Sequence -based GWAS for milk production traits, udder health and morphology in French dairy cattle

Thierry Tribout, Pascal Croiseau, Rachel Lefebvre, Anne Barbat, Mekki Boussaha, Sébastien Fritz, Didier Boichard, Chris Hozé, Marie-Pierre Sanchez

Sequence-based GWAS for milk production, udder health and morphology traits in French dairy cattle

T. Tribout[1], P. Croiseau[1], R. Lefebvre[1], A. Barbat[1], M. Boussaha[1], S. Fritz[2], D. Boichard[1], C. Hoze[2] and M.P. Sanchez[1]
[1]Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350, Jouy-en-Josas, France, [2]Allice, MNE, 75012, Paris, France; thierry.tribout@inrae.fr

Genome-wide association studies (GWAS) were performed at the sequence level in Montbéliarde (MO), Normande (NO) and Holstein (HO) dairy cattle breeds for 5 milk production traits, somatic cell scores, clinical mastitis, milking speed and 8 udder morphology traits. The number of bulls considered by trait varied from 1,857 to 2,515 in MO, from 1,427 to 2,203 in NO, and from 4,959 to 6,321 in HO. The variables were the bulls' daughter yield deviations (DYD), derived from the national genetic evaluations. Genotypes of the bulls for around 28 million sequence variants were imputed in 2 steps, using FImpute software: first from 50K level to HD level using 522 MO, 546 NO, and 776 HO HD genotyped bulls as a reference, and then to the sequence level using 1,147 sequenced bulls from the 1,000 bull genomes project (Run 4), including 28 MO, 24 NO and 288 HO bulls. GWAS were done independently within each breed and for each trait, using GCTA software, accounting for the population structure through a HD-based genomic relationship matrix. We detected 27, 14 and 49 significant QTL regions (-logP>8.2) in MO, NO and HO breeds, respectively. Numbers of QTL per trait and breed varied from 0 (protein yield in MO and NO, udder health in NO, 3 and 4 udder morphology traits in MO and NO, respectively) to 11 (protein content in HO). Some QTL were common for several traits within a breed, and/or for same traits in 2 or 3 breeds. Candidate variants (CV) were selected in the QTL, based on the significance of their effect and on their functional annotation. A total of 32,373 MO, 12,316 NO and 52,630 HO cows were genotyped or imputed for a 50K BeadChip augmented with the CV. A second set of GWAS was performed on these cows, considering their YD (ie adjusted own performances) for the same milk production and udder traits, these data being independent from the DYD of the first step bulls. The analyses confirmed 42 and 18 significant (-logP>6) QTL regions for production and udder traits, respectively. Except for 4 production and 4 udder QTL regions, the significance of estimated effect was higher for the CV than for the 50K variants.

Fine resolution CNV catalogue from deeply sequenced cattle genomes

Y.L. Lee[1], H. Takeda[2], G. Costa Monteiro Moreira[2], L. Karim[3], M. Bosse[1], A. Bouwman[1], E. Mullaart[4], W. Coppieters[3], M. Georges[2], T. Druet[2], C. Charlier[2] and GplusE Consortium[5]
[1]Wageningen University & Reserch, Animal Breeding and Genomics, P.O. Box 338, 6700 AH, Wageningen, the Netherlands, [2]University of Liège, Unit of Animal Genomics, GIGA-R & Faculty of Veterinary Medicine, Liège, Belgium, [3]GIGA Institute, GIGA Genomics Platform, Liège, Liège, Belgium, [4]CRV B.V., P.O. Box 454, 6800 AL, Arnhem, the Netherlands, [5]http://www.gpluse.eu/; younglim.lee@wur.nl

Genomes consist of various forms of variations that ultimately contribute to shaping phenotypes. Copy number variations (CNVs) are a form of genetic variation, and arise from gain or loss of DNA. CNV discovery depends on the quality of both sequencing data and reference genome, but in livestock an accurate CNV catalogue and an investigation of the functional impact of CNVs are lacking. We used deeply sequenced cattle genomes from 131 Dutch Friesian Holstein trios (mean coverage: 26x), mapped to the reference genome ARS1.2, to study CNVs. Harnessing the unique pedigree structure in livestock populations, we eliminated spurious CNVs, based on the Mendelian inheritance pattern. Offspring of the 131 probands (~5 animals/proband) enabled us to trace the inheritance of interesting CNVs observed in the probands. Among ~10,000 high quality CNVs, which were ascertained at base pair resolution, ~3,000 overlapped with the coding sequence. Subsequently, we used histone and chromatin modification assay data to investigate whether CNVs are overlapping with gene regulatory elements. Among the ~7,000 CNVs in non-coding regions, ~200 overlapped with putative regulatory elements such as enhancers, promoters, and open chromatin regions. These overlaps imply that CNVs can alter gene expression, either by directly affecting coding sequences or by interrupting gene regulatory elements. A highly interesting CNV in our catalogue is GC gene duplication, overlapping with the last exon of the GC gene. This duplication is located at ~36 kb distance from several known clinical mastitis QTLs. Using RNA seq data generated from liver tissues of Holstein cows (n=178), we confirmed that the duplication is associated with increased expression of GC. This valuable CNV catalogue warrants follow-up research on the functional impact of CNVs.