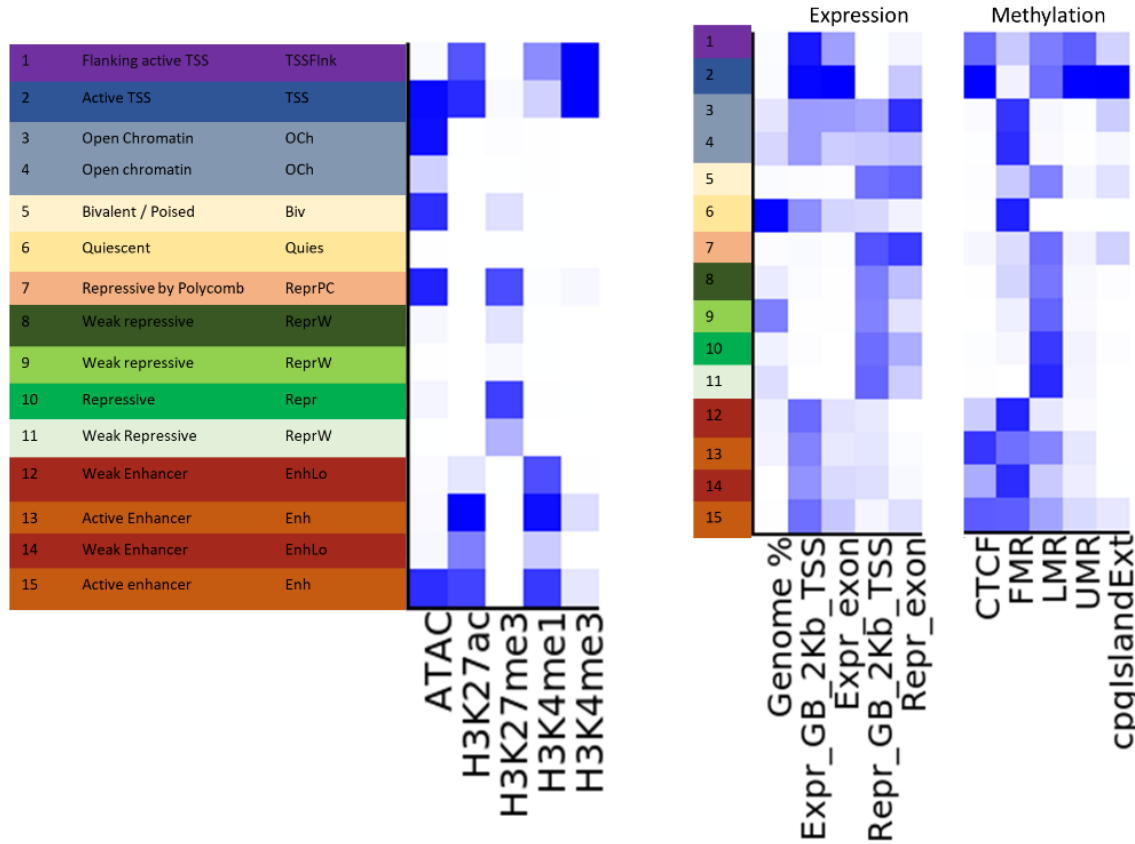
➤ Genome annotation: work in progress

- Establish the Ensembl Regulatory Build for pig genome
- Establish the Ensembl Regulatory Build for chicken genome
- Share regulatory annotation maps (pig and chicken) with WP4 (Aim B)



➤ Integrative data analysis: in progress

- The necessary bioinformatics tools have been tested with public domain mouse data and data from cultured chicken and pig cells (existing partner's data).



Integrative analysis for identification of epigenetic states in chicken cell line

J. de Vos et al. submitted



Jani de Vos
Martijn Derks
Ole Madsen



➤ Aim A: summary of progress

Improved annotation and bioinformatics tools enabling research and genomics application in pigs and chickens

- Bioinformatics pipelines for data analyses developed and released
- Primary analyses completed for all data but ChIP-seq
- First wave of improved annotation of pig and 3 chicken reference genome sequences

A lot of work still ongoing....

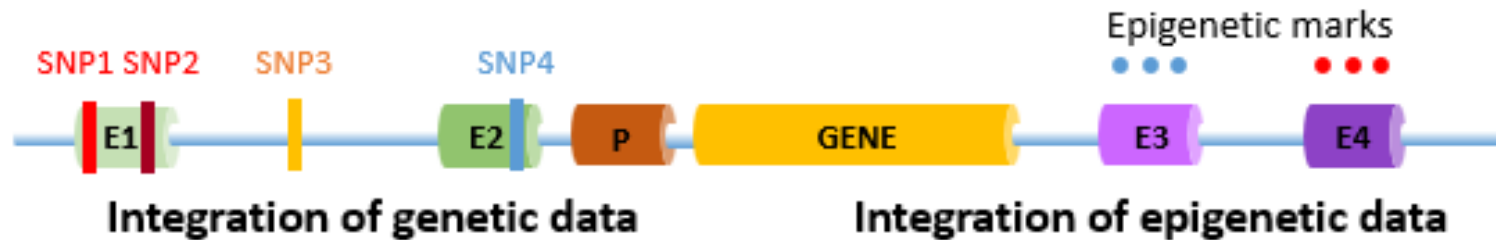
- Undertake integrative and comparative analyses once missing data (ChIP-seq) are available.
- Organizing manuscripts for publication describing switches in usage of regulatory sequences during development.

All of the GENE-SWitCH outputs are available from the GENE-SWitCH FAANG Data portal page (<https://data.faang.org/projects/GENE-SWitCH>) under the terms of the Fort Lauderdale agreement and Toronto Statement



➤ Aim B. Using functional annotation for precision animal breeding

Two different approaches:



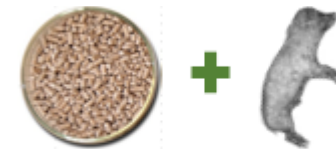
WP4



Genomic Selection

Can functional annotations enhance the prediction accuracy of breeding values in commercial populations?

WP5



Diet X epigenetics

Provide a basis for future studies focused on better farm management (e.g. using lower-quality, more sustainable feed)

➤ **WP4: Improving predictive models for genomic selection**

Led by:



Co-led by:



Aim: Extend genomic prediction models to exploit new annotation maps of pig and chicken generated in GENE-SWitCH.

This is achieved by:

- **Developing new genomic prediction models.**
- **Generating fine-mapped QTL and eQTL.**
- **Validating the models in large scale commercial data.**



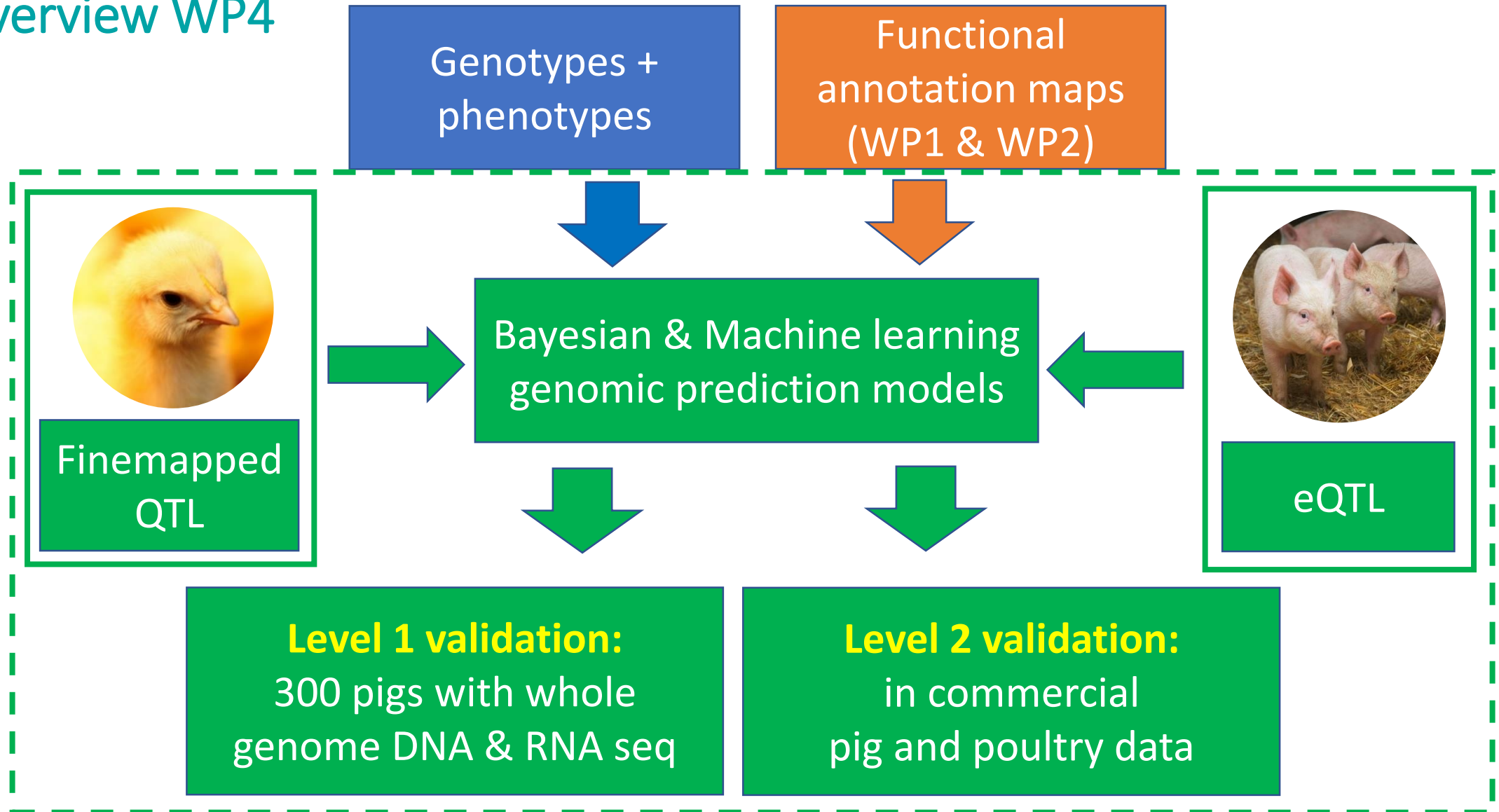
UPPSALA
UNIVERSITET



INRAE

PAG 2023 _ FAANG workshop
Elisabetta.giuffra@inrae.fr


➤ Overview WP4



➤ WP4: achievements

Genotypes + phenotypes

Functional annotation maps (WP1 & WP2)



Finemapped QTL



eQTL

Bayesian & Machine learning genomic prediction models

Level 1 validation:
300 pigs with whole genome DNA & RNA seq

Level 2 validation:
in commercial pig and poultry data

Crespo et al.: eQTLs associations & ASE and isoforms in (300) pigs and 3 tissues
Submitted

Ying, Carlborg, Rubin et al.
in prep.



UPPSALA
UNIVERSITET workshop
Elisabetta.giuffra@inrae.fr



GENE-SWITCH

➤ WP4: achievements



Fanny Mollandin
Andrea Rau
Pascal Croiseau (BovReg)
Helene Gilbert



OXFORD **G3** Genes | Genomes | Genetics

G3, 2021, 11(11), jkab225
DOI: 10.1093/g3journal/jkab225
Advance Access Publication Date: 7 July 2021
Investigation

An evaluation of the predictive performance and mapping power of the BayesR model for genomic prediction

Fanny Mollandin,^{1,*} Andrea Rau ,^{1,2} and Pascal Croiseau ¹

¹INRAE, AgroParisTech, GABI, Université Paris-Saclay, Jouy-en-Josas 78350, France
²BioEcoAgro Joint R
*Corresponding aut
¹Author contribute

<https://github.com/FAANG/BayesRCO>

Mollandin et al. *BMC Bioinformatics* (2022) 23:365
<https://doi.org/10.1186/s12859-022-04914-5> BMC Bioinformatics

RESEARCH Open Access


Accounting for overlapping annotations in genomic prediction models of complex traits

Fanny Mollandin^{1*}, H el ene Gilbert², Pascal Croiseau¹ and Andrea Rau^{1,3}

Two novel extensions of the Bayesian genomic prediction model (BayesRC π , BayesRC+) to account for overlapping functional categorizations obtained from prior knowledge (e.g., GENE-SWitCH functional annotation maps).

- **Modest gains in prediction for some traits when incorporating annotations.**
- **BayesRC π and BayesRC+ allow for improved prediction and prioritization of multi-annotated markers and can provide useful biological insight into the genetic architecture of traits.**

WP4: achievements



G3, 2022, 12(4), jkac039
<https://doi.org/10.1093/g3journal/jkac039>
 Advance Access Publication Date: 15 February 2022
 Genomic Prediction

Prediction performance of linear models and gradient boosting machine on complex phenotypes in outbred mice

Bruno C. Perez ¹, Marco C. A. M. Bink ¹, Karen L. Svenson ², Gary A. Churchill ² and Mario P. L. Calus ^{3,*}

¹Hendrix Genetics B.V., Research and Technology Center (RTC), 5830 AC Boxmeer, The Netherlands,
²The Jackson Laboratory, Bar Harbor, ME 04609, USA, and
³Wageningen University & Research, Animal Breeding and Genomics, 6700 AH Wageningen, The Netherlands

*Corresponding author: Wageningen University & Research, Animal Breeding and Genomics, P.O. Box 338, 6700 AH Wageningen, The Netherlands.
 Email: mario.calus@wur.nl



Bruno C. Perez, Marco Bink
 Mario Calus

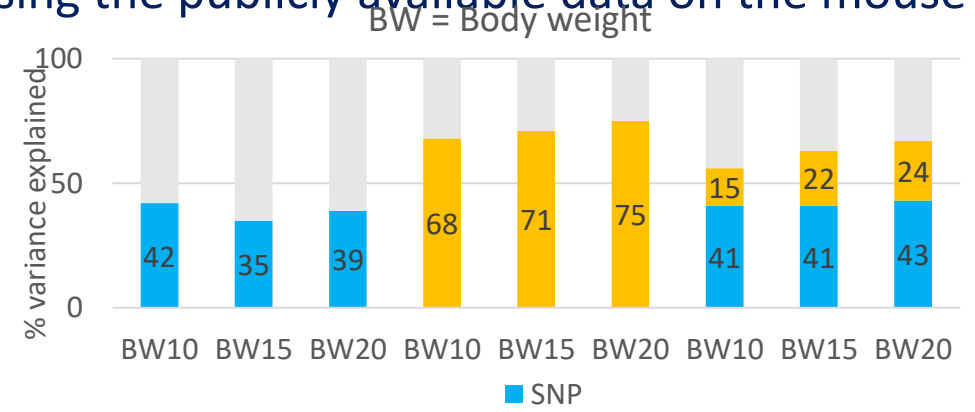



G3, 2022, 12(11), jkac258
<https://doi.org/10.1093/g3journal/jkac258>
 Advance Access Publication Date: 26 September 2022
 Investigation

Adding gene transcripts into genomic prediction improves accuracy and reveals sampling time dependence

Machine Learning approach (Gradient Boosting Machine - GBM) using the publicly available data on the mouse “collaborative cross” (Diversity Outbred Mouse data).

➤ The models with transcriptomics yielded higher proportions of explained variance.



➤ Traits measured closer to the gene-expression sampling date (week 24) had higher proportions of explained variance. This points to time-dependency for gene-expression measurements.

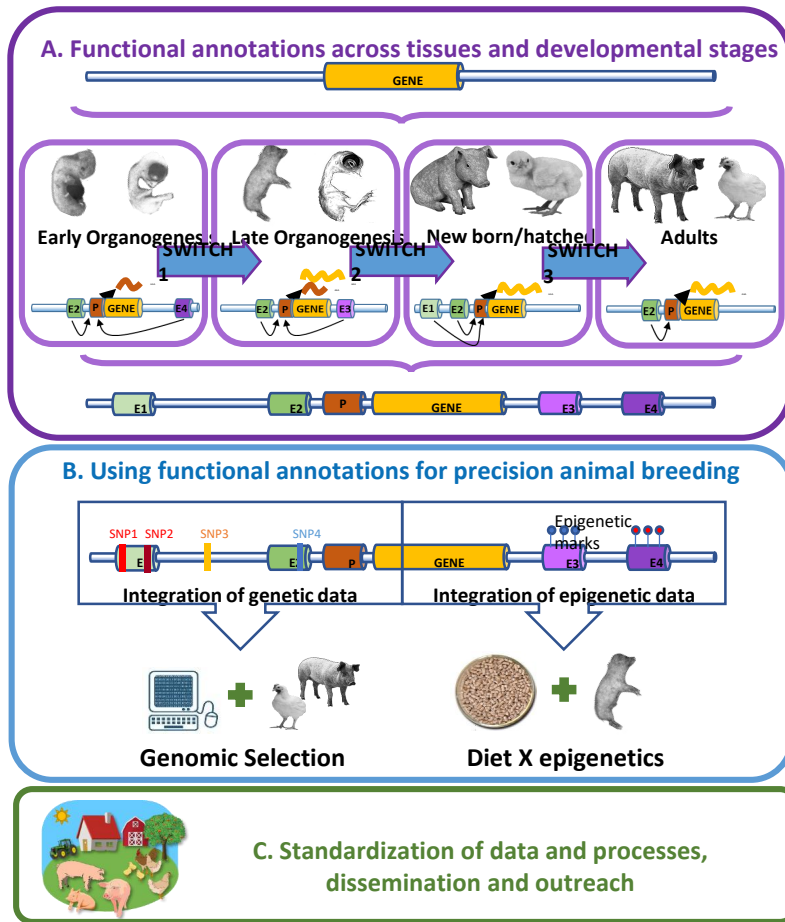
➤ WP4: summary of progress

Timely achievements in the development of new genomic prediction models.

Work in progress

- **Valorisation of results of the eQTL study in pigs and the fine QTL mapping in chickens**
- **Incorporating more annotations (from WP1 and WP2) in genomic models.**
- **Level 1 (300 pigs) versus level 2 (10,000s commercial animals) validation.**

➤ The expected impacts of GENE-SWitCH



- High-quality reference annotation maps for the whole research community.
- Cutting-edge research results: paving the way to further studies and strategies
- European stakeholders benefits: an increased understanding of the value of functional genome annotation to face current and future challenges to achieve sustainable productions.

EuroFAANG provides a great opportunity to increase these impacts in Europe and beyond.

EuroFAANG GENE-SWitCH

workshop: applications to attend online are open until 25th January!

20/12/2022



- Aimed at stakeholders for utilisation of GENE-SWitCH data and the genomic prediction models developed within the project.
- Extensive instruction on use of Ensembl to exploit pig and chicken gene, variation and regulatory annotations.



➤ With big thanks to all participants in GENE-SWitCH



With thanks
also to those
who joined
online

Hybrid annual meeting 2022 hosted by IRTA Torre Marimon