

ISGC call May 17th 2019

Attendees: Shannon Clarke, Rudiger Brauning, Kathryn McRae, John McEwan, Brenda Murdoch, Michelle Mousel, Kim Worley, Noelle Cockett, Carole Moreno-Romieux, Gwenola Tosser, Emily Clark

Agenda:

1. Sheep Genome DB run 3 update (Rudi Brauning)
2. Update from the Ovine FAANG project (Brenda Murdoch)
3. Mapping of ISGC SNP chip probes to Rambouillet (15K, 50K and HD chips) (Rudi Brauning)
4. Ovine SNP chips- what's available to the community
5. "best" genome assembly for functional assay mapping
6. Other business (successes, new projects, new tools?)

Meeting notes:

1. Sheep Genome DB update (Rudi Brauning):
 - Run 2
 - 935 animals
 - utilised GATK and Samtools with OAR v3.1 genome assembly
 - completed with results at EVA (<https://www.ebi.ac.uk/eva/?eva-study=PRJEB31241>).
 - Run 3
 - Utilising one pipeline (BWA alignment to Rambouillet v1; GATK haplotype caller for SNPs and indels)
 - WGS data to be deposited at SRA and project number sent to Rudiger Brauning at AgR for inclusion in Run 3.
 - ~300 new animals to be included
 - Aim is to have completed by August
 - Updates and the pipeline utilised will be shared via email and sheepgenomesdb.org
2. Ovine FAANG Project update (Brenda Murdoch)
 - Assay update
 - 56 tissues in total that will have functional assays performed
 - CAGE (Roslin) is almost complete on 56 tissues.
 - mRNA-seq (Baylor) completed on 30 tissues with the remaining 26 to be sequenced. A subset (8 tissues) also with ISO-Seq
 - CHIP-seq (UIDAHO): has completed 42 tissues for 4 antibodies. Samples from 38 tissues have been sent for sequencing. An additional 5 tissues will be completed bringing total to 47 tissues.
 - ATAC-seq (WSU): assay is still under development.
 - WGBS (AgResearch): WGBS underway for the subset of tissues that have mRNA-seq, CAGE, Iso-seq, and CHIP-seq (8 tissues). RRBS carried out on the remaining 48 tissues.
 - Rambouillet v1 genome used for all data
 - Current grant (4 year) - ending 2020.
3. Mapping of ISGC SNP chip probes to Rambouillet (15K, 50K and HD chips) (Rudi Brauning)

- Probe sequences were taken from the Illumina manifests and mapped onto the Rambouillet genome (GCA_002742125.1) using bwa mem v0.7.17-r1188 with default settings (Indels were ignored)
- For each SNP a probe pair was constructed by using AlleleA_ProbeSeq and appending either the reference or the alternative allele.
- Only probe pairs were accepted that passed following filters.
 - both probes are mapped
 - one probe is mapped with 0 mismatches
 - both probes are not multi-mapped
 - no indels were allowed
 - both probes had to map in the same orientation
 - both probes had to map to the same position
 - mapped probes had between them exactly 1 mismatch, the SNP.
- Mapping results for Rambouillet: 15K 87%; 50K 83% and HD 93%
- Rudi has also mapped to Oar_v3.1 and Oar_v4 to enable comparison of his mapping approach to NCBI and Ensembl.
- Approval given from participants on the call for SNP name, position and allele from the consortium arrays to be provided publicly via figshare. Aim is to have an RS number for all SNPs on all chips.

4. Ovine SNP chips- what's available to the community

- The Illumina ISGC 15K and 50K are listed products at Illumina
- The ISGC HD chip is also available but as a controlled chip and permission is required
- Several other low and mid density custom chips are available via different providers.
- Chinese group have an HD chip too-little overlap with the ISGC HD chip
- Possibility of the AgResearch Ovine LD chip released as a controlled ISGC chip moving forward. Requires commitment from consortium to place a combined order at next update of the chip.

5. "best" genome assembly for functional assay mapping

- Rambouillet v1 being used for the functional assays. NCBI has annotated it (and Oarv3.1, enabling comparison). Ensembl genebuild scheduled for later this year.
- Emily Clark informed of new funding to enable more than 1 reference genome at EMBL-EBI. Currently for pig but could also facilitate having more than one genome for sheep.

6. Other business

- Next USDA call-July 26th
- Discussion on potential ideas
 - Pan-genome
 - Structural variation (CNVs, segmental duplications...) onto annotation
 - Microbiome survey - international spread (countries, breeds, developmental stages, environments)
- Doodle poll to be sent for next call
- Catch-up at ISAG.