# A NEW STRATEGY TO IDENTIFY THE DISEASE CAUSING MUTATION FOR NEURONAL CEROID LIPOFUSCINOSIS IN SOUTH HAMSPHIRE SHEEP

**I.F. Mohd Ismail[1], J.A.L Cavanagh[1], N.L. Mitchell[2], P.J. Houweling[1], D.N. Palmer[2] and I. Tammen[1]**

[1]Reprogen, The University of Sydney, Camden, NSW, Australia
[2]Agriculture and Life Sciences Division, Lincoln University, New Zealand

## SUMMARY

The New Zealand South Hampshire sheep have been well characterised as an animal model for variant late-infantile neuronal ceroid lipofuscinosis (vLINCL) in humans. The disease causing gene has been identified as CLN6, but so far no mutation has been identified. A sheep BAC containing CLN6 and flanking region (~120kb) was sequenced at 13.49-fold sequence coverage using the 454 sequencing™ method from Roche® to complement the existing but incomplete public domain ovine sequence for the region of interest. The 454 sequence was assembled to bovine genomic sequence on chromosome 10 (BTA10). For mutation screening 15 long-distance PCR products from affected and normal South Hampshire sheep covering CLN6 as well as substantial 3' and 5' flanking sequence are currently optimized to be sequenced. Regulatory elements and/or mutations identified in CLN6 non-coding regions are likely to indicate positions for disease causing mutations not only in the South Hampshire sheep but also in human uncharacterised vLINCL patients.

## INTRODUCTION

The neuronal ceroid lipofuscinoses (NCL) are a group of inherited neurodegenerative diseases that occur in humans and several animal species. The NCLs are characterised by accumulation of fluorescent storage bodies in neurons and other cells with at least 8 different causatives genes identified so far. In South Hamsphire sheep the mode of inheritance is autosomal recessive and onset of clinical signs is about 8 months of age. Linkage mapping established South Hampshire (SH) sheep as a valuable animal model for human variant late-infantile NCL (vLINCL) (Broom *et al.* 1998), which is caused by mutations in the CLN6 gene (Wheeler *et al.* 2002; Sharp *et al.* 2003). Reduced expression of CLN6 in affected SH sheep tissues and absence of a disease causing mutation in the coding sequence suggests a mutation in a regulatory element (Tammen *et al.* 2006). These elements are found mostly in non-coding regions within the gene or regions upstream and downstream.

When compared to mouse models, large animal models have been found to be particularly valuable for these neurodegenerative diseases, as they resemble humans more closely in relation to brain size and structure, clinical signs and progression, and life expectancy.

## MATERIALS AND METHODS

**South Hampshire sheep**. South Hampshire (SH) sheep DNA samples were provided from the research flock at Lincoln University, Christchurch, New Zealand (Jolly *et al.* 1980, Tammen *et al.* 2006). The sheep were diagnosed as either normal, carrier or affected with NCL using clinical signs (blindness, seizures and behavioural changes), histology (fluorescent storage material in neurons and other cells) and pedigree information, as well as an indirect DNA test based on an A/G polymorphic allele in exon 7 of CLN6 (Tammen *et al.* 2006).

**Identification of conserved non-coding sequences.** The programs Vista® (Frazer *et al.* 2004) and GeneDoc® (Nicholas *et al.*1997) were used to align genomic and coding sequence of CLN6 in 9 species (human, cattle, dog, macaque, opossum, mouse, rat, chicken and Fugu fish) with the aim to identify conserved non-coding sequences (CNCS) with potential regulatory functions.

**Bacterial Artificial Chromosome (BAC) sequencing.** The sheep BAC clone 270H8 was provided by Dr. Daniel Vaiman (INRA, Jouy-en-Josas, France) after PCR screening of the sheep BAC library (Vaiman *et al.* 1999) with CLN6 specific primers. Further characterisation of the BAC revealed presence of the CALML 4 gene downstream of the CLN6 gene as well as a CNCS approximately 7kb upstream of CLN6 using primers designed from publically available sheep sequences from the International Sheep Genomics Consortium (ISGC) or CNCS sequences identified as part of this project.

The BAC DNA was purified using the QIAGEN Large-Construct® kit according to manufacturer's protocol. This method involves an ATP-dependent exonuclease digestion step for the selective removal of bacterial genomic DNA. A total of 3μg of purified BAC DNA at a concentration of 80ng/μl and an $OD_{260}/OD_{280}$ of 1.65 was submitted for 454 sequencing™ (Goldberg *et al.* 2006) to University of Otago, High Throughput DNA Sequencing Unit, New Zealand. **A** sequencing library was constructed and sequenced in a $1/16^{th}$ standard FLX plate equivalent.

Sequencing raw data was returned and contig assembly and individual reads were obtained in FASTA format using the whole genome shotgun and EST sequence assembler program MIRA (Chevreux *et al.* 2004). BLAST (Altschul *et al.* 1990) batch analysis was performed at the National Center for Biotechnology Information (NCBI) to identify *E.coli* and pBeloBACII contamination. The remaining contigs were aligned to ISGC sheep sequence version 1.5 using GeneDoc® (Nicholas *et al.*1997) as well as to bovine chromosome (BTA) 10 using the Ensembl Genome Browser (http://www.ensembl.org/).

**Targeted sequencing for mutation screening.** A total of 15 PCRs covering approximately 30kb of ovine genomic DNA are currently optimised for targeted sequencing and mutation screening in affected and normal South Hampshire sheep. This region covers genomic sequence including and surrounding CLN6 - starting from the CNCS 7kb 5' of CLN6 and ending in the 3' region of CALML4. Primers were designed from published CLN6 cDNA (Tammen *et al.* 2006) and ISGC sequence. PCR products from affected, carrier and normal South Hampshire sheep will be submitted for 454 sequencing™ (Goldberg *et al.* 2006).

## RESULTS AND DISCUSSION

**Initial bioinformatic analysis.** Bioinformatic analysis of VISTA® across 9 species revealed that parts of the 5'UTR, 3'UTR and intron 1, 2 and 3 of the CLN6 contain highly conserved regions (CNCS) with potential regulatory function (Mohd Ismail *et al.* 2006). Sequence information from regions conserved in cattle and human were used to design PCR primers for amplification of sheep specific sequence predicted to be located 7-10kb upstream of *CLN6* Exon 1. Furthermore, comparative information was used to predict the identity and distance of genes flanking ovine CLN6: FEM1-b was predicted to be located 40kb upstream and CALML 4 about 7kb downstream of CLN6.

**Sequencing of the sheep BAC.** 454 sequencing™ of the sheep BAC clone 270H8 resulted in approximately 2 million base pair (bp) sequence reads. 1248 asterix symbols were found throughout the 454 sequences which represent homopolymers. These homopolymers are

considered sequencing errors and they are randomly located with only 20% occurring before or after stretches of A/T.

Contig assembly from the raw data provided by the University of Otago sequencing unit was performed using the assembler program MIRA. Sequences were assembled into 114 contigs ranging from 42 to 18863bp each with an average read of 80 reads per contig and a 13.49 fold coverage as calculated by MIRA. BLAST analysis identified that approximately 30% of the total 114 sequence contigs contained BAC vector or bacterial genomic DNA sequences. However, these were smaller in length with an average size of 200 to 400 bp. BAC vector and *E.coli*- free contigs were aligned to the cattle genome (Btau4.0) at position 14,821,789-14,942,374 bp using the Ensembl Genome Browser to produce a visual representation as illustrated in Figure 1.

Additional editing based on ISGS sequence information as well as previously derived CLN6 sequences appears to allow bridging of most gaps but is not finalised. The sequence includes CLN6, CALML4, FEM1-b as well as parts of PIAS1. Two long terminal repeat (LTR) sequences ranging from 500bp to 1kb were matched to 10 to 15 regions of the cattle chromosome 10 which made assembly difficult.

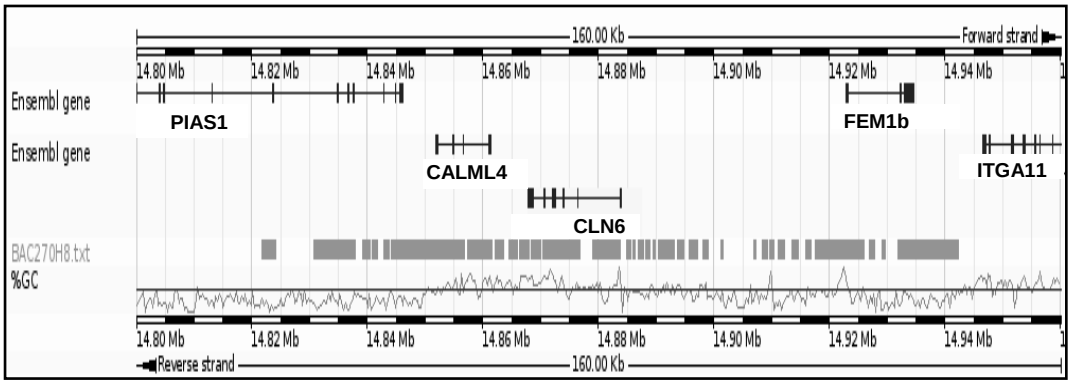


**Figure 2. Alignment of 454 sheep BAC 270H8 contigs against cattle chromosome 10: 14,800,000-14,960,000 (assembly Btau 4.0) using the Ensembl browser.**

**Targeted sequencing for mutation screening.** Ten out of 15 long distance PCR products have been optimised for amplification on sheep genomic DNA. The newly generated BAC sequence will allow for design of new primers for the remaining regions, which failed so far to amplify single band PCR products of predicted sizes. The resulting products will be individually multiplex identifier (MID) tagged and 454 sequenced.

## CONCLUSION

Whole genome sequencing has not been completed for the sheep and the publically available sequence information for our region of interest had large gaps. Traditional Sanger sequencing of the genomic CLN6 sequence, particularly in the very GC rich 5' region of the gene, has been very cumbersome (Tammen *et al.* 2006) and hindered the identification of a disease causing mutation for NCL in South Hampshire sheep. Identification of such a mutation is crucial, as these sheep are the most extensively characterised model for NCLs in general, and for vLINCL in particular. Identification of the disease causing mutation will not only assist greatly in the management of the research flock, but provide additional support for our claim that NCL in these sheep is caused by CLN6 down regulation, and thus be of great importance in relation to the evaluation of therapeutic approaches such as gene therapy.

The 454 sequencing of the sheep BAC containing CLN6 was found to be cost-effective and efficient for the generation of good coverage sequence for our region of interest. Initial contig assembly left some gaps, but these can be mostly bridged using existing sequence information or were caused due to differences between sheep and cattle sequences. This sequence provides the required backbone for the sequencing of long-distance PCR products that will be used for mutation screening. Depending on the types, number and locations of mutations identified by the proposed mutation screening approach, further experiments will be designed to assess the impact of mutations on gene expression. A number of human cases of vLINCL linked to the CLN6 gene appear to also have mutations in non-coding regions of CLN6. The identification of regulatory elements for CLN6 in sheep will hopefully indicate positions for further analysis in these patients and thus improve DNA diagnostics in patients and their relatives.

**ACKNOWLEDGMENTS**

This work has been funded by the BDSRA and NIH Grant (R01 NSO53559-01A1). We would like to thank Dr Kyall Zenger and Dr. Matthew Hobbs for assistance with bioinformatic analysis, Dr. John McEwan and the ISGC for early release of partial OAR7 sequence v1.5, and Dr. Daniel Vaiman (INRA, France) for providing the sheep BAC clone. I.F. Mohd Ismail is a sponsor student of Malaysia's Ministry of Higher Education and Universiti Putra Malaysia.

**REFERENCES**

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) *J. Mol. Biol.* **215**:403.

Broom, M. F., Zhou, C., Broom, J. E., Barwell, K. J., Jolly, R. D., and Hill, D. F. (1998) *Neuropathol. Appl. Neurobiol.* **35:**717.

Chevreux, B., Pfisterer, T., Drescher, B., Driesel, A. J., Mueller, W. E., Wetter, T. and Suhai, S. (2004) *Genome Res.* **14**:1147

Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., and Dubchak, I. (2004) *Nucleic Acids Res.* **32**:W273.

Goldberg, S.M**.,** Johnson, J**.,** Busam, D**.,** Feldblyum, T**.,** Ferriera,, S**.,** Friedman, R**.,** Halpern, A**.,** Khouri, H**.,** Kravitz, S.A**.,** Lauro, F.M**.,** Li. K**.,** Rogers, Y.H**.,** Strausberg, R**.,** Sutton, G**.,** Tallon, L**.,** Thomas, T**.,** Venter, E**.,** Frazier, M**.,** and Venter, J.C**.** (2006) *Proc. Natl. Acad. Sci. USA* **24:**103.

Jolly, R.D., Janmaat, A., West, D.M., and Morrison, I. (1980) *Neuropathol. Appl. Neurobiol.* **6:**195.

Mohd Ismail, I.F., Palmer, D.N., and Tammen, I. (2006) ICE-EM Summer Symposium in Bioinformatics, Canberra, 5-8.12.2006.

Nicholas, K.B, Nicholas, H.B. Jr., and Deerfield, D.W. (2007) *EMBNEW.NEWS* **4**:14.

Sharp, J.D., Wheeler, R.B., Parker, K.A., Gardiner, R.M., Williams, R.E., and Mole, S.E. (2003) *Hum. Mutat.* **22**:35.

Tammen, I., Houweling, P.J., Frugier, T., Mitchell, N.L., Kay, G.W., Cavanagh, J.A., Cook, R.W., Raadsma, H.W. and Palmer, D.N. (2006) *Biochim. Biophys. Acta* **1762**:898.

Vaiman, D., Billault, A., Tabet-Aoul, K., Schibler, L., Vilette, D., Oustry-Vaiman, A., Soravito, C., and Cribiu, E. P. (1999) *Mamm. Genome* **10:**585.

Wheeler, R. B., Sharp, J. D., Schultz, R. A., Joslin, J. M., Williams, R. E., and Mole, S. E. (2002) *Am. J. Hum. Genet.* **70**:537.