MOLECULAR INVESTIGATION OF SEVERAL EMERGING INHERITED DISEASES IN CATTLE AND SHEEP

S.A. Woolley[1], E.R. Tsimnadis[1], R.L. Tulloch[1], P. Hughes[1], B. Hopkins[1], S.E. Hayes[1], M.R. Shariflou[1], A. Bauer[2], I.M. Häfliger[2], V. Jagannathan[2], C. Drögemüller[2], T. Leeb[2], M.S. Khatkar[1], C.E. Willet[3], B.A. O'Rourke[4], I. Tammen[1]

[1]The University of Sydney, Faculty of Science, Sydney School of Veterinary Science, Camden, NSW, 2570 Australia
[2]The University of Bern, Institute of Genetics, Vetsuisse Faculty, Bern, Switzerland
[3]The University of Sydney, Sydney Informatics Hub, Core Research Facilities, Sydney, NSW, 2006 Australia
[4]The Elizabeth Macarthur Agricultural Institute, NSW Department of Primary Industries, Menangle, NSW, 2568 Australia

SUMMARY

Emerging inherited diseases can cause numerous issues for producers, including productivity loss, profit loss and animal welfare problems. Current collaborative efforts between the University of Sydney and the Elizabeth Macarthur Agricultural Institute, NSW Department of Primary Industries has resulted in the ongoing investigation of several inherited diseases using both SNP-based homozygosity mapping and whole genome sequencing approaches to identify positional candidate genes and likely causal variants. This paper serves as a brief update for eight of the investigated inherited diseases in cattle and sheep, with these studies aiming to identify positional candidate genes and causal variants to facilitate the improved management of at-risk populations for each inherited disease investigated.

INTRODUCTION

The advancement of livestock breeding has allowed for desirable traits and elite genetics to be disseminated throughout livestock populations within relatively short periods of time. Small effective population sizes and inbreeding poses a risk for the inheritance of deleterious alleles in homozygous form and can contribute to the increased observation of animals with recessive inherited diseases (Charlier *et al*. 2008; Groeneveld *et al.* 2010), especially when considering closed herds or flocks. The reporting of inherited diseases within Australian livestock is limited due to either misdiagnosis of a prospective inherited disease or concern for reputation damage and profit losses. Detailed clinical and phenotypic descriptions of suspected recessive inherited diseases is imperative to future molecular investigations. Without consistent reporting and detailed phenotype information, the molecular characterisation of emerging inherited diseases can be delayed due to resource loss or lack of key information such as pedigree data and clinical descriptions. This can therefore impact on the monitoring and management of the inherited disease in at-risk populations, especially if detailed pedigrees are unknown when genotyping tests become available (Man *et al.* 2007).

Collaborative projects between researchers at the University of Sydney and the Elizabeth Macarthur Agricultural Institute, NSW Department of Primary Industries (EMAI) has enabled the investigation of several emerging recessive inherited diseases in livestock. With an increasing number of suspected inherited disease cases being investigated, the use of SNP-chip based homozygosity mapping and whole genome sequencing approaches is becoming routine in identifying positional candidate genes, causal variants and for facilitating the development of genotyping tests for inherited diseases with little pedigree information or phenotypic descriptions. This paper serves as an update for eight of the emerging inherited diseases with a suspected recessive mode of inheritance currently under

investigation by the University of Sydney and EMAI. These emerging inherited diseases include: cardiomyopathy and woolly haircoat syndrome (CWH) in Hereford cattle, congenital mandibular prognathia (CMP) in Droughtmaster cattle, Niemann-Pick type C disease (NPC) in Angus cattle, new variants of ichthyosis fetalis (IF) in Hereford and Shorthorn cattle, the previously reported brachygnathia, cardiomegaly and renal hypoplasia syndrome (BCRHS) in Merino sheep (Shariflou *et al.* 2013), cervicothoracic vertebral subluxation (CVS) in Merino sheep, ovine dermatosparaxis (OD) in Merino sheep, and pulmonary hypoplasia with anasarca (PHA) in Persian sheep. The aim for these studies was to identify positional candidate genes and likely causal variants to facilitate improved management of at-risk populations for each inherited disease investigated.

MATERIALS AND METHODS

Analysis of SNP genotype data for carrier and affected animals (Table 1) using sliding windows of 25, 50 and 100 SNPs to identify runs of homozygosity (ROH) was previously conducted (Table 1) for affected animals using the bovine UMD3.1 genome assembly and the ovine Oarv1.0 genome assembly (Woolley *et al.* 2017). ROH were analysed using PLINK (Purcell *et al.* 2007) and were considered to be regions of interest if these regions were shared by all of the affected animals only. These regions were scanned for positional candidate genes based on gene function and comparative genomics methods.

Table 1. Number of affected and carrier DNA samples submitted for SNP chip genotyping and regions of homozygosity, including species specific OMIA ID

Disease	*OMIA ID*[1]	*Breed*	*Affected/ Carrier*	*SNP chip*
Cardiomyopathy and woolly haircoat syndrome	000161-9913	Poll Hereford	2/0	SNP80[2]
Congenital mandibular prognathia	-	Droughtmaster	9/4	SNP80[2]
Ichthyosis fetalis	000547-9913	Hereford	1/3	SNP80[2]
Niemann-Pick disease	-	Angus/Angus X	2/2	SNP80[2]
Cervicothoracic vertebral subluxation	000077-9940	Merino	14/2	SNP50[3]
Pulmonary hypoplasia with anasarca	000493-9940	Persian	5/5	SNP50[3]

[1]OMIA http://omia.angis.org.au, - indicates no species specific OMIA ID. [2]SNP80 = GeneSeek® Genomic Profiler Bovine HD Chip 80K chip (Neogen, NE, USA). [3]SNP50 = Illumina® OvineSNP50 Genotyping BeadChip (CA, USA).

Sanger sequencing for inherited diseases with identified positional candidate genes commenced but was cost and labour intensive. Whole genome sequencing (WGS) was conducted for affected animals for CMP, BCRHS, CVS and PHA (Woolley *et al.* 2017) with 150bp paired-end reads at an expected coverage of 20X or 30X (Table 2). Sequence reads were aligned with BWA-mem (Li 2013) to either the bosTau8 or oviAri3 reference genome assemblies and analysed for novel genetic variants using a modified GATK best practice pipeline (McKenna *et al.* 2010; DePristo *et al.* 2010). Large structural variant calling was completed using DELLY (version 0.7.6), LUMPY-sv (version 0.2.12) and LUMPY SVtyper (Rausch *et al.* 2012; Layer *et al.* 2014). WGS data generated at the University of Bern similarly applied standard bioinformatics pipelines using software and steps to process fastq files into bam and GVCF files in accordance to the latest 1000 Bulls processing guidelines (www.1000bullgenomes.com). For variant filtering, control genomes from other samples that were sequenced during this study were

used according to species and breed, and for the Shorthorn IF and OD samples, 341 control genomes of various cattle breeds and 16 control genomes of various sheep breeds were used to identify novel variants for affected animals only. Genetic variants were annotated using SnpEff for predicted effects and filtered using SnpSift (Cingolani *et al.* 2012). To predict the functional effects of candidate causal variants, both SnpEff and SIFT (Kumar *et al.* 2009; Cingolani *et al.* 2012) were used to assess whether candidate disease-causing variants were deleterious to protein function.

RESULTS AND DISCUSSION

As previously identified, homozygosity mapping was able to successfully reveal and/or exclude positional candidate genes for all of the inherited diseases investigated, with a likely causal variant in a positional candidate gene identified for NPC through Sanger sequencing of affected animals (Shariflou *et al.* 2013; Woolley *et al.* 2017). Affected samples for BCRHS, CMP, CWH, IF, CVS, OD and PHA were re-sequenced using WGS (Table 2) as either homozygosity mapping did not identify positional candidate genes of interest or Sanger sequencing of affected animals did not identify causal variants within candidate positional candidate genes. Preliminary quality control analysis of the WGS data was positive (Woolley *et al.* 2017), however WGS for CWH in Poll Hereford cattle and IF in Hereford cattle was unsuccessful due to inadequate DNA quality. Further investigation of other positional candidate genes and genomic regions of interest based on SNP genotyping data will be required for CWH and IF.

After application of filtering parameters on samples that were whole genome sequenced, numerous genetic variants that were homozygous for the alternate allele in the affected animal(s) only were identified either across the genome or within previously identified ROH (Table 2) (Woolley *et al.* 2017).

Table 2. Variants identified in affected animals for which each animal was homozygous alternate to the reference sequence

Disease	*Breed*	*Affected/ Carrier*	*No. homozygous alternate variants*	*Likely causal variant identified*
Brachygnathia, cardiomegaly and renal hypoplasia syndrome	Merino	1	215[1]	Yes
Cervicothoracic vertebral subluxation	Merino	2	Ongoing	Ongoing
Ovine dermatosparaxis	Merino	1	1864[2]	Yes
Pulmonary hypoplasia with anasarca	Persian	2/1	333[1,3]	Under validation
Congenital mandibular prognathia	Droughtmaster	2	5780[4]	Under validation
Ichthyosis fetalis	Shorthorn	1	298[2]	Yes

[1]Filtered for low, moderate and high impact with known dbsnps included.[2]Private homozygous alternate and heterozygous protein-changing variants with a moderate or high predicted impact.[3]At least one animal was homozygous alternate.[4]Includes SNPs and small indels.

Further manual filtering based on the predicted variant impact on protein function revealed candidate causal variants for BCRHS, PHA, CMP and IF in Shorthorn cattle (Table 2). A candidate causal variant with possible heterogeneity was identified for OD in Merino sheep and requires greater sample numbers to facilitate further validation. Genotyping assays were developed for these five inherited diseases, with preliminary validation results showing variant segregation with disease in related herds or flocks. The development of these genotyping assays has allowed for producers to facilitate forward planning breeding management strategies.

Despite small sample sizes, poor phenotypic descriptions and challenging sample types, candidate causal variants have been successfully identified through the combined use of genome wide SNP genotyping, homozygosity mapping and WGS. These approaches have successfully identified candidate genes and causal mutations in a range of recessive inherited diseases in cattle, including ichthyosis fetalis in Chianina cattle (Charlier *et al.* 2008). The reporting of the inherited diseases investigated in these studies has enabled for better screening and preliminary management and has showcased the ability to identify candidate causal variants using modern genomic technologies.

CONCLUSIONS

Despite the challenges surrounding insufficient sample numbers and poorly defined phenotypes, the results from these studies indicate that candidate causal variants can be identified by utilising targeted approaches. The identification of likely causal variants for BCRHS, OD with possible genetic heterogeneity, PHA, CMP, NPC and IF in Shorthorn cattle, has enabled for the development of genotyping assays that are able to successfully discriminate between homozygous wildtype, heterozygous and homozygous alternate genotypes. These assays are being used as a preliminary screen for related or founder herds or flocks and would prove to be a useful tool for screening wider populations to gain a more holistic understanding of population allele frequencies and future breed management strategies.

ACKNOWLEDGMENTS

The authors acknowledge and thank the producers and veterinarians for the submission of samples to EMAI and the University of Sydney. The authors thank the genetics laboratory staff at EMAI for their assistance. The authors acknowledge the Sydney Informatics Hub, a Core Research Facility at the University of Sydney for facilitating access to the High Performance Computer Artemis, where computational analysis in this study was performed. The University of Sydney Faculty of Veterinary Science provided research support for 3 honours projects (NN 2013, ET 2014 & SW 2015), and funding support from the Animal Welfare, FS Quiney and Dorothy Minchin Bequests. The Australian Wool Education Trust is acknowledged for the support of a research scholarship for RT (2016) and PH (2018) and the Australian Government for the Australian Government Research Training Program (RTP) Scholarship for SW. The WGS is funded by a Faculty of Veterinary Science - DPI collaborative compact fund and research grant CRSII3_160738 from the Swiss National Science Foundation to TL.

REFERENCES

Auwera G.A., Carneiro M.O., Hartl C., *et al.* (2013) *Curr Protoc Bioinformatics* **43**: 11.0.1
Charlier C., Coppieters W., Rollin F., *et al.* (2008) *Nat. Genet.* **40**: 449.
Cingolani P., Platts A., Wang le L., *et al.* (2012) *Fly (Austin)* **6**: 80.
DePristo M.A., Banks E., Poplin R., *et al.* (2011) *Nat. Genet.* **43**.
Groeneveld L.F., Lenstra J.A., Eding H., *et al.* (2010) *Anim. Genet.* **41**: 6.
Kumar P., Henikoff S. and Ng P.C. (2009) *Nat Protoc* **4**: 1073.
Layer R.M., Chiang C., Quinlan A.R. and Hall I.M. (2014) *Genome Biol.* **15**: R84.
Li H. (2013) arXiv:1303.3997v1 [q-bio.GN]
Man W.Y.N., Nicholas F.W. and James J.W. (2007) *Prev. Vet. Med.* **78**: 262.
McKenna A., Hanna M., Banks E., *et al.* (2010) *Genome Res.* **20**: 1297.
Purcell S., Neale B., Todd-Brown K., *et al*. (2007) *Am. J. Hum. Genet.* **81**.
Rausch T., Zichner T., Schlattl A., *et al.* (2012) *Bioinformatics* **28**: i333.
Shariflou M.R., Wade C.M., Kijas J., *et al.* (2013) *Animal Genet.* **44**: 231.
Woolley S.A., Tsimnadis E.R., Nowak N. *et al.* (2017) *Proc. Assoc. Advmt. Anim. Breed. Genet.* 15.