GENE-SWitCH

Newsletter - Issue 1

April 2020

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement N°817998.

Summary

Editorial

By Elisabetta Giuffra, GENE-SWitCH Project Coordinator (INRAE)

The term phenotype - from Greek *phainein*, meaning 'to show', and *typos*, meaning 'type' - indicates the observable traits or characteristics of an individual. Each phenotype results from the expression of the individual's genome (i.e. the whole hereditary information encoded in DNA) and its interaction with the environment. The challenge of the emerging field of predictive biology is to read the information encoded by the genome to predict phenotypes. The overarching goal is to gain the ability to predict complex outcomes as, for example, the efficacy of a drug, the susceptibility of an individual to cancer, or the breeding value of an elite boar or cockerel.

Eukaryotic genomes are extremely complex and at nearly 99% are made of non-coding regions. The whole genome harbors a rich array of functionally significant elements made of DNA as well as of epigenetic modifications (i.e. reversible changes that affect structure and function but not the sequence of DNA) that can be modulated during individual's development and in response to environmental perturbations. Thus, following the achievement of the first human genome sequence in 2004, the challenge became to identify and understand the instructions encoded in the genome and "epigenome". Large research consortia (e.g. ENCODE projects) have generated high-resolution, highly reproducible maps of DNA segments with biochemical signatures associated with diverse molecular functions, and all these data have been openly shared with the community. The availability of these annotated maps for humans and model animals has boosted research, leading the way to a much increased prediction power in human medicine such as facilitated personalized diagnostics, disease management, and biomarker discovery.

Since 2015, the Functional Annotation of Animal Genomes (FAANG) initiative has taken up this challenge for the genomes of domesticated animal species. It is expected that FAANG will boost both fundamental and applied research on farm and aquaculture species (spanning insects, fish, birds and mammals). On the applied side, innovative precise breeding relies on a better knowledge on structure and function of genomes and on their interaction with the environmental components of the production systems (e.g., nutrition), so that genomic selection and management practices can be optimized to improve performance. The functional annotation of these species' genomes will also facilitate the development of innovative guidelines for conservation of endangered populations that harbor genetic variation of potential importance for efforts to address global changes in animal production.

GENE-SWitCH "The regulatory GENomE of SWine and CHicken: functional annotation during development" is a 4-year H2020 project which started in July 2019 (www.gene-switch.eu). Our consortium brings together 11 partners representing pan-European excellence and world-leading animal breeding and biotech industry in a true co-creation effort. Our aim is to deliver new underpinning knowledge on the functional genomes of these two foremost sources of meat, as well as to enable immediate translation of this new knowledge to the pig

and poultry industry. We conduct GENE-SWitCH in the frame of the FAANG initiative. Datasets are being produced to fully conform to FAANG standards under the principles of International cooperation and Open Science, and in coordination and synergy with the other FAANG projects in EU and across the globe.

GENE-SWitCH partners

While some knowledge is available on the function and regulation of the genome in the tissues of adult animals, little is known on how the fetal genome integrates genetic information and environmental constraints. Thus, our first goal (Fig. 1A) is to characterize the functional elements of the genomes of pigs and chicken which are active, poised or repressed ("switches") during development *in utero* and *in ovo.* We have almost completed the collection of samples, and their molecular characterization is due to start soon *(see Fresh news from WP1 "Sample collection and assays-by-sequence"*) and the development and deployment of bioinformatic pipelines to analyse data has progressed well. These analyses will deliver new comprehensive annotated maps of pig and chicken genomes, as well as contribute to unravelling the evolutionary dynamic of putative functional elements and their patterns of conservation and variation by comparative analyses across other species.

Using functional annotations for precision animal breeding is our second goal. This is developed along two different axes: i) the development of innovative genomic predictive models for genomic selection, that will be followed by validation in commercial pig and poultry populations (Fig. 1B, left; *see Scientific Focus: Predictive Models for Genomic Selection*), and ii) exploring how different maternal diets contribute to establish the epigenetic profiles in the pig fetus, followed by the assessment of their possible persistence through the weaning stage (Fig. 1B, right); this kind of information paves the basis for the future improvement of breeding management (i.e. diet which may confer more robustness vs. piglet diseases) in a synergistic way to genomic selection The main animal trial for this "diet x epigenetics" study has successfully started in February 2020.

All activities rely on a large dedicated effort for the standardization of data and processes for the integration of new annotation maps in the FAANG data portal. This is at the heart of the FAANG concept and is reinforced by the ongoing clustering activities (e.g. for optimizing shared bioinformatic pipelines) with other H2020 and international projects. Extensive dissemination, communication and training activities will facilitate uptake of project's outcomes to a large audience of stakeholders (Fig.1C).

While writing this newsletter, the world is being heavily impacted by the COVID-19 pandemics. This crisis underlines once again the importance of innovative therapies, vaccines and improved diagnostics, as well as that saving the biodiversity of species and habitats is essential to reduce the chance of new pathogens to attain new hosts including humans. Ultimately, COVID-19 pandemics confirms the irreplaceable role of scientific research - in which the underpinning knowledge of genomes and their regulation is one of the main assets - to understand new problems and find innovative solutions across several fields. We are all coping as best we can with this situation, and we look forward to the near future with renewed motivations.

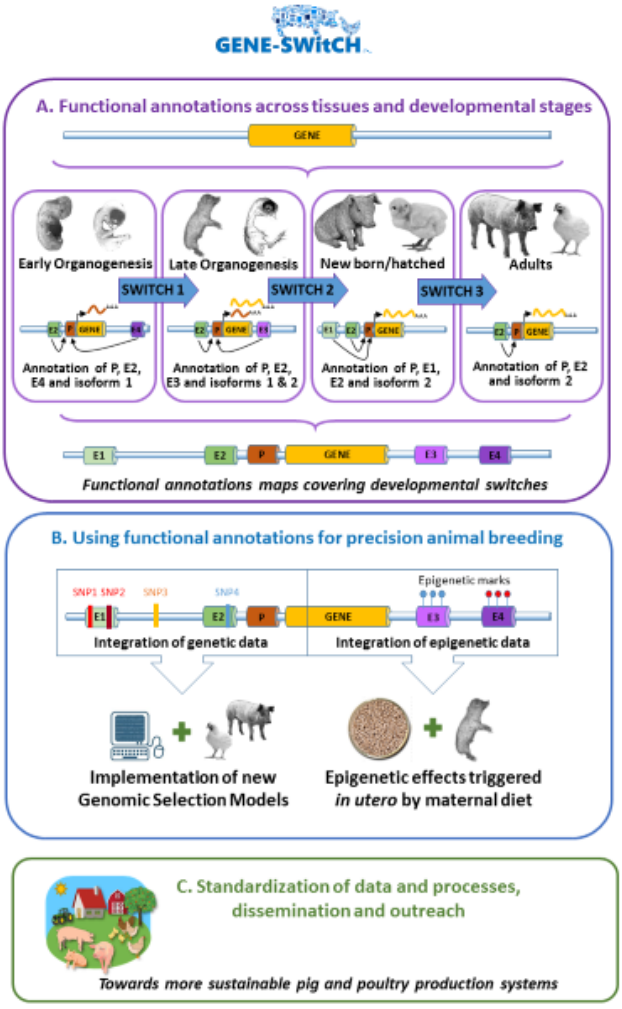

Figure 1. *The overall concept of GENE-SWitCH.*

Literature sources:

Kellis et al. Proc Natl Acad Sci U S A. 2014 Apr 29; 111(17):6131-8.
Andersson et al. Genome Biol. 2015 Mar 25;16:57.
Rexroad et al. Front Genet. 2019 May 16;10:327.
Giuffra, Tuggle et al. Annu Rev Anim Biosci. 2019 Feb 15;7:65-88.

News & Events

GENE-SWitCH Kick Off Meeting (10-11 September 2019)

By Elisabetta Giuffra (INRAE)

On September 10th and 11th, 2019, all project partners gathered together for the kick-off meeting of the GENE-SWitCH project, held in Paris at the AgroParisTech Alumni venue, facing the Louvre Museum. During the kick-off meeting, initial progress since the official start of the project (July 1st, 2019) was evaluated. We discussed and updated the action plans for delivering the sample collections and logistics; assays-by-sequence optimization; metadata and data coordination, curation, validation, archiving processes; as well as refining the transversal plan for standardization, dissemination, and outreach to the animal science community. More importantly, we advanced considerably the design of the "clustering plan" thanks to the presence of the coordinators the other two H2020 projects funded under the same topic call: AQUA-FAANG (on aquaculture fish species) and BovReg (on cattle), respectively. Overall, the meeting was highly productive and took place in a very convivial atmosphere.

ISAG 2019

At the 37th International Society for Animal Genetics (ISAG) Conference in Spain in July 2019, GENE-SWitCH project coordinator Elisabetta Giuffra (INRAE) presented "Introduction to FAANG (Functional Annotation of Animal Genomes" - Goals and Opportunities". Giuffra explained the FAANG Data Coordination Centre and FAANG Data Portal and presented 3 projects of the FAANG initiative BovReg, AQUA-FAANG and GENE-SWitCH. Giuffra's presentation covered the overall objective, concept and the consortium of GENE-SWitCH. The presentation is available on GENE-SWitCH's website.

GENE-SWitCH shines at PAG XXVIII Conference!

As every year, the International Plant and Animal Genome Conference (PAG XXVIII) took place in San Diago, CA, USA between 11-15 January 2020. PAG provides a forum on recent developments and future plans for plant and animal genome projects.

GENE-SWitCH project coordinator Elisabetta Giuffra made a poster presentation at the conference providing information about the aims, first achievements and perspectives of the GENE-SWitCH project. Peter Harrison (EMBL-EBI) presented "The FAANG Data Coordination Centre: New European Perspectives for Our Continued Global Effort", explaining the development of FAANG DCC and FAANG Data Portal. Peter also gave a presentation about "A Vision for Bioinformatics within the Global FAANG Project" conceived with Mick Watson (Roslin Institute). Peter explained the development of bioinformatics pipelines across the entire set of FAANG projects, based on the principles of open science, open source code and reproducible workflows and environments. The poster and the oral presentations are available on GENE-SWitCH's website.

FAANG-Europe COST Action Conference

The 2nd FAANG-Europe COST Action Conference on Functional Annotation of Animal Genomes (FAANG) took place in Prague, Czech Republic between 11 – 13 February 2020. The GENE-SWitCH project was outlined in a presentation by Elisabetta Giuffra. Peter Harrison from EMBL-EBI gave a talk about "The FAANG Data Coordination Centre: New European perspectives for our continued global effort", representing EMBL-EBI's Data Coordination Centre work on GENE-SWitCH, AQUA-FAANG and BovReg projects.

FAANG Shared Workshop: "Foundation for the Future Agenda"

FAANG shared workshop was held at the Rosalind Franklin Pavilion, Wellcome Genome Campus Conference Centre, Hinxton, UK on the 25th-27th February 2020.

The aim of the workshop was to establish coordination of plans for cross project (species) comparative analysis, standardised metadata, standardised FAANG analysis workflows/pipelines, planning for future shared dissemination, training and management meetings. An overview of GENE-SWitCH and the core annotation workplan ("GENE-SWitCH project overview, approach to bioinformatics and comparative analysis") was given by Sarah Djebali (INSERM, France).

This meeting was an excellent opportunity for cross project networking, with concrete actions such as the establishment of common working groups for bioinformatic pipelines validation and comparative genomics.

Fresh news from WP1 "Sample collection and assays-by-sequence"

By Herve Acloque (INRAE)

Sampling Strategy

Work Package 1 from the GENE-SWitCH project aims at (1) establishing a collection of cryopreserved tissues from pig and chicken at three developmental stages (four biological replicates: two males and two females), and (2) performing core molecular assays required for the functional annotation of the pig and chicken genomes on seven different tissues for each developmental stage.

The sampling strategy is summarized in Figure 1. Seven tissues (cerebellum, skin, kidney, skeletal muscle, lungs, liver, small intestine) have been selected for the FAANG core molecular assays. Six additional tissues (brain cortex, heart, spleen, gonads, stomach, large intestine) will be bio banked.

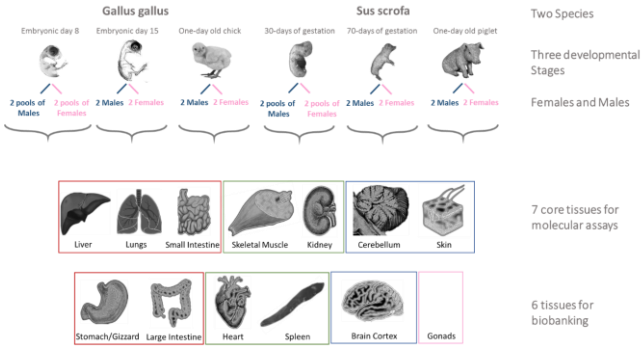

***Figure 1**: Rationale for the GENE-SWitCH sampling strategy*

Sampling of pig tissues was finished end of January, thanks to a nice team of collaborators from different INRAE laboratories and experimental units (GABI lab from Jouy en Josas, GenESI unit from Lusignan and Le Magneraud, PEAO unit from Tours). Protocols and metadata will be available soon on the FAANG data portal.

***Photo**: Our highly motivated experts ready for pig sampling!*

Sampling of chicken tissue has been achieved for the embryonic stages and is planned to be completed after the Covid-19 lockdown in Scotland for hatched chicks. The experts and nice team of Scottish embryologists from the Roslin Institute lent us their golden hands to perform precise dissections.

What's next for WP1?

- Submit the metadata and protocols to the FAANG data portal
- Complete DNA/RNA extraction for all the tissues
- Prepare the tender for RNA-seq and methylome assays
- Achieve standard protocols for ATAC-seq and ChIP-seq from frozen tissues

Many thanks to all our contributors and more news in the coming months!

	E8 chick embryo	E15 chick embryo	One-day old chick	D30 pig fetus	D70 pig fetus	One-day old pig
Sampling	Done		Planned	Done		
Metadata	To be completed			To be submitted to FAANG DCC		
DNA/RNA extraction	Planed			To be sent to the Partners		
mRNA-seq	Due date M18 (December 2020)					
Small RNA-seq	Due date M18 (December 2020)					
Iso RNA-seq	Due date M18 (December 2020)					
WGBS/RRBS	Due date M12 (June 2020). Tender in process					
ATAC-seq & ChIP-Seq	Due date M18 (December 2020) for ATAC-seq and M24 (June 2021) for ChIP-seq Optimization of procedures started					

***Figure 2**: Current Progress for Work Package 1*

Scientific Focus: Predictive Models for Genomic Selection

By Fanny Mollandin and Andrea Rau (INRAE)

What is genomic selection?

Within the past century, advances in agronomy and science have led to the development of sophisticated evaluation techniques for livestock that have contributed to marked improvements for specific traits of agricultural and economic interest. In the mid-1900s, with the introduction of the Best Linear Unbiased Predictor (BLUP) to predict and estimate breeding values of candidate sires, pedigrees came to be seen as a powerful tool for breeders to drive improvements on a genetic level. More recently, the widespread availability and decreasing costs of high-throughput genotyping and genomic sequencing technologies have paved the way for genomic evaluation methods that have accelerated the successful implementation of evaluation in livestock breeding for many species.

How do genomic prediction models work?

Among existing state-of-the-art genomic evaluation methods, several different approaches have been proposed. All share a common objective, namely to accurately estimate a phenotype or estimated breeding value as a combination of the effects of a set of single nucleotide polymorphisms (SNPs), which represent genetic sequence variation differentiating animals at a single nucleotide position (see example in Figure 1).

What are some of the state-of-the-art models currently in use?

One of the most widely used methods, Genomic Best Linear Unbiased Predictor (GBLUP), uses a linear combination of SNPs to obtain predictions and has the advantage of a fast computing time. As an alternative, the suite of nonlinear models in the so-called Bayesian alphabet represents an attractive approach, due in part to their flexibility and ability to automatically remove a considerable portion of non-informative SNPs from the model. An additional advantage in these Bayesian models is that they make use of *a priori* information within the model framework itself, which naturally facilitates the incorporation of known biological information when available. Although the prediction performances of these models are comparable to, or in some cases better than, linear models, their primary limitation lies in their computational complexity.

The promise of incorporating biological knowledge into genomic prediction models

Despite the initial hope that the use of whole genome sequencing (WGS) data would lead to significantly improved predictions thanks to the inclusion of true causative SNPs, prediction reliabilities in fact tend to be equivalent or worse than those from lower density genotypes, possibly due to the inclusion of a very large number of "noisy", or non-causative, SNPs. One potential avenue for improvement thus lies in the prioritization of potentially causative SNPs from WGS data. To this end, several international actions and projects, including GENE-SWitCH, have recently started to focus significant efforts to better characterize the intermediate functional actors connecting genotypes to quantitative phenotypes. In particular, the goal is to complement the wide availability of genotyping and genomic sequencing data with functional annotation data, such as gene expression and chromatin accessibility in relevant tissues and developmental stages, to better identify causal SNPs.

Targeting predictive power and interpretability in genomic prediction

In Fanny Mollandin's PhD research, our goal is to develop and validate Bayesian genomic prediction models able to weight SNPs according to information extracted from these functional annotations. In particular, we simultaneously aim for strong predictive power and better interpretability of the results. In this objective, we will start from two existing methods, BayesR (Erbe et al, 2012) and BayesRC (MacLeod et al, 2016), in which SNP signals are modelled as having a null, small, medium, or large effect. Although both of these models have been shown to improve prediction accuracy in some cases, a first open question we are seeking to address through an extensive set of simulations concerns the interpretability of the *a posteriori* SNP classifications obtained from these models. Results from these analyses will subsequently help inform the optimal strategy to be employed to incorporate *a priori* knowledge, such as that from the functional datasets, to best prioritize potentially causal SNPs in the model.

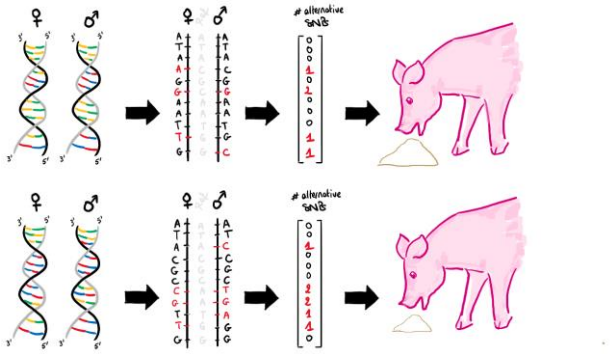

Figure 1: *Using genomic data for phenotype prediction. Each pig has a different genotype, which represents a combination of alleles inherited from its two progenitors. Using a reference genome as a baseline, we can identify the specific positions where genetic variations occur for each pig. These variations can be mathematically expressed as vectors indicating the number of polymorphisms (0, 1, or 2) observed at each position, which in turn will be used to construct a model to predict a quantitative phenotype, for example feed efficiency.*

For further reading:

Erbe et al. (2012) Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. Journal of Dairy Science, 95(7):4114-29. doi: 10.3168/jds.2011-5019.
MacLeod et al. (2016) Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. BMC Genomics, 17:144. doi: 10.1186/s12864-016-2443-6

Fanny Mollandin obtained her Master's degree in Applied Mathematics for Life Sciences at the University of Paris Descartes in 2019, and she started her PhD research at INRAE in Jouy en Josas, France, in October 2019. Fanny's PhD work is co-supervised by Dr. Andrea Rau and Dr. Pascal Croiseau at INRAE.

In the spotlight: GENE-SWitCH Coordination Team

Dr. Elisabetta Giuffra, INRAE

Project coordinator, WP7 Leader and WP5 Deputy Leader-Influence of maternal diet on epigenetic programming of offspring

E-mail: elisabetta.giuffra@inrae.fr

Dr. Elisabetta Giuffra (https://orcid.org/0000-0001-9568-2056) works as a senior scientist at INRAE_GABI unit since 2015, with a focus on the genomics of host-virus interaction in pigs. She has launched and co-coordinated the Fr-AgENCODE project (2014-2017) and has been Strategic Board Advisory member for the FAANG Data Coordination Centre project led by EMBL-EBI (2016-2019). She co-leads the FAANG Steering group and the ASA Committee of FAANG (www.faang.org) and has been Vice-Chair of the FAANG-Europe COST Action CA15112 (http://www.faang-europe.org). In GENE-SWitCH she leads the project Management (WP7) and co-leads WP5.

Dr. Hervé Acloque, INRAE

Project co-coordinator, WP1 Deputy Leader-Sample collection and assays-by-sequence

E-mail: herve.acloque@inrae.fr

Dr. Hervé Acloque is a cellular and molecular biologist at INRAE_GABI unit. He has a strong and recognized background in the field of animal pluripotent stem cells and embryonic and fetal development of vertebrate species, including pig and chicken. Since 2015, he is also an active contributor of the FAANG international initiative. He has been work-package leader in FR-AgENCODE to supervise molecular assays and to develop Hi-C procedure on animal tissues. He was also member of the management committee of the COST action FAANG-Europe CA15112. In GENE-SWitCH he will co-lead WP1 with DIAGENODE, contribute to WP4 (T4.3) and to project Management (WP7).

Camille Bénard, INRAE Transfert (IT)

WP7 Deputy Leader—Project management and consortium coordination

E-mail: camille.benard@inrae.fr

Camille Bénard is a project manager at INRAE Transfert. She holds a PhD in agricultural sciences. Since 2017, she is involved in the management or proposal setting up phases of several European collaborative and European or national infrastructure projects. In GENE-SWitCH relaying on INRAE Transfert expertise, she helps the smooth running of the project and supports the Coordinator, the governance bodies and the whole consortium partners, with administrative and financial issues, and meetings logistics. She also updates the project collaborative platform, an internal project repository to facilitate information and document sharing and storage between the project partners.

Upcoming Events

EVENT	DATE	LOCATION
GENE-SWitCH Annual Meeting	08-09 July 2020	Barcelona, SPAIN
EAAP 2020	31 August - 04 September 2020	Porto, PORTUGAL
ENCODE WORKSHOP	01-03 October 2020	Barcelona, SPAIN

For more information, visit our website and follow us on social media!

@GeneSwitch **@GENESWitCHEU**

gene-switch www.gene-switch.eu