0407 Genomic selection for methane emission.

Y. de Haas*[1], J. E. Pryce[2], E. Wall[3], S. McParland[4], C. I. V. Manzanilla Pech[1], G. Difford[5], and J. Lassen[5], *[1]Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, Netherlands, [2]Agribio, Department of Economic Development, Jobs, Transport and Resources and La Trobe University, Melbourne, Australia, [3]SRUC, Edinburgh, UK, [4]Teagasc, Moorepark, Fermoy, Co. Cork, Ireland, [5]Center of Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, Foulum, Denmark.*

Climate change is a growing area of international concern, and it is well established that the release of greenhouse gases (GHG) is a contributing factor. Of the various GHG produced by ruminants, enteric methane (CH_4) is the most important contributor. One mitigation strategy is to reduce methane emission through genetic selection. Our first attempt used beef cattle and a GWAS to identify genes associated with several CH_4 traits in Angus beef cattle. The Angus population consisted of 1020 animals with phenotypes on methane production (MeP), dry matter intake (DMI), and weight (WT). Additionally, two new methane traits: residual genetic methane (RGM) and residual phenotypic methane (RPM) were calculated by adjusting CH_4 for DMI and WT. Animals were genotyped using the 800k Illumina Bovine HD Array. Estimated heritabilities were 0.30, 0.19 and 0.15 for MeP, RGM and RPM respectively, and estimated genetic correlations of MeP with DMI and WT were 0.83 and 0.80, respectively. Strong associations with MeP were found on chromosomes 4, 12, 14, 19, and 30. We have recently tried another approach in dairy cattle, where we aimed to enlarge the reference population for genomic selection by combining data on methane emissions in dairy cattle using data from 5 countries (Australia, Denmark, Ireland, the Netherlands and UK). The total dataset consists of 3060 dairy cows, of which most were genotyped, but with various kinds of SNP chips. We ended up with a uniform set of SNPs for each cow. Even though three different types of measurement equipment (laser, sniffer and SF_6) and protocols (measuring for 3 d, 1 wk, multiple weeks) were used, these data will be analyzed jointly to establish genetic and genomic parameters for enteric methane. The average methane production was 448 g/d in Australia (354 cows); 554 g/d in Denmark (1769 cows); 381 g/d in IRL (260 cows); 549 g/d in NL (457 cows); and 325 g/d in UK (216 cows). This clearly shows that the populations and diets are different in addition to the equipment and protocol. Therefore, a multi-trait approach will be used in the analysis. Following the experiences of a similar project (gDMI), it is expected that each country will benefit for contributing to an international reference set with increased accuracies of the estimates.

Key Words: esnteric methane, genomic selection, international collaboration

0408 How is genomics changing cattle breeding?

D. Boichard*[1], V. Ducrocq[1], P. Croiseau[1], and S. Fritz[1,2], *[1]GABI, INRA, AgroParisTech, Universite Paris Saclay, Jouy-en-Josas, France, [2]Allice, Paris, France.*

Genomic selection offers considerable flexibility to increase genetic trends in dairy cattle breeding, through a decrease in generation interval, an increase in selection intensity, and an increase in reliability for females and for low heritability traits. It is also an opportunity for more sustainable breeding, in terms of breeding goal and genetic variability. With a shorter generation interval, there is a big risk of increasing inbreeding if semen dissemination policy of elite bulls is not changed. However, using a large number of young bulls both as service bulls and bull sires is a simple solution for both maximizing genetic trend while reducing inbreeding trend. Female genotyping is a key challenge for within herd selection and, simultaneously, for replacing current reference populations based on progeny tested bulls, assembling new ones in breeds of more limited size, and for selection of newly recorded traits. At a reasonable price and coupled with use of sexed semen, female genotyping is profitable for the farmers and is becoming a routine practice in an increasing number of herds. New applications are generated, such as renovated mating plans, efficient management of genetic defects, prediction of cows' future career and optimization of culling policy. With more diverse bulls on the market and with female genotyping, genomic selection also opens new avenues for more customized breeding across herds or production systems. A big challenge is to reduce the dependency of genomic predictions on relationship between candidates and the reference population. A strong effort is presently dedicated to integrating genome sequence information into predictions, to improve their accuracy and persistency. To increase the accuracy, within and especially across breeds, causal variants or very close proxies should be identified and included in the predictions, while discarding or limiting the weight of many other variants generating noise. In the longer term, further customization of selection will be possible by accounting for GxE interactions. Important developments are also necessary to decrease loss of favorable alleles through genetic drift.

Key Words: dairy cattle, genomic selection

0409 Genomic prediction using imputed sequence data in dairy and dual purpose breeds. M. Erbe[*1,2], M. Frischknecht[3,4], H. Pausch[5], R. Emmerling[1], T. H. Meuwissen[6], B. Gredler[3], B. Bapst[3], I. Consortium[7], K. U. Götz[1], and H. Simianer[2], *[1]Bavarian State Research Centre for Agriculture, Institute for Animal Breeding, Grub, Germany, [2]Georg-August-University, Department of Animal Sciences, Animal Breeding and Genetics Group, Göttingen, Germany, [3]Qualitas AG, Zug, Switzerland, [4]Bern University of Applied Sciences, School of Agricultural, Forest and Food Sciences HAFL, Zollikofen, Switzerland, [5]Technische Universität Müchen, Chair of Animal Breeding, Freising, Germany, [6]Norwegian University of Life Sciences, Department of Animal and Aquacultural Sciences, Ås, Norway, [7]Interbull Centre, Uppsala, Sweden.*

Technical progress has made it possible to re-sequence individuals within a reasonable time frame and at acceptable costs. However, as sequencing all individuals of a breeding population is still too expensive, only key individuals of a population contributing most to the genetic variation usually are chosen to be sequenced. All other individuals genotyped with common single nucleotide polymorphism (SNP) arrays are then imputed up to all known SNPs and possibly biallelic short insertions or deletions (indels) at sequence level. Different simulation studies have shown that using sequence data for genomic prediction can have a positive effect on the accuracy and the stability of marker effect estimates especially when using variable selection methods. We thus tested these hypotheses with two different data sets: one with over 6000 Fleckvieh bulls genotyped with 50k or 777k and one with over 2000 Brown Swiss dairy cattle genotyped with 30k, 50k or 777k, both imputed to sequence level with a reference set of 150 and 123 sequenced individuals, respectively. With the Fleckvieh data set, no or only very slightly higher prediction accuracies were found with imputed sequence data than with SNP array data for six different traits studied. This was true for different genomic BLUP models as well as for GBCPP, a fast EM-based variable selection method similar to Bayes Cπ. Attempts to reduce noise by modeling only specific subsets of SNPs (e.g., very accurately imputed SNPs, SNPs from genic regions) generally improved prediction compared with modeling all imputed SNPs. Sequence-based predictions did not appear to be more stable as prediction ability decreased similarly for both 50k and sequence data when sires and/or grandsires of candidates were removed from the calibration set. For Brown Swiss, a slight increase in prediction accuracy was found for non-return rate after 56 d in heifers when modeling all imputed SNPs with GBCPP compared with modeling only SNPs from the 50k array. Using prior biological information by modeling only the 50k most significant SNPs obtained from a genome-wide association study did not improve prediction accuracy, but outperformed prediction based on the 50k array. Possible explanations for the limited success of genomic prediction with sequence data are inaccuracies in imputed genotypes, especially for variants with small minor allele frequencies, lack of proper models to account for the underlying genetic architecture, and incompleteness of genome maps and structural annotation.

Key Words: genomic prediction, sequence data

0410 Multi-breed genomic evaluations for 1 million beef cattle in Ireland. A. Cromie[*1], R. Evans[2], F. Kearney[2], D. Berry[3], M. C. McClure[1], and J. McCarthy[4], *[1]Irish Cattle Breeding Federation, Bandon, Ireland, [2]Irish Cattle Breeding Federation, Bandon, Co. Cork, Ireland, [3]Teagasc, Moorepark Research Centre, Fermoy, Cork, Ireland, [4]Irish Cattle Breeding Federation, Cork, Ireland.*

Key stakeholders in Ireland (Irish Cattle Breeding Federation, Teagasc and the Department of Agriculture and Marine) are currently developing multi-breed genomic evaluations for some 1 million beef cattle. The project is co-funded through the EU's Rural Development Program, with the overall objective of increasing rates of genetic gain for key traits related to profitability and environmental sustainability within the Irish suckler beef herd. A total of $338 million has been allocated to the project over the 6-yr period (2015–2020), of which some 15% will be allocated toward the cost of genotyping and related genomic evaluations.

Phase 1 of the project is underway, with 300k beef animals genotyped in 2015. This is in addition to a further 100k animal's which were genotyped in 2013 and 2014, as part of an initial Irish government and industry funded initiative to help establish the required infra-structure (phenotypes and genotypes) for large scale multi-breed genomic evaluations. All animals have been genotyped on the International Dairy and Beef chip (IDB), with the latest version (the IDBv3) being a customised 54k chip developed in conjunction with Teagasc and Illumina Inc. It is anticipated that a further 300k animals/year will be genotyped in 2016 and 2017, resulting in a total of 1 million beef animals required for routine genomic evaluations by end 2017. Phase 2 of the project, will result in a further 1 million animals being genotyped in 2018–2020, bringing the total requirement for routine genomic evaluations to in excess of 2 million animals.

Analysis to-date has been based on a subsample of some 100k sires and cows with reliable evaluations for key profit traits. Single-step genomic evaluations using Mix99 software have been applied to the dataset, with an almost doubling of reliability from the current 20% to almost 40% for individual traits and the relevant economic indexes. Initial feedback from industry has been very positive, with an expectation that the evaluations will become official from August 2016, after which we anticipate running routine evaluations, based on the

increasing genotype and phenotype data every 2–3 mo.

Key Words: genomics, beef cattle, multibreed

FUNCTIONAL ANNOTATION OF ANIMAL GENOMES (FAANG) ASAS-ISAG JOINT SYMPOSIUM

0411 Important lessons from complex genomes.

T. R. Gingeras*, *Cold Spring Harbor Laboratory, Functional Genomics, NY.*

The approximately three billion base pairs of the human DNA represent a storage devise encoding information for hundreds of thousands of processes that can go on within and outside of a human cell. This information is revealed in the RNAs that are composed of 12 billion nucleotides considering the strandedness and the allelic content each of the diploid copies of the genome. Results stemming from the efforts to catalog and analyze the RNA products made by cells in the human (ENCODE), fly-worm (modENCODE) and mouse ENCODE projects have shed light on both the functional content and how this information is organized by various genomes. In human cells, a total of ~161,000 transcripts present within ~50,000 genic regions represent our previously best manually-curated annotation (based on v 7 Gencode) of the transcriptome. The results from the ENCODE project point to considerable supplementation of these data. Analyses of these transcriptome data sets have resulted in important and under appreciated lessons such as: (1) pervasive genome-wide transcription prompts a need to redefine the definition of a gene, (2) expression ranges follow transcript types and subcellular localization, (3) expression of isoforms of a gene by a cell do not follow a minimalistic strategy, and (4) genomic characteristics of potential *trans*-acting enhancer regions are distinguishable from other types of *cis*-acting regulatory regions. These and other lessons drawn from the landscape of both coding and non-coding RNAs present in eukaryotic cells have been used to assist in understanding and organizing what is often seen as dauntingly complex genomes.

Key Words: annotation, ENCODE, transcriptome

0412 Causal inference of molecular networks integrating multi-omics data.

F. Peñagaricano*, *University of Florida, Gainesville.*

Recent developments of massively parallel technologies allow assaying different biological molecules at very high throughput rates, including sequencing and genotyping of DNA, quantifying whole-genome gene expression, including measuring mRNA and microRNA abundance, identifying genome-wide epigenetic modifications, such as DNA methylation, and measuring different proteins and cellular metabolites. These advancements provide unprecedented opportunities to uncover the genetic architecture underlying phenotypic variation. In this context, the main challenge is to decipher the flow of biological information that lies between the genotypes and the phenotypes under study; in other words, the new challenge is to integrate multiple sources of molecular information, i.e., multiple layers of omics data, to reveal the causal biological networks that underlie complex traits. It is important to note that knowledge regarding causal relationships among genes and phenotypes can be used to predict the behavior of complex systems, as well as to optimize management practices and selection strategies. Here, we describe a multistep procedure for inferring causal gene-phenotype networks underlying complex phenotypes integrating multi-omics data. We initially assess marginal associations between genotypes and either intermediate phenotypes (such as gene expression) and endpoint phenotypes (such as carcass fat deposition and muscularity), and then, in those genomic regions where multiple significant hits co-localize, we attempt to reconstruct molecular networks using causal structural learning algorithms. These algorithms attempt to infer networks assuming that the pattern of conditional independencies observed in the joint probability distribution of these set of correlated variables are compatible with the unknown causal model. As a proof of principle of the significance of this integrative approach, we show the construction of causal molecular networks underlying economically relevant meat quality traits in pigs using multi-omics data obtained from an F2 Duroc x Pietrain resource population. Interestingly, our findings shed light on the mechanisms underlying some known antagonist relationships between important phenotypes, for instance, carcass fat deposition and meat lean content. More generally, the proposed methodology allows further learning regarding phenotypic and molecular causal structures underlying complex traits in farm species.

Key Words: causal inference, graphical models, systems biology

0413 Genotypes to phenotypes: Lessons from functional variation in the human genome and transcriptome.

B. E. Stranger*, *Section of Genetic Medicine, Department of Medicine, Institute of Genomics and Systems Biology, Center for Data Intensive Sciences, University of Chicago, IL.*

Complex trait association mapping in humans has successfully identified genetic loci influencing trait variation for hundreds of different phenotypes, including disease. The vast majority of associated loci localize to non-coding regions of the genome, suggesting possible effects on gene regulatory mechanisms. Without a clear understanding of the regulatory code of the human genome, deep characterization of the molecular function(s) of genetic variants in the human genome has become increasingly important for defining that code and for understanding genetic associations to complex traits. Studies of the human