GENE-SWitCH

BOOK OF ABSTRACTS

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

Table of Contents

Editorial by Elisabetta Giuffra, GENE-SWitCH project coordinator 5

Scientific Papers 7

1. The European Nucleotide Archive in 2019 7

2. From FAANG to fork: application of highly annotated genomes to improve farmed animal production 7

3. The European Nucleotide Archive in 2020 8

4. Ensembl 2021 8

5. The FAANG Data Portal: Global, Open-Access, "FAIR", and Richly Validated Genotype to Phenotype Data for High-Quality Functional Annotation of Animal Genomes 9

6. Microbial Regulation of Host Physiology by Short-chain Fatty Acids 10

7. An evaluation of the predictive performance and mapping power of the BayesR model for genomic prediction 10

8. Ensembl 2022 11

9. Prediction performance of linear models and gradient boosting machine on complex phenotypes in outbred mice 11

10. Accounting for overlapping annotations in genomic prediction models of complex traits 12

11. Adding gene transcripts into genomic prediction improves accuracy and reveals sampling time dependence 13

12. Ensembl 2023 13

14. Detailed molecular and epigenetic characterization of the pig IPEC-J2 and chicken SL-29 cell lines 14

15. nf-core/isoseq: simple gene and isoform annotation with PacBio Iso-Seq long-read sequencing ... 15

16. Copy number variation on ABCC2-DNMBP loci impacts the diversity and composition of the gut microbiota in pigs 15

17. Identification of transcriptional regulatory variants in pig duodenum, liver, and muscle tissues 16

18. Mutations on a conserved distal enhancer in the porcine C-reactive protein gene impair its expression in liver 16

19. TAGADA: a scalable pipeline to improve genome annotations with RNA-seq data 17

20. Genomic architecture of carcass and pork traits and their association with immune capacity 18

Oral presentations *19*

1. The regulatory GENomE of Swine and CHicken: functional annotation during development 19

2. The GENE-SWitCH H2020 project: follows up and perspectives of functional genome annotations .. 20

3. Progress of H2020 GENE-SWitCH: Functional Annotation of Pig and Chicken Genomes during Development 21

4. GENE-SWitCH: improving the functional annotation of pig and chicken genomes for precision breeding 22

5. Aims and Outcomes of H2020 GENE-Switch (The regulatory GENomE of SWineand CHicken: Functional Annotation during development) 22

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

6. Differences in Maternal Diet Fiber Content Influence Patterns of Gene Expression and Chromatin Accessibility in Fetuses and Piglets 23
7. Evaluating the interpretability of SNP effect size classes in Bayesian genomic prediction models 24
8. Intégration d'annotations biologiques complexes dans les modèles bayésiens de prédiction génomique: Evaluation et extension du modèle BayesRC 25
9. Exploiting prior knowledge in genomic prediction models with BayesRCO 25
10. Capitalizing on complex annotations in Bayesian genomic prediction for a backcross population of growing pigs 26
11. The FAANG Data Coordination Centre 27
12. A vision for Bioinformatics within the Global FAANG project 27
13. The FAANG Data Coordination Centre: New European Perspectives for Our Continued Global Effort 28
14. Importance of rich 'FAIR' validated metadata, standardized protocols and analyses, and open-access datasets 28
15. The Functional Annotation of Animal Genomes Project: Progress and challenges for our continued global effort 29
16. EuroFAANG: empowering the future of FAANG to Fork research and infrastructure in Europe and beyond 30
17. Harnessing the Ag Genomics Data torrent: A community-driven discussion on best practices for using and re-using genomics data 30
18. The Global FAANG Data Portal and its role in the future EuroFAANG infrastructure 31
19. The EU H2020 project GENE SWitCH The regulatory GENomE of SWine and CHicken: functional annotation during development 32
20. Midterm progress of H2020 GENE-SWitCH: samples collection and assays-by-sequence 33
21. Extensive functional genomics information from early developmental time points for pig and chicken 34
22. nf-isoseq pipeline provides a first insight onto chicken transcripts landscape 35
23. Automatic and simple Iso-Seq genome annotation with nf-core/isoseq 36
24. Exploring early transcripts landscape with Isoseq 36
25. Predictive ability of genomic prediction in layers when including CADD scores as genome function information 37
26. TAGADA: Transcripts and Genes Assembly, Deconvolution, Analysis; a Nextflow pipeline to improve genome annotations with RNA-seq data 39
27. Improving genome annotations with RNA-seq data: a scalable and reproducible workflow for Transcripts and Genes Assembly, Deconvolution, Analysis (TAGADA). 40
28. DNA methylation dynamics regulating early embryonic development in pig and chicken 41
29. Multi-breed, multi-tissue, and multi-omics aiding the quest for key porcine regulators 42
30. Extending Ensembl Regulatory Annotation to Farmed Animals 43
31. Accounting for overlapping annotations as biological priors in genomic prediction models of complex traits 44

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

32. Using functional annotation and individual omics in genomic prediction 45
33. Using methylation to improve genomic prediction of gene expression in pigs 46
34. Using convolutional neural networks for image-based genomic prediction in mice 47
35. Genome-wide association studies for body weight in broilers using sequencing and SNP chip data 47
36. Identifying muscle transcriptional regulatory elements in the pig genome 48
37. An in-depth analysis of regulatory polymorphisms in three pig transcriptomes 49
38. Deciphering genetic variants from whole genome affecting duodenum, liver and muscle transcriptomes in pigs 50
39. Epigenetic effects of maternal rich-fiber diets in pigs 50
40. Influence of maternal diet on the liver and skeletal muscle epigenomes of fetal and young pigs ... 51
41. Dynamic inclusion of functional genome annotations to improve accuracy of genomic 52
prediction in pigs 52
GENE-SWitCH Final conference 54
Day 1 - Session 1: Identification and characterization of functional genomic elements 54
Day 2 -Session 2: Implementing FAANG innovation for innovation breeding 54
Day 2 - Session 3 - Celebrating 10 years of FAANG 54
Summary talks: 54
Day 3 - Policy & Ethics Workshop - jointly organized with the GEroNIMO project 55

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

Editorial by Elisabetta Giuffra, GENE-SWitCH project coordinator

Back in 2017, the EC launched the Thematic Call "SFS-30-2018-2019-2020: Agri-Aqua Labs" within the Horizon 2020 Work Program 2018-2020, and our consortium came together for the first time to start preparing a research pre-proposal for this call. We were eleven partners representing pan-European excellence, including the world's leading animal breeding and biotechnology industries, some of us early co-founders of the global FAANG (Functional Annotation of the Animal Genomes) initiative. We set an ambitious goal: to provide new fundamental knowledge on the functional genomes of two major meat sources (pig and chicken), and to enable its immediate translation to the pig and poultry industries. In 2018, we jointly developed and submitted the full proposal, which was successfully selected later that year. Finally, GENE-SWitCH started in July 2019 as a four-year project. Due to the challenges we faced during the COVID pandemic, the project was extended by another six months and ended on 31 December 2023.

As little knowledge was available for annotating the chicken and pig genomes across development, our first goal was to identify and characterize the functional elements of the pig and chicken genomes that are active, poised or repressed ('switches') in utero and in ovo, as well as in the neonate. All datasets would be produced in full compliance with FAANG standards, under the principles of international collaboration and open science, and in coordination and synergy with the other FAANG projects in the EU and worldwide. This would contribute to the achievement of highly informative reference annotation maps, i.e. new Ensembl regulatory builds for these species, available to the whole community, as well as providing a solid basis for evolutionary and comparative analyses of these genomes across development.

To assess the interest in using this new knowledge for precision animal breeding, two approaches were taken, coupled with several dissemination and outreach activities aimed at transferring the results of the project to stakeholder communities. One approach focused on improving genomic tools to enable sustainable breeding and production. Genomic prediction models were extended by incorporating different functional and overlapping annotations, tested in targeted pig and chicken phenotype cohorts and finally validated in commercial populations of both species. In a second 'pilot' approach, a large-scale animal experiment was conducted to assess the epigenetic effects on the functional genome induced by diet, a major environmental factor that can be easily manipulated for farm animals. Specifically, we wanted to assess whether and to what extent the addition of non-digestible fibers (key components of more sustainable diets) to the maternal diet could affect the epigenetic profiles of the offspring at the fetal and post-weaning stages.

At the time of writing this book, work is still in progress on several forthcoming publications, but we can already be very pleased with the overall impact of GENE-SWitCH. High-quality reference annotation maps for these species have become available to the entire research community (with Ensembl release 111 on the way), cutting-edge research results have been achieved, and our results have been effectively communicated through several training, dissemination and outreach activities to European and worldwide stakeholders. Importantly, thanks to the extension of clustering activities with the parallel H2020 projects in other species, GENE-SWitCH has been a key co-founder of the ongoing EuroFAANG infrastructure project (eurofaang.eu). This infrastructure provides a great opportunity for future large-scale genotype-to-phenotype research in Europe and the perfect framework for several follow-up studies of our project.

Elisabetta Giuffra

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

GENE-SWitCH Coordinator

Elisabetta Giuffra

Elisabetta Giuffra (female), biologist, coordinates GENE-SWitCH. After her PhD (1993) she worked on different animal genetics and genomics topics across Italy, France, Sweden and U.K. She is a research director at Paris-Saclay University, INRAE, AgroParisTech, GABI (Jouy-en-Josas, France) since 2011. She works in the field of host (pig)-virus interactions by transcriptional genomics approaches, and in local (e.g. FR-AgENCODE French pilot project) and international efforts to improve the functional annotation of farmed animal genomes for genotype-to-phenotype research (FAANG initiative). She has been co-Chair of the FAANG-Europe COST Action CA15112 (2016-2020) and co-leads the FAANG Steering Committee. More recently she works with organoids to elucidate genotype-to-phenotype relationships for immune response related traits in pigs and co-leads a work package (Development of a framework for biobanking and use of *in vitro* cellular models) of the EuroFAANG research infrastructure EU project since 2023.

INRAE, France

E-mail: elisabetta.giuffra@inrae.fr

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

Scientific Papers

1. The European Nucleotide Archive in 2019

C. Amid[1], B.T.F. Alako[1], V. Balavenkataraman Kadhirvelu[1], T. Burdett[1], J. Burgin[1], J. Fan[1], P. Harrison[1], S. Holt[1], A. Hussein[1], E. Ivanov[1], S. Jayathilaka[1], S. Kay[1], T. Keane[1], R. Leinonen[1], X. Liu[1], J. Martinez-Villacorta[1], A. Milano[1], A. Pakseresht[1], N. Rahman[1], J. Rajan[1], K. Reddy[1], E. Richards[1], D. Smirnov[1], A. Sokolov[1], S. Vijayaraja[1], G. Cochrane[1]

1. European Molecular Biology Laboratory, European Bioinformatics Institute

Nucleic Acids Research, Volume 48, Issue D1, 8 January 2020
https://doi.org/10.1093/nar/gkz1063

The European Nucleotide Archive (ENA - https://www.ebi.ac.uk/ena) at the European Molecular Biology Laboratory's European Bioinformatics Institute provides open and freely available data deposition and access services across the spectrum of nucleotide sequence data types. Making the world's public sequencing datasets available to the scientific community, the ENA represents a globally comprehensive nucleotide sequence resource. Here, we outline ENA services and content in 2019 and provide an insight into selected key areas of development in this period.

2. From FAANG to fork: application of highly annotated genomes to improve farmed animal production

E. Clark[1], A. Archibald[1], H.D. Daetwyler[2], M.A.M. Groenen[3], P. Harrison[4], R.D. Houston[1], C. Kühn[5], S. Lien[6], D.J. Macqueen[1], J.M. Reecy[7], D. Robledo[1], M. Watson[1], C.K. Tuggle[7], E. Giuffra[8]

1. The Roslin Institute and R(D)SVS, University of Edinburgh
2. Agriculture Victoria, AgriBio Centre for AgriBioscience
3. Animal Breeding and Genomics Centre, Wageningen University and Research
4. European Molecular Biology Laboratory, European Bioinformatics Institute
5. Leibniz Institute for Farm Animal Biology (FBN), Institute of Genome Biology, Genome Physiology Unit
6. Centre for Integrative Genetics (CIGENE), Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences
7. Department of Animal Science, Iowa State University
8. Paris-Saclay University, INRAE, AgroParisTech, GABI

Genome Biology, 21, Article number: 285 24 November 2020
https://doi.org/10.1186/s13059-020-02197-8

The Food and Agriculture Organization of the United Nations (FAO) reports that by the year 2050 the global human population is likely to reach 9.7 billion, rising to 11.2 billion by 2100. This population growth poses several challenges to the global food system, which will need to produce more healthy food using fewer natural resources, reducing the environmental impact, conserving biodiversity and flexibly adjusting to changing societal expectations. Meeting this demand requires environmentally sustainable improvements to farmed animal health and welfare, and of efficiency and diversification (e.g. to include a broader range of locally adapted species) [1]. The changes in breeding strategies and management practices required to meet these goals will need to build on an improved ability to accurately use genotype to predict phenotype in the world's farmed animal species, both terrestrial and aquatic. Here we describe a set of research priorities to meet such present and future challenges that build on progress, successes and resources from the Functional

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

Annotation of ANimal Genomes (FAANG) project. The first stages of FAANG focused on foundational data generation to characterize expressed and regulatory genomic regions, curation and provision of annotated farmed animal genomes. These were largely based on individual level, high depth approaches. The primary challenge facing this community now is harnessing these resources to link genotype, phenotype and genetic merit in order to translate this research out of the laboratory and into industry application in the field. To achieve this effectively, we will need to generate functional genomic information for large populations of animals, rather than relying on a small number of deeply annotated individuals. Furthermore, to date, most of the datasets are from tissues consisting of heterogeneous cell populations, hindering the resolution of functional information and limiting our ability to understand the fundamental cellular and subcellular processes underlying phenotypes. Since the original FAANG white paper was published in 2015, exciting new opportunities have arisen to tackle these challenges. In the paper, we describe a set of research action priorities for FAANG for the next decade.

3. The European Nucleotide Archive in 2020

P. Harrison[1], A. Ahamed[1], R. Aslam[1], B.T.F. Alako[1], J. Burgin[1], N. Buso[1], M. Courtot[1], J. Fan[1], D. Gupta[1], M. Haseeb[1], S. Holt[1], T. Ibrahim[1], E. Ivanov[1], S. Jayathilaka[1], V. Balavenkataraman Kadhirvelu[1], M. Kumar[1], R. Lopez[1], S. Kay[1], R. Leinonen[1], X. Liu[1], C. O'Cathail[1], A. Pakseresh[1], Y. Park[1], S. Pesant[1], N. Rahman[1], J. Rajan[1], A. Sokolov[1], S. Vijayaraja[1], Z. Waheed[1], A. Zyoud[1], T. Burdett[1], G. Cochrane G[1]

1. European Molecular Biology Laboratory, European Bioinformatics Institute

Nucleic Acids Research, Volume 49, Issue D1, 8 January 2021
https://doi.org/10.1093/nar/gkaa1028

The European Nucleotide Archive (ENA - https://www.ebi.ac.uk/ena), provided by the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI), has for almost forty years continued in its mission to freely archive and present the world's public sequencing data for the benefit of the entire scientific community and for the acceleration of the global research effort. Here we highlight the major developments to ENA services and content in 2020, focusing in particular on the recently released updated ENA browser, modernization of our release process and our data coordination collaborations with specific research communities.

4. Ensembl 2021

K.L. Howe[1], P. Achuthan[1], J. Allen[1], J. Allen[1], J. Alvarez-Jarreta[1], M.R. Amode[1], I.M. Armean[1], A.G. Azov[1], R. Bennett[1], J. Bhai[1], K. Billis[1], S. Boddu[1], M. Charkhchi[1], C. Cummins[1], L. Da Rin Fioretto[1], C. Davidson[1], K. Dodiya[1], B. El Houdaigui[1], R. Fatima[1], A. Gall[1], C. Garcia Giron[1], T. Grego[1], C. Guijarro-Clarke[1], L. Haggerty[1], A. Hemrom[1], T. Hourlier[1], O.G. Izuogu[1], T. Juettemann[1], V. Kaikala[1], M. Kay[1], I. Lavidas[1], T. Le[1], D. Lemos[1], J. Gonzalez Martinez[1], J.C. Marugán[1], T. Maurel[1], A.C. McMahon[1], S. Mohanan[1], B. Moore[1], M. Muffato[1], D.N. Oheh[1], D. Paraschas[1], A. Parker[1], A. Parton[1], I. Prosovetskaia[1], M.P. Sakthivel[1], A.I.A. Salam[1], B.M. Schmitt[1], H. Schuilenburg[1], D. Sheppard[1], E. Steed[1], M. Szpak[1], M. Szuba[1], K. Taylor[1], A. Thormann[1], G. Threadgold[1], B. Walts[1], A. Winterbottom[1], M. Chakiachvili[1] A. Chaubal[1], N. De Silva[1], B. Flint[1], A. Frankish[1], S.E. Hunt[1], G.R. IIsley[1], N. Langridge[1], J.E. Loveland[1], F.J. Martin[1], J.M. Mudge[1], J. Morales[1], E. Perry[1], M. Ruffier[1], J. Tate[1], D. Thybert[1], S.J. Trevanion[1], F. Cunningham[1], A.D. Yates[1], D. Zerbino[1], P. Flicek[1]

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

1. European Molecular Biology Laboratory, European Bioinformatics Institute

Nucleic Acids Research, Volume 49, Issue D1, 8 January 2021
https://doi.org/10.1093/nar/gkaa942

The Ensembl project (https://www.ensembl.org) annotates genomes and disseminates genomic data for vertebrate species. We create detailed and comprehensive annotation of gene structures, regulatory elements and variants, and enable comparative genomics by inferring the evolutionary history of genes and genomes. Our integrated genomic data are made available in a variety of ways, including genome browsers, search interfaces, specialist tools such as the Ensembl Variant Effect Predictor, download files and programmatic interfaces. Here, we present recent Ensembl developments including two new website portals. Ensembl Rapid Release (http://rapid.ensembl.org) is designed to provide core tools and services for genomes as soon as possible and has been deployed to support large biodiversity sequencing projects. Our SARS-CoV-2 genome browser (https://covid-19.ensembl.org) integrates our own annotation with publicly available genomic data from numerous sources to facilitate the use of genomics in the international scientific response to the COVID-19 pandemic. We also report on other updates to our annotation resources, tools and services. All Ensembl data and software are freely available without restriction.

5. The FAANG Data Portal: Global, Open-Access, "FAIR", and Richly Validated Genotype to Phenotype Data for High-Quality Functional Annotation of Animal Genomes

P. Harrison[1], A. Sokolov[1], A. Nayak[1], J. Fan[1], D. Zerbino[1], G. Cochran[1], P. Flicek[1]

1. European Molecular Biology Laboratory, European Bioinformatics Institute

Frontiers in Genetics, Volume 12, 17 June 2021
https://doi.org/10.3389/fgene.2021.639238

The Functional Annotation of ANimal Genomes (FAANG) project is a worldwide coordinated action creating high-quality functional annotation of farmed and companion animal genomes. The generation of a rich genome-to-phenome resource and supporting informatic infrastructure advances the scope of comparative genomics and furthers the understanding of functional elements. The project also provides terrestrial and aquatic animal agriculture community powerful resources for supporting improvements to farmed animal production, disease resistance, and genetic diversity. The FAANG Data Portal (https://data.faang.org) ensures Findable, Accessible, Interoperable and Reusable (FAIR) open access to the wealth of sample, sequencing, and analysis data produced by an ever-growing number of FAANG consortia. It is developed and maintained by the FAANG Data Coordination Centre (DCC) at the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI). FAANG projects produce a standardized set of multi-omic assays with resulting data placed into a range of specialized open data archives. To ensure this data is easily findable and accessible by the community, the portal automatically identifies and collates all submitted FAANG data into a single easily searchable resource. The Data Portal supports direct download from the multiple underlying archives to enable seamless access to all FAANG data from within the portal itself. The portal provides a range of predefined filters, powerful predictive search, and a catalogue of sampling and analysis protocols and automatically identifies publications associated with any dataset. To ensure all FAANG data submissions are high-quality, the portal includes powerful contextual metadata validation and data submissions brokering to the underlying EMBL-EBI archives. The portal will incorporate extensive new technical infrastructure to effectively deliver and standardize FAANG's shift to single-cell omics, cell atlases, pangenomes, and novel

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

phenotypic prediction models. The Data Portal plays a key role for FAANG by supporting high-quality functional annotation of animal genomes, through open FAIR sharing of data, complete with standardized rich metadata. Future Data Portal features developed by the DCC will support new technological developments for continued improvement for FAANG projects.

6. Microbial Regulation of Host Physiology by Short-chain Fatty Acids

B. van der Hee[1], J. Wells[1]

1. Wageningen University & Research, Animal Breeding and Genomics

Trends in Microbiology, Volume 12, Issue 8, August 2021
https://doi.org/10.1016/j.tim.2021.02.001

Our ancestral diet consisted of much more non-digestible fiber than that of many societies today. Thus, from an evolutionary perspective the human genome and its physiological and nutritional requirements are not well aligned to modern dietary habits. Fiber reaching the colon is anaerobically fermented by the gut bacteria, which produce short-chain fatty acids (SCFAs) as metabolic by-products. SCFAs play a role in intestinal homeostasis, helping to explain why changes in the microbiota can contribute to the pathophysiology of human diseases. Recent research has shown that SCFAs can also have effects on tissues and organs beyond the gut, through their circulation in the blood. SCFAs not only signal through binding to cognate G-protein-coupled receptors on endocrine and immune cells in the body but also induce epigenetic changes in the genome through effects on the activity of histone acetylase and histone deacetylase enzymes. Furthermore, epigenetic imprinting likely occurs in utero, highlighting the importance of the maternal diet in early life. Here we review current understanding of how SCFAs impact on human and animal physiology and discuss the potential applications of SCFAs in the prevention and treatment of human diseases.

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

F. Mollandin[1], A. Rau[1], P. Croiseau[1]

1. INRAE, AgroParisTech, GABI, Université Paris-Saclay

G3, Volume 11, Issue 11, November 2021
https://doi.org/10.1093/g3journal/jkab225

Technological advances and decreasing costs have led to the rise of increasingly dense genotyping data, making feasible the identification of potential causal markers. Custom genotyping chips, which combine medium-density genotypes with a custom genotype panel, can capitalize on these candidates to potentially yield improved accuracy and interpretability in genomic prediction. A particularly promising model to this end is BayesR, which divides markers into four effect size classes. BayesR has been shown to yield accurate predictions and promise for quantitative trait loci (QTL) mapping in real data applications, but an extensive benchmarking in simulated data is currently lacking. Based on a set of real genotypes, we generated simulated data under a variety of genetic architectures and phenotype heritabilities, and we evaluated the impact of excluding or including causal markers among the genotypes. We define several statistical criteria for QTL mapping, including several based on sliding windows to account for linkage disequilibrium (LD). We

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

compare and contrast these statistics and their ability to accurately prioritize known causal markers. Overall, we confirm the strong predictive performance for BayesR in moderately to highly heritable traits, particularly for 50k custom data. In cases of low heritability or weak LD with the causal marker in 50k genotypes, QTL mapping is a challenge, regardless of the criterion used. BayesR is a promising approach to simultaneously obtain accurate predictions and interpretable classifications of SNPs into effect size classes. We illustrated the performance of BayesR in a variety of simulation scenarios, and compared the advantages and limitations of each.

8. Ensembl 2022

F. Cunningham[1], J.E. Allen[1], J. Allen[1], J. Alvarez-Jarreta[1], M.R. Amode[1], I.M. Armean[1], O. Austine Orimoloye[1], A.G. Azov[1], I. Barnes[1], R. Bennett[1], A. Berry[1], J. Bhai[1], A. Bignell[1], K. Billis[1], S. Boddu[1], L. Brooks[1], M. Charkhchi[1], C. Cummins[1], L. Da Rin Fioretto[1], C. Davidson[1], K. Dodiya[1], S. Donaldson[1], B. El Houdaigui[1], T. El Naboulsi[1], R. Fatima[1], C.G. Giron[1], T. Genez[1], J.G. Martinez[1], C. Guijarro-Clarke[1], A. Gymer[1], M. Hardy[1], Z. Hollis[1], T. Hourlier[1], Hunt T [1], T. Juettemann[1], V. Kaikala[1], M. Kay[1], I. Lavidas[1], T. Le[1], D. Lemos[1], J.C. Marugán[1], S. Mohanan[1], A. Mushtaq[1], M. Naven[1], D.N. Ogeh[1], A. Parker[1], A. Parton[1], M. Perry[1], I. Piližota[1], I. Prosovetskaia[1], M.P. Sakthivel[1], A.I.A. Salam[1], B.M. Schmitt[1], H. Schuilenburg[1], D. Sheppard[1], J.G. Pérez-Silva[1], W. Stark[1], E. Steed[1], K. Sutinen[1], R. Sukumaran[1], D. Sumathipala[1], M.M. Suner[1], M. Szpak[1], A. Thormann[1], F.F. Tricomi[1], D. Urbina-Gómez[1], A. Veidenberg[1], T.A. Walsh[1], B. Walts[1], N. Willhoft[1], A. Winterbottom[1], E. Wass[1], M. Chakiachvili[1], B. Flint[1], A. Frankish[1], S. Giorgetti[1], L. Haggerty[1], S.E. Hunt[1], G.R. IIsley[1] , J.E. Loveland[1], F.J. Martin[1], B. Moore[1], J.M. Mudge[1], M. Muffato[1], E. Perry[1], M. Ruffier[1], J. Tate[1], D. Thybert[1], S.J. Trevanion[1], S. Dyer[1], P. Harriso[1], K.L. Howe[1], A.D. Yates[1], D. Zerbino[1], P. Flicek[1]

1. European Molecular Biology Laboratory, European Bioinformatics Institute

Nucleic Acids Research, Volume 50, Issue D1, 7 January 2022
https://doi.org/10.1093/nar/gkab1049

Ensembl (https://www.ensembl.org) is unique in its flexible infrastructure for access to genomic data and annotation. It has been designed to efficiently deliver annotation at scale for all eukaryotic life, and it also provides deep comprehensive annotation for key species. Genomes representing a greater diversity of species are increasingly being sequenced. In response, we have focused our recent efforts on expediting the annotation of new assemblies. Here, we report the release of the greatest annual number of newly annotated genomes in the history of Ensembl via our dedicated Ensembl Rapid Release platform (http://rapid.ensembl.org). We have also developed a new method to generate comparative analyses at scale for these assemblies and, for the first time, we have annotated non-vertebrate eukaryotes. Meanwhile, we continually improve, extend and update the annotation for our high-value reference vertebrate genomes and report the details here. We have a range of specific software tools for specific tasks, such as the Ensembl Variant Effect Predictor (VEP) and the newly developed interface for the Variant Recoder. All Ensembl data, software and tools are freely available for download and are accessible programmatically.

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

B. Perez[1], M. Bink[1], K.L. Svenson[2], G.A. Churchill[2], M. Calus[3]

1. Hendrix Genetics B.V., Research and Technology Center (RTC)

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

2. The Jackson Laboratory
3. Wageningen University & Research, Animal Breeding and Genomics

G3, Volume 12, Issue 4, April 2022
https://doi.org/10.1093/g3journal/jkac039

We compared the performance of linear (GBLUP, BayesB, and elastic net) methods to a nonparametric tree-based ensemble (gradient boosting machine) method for genomic prediction of complex traits in mice. The dataset used contained genotypes for 50,112 SNP markers and phenotypes for 835 animals from 6 generations. Traits analyzed were bone mineral density, body weight at 10, 15, and 20 weeks, fat percentage, circulating cholesterol, glucose, insulin, triglycerides, and urine creatinine. The youngest generation was used as a validation subset, and predictions were based on all older generations. Model performance was evaluated by comparing predictions for animals in the validation subset against their adjusted phenotypes. Linear models outperformed gradient boosting machine for 7 out of 10 traits. For bone mineral density, cholesterol, and glucose, the gradient boosting machine model showed better prediction accuracy and lower relative root mean squared error than the linear models. Interestingly, for these 3 traits, there is evidence of a relevant portion of phenotypic variance being explained by epistatic effects. Using a subset of top markers selected from a gradient boosting machine model helped for some of the traits to improve the accuracy of prediction when these were fitted into linear and gradient boosting machine models. Our results indicate that gradient boosting machine is more strongly affected by data size and decreased connectedness between reference and validation sets than the linear models. Although the linear models outperformed gradient boosting machine for the polygenic traits, our results suggest that gradient boosting machine is a competitive method to predict complex traits with assumed epistatic effects.

10. Accounting for overlapping annotations in genomic prediction models of complex traits

F. Mollandin[1], H. Gilbert[2], P. Croiseau[1], A. Rau[1]

1. INRAE, AgroParisTech, GABI, Université Paris-Saclay
2. GenPhySE, INRAE, ENVT, Université de Toulouse

BMC Bioinformatics, 23, Article 365 - 6 September 2022
https://doi.org/10.1186/s12859-022-04914-5

It is now widespread in livestock and plant breeding to use genotyping data to predict phenotypes with genomic prediction models. In parallel, genomic annotations related to a variety of traits are increasing in number and granularity, providing valuable insight into potentially important positions in the genome. The BayesRC model integrates this prior biological information by factorizing the genome according to disjoint annotation categories, in some cases enabling improved prediction of heritable traits. However, BayesRC is not adapted to cases where markers may have multiple annotations. We propose two novel Bayesian approaches to account for multi-annotated markers through a cumulative (BayesRC+) or preferential (BayesRCπ) model of the contribution of multiple annotation categories. We illustrate their performance on simulated data with various genetic architectures and types of annotations. We also explore their use on data from a backcross population of growing pigs in conjunction with annotations constructed using the PigQTLdb. In both simulated and real data, we observed a modest improvement in prediction quality with our models when used with informative annotations. In addition, our results show that BayesRC+ successfully prioritizes multi-annotated markers according to their posterior variance, while BayesRCπ provides

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

a useful interpretation of informative annotations for multi-annotated markers. Finally, we explore several strategies for constructing annotations from a public database, highlighting the importance of careful consideration of this step. When used with annotations that are relevant to the trait under study, BayesRCπ and BayesRC+ allow for improved prediction and prioritization of multi-annotated markers, and can provide useful biological insight into the genetic architecture of traits.

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

B.C. Perez[1], M. Bink[1], K.L. Svenson[2], G.A. Churchill[2], M. Calus[3]

1. Hendrix Genetics B.V., Research and Technology Center (RTC)
2. The Jackson Laboratory
3. Wageningen University & Research, Animal Breeding and Genomics

G3, Volume 12, Issue 11, November 2022
https://doi.org/10.1093/g3journal/jkac258

Recent developments allowed generating multiple high-quality 'omics' data that could increase the predictive performance of genomic prediction for phenotypes and genetic merit in animals and plants. Here, we have assessed the performance of parametric and nonparametric models that leverage transcriptomics in genomic prediction for 13 complex traits recorded in 478 animals from an outbred mouse population. Parametric models were implemented using the best linear unbiased prediction, while nonparametric models were implemented using the gradient boosting machine algorithm. We also propose a new model named GTCBLUP that aims to remove between-omics-layer covariance from predictors, whereas its counterpart GTBLUP does not do that. While gradient boosting machine models captured more phenotypic variation, their predictive performance did not exceed the best linear unbiased prediction models for most traits. Models leveraging gene transcripts captured higher proportions of the phenotypic variance for almost all traits when these were measured closer to the moment of measuring gene transcripts in the liver. In most cases, the combination of layers was not able to outperform the best single-omics models to predict phenotypes. Using only gene transcripts, the gradient boosting machine model was able to outperform best linear unbiased prediction for most traits except body weight, but the same pattern was not observed when using both single nucleotide polymorphism genotypes and gene transcripts. Although the GTCBLUP model was not able to produce the most accurate phenotypic predictions, it showed the highest accuracies for breeding values for 9 out of 13 traits. We recommend using the GTBLUP model for prediction of phenotypes and using the GTCBLUP for prediction of breeding values.

12. Ensembl 2023

F.J. Martin[1], M.R. Amode[1], A. Aneja[1], O. Austine-Orimoloye[1], A.G. Azov[1], I. Barnes[1], A. Becker[1], R. Bennett[1], A. Berry[1], J. Bhai[1], S.K. Bhurji[1], A. Bignell[1], S. Boddu[1], P.R. Branco Lins[1], L. Brooks[1], S.B. Ramaraju[1], M. Charkhchi[1], A. Cockburn[1], L. Da Rin Fiorretto[1], C. Davidson[1], K. Dodiya[1], S. Donaldson[1], B. El Houdaigui[1], T. El Naboulsi[1], R. Fatima[1], C.G. Giron[1], T. Genez[1], G.S. Ghattaoraya[1], J.G. Martinez[1], C. Guijarro[1], M. Hardy[1], Z. Hollis[1], T. Hourlier[1], T. Hunt[1], M. Kay[1], V. Kaykala[1], T. Le[1], D. Lemos[1], D. Marques-Coelho[1], J.C. Marugán[1], G.A. Merino[1], L.P. Mirabueno[1], A. Mushtaq[1], S.N. Hossain[1], D.N. Ogeh[1], M.P. Sakthivel[1], A. Parker[1], M. Perry[1], I. Piližota[1], I. Prosovetskaia[1], J.G. Pérez-Silva[1], A.I.A.

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

Salam[1], N. Saraiva-Agostinho[1], H. Schuilenburg[1], D. Sheppard[1], S. Sinha[1], B. Sipos[1], W. Stark[1], E. Steed[1], R. Sukumaran[1], D. Sumathipala[1], M.M. Suner[1], L. Surapaneni[1], K. Sutinen[1], M. Szpak[1], F.F. Tricomi[1], D. Urbina-Gómez[1], A. Veidenberg[1], T.A. Walsh[1], B. Walts[1], E. Wass[1], N. Willhoft[1], J. Allen[1], J. Alvarez-Jarreta[1], M. Chakiachvili[1], B. Flint[1], S. Giorgetti[1], L. Haggerty[1], G.R. Ilsley[1], J.E. Loveland[1], B. Moore[1], J.M. Mudge[1], J. Tate[1], D. Thybert[1], S.J. Trevanion[1], A. Winterbottom[1], A. Frankish[1], S.E. Hunt[1], M. Ruffier[1], F. Cunningham[1], S. Dyer[1], R.D. Finn[1], K.L. Howe[1], P. Harrison [1], A.D. Yates[1], P. Flicek[1]

1. European Molecular Biology Laboratory, European Bioinformatics Institute

Nucleic Acids Research, Volume 51, Issue D1, 6 January 2023
https://doi.org/10.1093/nar/gkac958

Ensembl (https://www.ensembl.org) has produced high-quality genomic resources for vertebrates and model organisms for more than twenty years. During that time, our resources, services and tools have continually evolved in line with both the publicly available genome data and the downstream research and applications that utilize the Ensembl platform. In recent years we have witnessed a dramatic shift in the genomic landscape. There has been a large increase in the number of high-quality reference genomes through global biodiversity initiatives. In parallel, there have been major advances towards pangenome representations of higher species, where many alternative genome assemblies representing different breeds, cultivars, strains and haplotypes are now available. In order to support these efforts and accelerate downstream research, it is our goal at Ensembl to create high-quality annotations, tools and services for species across the tree of life. Here, we report our resources for popular reference genomes, the dramatic growth of our annotations (including haplotypes from the first human pangenome graphs), updates to the Ensembl Variant Effect Predictor (VEP), interactive protein structure predictions from AlphaFold DB, and the beta release of our new website.

14. Detailed molecular and epigenetic characterization of the pig IPEC-J2 and chicken SL-29 cell lines

J. de Vos[1], R.P.M.A. Crooijmans[1], M. Derks[1], S.L. Kloet[2], B. Dibbits[1], M.A.M. Groenen[1], O. Madsen[1]

1. Wageningen University & Research, Animal Breeding and Genomics
2. Leids Universitair Medisch Centrum

iScience, Volume 26, Issue 3, March 2023
https://doi.org/10.1016/j.isci.2023.106252

The pig IPEC-J2 and chicken SL-29 cell lines are of interest because of their untransformed nature and wide use in functional studies. Molecular characterization of these cell lines is important to gain insight into possible molecular aberrations. The aim of this paper is to provide a molecular and epigenetic characterization of the IPEC-J2 and SL-29 cell lines, a cell-line reference for the FAANG community, and future biomedical research. Whole genome sequencing, gene expression, DNA methylation, chromatin accessibility, and ChIP-seq of four histone marks (H3K4me1, H3K4me3, H3K27ac, H3K27me3) and an insulator (CTCF) are used to achieve these aims. Heteroploidy (aneuploidy) of various chromosomes was observed from whole genome sequencing analysis in both cell lines. Furthermore, higher gene expression for genes located on chromosomes with aneuploidy in comparison to diploid chromosomes was observed. Regulatory complexity of gene

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

expression, DNA methylation, and chromatin accessibility was investigated through an integrative approach.

15. nf-core/isoseq: simple gene and isoform annotation with PacBio Iso-Seq long-read sequencing

S. Guizard[1], K. Miedzinska[1], Ja. Smith[1], Jo. Smith[1], R.I. Kuo[2], M. Davey[1], A. Archibald[1], M. Watson[1]

1. The Roslin Institute and R(D)SVS, University of Edinburgh
2. Wobble Genomics, University of Edinburgh

Bioinformatics, Volume 39, Issue 5, May 2023
https://doi.org/10.1093/bioinformatics/btad150

Iso-Seq RNA long-read sequencing enables the identification of full-length transcripts and isoforms, removing the need for complex analysis such as transcriptome assembly. However, the raw sequencing data need to be processed in a series of steps before annotation is complete. Here, we present nf-core/isoseq, a pipeline for automatic read processing and genome annotation. Following nf-core guidelines, the pipeline has few dependencies and can be run on any of platforms. The pipeline is freely available online on the nf-core website (https://nf-co.re/iso seq) and on GitHub (https://github.com/nf-core/isoseq) under MIT License.

16. Copy number variation on ABCC2-DNMBP loci impacts the diversity and composition of the gut microbiota in pigs

Y. Ramayo-Caldas[1], D. Crespo-Piazuelo[1], J. Morata[2], O. González-Rodríguez[1], C. Sebastià[3], A. Castello[3], A. Dalmau[1], S. Ramos-Onsins[3], K.G. Alexiou[3], J.M. Folch[3], R. Quintanilla[1], M. Ballester[1]

1. Animal Breeding and Genetics Program, Institute of Agrifood Research and Technology
2. Barcelona Institute of Science and Technology
3. Consejo Superior de Investigaciones Científicas, Autonomous University of Barcelona

Microbiology Spectrum, Volume 11, No. 4, 31 May 2023
https://doi.org/10.1128/spectrum.05271-22

Genetic variation in the pig genome partially modulates the composition of porcine gut microbial communities. Previous studies have been focused on the association between single nucleotide polymorphisms (SNPs) and the gut microbiota, but little is known about the relationship between structural variants and fecal microbial traits. The main goal of this study was to explore the association between porcine genome copy number variants (CNVs) and the diversity and composition of pig fecal microbiota. For this purpose, we used whole-genome sequencing data to undertake a comprehensive identification of CNVs followed by a genome-wide association analysis between the estimated CNV status and the fecal bacterial diversity in a commercial Duroc pig population. A CNV predicted as gain (DUP) partially harboring ABCC2-DNMBP loci was associated with richness ($P = 5.41 \times 10–5$, false discovery rate [FDR] = 0.022) and Shannon α-diversity ($P = 1.42 \times 10–4$, FDR = 0.057). The in silico predicted gain of copies was validated by real-time quantitative PCR (qPCR), and its segregation, and positive association with the richness and Shannon α-diversity of the porcine fecal bacterial ecosystem was confirmed in an unrelated F1 (Duroc × Iberian) cross. Our results advise the relevance of considering the role of host-genome structural variants as potential modulators of microbial ecosystems and suggest the ABCC2-DNMBP CNV as a

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

host-genetic factor for the modulation of the diversity and composition of the fecal microbiota in pigs.

17. Identification of transcriptional regulatory variants in pig duodenum, liver, and muscle tissues

D. Crespo-Piazuelo[1], H. Acloque[2], O. González-Rodríguez[1], M. Mongellaz[2], M.J. Mercat[3], M. Bink[4], A.E. Huisman[5], Y. Ramayo-Caldas[1], J.P. Sánchez[1], M. Ballester[1]

1. Animal Breeding and Genetics Program, Institute of Agrifood Research and Technology
2. Paris-Saclay University, INRAE, AgroParisTech, GABI
3. IFIP-Institut du porc and Alliance R&D
4. Hendrix Genetics Research Technology & Services B.V.
5. Hypor B.V.

GigaScience, Volume 12 2023 - 24 June 2023
https://doi.org/10.1093/gigascience/giad042

In humans and livestock species, genome-wide association studies (GWAS) have been applied to study the association between variants distributed across the genome and a phenotype of interest. To discover genetic polymorphisms affecting the duodenum, liver, and muscle transcriptomes of 300 pigs from 3 different breeds (Duroc, Landrace, and Large White), we performed expression GWAS between 25,315,878 polymorphisms and the expression of 13,891 genes in duodenum, 12,748 genes in liver, and 11,617 genes in muscle. More than 9.68 × 1011 association tests were performed, yielding 14,096,080 significantly associated variants, which were grouped in 26,414 expression quantitative trait locus (eQTL) regions. Over 56% of the variants were within 1 Mb of their associated gene. In addition to the 100-kb region upstream of the transcription start site, we identified the importance of the 100-kb region downstream of the 3'UTR for gene regulation, as most of the cis-regulatory variants were located within these 2 regions. We also observed 39,874 hotspot regulatory polymorphisms associated with the expression of 10 or more genes that could modify the protein structure or the expression of a regulator gene. In addition, 2 motifs (5'-GATCCNGYGTTGCYG-3' and a poly(A) sequence) were enriched across the 3 tissues within the neighboring sequences of the most significant single-nucleotide polymorphisms in each cis-eQTL region. The 14 million significant associations obtained in this study are publicly available and have enabled the identification of expression-associated cis-, trans-, and hotspot regulatory variants within and across tissues, thus shedding light on the molecular mechanisms of regulatory variations that shape end-trait phenotypes.

18. Mutations on a conserved distal enhancer in the porcine C-reactive protein gene impair its expression in liver

C. Hernández-Banqué[1], T. Jové-Juncà[1], D. Crespo-Piazuelo[1], O. González-Rodríguez[1], Y. Ramayo-Caldas[1], A. Esteve-Codina[2-3], M.J. Mercat[4], M. Bink[5], R. Quintanilla[1], M. Ballester[1]

1. Animal Breeding and Genetics Program, Institute of Agrifood Research and Technology
2. CNAG-CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology
3. Universitat Pompeu Fabra
4. IFIP-Institut Du Porc and Alliance R&D
5. Hendrix Genetics Research Technology & Services B.V.

Frontiers in Immunology, Volume 14, 14 September 2023

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

https://doi.org/10.3389/fimmu.2023.1250942

C-reactive protein (CRP) is an evolutionary highly conserved protein. Like humans, CRP acts as a major acute phase protein in pigs. While CRP regulatory mechanisms have been extensively studied in humans, little is known about the molecular mechanisms that control pig CRP gene expression. The main goal of the present work was to study the regulatory mechanisms and identify functional genetic variants regulating CRP gene expression and CRP blood levels in pigs. The characterization of the porcine CRP proximal promoter region revealed a high level of conservation with both cow and human promoters, sharing binding sites for transcription factors required for CRP expression. Through genome-wide association studies and fine mapping, the most associated variants with both mRNA and protein CRP levels were localized in a genomic region 39.3 kb upstream of CRP. Further study of the region revealed a highly conserved putative enhancer that contains binding sites for several transcriptional regulators such as STAT3, NF-kB or C/EBP-β. Luciferase reporter assays showed the necessity of this enhancer-promoter interaction for the acute phase induction of CRP expression in liver, where differences in the enhancer sequences significantly modified CRP activity. The associated polymorphisms disrupted the putative binding sites for HNF4α and FOXA2 transcription factors. The high correlation between HNF4α and CRP expression levels suggest the participation of HNF4α in the regulatory mechanism of porcine CRP expression through the modification of its binding site in liver. Our findings determine, for the first time, the relevance of a distal regulatory element essential for the acute phase induction of porcine CRP in liver and identify functional polymorphisms that can be included in pig breeding programs to improve immunocompetence.

19. TAGADA: a scalable pipeline to improve genome annotations with RNA-seq data

C. Kurylo[1], C. Guyomar[1], S. Djebali[2], S. Foissac[1]

1. GenPhySE, Université de Toulouse, INRAE, ENVT
2. IRSD, Université de Toulouse, INSERM, INRAE, ENVT, UPS

NAR Genomics and Bioinformatics, Volume 5, Issue 4, 16 October 2023
https://doi.org/10.1093/nargab/lqad089

Genome annotation plays a crucial role in providing comprehensive catalog of genes and transcripts for a particular species. As research projects generate new transcriptome data worldwide, integrating this information into existing annotations becomes essential. However, most bioinformatics pipelines are limited in their ability to effectively and consistently update annotations using new RNA-seq data. Here we introduce TAGADA, an RNA-seq pipeline for Transcripts and Genes Assembly, Deconvolution, and Analysis. Given a genomic sequence, a reference annotation and RNA-seq reads, TAGADA enhances existing gene models by generating an improved annotation. It also computes expression values for both the reference and novel annotation, identifies long non-coding transcripts (lncRNAs), and provides a comprehensive quality control report. Developed using Nextflow DSL2, TAGADA offers user-friendly functionalities and ensures reproducibility across different computing platforms through its containerized environment. In this study, we demonstrate the efficacy of TAGADA using RNA-seq data from the GENE-SWiTCH project alongside chicken and pig genome annotations as references. Results indicate that TAGADA can substantially increase the number of annotated transcripts by approximately in these species. Furthermore, we illustrate how TAGADA can integrate Illumina NovaSeq short reads with PacBio Iso-Seq long reads, showcasing its versatility. TAGADA is available at github.com/FAANG/analysis-TAGADA.

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

20. Genomic architecture of carcass and pork traits and their association with immune capacity

T. Jové-Juncà[1], D. Crespo-Piazuelo[1], O. González-Rodríguez[1], M. Pascual[1], C. Hernández-Banqué[1], J. Reixach[2], R. Quintanilla[1], M. Ballester[1]

1. Animal Breeding and Genetics Program, Institute of Agrifood Research and Technology
2. Selección Batallé S.A

Animal, 28 November 2023
https://doi.org/10.1016/j.animal.2023.101043

Carcass and pork traits have traditionally been considered of prime importance in pig breeding programs. However, the changing conditions in modern farming, coupled with antimicrobial resistance issues, are raising the importance of health and robustness-related traits. Here we explore the genetic architecture of carcass and pork traits and their relationship with immunity phenotypes in a commercial Duroc pig population. A total of nine traits related to fatness, lean content and meat pH were measured at slaughter (~190d of age) in 378 pigs previously phenotyped (~70d of age) for 36 immunity-related traits, including plasma concentrations of immunoglobulins, acute phase proteins, leukocytes subpopulations and phagocytosis. Our study showed medium to high heritabilities and strong genetic correlations between fatness, lean content and meat pH at 24h post-mortem. Genetic correlations were found between carcass and pork traits and white blood cells. pH showed strong positive genetic correlations with leukocytes and eosinophils, and strong negative genetic correlations with hemoglobin, hematocrit and cytotoxic T cell proportion. In addition, genome-wide association studies (**GWAS**) pointed out four significantly associated genomic regions for lean meat percentages in different muscles, ham fat, backfat thickness, and semimembranosus pH at 24h. The functional annotation of genes located in these regions reported a total of 14 candidate genes, with *BGN*, *DPP10, LEPR, LEPROT, PDE4* and *SLC6A8* being the strongest candidates. After performing an expression GWAS (**eGWAS**) for the expression of these genes in muscle, two signals were detected in *cis* for the *BGN* and *SLC6A8* genes. Our results indicate a genetic relationship between carcass fatness, lean content and meat pH with a variety of immunity-related traits that should be considered to improve immunocompetence without impairing production traits.

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

Oral presentations

GENE-SWitCH Coordinator

Elisabetta Giuffra

Elisabetta Giuffra (female), biologist, coordinates GENE-SWitCH. After her PhD (1993) she worked on different animal genetics and genomics topics across Italy, France, Sweden and U.K. She is a research director at Paris-Saclay University, INRAE, AgroParisTech, GABI (Jouy-en-Josas, France) since 2011. She works in the field of host (pig)-virus interactions by transcriptional genomics approaches, and in local (e.g. FR-AgENCODE French pilot project) and international efforts to improve the functional annotation of farmed animal genomes for genotype-to-phenotype research (FAANG initiative). She has been co-Chair of the FAANG-Europe COST Action CA15112 (2016-2020) and co-leads the FAANG Steering Committee. More recently she works with organoids to elucidate genotype-to-phenotype relationships for immune response related traits in pigs and co-leads a work package (Development of a framework for biobanking and use of *in vitro* cellular models) of the EuroFAANG research infrastructure EU project since 2023.

INRAE, France
E-mail: elisabetta.giuffra@inrae.fr

1. The regulatory GENomE of Swine and CHicken: functional annotation during development

International Plant and Animal Genome Conference - Functional Annotations of Animal Genomes (FAANG) workshop - San Diego, USA - 10 January 2020

E. Giuffra[1], H. Acloque[1], C. Bénard[2], L. Tourneur[2], M. Watson[3], A. Archibald[3], M. Calus[4], J. Wells[4], G. Cochrane[5], F. Martin[5], C. Sabatel[6], C. Kaya[7], D. Eroglu[7], A. Rosati[8], M. Bink[9]

1. Paris-Saclay University, INRAE, AgroParisTech, GABI
2. INRAE Transfer
3. The Roslin Institute and R(D)SVS, University of Edinburgh
4. Wageningen University & Research, Animal Breeding and Genomics
5. European Molecular Biology Laboratory, European Bioinformatics Institute
6. Diagenode S.A., Epigenetics R&D
7. European Forum of Farm Animal Breeders
8. European Federation of Animal Science
9. Hendrix Genetics B.V., Research and Technology Center (RTC)

The H2020 project GENE-SWitCH started in July 2019 and will span 4 years. It 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 activation status of functional genome sequences varies across time and space, and in response to environmental perturbations. In full coordination and synergy with global effort and ongoing projects of the FAANG community, we will characterize the dynamics ("switches") of the functional genome from embryo (chicken) and fetus (pig) to adult life by targeting a panel of tissues relevant to sustainable production. New expression QTL data in pigs and existing high-resolution QTL data in chicken will be used for developing innovative genomic predictive models that integrate functional annotations, and these models will be validated in commercial pig and poultry populations. In addition,

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

nutritional epigenetic data will allow evaluation of the influence of maternal diet on the epigenome of the pig fetus and whether such effects persist until post-weaning. These open-shared datasets will conform fully with FAANG standards and add valuable knowledge on genetic and epigenetic variation of functional elements to FAANG. A comprehensive plan of dissemination and outreach activities to a large audience of stakeholders will be implemented. 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. This talk will provide an update of activities carried out in the first 6 months of the project.

2. The GENE-SWitCH H2020 project: follows up and perspectives of functional genome annotations

73rd EAAP Annual Meeting – Porto, Portugal - 5-9 September 2022

E. Giuffra[1], H. Acloque[1], A. Archibald[2], M. Bink[3], M. Calus[4], P. Harrison[5], C. Kaya[6], W. Lackal[7], F. Martin[5], A. Rosati[8], M. Watson[2], J. Wells[4]

1. Paris-Saclay University, INRAE, AgroParisTech, GABI
2. The Roslin Institute and R(D)SVS, University of Edinburgh
3. Hendrix Genetics B.V., Research and Technology Center (RTC)
4. Wageningen University & Research, Animal Breeding and Genomics
5. European Molecular Biology Laboratory, European Bioinformatics Institute
6. European Forum of Farm Animal Breeders
7. Diagenode S.A., Epigenetics R&D
8. European Federation of Animal Science

Session: 10: Novelties in genomics research and their impact on genetic selection

Since July 2019, the GENE-SWitCH (the regulatory GENomE of SWine and CHicken: functional annotation during development) consortium has been working to deliver underpinning knowledge on the pig and chicken genomes and to enable its translation to the pig and poultry sectors. Extensive functional genomics information has been produced for a panel of important tissues at early embryonic/fetal timepoints for both species. This information is being used to characterize the transcriptional and regulatory dynamics of the genome across development and is being integrated with existing data to deliver comprehensive genome annotation maps. Targeted studies aim to gain insights into the value of functional annotations for precision breeding, specifically by i) developing improved models for genomic prediction that use functional annotation combined with genotypic and phenotypic data to enhance the prediction accuracy of breeding values in commercial populations of both species, and (ii) assessing if and how varying fiber content in the maternal diet affects the epigenetic patterns of the pig fetus, and the possible persistence of epigenetic marks through weaning stage. Examples of current results will be illustrated along with the further needs and future strategies they highlighted.

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

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

International Plant & Animal Genome Conference - San Diego, USA - 13-18 January 2023

E. Giuffra[1], H. Acloque[1] and GENE-SWitCH Consortium

1. Paris-Saclay University, INRAE, AgroParisTech, GABI

The main aim of the EU-H2020 project GENE-SWitCH (the regulatory GENomE of Swine and Chicken: functional annotation during development) is to deliver new underpinning knowledge on the functional genomes of two main monogastric farm species (pig and chicken) and to enable its immediate translation to the pig and poultry sectors. Extensive functional genomics information has been produced at three developmental stages of chicken (E8, E15 and hatched) and pig (foetus at day 30 and day 70, and newborn) for a panel of 7 tissues selected from the project biorepository. The sequencing data were obtained by conventional short read RNA-seq, small RNA-seq, Iso-seq, ATAC-seq, whole genome bisulfite sequence (WGBS), reduced representation bisulfite sequence (RRBS) and promoter capture Hi C, while Chromatin immunoprecipitation sequence data (ChIP-seq of histone marks H3K4me1, H3K4me3, H3K27Ac, H3K27me3 and of CCCTC binding factor CTCF) is still in production. The Nextflow pipeline TAGADA was developed for the primary bioinformatic analyses of RNA-Seq data. For the other data, nf-core pipelines were either developed (for Iso-seq data) or refined, e.g. by developing an extension of the nf-core/methylseq pipeline (GSM) for downstream analyses. All available primary analysis outcomes and pipelines are shared in the dedicated FAANG Data portal page (https://data.faang.org/projects/GENE-SWitCH). Further analyses aim to identify tissue and time-point specific transcripts and regulatory regions that underly biological switches across development, with full integrative analyses scheduled later in 2023. GENE-SWitCH data have been processed already by partner EMBL-EBI to improve the annotation of protein-coding and non-coding RNA gene structures of the pig and chicken (both layer and broiler) genomes (https://projects.ensembl.org/gene-switch). The first Ensembl Regulatory Builds for the chicken and the pig will be key outputs from this project and will be delivered during 2023. To assess the value of using functional genome annotations for precision animal breeding, two approaches were taken. Genomic prediction models based on novel Bayesian and Machine Learning (Gradient Boosting Machine - GBM) approaches have been extended and tested to evaluate how prediction accuracy can be improved when including different functional and overlapping annotations. The validation of these models (in progress) makes use of the eQTL studies conducted in the project and of phenotyped swine and poultry commercial populations. A “diet x epigenetic” experiment on pigs has explored if and how varying fibre content in the maternal diet affects the epigenetic patterns of liver and muscle of the foetus (70 days), and the possible persistence of maternal diet effects in post-weaning piglets. Significant differences in chromatin accessibility were found prevalently at the piglet stage, with a strong representation of functions underlying metabolisms and immunity. Extensive dissemination and outreach activities have been implemented to transfer the outcomes of the project to stakeholder communities. GENE-SWitCH contributes to the global FAANG initiative and to EuroFAANG (https://eurofaang.eu), a synergy of European projects that share the common goal to discover links between genotype to phenotype in farmed animals and to meet global FAANG objectives.

The GENE-SWitCH project has received funding from the European Union’s Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

4. GENE-SWitCH: improving the functional annotation of pig and chicken genomes for precision breeding

74th EAAP Annual Meeting - Lyon, France - 26 August -1 September 2023

E. Giuffra[1], H. Acloque[1], A. Archibald[2], M. Bink[3], M. Calus[4], P. Harrison[5], C. Kaya[6], W. Lackal[7], F. Martin[5], A. Rosati[8], M. Watson[2], J. Wells[4]

1. Paris-Saclay University, INRAE, AgroParisTech, GABI
2. The Roslin Institute and R(D)SVS, University of Edinburgh
3. Hendrix Genetics B.V., Research and Technology Center (RTC)
4. Wageningen University & Research, Animal Breeding and Genomics
5. European Molecular Biology Laboratory, European Bioinformatics Institute
6. European Forum of Farm Animal Breeders
7. Diagenode S.A., Epigenetics R&D
8. European Federation of Animal Science

The H2020 GENE-SWitCH (the regulatory GENomE of SWine and CHicken: functional annotation during development) project has delivered extensive functional genomics information at three developmental stages (early and late organogenesis, newborn) for a panel of seven selected tissues. Data are publicly available (https://projects.ensembl.org/gene-switch); they are being processed to characterize biological switches during development and to generate the first Ensembl Regulatory Builds for the chicken (both layer and broiler) and the pig. To assess the interest of using genome annotations for precision animal breeding, two approaches were taken, coupled with dissemination and outreach activities aimed to transfer the outcomes of the project to stakeholder communities. A 'diet × epigenetic' experiment in pigs has investigated how the fiber content of the maternal diet can affect the epigenetic profiles of the liver and muscle of the offspring at the fetal and post weaning stages. In the second approach, genomic prediction models have been extended and tested to evaluate how prediction accuracy can be improved by incorporating different functional and overlapping annotations. The two project presentations will highlight the potential of this approach for swine and poultry breeding. GENE-SWitCH has received funding from the European Union's Horizon 2020 Research and Innovation Program under the grant agreement n° 817998.

5. Aims and Outcomes of H2020 GENE-Switch (The regulatory GENomE of SWineand CHicken: Functional Annotation during development)

International Plant & Animal Genome Conference - San Diego, USA - 12-17 January 2024

E. Giuffra[1], H. Acloque[1] and GENE-SWitCH Consortium

1. Paris-Saclay University, INRAE, AgroParisTech, GABI

The overall objective of H2020 GENE-SWitCH was to provide new fundamental knowledge on the functional genomes of two major monogastric farm species (pig and chicken) and to enable its immediate translation to the pig and poultry sectors. In order to characterize biological 'switches' during development and to generate the first Ensembl Regulatory Builds for chicken (both layer and broiler) and pig, we generated comprehensive functional genomic information at three developmental stages (early and late organogenesis, newborn) for a panel of seven selected tissues. Data, metadata, primary analyses and pipelines are open-shared in the GENE-SWITCH FAANG data portal(https://data.faang.org/projects/GENE-SWitCH) under the terms of the Fort Lauderdale Agreement and the Toronto Statement. To assess the interest in using genome annotations for precision animal breeding, two approaches were taken, coupled with several dissemination and

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

outreach activities aimed at transferring the results of the project to stakeholder communities. A 'diet ×epigenetics' experiment in pigs investigated how the fiber content of the maternal diet can affect the epigenetic profiles of the liver and muscle of the offspring at the fetal and post-weaning stages. In the second approach, genomic prediction models have been extended and tested to evaluate how prediction accuracy can be improved by incorporating different functional and overlapping annotations; an example of these results will be detailed in the next presentation. Overall, GENE-SWitCH has contributed to the generation of high-quality reference annotation maps for these two species and has produced innovative research results that support the value of functional genome annotation to address current and future challenges in sustainable production and pave the way for further genotype-to-phenotype research in livestock.

GENE-SWitCH has been funded by the European Union's Horizon 2020Research and Innovation Program under Grant Agreement No. 817998.

6. Differences in Maternal Diet Fiber Content Influence Patterns of Gene Expression and Chromatin Accessibility in Fetuses and Piglets

International Plant & Animal Genome Conference - San Diego, USA - 12-17 January 2024

S. Chalabi[1], L. Loonen[2], J. Boekhorst[2], H. Li[3], L. Fang[3], P. Harrison[4], W. Lakhal[5], J. Lluch[6], A. Sokolov[4], S. Djebali[7], A. Rau[1], Elisabetta Giuffra[1], Jerry Wells[2]

1. Paris-Saclay University, INRAE, AgroParisTech, GABI
2. Wageningen University & Research, Animal Breeding and Genomics
3. Center for Quantitative Genetics and Genomics, Aarhus University
4. European Molecular Biology Laboratory, European Bioinformatics Institute
5. Diagenode S.A., Epigenetics R&D
6. GeT-PlaGe, INRAE
7. IRSD, Université de Toulouse, INSERM, INRA, ENVT, UPS

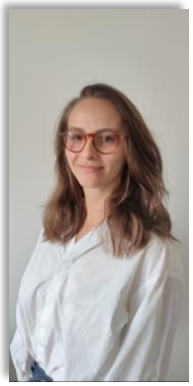
Pregnancy is a critical period of genome plasticity during which maternal diet may have long-lasting impact on the growth and health of offspring into adulthood. Short chain fatty acids (SCFAs) produced by gut microbiota can affect a wide range of physiological functions, and in human and other models have emerged as one clear link between the non-digestible fiber content of the diet and the epigenetic regulation of host cells and tissues, including muscle and liver. In this large-scale experimental study, we aimed to assess the effect of maternal diets varying in fiber content on the liver and skeletal muscle transcriptomes and epigenomes of fetuses and 10-week post-partum piglets. Three isocaloric diets (High Fiber-Beet pulp (HF-B), High Fiber-Pea (HF-P) and a Low Fiber control (LF)) were provided to groups of 7 sows each throughout two successive gestations. We established a collection of samples from 70-day-old fetuses and post-weaned piglets (2 males and 2females for each developmental stage per sow). The sow gut microbiota composition clustered as expected according to diet, but fecal SCFA levels displayed no significant differences. Piglet fecal microbiota and SCFA levels could only be partially explained by maternal diet. Liver and skeletal muscle of fetuses and piglets were profiled for chromatin accessibility (ATAC-seq) and gene expression (polyA+RNA-seq). Data quality checks and primary analyses were performed using the nf-core ATAC-seq and Nextflow TAGADA pipelines for ATAC-seq and RNA-seq data, respectively. As expected, principal components analyses revealed a strong separation among samples first by tissue, followed by developmental stage in both datasets. Differential analyses between pairs of diets were performed independently within each combination of tissue and developmental stage

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

using an empirical Bayes linear (ATAC-seq) or generalized linear (RNA-seq) model. A modest number of differential ATAC-seq peaks was found, most in the piglet muscle HF-B vs. LF diet comparison. A Transcription Factor (TF)foot printing analysis of the consensus ATAC-seq peaks revealed stage- and tissue- specific TF clusters with differential activity triggered by the HF-B diet. While no significantly differentially expressed genes were found in any combination of tissue and developmental stage, a gene set enrichment analysis (GSEA) of log-fold change expression values between the HF-B and control diets revealed an enrichment of metabolism and immunity related functions in both liver and muscle of piglets. A cell type deconvolution analysis on the bulk RNA-seq samples relying on a pig single cell atlas revealed the inter-individual variation in the proportion of cell types in liver and muscle under different diet and development groups and provided insights into targeted GSEA pathways across cell types.

These results provide a rich basis for future investigations aimed at identifying links between the functional genome and the phenotypes triggered by diet and other environmental factors in pigs. This study was funded by the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement No. 817998 (GENE-SWitCH).

Fanny Mollandin



Fanny Mollandin obtained her PhD in applied mathematics working on the integration of functional annotations in Bayesian genomic prediction models at the National Research Institute for Agriculture, Food and the Environment (INRAE). As part of the European GENE-SWitCH project, she developed and implemented the BayesRCO software for exploiting overlapping annotations to prioritize SNPs in order to improve genomic prediction and QTL mapping. More generally, she is interested in interdisciplinary research, combining statistics, computer science and genetics. She is continuing her post-doctoral research at the Center for Genomic Regulation in Barcelona, and is working on pancreatic islets regulation in diabetes context.

CRG, Spain

E-mail fanny.mollandin@crg.eu

7. Evaluating the interpretability of SNP effect size classes in Bayesian genomic prediction models

71st EAAP Annual Meeting - Virtual Meeting - 1-4 December 2020

F. Mollandin[1], P. Croiseau[1], A. Rau[1-2]

1. Université Paris-Saclay, INRAE, AgroParisTech, GABI
2. BioEcoAgro Joint Research Unit, INRAE, Université de Liege, Université de Lille, Université de Picardie Jules Verne

Non-linear Bayesian models represent an attractive approach to perform genomic prediction, due in part to their exibility and ability to perform variable selection. In recent years, a suite of models in the so-called Bayesian alphabet have been proposed to and a compromise between reality, which may correspond to an omnigenic model for complex traits, and computational convenience. In particular, the BayesR model (Erbe et al., 2012) strikes this balance by modeling single nucleotide polymorphisms (SNPs) as a mixture of markers with null, small, medium, or large variance. The

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

BayesRC model (MacLeod et al., 2016) in turn extends this approach by allowing different a priori categories of SNPs, as defined by previous biological knowledge, to be represented by different mixtures of effects. Although both of these approaches have been shown to achieve improved prediction accuracy in a variety of scenarios, there is still a need to evaluate the extent to which the assignment of SNPs to specific effect size classes (small, medium, large) reflects the true underlying genetic architecture and is meaningful for downstream SNP selection. In this work, we generate several sets of simulated data based on a real set of 50K genotype data in Montbeliarde bulls with a known pedigree, thus preserving the true linkage disequilibrium present in the data. We consider a wide variety of genetic architectures, phenotype heritability, and polygenic variances, and we illustrate under which conditions the BayesR and BayesRC models are able to recover known SNPs while also achieving high phenotypic prediction accuracy. Finally, we provide some insight into how the biological interpretation of the SNP classes identified by the BayesR and BayesRC models can be further refined through the incorporation of additional biological a priori information.

8. Intégration d'annotations biologiques complexes dans les modèles bayésiens de prédiction génomique: Evaluation et extension du modèle BayesRC

INRAE/GABI H2020 Virtual Seminar - 21 January 2022

F. Mollandin[1], P. Croiseau[1], A. Rau[1 2]

1. INRAE, AgroParisTech, GABI, Université Paris-Saclay
2. BioEcoAgro Joint Research Unit, INRAE, Université de Liège, Université de Lille, Université de Picardie Jules Verne

There is great interest in exploiting available biological information to guide genomic prediction models in the H2020 BovReg and GENE-SWitCH projects. In this talk, we discuss our recent work (1) evaluating the potential added value provided by functional annotations in cattle using the state-of-the-art model BayesRC (BovReg) and (2) developing a generic extension of BayesRC to account of multiple overlapping lists of functional annotations, implemented in the BayesRCO software (GENE-SWitCH).

9. Exploiting prior knowledge in genomic prediction models with BayesRCO

European Mathematical Genetics Meeting - Cambridge, United Kingdom - 21-22 April 2022

F. Mollandin[1], P. Croiseau[1], A. Rau[1-2]

1. INRAE, AgroParisTech, GABI, Université Paris-Saclay
2. BioEcoAgro Joint Research Unit, INRAE, Université de Liège, Université de Lille, Université de Picardie Jules Verne

The primary objective of genomic prediction is to use genomic variation, usually single nucleotide polymorphisms (SNPs), to predict phenotypes. However, some complex characters remain difficult to predict. In parallel, functional genomic knowledge obtained from omics studies holds promise for providing insight into relevant underlying cellular mechanisms and trait etiology. Integrating this information into genomic prediction models could potentially lead to a better understanding of complex traits. Bayesian models, through their use of priors, seem to be a straightforward way to introduce known functional information into genomic prediction models for complex traits, both to potentially improve prediction accuracy and give a better understanding of their genomic architecture. In particular, BayesRC (MacLeod et al., 2016) categorizes SNPs in multiple disjoint

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

annotations, allowing the proportion of QTLs to vary in each. Although this model shows good results, it is limited by the non-overlapping nature of the annotations, which prevents SNPs from belonging to more than one. As the number of available annotations increases, this becomes a strong limitation. We propose two novel Bayesian models to account for such multi-annotated SNPs, either attributing a cumulative (BayesRC+) or preferential (BayesRCπ) contribution across annotation categories. These two models are implemented in BayesRCO, an open source Fortran 90 software (https://github.com/fmollandin/BayesRCO). We show how the use of BayesRCπ and BayesRC+ through this software allows for a better exploitation of functional annotations and provides useful insight for complex traits.

10. Capitalizing on complex annotations in Bayesian genomic prediction for a backcross population of growing pigs

World Congress on Genetics Applied to Livestock Production - Rotterdam, The Netherlands - 3-8 July 2022

F. Mollandin[1], H. Gilbert[2], P. Croiseau[1], A. Rau[1,3]

1. INRAE, AgroParisTech, GABI, Université Paris-Saclay
2. GenPhySE, Université de Toulouse, INRAE, ENVT
3. BioEcoAgro Joint Research Unit, INRAE, Université de Liège, Université de Lille, Université de Picardie Jules Verne

Prior biological information has the potential to guide and inform genomic prediction models, but the BayesRC approach is currently limited to the use of disjoint categorizations of genetic markers. We propose two novel Bayesian approaches to model cumulative (BayesRC+) or preferential (BayesRCπ) contributions of multiple biological categories for multi-annotated SNPs. We illustrate the performance of these approaches on data from a backcross population of growing pigs in conjunction with several different sets of annotations related to multiple production traits constructed using the PigQTLdb.

Peter Harrison

Peter W. Harrison is the Genome Analysis Team Leader at EMBLs European Bioinformatics Institute. He leads GENE-SWitCH work package 3 that develops the FAANG Data Portal and GENE-SWitCH project page that presents the wealth of data generated by GENE-SWitCH to the world. Peter leads EMBL-EBIs efforts on animal agricultural and biodiversity coordination, as well as the Ensembl genome browser infrastructure and regulation teams. Peter is a member of the FAANG steering committee and chairs the metaFAIR Task Force, as well as a leading role in EuroFAANG coordination and communication.

EMBL-EBI, UK
E-mail: peter@ebi.ac.uk

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

11. The FAANG Data Coordination Centre

International Plant and Animal Genome Conference - San Diego, USA - 10 January 2020

P. Harrison

European Molecular Biology Laboratory, European Bioinformatics Institute

The Functional Annotation of Animal Genomes (FAANG) Project is a coordinated international effort to produce and collate high quality functional annotation of livestock genomes. The FAANG Data Coordination Centre (DCC) at EMBL-EBI has been developing the core infrastructure to support the global community to create this rich genome to phenome resource over a number of years. Three new European FAANG Horizon2020 projects started in the last year focusing upon cattle (BovReg), chicken and pigs (GENE-SWitCH) and Salmonids (AQUA-FAANG) that collectively provide a focal point for European FAANG research and are each supporting continued development of the FAANG DCC at EMBL-EBI. This is driving new requirements and development of FAANG metadata standards, validation and submission software and the FAANG data portal, of importance to the global FAANG community. The FAANG DCC also continue to be guided and active in the global FAANG steering committee and working groups. FAANG has a particular focus on ensuring high quality and rich supporting metadata to describe the project's samples and experimental assays. The DCC is launching a new improved metadata validation service that is hosted within the FAANG data portal (http://data.faang.org/home), rather than existing as a separate site, that will continue to ensure rich data submissions to the public archives. The documentation is also being shifted to the data portal so that all FAANG submission and retrieval data services are accessible from a single site. The data portal acts as a single access point for the wealth of livestock functional annotation data available from FAANG contributors combined with existing data available from public archives imported under legacy standards. The portal provides direct links for downloading data files direct from the public archives, with web-based bulk download support recently added, and supports programmatic API access. The DCC will continue to support the community with a dedicated helpdesk (dcc-faang@ebi.ac.uk) and is currently developing new software to assist the community in improving the suitability of ontologies for use in livestock context. Through effective standard driven metadata validation, a powerful search driven data portal and promotion of best practice in metadata implementation, the FAANG DCC aims to maximize effectiveness and inter-comparability of assay data, supporting the community to create a rich genome to phenome resource.

12. A vision for Bioinformatics within the Global FAANG project

International Plant and Animal Genome Conference - San Diego, USA - 10 January 2020

P. Harrison

European Molecular Biology Laboratory, European Bioinformatics Institute

The global FAANG project now consists of hundreds of researchers and multiple, separately funded projects, loosely coordinated by committees, mailing lists and informal communications. It has been an incredible success so far. One of the challenges of such a loosely coordinated project is to ensure reproducible research across multiple projects and species. Luckily, a suite of technologies now exist that enable reproducible research in bioinformatics and computational biology, at any scale. Workflow managers, software environments, software containers and cloud technologies ensure that we can all access the exact same pipelines, and ensure maximum reproducibility. Here we propose a vision for shared development of bioinformatics pipelines across the entire set of FAANG

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

(and associated) projects, based on the principles of open science, open source code and reproducible workflows and environments

13. The FAANG Data Coordination Centre: New European Perspectives for Our Continued Global Effort

2nd FAANG-Europe Workshop on Functional Annotation of Animal Genomes (FAANG) - Prague, Czech Republic - 11-13 February 2020

P. Harrison

European Molecular Biology Laboratory, European Bioinformatics Institute

The Functional Annotation of Animal Genomes (FAANG) Project is a coordinated international effort to produce and collate high quality functional annotation of livestock genomes. The FAANG Data Coordination Centre (DCC) at EMBL-EBI has been developing the core infrastructure to support the global community to create this rich genome to phenome resource over a number of years. Three new European FAANG Horizon2020 projects started in the last year focusing upon cattle (BovReg), chicken and pigs (GENE-SWitCH) and Salmonids (AQUA-FAANG), that collectively provide a focal point for European FAANG research and are each supporting continued development of the FAANG DCC at EMBL-EBI. This is driving new requirements and development of FAANG metadata standards, validation and submission software and the FAANG data portal, of importance to the global FAANG community. The FAANG DCC also continue to be guided and active in the global FAANG steering committee and working groups. With this new European focus, we are driving forward development of common coding standards, shared analysis pipelines and code sharing through the FAANG GitHub repository. FAANG has a particular focus on ensuring high quality and rich supporting metadata to describe the project's samples and experimental assays. The DCC is launching a new improved metadata validation service that is hosted within the FAANG data portal (http://data.faang.org/home), rather than existing as a separate site, that will continue to ensure rich data submissions to the public archives. The documentation is also being shifted to the data portal so that all FAANG submission and retrieval data services are accessible from a single site. The data portal acts as a single access point for the wealth of livestock functional annotation data available from FAANG contributors combined with existing data available from public archives imported under legacy standards. The portal provides direct links for downloading data files direct from the public archives, with web-based bulk download support recently added, and supports programmatic API access. The DCC will continue to support the community with a dedicated helpdesk (dcc-faang@ebi.ac.uk) and is currently developing new software to assist the community in improving the suitability of ontologies for use in livestock context. Through effective standard driven metadata validation, a powerful search driven data portal and promotion of best practice in metadata implementation, the FAANG DCC aims to maximize effectiveness and inter-comparability of assay data, supporting the community to create a rich genome to phenome resource.

14. Importance of rich 'FAIR' validated metadata, standardized protocols and analyses, and open-access datasets

G2P - AG2PI Field Day - Virtual Conference - 16 June 2021

P. Harrison

European Molecular Biology Laboratory, European Bioinformatics Institute

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

A hallmark of successful genetics/genomics consortia is a strong emphasis on data collection, description and storage that allows efficient and effective data sharing among the community members. FAANG has been a leader in emphasizing a strong and well-organized set of metadata standards and protocols from the beginning, and this talk will describe the components of this community resource developed at EMBL-EBI. The principles underlying this infrastructure are continuing to be developed through collaboration with three 2020-funded EuroFAANG projects (AQUA-FAANG, BovReg and GENE-SWitCH). EMBL-EBI recently appointed two new team leaders (Peter Harrison and Sarah Dyer) to collaboratively develop data resources supporting scientists using genomics to improve crop and animal agriculture, enhancing resources centered around the world-renowned Ensembl genome browser (https://www.ensembl.org/).

15. The Functional Annotation of Animal Genomes Project: Progress and challenges for our continued global effort

The 38th International Society for Animal Genetics Conference - Virtual - 26-30 June 2021

P. Harrison

European Molecular Biology Laboratory, European Bioinformatics Institute

The Functional Annotation of Animal Genomes (FAANG) Project is a coordinated international effort to produce and collate high quality functional annotation of animal genomes. It has a collaborative focus across multiple coordinated projects on the generation, annotation and presentation of FAIR open access and richly validated genotype and phenotype datasets. The datasets and accompanying supportive informatics infrastructure, that has been developed in recent years, advances the scope of comparative genomics and furthers the understanding of functional elements in these communities. The project provides terrestrial and aquatic animal agriculture communities powerful resources for supporting improvements to farmed animal production, disease resistance and genetic diversity. The FAANG Data Coordination Centre (DCC) at EMBL-EBI has been developing the core infrastructure to support the global community to create this rich genome to phenome resource. The focal point is the FAANG Data portal that hosts all of the rich datasets generated by the projects (http://data.faang.org). The EuroFAANG projects that focus upon cattle (BovReg), chicken and pigs (GENE-SWitCH) and Salmonids (AQUA-FAANG), collectively provide a focal point for European FAANG research and are supporting the continued development of the FAANG DCC infrastructure at EMBL-EBI. This talk focusses on the progress made within FAANG thus far, the main gaps in both datasets and infrastructure and the challenges community needs to address going forward. Some key infrastructure challenges include automated visualization, track hub enhancements, single cell atlases, cross referencing to biorepositories, and maintaining FAIR datasets at increasing scale. EMBL-EBI will continue to tackle these challenges with the FAANG community to develop the coordination and Data Portal technological developments for continued improvement in functional annotation of animal genomes. FAANGs recent whitepaper outlines the key scientific challenges and opportunities that the community needs to address to more accurately use genotype to predict phenotype in the world's farmed animal species, both terrestrial and aquatic to improve farmed animal health, welfare, efficiency and diversification.

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

16. EuroFAANG: empowering the future of FAANG to Fork research and infrastructure in Europe and beyond

International Plant & Animal Genome Conference - San Diego, USA - 8-12 January 2022

P. Harrison

European Molecular Biology Laboratory, European Bioinformatics Institute

FAANG (Functional Annotation of ANimal Genomes) is a coordinated international effort to produce and collate high quality functional annotation of animal genomes. To complement and accelerate this global effort, European FAANG (EuroFAANG; https://www.eurofaang.eu/) aims to harmonize FAANG research in Europe to create free resources that link genotype to phenotype (G2P) in order to tackle the current challenges in animal breeding, welfare and management to achieve sustainable productions. EuroFAANG brings together a wide range of expertise in farmed animal biology and breeding, genomics, bioinformatics, modelling and open data, as well as multiple platforms for dissemination and outreach, providing a focal point and close coordination of G2P research in Europe. EuroFAANG was founded by the three flagship Horizon 2020 funded projects AQUA-FAANG (https://www.aqua-faang.eu/), BovReg (https://www.bovreg.eu/) and GENE-SWitCH (https://www.gene-switch.eu/), and the FAANG Data Coordination Centre (DCC) at EMBL's European Bioinformatics Institute (EMBL-EBI). EuroFAANG is now expanding to include newly established FAANG-aligned projects funded by the Horizon 2020 and Horizon Europe programmes to exploit synergies and shared infrastructure and support G2P research. The existing EuroFAANG DCC infrastructure supports the community to openly share rich and standardized datasets to accelerate research. The data portal collates all of the comparable datasets into a single access point (https://data.faang.org/). EuroFAANG is already enabling scientists to collaborate on developing standardized protocols, analysis pipelines in particular in the nf-core (https://nf-co.re/) environment and coordinate comparative cross species analyses. The goals of EuroFAANG for the next decade rely on the scientific priorities outlined in the recent FAANG to Fork white paper by Clark et al. (https://doi.org/10.1186/s13059-020-02197-8). These research priorities include building single cell atlases, providing high throughput in vitro systems to screen large numbers of potentially causal variants, generating pangenomes, generating gene expression and phenotype information for large diverse populations of animals and creating in vitro biorepositories for organoids and cell lines. EuroFAANG is coordinating the development of a Europe-wide formalized infrastructure to facilitate the creation of common data structure and access for G2P research. This will be achieved by building on the joint support the FAANG DCC, Ensembl (https://www.ensembl.org), Elixir (https://elixir-europe.org) and EMBL-EBI cloud analysis infrastructures. A formalized infrastructure for EuroFAANG will also coordinate development, curation and biobanking of in vitro biological systems, expand capabilities in genome editing, and exploit synergies with existing European and global infrastructures to consolidate G2P research in farmed animals across Europe.

17. Harnessing the Ag Genomics Data torrent: A community-driven discussion on best practices for using and re-using genomics data

AG2PI Data Reuse Virtual Workshop - 9 February 2022

P. Harrison

European Molecular Biology Laboratory, European Bioinformatics Institute

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

The livestock and crop communities are accelerating the creation of functional genomics data to uncover the link between genotype and phenotype. The USDA-funded Agricultural Genome to Phenome Initiative is striving to identify the critical requirements to accelerate such research across these communities. We have identified that an important current and future requirement for this advancement is better methods for integrating disparate datatypes to allow new insights. For example, one specific goal of functional genomics is to predict regulatory elements across cells, tissues, individuals and populations. Tools are needed to easily help gather raw and create processed files and visualize the results of further analysis, such as a web-browser display of chromatin state maps, to extend the availability of such data to many more groups. In this Workshop, we will hear from speakers describing current solutions to data visualization and re-use in several genomics communities, as well as discuss as a group the paths to nurture a sustainable set of such tools. Our discussion is intended to include improving, standardizing and maintaining inputs (data and metadata), as well as creating, enhancing and supporting tools for mining such data.

18. The Global FAANG Data Portal and its role in the future EuroFAANG infrastructure

International Plant & Animal Genome Conference - San Diego, USA - 13-18 January 2023

P. Harrison

European Molecular Biology Laboratory, European Bioinformatics Institute

The Functional Annotation of Animal Genomes (FAANG) Project is a coordinated international effort to produce and collate high quality functional annotation of animal genomes. It has a collaborative focus across multiple coordinated projects on the generation, annotation and presentation of FAIR (findable, accessible, interoperable and reusable) open access and richly validated genotype and phenotype datasets. The datasets and accompanying supportive informatics infrastructure, that has been developed in recent years, advances the scope of comparative genomics, and furthers the understanding of functional elements in these communities. The project provides terrestrial and aquatic animal communities powerful resources for supporting improvements to farmed animal production, disease resistance and genetic diversity. Development of the FAANG Data Coordination Centre at EMBL-EBI has for several years been supported by the European Commission projects AQUA-FAANG, GENE-SWitCH, and BovReg. Collectively these projects have formed a coordinated European node of FAANG, that in 2022 was joined by three further projects GEroNIMO, RUMIGEN and HoloRuminant. This collaboration is now establishing itself as a future research infrastructure for Europe, EuroFAANG, with the overall aim to facilitate research and innovation for genotype to phenotype prediction in farmed animals (terrestrial and aquatic) to achieve sustainable, efficient and socially accepted farmed animal production. The FAANG DCC at EMBL-EBI will continue to develop and expand its services within this framework to coordinate large global FAANG 'omic datasets with a focus on FAIR principles as well as specific developments to support access and coordinate data for the emerging infrastructure. There is a continued need to create a common structure and data access service for all the datasets generated for farmed animals to consolidate research efforts and minimize redundancy. The Data Portal will also innovate new coordination and presentation solutions to support the emerging key scientific areas of the FAANG to Fork strategy such as pan-genomics and genome editing. The FAANG data portal (http://data.faang.org) acts as a single access point for the wealth of domesticated animal functional annotation data available from FAANG contributors combined with existing data available from public archives imported under legacy standards. I will highlight some of our recent developments such as the community ontology

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

improvement web service, and a recent portal solution to deal with the scale and complexity of track hubs beings generated by the FAANG community. The FAANG project, and the animal agriculture research community, shares many challenges with crop agricultural genomics, highlighting the importance of initiatives such as AG2PI and AgBioData. Some key infrastructure challenges include automated visualization, track hub enhancements, single cell atlases, cross referencing to biorepositories, standardization of ontologies, and maintaining FAIR datasets at increasing scale. EMBL-EBI will continue to tackle these challenges for animal and plant agricultural communities through the Ensembl project and our specific collaborations with consortia such as FAANG and Designing Future Wheat. It is important for FAANG members to engage with these global coordination initiatives. Through these collaborations we will enable researchers to more accurately use genotype to predict phenotype in the world's agricultural species, to improve health, welfare, efficiency and diversification.

GENE-SWitCH Deputy Coordinator

Hervé Acloque

Hervé Acloque is senior researcher in the INRAE Laboratory of Animal Genetics and Integrative Biology (GABI). After a PhD on the biology of chicken embryonic stem cells in Prof. Jacques Samarut's team in Lyon, he did a post-doctorate in developmental biology in Prof. Angela Nieto's team in Alicante (Spain) focusing on the molecular mechanisms controlling vertebrate gastrulation. He then obtained a permanent position at INRAE in 2010 and is currently developing research on innovative cellular models to understand the contribution of genetic variants to the construction of phenotypes in farmed species. Since 2014, he has been contributing to the international FAANG initiative to improve the functional annotation of animal genomes and since 2019, he has been co-coordinating the European GENE-SWitCH project to improve the functional annotation of pig and chicken genomes.

INRAE, France
E-mail: herve.acloque@inrae.fr

19. The EU H2020 project GENE SWitCH The regulatory GENomE of SWine and CHicken: functional annotation during development

International Plant & Animal Genome Conference - San Diego, USA - 11-15 January 2020

E. Giuffra[1], H. Acloque[1], C. Bénard[2], M. Watson[3], A. Archibald[3], M. Calus[4], J. Wells[4], P. Harrison[5], F. Martin[5], W. Lakhal[6], C. Kaya[7], D. Eroglu[7], A. Rosati[8], M. Bink[9]

1. Paris-Saclay University, INRAE, AgroParisTech, GABI
2. INRAE Transfert
3. The Roslin Institute and R(D)SVS, University of Edinburgh
4. Wageningen University & Research, Animal Breeding and Genomics
5. European Molecular Biology Laboratory, European Bioinformatics Institute
6. Diagenode S.A., Epigenetics R&D
7. European Forum of Farm Animal Breeders
8. European Federation of Animal Science
9. Hendrix Genetics B.V., Research and Technology Center (RTC)

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

The H2020 project GENE-SWitCH (2019-2023) aims to deliver new underpinning knowledge on the functional genomes of two main monogastric farm species (pig and chicken) in the framework of the Functional Annotation of Animal Genomes (FAANG) initiative and to enable immediate translation to the pig and poultry sectors. Inferences about the functional elements of a genome can be drawn from the analysis of a range of assays of both genomic DNA and transcribed RNA molecules. One main objective of GENE-SWitCH is to characterize the functional elements of the genomes of pigs and chicken that are active, poised or repressed ("switches") during development. The project will provide functional annotations maps covering the switches between early and late organogenesis and newborn animals in seven tissues (cerebellum, skin, kidney, skeletal muscle, lungs, liver, small intestine). For each stage and tissue, we will perform a set of 11 molecular assays (mRNA-seq, smallRNA-seq, Iso-seq, ATAC-seq, RRBS/WGBS, ChIP-seq against CTCF/H3K4me1/H3K4me3/ H3K27Ac/H3K27me3, promoter Capture Hi-C) on four biological replicates (2 males and 2 females). The production of these large sequence datasets is at advanced stage and analyses are in progress using dedicated tools to facilitate the identification of the functional elements in the pig and chicken genomes. We will present an overview of the objectives of the project and its current progress on this objective. A particular focus will be made on how we apply the FAIR principles (Findable, Accessible Interoperable, Reusable) to our samples, datasets and bioinformatics tools.

20. Midterm progress of H2020 GENE-SWitCH: samples collection and assays-by-sequence

International Plant & Animal Genome Conference - San Diego, USA - 8-12 January 2022

H. Acloque[1], W. Lakhal[3], P. Harrison[2], M. Davey[4], Ja. Smith[4], Jo. Smith[4], K. Miedzinska[4], M. Beinat[1], M. Mongellaz[1], A. Sokolov[2], O. Madsen[7], J. de Vos[7], S. Foissac[6], C. Kurylo[6], S. Djebali[5], S. Guizard[6], F. Martin[2], E. Giuffra[1], M. Watson[4]

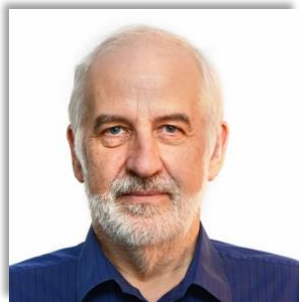
1. Paris-Saclay University, INRAE, AgroParisTech, GABI
2. European Molecular Biology Laboratory, European Bioinformatics Institute
3. Diagenode S.A., Epigenetics R&D
4. The Roslin Institute and R(D)SVS, University of Edinburgh
5. IRSD, Université de Toulouse, INSERM, INRA, ENVT, UPS
6. GenPhySE, Université de Toulouse, INRAE, ENVT
7. Wageningen University & Research, Animal Breeding and Genomics

GENE-SWitCH ("The regulatory GENomE of SWine and CHicken: functional annotation during development") 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 (https://www.gene-switch.eu). During development, several functional elements modulate the activity of the genome in order to reproduce a well-conserved developmental program that reflects inter-individual and sex-related differences as well as the adaptations to the constraints of in ovo and in utero environments. A central objective of GENE-SWitCH is to characterize the functional elements of the genomes of pigs and chicken which are active, poised or repressed ("switches") across three developmental phases (early and late organogenesis, and newborn). We are building functional annotation maps from seven tissues that represent the three layers of embryonic and fetal development (ectoderm: cerebellum, skin; mesoderm: kidney, skeletal muscle; endoderm: lung, liver, small intestine). Additional tissues have been bio banked. Sampling is complete and metadata have been deposited in BioSamples via the FAANG Data Portal. The generation of sequence data from a set of 11 molecular assays (RNA-seq, smallRNA-seq, Iso-seq,

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

ATAC-seq, RRBS / WGBS, ChIP-seq against CTCF / H3K4me1 / H3K4me3 / H3K27ac / H3K27me3, promoter Capture Hi-C) on four biological replicates per species (2 males and 2 females) for each stage and tissue is nearing completion. The data are currently being analyzed using standardized dedicated pipelines. We will present here an overview of this objective and its current progress. A particular focus will be made on how we apply the FAIR principles (Findable, Accessible Interoperable, Reusable) to our samples, datasets and bioinformatics tools.

Alan Archibald



Alan L. Archibald is a former Head of the Division of Genetics and Genomics at the Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh. He is internationally recognized in the field of farm animal genetics and genomics research. He co-led the first international farm animal genome project – PiGMaP. He played a leading role in the Swine Genome Sequencing Consortium that established the draft pig referenced genome sequence and co-led the generation of a revised high-quality pig reference genome sequence. He was a key initiator of the global Functional Annotation of ANimal Genomes project. His research is concerned with understanding the genetic control of complex traits, including production efficiency, product quality and host response to infectious disease, mainly in pigs. He is a Fellow of the Royal Society of Edinburgh which is Scotland's National Academy of Science and Letters and a Fellow of the International Society for Animal Genetics.

Roslin Institute, UK
E-mail: alan.archibald@roslin.ed.ac.uk

21. Extensive functional genomics information from early developmental time points for pig and chicken

World Congress on Genetics Applied to Livestock Production - Rotterdam, The Netherlands - 3-8 July 2022

H. Acloque[1], P. Harrison[2], W. Lakhal[3], F. Martin[2], A. Archibald[4], M. Beinat[1], M. Davey[4], S. Djebali[5], S. Foissac[6], S. Guizard[4], C. Guyomar[6], R. Kuo[4], C. Kurylo[6], O. Madsen[7], K. Miedzinska[4], M. Mongellaz[1], Ja. Smith[4], Jo. Smith[4], A. Sokolov[2], J. de Vos[7], E. Giuffra[1], M. Watson[4]

1. Paris-Saclay University, INRAE, AgroParisTech, GABI
2. European Molecular Biology Laboratory, European Bioinformatics Institute
3. Diagenode S.A., Epigenetics R&D
4. The Roslin Institute and R(D)SVS, University of Edinburgh
5. IRSD, Université de Toulouse, INSERM, INRA, ENVT, UPS
6. GenPhySE, Université de Toulouse, INRAE, ENVT
7. Wageningen University & Research, Animal Breeding and Genomics

The global Functional Annotation of Farm Animal Genomes initiative (FAANG) aims to improve animal breeding by improved genomic prediction via integration of functional genomics information. The GENE-SWitCH project has produced extensive functional genomics information for a variety of important tissues at early embryonic timepoints for both chickens and pigs. These datasets will be integrated to produce both tissue and time-point specific transcript, gene, and

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

regulatory annotation for both species. In this paper, we describe the aims of the project, and the initial release of both raw and processed data.

Sébastien Guizard

After having put my first foot in Biology in high school, I put the second in Bioinformatics from my third year of study. Driven by the search for answers, but above all by the search for new questions, I have enriched my knowledge over the years. First, in Biology, with a passage to the bench, in Computer science by learning to program, to tame Linux and finally in bioinformatics by deploying "whole genome" study projects. After a master degree in Bioinformatics, I worked for one year as an engineer on cereal paleo genomics before starting my PhD. It focused on the organization of the chicken genome through the annotation of repeated sequences. I started my career as Bioinformatician for a French international plant breeder, where I could continue genome analysis. Today, I'm a core scientist of Roslin Institute. I was in charge of the annotation of the pig and chicken genome with PacBio ISOseq.

Exseed, UK

E-mail: sguizard@exseed.ed.ac.uk

22. nf-isoseq pipeline provides a first insight onto chicken transcripts landscape

The 38th International Society for Animal Genetics - Virtual Conference - 26-30 July 2021

S. Guizard, K. Miedzinska, Ja. Smith, Jo. Smith, R. Kuo, M. Davey, M. Watson

The Roslin Institute and R(D)SVS, University of Edinburgh

The EU H2020 GENE-SWitCH project is a collaborative endeavor that will produce new genome information to enable the characterization of genetic and epigenetic determinants of complex traits in the two monogastric species (chicken and pig) that are the primary sources of meat worldwide. The characterization of genome functional elements is being made through a wide diversity of analysis methods (RNASeq, miRNA-Seq, ATAC-Seq, WGBS, ChIP-Seq, IsoSeq, Promoter Hi-C), on seven tissues at three developmental stages (pig: 30, 70 days post-fertilization, new born and chicken: E8, E15, hatched). Functional annotation of the genome partly relies on RNA sequencing. Illumina short read sequencing has been extensively used, as it is low cost, labour saving and rapid. However, short sequence lengths make isoform reconstruction challenging. This drawback can be overcome with long read data. Iso-Seq, with circular cDNA sequencing, generates accurate full-length transcripts (<1% error). It allows direct identification of gene intron-exon structure and thus better isoform detection. This method has growing utility and has been used to re-annotate genomes but no standard pipeline has yet been released. We have thus developed the nf-isoseq pipeline, which enables insight into the chicken transcriptomic landscape. nf-core is a community initiative working on the development of standard, portable and reproducible analysis pipelines. The pipeline generates CCS (Circular Consensus Sequence), maps transcripts to the genome and reduces gene model redundancy with TAMA collapse. Thanks to the nextflow language and containerization technologies, it can be easily run on a wide range of cluster configurations.

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

We processed the Iso-Seq data for each chicken sample to reveal the transcriptome for a given tissue at a given time point. Comparison of annotations gives a first look at transcriptional landscapes across tissues and the dynamic over time development.

23. Automatic and simple Iso-Seq genome annotation with nf-core/isoseq

International Plant & Animal Genome Conference - San Diego, USA - 13-18 January 2023

S. Guizard, K. Miedzinska, Ja. Smith, Jo. Smith, R. Kuo, M. Davey, M. Watson

The Roslin Institute and R(D)SVS, University of Edinburgh

The detection and annotation of transcribed and functional regions of a genome has been and remains a complex task. The invention of high-throughput short read sequencing technology, allowed easier genome annotation at the cost of heavy computation and non-exhaustive isoforms detection. Iso-Seq, with circular cDNA sequencing, generates accurate full-length transcripts (<1% error). It allows direct identification of gene intron-exon structure and thus better isoform detection. This method has growing utility and has been used to re-annotate genomes but no standard pipeline has yet been released. The nf-core/isoseq pipeline we developed, enables fast and simple Iso-Seq genome annotation. The pipeline generates FLNCs sequences, maps transcripts to the genome and reduces genes model redundancy with TAMA collapse. The nf-core bioinformatics community support portable and reproducible pipeline and provide development guidelines. The pipeline must be written in nextflow and each program must be available as a container. Thanks to these requirements, the pipeline can be easily run on a wide range of cluster configurations. The pipeline allows simple and fast Iso-Seq genome annotation and is freely available on nf-core website (https://nf-co.re/isoseq) and github (https://github.com/nf-core/isoseq).

24. Exploring early transcripts landscape with Isoseq

International Plant & Animal Genome Conference - San Diego, USA - 8-12 January 2022

S. Guizard, K. Miedzinska, Ja. Smith, Jo. Smith, R. Kuo, M. Davey, M. Watson

The Roslin Institute and R(D)SVS, University of Edinburgh

The EU H2020 GENE-SWitCH is a collaborative endeavor that will produce new genome information to enable the characterization of genetic and epigenetic determinants of complex traits in the two monogastric species (chicken and pig) that are the primary sources of meat worldwide. The characterization of genome functional elements is being made through a wide diversity of analysis methods (RNASeq, miRNA-Seq, ATAC-Seq, WGBS, ChIP-Seq, IsoSeq, Promoter Hi-C), on seven tissues at three developmental stages (pig: 30, 70 days post-fertilization, new born and chicken: E8, E15, hatched). Iso-Seq, with circular cDNA sequencing, generates accurate full-length transcripts (<1% error). It allows direct identification of gene intron-exon structure and thus better isoform detection. This method has growing utility and has been used to re-annotate genomes but no standard pipeline has yet been released. We are thus developing an nf-core isoseq pipeline, which enables insight into the chicken early transcriptomic landscape. The pipeline generates HIFI sequences, maps transcripts to the genome and reduces gene model redundancy with TAMA collapse. Thanks to the nextflow language and containerization technologies, it can be easily run on a wide range of cluster configurations. We processed the Iso-Seq data for each chicken sample to

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

reveal the transcriptome for a given tissue at a given time point. Comparison of annotations gives a first look at transcriptional landscapes across tissues and the dynamic over time development.

Marco Bink

Marco Bink was born in 1968 and obtained his MSc in Animal Science early nineties, and subsequently a PhD degree in Animal Breeding and Genetics from Wageningen University in the Netherlands. He continued at Wageningen University and Research while expanding his horizon towards statistical research in plant breeding and genomics. In many collaborative research projects, Marco very much enjoyed the interaction with clients and collaborators to better understand and respond to their needs and questions. In March 2016 he started as a senior quantitative geneticist at Hendrix Genetics, a world leading multispecies animal breeding company. In his present position as a team lead Genetics and Genomics he takes responsibility for providing and supporting software tools for fast and accurate genetic evaluations in all breeding programs of Hendrix Genetics.

Hendrix Genetics, Netherlands
E-mail: marco.bink@hendrix-genetics.com

25. Predictive ability of genomic prediction in layers when including CADD scores as genome function information

World Congress on Genetics Applied to Livestock Production - Rotterdam, The Netherlands - 3-8 July 2022

M. Bink[1], M. Calus[2], M. Derks[2], J. Visscher[3], B. Perez[1]

1. Hendrix Genetics B.V., Research and Technology Center (RTC)
2. Wageningen University & Research, Animal Breeding and Genomics
3. Institut de Selection Animale B.V.

Accuracy of genomic prediction is key for genetic progress in breeding programs and including genome functional annotation may help. Here we report on the added value of Combined Annotation Dependent Depletion (CADD) scores to weigh SNPs in GBLUP analyses. Multiple transformations of CADD scores were considered and empirically validated on a layer dataset including 5 traits and 18K animals with 27K SNP genotypes. For 3 traits, the use of (squared) CADD scores yielded marginally higher accuracy. We anticipate that the added value of CADD scores is hampered by the number of SNPs in this study and is likely more pronounced when using whole genome sequence SNPs.

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

Sylvain Foissac

Sylvain Foissac is a senior researcher with ~20 years of experience in omics data analysis. His expertise spans functional genomics, sequencing data analysis, gene expression regulation, and 3D genomics. His primary research interest now centers on unraveling the intricate links between genome structure and function. After completing his Ph.D. in Bioinformatics in Toulouse, France, Sylvain gained valuable experience through a postdoctoral position at the CRG in Barcelona, as well as positions in industry at Affymetrix in California, USA, and Integromics in Madrid, Spain. Following this international experience, Sylvain returned to Toulouse and has been serving as a Research Scientist at INRAE since 2012. Sylvain has made significant contributions to various genome annotation projects, including the ENCODE project and the FAANG action, in order to enhance our understanding of life at the molecular level. He is the author of about 30 peer-reviewed articles and book chapters in the field of genomics and computational biology.

INRAE, France
E-mail: sylvain.foissac@inrae.fr

Cervin Guyomar

Cervin Guyomar is an engineer initially trained in agronomical sciences and statistics. In 2018, he completed his PhD in bioinformatics in the Inria institute in Rennes, France. His researched there focused on the fine-scale characterization of insect microbiotas through metagenomics. Following this, he worked as a bioinformatician first at iDiv in Leipzig, Germany and then at INRAE in Rennes, France. During that time, he worked on different plant and animal models, on subjects including genome assembly, structural variant detection and microRNA annotation. In 2020, he was recruited as a permanent research engineer at INRAE Toulouse, where he serves on the Sigenae platform. His work now primarily revolves around the functional characterization of animal genomes, with a particular involvement on the GENE-SWitCH and AquaFaang projects.

INRAE, France
E-mail: cervin.guyomar@inrae.fr

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

Sarah Djebali

Initially trained in computer science, I did a phD in bioinformatics on eukaryotic gene finding. I then did a long postdoctoral stay at the Center for Genomic Regulation in Barcelona (Roderic Guigó), where I was deeply involved in several international functional genomic consortia. Our work in ENCODE allowed us to deeply characterize the human transcriptome, on many cell types and with several complementary techniques, and in the context of mouse ENCODE, we provided a deep comparison of the mouse and human transcriptomes and produced a list of tissues for which mouse was a good model for human. I then came back to France to work on the FR-AgENCODE project, to better understand genome function in four terrestrial livestock species. In 2019 I became a permanent INSERM researcher at IRSD, a Toulouse lab working on digestive health, where I am working on better understanding gene expression regulation, and its application to both human common diseases and agronomics.

INSERM, France
E-mail: sarah.djebali@inserm.fr

26. TAGADA: Transcripts and Genes Assembly, Deconvolution, Analysis; a Nextflow pipeline to improve genome annotations with RNA-seq data

PRBB-CRG Joint seminar - Barcelona, Spain - 6 May 2022

C. Kurylo[1], C. Guyomar[1], S. Djebali[2], S. Foissac[1]

1. GenPhySE, Université de Toulouse, INRAE, ENVT
2. IRSD, Université de Toulouse, INSERM, INRAE, ENVT, UPS

In large scale genome annotation projects, multiple RNA-seq datasets are typically generated and/or analyzed to characterize the transcriptome of various tissues or cell types. While many available RNA-seq pipelines can build disparate de novo transcript models or quantify transcript expression levels using a provided annotation, few allow cohesive transcript reconstruction and expression assessment from multiple RNA-seq datasets in a reproducible way. To fill this gap, in the context of the GENE-SWitCH project, we have developed TAGADA: Transcripts and Genes Assembly, Deconvolution, Analysis. TAGADA combines several reference RNA-seq bioinformatics tools into a streamlined pipeline. The current version aligns single-end and paired-end reads with STAR, reconstructs and quantifies transcripts and genes with StringTie, detects and characterizes long non-coding RNAs (lncRNAs) with FEELnc, and generates a summarized MultiQC report compiling multiple quality control figures and tables. Built with Nextflow DSL2, TAGADA is designed to be easy to use and flexible. It provides a containerized environment compatible with a variety of high-performance computing platforms and workload orchestrators. A minimal set of inputs is required: RNA-seq read files, a reference genome and a genome annotation. Optionally, a generic metadata file can be provided to describe the experimental design and seamlessly merge samples. Among the generated quality controls are interactive charts and tables with statistics and metrics for various steps of the workflow. Expression tables are also provided with read counts for annotated and predicted genes and transcripts, allowing further comparative expression analyses. We believe that the TAGADA pipeline offers a useful, flexible and reproducible way to process RNA-seq data, nicely complementing the existing nf-core catalog of bioinformatics tools and contributing to the FAANG global action.

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

27. Improving genome annotations with RNA-seq data: a scalable and reproducible workflow for Transcripts and Genes Assembly, Deconvolution, Analysis (TAGADA).

21st European Conference on Computational Biology - Sitges, Spain - 12-21 September 2022

C. Kurylo[1], C. Guyomar[1], S. Djebali[2], S. Foissac[1]

1. GenPhySE, Université de Toulouse, INRAE, ENVT
2. IRSD, Université de Toulouse, INSERM, INRAE, ENVT, UPS

Genome annotation aims to provide a comprehensive, accurate and nonredundant catalog of all reported genes and transcripts for a given species. Research projects worldwide routinely generate new transcriptome data to be integrated into existing annotations in a consistent manner. While many bioinformatics pipelines can build disparate de novo transcriptomes or quantify gene expression levels from a provided annotation, they usually do not properly update a reference annotation using new RNAseq data, in particular for large numbers of samples. Here we present TAGADA, an RNA-seq pipeline for Transcripts and Genes Assembly, Deconvolution, Analysis. Given a genomic sequence, a reference annotation and RNA-seq short reads, TAGADA extends the set of provided gene and transcript models with new ones, generating an improved annotation. In addition, it computes expression values for both the reference and the novel annotation, detects long noncoding transcripts (lncRNAs), and generates a comprehensive report with multiple quality controls. Built with Nextflow DSL2, TAGADA is easy to use and provides a containerized environment to ensure reproducibility across computing platforms. TAGADA has been used in the context of the FAANG action (Functional Annotation of ANimal Genomes) to detect thousands of new coding and non-coding.

Jani de Vos

Jani de Vos was born in South Africa. Because of her interests in Animals she enrolled in the NAimal Science Bachelor program in 2013 and obtained a BSc(Agric) Animal sciences undergraduate in 2016 at the University of Pretoria. She then continued with her MS, on the study of Nguni cattle where she wrote her thesis on "A Genome wide association study of carcass traits based on Real-time ultrasound in South African Nguni cattle". After graduating in 2018, Jani left South Africa, to start her PhD at Wageningen University. Here her PhD formed a part of the GENE-SWitCH consortium where she investigated functional genomics regulating development in pig and chicken. During her PhD she went on a research visit to a GENE-SWitCH partner: the European bioinformatics institute in Hinxton. There she developed a framework for integrating different functional data and for comparative research between pig and chicken. She submitted her thesis manuscript in August 2023 and will defend her thesis in December 2023.

Wageningen University & Research, Netherlands
E-mail: jani.devos@wur.nl

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

28. DNA methylation dynamics regulating early embryonic development in pig and chicken

The 39th International Society for Animal Genetics Conference - Cape Town, South Africa - 2-7 July 2023

J. de Vos[1], M F.L. Derks[1], H. Acloque[2], S. Djebali[3], S. Foissac[4], C. Guyomar[4], C. Kurylo[4], E. Giuffra[2], M A.M. Groenen[1], O Madsen[1]

1. Wageningen University & Research, Animal Breeding and Genomics
2. Paris-Saclay University, INRAE, AgroParisTech, GABI
3. IRSD, Université de Toulouse, INSERM, INRA, ENVT, UPS
4. GenPhySE, Université de Toulouse, INRAE, ENVT

The goal of the Functional Annotation of Farm Animal Genomes (FAANG) consortium is to improve animal breeding by incorporating functional genomics data into genomic prediction models. GENE-SWitCH is a project under the FAANG umbrella with the aim of providing the functional genomic annotation for a number of key tissues in both pigs and chickens during early embryonic stages. DNA methylation is an epigenetic modification which plays a crucial role in mammalian development, however there is still limited understanding of the overarching dynamics in the developing embryo. In this research we used both whole genome- and reduced representation bisulphite sequencing (WGBS and RRBS) data to evaluate the methylome dynamics during development at 30 days post fertilization (dpf), 70dpf, and new born (NB) of seven tissues (liver, kidney, brain, muscle, skin, small intestine and lung) in pig. Developmental transitions were investigated using a two-fold approach: 1) Dividing the methylome into unmethylated regions (UMR), indicative of promoters, and lowly methylated regions (LMR) indicating enhancers, and 2) a differential methylation analysis. The number of UMRs across developmental stages within the various tissues ranged from 10,822 to 14,680 and from 35,073 to 75,477 for LMRs. The defined methylation states (UMRs, LMRs and fully methylated regions) were used to define the dynamic changes of the methylome and cis-regulatory elements during embryological development. The most notable finding was the shift of methylation states, from hypo to hyper methylation, in liver 70dpf to NB. Lastly, differentially methylated regions were integrated and combined with differentially expressed genes from the same samples. The combined differentially methylated and expressed genes were involved in biological pathways related with general growth, i.e. regulation of developmental processes, during initial stages of development (30dpf to 70dpf), and tissue specific functions during maturation transition (70dpf to NB). This trend was observed for most tissues, except in liver and brain which showed tissue specific functions during early developmental stages like e.g. neurogenesis in brain.

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

Maria Ballester

Maria Ballester (DVM, PhD in Animal Production) is currently a senior researcher at the Animal Breeding and Genetics Program at IRTA and an associate professor in the Department of Animal and Food Sciences at UAB. She has a broad experience in animal genetics focused on immunogenetics, animal health, functional genomics, and bioinformatics. In recent years, she has established a research line on the genetic determinism of global immunocompetence in pigs with the final aim of identifying functional genetic variants for implementing genomic selection for robustness and disease resistance in pigs. During her career, she has participated in 14 national projects and 8 international projects (5 EU projects). She is also involved in different RDI contracts with the swine sector. She has co-authored more than 80 peer-reviewed scientific articles (H-index: 27), co-supervised eight PhD thesis (4 ongoing) and she is coinventor of two patents.

IRTA, Spain
E-mail: maria.ballester@irta.cat

29. Multi-breed, multi-tissue, and multi-omics aiding the quest for key porcine regulators

The 39th International Society for Animal Genetics Conference - Cape Town, South Africa - 2-7 July 2023

D. Crespo-Piazuelo[1], A. Reverter[2], Y. Ramayo-Caldas[1], R. Quintanilla[1], H. Acloque[3], M.J. Mercat[4], M. Bink[5], A.E. Huisman[5], M. Ballester[1]

1. Animal Breeding and Genetics Program, Institute of Agrifood Research and Technology
2. CSIRO Agriculture and Food
3. Paris-Saclay University, INRAE, AgroParisTech, GABI
4. IFIP-Institut du porc and Alliance R&D
5. Hendrix Genetics B.V., Research and Technology Center (RTC)

This study aims at identifying breed and tissue-specific key regulators in pigs by using co-expression and co-association gene network analysis. For that purpose, duodenum, liver, and muscle samples were obtained at slaughter from 300 pigs of three different breeds: Duroc, Landrace, and Large White (n=100 each). Whole-genome sequencing and RNA-seq were performed on the Illumina NovaSeq6000 platform. RNA counts were quantified by RSEM/1.3.0 and normalized by TMM (trimmed mean of M-values). Lowly expressed genes and those missing in more than 20% of the animals were removed, remaining 13,891 genes expressed in duodenum, 12,748 genes in liver and 11,617 genes in muscle. Genetic variant calling, conducted with GATK/4.1.8.0 HaplotypeCaller, resulted in 44,127,400 polymorphisms (SNPs and indels) among all the individuals. After removing those variants with a minor allele frequency below 5% and more than 10% missing genotype data on each breed, 25,224,146 polymorphisms were kept among the three breeds. eGWAS were conducted using the fastGWA tool from GCTA/1.93.2. Gene co-expression networks were inferred using the PCIT algorithm and the list of genes encoding transcription factors (n=1,109) and co-factors (n=869), and the 100 most expressed genes in each tissue. Thus, a total of 2,248 genes remained to generate regulatory gene networks. Across breeds the transcription factors CTCF, SP3 and TOX4 and the cofactors CHTOP, ELOB and NCL were identified as the highest connected regulators in the duodenum, liver and muscle co-expression gene networks, respectively. Within breeds, KHSRP, TOX4 and CSDE1 transcription factors and HNRNPU, CRK and CRK cofactors showed the highest

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

connectivity in Duroc, Landrace and Large White, respectively. Our results identified putative key regulatory genes that will be reassessed by the eGWAS analysis. These findings bring us closer to understanding gene expression regulation in porcine key tissues and will allow us to improve genomic evaluation procedures by considering this functional information.

Garth Ilsley

Garth Ilsley has led Ensembl Regulation since 2019. His team develops and runs bioinformatic workflows to produce regulatory annotation for a growing number of species. Applying data analysis methods to support workflow optimization and quality improvements are an integral part of the team. Prior to joining Ensembl, Garth worked as a researcher developing statistical and machine learning models of enhancer regulation using cellular resolution datasets of gene expression. He has more than 6 years' experience in industry as a project lead, business analyst and technical architect developing and delivering software solutions to clients in the financial and pharmaceutical sectors.

EMBL-EBI, UK
E-mail: garth.ilsley@ebi.ac.uk

30. Extending Ensembl Regulatory Annotation to Farmed Animals

The 39th International Society for Animal Genetics Conference - Cape Town, South Africa - 2-7 July 2023

G. Ilsley, G.A. Merino, P.R. Branco Lins, M. Perry, D. Urbina-Gómez, P. Harrison

European Molecular Biology Laboratory, European Bioinformatics Institute

Ensembl (https://www.ensembl.org/) is a widely used genome browser that has assisted the scientific community in interpreting the genome for more than twenty years. Ensembl's regulatory annotation identifies regions in the genome that might regulate and control the expression of nearby genes. These regulatory features can be used to filter and identify regions of the genome where non-coding variation or targeted mutations could have important consequences for gene expression and hence phenotype. Ensembl has well established regulatory builds for human and mouse, which grew out of work on the ENCODE and BLUEPRINT projects. Now for the first time, we are extending this analysis beyond human and mouse. In collaboration with the GENE-SWitCH and AQUA-FAANG consortia, Ensembl is making regulatory annotation available for farmed animals including Pig, Chicken, Atlantic salmon and Turbot. EMBL's European Bioinformatics Institute also develops the FAANG Data Portal (https://data.faang.org/), which together with its rich and standardized metadata enables a wide range of experimental datasets for farmed animals to be discovered. Ensembl Regulation's computational workflow relies on this standardized metadata to retrieve and process primary datasets to identify open chromatin (ATAC-seq) and histone marks (ChIP-seq). These are then combined to produce genomic-level regulatory annotation. The analysis process will be described, along with examples of the resulting annotation and how it might be interpreted to guide further studies. Work is ongoing to add additional capabilities and further farmed animal species, and we are seeking feedback and collaborations with the community.

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

Andrea Rau

Andrea Rau is a research director and biostatistician at the French National Research Institute for Agriculture, Food, and the Environment (INRAE) in the Animal Genetics and Integrative Biology (GABI) research unit in Jouy-en-Josas, France. Her research centers on methodological developments for the analysis of complex, high-dimensional molecular data, in particular for the analysis of transcriptomic data, multi-omic integration, and biologically-informed genomic prediction models. She enjoys working on interdisciplinary problems at the interface of statistics and biology, in close collaboration with experts in animal, plant, and human genomics.

INRAE, France
E-mail: andrea.rau@inrae.fr

31. Accounting for overlapping annotations as biological priors in genomic prediction models of complex traits

AQUA-FAANG Project Final Meeting - Edinburgh, United Kingdom - 11-13 October 2023

F. Mollandin[1], H. Gilbert[3], P. Croiseau[1], A. Rau[1 2]

1. INRAE, AgroParisTech, GABI, Université Paris-Saclay
2. BioEcoAgro Joint Research Unit, INRAE, Université de Liège, Université de Lille, Université de Picardie Jules Verne
3. GenPhySE, INRAE, ENVT, Université de Toulouse

It is now widespread in farm animal and plant breeding to use genotyping data to predict phenotypes with genomic prediction models. Functional genomic annotations (e.g., the accessibility of chromatin or methylation status in relevant tissues), have the potential to provide valuable insight into the location and effect size of causal genetic variants underlying complex traits. Developing and validating genomic prediction models able to fully leverage such complex functional annotations for improved accuracy and interpretability was one of the aims of the H2020 GENE-SWitCH project. To this end, we defined and implemented a flexible framework for genomic prediction called BayesRCO to simultaneously take advantage of the availability of multiple functional genomic annotations. In this talk, I will describe the intuition behind our proposed model and discuss some of our key take-away messages from early results.

Mario Calus

Mario Calus is an associate professor at Wageningen University, and his main area of expertise is on genomic prediction and genomic selection of a variety of livestock species. Since obtaining his PhD degree in 2006, he has 17 years of experience in the scientific development of genomic selection, and its implementation in collaboration with the breeding industry. He has been involved, as PIs, in the Dutch public-private partnership Breed4Food since 2012, and is currently involved in the Horizon 2020 projects GENE-SWitCH and GEroNIMO. He has supervised 11 PhD students, with an additional 7 ongoing, and has trained 7 postdocs. He

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

has (co)authored 159 peer reviewed articles. He is Senior Editor for the section Complex Traits of GENETICS, and is one of the co-Editors-in-Chief of Genetics Selection Evolution.

Wageningen University & Research, Netherlands
E-mail: mario.calus@wur.nl

32. Using functional annotation and individual omics in genomic prediction

The 39th International Society for Animal Genetics Conference - Cape Town, South Africa - 2-7 July 2023

M. Calus[1], B. Perez[2], J. de Vos[1], O. Madsen[1], L. Ayres[1], H. Bovenhuis[1], M. Ballester[3], M.J. Mercat[4], M. Bink[2]

1. Wageningen University & Research, Animal Breeding and Genomics
2. Hendrix Genetics B.V., Research and Technology Center (RTC)
3. Animal Breeding and Genetics Program, Institute of Agrifood Research and Technology
4. IFIP-Institut du porc and Alliance R&D

Since the advent of genomic prediction, considerable efforts have been made to improve its accuracy by aligning the model more closely with the underlying biology of complex traits. In line with research aims to close the genotype-phenotype gap, intermediate omics have been used as additional information in genomic prediction. Broadly speaking, there are 2 approaches to do this: 1) generate functional annotation information at the species level, or 2) measure intermediate omics for individual animals. These approaches are investigated within GENE-SWitCH and GEroNIMO projects. Functional annotations may indicate regulatory elements that affect the level of gene expression for different genomic regions. This prior information on the importance of genomic regions can be used in genomic prediction models. Measuring intermediate omics of individuals gives a better handle on individual differences between animals and may help to explain a larger proportion of the phenotypic variance. It may not only help to better predict phenotypes, but potentially can also help to predict breeding values more accurately. By applying the first approach to predict gene expression at 10 genes of 300 pigs using SNPs and methylation status of regulatory regions, we showed that using methylation status in the tissue for which gene expression was predicted tended to yield higher prediction accuracy. By using the second approach to predict phenotypes of 478 mice, we showed that individual gene expression explained considerably more phenotypic variation than SNP genotypes alone. In addition, more phenotypic variance was explained if gene expression and phenotypes were measured closer to each other in time. The use of gene expression considerably increased the accuracy of predicting phenotypes and the accuracy of predicting breeding values for 9 out of 13 traits.

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

Bruno Perez

I graduated as a Veterinary in 2013 and obtained my MSc-degree in Animal Breeding at the University of São Paulo (USP) in 2015. During my MSc, I have spent some time at Embrapa (Brazilian Agricultural Research Corporation, Dairy Cattle department), investigating the prospect of selection for in-vitro embryo production performance in the Guzerá (Bos indicus) breed. In 2016 I started a PhD in animal breeding also at USP, with focus on investigating genotyping strategies for small dairy cattle breeds to allow a faster and less costly implementation of genomic selection. After that, I joined Hendrix Genetics B.V. as a geneticist. In this position, I am part of the global research and development team which focus on providing solutions to all our different breeding programs. Within GENE-SWitCH, I work on the development genomic prediction models that incorporate functional annotations and their application for pig breeding.

Hendrix Genetics, Netherlands
E-mail: marco.bink@hendrix-genetics.com

33. Using methylation to improve genomic prediction of gene expression in pigs

74th EAAP Annual Meeting - Lyon, France - 26 August -1 September 2023

B. Perez[1], J. de Vos[3], D. Crespo-Piazuello[2], M Bink[1], M. Ballester[2], O. Madsen, M. Calus[3]

1. Hendrix Genetics B.V., Research and Technology Center (RTC)
2. Animal Breeding and Genetics Program, Institute of Agrifood Research and Technology
3. Wageningen University & Research, Animal Breeding and Genomics

(De)Methylation of DNA has been extensively linked to variation in gene expression and thus could be used to improve prediction of gene transcripts. The aim of this study was to incorporate genome-wide DNA methylation in genomic prediction models to improve predictive accuracy for gene expressions. Data used in the analyses comprised of whole genome sequence and gene transcripts from 10 genes (in 3 tissues, totalizing 30 traits) from 300 pigs (100 Duroc, 100 Landrace and 100 Large White). We considered SNP variants placed within 2Kb upstream and 0.2Kb within genes as possible regulatory regions, and for each gene SNP placed within this range were subset, resulting in a panel of 65,486 variants. Methylation states at the loci level from the same tissues and two time points (30 days post-fert. and newborn) were used to weight SNP accordingly in the genomic relationship matrix (GRM). These GRM were used in GBLUP models to compare predictive accuracies for gene-expressions with either including (mGBLUP) or not including (stdGBLUP) methylation information as functional annotation. Model performance was measured by the correlation between solutions and gene-expressions pre-corrected for the effects of sex and breed using a 5-fold cross validation scheme. Results showed the performance of mGBLUP was highly dependent on the gene and tissue analyzed, together with the methylation state considered as SNP weights. In general, best results for mGBLUP were obtained when predicting gene-expressions on the same tissue as methylation information was used. Across the 10 genes, 3 tissues and 2 time-points considered, average difference in predictive accuracy between mGBLUP and stdGBLUP was +2.6% ± 3.9%. Although mGBLUP showed competitive predictive performance when compared to stdGBLUP when predicting gene-expression for some genes, this pattern was not consistent across all gene-tissue combinations. Further studies will investigate the used of the same framework for prediction of complex phenotypes instead of gene-expressions.

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

34. Using convolutional neural networks for image-based genomic prediction in mice

World Congress on Genetics Applied to Livestock Production - Rotterdam, The Netherlands - 3-8 July 2022

B. Perez[1], A. Savchuk[1], P. Duenk[2], M. Calus[2], M. Bink[1]

1. Hendrix Genetics B.V., Research and Technology Center (RTC)
2. Wageningen University & Research, Animal Breeding and Genomics

Convolutional neural networks (CNN) are well suited image recognition tools for their ability to recognize latent patterns from images. Here, we investigated whether CNN can be used for genomic prediction. We created genomic images from genotype data and used them to predict phenotypes in mice. This approach was compared with traditional GBLUP and gradient boosting machine (GBM) models. For the two traits analyzed, CNN was competitive in terms of predictive performance. The resolution of genomic images impacted model performance where, for this dataset, optimum results were obtained with 100×100 pixels. These first results demonstrate the potential of genomic images for genomic prediction using CNNs and merit investigation on adding layers of information to further increase accuracy of prediction.

Andreas Kranis

Roslin Institute, UK
E-mail: andreas.kranis@roslin.ed.ac.uk

35. Genome-wide association studies for body weight in broilers using sequencing and SNP chip data

74th EAAP Annual Meeting – Lyon, France - 26 August -1 September 2023

E. Tarsani[1], V. Riggio[1], A. Kranis[1-2]

1. The Roslin Institute and R(D)SVS, University of Edinburgh
2. Aviagen Ltd.

Body weight (BW) is an economically important trait for the chicken industry. Detecting which variants affect chicken growth is of great benefit to the genetic improvement of this important agricultural species. A powerful resource for identifying trait-associated variants in genome-wide association studies (GWAS) is using whole genome sequencing (WGS) data. In this study, we used a dataset comprising of 60,558 chickens genotyped with a 50k SNP chip. These SNP genotypes were imputed to sequence level for a total of 12.8 million variants with high levels of accuracy (>0.95). To evaluate the power of sequencing data, we conducted two GWAS for BW using the chip and WGS genotypes, respectively. When using the 50k SNP chip data, we obtained high genomic heritability (0.40) and identified 18 SNPs across eight autosomes (2-6, 8, 20 and 23) reaching genome-wide significance (FDR p-value < 0.10) for BW. The majority of those SNPs were introns. With the use of sequencing data, results were similar to those obtained with SNP chip data. However, additional GWAS signals were found, including variants in regions harboring functional elements for BW. The results suggest an improvement in power and precision when using WGS data. A significant increase in computational time to conduct such a large-scale WGS GWAS was noted, which can limit the number of animals to include in the study. Taken together, the present study enriched our knowledge of the genetic mechanisms regulating chicken growth traits. Next steps will focus on

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

improving the computational tools to enable larger scale analyses to further increase the statistical power.

Daniel Crespo Pizuelo

Daniel Crespo Piazuelo is a Doctor in Veterinary Medicine by the University of Zaragoza (2013). To continue his academic career, 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). Then, he received his PhD degree in Animal Production (UAB) in 2018. After his PhD, he worked as a postdoctoral researcher in the Pig Development Department at Teagasc in Ireland for two years. For his second post-doc, he joined the Institut de Recerca i Tecnologia Agroalimentàries (IRTA) in December 2020, working in the GENE-SWitCH project under the supervision of Maria Ballester (IRTA). Within the GENE-SWitCh project, he analyzed the relationship between the genome and the transcriptome of three different pig populations. Nowadays, he is working as a researcher in Cuarte SL (Grupo Jorge), one of the biggest pig production companies from Spain.

IRTA, Spain
E-mail: daniel.crespo@irta.cat

36. Identifying muscle transcriptional regulatory elements in the pig genome

The 38th International Society for Animal Genetics Conference - Virtual - 26-30 June 2021

D. Crespo-Piazuelo[1], O. González-Rodríguez[1], M. Mongellaz[2], H. Acloque[2], M.J. Mercat[3], M. Bink[4], A.E. Huisman[4], Y. Ramayo-Caldas[1], J.P. Sánchez[1], M. Ballester[1]

1. Animal Breeding and Genetics Program, Institute of Agrifood Research and Technology
2. Paris-Saclay University, INRAE, AgroParisTech, GABI
3. IFIP-Institut du porc and Alliance R&D
4. Hendrix Genetics B.V., Research and Technology

The present work is part of H2020 GENE-SWitCH project which aims to deliver new functional genome information of two main monogastric farm species (pig and chicken) to better understand the genetic determinants of complex traits. Specifically, in this work, we aimed to identify regulatory elements of muscle transcriptome in the genome of 300 pigs (100 Duroc, 100 Landrace and 100 Large White) using expression genome-wide association studies (eGWAS). For that purpose, muscle samples were obtained at slaughter and RNA was extracted using spin column-based kit and sequenced on the Illumina NovaSeq6000 platform. Counts were quantified by RSEM/1.3.0 and normalized by TMM (trimmed mean of M-values). Low expressed genes and those not present in at least 5% of the animals were removed, remaining 13,887 genes for further analysis. Through whole genome sequencing (NovaSeq6000 platform), 44,127,400 polymorphisms (SNPs and indels) were found among all the individuals. A total of 25,315,878 polymorphisms were kept after filtering out those with missing genotype data >0.1 and minor allele frequency <0.05. eGWAS were conducted between the filtered polymorphisms and the normalized expression data using the fastGWA tool from GCTA/1.93.2. After Bonferroni correction, a total of 8,099,604 significant associations were found between 4,814,732 polymorphisms and the expression of 7,496 genes. A total of 25,642 polymorphisms were associated with more than 10 genes and were considered hotspot regulatory

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

elements. Regarding their genomic position, 3,283,835 polymorphisms (68%) were annotated as cis-regulatory elements, as they were located at 1Mb or less than their associated gene. The most significantly associated variant (adj.p-value=2.66×10-178) was located on an intronic region of the SRSF Protein Kinase 3 (SRPK3) gene, which encodes a protein associated to muscle development. Our results identified key regulatory elements associated with gene expression in muscle that may be included in new predictive models to increase the accuracy of genomic predictions, speeding up the rate of genetic improvement of economically important traits in pigs.

37. An in-depth analysis of regulatory polymorphisms in three pig transcriptomes

XX Reunión Nacional de Mejora Genética Animal - Madrid, Spain - 1-3 June 2022

D. Crespo-Piazuelo[1], O. González-Rodríguez[1], M. Mongellaz[2], H. Acloque[2], M.J. Mercat[3], M. Bink[4], A.E. Huisman[4], Y. Ramayo-Caldas[1], J.P. Sánchez[1], M. Ballester[1]

1. Animal Breeding and Genetics Program, Institute of Agrifood Research and Technology
2. Paris-Saclay University, INRAE, AgroParisTech, GABI
3. IFIP-Institut du porc and Alliance R&D
4. Hendrix Genetics B.V., Research and Technology

Over the last decade, genome-wide association studies (GWAS) have reported that 90% of the phenotype-associated loci fell outside protein-coding regions. Thus, these genetic variants may be explaining phenotypic variations due to gene regulation mechanisms. The objective of the present study was to discover genetic variants that could be modifying gene expression levels across duodenum, liver, and muscle transcriptomes through the regulation of specific genes. For that purpose, duodenum, liver, and muscle samples were collected at slaughter from 300 pigs of three different breeds (Duroc, Landrace, and Large White; n=100/per breed). DNA and RNA were paired-end (2×150bp) sequenced in an Illumina NovaSeq6000 platform and sequences were mapped against the reference genome assembly (Sscrofa11.1) with BWA-MEM and STAR aligners, respectively. A total of 25,315,878 genetic variants were obtained with GATK after keeping those with a call rate higher than 10% and a minor allele frequency greater than 5%. After quantifying RNA counts with RSEM and normalizing them, 16,753 genes were kept in duodenum, 15,710 in liver, and 13,887 in muscle. Expression GWAS were carried out between the filtered polymorphisms and the expressed genes on each tissue, performing more than 1.17×1012 combinations with the fastGWA tool from GCTA. Bonferroni correction was applied per tissue, resulting in 19,926,590 significantly associated polymorphisms (adjusted p-value≤0.05) with 13,483 different genes. Out of these significant polymorphisms, 52.6% were cis-regulatory elements, i.e., located at less than 1Mb from the associated gene. Then, eQTL regions were defined by grouping the significant variants located at less than 2Mb between them. Thereafter, eQTL regions (46.2%) that consisted of only one single polymorphism were filtered out and 68,610 eQTL regions remained, including 7,288 cis-eQTL regions (10.6%). Considering exclusively the most significant polymorphisms of each eQTL region and those associated with more than 10 genes, 238 hotspots were found in duodenum, 296 in liver, and 274 in muscle. Out of these hotspots, there were 33 cis-regulatory elements in duodenum, 14 in liver, and 25 in muscle. Some of the genes associated in cis with these variants were transcription factors and cofactors: CCAR2, GMEB2, SLC2A4RG, TERF2, ZGPAT, and ZNF512B in duodenum; CHD7, CHD8, and LHX6 in liver; and APEX1, and NME2 in muscle. Pathway analyses were conducted between each cis-associated gene and the rest of the genes associated in trans with the same variant. Notably, the CHD8 cis-polymorphism was associated with 25 genes involved in gene expression regulation, and a high expression of CHD8 in liver cancer has been associated with a

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

worse prognosis in humans. In muscle, APEX1 plays a role in DNA repair and cellular response to oxidative stress, and it was associated with 49 genes involved in muscle structure development. In conclusion, our work has allowed the identification of genetic variants and candidate genes that will help to understand the molecular mechanisms behind gene regulation and their potential effect over end-trait phenotypes.

38. Deciphering genetic variants from whole genome affecting duodenum, liver and muscle transcriptomes in pigs

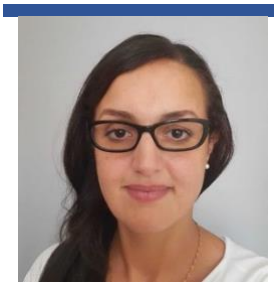
World Congress on Genetics Applied to Livestock Production - Rotterdam, The Netherlands - 3-8 July 2022

D. Crespo-Piazuelo[1], O. González-Rodríguez[1], M. Mongellaz[2], H. Acloque[2], M.J. Mercat[3], M. Bink[4], A.E. Huisman[4], Y. Ramayo-Caldas[1], J.P. Sánchez[1], M. Ballester[1]

1. Animal Breeding and Genetics Program, Institute of Agrifood Research and Technology
2. Paris-Saclay University, INRAE, AgroParisTech, GABI
3. IFIP-Institut du porc and Alliance R&D
4. Hendrix Genetics B.V., Research and Technology

With the aim of deciphering genetic variants affecting gene expression levels of duodenum, liver and muscle, the DNA and RNA of 300 pigs from three different breeds (Duroc, Landrace and Large White) was extracted and sequenced. After filtering and normalizing both the genomic and the transcriptomic datasets, expression genome-wide association studies (eGWAS) were conducted between 25,315,878 polymorphisms and the expression of 16,753 genes in duodenum, 15,710 genes in liver, and 13,887 genes in muscle. In total, more than 1.17×1012 combinations were performed, obtaining 19,926,590 significantly associated polymorphisms, which were grouped in 148,639 expression quantitative trait locus (eQTL) regions. Half of the polymorphisms were located in close proximity to their associated gene, but only 7,773 cis-eQTL regions were defined, the remaining being trans-eQTLs regions. Out of the cis-eQTL regions, 18 shared the same top polymorphism among the three tissues; thus, they were considered the strongest candidates for gene expression regulation.

Smahane Chalabi



I am postdoc researcher in bioinformatics and biostatistics. I am interested in data analysis, and more particularly functional genomics, to identify the mechanisms underlying complex disease or system. As part of the European GENE-SWitCH project, at the National Research Institute for Agriculture, Food and the Environment (INRAE), I studied the impact of the pig maternal diet on the transcriptome and epigenome of fetus and newborn, on different tissues.

INRAE, France
E-mail: smahane.chalabi@gmail.com

39. Epigenetic effects of maternal rich-fiber diets in pigs

INRAE/GABI H2020 Virtual Seminar - 21 January 2022

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

S. Chalabi[1], L. Loonen[2], W. Lakhal[3], J. Lluch[4], H. Acloque[1], J. Boekhorst[2], A. Sokolov[5], P. Harrison[5], H. Li[7], L. Fang[7], A. Rau[1], S. Djebali[6], E. Giuffra[1], J. Wells[2] and GENE-SWitCH Consortium

1. Paris-Saclay University, INRAE, AgroParisTech, GABI
2. Wageningen University & Research, Animal Breeding and Genomics
3. Diagenode S.A., Epigenetics R&D
4. INRAE, Plateforme Génomique, Génopole
5. European Molecular Biology Laboratory, European Bioinformatics Institute
6. IRSD, Université de Toulouse, INSERM, INRAE, ENVT, UPS
7. Center for Quantitative Genetics and Genomics, Aarhus University

The short chain fatty acids (SCFAs) produced by gut microbiota can affect a wide range of physiological functions related to metabolism and immunity, and have emerged as one clear link between the non-digestible fibre content of the diet and the epigenetic regulation of host cells and tissues, including muscle and liver. Pregnancy is a critical period of genome plasticity during which maternal diet may have long-lasting impact on the growth and health of offspring into adulthood. We hypothesized that maternal effects on the functional genome triggered by maternal diets enriched in fibers can be detected and assessed in utero and/or post-partum using a sufficiently large experimental pig population. In an exploratory “diet x epigenetics” study designed within the H2020 GENE-SWitCH project (https://gene-switch.eu), three isocaloric diets (high fibre-Beetpulp (HF-B), high fibre-Pea (HF-P) and a low fiber control (LF)) were each provided to a group of 7 sows throughout gestation. A collection of samples of 70-day-old fetuses and post-weaned piglets (2 males and 2 females per sow) was established and stored. Liver and skeletal muscle were profiled for chromatin accessibility (ATAC-seq) and gene expression (polyA+ RNA-seq). Data quality checks and primary analyses were performed using the nf-core ATAC-seq and Nextflow TAGADA pipelines for ATAC-seq and RNA-seq data, respectively. Sampling procedures, metadata and data have been shared on the FAANG Data portal page (https://data.faang.org/projects/GENE-SWitCH). As expected, principal components analyses revealed a strong separation among samples first by tissue and followed by developmental stage in both datasets. Differential analyses between pairs of diets were performed independently within each combination of tissue and developmental stage using an empirical Bayes linear (ATAC-seq) or generalized linear (RNA-seq) model. A modest number of differential ATAC-seq peaks was found, most in the piglet muscle HF-B vs. LF diet comparison. Although we found no significantly differentially expressed genes between pairs of diets in all tissue and developmental stage combinations, gene set enrichment analysis of expressed genes in the RNA-seq data revealed an enrichment of metabolism and immunity related functions in both piglet liver and muscle. Integrated analyses of RNA-seq and ATAC-seq revealed TF footprints with differential activity and significantly differentially expressed TF-targeted genes. These results thus provide an intriguing basis for future studies aimed at identifying links between epigenetic traits and phenotypes in pigs. GENE-SWitCH contributes to the global FAANG initiative and to EuroFAANG (https://eurofaang.eu), a synergy of European projects that share the common goal to discover links between genotype to phenotype in farmed animals and to meet global FAANG objectives.

40. Influence of maternal diet on the liver and skeletal muscle epigenomes of fetal and young pigs

Virtual FAANG Workshop - 16 February 2022

S. Chalabi[1], L. Loonen[2], W. Lakhal[3], J. Lluch[4], H. Acloque[1], J. Boekhorst[2], S. Djebali[5], A. Rau[1-6], E. Giuffra[1], J. Wells[2]

The GENE-SWitCH project has received funding from the European Union’s Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

1. Paris-Saclay University, INRAE, AgroParisTech, GABI
2. Wageningen University & Research, Animal Breeding and Genomics
3. Diagenode S.A., Epigenetics R&D
4 INRAE, Plateforme Génomique, Génopole
5. IRSD, Université de Toulouse, INSERM, INRAE, ENVT, UPS
6. BioEcoAgro Joint Research Unit, INRAE, Université de Liège, Université de Lille, Université de Picardie Jules Verne

The short chain fatty acids (SCFAs) produced by gut microbiota have emerged as one clear link between the non-digestible fiber content of the diet and the epigenetic regulation of host cells and tissues. Pregnancy is a critical period of genome plasticity during which maternal diet may have long-lasting impact on the growth and health of offspring into adulthood. We aimed (i) to assess if varying fiber content in maternal diet influences the epigenome and the transcriptome in two target tissues (liver and skeletal muscle) of pig fetuses; and (ii) to explore the possible persistence of epigenetic marks through the weaning stage. Three isocaloric diets (high-fiber, pea fiber + acetate, and no-fiber control) were provided to three groups of sows (7 sows each) throughout gestation. Metadata, tissues and cell samples of all fetuses and piglets (2 males and 2 females per sow) were collected by FAANG protocols. Liver and skeletal muscle of 84 fetuses and 76 piglets were profiled for chromatin accessibility (ATAC-seq) and polyA+ RNA-seq. The 632 ATAC-seq experiments (30 million 150bp paired-end reads) were processed with the nfcore ATAC-seq pipeline and results showed a very high mapping rate, a fraction of reads in peaks (FRIP) above 0.3 and an accumulation of mapped reads at the transcription start site (TSS) of known genes. Exploratory and differential analyses on the 245,452 peaks quantified in each sample are currently in progress.

Derek Bickhart

Derek Bickhart was born in the US and received his PhD in Genetics and Genomics from the University of Connecticut in 2010. He completed a Postdoctoral position in George Liu's lab at the United States Department of Agriculture's (USDA) Beltsville research campus, implementing the means to characterize structural variants using whole genome sequence data. He has held several positions as a Research Bioinformatician and Microbiologist at the USDA since 2012 where he has led or contributed to the development of several major agricultural reference genome assemblies, specifically for the Goat, Cow, Water Buffalo, and Pig. In August 2022 he started as a Bioinformatician at Hendrix Genetics, a world leading multispecies animal breeding company. In his present position, he informs the company on new technological developments, improves the state of genotyping platforms, and conducts analysis on datasets with high degrees of dimensionality.

Hendrix Genetics, Netherlands
E-mail: Derek.Bickhart@hendrix-genetics.com

41. Dynamic inclusion of functional genome annotations to improve accuracy of genomic prediction in pigs

International Plant & Animal Genome Conference - San Diego, USA - 12-17 January 2024

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

D. Bickhart[1], B. Perez[1], G. Vaccari[2], S. Radstaak[2], M. Calus[2], M. Bink[1]

1. Hendrix Genetics B.V., Research and Technology
2. Wageningen University & Research, Animal Breeding and Genomics

Genomic Best Linear Unbiased Prediction (GBLUP) model. However, this model assumes equal importance for all Single Nucleotide Polymorphisms (SNPs), irrespective of their genomic positions and functions. In this study, we aimed to augment prediction accuracies by introducing high-quality functional genome annotation maps. Our customized pipeline, based on Functional-And-Evolutionary Trait Heritability (FAETH) scores, dynamically assessed the relevance of variants in the whole-genome sequence (WGS) of Sus Scrofa. Thirty-two functional annotation maps were incorporated, encompassing various aspects such as expression associations, chromatin accessibility, differential methylation/expression, selection signatures, and variant annotations in the pig genome. Out of 15 million SNPs in the WGS dataset, 6.2 million had non-neutral FAETH scores. Six traits related to reproduction, production, and health were considered, of which heritability varied from 0.10 to 0.38. A forward validation approach, utilizing animals born after July 2020, was employed. The reference datasets ranged from 7,000 to 22,000 observations, with approximately 1,000 animals in the validation set per trait. The results suggest the relevance of specific annotation maps, with lncQTL, sQTL, and eQTL demonstrating high informativeness across the traits. Our study suggests that FAETH scores effectively rank SNP variants informative across traits, however, the impact on prediction accuracy varied across traits analyzed. Our approach is dynamic on the process of updating FAETH scores when new annotation maps emerge as well as accumulation of data on the reference population. Incorporating high FAETH ranking variants into commercial SNP chips holds potential for enhancing predictive accuracy in genomic selection for pig breeding. “GENE-SWitCH has received funding from the European Union's Horizon 2020 Research and Innovation Program under the Grant Agreement n° 817998”.

The GENE-SWitCH project has received funding from the European Union’s Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

GENE-SWitCH Final conference

Day 1- Session 1: Identification and characterization of functional genomic elements

- Welcome by Elisabetta Giuffra (coordinator) & Hervé Acloque (deputy coordinator) - INRAE
- Introduction by Alan Archibald (UEDIN)
- GENE-SWitCH: samples collection and assays-by-sequence provided to the scientific community - Hervé Acloque (INRAE)
- Profiling chromatin accessibility and gene expression during pig and chicken development - Sylvain Foissac (INRAE) and Sarah Djebali (INSERM)
- Comparative analysis of methylation and epigenomic states across developmental stages - Martjin Derks (WUR)
- Uncovering early transcriptomic landscape with ISOseq - Sebastien Guizard (UEDIN)

Day 2-Session 2: Implementing FAANG innovation for innovation breeding

- Part 1 Introduction by Elisabetta Giuffra (INRAE)
- Long-term effects of high-fiber maternal diets on the functional genome of pig offspring - Andrea Rau (INRAE)
- Part 2 Introduction by Marco Bink (Hendrix Genetics)
- Using Bayesian variable selection, Machine Learning and individual omics for genomic prediction - Mario Calus (WUR)
- Identification of transcriptional regulatory variants in pig duodenum, liver, and muscle tissues - Maria Ballester (IRTA)
- Dynamic inclusion of functional genome annotations to improve accuracy of genomic prediction in pigs - Bruno Perez (Hendrix Genetics)
- Fine-mapping growth-related haplotypes in chickens - Carl-Johan Rubin (UU)
- Sub setting GWS SNPs based on functional annotations in chickens - Andreas Kranis (UEDIN)

Day 2- Session 3- Celebrating 10 years of FAANG

- Introduction by FAANG coordinators - Elisabetta Giuffra (INRAE) - Emily Clark (UEDIN) - Huaijun Zhou (University of California), Chris Tuggle (Iowa State University)
- From FAANG to Fork: Highly annotated genomes as resources to improve farmed animal production - Emily Clark (UEDIN)

Summary talks:

- Progress of integrated and comparative analyses - Martijn Derks (WU)
- Ensembl gene and regulatory annotation for pig and chicken - Garth Ilsley (EMBL-EBI)
- Using functional and other biological data in genomic prediction - Mario Calus (WUR)
- Epigenetic effects of maternal high fiber diets in utero and the offspring - Jerry Wells (WUR)

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

Round table: The functional genome for future Genotype-to-Phenotype research and applications in swine and poultry

- Session 1: Genome annotations across developmental phases

Chairs: Hervé Acloque (INRAE) & Megan Davey (UEDIN)

Panelists: Huaijun Zhou (University of California), Ramiro Alberio (University of Nottingham), Gabriel Costa (University of Liege/BovReg), Dan McQueen (UEDIN/AQUAFAANG)

- Session 2: The epigenetic impact of diets

Chairs: Elisabetta Giuffra & Andrea Rau (INRAE) & Jerry Wells (WUR)

Panelists: Anne Gabory (INRAE), Carlos Guerrero Bosagna (Uppsala University), Chris Tuggle (Iowa State University, Pieter Knap (PIC)

- Session 3: Genomic selection of swine and poultry

Chairs: Marco Bink (Hendrix Genetics) & Mario Calus (WUR)

Panelists: Romi Peni Subira (University of Lleida), Eli Grindflek (Topigs Norsvin), Rachel Hawken (Cobb)

- Conclusive remarks by Elisabetta Giuffra and Hervé Acloque (INRAE)

Day 3- Policy & Ethics Workshop- jointly organized with the GEroNIMO project

- GENE-SWitCH: achievements and perspectives for stakeholders - Elisabetta Giuffra (INRAE)
- What's new in GEronIMO - Tatiana Zerjal (INRAE) and Mario Calus (WUR)
- Introduction to Policy & Ethics Workshop: on the need to include stakeholder in the debate on ethics- Franck Meijboom & Koen Kramer (UU)
- Interactive Session in subgroups led by Franck Meijboom & Koen Kramer (UU)
- Summary of each group discussion
- Closure of the interactive session - Franck Meijboom (UU)
- Conclusions - Elisabetta Giuffra, Hervé Acloque (INRAE) and Tatiana Zerjal (INRAE)

The GENE-SWitCH project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817998. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.