GENE-SWitCH

Newsletter - Issue 3

July 2021

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 Elisabetta

GENE-SWitCH has just reached the turning point of the second year. We are proud to say that, despite the difficulties imposed by the pandemic for more than one year, we are on good schedule with the workplan and excited to have entered the core of activities in all work packages.

In this issue, Dr. Wassim Lackal (Research Scientist at Diagenode) explains the "ChIPmentation" technique that is being used in GENE-SWitCH (WP1) to generate ChIP-seq data for several epigenetic marks, from different pig and chicken tissues at different developmental stages (see page 7. We also introduce Dr. Daniel Crespo-Piazuelo (postdoc at IRTA since Dec. 2020) who explains the conceptual basis of the eQTL approach in WP4 (page 4), followed by a brief update of the good progress made by the WP4 taskforce. More people have joined/are now joining our consortia for WP2 and WP5 and will be introduced soon.

Communications and dissemination activities (WP6) increase and diversify in pace with the project's progress. We strongly believe in the mission of science towards all stakeholders and society as a whole. We are proud to have two EMABG members in our Scientific Advisory Board and look forward more interactions for training of young students and future scientists (see page 3). At page 2 we present the video (relaunched by our twitter channel) by Jani de Voos - PhD student, Wageningen University - where she explains her thesis work. New videos are being produced, keep tuned with our channels!

Finally, EuroFAANG is further developing (page 2) and will be actor at next international occasions (the first being EAAP 2021). We hope it will further expand and we look forward the several opportunities for research, teaching and dissemination that this "node" of the global FAANG framework will allow achieving for European research.

All this and more will be discussed and enriched at our second Annual Meeting (5-6-7 July), that will be still held online for safety reasons. We got used and now handle quite well virtual communications, but we definitely look forward to travel again!

News

GENE-SWitCH Fact Sheet now comes in different languages

Our fact sheet is an excellent tool to easily explain what we are working on with GENE-SWitCH. The factsheet is built up of different sections explaining the background, the project results, and the impact. On top of that, the fact sheet introduces eight different UN goals that can all be directly related to what GENE-SWitCH is working towards. With collaboration from all the project partners, we have managed to translate the factsheet into six different languages. This is a huge asset in terms of communicating to different stakeholders and gaining more engagement on a national level.

The translated fact sheet is also a strong communication tool in terms of interacting more with the press and helping to simplify the area we are researching in. The different fact sheets have been communicated via social media in order to make people aware that they exist. We kindly invite everybody to engage with the factsheet and share them amongst your network in your preferred language. The fact sheet is currently available in English, French, Dutch, Italian, Spanish and Swedish. We are still open for more translations.

Find the different factsheets here.

EuroFAANG webpage has been launched

As announced in the previous newsletter (February 2021), the EuroFAANG page has officially been launched and is now accessible via AQUA-FAANG, BovReg and GENE-SWitCH website. EuroFAANG brings together a wide range of expertise in farmed animal biology and breeding, genomic, bioinformatics, modelling and open data, as well as multiple platforms for dissemination and outreach, with a common goal to discover links between genome and phenome (i.e. G2P) in the frame of the FAANG to Fork strategy. GENE-SWitCH. We are very excited about this close collaboration and invite everybody to stay updated on news and other relevant information via the different project's main websites and via social media.

EuroFAANG: AQUA-FAANG, BovReg and GENE-SWitCH will attend the 72nd EAAP Annual Meeting held in Davos from 30th August to 3rd September 2021.

The three projects have organized a joint session on 30th August during which project representatives will present their latest activities and achievements.

Video interview with a young researcher

It is very important for us to differentiate our communication actions for GENE-SWitCH to create engagement with different stakeholders and communicate our research. We have started a new video initiative called "Interviews with Young Researchers". This is a series where PhD students explain their area of study and what they are working on in a language that is understandable for people outside the bubbles of scientific exchange. In our first video we meet young scientist Jani de Vos from Wageningen University in the Netherlands. Jani's PhD is part of the GENE-SWitCH project, aiming to explore the regulation of genes for pigs and poultry. Jani explains her PhD in a 2.15 min long video and illustrates how and what specifically she is working on. In the video Jani uses "Dynamic Switches" as a metaphor to explain how genes are expressed. Jani explains how she is looking into the embryo of a chicken and the foetus stage of a pig whilst comparing this to the adult species. Jani explains the data, her project and the aim in a very good way and the video had more than 100 views after only one week. The video was shared on social media channels and is also available on the

main website. Overall the conclusion of these type of videos is that they are a strong tool and an effective way for us to engage with new stakeholders and encourage young graduates to continue their career with EU funded projects in Horizon 2020 areas.

Watch the video here.

European Master's in Animal Breeding and Genetics (EMABG) course announcement

We are very pleased to have two leaders of EMABG in the Scientific Advisory Board of GENE-SWitCH as we work to establish fruitful interactions with programs working in the same research area. The focus of the European Master in Animal Breeding and Genetics (EMABG) is on the development of sustainable breeding programs for farm animals, fish and companion animals. The program is designed to answer the scientific, practical, and societal challenges of animal breeding and genetics. It offers high quality training in the field of animal breeding and genetics. offered jointly by a Consortium of six leading European universities: University of Natural Resources and Life Sciences, Vienna, BOKU (Austria) as Coordinator, Wageningen University, WUR (the Netherlands), Georg-August-University Goettingen, UGOE (Germany), Institut des sciences et industries du vivant et de l'environment (AgroParisTech/Université Paris-Saclay), APT (France), Swedish University of Agricultural Sciences, SLU (Sweden) and The Norwegian University of Life Sciences, NMBU (Norway).

Specializations are offered in one of four areas:

- Breeding programs in low-income countries (WUR – BOKU)
- Biological and societal context of breeding (WUR – SLU or UGOE – SLU)
- Genomic selection and precise phenotypes (WUR – NMBU or UGOE – NMBU)
- Integrative biology (UGOE – APT)

The EMABG is an Erasmus Mundus Joint Master Degree (EMJMD), co-funded by the Erasmus+ Program of the European Union. The scholarship covers all university fees, travel costs, insurance for the two years of studies and an additional € 1000 per month for living costs. The MSc course is open for motivated students with a background in animal sciences, biology or related fields, who would like to specialize in the field of animal breeding and genetics. We welcome applicants of any nationality, economic or social background who meet the admission requirements. The EMABG application period for student intakes in 2022/2023 (start of studies in August 2022) will begin on October 1st. Applications are open until January 15th 2022.

All information, admission criteria and application guidelines can be found on https://www.emabg.eu/

In case of questions, please contact the EMABG Central Office at emabg@boku.ac.at

Identify regulatory regions in the pig genome *by Daniel Crespo (IRTA)*

Daniel Crespo Piazuelo is a Doctor in Veterinary Medicine by the University of Zaragoza (2013). During 2013 and 2014 he did a MSc in Molecular Biotechnology at the University of Barcelona (UB), and a MSc in Bioinformatics at the Autonomous University of Barcelona (UAB). In 2018, he received his PhD degree in Animal Production (UAB). From 2018 to 2020, he worked as a postdoctoral research associate in the Pig Development Department at Teagasc in Ireland. He joined Institut de Recerca i Tecnologia Agroalimentàries (IRTA) in December 2020 as a postdoctoral researcher working under the supervision of Maria Ballester (IRTA) in activities related to WP4 in GENE-SWitCH.

Introduction to eQTLs

In the last 10 years, the reduction in costs of sequencing the whole genome of an individual made genomic selection a powerful tool for animal breeding. This tool takes advantage of the millions of genetic variations (polymorphisms) that are present in the genome of an individual and makes it different from the rest. Based on the idea that the differences in the phenotypes of interest (e.g., carcass weight, coat color, or levels of white blood cells) are driven by these genetic variants and as such, they can be transmitted to the next generation. Sometimes, a polymorphism falls inside a gene region that modifies the protein that its encoded by this gene, making it clearly visible on the phenotype. However, the vast majority of polymorphisms that are associated with the traits of interest aren't located in the coding regions of a gene. These other regions of the genome can regulate the activity of genes and can therefore have an impact over the phenotypes. The genetic variants that are significantly associated with a trait of interest are called quantitative trait loci (QTL), and they are named expression-QTLs (eQTLs) if the association is found between the polymorphism and the expression of a gene. In addition, we talk about cis-eQTL if the variant is associated to the expression levels of a nearby gene, and trans-eQTL if the variant is far away from the gene or even in another chromosome.

Identifying eQTL in key porcine tissues

Inside the GENE-SWitCH project, we are in charge of identifying eQTLs in three porcine tissues (i.e., small intestine, liver and muscle). For that purpose, we sequenced all expressed messenger RNA sequences from the genes of 300 pigs from three different breeds (Duroc, Landrace and Large White) in those three tissues. We also generated the whole genome of the 300 pigs to extract all the genetic variants within their genomes. Using these two datasets, we are currently conducting genome wide association studies between the expression (eGWAS) of genes and genetic markers. This method has the advantage that can be easily parallelizable, as the associations are performed individually, reducing computing times considerably. For example, in the last eGWAS we conducted for the genes expressed in muscle, we assessed the association between more than 13 thousand genes and 34 million mutations, which is more than 475 billion combinations! A clear example of the analysis for the expression of one gene can be seen in **Figure 1**, which is called a Manhattan plot due to the peaks resembling skyscrapers.

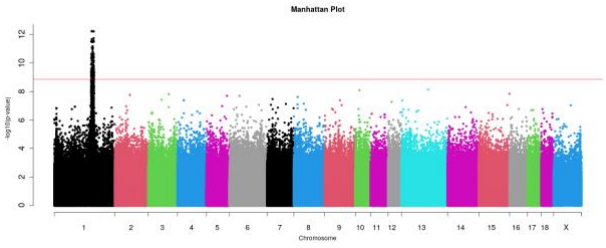


Figure 1: Example of Manhattan plot for the expression of a gene. Points above the red line are significantly associated polymorphisms.

The higher the peak, the greater the statistical significance of the association. Surrounding polymorphisms to the most associated one are less significant, but greater than the rest of the variants. This is based on the premise of linkage disequilibrium, in which the closer two polymorphisms are within the same chromosome, the easier to be inherited together, because the chance of recombination between them is lower. Nonetheless, eGWAS has the drawback that it does not consider the position of the other polymorphisms across the genome, so additional methods for dealing with the linkage between variants will be required. In addition, we will also report regions of the genome that regulate multiple genes, also called hotspot regions. Altogether, the identified eQTLs in the three tissues from our work will be included in the genomic prediction models developed during the course of the project to increase the accuracy of genomic evaluation of target phenotypes, on the basis that gene expression data can be considered as "intermediate phenotypes", so it is expected to be more closely related to genetic data than the phenotypes.

WP4 is making good progress

By the WP4 project team

WP4, entitled "Improving predictive models for genomic selection", is developing and validating new genomic prediction models that use functional information in addition to the commonly used phenotypic and genotypic data, aiming to improve prediction accuracy and persistency of predictions over time. The rationale and coherence between the tasks in WP4 are illustrated in **Figure 1**. In the following, we provide a brief update from WP4.

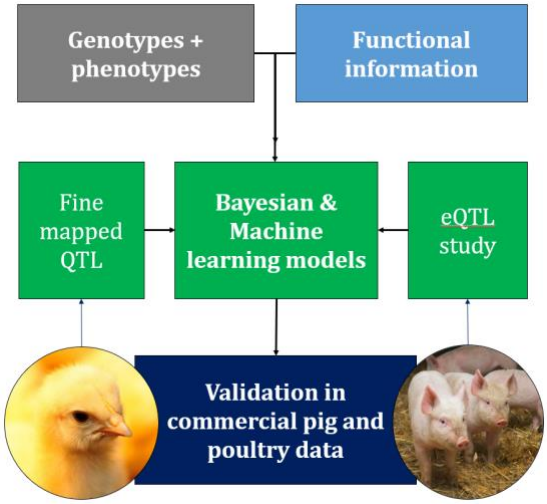


Figure 1: Rationale of Work Package 4 where two tasks focus on deriving new statistical models (T4.1 & T4.2); one task on expression-QTL in pigs (T4.3); one task on fine-mapping of QTL in broilers (T4.4); and two tasks on the validation of new models on commercial breeding data (T4.5 & T4.6).

Bayesian models

Bayesian models allow including functional information as prior information in the model. In the 1st GENE-SWitCH newsletter, Fanny Mollandin and Andrea Rau gave a description of their work on their development and validation of Bayesian models. Initial activities empirically evaluated the performance of the BayesR model for simultaneous phenotypic prediction and QTL mapping based on simulated datasets with a variety of genetic architectures and phenotypic heritabilities (Mollandin et al., 2020). This study lay the groundwork for the development of extensions to the related BayesRC model, which incorporates disjoint prior categorizations of variants to guide genomic prediction. In Task 4.1, two models (BayesRC+ and BayesRCπ) were recently developed to incorporate multiple complex, potentially overlapping prior categorizations of variants, such as those that will be constructed based on WP1 and WP2 annotation data. These models are implemented in a documented Fortran software package called BayesRCO (see **Figure 2**), now available on GitHub: https://github.com/fmollandin/BayesRCO

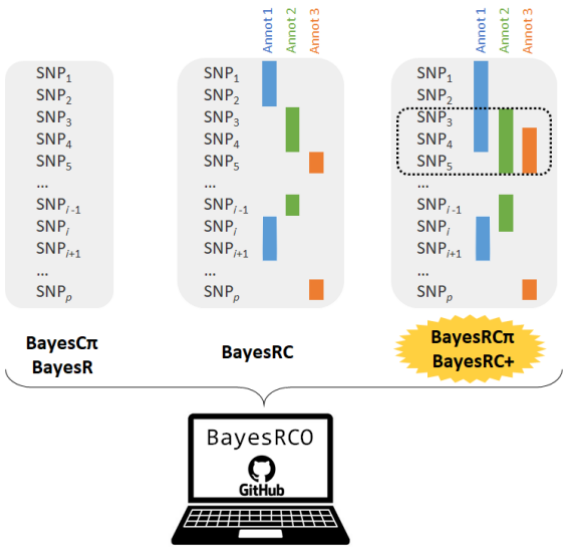


Figure 2. BayesRCO implements five Bayesian hierarchical models for complex trait prediction: two that are blind to prior categorizations of variants (BayesCπ, BayesR), one that incorporates disjoint prior categorizations (BayesRC), and two novel extensions (BayesRCπ, BayesRC+) able to handle more complex, overlapping prior knowledge.

Machine Learning models

Task 4.2 develops models with the same aim as Task 4.1, but uses machine learning models that were chosen for their flexibility in including different sources of data. In the 1st GENE-SWitCH newsletter, Bruno Perez described the background of his work in the Young researcher section. The first activity compared the machine learning model gradient boosting machine (GBM) to the Elastic Net, and the traditionally used linear models GBLUP and BayesB. Analyses were performed in collaboration with the Jackson lab in the US on mouse data, generated for high resolution mapping of more than 50 traits. For polygenic traits, GBLUP and BayesB tended to outperform GBM, while GBM was superior for traits affected by epistatic effects. Following up from this, predictions were performed using genomic or gene expression data or both. Results indicated that for most traits, individual gene expression data was able to explain more phenotypic variance than genomic (SNP) data, while predicted phenotypes were generally more accurate when based on gene expression (TBLUP; Figure 3), compared to being based on SNP genotypes (GBLUP; **Figure 3**).

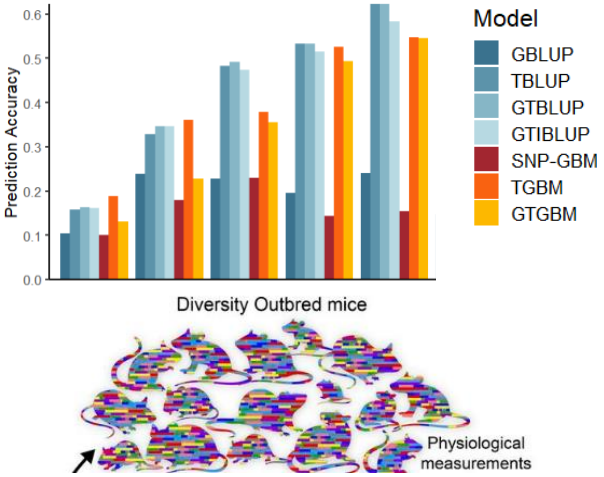


Figure 3. Accuracies of predicting phenotypes in mice for bone mineral density (bmd) at two time points, and body weight (bw) at 10, 15 and 20 weeks of age.

Gene expression QTL in pigs

Task 4.3 has generated functional information based on gene-expression data (RNA-seq) on three experimental pig populations of 100 animals each (**Figure 4**) to identify e-QTL in three key porcine tissues: small intestine, skeletal muscle, and liver. Quality control on the gene expression data has been completed, showing that the data is of good quality. Preliminary results of e-QTL analysis using transcriptome data of muscle and genetic markers (SNPs and indels), generated by WGS and distributed throughout the pig genome, identified key regulatory markers in common for the three breeds influencing gene expression levels. A greater number of trans- compared with *cis*-acting variants were identified. Additionally, regulatory hotspots governing the expression of multiple genes were found. For instance, a 3'UTR variant in the Calmodulin gene, which encodes a muscle regulatory protein, was associated with the expression variation of 1,387 genes. In the Young researcher section of this 3rd newsletter, Daniel Crespo-Piazuelo describes in more detail the background of this work.

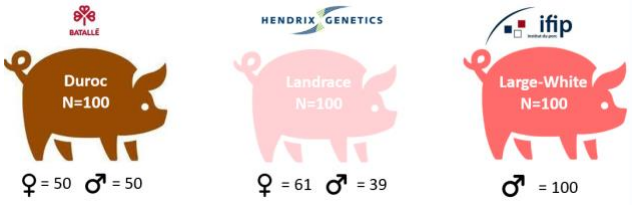


Figure 4: Overview of the origin of the pigs used for RNA-seq in Task 4.3.

Fine mapping QTL in broiler chickens

Task 4.4 focusses on fine-mapping of QTL in chickens for body weight at 8 weeks of age. This task relies on data of an advanced intercross line between two chicken populations divergently selected for this trait from a common base-population. A novel fine-mapping approach is developed and used, in which haplotypes are traced back to the ancestral founder populations of the White Plymouth Rock (WPR) chicken breed. This fine-mapping approach using ancestry tag-SNPs, is expected to provide a resolution to ancestry blocks of ~10-100 kb that can later be evaluated in other experimental and commercial chicken populations originating from the WPR or its ancestors. Most of the analyses are mid-way and the majority of results will be delivered in the coming year.

Validation in commercial breeding data

We are on the verge to commence the two tasks pertaining to the empirical validation of the improved prediction models, combining the functional annotation and experimental data generated in the project, with large scale data from two commercial breeding programs. These tasks are very important to demonstrate the added value of newly generated functional annotation data within GENE-SWitCH to practical breeding programs. In the next two years, we will keep you informed about the results!

Further reading

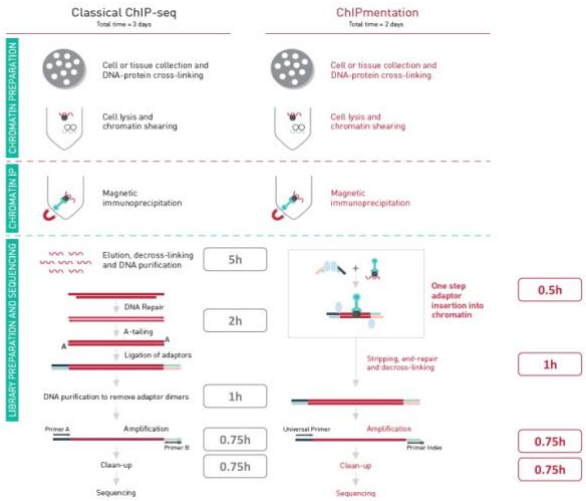
Mollandin, F., Rau, A., and Croiseau, P. (2021) An evaluation of the predictive performance and mapping power of the BayesR model for genomic prediction. G3: Genes, Genomes, Genetics, in press.

In the spotlight

Chipmentation technology

By Wassim Lakhal

ChIPmentation is an exclusive Diagenode technology for studying **Epigenetic** regulation in living organisms (plants, animals, Human …). **Epigenetics** is the science studying the change of the phenotype that does not involve alterations in the DNA sequence but are subject to environment factors (climate, diet, etc.). ChIPmentation is a method that combines **chromatin immunoprecipitation (ChIP)** and **library preparation** in the same experiment resulting on an easier and shorter protocol.



Chromatin immunoprecipitation is a technic that studies the interaction of DNA with regulatory proteins like histones and transcription factors. This method can be followed by quantitative PCR (we talk about ChIP-qPCR) or by **Next Generation Sequencing** (we talk in this case about **ChIP-seq**). The GENE-SWitCH consortium is rather interested by the second type which is ChIP-seq because the generated sequencing data provides usually generous biological information which is usually analyzed by bioinformaticians and exploited by researchers to understand various phenomena. However, before the sequencing, we need to prepare what we call "**Libraries**" which are target DNA molecules that contain a **barcode/adaptor** DNA sequence. The experiment needed to add these barcode

sequences is called **Library preparation**. Standard **library preparation** techniques require usually many ligation steps and can be long and delicate. Here comes the interest of ChIPmentation method on which the **Library preparation** step is done right after the **Chromatin immunoprecipitation** in the same experiment. What allows this shortcut is an Enzyme called **Tn5 transposase.** In nature, this Enzyme is known to cut/paste DNA fragments from a location to another in the genome. For example, some Bacteria like Escherichia will acquire antibiotic resistance thanks to Tn5. The Tn5 used in ChIPmentation, is engineered to insert the **barcode/adaptor** DNA sequence into the chromatin by a one step process. ChIPmentation is, thus, an easier and shorter method comparing to the standard workflow.

Diagenode's automated ChIPmentation won the CiteAb Award in 2018 as the innovative product of the year.

ChIPmentation will be used on GENE-SWitCH project within the WP1 to generate ChIP-seq data for hundreds of pig and chicken different tissues from different developmental stages using several Epigenetic marks. The advantages of ChIPmentation will help to get a high-quality data with an easier and shorter workflow.

WP leaders: Dr. Wassim Lakhal (WP1)

Dr. Wassim Lakhal is a research Scientist at Diagenode in the R&D Epigenetics department. After Agronomical studies, he obtained a PhD in

Molecular and Cellular Biology at the University of Orléans (France) and he was a Post-Doc fellow in EpiTRAITS network. He has a strong background in Plant science, especially trees but also in Epigenetics. He is a recognized expert in Biotech instruments development related to NGS and Long Read Sequencing. He is also specialized in protocols engineering and automation.

Since 2020, he is the leader of WP1 of GENE-SWitCH and generated ATAC-seq libraries for WP1 and WP5. Now he is focusing at optimizing a quick high throughput automated ChIPmentation protocol to complete the task 1.3 in WP1.

Publications

Amid C., *et al* (2019), The European Nucleotide Archive in 2019, Nucleic Acids Research, https://doi.org/10.1093/nar/gkz1063

Harrison PW., *et al* (2020), The European Nucleotide Archive in 2020, Nucleic Acids Research, https://doi.org/10.1093/nar/gkaa1028

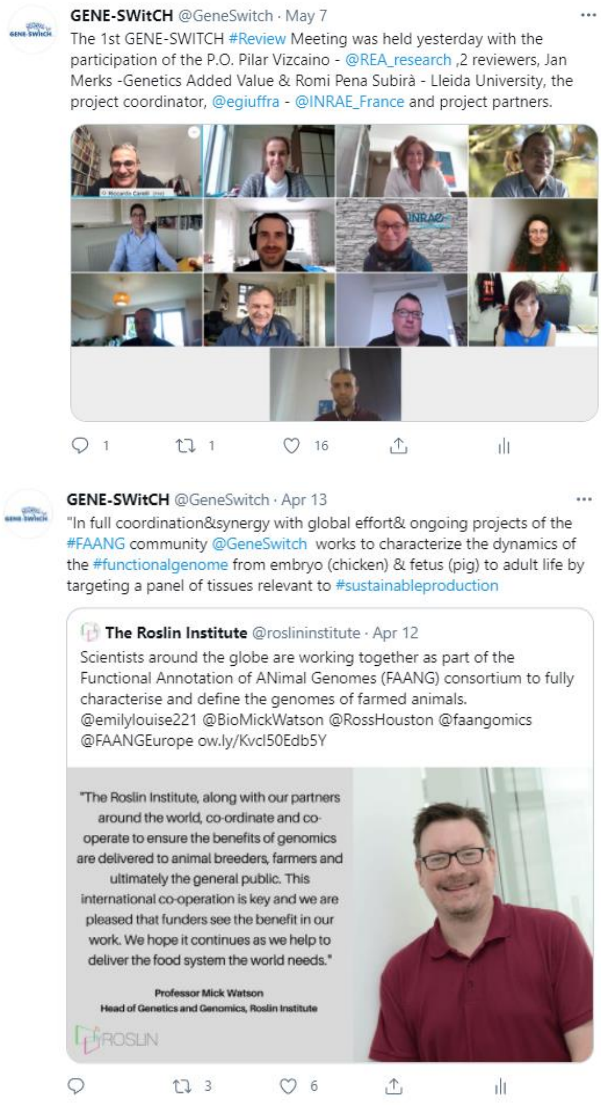
Clark E., et al (2020), From FAANG to fork: application of highly annotated genomes to improve farmed animal production, Genome Biology, https://doi.org/10.1186/s13059-020-02197-8

Howe K.L., et al (2020), Ensembl 2021, Nucleic Acids Research, https://doi.org/10.1093/nar/gkaa942

Wells J.M., van der Hee B., Microbial Regulation of Host Physiology by Short-chain Fatty Acids, Trends in Microbiology, https://doi.org/10.1016/j.tim.2021.02.001

Mollandin, F., Rau, A., and Croiseau, P. (2021) An evaluation of the predictive performance and mapping power of the BayesR model for genomic prediction. G3: Genes, Genomes, Genetics, in press.

Tweets



Upcoming events

Due to the COVID-19 pandemic, events in 2021 have been either postponed to a later date or moved online. Physical conferences are continuously following the evolution of the COVID-19 situation and updating information regularly on the relevant websites of the events. Below is the list of the upcoming international conferences and events in 2021/2022. Please check the websites of the events for the most updated information.

Event	Date	Location
38th International Society for Animal Genetics Conference (2021)	26-30 July 2021	Virtual
72nd EAAP Annual Meeting 2021	30 August – 3 September 2021	Davos, Switzerland
International Plant & Animal Genome XXIX	8-12 January 2022	San Diego, CA, USA
World Congress on Genetics Applied to Livestock Production	3-8 July 2022	Rotterdam, the Netherlands
World's Poultry Congress	08-12 August 2022	Paris, France

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

www.GENE-SWitCH.eu

@GeneSwitch

@GENESWitCHEU

GENE-SWitCH