THE UNIVERSITY OF
SYDNEY

IDENTIFICATION OF A NOVEL MUTATION IN THE CLN6 GENE (*CLN6*) IN SOUTH HAMPSHIRE SHEEP AFFECTED WITH NEURONAL CEROID LIPOFUSCINOSIS

Izmira Farhana Mohd Ismail

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Faculty of Veterinary Science

University of Sydney

Australia

February, 2015

SUMMARY

Neuronal ceroid lipofuscinoses (NCL/Batten disease) are a group of fatal inherited neurodegenerative diseases that occur in many species including humans, sheep, dogs and cattle. Typical NCL symptoms include progressive loss of vision, regression of mental and motor development, epileptic seizures and premature death. Currently there is no effective treatment or cure for NCL, with the underlying disease mechanisms still poorly understood. Advances in molecular genetics in recent years have allowed the characterisation of hundreds of causative mutations and polymorphisms in at least 17 disease-causing genes across all species.

For some species, research colonies have been established for studies relevant to the corresponding human NCL variants. Best characterised of all animal models is the New Zealand South Hampshire (SH) sheep which is a model for the human variant late-infantile form of NCL (vLINCL). Past studies have revealed the ovine CLN6 gene (*CLN6*) as a strong candidate gene for this disease in South Hampshire sheep. No disease-causing mutation was identified in the gene coding region, however quantitative PCR revealed reduced *CLN6* messenger RNA (mRNA) expression in affected sheep compared to normal healthy sheep. The main objective of the present thesis is the identification and characterisation of the mutation responsible for NCL in the South Hampshire sheep. It was proposed that the mutation lies in the non-coding regions within or flanking the gene and that this mutation affects gene regulation.

In Chapter 3, bioinformatic tools were used to identify conserved non-coding sequences (CNCS) which are deemed potential regions of interest for regulatory mutations. Due to the limited ovine genome resource available when the study was commenced in 2006, orthologous sequences from nine other species (mouse, rat, human, cattle, macaque, dog, opossum, chicken and fugu fish) were aligned against sheep using the VISTA suite of programs to detect CNCS. These analyses resulted in the identification of five highly conserved regions in the 5' UTR, 3'UTR and introns 1, 2 and 6 of the *CLN6* orthologs. Of the five identified CNCS, the region upstream of *CLN6* and intron 1 were considered priorities for sequencing as they were more likely to contain transcriptional regulatory elements and had not been sequenced previously (region upstream of *CLN6*) or only partially sequenced

(intron 1) in sheep. An ovine BAC clone containing these CNCS regions was used as template for sequencing using the conventional Sanger method and generated 1,450 bp new ovine sequence (Chapter 4). Given that the Sanger sequencing method was laborious and time-consuming, and that there was rapid development of the next-generation sequencing (NGS) technology; the Sanger sequencing approach was abandoned and NGS utilised for the following studies.

In Chapter 5, 454 Pyrosequencing NGS technology (Roche) was used to sequence the complete ovine BAC to generate a reference sequence for mutation screening approaches. NGS sequencing of the ovine BAC method was successful and generated approximately 120kb of normal sheep genomic sequence at ~14X coverage that spanned *CLN6* and flanking genes *CALML4, FEM1b* and *PIAS1*. Two distinct mutation screening approaches were implemented using consensus sequence obtained from alignment of the ovine 454 NGS sequence (Chapter 5) with known sheep sequences from published and unpublished sources. The first mutation screening approach, sequence capture and targeted sequencing approach (Chapter 6) failed; however, the second approach involving sequencing of long range PCR (LR-PCR) products (Chapter 7), successfully identified the disease-causing mutation.

The sequence capture approach isolated and enriched a specific ovine genomic region using hybridization on the customised probes on the μParaflo microarray chip (LC Sciences). This method appeared to have not captured the targeted sequence successfully, with only 10% of captured sequences mapping back to the reference sequence. LR-PCR amplification of 14 regions within the ovine genome region followed by SOLID sequencing-by-ligation NGS method (Applied Biosystems) identified the disease-causing mutation as a 402bp deletion and 1bp insertion in ovine CLN6, namely g.-251_+150del and g.+150_151insC. The mutation is predicted to lead to the deletion of the whole of exon 1 and the ATG start codon as well as flanking non-coding sequence.

Identifying the disease-causing mutation for NCL in SH sheep provides the long-awaited confirmatory evidence that ovine CLN6 is the causative gene for NCL in SH sheep. It was anticipated that discovery of the mutation would lead to development of a DNA test to screen the SH sheep flock and a wider population of sheep. However, developing a direct DNA test using PCR has been difficult. This is because the identified deletion is in a region that is extremely GC rich and composed of highly repetitive DNA, thus the PCR-based test

produced inconsistencies in amplification and preferential amplification of one allele relative to the other in carriers. However, there might not be a great need to generate a commercial DNA test for the SH sheep as the breed is unique, not widely used for production and localised only in New Zealand. Nonetheless there is a continued effort towards developing a direct DNA test to enable more reliable and effective screening of this mutation in the current SH experimental flock. The information gained in this research confirmed that South Hampshire sheep represent a model for the CLN6 variant in humans. Future research in this large animal model will allow for more effective strategies for developing therapeutic approaches for NCL in humans and further strengthens the invaluable role of this animal model for NCL studies.

DECLARATION

I certify that the content of this thesis is the result of work I have conducted since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution.

I certify that to the best of my knowledge any help received in preparing this thesis, and all the sources used, have been properly acknowledged in this thesis.

Izmira Farhana Mohd Ismail

ACKNOWLEDGEMENTS

First and foremost, I gratefully acknowledge the scholarship bodies: Universiti Putra Malaysia (UPM), Malaysia Ministry of Higher Education and the Faculty of Veterinary Medicine, UPM for giving me the opportunity and financial support to pursue a PhD in Australia.

I would like to convey my sincere gratitude to my supervisor Assoc. Prof. Imke Tammen for her outstanding supervision and contributions of ideas, time and funding towards my PhD journey. I am especially grateful that over the years she has not only been a supervisor to me, but also mentor, colleague and friend. I find her diligence and dedication to the genetics discipline and teaching truly inspirational and I aspire to be an academician and researcher like her.

A heartfelt thank you to my associate supervisor, Prof. David Palmer from Lincoln University, Christchurch, New Zealand, for his guidance, review of my thesis and for providing valuable insights from his decades of involvement in NCL research. I truly appreciate that Dave and his wife Jeanette offered their home and warm hospitality whenever there was a need to conduct research at Lincoln University.

I thank Dr. Michelle Hyde for editorial proof reading of Chapters 2 to 8 and Prof. Frank Nicholas for his continuous support in the genetics field. I would also like to thank Nadia Mitchell at the Lincoln University, Christchurch who has been very helpful, particularly in providing DNA samples and related information from the South Hampshire sheep flock. Her collaboration has made it possible for my research to progress. I would also like to thank my colleague and friend Peter Houweling for taking me under his wing early in my candidature and for showing me the ropes around the genetics laboratories.

Thank you to A. Prof Peter Thomson and Drs. Matthew Hobbs, Julie Cavanagh, Kyall Zenger, Ivonne Petermann and John Davis who have greatly assisted in bioinformatic analysis of the next-generation sequencing data. Peter also kindly provided feedback for Chapter 9 and Julie provided feedback on Chapters 2 and 5.

I have also benefited by outstanding work and help from the University of Sydney Camden technical staff. Thank you to the dynamic duo Gina Attard and Marilyn Jones for providing diligent assistance and much humour in the laboratory; and to David Palmer, Renee Dever, Joe Davis and Doug Raadsma for managing and caring for the Merino NCL research sheep flock. My thanks also to Marie Wildridge for her kindness and exceptional administration assistance.

Thank you to my family and in-laws, mainly my seven wonderful siblings, their spouses and children who have always been supportive and encouraging every step of the way. A special thank you to my beloved parents; Hj. Mohd Ismail Ridzwan and Hjh. Alijah Hassan for inspiring me every single day and for all their love and support in my pursuits. I am truly blessed to have such an incredible family to cheer me on to the PhD finish line.

Last but not least, my heartfelt appreciation for my boys. For my amazing husband Tengku Adam Setia for his remarkable patience, encouragement and unwavering love throughout my PhD journey. For my amazing son Tengku Johan Mikail for being the wonderful little person that he is and for teaching me the meaning of unconditional love.

This thesis is lovingly dedicated to the four most important people in my life; my parents, husband and son.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ABI	Applied Biosystems
AGRF	Australian Genome Research Facility
agRNAs	antigene RNAs
AI	artificial insemination
ANCL	adult neuronal ceroid lipofuscinosis
ANGIS	Australian National Genomics Information Service
ANOVA	analysis of variance
ATP	adenosine triphosphate
BAC	bacterial artificial chromosome
BAM	binary alignment map
BBB	blood brain barrier
BDSRA	Batten Disease Support and Research Association
BLAST	basic local alignment search tool
BLAT	basic local alignment tool
bp	base pair
BSA	bovine serum albumin
BTA	bos taurus
CALML4	calmodulin-like 4
ChIP	chromatin immunoprecipitation
CL	curvilinear profile
CLEAR	coordinated lysosomal expression and regulation
CNCS	conserved non coding sequence
CNS	central nervous system
CONCL	congenital neuronal ceroid lipofuscinosis
contig	contiguous
CSIRO	Commonwealth Scientific and Industrial Research Organisation
ddNTP	dideoxynucleotide triphosphate
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNAJC5	DnaJ homolog subfamily C member 5
dNTP	deoxynucleotide triphosphate
dsDNA	double stranded DNA
DSH	domestic short hair
E value	expect value

FP	forward primer
FPR	fingerprint profile
GB	gigabyte
GROD	granular osmiophilic deposits
GS	genome sequencer
HSA15	human chromosome 15
HSCT	hematopoietic stem cell transplantation
ICD	International Classification of Diseases
IGV	Integrative Genomics Viewer
INCL	infantile neuronal ceroid lipofuscinosis
indels	insertion/deletion
ISGC	International Sheep Genomics Consortium
IUB	International Union of Biochemistry
JNCL	juvenile neuronal ceroid lipofuscinosis
kb	kilobase
kg	kilogram
KO	knockout
l	litre
LFB	luxol fast blue
LINCL	late-infantile neuronal ceroid lipofuscinosis
LINE	long interspersed nuclear elements
LOD	logarithm of the odds
LR-PCR	long-range PCR
LSD	lysosomal storage disease
LTR	long terminal repeat
M	molar
MFSD8	major facilitator superfamily domain containing protein 8
Mg	miligram
MID	multiplex identifier
min	minute
miRNA	microRNA
MITF	microphthalmia-associated transcription factor
ml	millilitre
MOET	multiple ovulation and embryo transfer
MPS	mucopolysaccharidoses
mRNA	messenger ribonucleic acid
NCBI	National Center for Biotechnology Information

NCL neuronal ceroid lipofuscinosis
ng nanogram
NGS next generation sequencing
NMD nonsense-mediated mRNA decay
nmol nanomol
OAR ovis aries
oCLN6 ovine *CLN6* gene
OD optical density
OMIA Online Mendelian Inheritance in Animals
OMIM Online Mendelian Inheritance in Man
P value probability value
PAS periodic acid-Schiff
PCR polymerase chain reaction
post-hyb post-hybridized
PPT1 palmitoyl-protein thioesterase 1
pre-hyb pre-hybridized
QC quality control
q-PCR quantitative PCR
TAE tris-acetate-EDTA
RAM random access memory
RC rectilinear complex
RFLP restriction fragment length polymorphism
RP reverse primer
RT room temperature
RT-PCR real-time polymerase chain reaction
rVISTA regulatory VISTA
SAP saposin/sphingolipid activator protein
SB sudan black
SCMAS subunit c of mitochondrial ATP synthase
sec seconds
SFF standard flowgram format
SH south hampshire
SINE short interspersed nuclear element
SNP single nucleotide polymorphism
ssDNA single stranded DNA
SUPAMAC Sydney University Prince Alfred Molecular Analysis Centre
T_a annealing temperature

TBE	tris-Borate-EDTA
TE	tris-EDTA
TF	transcription factor
TFBS	transcription factor binding site
TPP1	tripeptidyl peptidase 1
TSS	transcription start site
UTR	untranslated regions
UV	ultraviolet
V	volt
v/v	volume/volume
w/v	weight/volume
WGS	whole genome shotgun
YAC	yeast artificial chromosome
µg	microgram
µl	microliter

CHAPTER 1: LITERATURE REVIEW

1.1 Neuronal ceroid lipofuscinoses

The Neuronal ceroid lipofuscinoses (NCL/Batten disease) are a group of fatal inherited neurodegenerative diseases of humans and animals. Common NCL characteristics include progressive loss of vision, mental and motor deterioration, epileptic seizures and premature death. In most variants, onset of disease is in childhood and most affected children die before they reach adulthood. Some rarer variants have adult onset (Zeman and Dyken, 1969). All forms of NCL share varying degrees of brain and retinal atrophy, as well as accumulation of fluorescent storage bodies in neurons and many other cells throughout the body. These florescent storage bodies stain positive with periodic acid-Schiff (PAS) and Sudan black (SB), and were initially referred to as 'ceroid-lipofuscin lipopigments' based on their staining properties (Zeman and Rider, 1975). Later it was identified that all variants of the disease have protein based storage material. The storage material was identified to consist largely of either subunit c of mitochondrial ATP synthase (SCMAS; Palmer et al., 1989a, 1989b) or sphingolipid activator proteins A and D (Saposins A and D; Tyynelä et al., 1993). Despite identification of the composition of these storage materials, the term 'ceroid lipofuscinoses' continues to describe this group of diseases.

NCLs are classified as amongst the lysosomal storage diseases (LSDs; Futerman and van Meer, 2004; Parenti et al., 2013). LSDs are a group of inherited metabolic disorders, each resulting from a deficiency of a particular lysosomal protein activity resulting in the intra-lysosomal storage of a variety of substrates in multiple tissues and organs (Parenti et al., 2013). The LSD family encompasses around 50 distinct genetic diseases, with the known LSDs grouped according to various classifications (Futerman and van Meer, 2004). One method of LSD classification is based on the substrates accumulated in the lysosomes; e.g. sphingolipidoses, mucopolysaccharidoses (MPS) and oligosaccharidoses; while others are named after the enzyme deficiency e.g. galactosialidosis and multiple sulphatase deficiency. The overall prevalence of LSDs is relatively high when compared to other groups of rare diseases, with an estimation of 1 in 8,000 live births (Fuller et al., 2006). Although there is more than a century of study of the

genetic and molecular basis of LSDs, little is known about the events that lead from intra-lysosomal accumulation to the disease pathology of a distinct LSD, particularly neurological diseases, and this holds true for NCLs as well. NCL disease variants, classified as CLN1, CLN2, CLN10 and CLN13, are caused by defects in the lysosomal enzymes PPT1, TPP1, Cathepsins D and F, respectively (Table 1.1). For the other NCL genes, the links between the protein defect, lysosomal storage and pathogenesis are not known.

International research efforts, often coordinated and supported by parent support groups such as the Batten Disease Support and Research Association (http://www.bdsra.org/), have led to an enormous increase in understanding of NCL, and the information is well communicated at national and international research and parent conferences. An abundance of scientific publications are available that describe the genetics, clinical signs and pathology of different variants, proposed disease mechanisms, current and proposed therapeutic approaches, animal models and many other different aspects of NCL. Such publications are in the form of review papers including Haltia, 2006; Kohlschutter and Schulz, 2009; Bennett and Rakheja, 2013; special sections/ issues in journals arising from conferences on NCL e.g. Biochimica Biophysica Acta-Molecular Basis of Disease 1832 (2013), Biochimica Biophysica Acta-Molecular Basis of Disease issue 1762 (2006) and Molecular Genetics and Metabolism issue 66 (1999), Journal of Inherited Metabolic Diseases issue 16 (1993), American Journal of Medical Genetics issue 42 (1992) and Advances in Experimental Medicine and Biology issue 266 (1990) as well as books e.g. The Neuronal ceroid lipofuscinoses (Batten disease), Mole et al., 2011; Age Pigments, Sohal, 1981; Ceroid-lipofuscinosis (Batten's disease), Armstrong et al., 1982 and Lipofuscin: State of the Art, Z-Nagy, 1988. The following review aims to summarise aspects of this broad area of research that are most relevant to this study.

Table 1.1 Classification and characteristics of the genes causing NCL in humans (http://www.ucl.ac.uk/ncl/mutation.shtml) and review papers by Haltia and Goebel (2013), Schulz et al. (2013) and Warrier et al. (2013).

Gene	No. of mutations reported*	Protein	Eponym	Genotype-phenotype correlation[e]	Stored protein[f]	Ultrastructure[g]	OMIM/ References[h]
CLN1/PPT1	64	PPT1[a]	Halta-Santavouri	**Infantile** Late infantile Juvenile Adult	SAPs	GRODs	256730
CLN2/TTP1	109	TPP1[a]	Jansky-Bielschowsky	**Late infantile** Juvenile Protracted	SCMAS	CL	204500
CLN3	57	Transmembrane protein	Spielmeyer-Sjogren	Juvenile Protracted	SCMAS	FPR (CL, RL)	204200
CLN4B/DNAJC 5	2	Soluble cysteine string protein α	Parry	Adult autosomal dominant	SAPs	GRODs	162350
CLN5	36	Soluble lysosomal protein	Finnish vLINCL	**Late infantile** Juvenile Protracted Adult	SCMAS	RL, CL, FPR	256731
CLN6	68	Transmembrane protein	Lake-Cavanagh early juvenile/Indian vLINCL, adult Kufs type A	**Late infantile** Protracted Adult Kufs type A	SCMAS	RL, CL, FPR	601780
CLN7/MFSD8	31	Transmembrane protein	Turkish vLINCL	**Late infantile** Juvenile protacted	SCMAS	RL, FPR	610951
CLN8	24	Transmembrane protein	Northern epilepsy/ progressive epilepsy with mental retardation	**Late infantile** Protracted **EPMR/Northern eplipepsy**	SCMAS	CL-like, FPR granular	610003
CLN9			Juvenile variant	Juvenile variant			609055

Gene	No. of mutations reported*	Protein	Eponym	Genotype-phenotype correlation[e]	Stored protein[f]	Ultrastructure[g]	OMIM/ References[h]
				Congenital			
				Late infantile			
				Juvenile			
CLN10/CTSD	7	Cathepsin D[a]	Congenital	Adult	SAPs	GRODs	610127
CLN11/GRN	2	Progranulin[b]	Adult variant	**Adult**		FPR	614706
CLN12/ATP13A2	1	ATPase type 13A2[c]	Juvenile NCL	**Juvenile**		FPR	606693/ Bras et al., 2012
CLN13/CTSF	5	lysosomal enzyme	Adult Kufs type B	**Adult Kufs type B**		FPR	615362/ Smith et al., 2013
CLN14/KCTD7	1	Potassium channel tetramerization domain containing protein type 7[d]	Infantile	**Infantile**			Staropoli et al., 2012
				Adult			
SGSH	2			**Late infantile MPSIIIA**			Sleat et al., 2009
CLCN6	2			Adult (only found in heterozygous form to date)			Poet et al., 2006

* No. of disease causing mutations. Other sequence variations have also been reported.
[a] Lysosomal enzymes
[b] *GRN* mutations also cause Frontotemporal lobar degeneration with TDP43 inclusions OMIM 607485.
[c] *ATP13A2* mutations also cause Kufor-Rakeb syndrome (KRS, Parkinson disease 9) OMIM 606693
[d] *KCTD7* mutations also cause progressive myoclonic epilepsy type 3 (EPM3) OMIM 611726 and Opsoclonus-myoclonus ataxia-like syndrome
[e] Bold = phenotype caused by complete loss of gene function
[f] SCMAS, subunit c of mitochondrial ATP synthase; SAPs, sphingolipid activator proteins.
[g] GRODs, granular osmiophilic deposits; CL, curvilinear profiles; FPR, fingerprint profiles; RL, rectilinear profiles.
[h] Due to the large number of relevant references to be listed per disease, an OMIM ID which links to these references is listed. References are listed for NCL variants without an OMIM ID.

1.1.2 NCL in humans

1.1.2.1 History of NCL

The NCL have been recognised for nearly 200 years, with the first probable instances of the disease described in four affected siblings in 1826 by Dr. Otto Stengel (Brean, 2004). Related publications appeared almost a century later, and the disorder was defined in 1903 as Batten disease (Batten, 1903), whilst also being classified under a generalised heading of amaurotic family idiocy (Sachs, 1896) due to its clinical resemblance to Tay-Sachs disease. Later reports disassociated the relationship between Batten disease and family amaurotic idiocy, and introduced the term neuronal ceroid lipofuscinosis (Zeman and Dyken, 1969).

1.1.2.2 Nomenclature

There have been several methods used to classify the different NCL variants in humans. BDSRA, an organisation which supports families of persons with NCL, follows the traditional classification introduced by Zeman (1976) based on the clinical onset of the diseases. In NCL, the earlier the onset of disease, the more severe is the progression of the disease and the shorter is the life expectancy. In brief, classical infantile (INCL, Santavuori-Haltia) onset begins at 8 to 18 months, late-infantile (LINCL, Jansky-Bielschowsky) has an onset between the ages of 2 and 4 years, juvenile NCL (JNCL, Batten disease, Spielmeyer-Vogt) usually commences between the ages of 5 and 8 years, and in adult NCL (ANCL, Kufs disease) clinical signs start in adulthood, typically around the age of 30. In addition to those variants described by Santavuori et al. (1973) and Zeman (1976), some cases with congenital variants of NCL have also been described (Siintola et al., 2006). The term 'Batten disease' was conventionally regarded as the juvenile form of NCL; however, the term is often used to describe all forms of NCL in both humans and animals.

Advances in molecular genetics have led to a re-classification of the nomenclature for NCL according to predicted or confirmed genes, with disease variants assigned to '*CLN*' symbols prior to the gene names e.g. *CLN1* for INCL. The online NCL mutation database (http://www.ucl.ac.uk/ncl/mutation.shtml) which is the database curated scientifically by a

leading expert in NCL research (Dr Sara Mole) has classified NCL into 13 genetic forms (*CLN1*, *2*, *3*, *4*, *5*, *6*, *7*, *8*, *10*, *11*, *12*, *13* and *14*), with potentially 2 additional genes; *CLCN6* and *SGSH* causing NCL in humans (Table 1.1). Lack of information about the ultrastructure and type of storage bodies in some of the more recent proposed variants makes their clear characterisation of NCL difficult.

The most recently suggested nomenclature proposes an axial diagnostic classification system similar to those used for the epilepsies and for mental health disorders in children (International Classification of Diseases-10 or ICD-10; http://www.who.int/classifications/icd/en/) and better reflects the complexity of this group of diseases. This new nomenclature is depicted in detail by Williams and Mole (2012). Briefly, the axial diagnostic system comprises of 7 axes describing the affected gene, mutation diagnosis, biochemical and clinical phenotype, ultrastructural features of the storage bodies, functionality and other remarks.

1.1.2.3 Incidence

NCLs have a worldwide distribution, and are often described as the most common inherited progressive neurodegenerative diseases of childhood (Santorelli et al., 2013). Reported epidemiological data are likely to be underestimates, because of missed diagnoses and non-reporting of this group of rare diseases; and the following published incidence and prevalence rates should be considered in that context. Incidence rates are reported to vary between countries and geographical regions, with 1:67,000 in Italy and Germany to 1:14,000 in Iceland and their prevalence rates vary from 1:1,000,000 in some regions to 1:100,000 in the Scandinavian countries (Uvebrant and Hagberg, 1997). Historically, higher incidences have been described in the Scandinavian countries (Mole, 2011). A growing awareness of diseases, as well as improvements in DNA and other diagnostics procedures, have led to increase in cases identified and reported in recent years. Advances in genetics in recent years, specifically, use of molecular genetics has led to identification of the NCL disease associated genes and detection of a spectrum of mutations in these causative genes to support clinical diagnoses (Santorelli et al., 2013). Despite these improvements in diagnostics, patients and their families often experience a long lasting odyssey involving delays in diagnosis, which is also commonly observed for many other rare diseases (Anderson et al., 2013).

1.1.2.4 Clinical signs and pathology

In almost all NCL variants the patients are initially healthy and have a normal developmental profile before exhibiting clinical signs. The key clinical signs of NCLs are progressive loss of vision, behavioural changes, regression of mental and motor development, epileptic seizures and premature death. The order of appearance of these clinical signs varies depending on the age at onset and the genetic form. Recently, a study has shown that some patients with the CLN3 variant can also have cardiovascular system abnormality (Cotman and Staropoli, 2012).

Macroscopically, the brain is almost always reduced in size, mainly due to cerebral cortical atrophy (Haltia, 2003). Light microscopy examination reveals loss of neurons, mostly in the cerebral and cerebellar cortices, with depletion of neurons usually accompanied and followed by fibrillar gliosis. Atrophy of the retina is observed in some variants (Goebel, 1992; Goebel et al., 1998). The accumulated storage bodies in the nerve cell appear brownish in haematoxylin-eosin stained sections, with a mild brownish shade to the grey matter at gross inspection and stains with periodic acid-Schiff (PAS), luxol-fast blue (LFB) and Sudan black. Electron microscopy reveals various ultrastructural patterns including granular osmiophilic deposits (GRODs), curvilinear profiles (CL), rectilinear complex (RC) and fingerprint profiles (FPR) (Mole et al., 2011). Furthermore, a combination of FPR and CL/RL has been reported in patients with CLN3, CLN5 and CLN6 forms (Table 1.1).

1.1.2.5 Genetics

By early 2014, there were at least 15 genes reported to be associated with NCL in humans. These genes are listed in the NCL resource and mutation database maintained by Sara Mole and colleagues (http://www.ucl.ac.uk/ncl/mutation.shtml) and summarised in Table 1.1. NCLs follow an autosomal recessive mode of inheritance, with the exception of ANCL (Table 1.1) which has both a dominant and recessive form (Haltia, 2006; Arsov et al., 2011).

The first causative gene identified via linkage analysis was the *CLN1/PPT1* (Vesa et al., 1995), closely followed by *CLN3* (International Batten disease consortium, 1995), *CLN2/TPP1* (Sleat et al., 1997), *CLN5* (Savukoski et al., 1998), *CLN8* (Ranta et al., 1999), *CLN6* (Wheeler et al., 2002; Gao et al., 2002), *CLN10/CTSD* (Siintola et al., 2006; Steinfeld et al., 2006), *CLN7/MFSD8* (Siintola et al., 2007) and *CLN4/DNAJC5* (Nosková et al., 2011). Four causative genes have been reported recently; *CLN11/GRN* (Smith et al., 2012), *CLN12/ATP13A2* (Bras et al., 2012), *CLN13/CTSF* (Smith et al., 2013) and *CLN14/KCTD7* (Staropoli et al., 2012). No gene causing the CLN9 variant has been identified (Warrier et al., 2013).

Conventionally, NCL genes were identified using linkage analysis. In more recent studies, linkage disequilibrium, gene expression, homozygosity mapping, and exome sequencing were utilised (Warrier et al., 2013). *CLN2/TPP1* was identified using a biochemical approach (Sleat et al., 1997). The *CLN8* (Ranta et al., 1999) and *CLN10/CTSD* genes (Siintola et al., 2006; Steinfeld et al., 2006) were identified using animal models. *CLN10/CTSD* was first identified to cause a congenital NCL in Swedish landrace sheep (Tyynelä et al., 2000) prior to the identification of human cases with congenital onset. For the recently proposed genes, only 1 or 2 disease causing mutations have been described, but for all other genes many different disease causing mutations are known (Table 1.1) (Warrier et al., 2013). Different mutations in the same gene can cause different onset (Table 1.1). Interestingly, some of the recently proposed genes: *CLN12/ATP13A2*, *CLN14/KCTD7, SGSH* and *CLCN6* have been reported to also cause non-NCL diseases (Table 1.1).

1.1.2.6 Disease mechanism

In spite of all of these research efforts to study the NCL proteins, and recent advances in molecular genetics, it remains unclear if all or some of the different genes identified so far interact or participate in common pathways relative to each other. Interactions between several NCL proteins have been reported, between CLN5 and CLN1/PPT1, CLN2/TPP1, CLN3, CLN6 and CLN8 (Vesa et al., 2002; Lyly et al., 2009), suggesting a modifying role for these proteins in the pathogenesis of individual NCL disorders (Lyly et al., 2009). The functions and localisations of several of the normal proteins coded for by NCL genes are not known and require further studies.

Researchers currently lack a good understanding of the underlying NCL disease mechanisms, with unclear pathways between the identified proteins to the defining pathology (Palmer et al., 2013).

In the early years, it was proposed that accumulated storage bodies were formed due to an increased rate of peroxidation of fatty acids (Zeman and Rider, 1975). This concept of peroxidation persisted until the end of the 1980s when it was established that either subunit c of mitochondrial ATP synthase (Palmer et al., 1989a) or sphilongolipid activator proteins (Tyynelä et al., 1993) were the major components of these storage bodies (Haltia and Goebel, 2013).

A review by Palmer et al. (2013) discussed a range of potential disease mechanisms that are proposed to be involved in NCL disease. It was highlighted that there has been a comprehensive increase in knowledge of NCL genetics since the 1990s, with discovery of disease causing mutations for *CLN1, CLN2, CLN3, CLN5, CLN6, CLN7, CLN8* and *CLN10*. Other genes have been proposed recently, including *CLN4/DNAJC5, CLN11/GRN, CLN13/CTSF, CLN14/KCTD7* and *CLN12/ATP13A2*. Histopathological studies in the ovine models identified a possible important role of inflammation in early stages of disease: neuroinflammation was present well in advance of clinical signs and preceded neurodegeneration in the disease pathogenesis (Oswald et al., 2005; Tammen et al., 2006). Immunohistochemical analyses revealed that glial activation begins prenatally, long before significant storage body accumulation or neuron loss (Oswald et al., 2005; Kay et al., 2006). Furthermore, it was discussed that a recent study proposed the involvement of metal dysregulation in the disease process (Kanninen et al., 2013a, 2013b). However, it remains unclear if this is a causative mechanism or secondary responses to the disease.

The review paper concluded that currently no single theory or disease mechanism exists to determine the pathogenesis of NCL. It is hoped that an understanding of this complex group of diseases can be achieved through collective consideration of multiple disciplines in the future. Animal models were proposed to play important roles in such studies.

1.1.2.7 Therapeutic approaches

Currently there is no effective treatment or cure for NCL (Bennett and Rakheja, 2013; Augustine et al., 2013). In the past, therapies were restricted to symptomatic approaches, which include antileptics for seizure management, physical/occupational therapy and medications to address motor impairment and movement disorders (Augustine et al., 2013). Supportive care led to a patient's relatively prolonged life expectancy but did not address the cause of the disease, halt or suppress disease progression nor replenish depleted or compensate the defective gene/enzyme.

The improved knowledge of the pathophysiology of the lysosomal storage diseases (LSD) generally, and NCL, specifically, in recent years, has spurred development of a cure versus attempting to treat clinical signs of the disease. Prior to recent advancement in knowledge, experimental treatments for NCL have been explored. Based on the misdiagnosis of the storage material as a lipofuscin, dietary intervention using anti-oxidant supplementation was explored (Santavuori and Moren, 1977; Santavuori et al., 1988). The efficacy of this therapy demonstrated no significant effect on disease outcome (Santavuori et al., 1989). More recent therapeutic approaches indicated for LSD have been hematopoietic stem cell transplantation (HSCT) using bone marrow transplants, which relies on transplantation of hematopoietic stem cells for repopulation of specific tissues and secretion of the functional enzyme into the blood circulation (Parenti et al., 2013). Such treatment did not seem to be beneficial for NCL patients (Lönnqvist et al., 2001) but appeared to work for the treatment of the mucopolysaccharidosis (MPS) group of LSD (Valayannopoulos and Wijburg, 2011).

Similar to other LSDs, therapeutic approaches that replace or increase the residual activity of the defective enzyme are applicable to the soluble lysosomal forms of NCL enzymes (CLN1/PPT1, CLN2/TPP1, CLN5 and CLN13/CTSF) (Table 1.2). The therapeutic approaches are enzyme replacement therapy (infusions of recombinant wild type enzymes to replace the defective enzymes), gene therapy (*in vivo* gene transfer to deliver a normal copy of the defective gene able to synthesis normal enzyme), substrate reduction therapy (administration of specific drugs to inhibit specific biosynthetic pathways of substrates), stem-cell therapy (transfer of stem cells to produce the required enzyme via cross-correction and integration into the diseased CNS to

replace lost cells), pharmacological chaperones (administration of drugs to interact with the mutant protein and increase enzyme activity) and various other pharmacological approaches. Other experimental therapies are stop-codon readthrough technology (Sleat et al., 2001) and manipulation of the autophagic pathway (Raben et al., 2010). Combination of various therapeutic approaches might be necessary for best results based on research in animal models (Porto et al., 2009).

Some therapeutic strategies applied to other LSDs are difficult to apply in NCL due to the necessity of passing molecules through the blood brain barrier (BBB) to prevent neurodegeneration, and repairing or restoring CNS functionality is difficult once neurons are lost (Kohan et al., 2011). The membrane-bound NCL proteins (CLN3, CLN6, CLN7/MFSD8, and CLN8) are even more challenging than the soluble lysosomal forms as there are currently no effective methods to introduce a normal membrane protein to a large number of cells. Despite this, therapies have been trialled in variant CLN3 patients and mouse models (Seehafer et al., 2011; Kovács et al., 2012) and the CLN6 variant sheep models (Kay and Palmer, 2013; Linterman et al., 2011). Small molecular weight molecules pharmacologically chaperoned are also potentially attractive therapeutics for both the soluble and membrane-bound proteins because they cross the BBB (Dawson et al., 2010).

A range of clinical trials described in the American clinical trials website (www.clinicaltrials.gov) and NCL online website http://www.ucl.ac.uk/ncl/treatment.shtml) are summarised in Table 1.2 and have been conducted, are in progress or are planned. These approaches involve gene therapy, stem-cell therapy and pharmacological approaches.

Most, if not all of the aforementioned therapies are still in their early stages and do mostly promise a slowing down of disease progression. Therefore, early diagnosis and intervention before the onset of irreversible pathology will be essential to benefit patients. Future directions of therapeutic development depend on improved understanding of the human and animal model phenotypes and comprehension of underlying disease mechanisms to achieve significant therapeutic advances. It is hoped that a combination of various approaches might ultimately help to develop a cure.

Table 1.2: Summary of information on human clinical trials for NCL. Information has been adapted from the American clinical trials website (www.clinicaltrials.gov) and NCL online database (http://www.ucl.ac.uk/ncl/).

Gene/ NCL variant	Clinical trials number	Patients' age	Number of patients	Country of operation	Trial progress	Trial phase	Trial description	Outcome	References
CLN1/ INCL	NCT00028262	6 months to 3 years	9	USA	completed	2	**Pharmacological approach:** Test of the effectiveness of a drug called Cystagon.	Preliminary results showed that although several parameters of disease progression were slowed due to the treatment, it did not completely arrest the neurodegenerative process	Gavin et al., 2013
CLN1/ INCL	NCT01238315			USA	withdrawn prior to enrollment		**Stem-cell therapy:** Safety and efficacy study of human central nervous system stem cells (HuCNS-SC)		
CLN1/ INCL or *CLN2*/ LINCL	NCT00337636	18 months to 12 years	6	USA	completed	1	**Stem-cell therapy:** Study of human central nervous system stem cells (HuCNS-SC) cells.	Trial was in phase 1 with no outcome measures provided.	Selden et al., 2013
CLN2/ LINCL	NCT00151216	3 to 18 years	11	USA	active	1	**Gene transfer:** Safety study of a gene transfer vector encoding the normal human *CLN2* (AAV2CUhCLN2).	Currently in phase 1 and expected to be completed in June 2019.	Worgall et al., 2008

Gene/ NCL variant	Clinical trials number	Patients' age	Number of patients	Country of operation	Trial progress	Trial phase	Trial description	Outcome	References
CLN2/ LINCL	NCT014 14985	3 to 18 years	-	USA	recruiting	2	**Gene transfer:** Administration of adeno-associated virus gene transfer vector serotype rh.10 (AAVrh.10) to express the human *CLN2* cDNA.The study will assess the safety and efficacy of the virus AAVrh.10 to deliver the *CLN2* gene to children with late infantile neuronal ceroid lipofuscinosis with uncommon genotypes or moderate/severe impairment .	Currently in phase 2 and expected to be completed in December 2022.	
CLN2/ LINCL	NCT011 61576	2 to 18 years	-	USA	recruiting	1	**Gene transfer:** Safety study of a gene transfer vector using a virus called AAVrh.10 to determine if the transfer slows the rate of progress of the disease.	Currently in phase 1 and expected to be completed in August 2032.	

Gene/ NCL variant	Clinical trials number	Patients' age	Number of patients	Country of operation	Trial progress	Trial phase	Trial description	Outcome	References
CLN3/ JNCL	NCT013 99047	6 to 25 years	-	USA	recruiting	2	**Pharmacological approach:** Establishing the safety and tolerability of short-term (8 weeks) administration of mycophenolate mofetil in ambulatory children.	Currently in phase 2 and expected to be completed in August 2015.	
CLN2/ LINCL	NCT019 07087	3 to 16 years	-	Germany, UK	recruiting	2	**Pharmacological approach:** Safety and Efficacy Study of BMN190 for the Treatment of CLN2 Patients	Currently in phase 2 and expected to be completed in March 2016.	

1.2 NCL in species other than humans

1.2.1 Introduction

Various forms of NCL have been diagnosed in at least 12 different animal species other than humans (www.ucl.ac.uk/ncl/animal.shtml) (Table 1.3), including sheep (Jolly and West, 1976; Järplid and Haltia, 1993; Edwards et al., 1994; Cook et al., 2002; and Jolly et al., 2002a). In some species, animals have been genetically modified to create models of NCL disease. Reviews of the various animal models for NCL have been published (Mole et al., 2011; Bond et al., 2013).

Some of the natural occurrences of NCL in animals have been well characterised and research populations established as they are considered valuable models for the corresponding human NCL variants. Currently, maintained research populations exist for dogs and sheep (Bond et al., 2013). Mouse models have been developed for most of the characterised human NCL variants (Cooper et al., 2006), and these are described in the NCL mouse model database (http://www.ucl.ac.uk/ncl-models) as well as in the review paper by Bond et al. (2013). These existing mouse models have either naturally occurring mutations (Bronson et al., 1993; 1998) or have undergone genetic modifications (Saftig et al., 1995; Katz et al., 1999; Mitchison et al., 1999; Gupta et al., 2001; Cotman et al., 2002; Kopra et al., 2004; Sleat et al., 2004; Jalanko et al., 2005; Eliason et al., 2007). Efforts are in progress to establish mouse models for each variant from the same C57BL/6 strain backgrounds to facilitate direct comparisons (Bond et al., 2013). In recent years, experimental unicellular (yeast) (Haines et al., 2009) and simple animal models such as fly (Myllykangas et al., 2005), nematode worm (Jansen et al., 1997) and zebrafish (Zon and Peterson, 2005) have also been developed for NCL research. Among the diverse animal models, naturally occurring NCL in some sheep, cattle and dog breeds as well as experimentally induced mouse models are the best documented cases of NCL by far.

Table 1.3: Summary of non-laboratory animal models with confirmed or suggested NCL. Information in the table has been adapted from a review paper by Bond et al. (2013).

Species/ Breed (Life span*)	**Gene**	**Genetic mutation**	**Molecular defect**	**Onset of clinical signs**	**Retinal defect**	**Mortality**	**Clinical Signs *****				**Storage material ******				**Original References**
							VI	BC	MD	S	F	US	Subunit c, SAPs	D	
Sheep (20)															
South Hampshire	*CLN6*	deletion of exon 1	lack of protein	7-12m	yes	25-30m	yes	yes	yes	yes	yes	lamellar	subunit c	general	Jolly et al., 1980; Broom et al., 1998; Tammen et al., 2006; Tammen, unpublished
Swedish Landrace	*CTSD/ CLN 10*	G**>A	Asp**>Asn non-functional protein	at birth	yes	<1m	nr	yes	yes	yes	yes	GRODs	SAPs	general	Järplid and Haltia, 1993; Tyynelä et al., 2000
Merino	*CLN6*	c.184C>T	p.Arg62Cys non-functional protein	7m	yes	19-27m	yes	yes	yes	yes	yes	lamellar	subunit c	general	Cook et al., 2002; Tammen et al., 2001; Tammen et al., 2006
Borderdale	*CLN5*	c.571+1G>A	splicing variant, truncated protein	10-11m	yes	24m	yes	yes	nr	nr	yes	lamellar	subunit c	general	Jolly et al., 2002a; Palmer, unpublished; Frugier et al., 2008
Rambouillet	n.d.			8m	yes	24m	yes	yes	yes	nr	yes	nr	nr	neuronal	Edwards et al., 1994; Woods et al., 1993
Cattle (20)															
Devon	*CLN5*	c.662dupG	p.Arg221GlyfsX6 truncated protein	9m	yes	39m	yes	yes	yes	no	yes	lamellar	subunit c	general	Harper et al., 1988; Jolly et al., 1992; Houweling et al., 2006a, 2006b
Beefmaster				12m	nr	18m	yes	yes	nr	yes	nr	lamellar	nr	general	Read and Bridges, 1969
Holstein				nr	yes	18m	yes	nr	nr	nr	yes	lamellar	nr	general	Hafner et al., 2005

Species/ Breed (Life span*)	**Gene**	**Genetic mutation**	**Molecular defect**	**Onset of clinical signs**	**Retinal defect**	**Mortality**	**Clinical Signs *****				**Storage material ******				**Original References**
							VI	BC	MD	S	F	US	Subunit c, SAPs	D	
Ferret (12)															
Domestic				> 3y	yes	nr	yes	yes	yes	nr	yes	lamellar	subunit c	general	France et al., 1999; M. France pers. comm.
Domestic				3m	nr	4m	nr	yes	yes	nr	yes	GRODs	SAPs	general	Nibe et al., 2011
Cat (34)															
DSH				8.5m	nr	9m	yes	nr	yes	yes	yes	lamellar	subunit c	neuronal	Weissenböck and Rössel, 1997
Siamese				< 22m	nr	23m	yes	yes	yes	yes	nr	lamellar	nr	general	Green and Little, 1974
DSH				15m	yes	20m	yes	yes	no	yes	yes	lamellar	nr	neuronal	Bildfell et al., 1995
Japanese				7m	nr	11m	nr	nr	yes	yes	yes	GRODs	nr	general	Nakayama et al., 1993
Horse (50)															
Aegidienberger				6m	nr	24m	yes	yes	yes	nr	yes	lamellar	subunit c	general	Url et al., 2001
Goat (20)															
Nubian				10-18m	nr	2-4y	nr	yes	yes	nr	yes	lamellar	nr	neuronal	Fiske and Storts, 1988
Pig (20)															
Vietnamese pot-bellied				2y	nr	2.5y	no	no	yes	no	yes	lamellar, GRODs	nr	neuronal	Cesta et al., 2006
Parrot															
Lovebird *(12)*				< 9m	nr	9m	nr	nr	yes	yes	yes	nr	nr	neuronal	Reece and MacWhirter, 1988

Species/ Breed (Life span*)	**Gene**	**Genetic mutation**	**Molecular defect**	**Onset of clinical signs**	**Retinal defect**	**Mortality**	**Clinical Signs *****				**Storage material ******				**Original References**
							VI	BC	MD	S	F	US	Subunit c, SAPs	D	
Monkey															
Cynomolgus monkey (37)				Pre-clinical at 7y							yes	variable	nr	general	Jasty et al., 1984
Duck (29)															
Mallet Duck				1 y	nr	3y	yes	yes	yes	yes	yes	nr	nr	neuronal	Evans et al., 2012
Dogs															
Miniature dachshund	*PPT1/ CLN1*	c.736_737insC	frameshift after Gly245 with stop codon at position 276	9 months	yes	nd	yes	yes	yes	no	yes	GRODs	nr	neuronal	Sanders et al., 2010
Longhaired Dachshund	*TPP1/ CLN2*	c.325delC	frame shift after amino acid 107 (exon 4) with stop codon at position 114	7-9 months	yes	12 mo	yes	yes	yes	yes	yes	lamellar	nr	neuronal	Awano et al., 2006a; Katz et al., 2008; Sanders et al., 2011; Vuillemenot et al., 2011;
Border collie	*CLN5*	c.619C>T	truncated protein (stop codon Q206X)	16-23 months	yes	28 mo	yes	yes	yes	yes	yes	lamellar	nr	neuronal	Taylor and Farrow, 1988, 1992; Studdert and Mitten, 1991; Franks et al., 1999; Melville et al., 2005
Australian Shepherd	*CLN6*	c.829T>C	missense Trp277Arg	1-2 years	yes	nr	yes	yes	yes	nr	yes	lamellar	subunit c SAPs	neuronal	O'Brien and Katz., 2008; Katz et al., 2011
English setter	*CLN8*	c.491T>C	missense L164P	1-2 years	yes	2y	yes	yes	yes	yes	yes	lamellar	subunit c	neuronal	Koppang, 1970, 1988; Katz et al., 2005a

Species/ Breed (Life span*)	**Gene**	**Genetic mutation**	**Molecular defect**	**Onset of clinical signs**	**Retinal defect**	**Mortality**	**Clinical Signs *****				**Storage material ******				**Original References**
							VI	BC	MD	S	F	US	Subunit c, SAPs	D	
American bulldog	*CTSD/ CLN10*	c.?G>A	missense Met199I1e	1-3 years	no	7y	no	yes	yes	no	yes	round uniformly staining inclusions embedded within granular matrixes	nr	neuronal	Evans et al., 2005; Awano et al., 2006b
Tibetan terrier	*ATP13A2/ CLN12*	1620delG	skipping of exon 16 leading to shortened protein	5-7 years	yes	nr	yes	yes	yes	yes	yes	lamellar granular contents	nr	neuronal	Riis et al., 1992; Farias et al., 2011; Wöhlke et al., 2011
Australian Shepherd (not CLN6)	n.d	n.d	n.d	15 months	yes	nr	yes	yes	yes	nr	yes	lamellar	nr	neuronal	O'Brien and Katz, 2008
Polish owczarek Nizinny/ Polish Lowland sheepdog	n.d	n.d	n.d	6 months-4 years	yes	8y	yes	yes	yes	no	yes	GRODs	SAPs	general	Narfström and Wrigstad, 1995; Wrigstad et al., 1995; Nilsson and Wrigstad, 1997; Narfström et al., 2007
Miniature schnauzer	n.d	n.d	n.d	2 years	yes	nr	yes	yes	yes	no	yes	GRODs	SAPs	neuronal	Smith et al., 1996; Jolly, 1997; Palmer et al., 1997a, 1997b
Chihuahua	n.d	n.d	n.d	16-18 months	yes	24 mo	yes	yes	yes	no	yes	lamellar	nr	general	Kuwamura et al., 2003; Nakamoto et al., 2011
Cocker spaniel	n.d	n.d	n.d	18 months	yes	6y	yes	yes	yes	yes	yes	lamellar	nr	neuronal and general	Nimmo Wilkie and Hudson, 1982; Jolly et al., 1994b; Minatel et al., 2000

Species/ Breed (Life span*)	**Gene**	**Genetic mutation**	**Molecular defect**	**Onset of clinical signs**	**Retinal defect**	**Mortality**	**Clinical Signs *****				**Storage material ******				**Original References**
							VI	BC	MD	S	F	US	Subunit c, SAPs	D	
Dalmation	n.d	n.d	n.d	6 months	yes	6y	yes	yes	yes	yes	yes	lamellar	nr	general	Goebel and Dahme, 1986; Goebel et al., 1988
Japanese retriever	n.d	n.d	n.d	3 years	nr	nr	nr	nr	nr	yes	yes	n.d	nr	general	Umemura et al., 1985
Welsh corgi	n.d	n.d	n.d	6-8 years	yes	nr	yes	yes	nr	yes	nd	n.d	nr	nr	Jolly et al., 1994a
Labrador retriever	n.d	n.d	n.d	7 years	no	nr	no	no	yes	yes	nr	lamellar	nr	neuronal	Rossmeisl et al., 2003
Australian cattle dog	n.d	n.d	n.d	1 year	yes		yes	no	yes	no					Wood et al., 1987; Sisk et al., 1990
Saluki	n.d	n.d	n.d	1 year	no	2yr	no	yes	yes	no	yes	lamellar	nr	neuronal	Appleby et al., 1982

*maximum life span in years (http://www.demogr.mpg.de/cgi-bin/longevityrecords/entry.plx/);
** corresponding to G934 and Asp295 of corresponding human CTSD sequence
***VI = visual impairment, BC = behavioural changes, MD = motor deficits, S = seizures
****F = fluorescent, US = ultrastructure, D = distribution
only central nervous system investigated, nr = not reported

Animal models for the human forms of NCL are more or less clinically, ultrastructurally, biochemically and genetically matched to human NCL. Large domestic animal models such as sheep and cattle are considered particularly valuable because their relatively larger brain size and structures are considered more human-like, their longer life expectancy allows for investigation of long-term effects of treatments, and their clinical progression more closely resembles that in humans (Houweling et al., 2005; Mole et al., 2011). The ovine and canine models are particularly relevant from the genetic perspective of NCL due to the high homology between ovine and canine genes and their human orthologs. Accumulating evidence suggests that mutations in orthologous genes yield similar disease phenotypes (Broom et al., 1998; Katz et al., 2011). The use of livestock such as sheep in research is considered by many as less controversial compared to dogs, which are companion animals.

Animal models allow for well designed studies to investigate the many unanswered questions in relation to NCL pathomechanisms and the evaluation of therapeutic approaches, as it is possible to acquire tissue samples from statistically meaningful numbers of affected and age-matched control animals at different time points along the course of the disease (Jolly, 1993). These animals often share the same strain or breed and genetic background (Järplid and Haltia, 1993; Tammen et al., 2006; Katz et al., 2008), and the same environment, which facilitates phenotypic comparisons. There are limited opportunities to access samples from human patients during disease progression and autopsy.

Despite the aforementioned benefits, interspecies differences between humans and animals lead to some extent to, limitations and a degree of variation in the validity of animal models. These differences may reflect the interspecies diversity, for instance, in terms of neuroanatomical structures and behaviour pattern which result in minor modifications to the clinical expression of the defect such as seen in mice (Bond et al., 2013) and sheep (Jolly et al., 1992) as compared to humans. The reliability of information obtained from studies in animal models is occasionally questioned when predicting the effectiveness of treatment strategies in human clinical trials, which effectiveness sometimes can be attributed to factors such as insufficient statistical power to detect the true benefit of the treatment and presence of a range of physiological variables in the patients which may affect outcome (van der Worp et al., 2010).

Despite these limitations, knowledge gained in NCL animal models as outlined in the following paragraphs has been essential in increasing understanding of the genetics, biochemistry and pathomechanism of NCL, and is crucial in the development and preclinical evaluation of proposed therapeutic approaches (Bond et al., 2013). Ultimately, appropriate caution needs to be taken when drawing conclusions and translating information obtained from studies of animal models.

The ovine models will be the focus of the following review on NCL animal models. Information on NCL in other species will be briefly presented and is summarised in Table 1.3. An outline of all the animal models can be found in reviews by Mole et al., 2011 and Bond et al., 2013.

1.2.2 NCL in Sheep

Sheep are one of the most abundantly researched animal species for the purpose of NCL animal models. To date, there are at least 3 naturally occuring mutations in 3 different genes (*CLN6, CLN5, CTSD/CLN10*) identified in sheep causing NCL in Merino, Borderdale and Swedish Landrace sheep, a second, so far unidentified, mutation is proposed for South Hampshire sheep in the *CLN6* gene, and no gene has been assigned for NCL in Rambouillet sheep (Table 1.3). The following ovine NCL models are listed according to their respective breed.

1.2.2.1 South Hampshire sheep

The best described NCL sheep model is the South Hampshire (SH) sheep, a breed used intensively in NCL research in New Zealand. The earliest cases were reported by Jolly and West (1976) in two 18 month old rams from the same flock that exhibited the abnormal behaviour of unusual docility, blindness, mild ataxia and generalised tremors when excited. An experimental research flock for the South Hampshire sheep was established at Massey University, New Zealand (Jolly et al., 1980) with animals derived from the original flock (Jolly and West, 1976) and control sheep comprised of age-matched New Zealand Romney or unaffected South Hampshires (Jolly et al., 1980). This experimental flock has since undergone changes and is

described in detail in Chapter 2.1.1, with the animals now maintained at the Lincoln University, Christchurch, New Zealand by Prof. David Palmer.

Affected SH sheep are generally normal at birth, but develop clinical symptoms between 9 and 12 months of age, notably blindness (Jolly et al., 1980). Other clinical signs include nibbling at air, intermittent episodes of face (lips, ears), head, neck and occasionally limb musculature tremors (Jolly et al., 1989, 1992). The most significant gross pathological change noted in affected South Hampshire sheep is atrophy of the brain (Jolly et al., 1989), notably the cerebral cortex, by 12 months of age (Oswald et al., 2005). Histopathological examination showed storage bodies in the neurons and macrophages throughout the central nervous system, retinal photoreceptors and a wide variety of other cells within the body such as the cardiac muscle and pancreatic epithelial cells (Jolly and West, 1976; Jolly et al., 1990; Goebel, 1992; Kay et al., 1999). Accumulation of storage bodies in the cerebral cortex is largely confined to cells with neuronal morphology for the first 6 months and progressed to non-neuronal cells beyond this age (Oswald et al., 2005). These storage bodies stained positive on Periodic acid-Schiff (PAS), Sudan black, luxol fast blue (LFB) stains and emitted bright yellow/green fluorescence under ultraviolet light (Jolly and West, 1976; Jolly et al., 1989). Degenerative changes were observed initially in the visual cortex with loss of layer definition by 12 months and very few neurons remaining after 19 months (Jolly and Walkley, 1999; Oswald et al., 2005).

Both neurodegeneration and neuroinflammation, which are hallmarks of NCL, become generalised and more severe with increasing age and clinical severity (Oswald et al., 2008; Kay and Palmer, 2013). Recent time course studies presented neuroinflammation as preceding neurodegeneration (Oswald et al., 2005). Mechanisms of neurodegeneration have been postulated by Jolly and Walkley (1999) as being caused by energy-linked excitotoxicity, whereas Lane et al. (1996) established apoptosis as the mechanism of neuronal and photoreceptor cell death in both humans and animal models. Other aspects of neurodegeneration in affected NCL sheep brain have also been studied (Kay et al., 1999, 2011).

Electron microscopy disclosed that the storage bodies in the ganglion cells, bipolar cells and photoreceptors are comprised of lamellar and fingerprint profile, as well as curvilinear profile ultrastructures (Goebel, 1992).

In the retina, most cells showed accumulation of fluorescent storage bodies, most prominently in the ganglion cells (Graydon and Jolly, 1984). The electroretinogram showed that photoreceptor cells underwent reduction in both rod and cone b-wave amplitudes, with rod b-wave changes preceding those of cones which begin in the central areas of the retina (Graydon and Jolly, 1984; Mayhew et al., 1985). Photoreceptor cells became virtually absent at the terminal stage of disease (Jolly et al., 1989) which was around 2 years of age (Goebel, 1992). In addition to these changes, electron microscopy showed the formation of abnormal dystrophic rod and cone outer segments in photoreceptor cells (Jolly et al., 1989).

Direct protein sequencing has established subunit c of mitochondrial ATP synthase (SCMAS) as the major component of the storage bodies (Palmer et al., 1989a). This was the first time that the storage material for an NCL was characterised, and SCMAS was later identified to be the main storage material for many other variants in humans and animals. Protein transcripts for SCMAS appeared to be expressed normally in the mitochondria prior to storage in the lysosomes, which suggests that NCL may occur due to a defect in turnover of protein inserted into the mitochondrial inner membrane (Hughes et al., 2001). The occurrence of mitochondrial dysfunction (Jolly et al., 2002b) was implied by variable ATP synthase activity measured in affected lambs with NCL (Jolly et al., 2001) and evidence of anomalies of mitochondrial ATP synthase regulation (Das et al., 1999).

The mode of inheritance of NCL in the SH sheep is autosomal recessive based on pedigree analysis. Linkage mapping using four microsatellite markers (BMS2635, BM1237, BMS528, and BMS2349) localised the gene causing NCL to ovine chromosome 7q13-15; with this region showing conserved synteny to human chromosome 15q21-23 (Broom et al., 1998; Broom and Zhou, 2001). This suggested that NCL in SH sheep and the CLN6 variant in humans (Sharp et al., 1999) represent mutations in orthologous genes. The *CLN6* gene contains 7 exons, encoding a predicted 311 amino acid transmembrane protein of unknown function (Tammen et al., 2006).

Previous studies did not identify a causative mutation in the coding sequence of the gene but linkage analysis using a silent mutation (c.822G>A) in the *CLN6* gene confirmed *CLN6* as the most likely candidate gene in the SH sheep (LOD score of 13.3 and Theta = 0.01) (Tammen et al., 2006). Identification of the silent mutation was useful as it allowed the development of an indirect DNA test for preclinical diagnosis of SH sheep and is now routinely used for diagnosis of NCL in the experimental flock (Tammen et al., 2006). The SH animals in this flock are configured so that all normal sheep are genotyped GG, heterozygotes AG and affected sheep AA. Prior to the indirect test, NCL diagnosis was established by histopathology of needle biopsies from brain taken under general anaesthesia at 10-12 weeks of age (Dickson et al., 1989), observation of clinical signs upon disease onset at approximately 10 months of age and typical NCL pathological findings during post-mortem.

The absence of any mutations in the coding region of the affected SH sheep incited a study to determine if *CLN6* mRNA expression or stability was affected in these sheep (Tammen et al., 2006). This study detected a reduction in *CLN6* expression in affected SH sheep to less than 1/3 of those of the control sheep tissues, with intermediate levels measured in the heterozygotes. A mutation in regulatory elements was postulated based on down-regulation of these *CLN6* transcripts (Tammen et al., 2006).

Attempts at therapeutic approaches have been carried out on the SH sheep. Hematopoietic cells transplant into SH fetuses was not beneficial (Westlake et al., 1995). The identification that neuroinflammation precedes neurodegeneration (Oswald et al., 2005) has led to a therapeutic study to investigate if anti-inflammatory drugs can have an effect on disease progression (Kay and Palmer, 2013). However, oral administration of the anti-inflammatory drug minocycline to affected lambs did not halt inflammation nor change the neuronal loss and clinical course of the disease (Kay and Palmer, 2013). Current research focuses on gene-therapy trials in SH sheep with a study by Linterman et al. (2011) providing the first evidence of lentiviral-mediated gene transfer to affected sheep brain via direct stereotactic injection.

1.2.2.2 Merino sheep

NCL in the Merino sheep in Australia was first documented from two commercial Merino flocks between 1996 and 1998, with affected sheep exhibiting mild behavioural changes of reduced herding instinct and startle response to visual and auditory stimuli at 8 to 12 months of age (Cook et al., 2002). After 12 months of age, clinical symptoms progressed to obvious changes in mental state (e.g. staring into space and circling), blindness, seizures, anorexia and recumbency developed, with death by 24 months of age (Tammen et al., 2001; Cook et al., 2002). An experimental research flock for the Merinos is currently maintained at the University of Sydney, Australia by Assoc. Prof. Imke Tammen. The Merinos have similar NCL clinical signs and pathology to the South Hampshires, although with a slightly earlier onset of disease (Graydon and Jolly, 1984; Mayhew et al., 1985; Jolly, 1995; Cook et al., 2002).

NCL in Merino sheep is inherited in an autosomal recessive mode of inheritance based on pedigree analysis. Based on the mapping of NCL in SH sheep, a homozygosity mapping approach localised the gene causing the disease in Merino sheep to the same chromosomal region (OAR 7q13-15) associated with NCL in South Hampshire sheep (Cook et al., 2002; Tammen et al., 2001). Tammen et al. (2006) identified a causative mutation in exon 2 of the ovine *CLN6* gene (c.184C>T) which is predicted to code for a major amino acid exchange (p.Arg62Cys) in the putative protein. Following this finding, a Restriction Fragment Length Polymorphism Analysis of PCR-Amplified Fragments (PCR-RFLP) was developed as a direct DNA test for diagnosis of NCL in the Merino sheep, which is beneficial for preclinical and prenatal investigations (Tammen et al., 2006).

1.2.2.3 Borderdale sheep

The New Zealand Borderdale sheep derive from initial crosses between Border Leicester and Corriedale sheep (Jolly et al., 2002a). NCL in Borderdales was initially observed over a period of 6 years in a flock of commercial sheep (Jolly et al., 2002a). Clinical signs for NCL in Borderdale sheep are to some degree different to those of affected Merino and SH sheep, with visual impairment as the first clinical sign noticed at approximately 15 months of age. Although visually impaired, the sheep had a normal gait but tended to walk in circles when prompted

(Jolly et al., 2002a). Research flocks have been established for the Borderdale (Jolly et al., 2002a), and are maintained by Prof. David Palmer at Lincoln University, New Zealand.

Pathological findings of sheep euthanised at 21 months of age revealed gross severe atrophy of the cerebral cortex in all affected animals (Jolly et al., 2002a). The neurons contained large numbers of round fluorescent emitting storage bodies. Subunit-c of mitochondrial ATP synthase was positively identified as the main storage material (Jolly et al., 2002a). Electron microscopy showed numerous round cytoplasmic inclusions surrounded by bilayer membranes in surviving neurons of the cerebral cortex (Jolly et al., 2002a). Three types of multi-lamellar ultrastructures were identified for the storage bodies: rectilinear, curvilinear and small circular profiles. NCL in Borderdale sheep exhibits some pathological differences from the Merino and SH (Jolly et al., 2002a). The Borderdale sheep appeared to have more severe brain atrophy and less intense autofluorescence of storage materials compared to the SH and Merino which, being both CLN6 variants, are caused by homologous genes (Jolly et al., 2002a). Dissimilar to the SH and Merinos, there was no report of photoreceptor cells damage or loss (Jolly et al., 2002a), thus it was assumed that the blindness could be triggered by the neurodegeneration of the visual cortex which is the part of the brain cerebral cortex associated with vision.

NCL in Borderdale sheep is inherited in an autosomal recessive mode, based on pedigree analysis. Absence of the disease causing mutation in the Merinos (c.184C>T), and more importantly, absence of linkage between the disease phenotype in Borderdale sheep to the polymorphism associated with NCL disease in the South Hampshires (c.822G>A) excluded *CLN6* as the locus responsible for NCL in the Borderdales (Tammen et al., 2006). Linkage analysis mapped NCL in Borderdale sheep to four microsatellite markers in the *CLN5* region: ILSTS056, OARHH41, BMS585 and BMS975 (Frugier et al., 2008). Sequencing of the ovine *CLN5* gene exploited the high degree of homology between ovine and bovine *CLN5* by using primers designed from the bovine sequence and ovine BAC DNA as template to generate ovine specific sequences (Frugier et al., 2008). The ovine *CLN5* gene contains 4 exons and encodes a predicted 361 amino acid protein with unknown role (Frugier et al., 2008). The disease causing mutation for NCL in the Borderdale sheep was identified in intron 3 of the ovine *CLN5* gene (c.571+1G>A) which leads to the splicing out of exon 3 and a shortened putative protein

unlikely to be functional or stable (Frugier et al., 2008). Whereas the CLN6 protein responsible for NCL in the South Hampshires and Merinos is an intracellular membrane protein predicted to reside in the endoplasmic reticulum, the CLN5 is a soluble lysosomal protein. (Frugier et al., 2008).

A recent gene therapy approach by Hughes et al. (2014) using a lentiviral vector in prenatal *CLN5*-deficient sheep neural cultures, showed that storage body accumulation is reversible *in vitro*. This supports the use of these cultures for testing of therapeutics prior to whole animal studies. An *in vivo* therapeutic trial in CLN5 variant NCL lambs is currently underway (Hughes et al., 2014).

1.2.2.4 White Swedish Landrace sheep

The earliest onset of NCL in sheep was observed in the White Swedish Landrace sheep. The first cases of this variant of NCL were reported in 17 newborn lambs from a flock of Swedish sheep maintained on an experimental farm in Northern Sweden (Järplid and Haltia, 1993). This novel form of congenital NCL (CONCL) was clinically characterised by extreme weakness and trembling with inability to rise and support their body after birth, however some lambs were able to support their head and survived for a few weeks when bottle-fed (Järplid and Haltia, 1993).

Gross pathology revealed that the brains of affected Swedish lambs were strikingly small (half the weight of brains of normal littermates), with severely reduced thickness of cerebral cortex (Tyynelä et al., 2000). Visceral tissues examined e.g. heart, kidney and liver appeared macroscopically unaffected (Tyynelä et al., 2000). Microscopically, there was marked loss of neurons and infiltration of macrophages in the cerebral cortex of the affected lambs (Tyynelä et al., 2000). Järplid and Haltia (1993) further observed generalised autofluorescent storage bodies in neurons in the cerebral cortex of affected lambs, with small amounts of storage material found in many extraneural cell types, but without evidence of further parenchymal damage.

The major component of the storage bodies was not subunit c of mitochondrial ATP synthase as evident in the South Hampshires and Merinos (Tyynelä et al., 1993). Immunohistochemical studies by Tyynelä et al. (2000) revealed that amounts of SAPs A and D, which are normal

lysosomal constituents, were greatly elevated in CONCL neurons. In contrast; the level of steady-state level of cathepsin D protein, attained when its rate of synthesis and its rate of degradation are equal, was markedly higher in the CONCL lambs than in controls. Thus, an enzymatically inactive, yet stable, and apparently normally processed cathepsin D was present in CONCL affected lambs (Tyynelä et al; 2000, 2001).

The mode of inheritance for NCL in the White Swedish Landrace sheep is autosomal recessive based on breeding information. Sequencing of RT-PCR Cathepsin D products derived from total RNA purified from brain samples in both control and CONCL sheep identified the causative mutation as a missense mutation in the *Cathepsin D* gene (c.934G>A). The mutation results in substitution of an asparagine for aspartate that corresponds to Asp295 of human cathepsin D and Asp215 of human pepsin (Tyynelä et al., 2000). This was the first reported disease arising from a naturally occuring *Cathepsin D* mutation (Tyynelä et al., 2000). Since then, NCL caused by *cathepsin D* mutations has been identified in humans with congenital (Siintola et al., 2006; Fritchie et al., 2009), late-infantile (Siintola et al., 2006) and juvenile onset (Steinfeld et al., 2006; Kousi et al., 2012); as well as dogs with congenital onset (Awano et al., 2006b).

1.2.2.5 Rambouillet sheep

The first cases of NCL in Rambouillet sheep were reported in Texas, USA, with two 8-month-old Rambouillet sheep presented with signs of visual deficits, but otherwise normal physically (Edwards et al., 1994). On further examination, the lambs were found to have decreased mentation, loss of herding instinct and slow response to auditory stimuli. Another 12 affected sheep in Texas, USA were also detected from 4 closely related flocks after observation of abnormal behaviour; abnormal head and ear carriages (head held low, drooping ears). As the disease progressed, the sheep eventually wandered aimlessly, frequently became isolated from the flock and were ‘star-gazing’ (Woods et al., 1994).

The pathology and neuronal lesions in these sheep were somewhat typical of those described in the South Hampshires and Merinos; including marked atrophy of the cerebrum and slight atrophy of the cerebellum, neurodegeneration of the brain cortex, retinal lesions, autofluorescent storage bodies in neurons of sections of the brain and spinal cord, and positive staining of these

storage materials with PAS, sudan black and luxol blue stains (Woods et al., 1993, 1994; Edwards et al., 1994). However, the lack of storage bodies in tissues outside of the central nervous system suggested a different biochemical defect in these sheep to that in other NCL sheep models (Edwards et al., 1994). The autosomal recessive mode of inheritance of NCL in the Rambouillet sheep was confirmed using controlled-breeding data (Edwards et al., 1994). There are no follow up studies of the disease in the Rambouillet sheep to date, thus no causative gene or mutation has been identified and no information about the ultrastructure or composition of storage bodies has been described.

1.2.3 NCL in species other than sheep

Apart from sheep, naturally occurring NCL have also been described in many other non-human animal species and comprehensively studied in other large animal models, namely, cattle and dogs. Amongst the small animal models, mice are used intensively for modelling of NCL, using both naturally occurring and genetically engineered mutant mouse strains with mutations in many of the identified NCL genes. A summary of naturally occurring NCL in non-laboratory animal models is shown in Table 1.3 and information on mouse models and other animal models is presented in Table 1.4.

1.2.3.1 Cattle

NCL have been described in 3 breeds of cattle, namely; Beefmaster (Read and Bridges, 1969), Australian Devon (Harper et al., 1988; Martinus et al., 1991; Jolly et al., 1992; Tammen et al., 2002) and Holstein Friesian (Hafner et al., 2005). In earlier studies, NCL in cattle was described as neuronal lipodystrophy (Read and Bridges, 1969), neurovisceral ceroid-lipofuscinosis (Harper et al., 1988) and bovine ceroid lipofuscinosis (Jolly et al., 1992).

Within the various breeds, affected animals are born normal and initially develop progressive blindness, followed by other signs with a general onset at about 9 months of age or older. Read and Bridges (1969) reported NCL occurrence in a herd of Beefmaster cattle and described a case of an 18-month old Beefmaster bull with early signs of nervous system disorder, blindness and

intermittent circling. These signs persisted for approximately 6 months, before the animal became comatose (interrupted by periodic clonic convulsion) prior to death. A single Holstein steer aged at 15-18 months old was reported to have a history of progressive blindness, but otherwise appeared to be afebrile, placid and well conditioned (Hafner et al., 2005). Ten Devon cattle were reported to show signs of progressive blindness at an average age of 14 months old; of these animals, six were examined in detail (Harper et al., 1988; Martinus et al., 1991). These animals collided with obstacles, and tended to walk or trot in a circle when disturbed, frequently with a mild head tilt (Harper et al., 1988). In the affected herd, the animals maintained an overall good condition before they eventually died from misadventure within 2 years of clinical onset (Harper et al., 1988). The clinical signs are similar in both the Beefmaster and Australian Devon breeds described above, with less detailed records for the Beefmaster cattle. In comparison, the single Holstein steer only displayed progressive blindness with absence of the other signs shown in the Beefmaster and Australian Devon cattle. However, other clinical signs might have developed if the animal had not been sent for slaughter.

During necropsy, macroscopic examination revealed a slight decrease in the size of brains of affected Devon cattle; with a varying extent of yellow discoloration present (Harper et al., 1988, Jolly et al., 1992). Furthermore, cerebral atrophy of the posterior half of the cerebral cortex was particularly severe in the occipital area. Another macroscopic change observed in the Devons was mild atrophy of the cerebellum (Jolly et al., 1992). The eyes of affected Devon exhibited severe retinal atrophy characterised by complete loss of the photoreceptor cell layers (Harper et al., 1988; Jolly et al., 1992). Degeneration of the ophthalmic layer such as the outer and inner retinal nuclear layer, with some loss of the ganglion cells was also observed. Retinal changes in the Holstein steer (Hafner et al., 2005) were similar to those described in Devons (Harper et al., 1988; Jolly et al., 1992). Read and Bridges (1969) did not report any macroscopic lesions in the affected Beefmaster cattle.

In affected Devon and Holstein cattle the cytoplasm of cerebral neurons contained the autofluorescent storage material characteristic of NCL. Storage material, which was also found in various organs of all 3 breeds of cattle, stained blue with Luxol fast blue (LFB), black with Sudan black (SB). Storage material in the Beefmaster did not stain pink with Periodic acid-Schiff

(PAS) but this stain was positive for both Devons and the single Holstein steer. Examination using the electron microscope identified that the ultrastructures of storage materials in affected Devons were membrane-bound curvilinear bodies (Harper et al., 1988); discrete, partially membrane bound and multilamellar in the Holstein (Hafner et al., 2005); and membrane bound, lamellar and identical to the curvilinear profile in the Beefmasters (Read and Bridges, 1969).

Surviving retinal cells in both the Devons and Holstein steer contained a moderate amount of autofluorescence storage material characteristic of NCL (Jolly et al., 1992; Harper et al., 1988, Hafner et al., 2005). Read and Bridges (1969) reported that neurons of the ganglion cell layer of the eyes contained similar storage material. It is not known whether the onset of clinical blindness was due to the retinopathy, the atrophy of the visual cortex, or to concurrent lesions at both sides (Harper et al., 1988). The comparative study of NCL demonstrated that the retinal degeneration was more severe in cattle than in South Hampshire sheep, with cattle showing a possible complete loss of the photoreceptor cells in terminal stage of disease (Jolly et al., 1992).

In Holstein cattle, Hafner et al. (2005) reported generalised deposition of cytoplasmic storage material, with presence of small amounts of autofluorescent substance found in the cortical tubules of the kidney and macrophages in the liver's fibrotic portal triads. Likewise, Harper et al. (1988) observed storage product in splenic and lymphoid histiocytes, renal tubular epithelium and hepatocytes of affected Devon cattle. Thus, in all three cattle breeds affected animals display pathologic findings characteristic for NCL. The limited data available for Holstein and Beefmaster cattle make it difficult to assess whether these breeds represent similar or different variants/forms of NCL. Using amino acid sequence and mass spectroscopy of isolated storage material, Martinus et al. (1991) identified that the storage bodies in affected Devons are composed of the hydrophobic protein subunit c of mitochondrial ATP synthase (SCMAS). In both Holstein and Beefmaster cattle the storage material was not further characterised, however, the histochemical and ultrastructural lamellar profiles suggest SCMAS, as compared to GRODs which are indicative of SAPs is more likely to be the storage material (Hafner et al., 2005; Read and Bridges, 1969).

Read and Bridges (1969), Harper et al. (1988) and Jolly et al. (1992) implied inheritance of NCL as an autosomal recessive trait, but the pedigree information available at that time was based on insufficient or incomplete breeding data. Years later, the autosomal recessive mode of inheritance in Devon cattle was substantiated by Tammen et al. (2002) using pedigree analysis of 1248 animals with 24 affected and 28 obligate carriers.

Tammen et al. (2002) listed potential NCL candidate genes in Devons as *CLN3, CLN5, CLN6* and *CLN8,* after excluding *CLN1, CLN2* and *Cathepsin D,* based on their respective clinical, pathological and biochemical characterisation. Three of the candidate genes were mapped using radiation hybrid mapping (Houweling et al., 2006a) and a homozygosity mapping approach was implemented (Tammen et al., 2002). The *CLN5* gene was selected as the most likely candidate gene for NCL in Devon cattle on the basis of phenotypic and pathologic similarities between affected cattle (Harper et al., 1988; Jolly et al., 1992) and CLN5 variant NCL affected humans (Savukoski et al., 1994)**.** A study (Houweling et al., 2006b) identified a single base duplication (c.662dupG) in bovine *CLN5* as the cause for NCL in Devon cattle. The mutation causes a frame-shift and premature termination (p.Arg221GlyfsX6) which is predicted to result in a severely truncated protein. However further studies will be required to know the exact function of *CLN5* in both humans and animals. Based on the clinical, pathological and recently identified molecular aspects of the disorder in Devon cattle, NCL in Devon cattle is considered as an animal model for human Finnish variant LINCL (fVLINCL; *CLN5*) (Houweling et al., 2005). It appears that the disease is not a problem for the wider Australian Devon population as a recent study estimated the allele frequency for the disease allele to be 0 or close to 0 (Okazaki et al., 2013).

1.2.3.2 Dogs

Canine NCL have been reported in at least 20 different breeds. In most cases, identification of NCL in a particular breed was originally made in the veterinary literature. These occurrences often remain as independent cases with no close examination or follow up studies (Mole et al., 2011).

Studies on NCL in dogs have significantly advanced the repertoire of genes that are known to cause NCL when mutated, provided mechanistic insights and enabled the testing of experimental therapies in a relatively large brain (Bond et al., 2013). As of 2014, NCL disease causing mutations have been identified in 8 canine genes (*CLN8*, *CLN5*, *CLN10/CTSD*, *CLN2/TPPI*, *CLN1/PPT1*, *ARSG*, *ATP13A2* and *CLN6*) (Table 1.3). DNA based tests for NCL diagnoses of affected animals and identification of carriers to reduce breeding of carriers dogs are available for some breeds, namely Tibetan terriers, Dachshunds, American bulldogs and English setters (http://www.caninegeneticdiseases.net/CL_site/mainCL.htm) and Border collies (http://www.bordercolliehealth.com/). Generally, an initial diagnosis of NCL is considered when dogs develop a progressive neurodegenerative disease with loss of vision and motor disturbances (ataxia, tremor) and behavioural abnormalities, such as aggression, fearfulness and compulsive activity (Jolly et al., 1994a). Affected dogs are usually euthanised due to a poor prognosis.

The following canine NCL models are listed according to their respective breed with a focus on those with assigned NCL genes (Table 1.3). NCLs in dogs are associated with autosomal recessive mode of inheritance based on pedigree information.

The Norwegian English Setter was the first canine model established for NCL research (Koppang, 1988). A research colony for this breed was maintained in Norway and has been studied extensively since the 1960s. Affected animals exhibit no symptoms from birth to 12 - 14 months of age, and then from about 12 months develop clinical signs of reduced vision and mental deterioration (Koppang, 1992). Convulsions are frequently observed within months of symptom onset and persist until death (Koppang, 1992). There is rare survival beyond 2 years of age (Koppang, 1992). Macroscopically, typical NCL brain changes of marked and moderate cerebral and cerebellum atrophies, respectively, with yellowish-brown discolouration was observed (Koppang, 1992). Microscopically, autofluorescent storage materials were found in the neurons of all regions of the central nervous system and almost every organ (Koppang, 1992). NCL in the English Setters was mapped to canine chromosome 37 (CFA37) orthologous to human chromosome 2 (HSA2) (Lingaas et al., 1998). Sequencing of the canine genome revealed that canine *CLN8* is located on canine chromosome 37 (CFA37) (Katz et al., 2005a). Sequencing analysis identified a c.491T>C missense mutation predicted to cause an amino acid change

(p.L164P) in the canine *CLN8* (Katz et al, 2005a). NCL in the English Setters is a model for human JNCL (Koppang, 1988). This research colony has since been discontinued as no funding support could be obtained for maintaining a small research colony. However, carrier dogs from the original research population and frozen semen is available for research purposes (Mole et al., 2011).

Border collies with NCL were first reported in 1988 (Taylor and Farrow, 1988). Clinical onset in these dogs were observed between 16 - 23 months of age, with progressing signs of behaviour changes including hyperactivity and aimless wondering, motor abnormalities, and visual deficits (Taylor and Farrow, 1992). Visual deficits or blindness often developed at approximately 21 months of age (Studdert and Mitten, 1991). Light and electron microscopic examination revealed neuronal degeneration with storage accumulation in neurons of the CNS, in ganglia of the peripheral nervous system, retina, and in several non-nervous tissues (Franks et al., 1999). Accumulation of SCMAS was reported (Jolly et al., 1994a) in the storage bodies. A combination of linkage analysis and comparative genomics identified the canine *CLN5* gene to be responsible for NCL in the Border collies (Melville et al., 2005). This gene is orthologous to the gene responsible for the human Finnish vLINCL (Melville et al., 2005). A nonsense point mutation (c.619C>T) that leads to a truncated protein (stop codon Q206X) was identified as the disease causing mutation in this breed (Melville et al., 2005). A recent paper demonstrated a high frequency of 8.1% of the mutant allele in dogs in Japan using novel rapid genotyping assays (Mizukami et al., 2011). No affected dogs are currently being maintained as research colonies.

NCL in American Bulldogs was described in a group of related dogs between 2001 and 2003 (Evans et al., 2005). Clinical onsets in these dogs were within the range of 1 – 3 years of age and consist mainly of motor abnormalities that included progressive ataxia and hypermetria in all four limbs (Evans et al., 2005). Other signs observed were conscious proprioception deficits and wide-based stance in all 4 limbs. Most of the affected dogs were euthanised by 4 – 6 years of age due to the severity of the symptoms. The only gross pathological change observed in the brain was that it had light brown hues (Evans et al., 2005). Storage bodies found in the brain, retina and liver exhibited golden-yellow autofluorescence characteristic of the NCL (Evans et al., 2005). Electron microscopy identified granular and lipoidal ultrastructures (Evans et al., 2005).

Sequencing analysis identified canine *CTSD/CLN10* as the causative gene for NCL in the American Bulldogs (Koike et al., 2000; Tyynelä et al., 2000). The mutation responsible for the disease in this breed is a G>A mutation which predicts the conversion of methionine-199 to an isoleucine (Awano et al., 2006b). This amino acid change results in a substantial reduction, but not a complete loss of, cathepsin D enzyme activity in the brain (Awano et al., 2006b). No affected dogs are currently being maintained as research colonies.

Isolated reports of NCL in Dachshunds have been published in the past (Cummings and de Lahunta, 1977; Vandevelde and Fatzer, 1980). Reports to be described here show two genetically distinct forms of NCL, which differ from previously reported Dachshund diseases in that these cases have a much earlier onset and are more rapidly progressive.

In longhaired Dachshunds, initial clinical signs were noted at 7 - 9 months of age and included vomiting, mental dullness and unresponsiveness to previously learned commands (Awano et al., 2006a). Signs developed in the coming months were progressive ataxia, visual deficits, generalised myoclonic seizures and circling. Affected dogs died at 12 months of age. Autofluorescent storage bodies characteristic of NCL were present in all examined regions of the central nervous system (Awano et al., 2006a). Ultrastructural analyses indicated that the storage body contents in all of the neural tissues consisted of curvilinear forms characteristic of those that accumulate in the human CLN2 variant (Awano et al, 2006a). Resequencing of the canine orthologue *TPP1* revealed a single nucleotide deletion (c.325delC) which was predicted to cause a frame shift after amino acid 107 (exon 4) with a stop codon at position 114 (Awano et al, 2006a). A research colony for the longhaired Dachshunds has been established as a model for the human CLN2 variant (Awano et al., 2006a; Katz et al., 2008).

A Miniature Dachshund with NCL was initially presented with kyphosis and stiffness of gait at 9 months of age (Sanders et al., 2010). Within months, the disease progressed to include uncontrolled rhythmic head movements, tremors and loss of coordination and vision. The dog was euthanised at 14 months of age due to the severity of the neurological signs (Sanders et al., 2010). Microscopic examination of fixed tissues showed massive accumulation of autofluorescent storage material characteristic of NCL in neurons of the retina, cerebellum, and

cerebral cortex (Sanders et al., 2010). Ultrastructural examination of the cerebral cortex and cerebellum indicated that the storage bodies consisted largely of uniform granular appearing material (Sanders et al., 2010), typical of the GROD described in some human NCLs, primarily classical INCL (Tyynelä et al., 2000). Resequencing of the *PPT1/CLN1* revealed a c.736_737insC mutation causing a frame shift which alters the predicted amino acid coding after Gly245, leading to a premature stop codon at position 276 in the altered protein (Sanders et al., 2010). As a result, this mutation was predicted to encode a polypeptide with an altered and truncated C-terminal end (Sanders et al., 2010). A research colony has been established at the University of Missouri (Sanders et al., 2011).

Australian Shepherds have potentially two genetically distinct forms of NCL. A report by O'Brien and Katz (2008) described 3 Australia Shepherd littermates with a history of progressive loss of vision and tremors that showed nervousness and a wide-based stance in the hind limbs and mild hypermetria upon examination. Magnetic resonance imaging (MRI) identified cerebral atrophy. There was a large accumulation of autofloresecent storage bodies characteristic of NCL found throughout the cells in the brain and central nervous system. Only paraffin-embedded tissues were available from these dogs and no attempt was made to identify the underlying mutation (O'Brien and Katz, 2008). Subsequently, tissues from an unrelated Australian Shepherd that was euthanised after exhibiting similar symptoms of vision loss and neurological signs at 19 months of age were taken for further analyses (Katz et al., 2011). Electron microscopy revealed large amounts of autofluorescent bodies in the cerebral cortex, cerebellum, and retina of the affected dog. Sequencing analysis identified the disease causing mutation as a c.829T>C transition in the *CLN6*, which constitutes a missense mutation producing a CGG arginine codon instead of the common TGG codon for tryptophan (Katz et al., 2011). The T to C transition results in a tryptophan to arginine amino acid change in the predicted protein sequence. Because of the relatively unknown function of the CLN6 protein, it is not possible to determine whether the mutation results in functional alterations.

Interestingly, two additional genes (*ARSG and ATP13A2*) proposed to cause NCL in dogs have so far not been associated with NCL in humans. These genes are discussed in the following paragraphs.

NCL in American Staffordshire Terriers has a relatively later onset than those described in the other breeds of dogs. Affected dogs have adult onset at 3 to 5 years of age, with signs of locomotor disabilities and ataxia (Abitbol et al., 2010). Through combined association, linkage, and haplotype analyses, NCL in this breed was mapped to a single region of canine chromosome 9. The disease causing mutation was postulated as 296G>A of the Arylsulfatase G (ARSG) gene (Abitbol et al., 2010). This missense change leads to a 75% decrease in sulfatase activity, providing a functional confirmation that the variant might be the NCL-causing mutation (Abitbol et al., 2010).

Another breed of dog with an assigned gene is the Tibetan terrier. NCL in this breed exhibits adult onset with visible signs at approximately 5 -7 years of age, with signs of anxiety, sensitivity to noise, ataxia, tremors and seizures (Katz et al., 2005b). The disease progresses slowly until euthanasia at 8 -10 years of age due to disease-related debility (Katz et al., 2005b). The most significant gross pathological change observed in affected dogs is a severe reduction of the cerebellum. Genome wide association analyses and mixed model analysis mapped NCL to dog chromosome 2 (CFA2) (Wöhlke et al., 2011). A mutation analysis identified the disease causing mutation as a single base pair deletion (c.1620delG) in the *ATP13A2* gene which causes skipping of exon 16 and results in a lack of 69 amino acids (Farias et al., 2011). Interestingly, other mutations in this gene have been described in Kufor-Rakeb syndrome (KRS) patients, a familial form of Parkinson disease (PARK9) (Ramirez et al., 2006). It has been suggested that KRS is a form of NCL (Farias et al., 2011; Mole et al., 2011) due to a single gene causing two distinctly rare neurodegenerative diseases. Analysis of KRS brain tissue will be needed to confirm this prediction (Farias et al., 2011).

At least 11 breeds of dogs with unassigned NCL genes (Polish Lowland Sheepdog, Miniature Schnauzer, Chihuahuas, Cocker spaniel, Dalmatian, Japanese Retriever, Welsh corgi, American Pit Bull Terrier, Labrador retriever, Golden retriever, Australian Cattle Dog) are summarised in Table 1.3.

1.2.3.3 Mice

Apart from sheep, cattle and dogs, mice are intensively used for modelling the biology, disease progression and therapeutic approaches of NCL. The existing NCL mouse models have mostly been developed through genetic modifications targeting a specific gene or gene locus to accurately mimic human genotypic defects (Cooper et al., 2006; Mole et al., 2011; Bond et al., 2013). Although mouse models of NCL are providing significant clues to the underlying biological basis of the disease, such studies in mice can be relatively slow, laborious and expensive to perform compared to the zebrafish (Cooper et al., 2006). A table which was adapted from a review paper by Bond et al. (2013) summarising the NCL genes for which a mouse model exists is shown here (Table 1.4).

1.2.3.4 Cats

Four independent case studies were documented for feline NCL in the 1990s, with 3 cases identified in domestic short hairs (DSH) (Nakayama et al., 1993; Bildfell et al., 1995; Weissenbock and Rossel, 1997) and one in a Siamese cat (Green and Little, 1974). The onset of disease varied between 7 months (Nakayama et al., 1993) and 22 months (Green and Little, 1974) of age. All affected cats were presented with various behavioural changes and neurological signs including uncoordinated gait (Nakayama et al., 1993; Weissenbock and Rossel, 1997), reduced vision and seizures (Weissenbock and Rossel, 1997), altered mentation and complete blindness, (Bildfell et al., 1995), as well as hyperesthesia, mania and photophobia (Green and Little, 1974). In all of these cases the progress of the disease was rapid and ultimately resulted in premature death or euthanasia.

Macroscopic examination showed similar pathological findings of atrophy and lesions of the brain, particularly of the cerebral hemispheres (Nakayama et al., 1993; Weissenbock and Rossel, 1997) with presence of yellowish brown discolouration in the brain and liver (Nakayama et al., 1993). Green and Little (1974) and Bildfell et al. (1995) did not report of any macroscopic lesion. All 4 studies documented multiple eosinophilic cytoplasmic storage materials of various sizes in the neurons throughout the brain and spinal cord. Positive staining of these materials with Luxol fast, Sudan black and PAS were characteristic of NCL. Generally, the storage

materials were shown to emit yellowish green autofluorescence under a fluorescent microscope, but fluorescence was not mentioned in the Siamese cats (Green and Little, 1974). Ultrastructurally, neuronal lysosomes contained multilamelar arrays consistent with curvilinear (Green and Little, 1974; Nakayama et al., 1993; Bildfell et al., 1995; Weissenbock and Rossel, 1997) or fingerprint structures (Weissenbock and Rossel, 1997). The accumulation of storage material was not detectable in extraneural tissues examined in two reports (Bildfell et al., 1995; Weissenbock and Rossel, 1997).

Only one study (domestic breed) reported diffuse retinal degeneration characterised by thinning of the photoreceptor layers and loss of ganglion cells (Bildfell et al., 1995). It was not stated if the retina was examined in the other NCL cats. Immunohistochemistry demonstrated the storage material in one cat to be composed of SCMAS (Weissenbock and Rossel, 1997). Histochemical and ultrastructural profiles of the storage material in the other cats suggest SCMAS (Green and Little, 1974; Nakayama et al., 1993; Bildfell et al., 1995). Pedigree information was not available for any of the cases but an autosomal recessive mode of inheritance was suggested (Green and Little, 1974).

A new report of feline NCL has emerged fifteen years after the last documented case (Kuwamura et al., 2009). The Japanese DSH cat showed almost identical clinical signs and GRODs ultrastructures to the Japanese DSH cat described earlier (Nakayama et al., 1993). The GROD ultrastructures and cytoplasmic vacuoles in some lymphocytes seen in the cat were postulated to be features of human NCL caused by defects of the *CLN3* gene (Furusawa et al., 2012). This hypothesis prompted analysis of the feline *CLN3* to identify a potential disease causing mutation (Furusawa et al., 2012). The direct DNA sequencing analysis revealed 17 variants in the *CLN3* whole gene and regulatory regions, from the cat with NCL, when compared with *CLN3* in the GenBank database and healthy control cats. None of these gene variants appeared to be disease causing, thus it was likely that *CLN3* was not the causative gene (Furusawa et al., 2012).

The most recent study on feline NCL has been documented in three DSH cats presented with a history of chronic progressive neurological clinical signs including partial (facial) and generalised seizures and variable visual deficits (blindness, absent menace reflex) (Chalkley et

al., 2013). Clinical onset varied between these cats (6 months, 1.5 years and between 1 and 2 years of age) and the progression from onset of disease to euthanasia was reported to be approximately 1 year. Macroscopic and microscopic examination at necropsy revealed typical NCL characteristics seen in earlier feline NCL cases; namely brain atrophy, neuronal loss and autofluorescence storage materials found throughout the brain and spinal cord. Although all 3 cats shared a similar distribution and pattern of lesions, the intensity and severity of the lesions varied from region to region throughout the CNS (Chalkley et al., 2013). Ultrastructural examination of the storage materials were slightly different between the cats; with GRODs, rectilinear (RL), fingerprint (FPR) and curvilinear profiles (CP) observed in one cat and only FPR and RL in the remaining two cats. Differences in clinical presentation and neurological lesions as well as storage material ultrastructures suggest that the 3 cats may have had different variants of NCL (Chalkley et al., 2013). NCL candidate gene mutational analysis was performed for only 1 cat due to limited resources. Of the 5 candidate genes evaluated (*CLN1/PPT1, CLN3, CLN5, CLN8, and CLN10/CTSD*), a number of sequence variants were identified when compared with the cat reference sequence, but none appeared likely to be disease causing (Chalkley et al., 2013).

1.2.3.5 Monkeys

A possible case of NCL was reported in a cynomolgus monkey (*Macaca fascicularis*) of 7 or 8 years of age (Jasty et al., 1984). The monkey was used as a control animal in an unrelated study and did not exhibit any clinical signs of illness during 3 months of observation prior to sacrifice. Light microscopy examination showed that various tissue cells including cells of the salivary gland, bile ducts, sweat glands, skeletal and smooth muscles, and neuronal cells of the CNS contained intracytoplasmic storage material. These granules stained weakly with Sudan black and PAS, and were autofluorescence under fluorescence microscopy (Jasty et al., 1984). Electron microscopy revealed that the granules were presented in various forms of profiles exclusive to the different types of tissue cells; including fingerprint structures in Purkinje's cells of the cerebellum and other neuronal cells of the brain, and lamella arranged in a vaguely distinctive concentric pattern in the ductal cells of the salivary glands (Jasty et al, 1984). This case study of possible NCL in the monkey was considered unique because the granules were present in abundance in almost every type of tissue (Jasty et al., 1984). Several findings in this case were

uncharacteristic of NCL such as the granules appearing bright pink when viewed with a light microscope and that they weakly stained with Sudan black and PAS. Apart from NCL, other potential causes for the unusual generalised distribution of the granules were old age and nutritional deficiency, however in the absence of supporting evidence the exact aetiology will be difficult to identify (Jasty et al., 1984).

1.2.3.6 Pigs

The only report of NCL in a pig was documented in a pig of 2 years of age that initially manifested minimal hind limb ataxia that progressed to tetraparesis with frequent stumbling and falling after 4 months (Cesta at al., 2006). Within a week the pig's condition deteriorated rapidly until it was unable to eat or drink independently and had a staggering gait (Cesta et al., 2006). Other signs noted were a slight head tilt and intermittent nystagmus. Upon presentation at the veterinary hospital, the pig was nonambulatory tetraparetic with a mild head tilt to the right with normal mentation and attitude. Neurological examination revealed nystagmus, rolling of the animal to the right when prompted to move, absence of postural reactions and conscious proprioception, and normal withdrawal reflexes (Cesta et al., 2006).

There was absence of gross lesions at necropsy. Microscopic lesions were restricted to the CNS with most neurons containing various storage materials (Cesta et al., 2006). The cerebral cortex contained the most storage materials. There was diffuse, mild to moderate neuronal loss in the cerebellar nuclei, cochlear nuclei and ventral horn of the spinal cord, as well as regions of the most severe neuronal degeneration, which includes the hippocampus, cerebral cortex and cerebellum (Cesta et al., 2006). Marked loss of cerebellar Purkinje cells was observed with the remaining cells containing small amounts of storage materials. These materials stained positive with PAS, luxol blue and Sudan black; and autofluorescence under fluorescence microscopy. The ultrastructures of the storage material were reported as multilamellar profiles of up to 5 layers consisting of curvilinear and GRODs; which appeared consistent with NCL associated profiles for SCMAS and SAPs (Cesta et al., 2006). Although there were no further studies (including genetic data) to verify NCL as the underlying disease in this pig, the morphological, histologic and ultrastructural properties of the storage materials were typical of NCL.

1.2.3.7 Horses

The only reported cases of NCL in horses have been documented in 3 distantly related Icelandic horse and Peruvian paso horses in Austria (Url et al., 2001). All three horses showed developmental retardation, slow movements and loss of appetite at six months of age. This was followed by an onset of neurological symptoms at about 1 year of age; with characteristics of torticollis, ataxia, head tilt and visual failure (observed in 1 horse). Gross pathological findings included slight flattening of the gyri in all 3 horses and yellow-brownish discoloration observed in the brain of 2 horses. Histological findings showed massive loss of neurons of all cortical layers of the cerebrum. Storage bodies of autofluorescence material were found in the majority of neurons in the cerebral cortex and with less frequency in neurons of other brain regions, spinal cord and in retinal cells (Url et al., 2001). Immunohistochemistry revealed the presence of large amounts of subunit c of mitochondrial ATP synthase and small amounts of SAPs in the storage bodies. Ultrastructurally, the storage bodies were present in various forms of lamellar profiles; fingerprints, curvilinear and rectilinear formations. The genetic basis for NCL in equines has yet to be determined, however, autosomal recessive inheritance was suggested by the existence of healthy parents and siblings that did not exhibit any neurological symptoms (Url et al., 2001).

1.2.3.8 Goats

NCL was reported in two closely related female Nubian goats that presented different onsets of clinical signs (Fiske and Storts, 1988). One showed progressive ataxia and hind quarter paresis at 18 months of age, and the other displayed neurological condition affecting the gait at 10 months, that developed to progressive ataxia and paresis in the months preceding death at 4 years of age (Fiske and Storts, 1988). Microscopic lesions in both goats showed brightly autofluoresecent storage material in all affected neurons. The physical appearance of the material ranged from very fine to distinct granules. The highest proportion of affected neurons was in areas of the central nervous system, specifically the ventral horns of all spinal cord segments and the brain stem (Fiske and Storts, 1988). Less severe lesions were found in the cerebellum and hypothalamus; and even milder ones in the cerebral cortex. Ultrastructurally, the storage material

within neurons consisted of spherical, concentrically-laminated membranous bodies, some of which had a fingerprint pattern (Fiske and Storts, 1988). The storage materials were characteristic of NCL: they showed autofluorescence with ultraviolet light and responded to other staining techniques e.g. PAS and Sudan black. However, no further studies or observation has been made into NCL in goats, and the biochemical properties of the storage materials have yet to be identified. The genetic basis for NCL in goats was considered to be autosomal recessive based on preliminary pedigree analysis (Fiske and Storts, 1988).

1.2.3.9 Birds

Reece and Macwhirter (1988) described a possible case of NCL in a peach-faced lovebird (*Agaponis roseicollis*) of nine months of age. The bird presented with signs of incoordination, loss of balance and intermittent convulsive seizures followed by death after a convulsion (Reece and Macwhirter, 1988). At necropsy, there were no gross lesions. Microscopically, the neurons in the brain stem and spinal cord showed yellow-gold granules in their cytoplasm. Many of the affected neurons were degenerate (Reece and Macwhirter, 1988). The yellow-gold granules showed autofluorescence under ultraviolet light and stained characteristically for NCL. The biochemical properties of the granules were not identified. However, a tumour found in close proximity to the brain stem in this bird could also have caused the neurological signs (Reece and Macwhirter, 1988).

1.2.3.10 Ducks

The second case report of NCL in an avian species and the first in a duck was reported in a 3 year old privately owned duck (Evans et al., 2012). The duck presented with a history of bilateral pododermatitis that progressed within 2 years, to bilateral nasal discharge, acute open-mouth breathing and nonambulatory paresis. Gross lesions were unremarkable. Storage materials were found within neurons and spinal cord, and stained positive for periodic acid-Schiff (PAS) and

luxol-fast blue (LFB) (Evans et al., 2012). NCL occurrence in this bird differed from the lovebird (Reece and Macwhirter, 1988) as the clinical presentation was chronic and lasted for 2 years after initial presentation. Different times of onset and rates of progression are not unusual for different variants of the disease and suggest different variants present between the avian species (Evans et al., 2012).

1.2.3.11 Ferrets

The only published report of NCL in a ferret was described in a 4 month old ferret with a history of ataxia at 3 months of age (Nibe et al., 2011). Additional signs observed in the ferret were weight loss, dysphagia, tremors, diarrhoea and dysstasia. Severe gross lesions found in the cerebrum were typical of NCL. Storage materials found in the neuronal cells exhibited fluorescence under fluorescence microscopy and stained positive for PAS, Sudan Black and LFB (Nibe et al., 2011). Electron microscopy examinations revealed the storage materials to be similar to GRODs ultrastructures. Results of immunohistochemical studies indicated SAPs as the major storage materials. The clinical and pathological features in the case study suggested that the ferret was affected with NCL (Nibe et al., 2011).

1.2.3.12 Other traditional animal models

In recent years, zebrafish (Cooper et al., 2006; Wang et al., 2007; Mahmood et al., 2013), nematode worm (*Caenorhabditis elegans*), fruitfly (*Drosophilia melanogaster)* and yeast (*Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*) have been genetically engineered to be used as models for NCL. Reviews of these NCL models are available (Bond et al., 2013 and Mole et al., 2011).

Briefly, zebrafish are traditionally used for mutational analysis, with studies establishing the functions of the proteins encoded by each gene that is mutated (Wang et al., 2007). This model

has certain distinct advantages including a short generation time, large brood size and increasingly well-understood anatomy, physiology and genetics (Cooper et al., 2006).

Studies in the fruitfly discovered that a large number of genes are homologues to genes in higher organisms including humans, and the resulting proteins have similar functions, thus studies in fruitfly can contribute to the identification of pathways underlying a variety of biological processes (Bond et al., 2013). The model is used to aid the study of NCL where it has contributed to the growing understanding of the molecular basis of the disease and may also provide an additional platform for the development of therapies.

Another model used for NCL is yeast, which has been found to be highly amenable to genetic manipulation and analysis, has short generation times and reproduces in a genetically stable manner. Yeast also recapitulates many fundamental aspects of mammalian cell biology (Bond et al., 2013) and is rapidly emerging as a powerful model for the cell biology of neurodegeneration (Khurana and Lindquist, 2010).

However, zebrafish, nematode worm, fruitfly and yeast have clear limitations (e.g. lack of a complex nervous system, short lifespan) when compared to mouse and large animal models for NCL (Bond et al., 2013, Mole et al., 2011).

A table listing naturally occurring and genetically engineered model organisms for human NCL genes and variants is shown below (Table 1.4).

Table 1.4: Summary of naturally occurring and genetically engineered model organisms with mutations in orthologues of human NCL genes and disease variants. Information has been adapted from Bond et al. (2013) and Cooper et al. (2006).

Model organism	Human NCL gene (italicized) or variants (not italicized)
Mouse	CLN1, CLN2, CLN3, CLN5, CLN6*, CLN8* , CLN10, Cathepsin D
Zebrafish	*CLN1*, *CLN2*, *CLN3*, *CLN4*, *CLN5*, *CLN6*, *CLN7*, *CLN8*, *CLN10* and *CLN11*
Drosophilia	*CLN1*, *CLN3* and *CLN10*
Yeast	*CLN1*, *CLN3* and *CLN10*

*naturally occurring NCL variant

Ultimately, NCL research in animal models has increased the understanding of NCL in humans. Discovery of the subunit c of mitochondrial ATP synthase (SCMAS) as one of the major storage body components was initially made in the South Hampshire sheep (Palmer et al., 1989a). Subsequent studies showed that SCMAS is also stored in brains of patients with late-infantile, juvenile and adult onset (Kufs) forms of NCL (Hall et al., 1991). The fact that mutations in *Cathepsin D* can cause NCL disease was originally discovered in Swedish landrace sheep (Tyynelä et al.. 2000), and subsequently found to be also involved in human patients (Steinfeld et al., 2006) and dogs (Awano et al., 2006b) with the CLN10 form of NCL. A study of a mouse model for human JNCL documented time course changes in retinal tissues (at birth, 3 months and 1 year), which resulted in valuable understanding of the mechanism for visual deficits (Weimer et al., 2006). An insight into preclinical course of the NCL disease in sheep identified early changes in the brain prenatally (Kay et al., 2006) and showed neuroinflammation preceded neurodegeneration in disease progression (Oswald et al., 2005). Recent studies in animal models have identified genes that are expected to cause NCL disease but for which so far no human cases have been identified (Abitbol et al., 2010; Farias et al., 2011 and Wöhlke et al., 2011).

1.3 Research objectives

South Hampshire sheep with NCL are recognised as one of the best characterised animal models for NCL in humans. Despite the identification of *CLN6* as a strong candidate gene for this disease in South Hampshire sheep a disease causing mutation has not been identified. A mutation affecting gene regulation has been proposed and lack of ovine genomic sequence information for the identified candidate gene region was identified as a major hindrance to detect such a mutation.

The main objective of the research presented in this thesis is the identification and characterisation of the mutation responsible for NCL in the South Hampshire sheep.

Generation of genomic sequence for the area of interest in normal sheep followed by mutation screening in affected sheep are proposed to achieve the research objectives. Due to changes in sequencing technology during this research the methods used to achieve these outcomes evolved.

Additional research objectives are the development of a diagnostic test if the South Hampshire NCL mutation can be identified.

1.4 Gene regulation

1.4.1 Introduction

Gene regulation relates to the regulation of the amount, location and timing of appearance of the functional product of a gene. The regulation process follows the fundamental dogma of molecular biology, where DNA is transcribed into RNA which in turn, translates into protein. Any step of gene expression may be modulated, from the DNA to RNA transcription step, to post-translational modifications of a protein.

Linkage analysis identified ovine *CLN6* as the candidate gene for NCL in South Hampshire sheep. While sequencing of the *CLN6* coding sequence in NCL affected SH sheep did not identify any disease causing mutation, quantitative PCR revealed reduced *CLN6* messenger RNA (mRNA) (Tammen et al., 2006). Therefore mutations in the non-coding regions within or flanking this gene, with an effect on the level of *CLN6* mRNA were postulated to cause NCL disease in these sheep (Tammen et al., 2006). Changes to regulatory sequences from these genome regions potentially affect the expression pattern or expression level of the gene they normally regulate (Loots, 2008). The following section briefly summarises information on regulation at transcription and post-transcription levels.

1.4.2 Transcriptional regulation

Gene transcription is complex in eukaryotic cells, with distinct classes of genes transcribed from DNA into RNA by multiple different RNA polymerase enzymes that interact with a variety of additional proteins to initiate transcription (Cooper, 2000). This sophisticated regulation of gene expression directs the activities of the many different cell types of multicellular organisms (Cooper, 2000). Transcriptional control of gene expression is achieved through 3 different mechanisms; chromatin structural regulation prior to onset of transcription, interaction of DNA sequence elements predominantly upstream to the gene transcription start site (TSS), and binding of the DNA elements to transcription factors that either activate or repress transcription (Latchman, 2005).

1.4.2.1 Chromatin structure

Chromatin is the complex of DNA and proteins found within the nucleus of a cell. Alteration in the chromatin structure of a gene is one of the three mechanisms of transcriptional control of gene expression. Tightly dense solenoid structured DNA is incapable of being transcribed, whereas open/loosely structured chromatin and nucleosome-free/ structurally altered nucleosomes are more accessible for binding of regulatory proteins to occur, which in turn initiates transcription (Latchman, 2005).

1.4.2.2 DNA sequence elements

The actual onset of transcription takes place through interaction of transcription factors (TF) with specific DNA sequence elements called promoters, which are adjacent to or at a distance from the target gene transcription start site (TSS) (Latchman, 2005). These promoters coordinate expression of protein coding genes and are recognised by the presence of known promoter elements and their consensus sequences, such as Cpg islands which are stretches of DNA in which the frequency of the CG sequence is higher than other regions (Carninci et al., 2006) and

CAAT boxes which are distinct pattern of nucleotides with a GGCCAATCT consensus sequence (Zhu et al., 2012).

Eukaryotic promoters are more diverse and complex when compared to promoters in prokaryotes and studies have shown that more than 10 different classes exist (Gagniuc and Ionescu-Tirgoviste, 2012). Alteration of these promoters effectively modifies transcription and potentially contributes to disease (Zhao et al., 2016). Promoter deacetylation has been observed to cause dysregulation in experimental mouse models of Huntington disease (Guiretti et al., 2016). Similar to NCL, Huntington disease is a fatal neurodegenerative condition.

Enhancers are elements that can activate transcription of their target genes in a tissue-specific manner, independent of distance (ranging from several to hundreds, in rare cases even thousands, of kilobases) and orientation (Calo and Wysocka, 2013). Enhancers have long been thought to bind transcription factors and be primarily active at the DNA level, however recent work has identified long non-coding RNAs (ncRNA) transcribed from active enhancers as important players in enhancer activity and function (Vucicevic et al., 2015). In contrast, silencers are elements that have an inhibitory effect on gene expression levels (Latchman, 2005). Silencers correlate with RNA production and comprise of both long and short RNA molecules; namely antigene RNAs (agRNAs) and microRNAs (miRNAs) (Kolovos et al., 2012). MicroRNAs (miRNAs) are an average of 22 nucleotides in length, and are among the shortest functional eukaryotic RNAs (Ameres and Zamore, 2013). These miRNAs repress most of the genes they regulate by just a small amount, yet they can potentially target hundreds of genes with different biochemical and biological functions (Ameres and Zamore, 2013). Because genes can have many regulatory elements, it is difficult to determine the relative effect of each element in influencing gene expression levels (Rajagopal et al., 2016).

Insulators act as boundary elements that prevent action of irrelevant enhancers and silencers (Kolovos et al., 2012). Insulators have been suggested to evolve from a class of promoters binding a specific subset of TF (Raab and Kamakaka, 2010), with many marked by DNase hypersensitivity and presence of bound RNA Polymerase II (RNAPII) (Kolovos et al., 2012).

1.4.2.3 Transcription factors (TF)

DNA sequence elements affect the rate of gene transcription by binding to regulatory proteins known as transcription factors (TF) (Latchman, 2005). Following binding of a TF to a particular DNA element (i.e. a transcription factor binding site (TFBS)), the TF can recruit the RNA polymerase and thus activate transcription or repress transcription. Transcription factors can be recognised based on their respective TFBS consensus sequence. The TRANSFAC 7.0/2005 database (http://www.gene-regulation.com/pub/databases.html; Wingender et al., 1996) is a public database that represents the largest repository for experimentally derived TFBS and contains data on transcription factors, their experimentally-proven binding sites and regulated genes. Although no *CLN6* TFBS are documented in TRANSFAC to date, predicted TFBS for *CLN6* are listed in the following link: http://www.genecards.org/cgi-bin/carddisp.pl?gene=CLN6&search=CLN6.

1.4.3 Methods to identify regulatory elements

Computationally predicting TFBSs that are functionally significant is a great challenge because they are short (6 – 12 bp), highly degenerate sequence motifs that occur very frequently in a genome, thus generate predictions with high rates of false positives when used in whole-genome analysis (Loots, 2008). It is not yet known how many different TFs need to synergistically cooperate to initiate expression. It has been suggested that a typical enhancer contains a minimum of 10 TFBSs for at least 3 different TFs (Levine and Tjian, 2003). What makes transcriptional genomics in vertebrates highly intricate stems from two recent observations: firstly, all regulatory elements associated with a transcript can be scattered over great distances that can reach megabases (Mb) in length (Nobrega et al., 2003; Sagai et al., 2005), and secondly, some regulatory elements are capable of controlling multiple transcripts, skip intercalating genes, or regulate one transcript while being positioned within a different transcript (Loots et al., 2000; Zuniga et al., 2004).

1.4.3.1 Computational method for predicting regulatory elements

The majority of available computational tools for predicting regulatory elements are based on aligning orthologous sequences and detecting TFBS. Loots (2008) discussed three different approaches that use pattern recognition. These approaches are, prediction using a library of known motifs (rVISTA; Loots et al., 2002) or conserved sequence blocks (FootPrinter; Blanchette and Tompa, 2002), identification of TFBS clusters (Berman et al., 2002) and prediction of conserved sequences based on specific regulatory properties shared by multiple functionally related sequences from the same organism (CRÈME; Sharan et al., 2004).

It has been shown that combination of pattern recognition with comparative sequence analysis dramatically reduces the number of false positives (Loots, 2008). rVISTA (Loots et al., 2002) and Consite (Sandelin et al., 2004) are programs that combine TFBS motif searches and cross-species sequence analysis. Considering that the ovine genome sequence was not available when this study commenced in 2006, flanking genome sequences from other species were proposed to be utilised to predict regulatory elements for the ovine *CLN6*.

1.4.3.2 Functional analyses for validating computationally predicted elements

Functional analyses validate computationally predicted TFBS and reveal whether a given binding site has the effect of activating or repressing transcription (Whitfield et al., 2012). The reporter gene or construct, which is a gene or construct attached to a regulatory sequence of another gene to indicate gene expression, is used in most of these analyses. A review of functional analyses approaches has been described by Loots (2008). These analyses include using transient transgenic mice embryos injected with a reporter construct and then examined for expression, and mutating candidate regulatory elements in engineered mice using either random mutagenesis or targeted knockout (KO) mice to detect loss-of-function outcome.

The NCLs are classified as lysosomal storage diseases. Many genes coding for lysosomal proteins share a sequence motif in their promotor region that acts as a binding site for transcription factor EB (TFEB, Sardiello et al., 2009). TFEB has been proposed to be a master regulator of these lysosomal genes and this gene network has been named CLEAR (Coordinated Lysosomal Expression and Regulation). The CLEAR gene network is proposed to regulate lysosomal biogenesis and function (Palmieri et al., 2011). There is a possible association between the CLEAR network and mechanisms causing lyosomal dysfunctions in lysosomal storage diseases (Settembre et al., 2013) and members of the CLEAR network represent potential therapeutic targets for lysosomal storage diseases (Palmieri et al., 2011). Several NCL genes (i.e. *PPT1* (*CLN1*), *TPP1* (*CLN2*), *CLCN7*, *SGSH* and *CLN3)* have been proposed to be members of the CLEAR network based on the fact that they share the common TFEB binding site (Palmieri et al., 2011). The *CLN6* gene is not known to be a member of this gene network.

Computational programs like rVISTA (Loots et al., 2002) and Consite (Sandelin et al., 2004) can be used to identify the presence of known transcription factor binding sites (e.g. the motif shared by the genes in the CLEAR network) in the promoter region of *CLN6*. Mutations affecting these binding sites are likely to affect the regulation of *CLN6* and could be disease causing.

Another more recently developed method to identify how transcription factors and other chromatin-associated proteins influence phenotype-affecting mechanisms is ChIP-sequencing, also known as ChIP-seq (Johnson et al., 2007). This is a method that uses chromatin immunoprecipitation (ChIP) and massively parallel DNA sequencing to identify the binding sites of DNA-associated proteins. It can be used to map global binding sites precisely for any protein of interest and could be used to identify if TFEB and MITF would play a role in *CLN6* regulation (Palmieri et al., 2011).

1.4.4 Post-transcriptional regulation

1.4.4.1 Nonsense-mediated mRNA decay (NMD)

The nonsense-mediated mRNA decay **(**NMD) surveillance pathway occurs post-translationally and functions in reducing errors in gene expression by eliminating aberrant mRNAs that encode incomplete polypeptides (Baker and Parker, 2004). These mRNAs are transcribed from alleles carrying nonsense mutations which contain premature translation-termination codons (PTCs) (Brogna and Wen, 2009). Considering that sequencing of the ovine *CLN6* coding sequence did not report of such mutations, this process is not likely to be of concern in relation to NCL in SH sheep.

1.4.4.2 Alternative splicing

In some genes, a process called alternative splicing occurs where a single gene is transcribed differently to yield different functional messenger RNAs in different tissues (Latchman, 2005). In many cases these RNAs are translated to yield different protein products. Alternative splicing has been shown to occur in the ovine *CLN6* gene with transcripts lacking exon 5 generating a shorter product than that of the gene with retained exon 5; however this occurrence was not associated with the disease (Tammen et al., 2006). The same study revealed no other splicing variants; however, the existence of other spicing variants was not excluded. Evidence of alternative splicing affecting exon 1 in one of the haplotypes encoding human *CLN6*, was proposed however the case was not further investigated (Dr. S. Mole pers. comm.). Cases of alternative splicing have also been reported in human *CLN2* (Kohan et al., 2013) and ovine *CLN3* (Oswald et al., 1999).

1.5 Sheep genetic resources

Any research that aims to identify the disease casing mutation for NCL in South Hampshire sheep would be greatly facilitated by access to genetic tools and especially a well annotated and complete ovine reference genome. Development of genetic resources for sheep in general, and especially efforts in relation to an ovine genome assembly have been lacking behind efforts in other animal species that are of interest as companion animals or livestock (Archibald et al., 2010).

At the beginning of this research in 2006 only limited ovine genetic resources were available (Table 1.5). The existence of the ovine linkage maps had allowed the mapping of NCL in South Hampshire sheep to OAR7 (Broom et al., 1998). Comparative genetics information (e.g. based on conserved synteny between cattle and sheep, sequence homologies between cattle and sheep genes, fluorescence in situ hybridization FISH (Langer-Safer et al., 1982) and Oxford Grids http://oxgrid.angis.org.au/ facilitated early research in sheep genetics in general and in relation to NCL research (Broom et al., 1998; Tammen et al., 2001).

As the research described here progressed, additional resources were developed that were largely driven by the International Sheep Genomics Consortium (ISGC; http://www.sheephapmap.org/). In particular, the continuous advancement of the ovine reference genome (Archibald et al., 2010) (Figure 1.1, Table 1.6) proved to be very beneficial. The construction of the sheep reference genome by ISGC and its collaborators has been developed with sequence data generated using both Sanger and next generation sequencing (NGS) platforms. The reference genome is based on sheep specific sequence contigs built on the same structures used by the humans and cattle genome framework. A flowchart for the sheep genome assembly is shown in Figure 1.1.

Table 1.5: Genetic and genomic resources for sheep (modified from Archibald et al., 2010).

Resource	Year	Description	Reference
Linkage Map	1995, 1998, 2000, 2001	Development of various sheep linkage maps	Tabet-Aoul et al., 2000; Maddox et al., 2001; de Gortari et al., 1998; Crawford et al., 1995 http://rubens.its.unimelb.edu.au/~jillm/jill.htm). Cockett, 2003
CHORI-243 BAC library	1998	CHORI-243 Ovine BAC library. End sequencing of the library formed the basis of the virtual sheep genome.	Osoegawa et al., 1998.
Ovine BAC library	1999	A sheep BAC library of over three genome equivalents	Vaiman et al., 1999
INRA 1200-rad RH Panel	2007	RH panel used for the assignment of sequence-tagged sites.	Laurent et al., 2007
Ovine SNP50 BeadChip	2009	Illumina Infinium-based platform for genotyping 50, 000 SNPs distributed across the sheep genome	http://www.illumina.com/products/ovinesnp50_dna_analysis_kit.ilmn

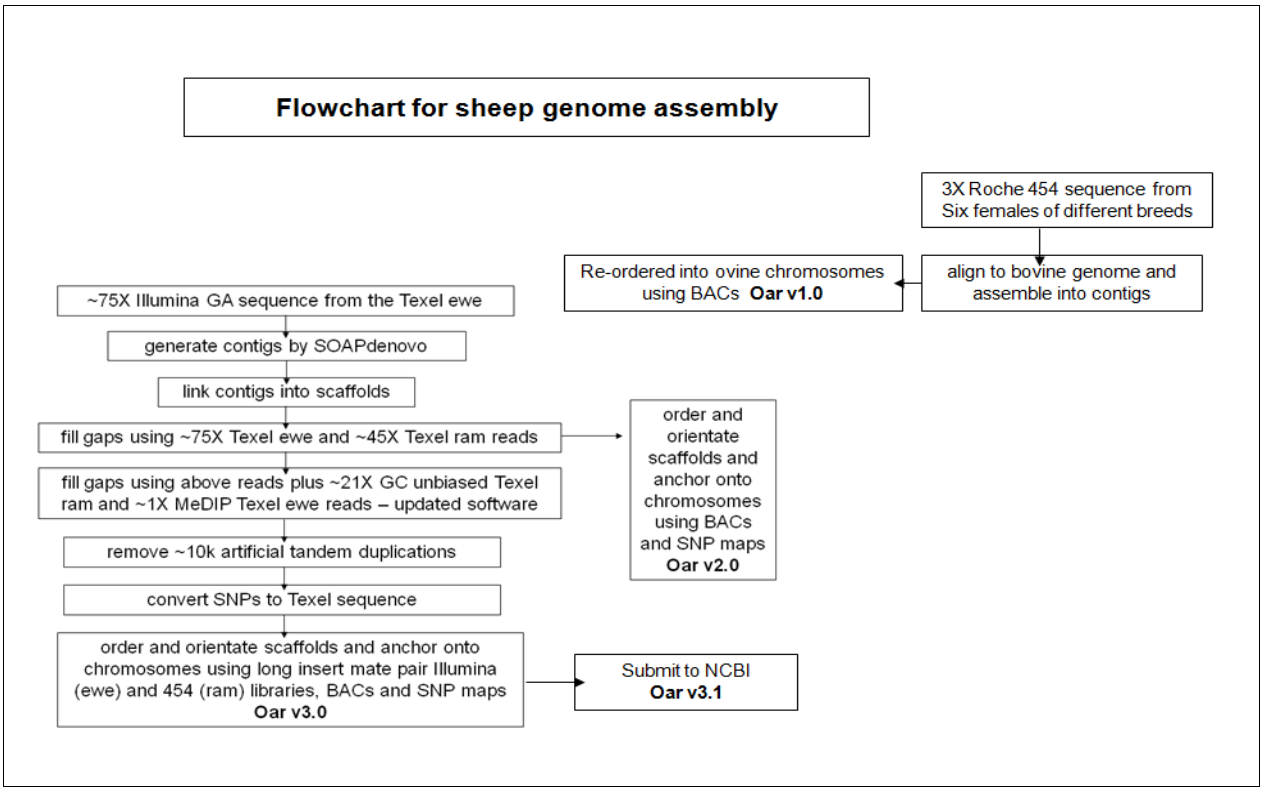


Figure 1.1: Flowchart for progress of the sheep genome assembly taken from http://www.livestockgenomics.csiro.au/sheep/oar3.1.php.

Table 1.6. Timeline for progress of the sheep genome assembly.

Sheep genome version	Description	Released year	Reference
Virtual sheep genome	Comparative analysis of BAC end sequences to generate a virtual genome assembly	2007	Dalrymple et al., 2007; http://www.livestockgenomics.csiro.au/vsheep/
Oar v1.0	First sheep genome assembly based on ovine contigs built onto the bovine sequence framework	2009	Archibald et al. (2010) Animal Genetics 41: 449. 453; http://www.livestockgenomics.csiro.au/sheep/oar1.0.php
Oar v2.0	Working draft release that contains both known and unknown errors and discrepancies	2011	Archibald et al. (2010) Animal Genetics 41: 449 . 453; http://www.livestockgenomics.csiro.au/sheep/oar2.0.php
Oar v3.0	Improved version of the Oar v2.0 and Oar v3.0 formed the basis of ISGC publications on the sheep genome	2011	Archibald et al. (2010) Animal Genetics 41: 449 .453; http://www.livestockgenomics.csiro.au/sheep/oar3.0.php
Oar v3.1	Sheep reference genome	2012	Archibald et al. (2010) Animal Genetics 41: 449 . 453; Jiang et al., 2014 http://www.livestockgenomics.csiro.au/sheep/oar3.1.php
Oar v4.0	Expected to be released mid to late 2014	2014	

1.6 Sequencing

Sequencing is the process used to determine the order of DNA nucleotides within a DNA molecule. Traditional Sanger sequencing historically was the most commonly used method for DNA sequencing, but more recently, a second generation of sequencing technologies, also known as next generation sequencing (NGS) has been developed that allows high throughput sequencing and massively parallel analysis. This technology has revolutionised sequencing. The Sanger and NGS technologies vary in many aspects including sequencing mechanism, read length (the actual number of continuous sequenced bases) and application. The analysis of the huge data sets obtained from the NGS platforms also brings unique challenges.

1.6.1 Sanger sequencing

The Sanger method of sequencing is based on the chain-terminating method (Sanger and Coulson, 1975, 1977). This method is based on the DNA polymerase-dependent synthesis of a complementary DNA strand in the presence of natural 2´-deoxynucleotides (dNTPs) and 2,3′ -

dideoxynucleotides (ddNTPs) that serve as non-reversible synthesis terminators (Sanger et al., 1977). Following the sequencing reaction, the products are separated by size using polyacrylamide gel electrophoresis and visualised via radioactive or fluorescence labels or silver staining. Over time, various improvements have been made in Sanger sequencing which contributed to the current modern Sanger sequencers which are faster and more accurate than their predecessors (Morozova and Marra, 2008; Carrilho, 2000). Modern Sanger sequencers are known for their low error rate and long read length of up to approximately 1,000 bp (Morozova and Marra, 2008; Zhang et al., 2011). Sanger sequencing was the main method for completion of the human genome project in 2001 (Collins et al., 2003) and due to the above mentioned improvements has generated an exponential increase of sequence information deposited into Genbank, a database containing an annotated collection of publicly available DNA sequences (Benson et al., 2005).

1.6.2 Next generation sequencing (NGS)

NGS has been reviewed in an abundance of papers (Chan, 2005; Mardis, 2008, 2013; Morozova and Marra, 2008; Liu et al., 2012). These papers have described the three main platforms for NGS which were available during the time of this project: Roche 454 GS FLX, Illumina/Solexa genome analyzer and Applied Biosystems (ABI) SOLiD system. All of these systems are rapidly evolving and have seen improvements in regards of key characters over the last decade. Apart from standard sequencing applications such as genome sequencing and resequencing, these technologies have also been applied in a variety of contexts including transcriptome analysis, discovery of non-coding RNAs and providing insights into transcription factor binding sites (Mardis, 2008). Generally, the read length for NGS is much shorter than that attained by Sanger sequencing (Zhang et al., 2011). Due to their relatively short reads, sufficient coverage (number of short reads that overlap each other within a specific genome) is very important for accurate assembly (Zhang et al., 2011).

The following review discusses three commercial NGS technologies available for the present study, with a summary of key features in comparison to Sanger sequencing presented in Table

1.7. It has been announced that the Roche 454 sequencing platform will be discontinued in 2016.

Table 1.7: Summary of key features of Sanger sequencing and three commonly used NGS sequencing platforms. Information is adapted from a review paper by Morozova and Marra, 2008; Liu et al., 2012 and Frey et al., 2014.

Sequencer	**454 GS FLX**	**Illumina/Solexa**	**SOLiDv4**	**Sanger 3730xl**
Sequencing mechanism	Pyrosequencing	Sequencing by synthesis with reversible terminators	Sequencing by ligation	Dideoxy chain termination
Average read length (bp)	Up to 1, 000 bp	2X 100 bp reads (PE)*	35-50	900
Company name and website	Roche applied science (https://lifescience.roche.com/shop/home)	Illumina (http://www.illumina.com)	Applied biosystems (http://www.lifetechnologies.com/au/en/home/brands/applied-biosystems.html)	Applied biosystems (http://www.lifetechnologies.com/au/en/home/brands/applied-biosystems.html)
Advantage	Read length, fast compared to Sanger	Capable of determining homopolymer length (compared to the 454 platform), cheaper than other NGS platforms at $1,000 per genome	Accuracy in determining bases, widely used for detecting genetic variation	High quality, long read length
Disadvantage	Prone to homopolymer length inaccuracies, relative high cost	Short read length, more challenging for assembly due to the short read length	Short read length, more challenging for assembly due to the short read length	High cost, low throughput
Resequencing	No	Yes	Yes	No
De novo	Yes	Yes	Yes	No
High GC sample	Yes	Yes	Yes	Yes
Large genome	Yes	Yes	No (short reads make assembly very difficult)	No (too expensive)
Mutation detection	Yes	Yes	Yes	Yes

*PE: pair-end reads

1.6.2.1 454 GS FLX sequencer

The 454 GS FLX sequencer, which uses pyrosequencing technology, was the first commercial NGS sequencer introduced into the market (2004) (Mardis, 2008). Its pyrosequencing technology relies upon enzyme cascades and CCD luminescence detection capabilities; where each nucleotide incorporated by the DNA polymerase results in the release of inorganic pyrophosphate that emits luminescence light (Ronaghi et al., 1998). Throughout the various sequencer releases, initial read lengths of 100 bp increased to 600 bp (Zhang et al., 2011) and have thus approached the lower end of read lengths obtained from traditional Sanger sequencing. The 454 reads are of sufficient length for *de novo* genome assembly (Mardis, 2008). This technology allows for hundreds of thousands of pyrosequencing reactions to be carried out in parallel, thus massively increasing sequencing throughput (Margulies et al., 2005). The 454 sequencing efficacy has been demonstrated by resequencing of the human genome at a fraction of the cost and time (Wheeler et al., 2008), compared to the initial sequencing using the Sanger method (Lander et al., 2001; Venter et al., 2001). However, since the luminescence intensity relies on the number of bases incorporated this technology is prone to errors resulting from incorrect estimation of length for stretches of sequence with an identical base (homopolymer) (Huse et al., 2007; Balzer et al., 2011; Gilles et al., 2011). Determining precise homopolymer length is particularly vital for detection of insertion-deletions (indels). Cost-wise, the 454 platform is more expensive compared to the Illumina and SOLiD systems.

1.6.2.2 Illumina/ Solexa genome analyser

The Illumina genome analyser was the second platform to reach the market (2006) and is currently the most widely used system (Zhang et al., 2011). The Illumina platform utilises a sequencing by synthesis approach in which all four nucleotides are added simultaneously into oligo-primed cluster fragments in flow-cell channels along with the DNA polymerase (Mardis, 2008). At present, the analyser is capable of producing single reads of 2X 100 bp (pair-end reads) (Zhang et al., 2011). The Illumina approach enables sequencing in a massively parallel

method and has been found to be more effective at sequencing homopolymer stretches than the 454 platform (Morozova and Marra, 2008). The disadvantage of this technology is the shorter length of sequence reads, which results in difficulties in resolving short sequence repeats and creates challenges for *de novo* genome assembly (Mardis, 2013).

1.6.2.3 ABI SOLiD system

The ABI SOLiD system uses a unique sequencing by ligation approach in which it uses an emulsion PCR with small magnetic beads to amplify the DNA fragments for parallel sequencing (Zhang et al., 2011). The platform uses a two-base encoding scheme which enables the distinction between a sequencing error and a sequence polymorphism (Morozova and Marra, 2008). In this scheme, an error will be detected in only one particular ligation reaction, whereas a polymorphism would be detected in both. The SOLiD platform has the highest accuracy especially when the coverage is more than 30X (Liu et al., 2012). The high accuracy makes it widely used in detecting genetic variations such as indels and copy number variation (CNV) (Liu et al., 2012). A read length of only 50 bp makes it the NGS platform with the shortest read length (Zhang et al., 2011). Like the Illumina platform, the short sequence reads make the determination of short sequence repeats and genome assembly difficult.

CHAPTER 2: GENERAL MATERIALS AND METHODS

2.1 Research sheep

2.1.1 South Hampshire and Coopworth sheep

The New Zealand South Hampshire (SH) and Coopworth (CPW) sheep used in studies described in consequent chapters were part of the unique South Hampshire CLN6 variant research flock currently maintained by Prof. David Palmer at Lincoln University, Christchurch, New Zealand. The flock is maintained under standard New Zealand pastoral conditions on a 250 hectare research farm at Lincoln University with management and husbandry performed according to NIH guidelines and the New Zealand Animal Welfare Act (1999).

NCL in South Hampshire sheep was first described by Jolly and West (1976) in two affected rams. An experimental flock was established (Jolly et al., 1980) and maintained by crossing homozygous affected rams with heterozygous carrier ewes resulting in 50% affected and 50% carrier offspring each year. The flock has undergone two outbreeding programs in the past 30 years with the introduction of heterozygous ewes attained from crosses between normal Friesian/ Finn and Coopworth ewes and affected South Hampshire rams to increase fecundity, and improve the health and reproductive performance of the sheep flock. The first outcross occurred early in the establishment of the research flock and included the introduction of Friesian and Finn crossbred ewes, whereas the second outcross occurred in 2001 and included Coopworth crossbred ewes (N. Mitchell, pers. comm.).

The sheep from the New Zealand CLN6 variant research flock will be described in the following as South Hampshire despite these previous outbreeding programs. As no normal SH sheep are born in the research flock, CPW sheep maintained at the research farm at Lincoln University Lincoln University under the same conditions are used as normal controls.

The SH and CPW sheep are characterised as either 'normal', 'carrier' or 'affected' for NCL by using a combination of pedigree information, observation of emerging clinical signs and an indirect DNA test (Tammen et al., 2006). Affected NCL sheep were observed for clinical signs at approximately 10 to fourteen months of age, which is the stage when animals tend to develop blindness and neurological signs (Jolly et al,, 1980). Histopathology analysis of needle brain biopsy samples from lambs at 2 to 3 months of age was used for diagnosis (Dickson et al., 1989) prior to development of the DNA test.

The indirect DNA test used to detect NCL was based on a silent mutation (c.822G>A) identified in exon 7 of the *CLN6* gene coding sequence (Tammen et al., 2006) where the 'A' allele is closely linked to the disease mutation with LOD score of 13.3 (θ=0.01) (Tammen et al., 2006). The control sheep is considered to be unaffected or normal based on a lack of clinical phenotype and histopathology results, if available (Tammen et al., 2006).

A total of 14 South Hampshire and 3 Coopworth sheep were selected for this study with the DNA samples provided by Prof. David Palmer from Lincoln University, Christchurch, New Zealand.

2.1.2 Merino sheep

Two Merino sheep were selected for this study from the NCL research flock at Camden Campus at the University of Sydney, Australia. These animals were maintained under standard Australian pastoral conditions. The flock was established in 1998, initially using advanced reproductive technologies (Multiple Ovulation and Embryo Transfer - MOET, Artificial Insemination - AI and/or estrus synchronisation (Cook et al., 2002). In more recent years the flock has been maintained largely by using natural mating between carrier animals. Concerns about high levels of inbreeding resulted in outcrossing of carrier rams to unrelated Merino ewes obtained from the University of Sydney farms in the year 2008. Procedures performed on the sheep throughout the research project were approved by the University of Sydney Animal Ethics Committee (AEC)

and are in accordance with NIH guidelines, the NSW Animal Research Act (1985) and the Australia Code of Practice for the Care and Use of Animals for Scientific Purposes 7th Edition (NHMRC 2004).

The Merino sheep are characterised as either 'normal', 'carrier' or 'affected' for NCL within the first two months of birth using a direct DNA test (Tammen et al., 2006). The test was developed based on the finding of the disease causing mutation (c.184C>T) in exon 2 of *CLN6*, which results in a major amino acid exchange (p.Arg62Cys). The DNA test comprised of blood collection, DNA extraction, PCR amplification, enzyme cleavage of amplicons and agarose gel separation to visualise PCR products. As well as the DNA test, the demeanor and clinical signs of sheep were closely observed throughout their lifetime.

Although research efforts were mainly focused on the SH breed of sheep, the Merino sheep flock was also involved, particularly for use as control animals when comparing sequence and genomic information (Chapter 7). The routine husbandry procedures performed on the Merino sheep, which were assisted by the author, included ear tagging newborns for identification, vaccination, bleeding and tail docking within the first month. Blood for DNA testing was collected into 15 mg ethylenediaminetetraacetic acid **(**EDTA) treated vacutainer tubes (Becton Dickinson, USA) via jugular venipuncture and samples were kept frozen until processed. Animals were weaned at 2 to 3 months of age and subsequently separated according to sex. At 6 to 8 months of age those lambs that were identified as 'affected' based on the direct DNA test were separated from the flock and housed in small groups in outdoor or indoor pens to best allow daily observation for behavioural changes and signs of disease progression.

Animals were euthanised at a range of ages (e.g. 3, 6 or 8 months), chosen to reflect the progress of the disease (Cook et al., 2002) using intravenous injection of an overdose of Lethabarb (Sodium penthobarbitone) (60 mg/kg). During post-mortem, weights of brain, liver and kidney were recorded and tissue samples from various organs were collected, preserved or frozen, and stored for future analysis.

2.2 DNA extraction

2.2.1. DNA extraction from blood

The South Hampshire sheep genomic DNA (gDNA) samples were extracted by Nadia Mitchell of Lincoln University from pelleted sheep blood leucocytes using the Qiagen QIAamp DNA mini kit (Qiagen, Hilden, Germany). DNA was quantitated using the NanoDropTM ND-1000 spectrophotometer (NanoDrop Technologies Inc, Thermo Scientific, USA) and dried by freezing in a high vacuum (lyophilised) into a DNA pellet before transport to the University of Sydney. Each DNA pellet was resuspended in TE (Tris-EDTA, pH 8.0, Amresco, Ohio, USA) to a required concentration prior to use.

The Merino sheep gDNA was extracted from EDTA treated vacutainer tubes using either one of the following three kits; the QIAamp DNA blood mini extraction kit (Qiagen, Hilden, Germany), QIAamp 96 DNA Blood kit or the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). These kits share a similar principle of lyse-bind-wash-elute spin techniques. The starting amount for all samples was 100 to 200 µl of fresh or defrosted blood with final elution of DNA in similar volume in Qiagen supplied Buffer AE (10 mM Tris·Cl; 0.5 mM EDTA; pH 9.0). DNA extraction with the Qiagen kits was performed according to the respective Qiagen handbooks. Centrifuging of samples was conducted using the tabletop laboratory centrifuge (4K15 model, Sigma, Harz, Germany) and microcentrifuge (5415D model, Eppendorf AG, Hamburg, Germany).

2.2.2 DNA extraction from tissue

During sheep necropsy, part of the liver and kidney were excised with a size 22 scalpel blade (Swann-Morton, Sheffield, England) in small pieces and stored immediately at -80˚C. The liver sample was then defrosted and DNA extracted using the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) according to the Qiagen Handbook. The kit uses similar principles to DNA extraction from blood, which is lyse-bind-wash-elute spin technique. The starting amount for all

samples was approximately 25 mg in wet weight for liver and kidney with a final elution of DNA in 200 µl of Qiagen supplied Buffer AE (10 mM TrisCl; 0.5 mM EDTA; pH 9.0).

2.3 Bacterial Artificial Chromosome (BAC)

Three sheep BAC clones (35B7, 35C9 and 270H8) were kindly provided by Dr. Daniel Vaiman (INRA, Jouy-en-Josas, France) after PCR screening of an ovine BAC library (Vaiman et al., 1999) with *CLN6* specific primers (Table 2.1).

Table 2.1: Screening of ovine BAC clones 35B7, 35C9 and 270H8 using *CLN6* specific primers

Primer	Direction	Sequence 5'>3'
E3F1	Forward	ATCGCCGTGGCTGAGA
E4F1	Forward	ATACAGGTTTCGGGGAGCC
E6F1	Forward	CGAGTGGGCGAGGAAAC
E4R2	Reverse	GAGCGCAGCAGATCCCA

2.3.1 DNA extraction from cultured BAC clones

Culture and purification of ovine BAC clones were performed using the QIAGEN Large-Construct kit protocol with a minor modification where the QIAGEN-tip was washed with 30 ml Buffer QC repeated 3 times instead of twice as outlined in the protocol, for complete removal of remaining RNA and protein contaminants. The purification technique with this kit involved an ATP-dependent exonuclease digestion step for efficient removal of genomic DNA contamination to yield ultrapure, genomic DNA-free BAC DNA. The summarised protocol is described below.

Single colonies were obtained by streaking BAC cultures onto LB agar plates containing 12.5 µg/ml Chloramphenicol (Sigma-Aldrich). Following overnight incubation at 37°C, a single colony was selected and inoculated in a starter culture of 5 ml LB medium containing Chloramphenicol at 25 mg/ml and incubated for 8 hours at 37°C in a shaking incubator (300 rpm). 200 µl of the starter culture was diluted into 200 ml selective LB Medium (1/1000 dilution) and further grown in a shaking incubator (300 rpm) at 37°C for 16 hours. Two ml

aliquots were transferred into NUNC cryovials (Nalgene, NY, USA) containing 2 ml of BAC storage solution (65 %glycerol, 0.25 M Tris-HCl 8.0 pH and 0.1 M MgSO4 heptahydrate) and stored at - 80˚C.

The remaining culture was placed in a 500 ml size centrifuge tube and centrifuged at 6000 rpm at 4˚C for 15min before removing the supernatant to leave a bacterial pellet at the bottom of the tube. This pellet was resuspended in 20 ml of Buffer P1 (containing RNase A solution) and transferred to a clean 60 ml tube before 20 ml of buffer P2 was added and the contents thoroughly mixed. The lysate was incubated at room temperature (RT) for 5 min before 20 ml of chilled buffer P3 was added and the contents immediately and gently mixed by inverting 4 to 6 times. This was followed by incubation on ice for 10 min to produce a fluffy white precipitate containing the genomic DNA, proteins, cell debris and potassium dodecyl sulphate.

The lysate was centrifuged at 16,000 rpm at 4˚C for 30 min before the supernatant was discarded and the lysate transferred through a folded filter paper pre-wetted with distilled water. The DNA was precipitated by adding 36 ml isopropanol at RT. The mixture was centrifuged immediately at 14,000 rpm at 4˚C for 30 min. The DNA pellet was washed with 5 ml of 70% ethanol (RT) then centrifuged at 14,000 rpm for 15 min and the supernatant discarded.

The tube containing the DNA pellet was placed upside down on a paper towel and the DNA allowed air-drying for 2 - 3 minutes. Any visible liquid around the tube opening was carefully removed and the DNA dissolved in 9.5 ml Buffer EX with gentle shaking. 200 µl of ATP-Dependent Exonuclease (completely dissolved in Exonuclease Solvent) and 300 µl of ATP solution were added to the dissolved DNA, gently mixed and incubated in a water bath at 37˚C for 60 min. In this stage, the genomic DNA and nicked DNA were digested by the exonuclease, leaving the remaining supercoiled DNA for further purification.

During the 60 min incubation, a Qiagen-tip 500 column was equilibrated with 10 ml QBT Buffer and the column allowed emptying by gravity. 10 ml Buffer QS was added to the incubated DNA sample and the whole sample was applied to the column and allowed to enter the resin by gravity flow. The column was washed with 30 ml QC Buffer 3 times for complete removal of RNA and

protein contaminants and the DNA was eluted with 15 ml QF Buffer pre-warmed to 65°C. The eluted DNA was precipitated with 10.5ml isopropanol at RT, then mixed and centrifuged immediately at 15,000 rpm at 4°C for 30 min, with the tubes pre-marked to indicate where the glassy pellet would accumulate. The supernatant was carefully decanted to prevent removal of the loosely attached pellet. This pellet was washed with 5 ml 70% ethanol at RT and centrifuged at 15,000 rpm at RT for 15 min. The pellet was air-dried for 10min then dissolved overnight in 250 µl TE (pH8.0, 10mM Tris-HCl).

2.4 Quantitation and quality assessment of DNA templates

DNA concentration of purified genomic and BAC DNA was determined using either one of the following methods; FLUOstar OPTIMA® spectrophotometer analysis (BMG Labtech, Japan), Quant-iT PicoGreen dsDNA fluorescence absorbance measurement, NanoDrop ND-1000 microvolume spectrophotometry (NanoDrop Technologies Inc, Thermo Scientific, USA) or the Agilent 2100 Bioanalyzer (Agilent Technologies, USA) according to their respective handbooks and protocols. The FLUOstar and PicoGreen methods were used for quantitation of gDNA (Chapter 4), whereas the ND-1000 and Agilent 2100 were used for quantitating gDNA and PCR products (Chapters 6 and 7). Quantitation enabled the use of appropriate quantities of DNA in PCR and sequencing.

2.5 Polymerase chain reaction (PCR)

2.5.1 PCR

Polymerase Chain Reaction (PCR) was used extensively for NCL genotyping of Merino, South Hampshire and Coopworth sheep and the preparation for different types of sequencing templates for Sanger sequencing (Chapters 4 and 8) as well as next generation sequencing (Chapters 5, 6 and 7). For each application, detailed PCR protocols and conditions are described in the materials and methods section of each specific chapter.

The HotStarTaq DNA Polymerase kit (Qiagen, Hilden, Germany) and dNTPs (Astral Scientific Pty Ltd, Australia) were used for conventional PCR, whereas the KOD Hot Start DNA Polymerase kit (Merck Biosciences Limited, Australia) was used to amplify long-range PCR products (Chapter 7). For G-C rich templates, additives such as Dimethyl sulfoxide (DMSO) (Amresco Inc, USA) at 5-10% concentration was added to reduce and/or inhibit DNA secondary structures.

PCR mixtures were prepared in 0.2 ml strip tubes (Eppendorf AG, Hamburg, Germany) and 96-well plates (Eppendorf AG, Hamburg, Germany). Two 96-well CG1-96 Thermocyclers (Corbett Research, Sydney, Australia) equipped with built in heated lids to prevent evaporation of the PCR reaction were used for all the PCR reactions. Details of the PCR cycles are presented in specific chapters.

2.5.2 Primers

Primers were designed using online softwares such as Primaclade http://www.umsl.edu/services/kellogg/primaclade.html, NetPrimer http://www.premierbiosoft.com/netprimer/index.html, and Primer3 (Koressaar and Remm, 2007; Untergasser et al., 2012). Both Primaclade and NetPrimer are free web based applications. Primer3 was a program available on the Australian National Genomics Information Service (ANGIS) website which has now been discontinued.

Primaclade was used for choosing primers in conserved non-coding sequences across multiple species (Chapter 3) whereas Primer3 selected primers from the user's single nucleotide sequence input (Koressaar and Remm, 2007; Untergasser et al., 2012). Primers designed using Primaclade and Primer3 were further analysed with NetPrimer for the presence of secondary structures such as hairpins and cross-dimers in primer pairs. In most occurrences, only primers with zero or minimal secondary structures were ordered for synthesis.

The majority of primers were ordered from either Sigma (Sigma-Aldrich Corporation, Australia) or Invitrogen (Life Technologies, Australia) and synthesised at 0.05 µmol synthesis scale and purified by the desalting method. These primers were used for sequencing with the Sanger method performed in house or outsourced to other sequencing service providers (to be further described in Chapter 2.7). Other types of oligonucleotides used were the 5' amine-modified custom oligonucleotides and non-modified barcodes which were ordered from Invitrogen (Life Technologies, Australia) (Chapter 6). These primers were synthesized at 1 µmol synthesis scale and purified by High Performance Liquid Chromatography-HPLC method.

Primers designed for each step of the study are listed in the specific material and method sections.

2.6 Agarose gel electrophoresis to visualise DNA bands

Agarose gel electrophoresis was used to separate and visualise DNA and PCR products of various sizes. The appropriate percentage of agarose gel to use was determined based on the size of DNA to be separated. Higher agarose concentration facilitates separation of small DNA/ PCR products while low agarose concentrations allow resolution of larger DNA/ PCR products. Agarose concentrations of 1 % to 2.5 % were used in this research.

The gels were prepared using Agarose I gelling agarose powder (Amresco, Ohio, USA) mixed with 1XTris-borate-EDTA (TBE) (Amresco, Ohio, USA) buffer heated in a microwave oven until completely dissolved. The mixture was cooled to about 60ºC, and in a fume hood, ethidium bromide (10 mg/ml stock concentration, Amresco, Ohio, USA) was added to a final concentration of 0.5 ug/ml. The mixture was poured into a casting tray containing a sample comb/s (each comb with between 10 to 30 wells for sample loading) and after the gel had solidified it was moved out of the fume hood.

The casting tray was inserted into an electrophoresis chamber and covered with 1X TBE buffer. After removal of the comb, 5 µl of each sample was mixed with 1 µl of agarose gel loading

buffer (15 % Ficoll PM 400 (Sigma-Aldrich, USA) and 0.25 % bromophenol blue) and loaded into wells. Various molecular size markers were used to enable estimation of size and concentration of DNA or PCR products.

The gel in the electrophoresis chamber was applied with electric currents of various voltages and times and the negatively charged DNA samples migrated towards the positive electrode of the chamber. Samples could be visualised as the ethidium bromide within the gel would intercalate with the DNA during electrophoresis and DNA could thus be observed using a BioRad ultraviolet transiluminator (76 S Geldoc, Bio-Rad Lab Pty Ltd, California, USA) and documented with the GelDoc XR Quantity One analysis software (Bio-Rad Lab Pty Ltd, California, USA).

2.6.1 Gel extraction from agarose gels

In some instances where spurious or non specific bands were present on the agarose gel, the band/s of interest was excised using sterile size 22 scalpel blades (Swann-Morton, Sheffield, England) or the GeneCatcher PK B4.0 cutting tips (GeneCatcher, CA, USA) fitted into a standard 1000 µl pipette (Gilson Pipetman, USA). The removed gel was purified using Qiagen QIAquick gel extraction kit (Qiagen, Hilden, Germany) using the spin technique following the manufacturer's protocol. Purified DNA was eluted in 30 µl of Qiagen supplied buffer EB (10mM Tris-Cl,pH 8.5) and analysed on an agarose gel to examine the quality and presence of the DNA.

2.7 Sequencing

2.7.1 Sanger sequencing of PCR products

Early sequencing work, taking place in the years 2007 to 2009, was mostly performed using the Sanger method (Sanger and Coulson, 1975, 1977). The quantity of DNA template used in sequencing reactions varied according to the expected length of the PCR products. For most of

the sequencing conducted in-house, 50 ng (based on BigDye Terminator v3.1 Cycle Sequencing Kit, Applied Biosystems, USA) of DNA template was required to sequence approximately 2,000 bp PCR products. The PCR products were purified from residual primers and nucleotides using ExoSAP-IT (USB, Ohio, USA) with a volume of 1 µl ExoSAP-IT per 5 µl PCR product, following a modified manufacturer's protocol of incubation at 37 C for 60 min and 80 C at 15 min on the thermocycler. An extension of time during the first incubation step was needed due to the reduced amount of ExoSAP enzyme (1 µl compared to manufacturer's recommendation of 2 µl) added to the PCR products.

Purified PCR products were sequenced using the BigDye Terminator v 1.1/3.1 Sequencing Kit (Applied Biosystems, USA) in a 20 µl sequencing reaction. The sequencing mix comprised of 1 - 5 µl of purified PCR template (corresponding to approximately 1 - 50 ng of DNA), 4 µl of the 2.5X Ready Reaction mix (containing premixed fluorescently labeled ddNTPS), 2 µl of the 5X BigDye Sequencing Buffer, 4 pmol primer (forward or reverse) and molecular grade water to 20 µl of final solution. The mix was placed into a thermocycler under the following general conditions: 96°C for 1 min; 25 cycles at 96˚C for 10 sec, 50˚C for 5 sec; and 60˚C for 4 min.

2.7.2 Direct sequencing of BAC DNA

Direct sequencing of BAC DNA was attempted with primers that failed to amplify a single band PCR product. Purified DNA of BAC clones was sequenced using BigDye Terminator v 1.1/3.1 Sequencing Kit (Applied Biosystems, USA) in 20 µl sequencing reaction.

The sequencing mix comprised of 1 - 5 µl of purified DNA template (corresponding to approximately 0.5 – 1.0 µg of DNA), 4 µl of the 2.5X Ready Reaction mix (containing premixed fluorescently labeled ddNTPS), 2 µl of the 5X BigDye Sequencing Buffer, 4 pmol primer (forward or reverse) and molecular grade water to 20 µl of final solution. The mix was placed into a thermocycler with the following conditions: 95°C for 5 min; 30 cycles at 95˚C for 30 sec, 50 - 55˚C for 10 sec; and 60˚C for 4 min.

Extension products were purified using the Ethanol/EDTA/Sodium Acetate Precipitation method as described in the BigDye Terminator v 1.1/3.1 Sequencing Kit user manual (Applied Biosystems, USA) to remove excess terminators and salts that might have interfered with capillary electrophopresis.

Purified products were analysed and visualised using the ABI 3130 Genetic Analyzer (Applied Biosystems, USA) when performed in-house or ABI 3730 (Applied Biosystems, USA) when outsourced to Sydney University Prince Alfred Molecular Analysis Centre (SUPAMAC, Sydney, Australia), Australian Genome Research Facility (AGRF, Melbourne, Australia) or Westmead Millennium Institute (Sydney, Australia). Preparation of samples for both in-house and outsource sequencing is described in the relevant sections of each chapter.

2.7.3 Next generation sequencing (NGS) methods

Other than the conventional Sanger sequencing, two NGS methods; Roche 454 Pyrosequencing and ABI SOLiD sequencing platforms were also used to provide sequences for the genome regions of interest. Principles and detailed information on these methods are described in Chapters 5, 6 and 7.

2.8 Bioinformatic analysis

Published and known genomic sequences from sheep and various species were obtained from publicly available genome databases such as Genbank (http://www.ncbi.nlm.nih.gov/genbank/) and Ensembl (http://www.ensembl.org/index.html). Sequence multiple alignment was performed using either one or all of the following programs: VISTA (Mayor et al., 2000), GeneDoc (Nicholas et al., 1997) and Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) (Goujon et al., 2010; Sievers et al., 2011). Two genome browsers; The University of California Santa Cruz (UCSC) and password protected GBrowse (http://crcidp.vetsci.usyd.edu.au/cgi-bin/gb2/gbrowse/NCL/) were used for mapping and visualisation of sequence against the bovine genome (Chapter 4) and ovine reference sequences (Chapters 5, 6, 7 and 8). The Basic local alignment search tool (BLAST) (Altschul et al., 1990) and Basic local alignment tool (BLAT) on

the National center for biotechnology information (NCBI) website (http://www.ncbi.nlm.nih.gov/) were used extensively to search for user's defined sequence/s in the published Genbank genome databases. Other programs used included Gumby (Prabhakar et al., 2006), which identifies statistically significant conservation sequences (Chapter 3), ExPASy Translate tool (Gasteiger et al., 2003) that translates *CLN6* transcript sequences to protein sequences and GREMET (Athanasiadis et al., 2013) to predict effects on gene regulation.

CHAPTER 3: IDENTIFICATION OF CONSERVED NON - CODING SEQUENCES (CNCS) IN THE *CLN6* GENOMIC REGION

3.1 Introduction

Linkage analysis has mapped the NCL in the South Hampshire sheep to a region on ovine chromosome 7 (OAR7) containing the *CLN6* (Broom et al., 1998; Tammen et al., 2006). No disease causing mutation was identified in the gene coding sequence, but mRNA concentrations are reduced in affected SH sheep (Tammen et al., 2006). The research hypothesis states that, firstly, the NCL causative mutation is in the ovine *CLN6* non-coding regions or sequence flanking the gene and secondly, that such a mutation would have an effect on gene regulation. This chapter outlines an *in silico* analysis approach to identify conserved non-coding sequences (CNCS) as potential regions of interest for such mutations.

3.1.1 Identification of CNCS

3.1.1.1 Conserved non-coding sequences (CNCS)

Sequences that do not code for protein make up the majority of the mammalian genome (Smith et al., 2004). The proportion of coding and non-coding DNA varies widely between organisms and non-coding DNA makes up more than 98% of the human genome (Elgar and Vavouri, 2008). In the past, non-coding DNA was labelled 'junk DNA' or non-functional DNA sequences that divide protein-coding exons (Ohno, 1972). However, recent studies, as will be shown in the following paragraphs, have documented that non-coding DNA sequences represent an extraordinary trove of information about biological processes.

Non-coding DNA sequences have been shown to play important roles in the regulation of gene expression. Regulatory elements in 5' and 3' untranslated regions (UTR) have been identified as protein binding sites that control regulation at translational level (Mignone et al., 2002), whereas introns have been shown to contain independent transcriptional units that contribute to gene regulation at the transcriptional level (Shabalina and Spiridonov, 2004).

Conserved non-coding sequences (CNCS) are regions of the genome that contain a high level of non-coding sequence similarity when aligned between species (Hardison, 2000). It is predicted that CNCS are preserved interspecies under functional and evolutionary constraints (Ludwig, 2002) that evolved naturally as a result of negative selection (Shabalina and Spiridonov, 2004) because they represent sequences with biologically important elements (Frazer et al., 2004).

3.1.1.2 Cross-species comparative analysis of DNA sequences

Comparison of long genomic segments across multispecies has been found to be a powerful approach for identifying functional segments of the non-coding regions, such as gene regulatory elements (Hardison, 2000). When the current research project began there was only limited available ovine genome sequence information for regions flanking the *CLN6*. This was because the sheep genome assembly Oar v1.0 released in 2009 that mapped ovine scaffolds onto the bovine genome framework was far from complete and contained many sequence gaps or unsequenced regions. Due to this limitation, orthologous counterparts from representative members of rodent (mouse, rat), other mammal (human, cattle, macaque, dog, opossum), aves (chicken), and fish (Fugu fish, pufferfish) lineages were aligned to detect CNCS.

The chromosomal location of the ovine *CLN6* orthologs were chromosome (Chr) 15 in human, Chr 7 in macaque, Chr 9 in mouse, Chr 10 in cattle, Chr 30 in dog, Chr 10 in chicken, Chr 1 in opossum and Chr 8 in rat. The *CLN6* orthologue was not assigned to a chromosome in the Fugu fish but sequence was provided in scaffold_1 of the sequence assembly. These nine species used for alignment against the sheep sequence were chosen based on availability of sequence

information and their different positions on the phylogenetic tree compared to sheep. They were aligned because sequences that remain highly conserved between divergent organisms are likely to be functional (Elgari and Vavouri, 2008).

The phylogenetic tree (Figure 3.1) was generated with the NCBI taxonomy browser (http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/) using information from the NCBI taxonomy database. The database does not follow a single taxonomic treatise but rather incorporates phylogenetic and taxonomic knowledge from a variety of sources, including the published literature, web-based databases, and the advice of sequence submitters to NCBI and outside taxonomy experts (Sayers et al., 2009; Benson et al., 2008).

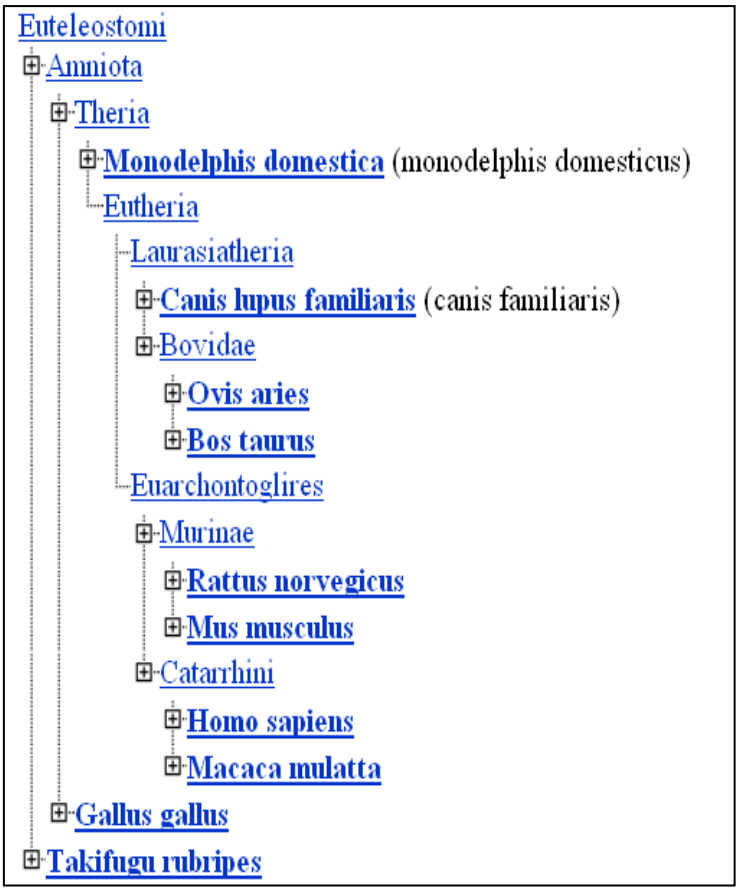


Figure 3.1 indicate positions of the 10 species (shown in underlined bold blue text) used for cross-species genomic multialignment visualised using the NCBI taxonomy browser (http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/).

3.2 Materials and methods

3.2.1 Cross-species sequence analysis

Complete or draft sequences for ten different species were obtained from the publicly available genome databases Genbank (http://www.ncbi.nlm.nih.gov/genbank/) and Ensembl (http://www.ensembl.org/index.html). A summary of the sequence information obtained appears below (Table 3.1). Eight sequences were downloaded from Ensembl (version 47, release date October 2007); human '*Homo sapiens*' (Ensembl Gene ID: ENSG00000128973), cattle '*Bos taurus*' (Ensembl Gene ID: ENSBTAG00000005565), dog '*Canis familiaris*' (Ensembl Gene ID: ENSCAFG00000017473), macaque '*Macaca mulatta*' (Ensembl Gene ID: ENSMMUG00000011703), Fugu fish (pufferfish) '*Takifugu rubripes*' (IMCB Gene ID: NEWSINFRUG00000122292), house mouse '*Mus musculus*' (Ensembl Gene ID: ENSMUSG00000032245), opossum '*Monodelphis domestica*' (Ensembl Transcript ID: ENSMODG0000000942) and rat '*Rattus norvegicus*', (Ensembl Gene ID :ENSRNOG00000007164). Chicken genomic sequence was obtained from Genbank: '*Gallus gallus*' (GenBank: NW_060444), as was ovine '*Ovis aries*' *CLN6* mRNA sequence (GenBank: DQ458790.1), which was supplemented with unpublished genomic sequence provided by Nadia Mitchell, Lincoln University (Sequence of 12,601 kb is shown in Appendix 1).

Table 3.1: Information of the 10 genomic sequences obtained from genomic databases Ensembl and Genbank. The only genomic sequence obtained from Genbank is marked with the symbol *

Species	Gene ID	Gene ortholog	Genome assembly	Transcript ID	Chr	Exons	Start sequence	End sequence	Finished /Draft	Total length(bp)
Human	ENSG00000128973	CLN6	NCBI 36	ENST00000249806	15	7	66,286,386	66,309,102	Finished	42717
Macaque	ENSMMUG00000011703	CLN6	MMUL 1.0	ENSMMUT00000016395	7	6	46,797,911	46,828,306	Finished	30396
Mouse	ENSMUSG00000032245	Cln6	NCBI m37	ENSMUST00000034776	9	7	62,626,793	62,660,008	Finished	33216
Cattle	ENSBTAG00000005565	CLN6	Btau_3.1	ENSBTAT00000007315	10	6	9,884,836	9,912,794	Draft	27959
Dog	ENSCAFG00000017473	CLN6_ CANFA	CanFam 2.0	ENSCAFT00000027690	30	6	35,182,226	35,211,895	Finished	29670
Chicken*	Accession no:NW_060444	CLN6	none	Locus: NW_060444	10	none	1,065,331	1,101,231	Finished	35,900
Opossum	ENSMODG00000009425	CLN6	MonDom5	ENSMODT00000011999	1	6	155,193,191	155,228,317	Finished	35127
Rat	ENSRNOG00000007164	Cln6	RGSC 3.4	ENSRNOT00000034926	8	7	66,977,073	67,010,998	Finished	33926
Fugu fish	NEWSINFRUG00000122292	CLN6	FUGU 4.0	NEWSINFRUT00000129135	scaffold_1	8	3,061,164	3,085,822	Finished	24659
Sheep*	Accession no:DQ458790 (mRNA)	CLN6	-	Locus: DQ458790	7		1	933	Draft	933

3.2.2 Sequence alignments and visualisation using bioinformatic programs

The VISTA program (Mayor et al., 2000) was used for comparative analysis and alignment visualisation of the ten sequences. The global alignment algorithm (Needleman and Wunsch 1970) used in VISTA incorporates an assumption that highly similar regions in sequences from related organisms appear in the same order and orientation (Bray et al., 2003). Biological sequences from related organisms such as those used in this study satisfy this assumption. DNA sequences were uploaded into the VISTA server (http://genome.lbl.gov/vista/index.shtml) in individual single contig FASTA files. The sheep and cattle genome sequences used were draft sequences. The other eight genome sequences; human, macaque, dog, mouse, rat, chicken, opossum and Fugu fish, were obtained from completed eukaryotic genomes ((Lander et al., 2001 (humans) ; Gibbs et al., 2007 (macaque); Lindblad-Toh et al., 2005 (dog); Waterston et al., 2002 (mouse); Gibbs et al., 2004 (rat); Hillier et al., 2004 (chicken); Mikkelsen et al., 2007 (opossum); Aparicio et al., 2002 (Fugu fish)). However, it needs to be acknowledged that reported 'completed genomes' still contain gaps, misarrangements and incomplete annotations.

Sequences were aligned using a multiple alignment method AVID (Bray et al., 2003) provided in VISTA. AVID was the only alignment program available through this program server with the ability to align both draft and complete sequences by ordering and orienting the former sequence to the latter. The other two alignment programs LAGAN and Shuffle-LAGAN (Brudno et al.,

2003) perform alignments for complete genomes. Conservation parameters for calculating CNCS on the VISTA plot were maintained at default values of a minimum 70% identities at 100 bp, generated by sliding the default window of 100 bp long along each pairwise sequence alignment and calculating the percent identity at each base pair position (Frazer et al., 2004).

Human *CLN6* was used as the reference sequence as it encompasses a complete gene sequence with known annotations. The human gene annotation file was submitted for mVISTA (multiple VISTA for comparing two or more species DNA sequences) analysis in a simple plain text format (Figure 3.2). The human *CLN6* is defined by its start and end coordinates on the sequence. A greater than (>) was placed in the first line to indicate plus strand and the numbering was displayed according to the plus strand as well. The exons are listed individually starting with the label 'exon', after the start and end coordinates of each exon. UTRs are defined here as sequences upstream to exon 1 (coordinates 1 – 10126) and downstream to exon 7 (31572 – 42717) of the *CLN6*.

```
>10127 31571 Human
HS
10127 10209 exon 1
21061 21175 exon 2
25323 25421 exon 3
27848 28036 exon 4
28393 28448 exon 5
29952 30074 exon 6
```

Figure 3.2: The human *CLN6* annotation file uploaded as an input for mVISTA analysis. The gene is defined by its start 'exon' coordinates on the plus DNA strand. The annotation file was used as reference sequence for cross-species sequence comparison of the nine multispecies genomic sequences against the human reference sequence.

To determine whether the alignment results of the 10 species conform to the NCBI phylogenetic tree in Figure 3.1, another cross-species comparison was analysed using the LAGAN alignment program. LAGAN was used instead of AVID as it was the only program that calculates a phylogenetic tree. Input sequences uploaded into VISTA were similar to those described earlier for AVID.

3.3 Results

3.3.1 Identification of CNCS across *CLN6* orthologs

The VISTA plot of the *CLN6* orthologs from human, macaque, cattle, sheep, dog, mouse, rat, chicken, opossum and Fugu fish is presented below (Figure 3.3). The x-axis represents the human reference sequence (kb) and the y-axis represents the percent identity (%). The human exons and UTRs are marked above the plot based on the annotation file uploaded into the program (Figure 3.2).

Exon 1 was shown to be conserved only between the species of human and sheep (Figure 3.3; Table 3.2) whereas exons 2 to 7 were highly conserved in most species with these exons containing almost similar lengths to the human reference sequence. The largest CNCS were identified in introns 1, 2 and 6 as well as flanking sequences upstream (5') and downstream (3') to the *CLN6*. A region of particular interest was identified in position 7,885 - 8,253 kb approximately 2 kb upstream of *CLN6* which showed greater than 70% sequence identity between human/ macaque/ dog/ mouse/ rat (genomic sequence for cattle, sheep, chicken, opossum and Fugu were lacking for this area). Given that transcriptional control of gene expression is likely to be coordinated by DNA sequences upstream of the transcription start site of genes (Latchman, 2005), this CNCS upstream of the *CLN6* was considered as a high priority for sequencing and mutation screening.

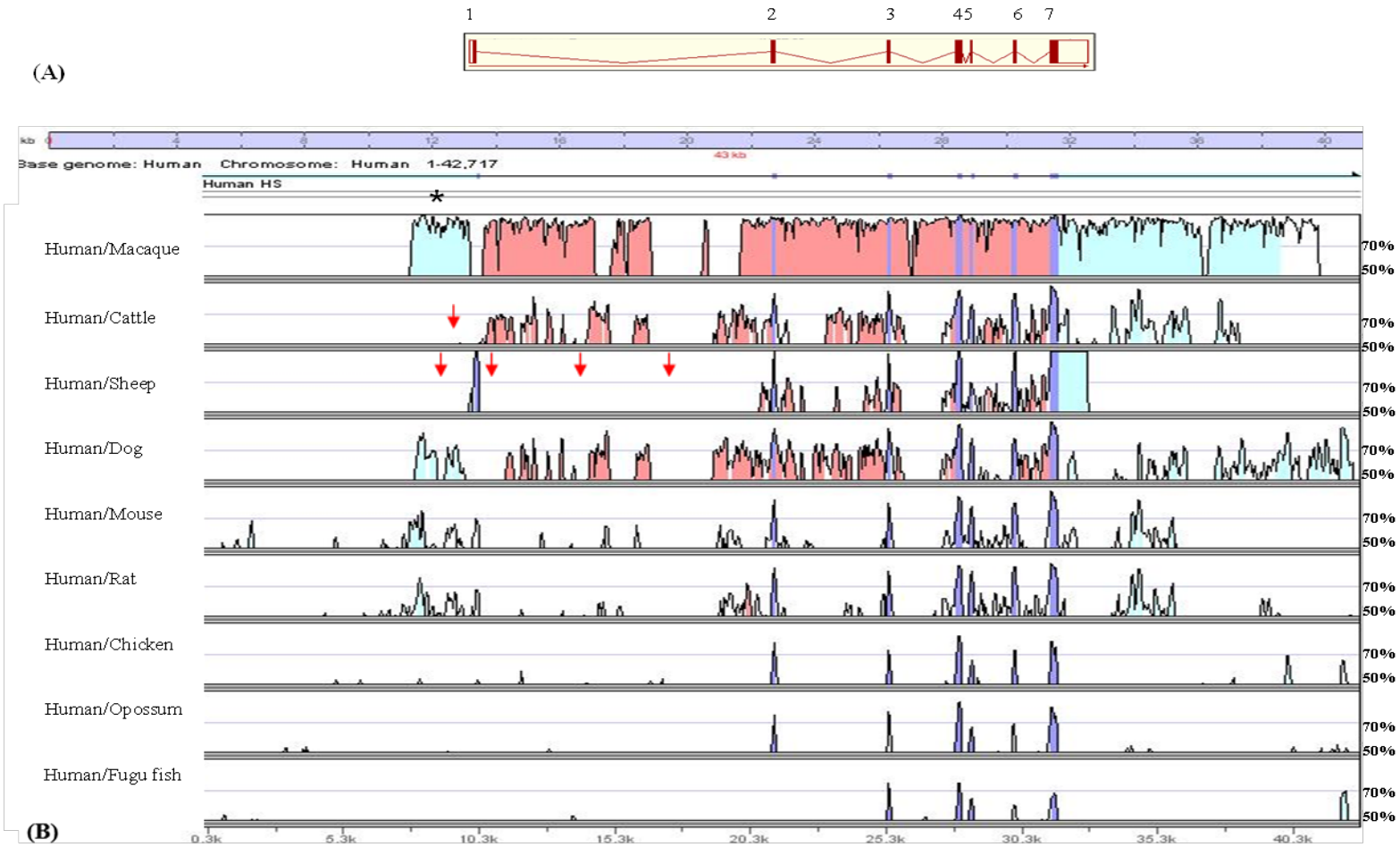


Figure 3.3: (A) Diagram of the *CLN6* sense transcript showing Exons 1 to 7 (modified diagram from cattle transcript taken from Ensembl Transcript: CLN6-201 and Gene: ENSBTAT00000007315), (B) VISTA plot for *CLN6* orthologs between human and macaque, cattle, sheep, dog, mouse, rat, chicken, opossum and Fugu fish. Peaks are shown relative to their position in the human reference sequence (x-axis) and their percent identities are indicated on the y-axis. The colouring of the different conserved regions correspond to the human *CLN6* gene annotation as shown in Figure 3.2; pink regions are conserved non-coding sequences (CNCS), the dark blue regions which conform to the gene transcript diagram above (A) denote coding sequences (exons) and the turquoise regions indicate UTRs. The red arrows are added manually to signify gaps in species with draft sequences. * marks the CNCS region identified upstream to the *CLN6* and considered priority for sequencing in Chapter 4.

Table 3.2: VISTA summary of conserved regions in the *CLN6* coding sequences between human and macaque, cattle, sheep, dog, mouse, rat, chicken, opossum and Fugu fish. The exon length and percentage identities shown in the other species are relative to the human *CLN6*.

Species	Exon 1	Exon2	Exon3	Exon4	Exon5	Exon6	Exon7
Human	83 bp	115 bp	99 bp	189 bp	56 bp	123 bp	271 bp
Macaque	-	96.5% ; 115 bp	97.0% ; 99 bp	97.4% ; 189 bp	98.2% ; 56 bp	95.9% ; 123 bp	98.5% ; 271 bp
Cattle	-	87% ; 115 bp	91.9% ; 99 bp	91.0% ; 189 bp	92.9% ; 56 bp	91.1% ; 123 bp	93.4% ; 271 bp
Sheep	83.1% ; 83 bp	87.8% ; 115 bp	92.9% ; 99 bp	91% ; 189 bp	92.9% ; 56 bp	90.2% ; 123 bp	92.3% ; 271 bp
Dog	-	87.0% ; 115 bp	90.9% ; 99 bp	90.5% ; 189 bp	91.1% ; 56 bp	81.3% ; 123 bp	93% ; 271 bp
Mouse	-	87% ; 115 bp	86.9% ; 99 bp	86.2% ; 189 bp	91.1% ; 56 bp	87.0% ; 123 bp	90.0% ; 271 bp
Rat	-	85.2% ; 115 bp	86.9% ; 99 bp	86.8% ; 189 bp	92.9% ; 56 bp	87.8% ; 123 bp	87.8% ; 271 bp
Chicken	-	77.4% ; 115 bp	75.8% ; 99 bp	84.7% ; 190 bp	80.4% ; 56 bp	75.4% ; 122 bp	78.4% ; 268 bp
Opossum	-	75.4% ; 114 bp	-	82.0% ; 189 bp	83.9% ; 56 bp	-	79.0% ; 271 bp
Fugu fish	-	-	78.8% ; 99 bp	71.5% ; 186 bp	80.4% ; 56 bp	-	69.0% ; 171 bp

3.3.2 Phylogenetic tree generated by LAGAN alignment program

The LAGAN program calculated phylogenetic tree (Figure 3.4) was mostly in agreement with the NCBI phylogenetic inference made using the NCBI taxonomy browser (Figure 3.1).

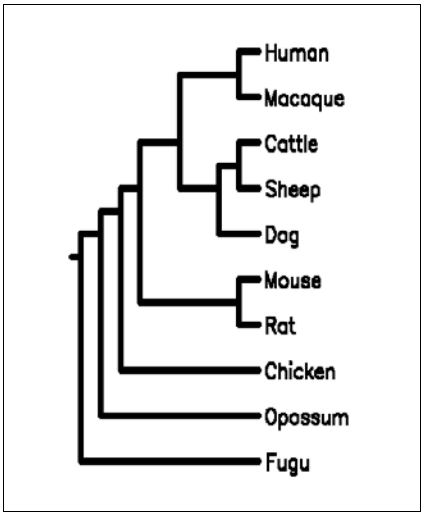


Figure 3.4: Phylogenetic tree generated by LAGAN cross-species sequence multialignment.

3.4 Discussion

Most of the *CLN6* orthologs share structural similarity to humans, which is not surprising considering orthologs share a closer common ancestor, and have sequence and potentially, functional similarity (Peterson et al., 2009). The significantly conserved region in 5' upstream to the *CLN6* exon 1 found between species human/ macaque/ dog/ mouse/ rat but absent in cattle and sheep are likely due to corresponding segments of the DNA that have not been sequenced in the cattle and sheep draft genome assemblies.

Exon 1 was shown to be conserved only in human/ sheep. This may be due to several factors including the VISTA default parameters that only identify conserved regions at or higher than 70 % identify, exon 1 is missing in several species or because these corresponding regions have not been sequenced in the draft assemblies (e.g. cattle). This variation in the existence of exon 1 is very interesting and will be further investigated (Chapter 7). Sequence flanking exon 1 was well conserved such as seen in intron 1 (human/ macaque/ cattle/ dog) and upstream sequences (human/ macaque/ cattle/ dog/ mouse/ rat) (Figure 3.3).

Sequences conserved across species of different evolutionary distances highly suggest a conservation of function (Thomas et al., 2003; Larney et al., 2015). The contrast in sequence conservation when comparing Fugu fish and mammalian DNA, and mammalian against mammalian DNA where they are closely related, is striking (Lettice et al., 2003; He and Zhu, 2011). Higher conservation between human/ macaque/ cattle/ sheep sequences compared to human/ Fugu fish were observed in this study.

Computational approaches to sequence alignment generally fall into two categories, global alignments and local alignments. The choice of an alignment program depends on the sequence set to be aligned (Thompson et al., 1999) and many studies have used either method to align multispecies sequences. Examples are seen in the identification of conserved regulatory elements in Pax6 intron 7 involved in (auto) regulation and alternative transcription (Kleinjan et al., 2004) using global alignment; and identification of binding motifs of sequences flanking the testis-determining gene Sry in 17 mammalian species using both global and local sequence alignment

methods (Larney et al., 2015). In both studies, most of the candidate regulatory elements were experimentally demonstrated to be functional. VISTA was used for the present study because of the better performance of global alignment than local methods in aligning divergent sequence to a set of closely related sequences (Thompson et al., 1999); seen here with Fugu fish/ chicken/ opossum compared to human/ macaque/ dog/ cattle/ sheep. VISTA was also chosen because of the user-friendly interface and customised visualisation mode.

In conclusion, an *in silico* approach was used to identify regions of interest surrounding and within the ovine *CLN6*. As described earlier, CNCS have been identified as potential regions of interest for mutations that may have an effect on *CLN6* expression. Although long-range regulatory elements have been found to modify gene expression from a distance of up to 1Mb from the target gene (Lettice et al., 2003) it is more likely that the mutation causing NCL in the South Hampshire sheep is relatively close to the gene itself due to the strong linkage between a polymorphism in exon 7 of *CLN6* and NCL in the South Hampshire sheep (Tammen et al., 2006). Highly conserved regions in parts of the 5' UTR, 3' UTR and introns 1, 2 and 6 of the *CLN6* orthologs suggest a functional importance and are predicted to contain the disease causing mutation in South Hampshire sheep. Sequencing of these identified CNCS will be discussed in the following chapter (Chapter 4).

CHAPTER 4: SANGER SEQUENCING OF IDENTIFIED CNCS REGIONS WITHIN AND FLANKING THE OVINE *CLN6*

4.1 Introduction

Five conserved non-coding sequences (CNCS) in introns 1, 2 and 6, and parts of the flanking sequence upstream (5') and downstream (3') of the *CLN6* were identified in previously described work (Chapter 3) as potential regions of interest for housing the NCL causative mutation in the South Hampshire sheep. The regions upstream of *CLN6* and intron 1 were considered a priority for sequencing, as they were likely to contain transcriptional regulatory elements and had not been sequenced previously (region upstream of *CLN6*) or only partially sequenced (intron 1) in sheep. This chapter describes Sanger sequencing of an ovine BAC clone containing these regions, the aim being to generate *de novo* ovine specific sequence.

4.2 Materials and methods

4.2.1 Ovine BAC clone

DNA from ovine BAC clone 35C9, extracted during a previous project (Houweling, 2009) using methods previously described (Chapter 2.3.1) was chosen as the template for Sanger sequencing.

4.2.2 Primers and PCRs

4.2.2.1 Sequencing of PCR products with primers derived from cross-species alignment for 5'-CNCS

The 369 bp length CNCS (Appendix 2) found 1,906 bp 5' upstream of the *CLN6* (5'-CNCS) of the *CLN6* start codon in the human genome was identified using the VISTA multialignment

program. This CNCS is located between positions 7,885 - 8,253 bp in the human reference sequence (Figure 3.3). A GeneDoc (Nicholas et al., 1997) multialignment of this region in human, macaque, dog, mouse and rat is presented in Figure 4.1. This cross-species sequence alignment file was used to input into Primaclade software (http://www.umsl.edu/services/kellogg/primaclade.html) to identify a set of PCR primers that should bind in all species used in the alignment. Primers were further analysed with NetPrimer (http://www.premierbiosoft.com/netprimer/index.html) to check for secondary structures, and only primers with no or minimal secondary structures were ordered for synthesis. The sequences of the four primers identified appear below (Table 4.1) and the predicted positions in the cross-species sequence alignment and in the ovine genome as well as directions of primers can also be visualised (Figures 4.1 and 4.2).

Table 4.1 Primers derived from VISTA cross-species alignment for 5' -CNCS

Primer	Direction	Sequence 5'>3'
5UTRF1	Forward	TGTTCATTCAGAAAGGCCC
5UTRR2	Reverse	GCTTCCAGCCATCAGAGG
5UTRF3	Forward	GCCTCATCCTCTGATGGCT
5UTRR4	Reverse	ACCAGAGAAGAAGGATTGAGG

A PCR with a total reaction volume of 20 µl, containing QBuffer (Qiagen, Hilden, Germany), 1.5 mM $MgCl_2$, 0.2 µM of each dNTP (Astral Scientific Pty Ltd, Australia), 20 pmol of each primer (using all possible combinations of primers in Table 4.1), 0.5 U of HotStar Taq DNA polymerase (Qiagen, Hilden, Germany) and 50 ng of BAC DNA was performed under the following conditions: 95°C for 15 min, 40 cycles at 95°C for 30 sec, various annealing temperatures T_a (50°C to 58°C) for 30 sec, 72°C for 1 min and cooling at 4°C for 10 min. PCR products were visualised using agarose gel electrophoresis as described earlier (Chapter 2.6).

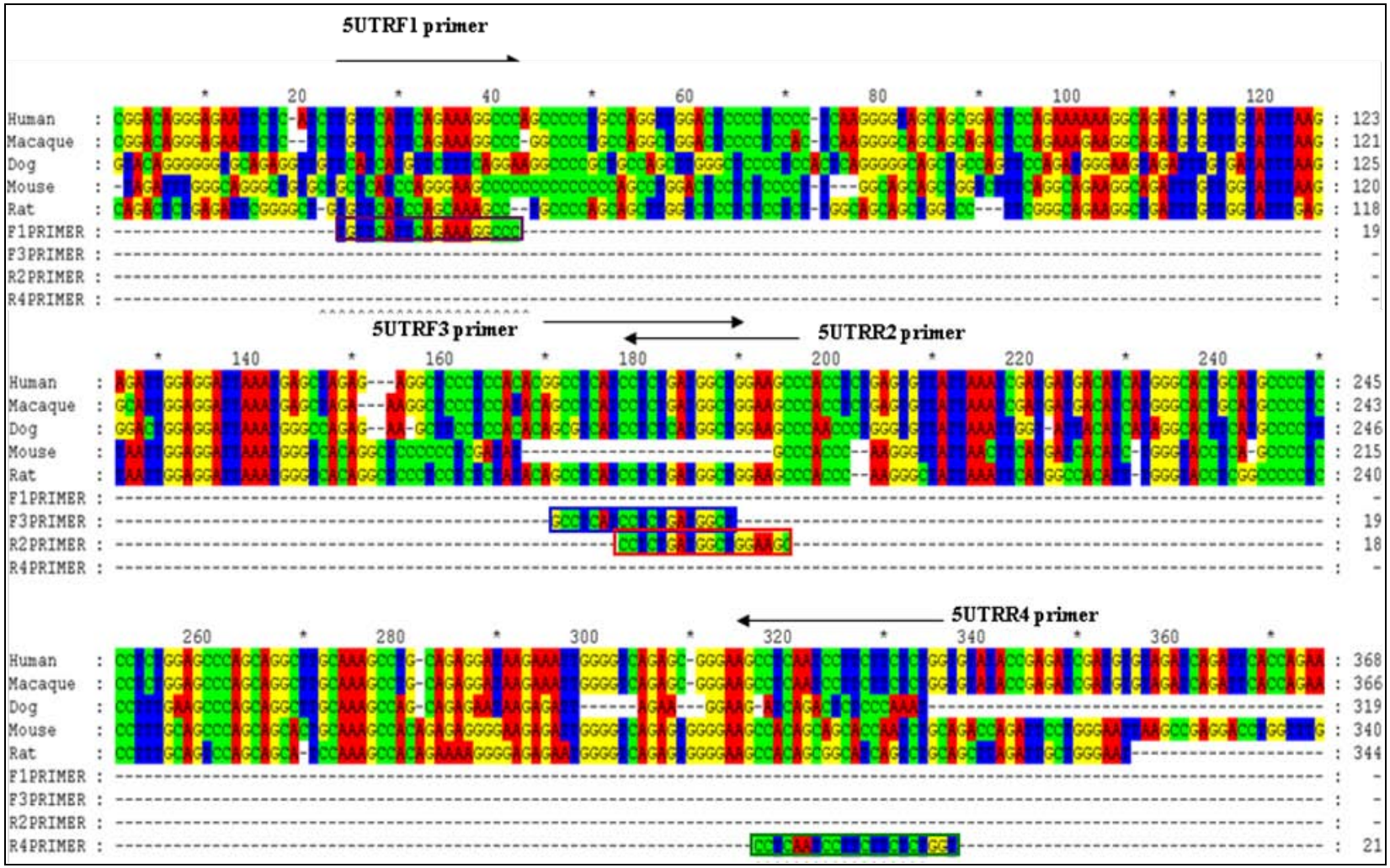


Figure 4.1: Alignment of macaque, dog, mouse and rat CNCS sequences 5' upstream to *CLN6* (5' –CNCS) with a human *CLN6* reference sequence (position 7,885-8,253) for 5'UTR-primer identification. Primaclade (http://www.umsl.edu/services/kellogg/primaclade.html) software identified four PCR primers that should bind across the cross -species alignment. The positions and names of each primer (5'UTRF1 primer, 5'UTRF3 primer, 5'UTRR2 primer, 5'UTRR4 primer) are shown using coloured boxes along with the directions of primers represented by black arro

Only the primer combination 5UTRF1/ 5UTRR4 generated a single PCR product of approximately 200 bp length at a T_a of 55°C. This was sequenced as described in Chapter 2.7.1. The four primers were also used for direct sequencing of BAC DNA as previously described (Chapter 2.7.2), but direct sequencing was only achieved with the primer 5UTRR4.

4.2.2.2 Ovine and human specific primers for direct BAC sequencing of the 5'-CNCS

The sequence generated using the primer combination 5UTRF1/ 5UTRR4 (Chapters 4.2.2.1 and 4.2.3.1) was used to design ovine specific primers for direct sequencing using the Primer3 software (Rozen and Skaletsky, 2000) available at the now discontinued Australian National Genomics Information Service (ANGIS) website, and further analysed with NetPrimer. The sequences of the four ovine specific primers identified are shown below (Table 4.2) as are the predicted positions in the ovine genome and directions of primers (Figure 4.2). An additional primer, 5UTRR8, was designed using the 5' sequence upstream of the *CLN6* start codon in the human genome (Table 4.2).

Table 4.2 Ovine and human specific primers for 5' -CNCS

Primer	Direction	Sequence 5'>3'
5UTRR6	Reverse	CTGTATGTCCCTTCCTTCGG
5UTRF7	Forward	GATGGGACAGGAAAGGGAGA
5UTRR8	Reverse	AGGAAGAGACCGGTTCAGCTC
5UTRF9	Forward	ACTCTGACCCCAATCTCATTCTCT
5UTRF11	Forward	CCTTTCTCTCATTTGCTCCTCAC

4.2.2.3 PCRs with primers for intron 1 of ovine CLN6

Seven primers (I1F1, I1R2, I1F3, I1R4, I1F5, I1R6 and I1F7) within intron 1 were designed using the ovine intron 1 sequence previously generated by Nadia Mitchell (Lincoln University, New Zealand, pers. comm.). Primers were designed using the software packages (Primer3 and NetPrimer) mentioned above and paired during PCR using all possible combinations. The

sequences of the seven primers are shown below (Table 4.3) with the predicted binding positions in relation to ovine *CLN6* and direction of primers also shown (Figure 4.2). An additional primer I1R8 was designed using the bovine *CLN6* intron 1 sequence for direct sequencing of BAC DNA (Table 4.3).

Table 4.3 Ovine and bovine specific primers for intron 1 of ovine *CLN6*

Primer	Direction	Sequence 5'>3'
I1F1	Forward	ATACGACACTGCCCGCTCAAATAG
I1R2	Reverse	AACACTTTCTACATTCCAGGCACTC
I1F3	Forward	TGGCCGGAAGGTTAGCCTGGA
I1R4	Reverse	TCCAGAACCCAGTTCTGCAGAGTG
I1F5	Forward	GACCAGAGTGAGAGCGTAGAGTG
I1R6	Reverse	AACTAGGGTGAGGCAAGTGAGA
I1F7	Forward	CCAGAGAAAGAACGAGGAAGG
I1R8	Reverse	CAGTCAGCCCTTCTCTACTCCA

As PCRs were designed mostly for long-range amplification, the KOD Hot Start DNA Polymerase kit (Merck Biosciences Limited, Australia) and supplied reagents were used to amplify PCR products in this region. PCRs with total reaction volumes of 50 µl were prepared using reagents KOD Buffer, 2.5 mM $MgSO_4$ and 0.2 µM dNTP supplemented with the kit, as well as 20 pmol of each primer (using all possible combinations of the ovine specific primers in Table 4.3), 1 U of KOD Hot Start DNA polymerase (Merck Biosciences Limited, Australia) and 200 ng of BAC DNA. Long-range PCR (LR-PCR) reactions were performed under the following conditions: 95°C for 2 min, 35 cycles at 95°C for 20 sec, various annealing temperatures (60°C to 71°C) for 10 sec, 70°C for 2 min, an additional extension time at 72°C for 2 min and cooling at 4°C for 15 min.

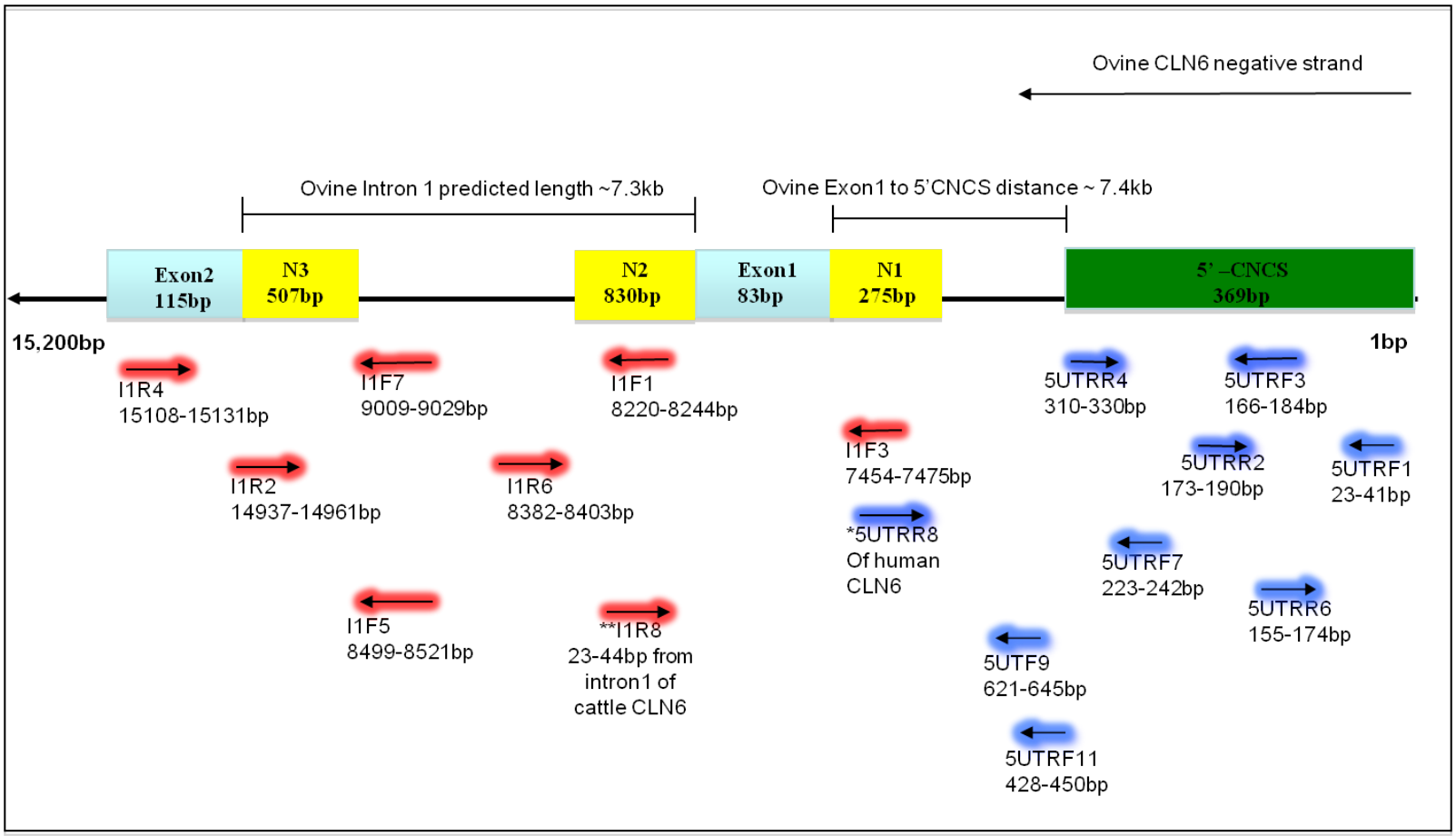


Figure 4.2: Schematic diagram of approximately 15.2 kb of the *CLN6* gene transcript in accordance with the ovine and bovine genome browsers. The length of intron 1 and distance between 5'-CNCS and exon 1 are predicted based on the bovine Btau4.0 genome assembly. The blue and yellow boxes represent previously known ovine sequences which are published exons and non-coding sequence (N1-N3) provided by collaborator Nadia Mitchell. The green box represents the 369 bp 5'-CNCS identified using VISTA. Primers generated to amplify 5'- CNCS are represented by blue arrows and the primers generated to amplify intron 1 of ovine *CLN6* by red arrows. The positions of primers are based on a predicted length of ~7.4 kb for the distance between 5'-CNCS and exon1 and of ~7.3 kb for intron1. The symbols * and ** represent primers designed from 5' sequence upstream of the *CLN6* start codon in the human genome and intron 1 sequence from bovine *CLN6*, respectively.

4.2.3 DNA sequencing

Sequencing was performed using either PCR products as DNA templates for bidirectional sequencing, or direct sequencing of BAC DNA 35C9, both using a single primer. Sequencing was conducted at a commercial sequencing facility (SUPAMAC, Sydney University) and at the University of Sydney using an ABI 3730 (Applied Biosystems, USA) and an ABI 3130 (Applied Biosystems, USA) genetic analyser, respectively.

4.2.3.1 Bidirectional sequencing of PCR products

Bidirectional sequencing was performed using PCR products amplified with primer combinations 5UTRF1/ 5UTRR4, I1F1/ I1R2, I1F3/ I1R4, I1F7/ I1R6, I1F5/ I1R4, I1F7/ I1R4 and I1F7/ I1R2 according to the protocol described in Chapter 2.7.1. In some instances where non-specific PCR bands were revealed by agarose gel electrophoresis, the bands of interest were excised and purified (Chapter 2.6.1). Sample 5UTRF1/ 5UTRR4 was sequenced in-house (Chapter 2.7.1). The remaining samples comprised of 16 µl mixtures containing purified PCR products (IIF1/ IIR2, IIF3/ IIR4, I1F7/ I1R6, I1F5/ I1R4, I1F7/ I1R4 and I1F7/ I1R2), 4 pmol primers (forward or reverse) and molecular grade water (Eppendorf, Hamburg, Germany) were sent at room temperature to SUPAMAC for sequencing.

4.2.3.2 Direct sequencing method

Direct sequencing of the ovine BAC clone was performed in-house with primers 5UTRR4, 5UTRR6 and 5UTRF7 as described previously (Chapter 2.7.2). Samples from primers 5UTRR8, 5UTRF9, 5UTRI1F11 and I1R8 were sent to SUPAMAC for sequencing.

4.2.4 Alignment and mapping of sequence to genome assemblies

New sequences generated from this study, and known ovine sequences from both published and unpublished sources (Genbank and Nadia Mitchell, Lincoln University, New Zealand, respectively) were mapped to genome assembly versions that were available when the current study began in 2007. The assemblies were *Bos taurus* Baylor draft assembly or Btau 4.0 (released in 2007) and Ovis aries ISGC draft assembly version 2.0 or OARv2.0 (released in 2010) accessible via the UCSC genome browser (http://genome.ucsc.edu/cgibin/hgGateway?hgsid=201629225&clade=mammal&org=Cow&db=0) and the GBrowse interface hosted at the CSIRO's Livestock Genomics site (http://www.livestockgenomics.csiro.au/cgi-bin/gbrowse/oarv2.0/), respectively. The bovine chromosome 10 (BTA10) of Btau 4.0 was used because of its homology with the ovine chromosome 7 (OAR7) of OARv2.0 containing the ovine *CLN6*.

Sequence alignments were performed using BLAT (Kent, 2002) and BLAST (Altschul et al., 1990) analysis tools available on the respective browsers websites mentioned above. Sequence information was uploaded into the genome browsers in general feature format (GFF) and big bed (BED) data files based on results from BLAT analysis against BTA10 and BLAST analysis against OAR7, respectively. Sequences were presented on the respective genome browsers using the following user custom browser tracks: known ovine *CLN6* sequence (oCLN6), CNCS sequence generated in the 5'-CNCS (5primeoCLN6) and sequence generated in intron 1 of *CLN6* (I1oCLN6).

4.3 Results

4.3.1 PCR primers for amplifying 5'CNCS and intron 1 of ovine CLN6

A total of seventeen primers (9 for the 5'-CNCS and 8 for intron 1) were generated using the *CLN6* sequence obtained from the cross species sequence alignment as well as from the known ovine, bovine and human sequences. As sequence information in the upstream of ovine CLN6 and in intron 1 was limited, the positions of the primers were predicted from the homologous

region on BTA10. A schematic diagram containing the predicted location of primers and flanking sequence from published exons 1 and 2 of ovine CLN6 (Tammen et al., 2006) and sequence provided by Nadia Mitchell (Lincoln University, New Zealand, pers. comm.) is presented (Figure 4.2).

4.3.2 PCR products for sequencing

Only one combination of primers derived from cross-species alignment in the 5'-CNCS generated a PCR product suitable for sequencing using the ovine BAC DNA as a template. The primer combination 5UTRF1/ 5UTRR4 initially generated three non-specific PCR products (approximately 200 - 300 bp) at annealing temperatures (T_a) 50°C and 52°C, respectively (Figure 4.3). A single PCR product of approximately 200 bp was then produced when T_a was increased to 55°C (Figure 4.3). The resultant sequence using this PCR product was used to create ovine specific primers.

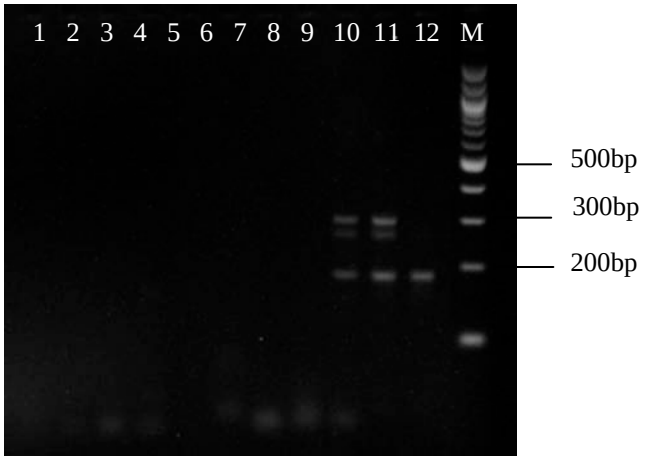


Figure 4.3: Optimisation of PCR conditions for primers derived from cross-species alignment for 5'-CNCS. Each primer combination was tested at three annealing temperatures: 50°C, 52°C and 55°C**.** (M) New England BioLabs 100 bp DNA Ladder, (1-3) 5'UTRF3/ 5'UTRR2, (4-6) 5'UTRF3/ 5'UTRR4, (7-9) 5'UTRF1/ 5'UTRR2, (10-12) 5'UTRF1/ 5'UTRR4 (1% agarose gel).

For intron 1, six of the 12 tested primer combinations generated PCR products suitable for sequencing. Combinations I1F1/ I1R2, I1F3/ I1R4, I1F7/ I1R6, I1F5/ I1R4, I1F7/ I1R4 and I1F7/ I1R2 generated PCR products of approximately 7 kb, 1.3 kb, 700 bp, 2 kb, 2 kb and 2 kb, respectively (Figure 4.4a and b). Three primer combinations (I1F5/ I1R4, I1F7/ I1R4, I1F7/ I1R2) produced multiple bands, which were extracted from agarose gels prior to sequencing (Figure 4.5). Most of these PCR products were shorter than predicted from the known bovine sequence information, but, as the true length of ovine intron 1 was unknown when the study was conducted; all purified PCR products were sequenced.

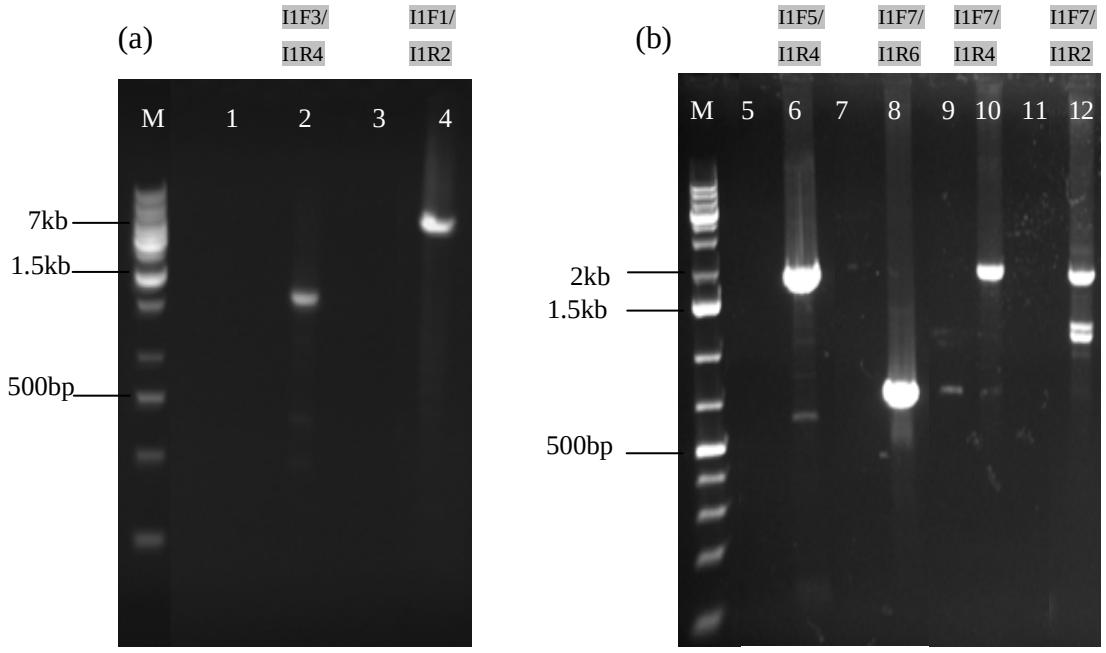


Figure 4.4: PCR products for intron 1 of ovine *CLN6* amplified with primers (a) I1F1/ I1R2, I1F3/ I1R4 and (b) I1F7/ I1R6, I1F5/ I1R4, I1F7/ I1R4 and I1F7/ I1R2. (M) Fermentas GeneRuler 1 kb DNA ladder plus, (1, 3, 5, 7, 9 and 11) no template control, (2) I1F3/ I1R4 (~1.3 kb), (4) I1F1/ I1R2 (~7 kb), (6) I1F5/ I1R4 (~2 kb), (8) I1F7/ I1R6 (~700 bp), (10) I1F7/ I1R4 (~2 kb), (12) I1F7/ I1R2 (~2 kb, ~1.3 kb) (1% agarose gel).

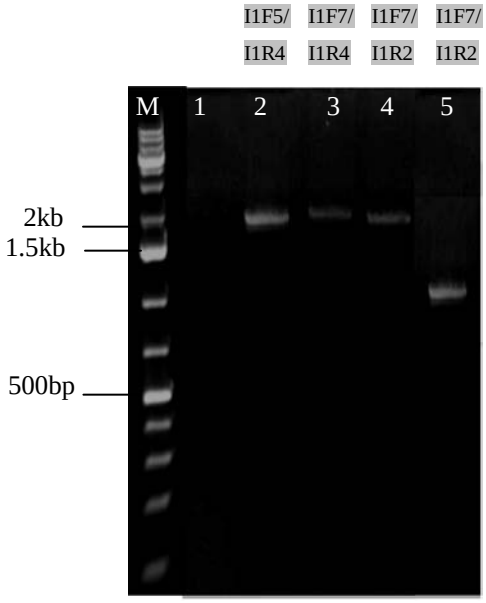


Figure 4.5: Agarose gel purified PCR products from intron 1 of ovine *CLN6* amplified with primers I1F5/ I1R4, I1F7/ I1R4 and I1F7/ I1R2. (M) Fermentas GeneRuler 1 kb DNA Ladder Plus, (1) no sample, (2) I1F5/ I1R4 (2 kb), (3) I1F7/ I1R4 (2 kb), (4,5) I1F7/ I1R2 (2 kb, 1.3 kb) (1% agarose gel).

4.3.3 Sequences generated in the 5' region and intron 1 of ovine *CLN6*

The sequencing of CNCS in the 5' region of ovine CLN6 was conducted in two phases. In phase 1, four primers, namely 5UTRF1, 5UTRF3, 5UTRR2 and 5UTRR4, designed from cross-species alignment, were used for direct sequencing of the BAC and PCR amplification. Only primer 5UTRR4 successfully generated sequence (232 bp) while PCR product using primers 5UTRF1/ 5UTRR4 generated a 195 bp sequence. In phase 2, the ovine specific primers 5UTRR6, 5UTRF7, 5UTRF9, 5UTRF11 and the human specific primer 5UTRR8 were used for direct sequencing of the ovine BAC, generating 635, 869, 846, 859 and 1028 bp, respectively. In summary, there was a total of 4, 664 bp sequence generated in the 5' region of ovine *CLN6*.

Sequencing of intron 1 of ovine CLN6 was conducted using six PCR products derived from long-range amplification. Only a segment of each PCR product was obtained in a single sequence read. For five of the products only one of the two primers used for each sequencing reaction generated sequence. The PCR products I1F1/ I1R2, I1F3/ I1R4, I1F7/ I1R6, I1F5/ I1R4, I1F7/ I1R4 and I1F7/ I1R2 generated 950, 800, 627, 757, 471 and 886 bp sequence, respectively. Direct sequencing of the bovine specific primer I1R8 generated an 880 bp sequence. In summary, there were a total of 5,371 bp sequences generated in intron 1 of ovine *CLN6*.

Individual raw sequences generated are presented in an appendix to this work (Appendix 3).

4.3.4 Alignment of sequences against BTA10

New non-coding sequences generated, and the previously known ovine *CLN6* sequences (exons 1 to 7: GenBank GeneID: DQ458790.1), 275 bp of 5' sequence, 1,337 bp of intron 1, as well as the sequence in introns 2 to 6 (provided by Nadia Mitchell, Lincoln University, New Zealand: pers. comm.) were aligned against bovine BTA10 using the BLAT analysis tool. The quality (percentage identity and scores based on web-based BLAT calculations) and location on BTA10 of the strongest BLAT hits for all query sequences are shown in tabular (Table 4.4) and visual form (Figure 4.6).

Table 4.4: Location and quality of the strongest BLAT hits of sequences against bovine BTA10.

	Chromosome	Score	Percentage identity	Location	
				From	to
CNCS sequences					
5UTRF1R4seq	10	171	96.8	14890986	14891287
5UTRR4seq	10	195	94.6	14891193	14891416
5UTRR6seq	10	406	96	14890588	14891041
5UTRF7seq	10	602	94.4	14890627	14891307
5UTRF9seq	10	733	95.6	14891067	14891901
5UTRF11seq	10	768	95.7	14890865	14891725
5UTRR8seq	10	771	91.3	14892668	14893619
I1R8seq	10	522	93.7	14876590	14877488
I1F1R2seq	10	782	93.8	14882541	14883486
I1F3R4seq	10	666	94	14882714	14883493
I1F7R6seq	10	491	92.6	14882687	14883289
I1F5R4Rseq	10	316	92.6	14876412	14876791
I1F5R4Fseq	10	222	91.7	14882477	14882766
I1F7R4seq	10	232	91.2	14879022	14879344
I1F7R2Fseq	10	267	95.2	14882972	14883289
I1F7R2Rseq	10	290	93.5	14877109	14877464
Known *CLN6* sequences					
Exon 1	10	75	95.2	14883808	14883890
Exon 2	10	111	98.3	14876348	14876462
Exon 3	10	97	99	14873865	14873963
Exon 4	10	185	99	14872393	14872581
Exon 5	10	56	100	14872061	14872116
Exon 6	10	117	97.6	14870546	14870668
Exon 7	10	265	98.9	14868505	14868775
Intron 2	10	2013	92.3	14873964	14876340
Intron 3	10	1065	92	14872582	14873849
Intron 4	10	231	92.8	14872117	14872392
Intron 5	10	1084	92.2	14870669	14872060
Intron 6	10	1466	91.8	14868776	14870545
Partial_3UTR	10	1052	93	14867033	14868504
Partial_5UTR	10	187	89.8	14883883	14884106
Partial_I1A	10	528	95.6	14883173	14883884
Partial_I1B	10	327	94.5	14876463	14877129

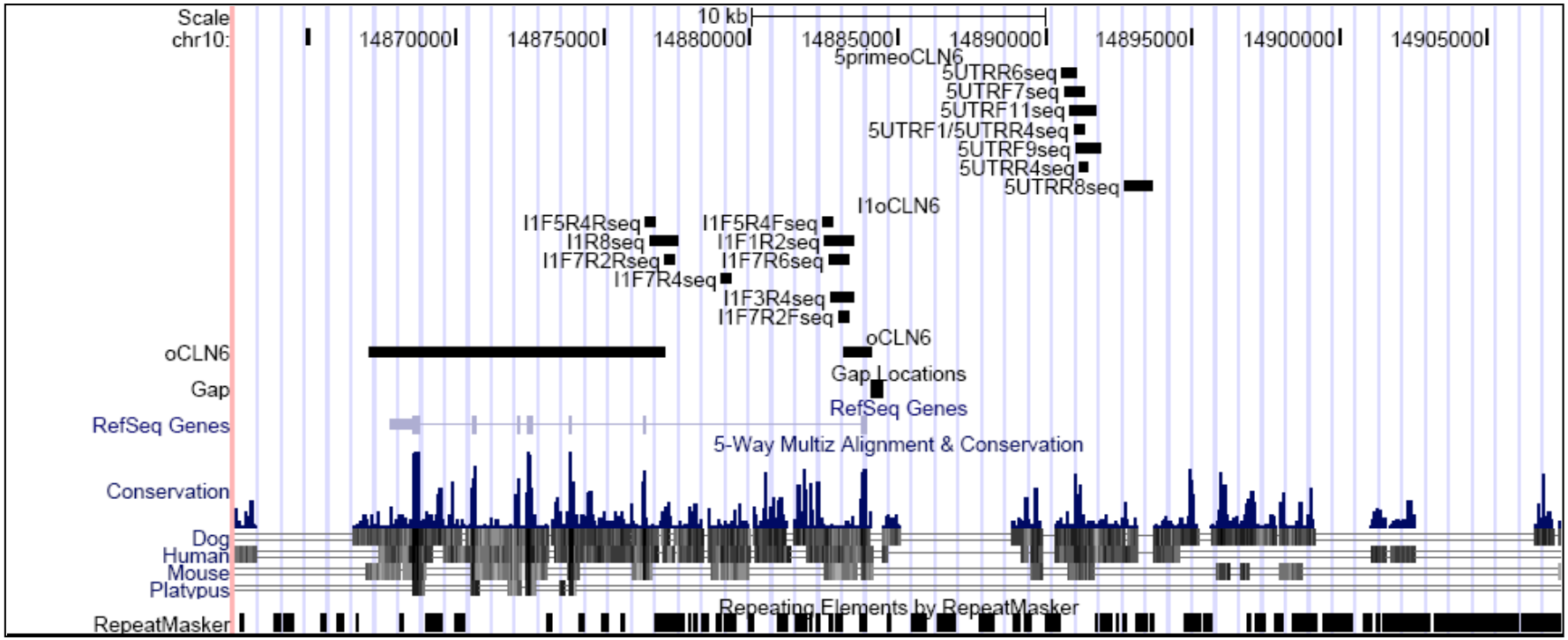


Figure 4.6: Alignment of known ovine sequences (track 'oCLN6', set to dense mode) and new non-coding sequences (track '5primeroCLN6' and 'I1oCLN6' are set to 'pack' mode) against BTA10: 14,862,499 - 14,907,500 (assembly Btau 4.0) using the UCSC genome browser. The blue shaded regions (represented by track 'Conservation') are sequence conserved between dog, human, mouse and platypus. The locations of the seven exons of bovine *CLN6* and bovine repeats are shown in the track 'RefSeq Genes' and 'RepeatMasker', respectively. A gap in the bovine sequence exists 5' of bovine *CLN6*, as is shown in the 'Gap' track.

Known ovine sequences (‘oCLN6’**)** were aligned as expected against bovine *CLN6* (Figure 4.6) (‘RefSeq Genes’ track). The *CLN6* BTA10 coding sequence is on the negative strand.

The ‘conservation’ track showed the presence of conserved sequences in the human, mouse, dog and platypus, which further support identification of CNCS in the 5’ upstream of *CLN6* (Chapter 3). The distance between the 5’-CNCS and the *CLN6* start codon in exon 1 is approximately 5 kb in cattle BTA10 and 2 kb in human HSA15. Most of the sequences generated for the 5’-CNCS (‘5primeoCLN6’ track) aligned with and extended the 5’-CNCS region as expected, to BTA10: 14,890,588 - 14,891,901. Sequences in this region encompassed the overlapping sequences of ‘5UTRR6seq’ (direct sequencing with the 5UTRR6 primer), ‘5UTRF7seq’ (direct sequencing with the 5UTRF7 primer), ‘5UTRF11seq’ (direct sequencing with the 5UTRF11 primer), ‘5UTRF1/ 5UTRR4’ (PCR primer combinations 5UTRF1/ 5UTRR4), ‘5UTRF9seq’ (direct sequencing with the 5UTRF9 primer) and ‘5UTRR4seq’ (direct sequencing with 5UTRR4 primer). However, the sequence ‘5UTRR8seq’ unexpectedly aligned to BTA10: 14,892,668 - 14,893,619 approximately 7 kb upstream of the expected region, and not in the proximity of the predicted primer location in the 5’-CNCS region (Figure 4.2).

Although most of the PCR products generated for intron 1 of ovine *CLN6* were shorter than expected (Figures 4.4 and 4.5) as compared to the 7 kb predicted based on BTA10, these generated sequences aligned to three different regions in the bovine intron 1 (‘I1oCLN6’ track). The first sequence which aligned to BTA10:14,876,412 - 14,877,464 encompassed the overlapping sequences of ‘I1F5R4_Rseq’ (PCR primer combination I1F5/ I1R4, sequencing primer I1R4), ‘I1R8seq’ (direct sequencing with I1R8 primer) and ‘I1F7R2_Rseq’ (PCR primer combination I1F7/ I1R2, sequencing primer I1R2). The second sequence: ‘I1F7R4’ (PCR primer combination I1F7/ I1R4, sequencing primers I1F7 and I1R4) aligned to BTA10:14,879,022 - 14,879,344. The third sequence which encompassed the overlapping sequences of ‘I1F5R4_Fseq’ (PCR primer combination I1F5/ I1R4, sequencing primerI1F5), ‘I1F1R2seq’ (PCR primer combination I1F1/ I1R2, sequencing primers I1F1 and I1R2) , ‘I1F7R6_seq’ (PCR primer combination I1F7/ I1R62, sequencing primers I1F7 and I1R62), ‘I1F3R4seq’ (PCR

primer combination I1F3/ I1R4, sequencing primer I1R4) and 'I1F7R2_Fseq' (PCR primer combination I1F7/ I1R2, sequencing primer I1F7) aligned to BTA10: 14,882,477 - 14,883,289.

4.3.5 Alignment of sequences against OAR7

After the initial BLAT analysis against the bovine sequence, and while the investigation was underway, the International Sheep Genome Consortium (ISGC) published a draft ovine genome sequence (OARv2.0; Archibald et al., 2010). This, along with the 5' and intron 1 sequences generated here, as well as known ovine *CLN6* sequences were aligned and mapped against ovine OAR7 using results from BLAST analysis (http://www.livestockgenomics.csiro.au/cgi-bin/gbrowse/oarv2.0/). The quality (Expect value {E value} and score) and the location of the best BLAST hits containing the lowest E values are shown below in tabular (Table 4.5) and visual (Figure 4.7) form. The E-value gives an indication of the number of hits one can expect to see by chance (the lower the E-value the more significant the hit) and the score (bit) indicates how good an alignment is (the higher the score the better the alignment).

Table 4.5: Location and quality of the strongest BLAST hits of sequences against ovine OAR7.

CNCS sequences	Chromosome	Score (bits)	E-value	Location	
				From	To
5UTRF1R4	7	319 bits (161)	3e'-85	14858684	14858867
5UTRR4seq	7	422 bits (213)	e-116	14858894	14859114
5UTRR6seq	7	817 bits (412)	0	14858742	14858310
5UTRF7seq	7	1124 bits (567)	0	14859012	14858334
5UTRF9seq_p1	7	1158 bits (584)	0	14858770	14859362
5UTRF9seq_p2	7	402 bits (203)	e-109	14859389	14859597
5UTRF11seq	7	1556 bits (785)	0	14858566	14859362
5UTRR8seq	7	1675 bits (845)	0	14861335	14860356
I1F1R2	7	1725 bits (870)	0	14851333	14850389
I1F3R4	7	1473 bits (743)	0	14850568	14851340
I1F7R6	7	1102 bits (556)	0	14851136	14850534
I1F5R4R	7	541 bits (273)	e-152	14850622	14850326
I1F5R4F	7	688 bits (347)	0	14845562	14845952
I1F7R4	7	507 bits (256)	e-141	14847175	14846836
I1F7R2F	7	601 bits (303)	e-169	14851136	14850818
I1F7R2R	7	666 bits (336)	0	14846008	14846352
I1R8seq	7	1181 bits (596)	0	14845732	14846419
Known *CLN6* sequences					
Exon 1	14	36.2	2.1	53893703	53893720
Exon 2	7	228 bits (115)	4e-58	14845605	14845491
Exon 3	7	196 bits (99)	1e-48	14842069	14841971
Exon 4	7	375 bits (189)	e-102	14840667	14840479
Exon 5	7	111 bits (56)	3e-23	14840201	14840146
Exon 6	7	188 bits (95)	4e-46	14838319	14838225
Exon 7	7	484 bits (244)	e-134	14836413	14836162
Intron 2_a	7	1739 bits (877)	0	14844614	14845490
Intron 2_b	7	1223 bits (617)	0	14843924	14844540
Intron 2_c	7	543 bits (274)	e-152	14843479	14843752
Intron 2_d	7	1125 bits (567)	0	14842906	14843478
Intron 2_e	7	1618 bits (816)	0	14842905	14842070
Intron 3	7	2493 bits (1254)	0	14841970	14840668
Intron 4_p1	7	137 bits (69)	3e-30	14840270	14840202
Intron 4_p2	7	131 bits (66)	2e-28	14840478	14840413
Intron 5_p4	7	549 bits (273)	e-153	14839563	14839259
Intron 5_p1	7	416 bits (207)	e-113	14838575	14838320
Intron 5_p3	7	410 bits (204)	e-112	14839198	14838971
Intron 5_p5	7	304 bits (151)	1e-79	14840145	14839995
Intron 5_p2	7	211 bits (105)	7e-52	14838828	14838724
Intron 6	7	3289 bits (1658)	0	14838086	14836414
Partial_3UTR_p1	7	1861 bits (939)	0	14834948	14835904
Partial_3UTR_p2	7	373 bits (188)	e-100	14835906	14836136

Known *CLN6* sequences	Chromosome	Score (bits)	E-value	Location	
				From	To
Partial_5UTR	4	38.2 bits (19)	2.1	46088674	46088696
Partial_I1A1_p1	7	412 bits (207)	e-112	14851406	14851014
Partial_I1A1_p2	7	239 bits (120)	2e-60	14851526	14851407
Partial_I1B	7	817 bits (409)	0	14846034	14845513

In Table 4.5, 'Intron 2' which is 3,476 bp in length was manually subdivided into several smaller parts noted by the addition of label '_a' – '_e' after each sequence name. These were used as inputs for BLAST analysis to overcome a 2 kb limit for input sequences on the CSIRO server and problems with unwanted numerous repeats found during BLAST searches, which delayed and often halted analysis. Sequences '5UTRF9seq', 'Intron 4', 'Intron 5', 'Partial_3UTR' and 'Partial_I1A1' were subdivided during BLAST analysis by the addition of labels '_p1' – '_p5' after each sequence name, as there were gaps in the sequence alignment.

Most of the known ovine sequences (except for 'exon 1' and the 'partial_5UTR') aligned as expected (Figure 4.7) to the predicted *CLN6* location ('Cow Refseqs' track). BLAST analysis aligned exon 1 to chromosome 14 (OAR14:53,893,703 - 53,893,720 bp, Table 4.5) but the high E-value of 2.1 indicates that this sequence is likely to be incorrectly assigned. The other unassigned sequence: 'partial_5UTR' aligned to multiple regions across many chromosomes and aligned less than 20 bp at a high E-value of 2.1. The failure to align these two known sequences to the predicted region on OAR7 could have arisen because the version used in the analysis (oarv2.0) contains a gap within the scaffold (scaffold476), located between 'contigOAR.1099' and 'contigOAR7.110' and spanning approximately 900 bp (OAR7:14,851,500 - 14,852,400 bp), which is the region expected to contain exon 1 of *CLN6*. This also likely caused the 'Cow_RefSeqs' track (Figure 4.7) to not display the position of exon 1 in bovine *CLN6*. For visualisation purposes, positions of the previously known sequences 'exon 1' and 'partial_5UTR' were manually added to the genome browser at location OAR7:14,851,527 - 14,851,610 and OAR7:14,851,611 - 14,851,886, respectively, using flanking sequences to predict these locations.

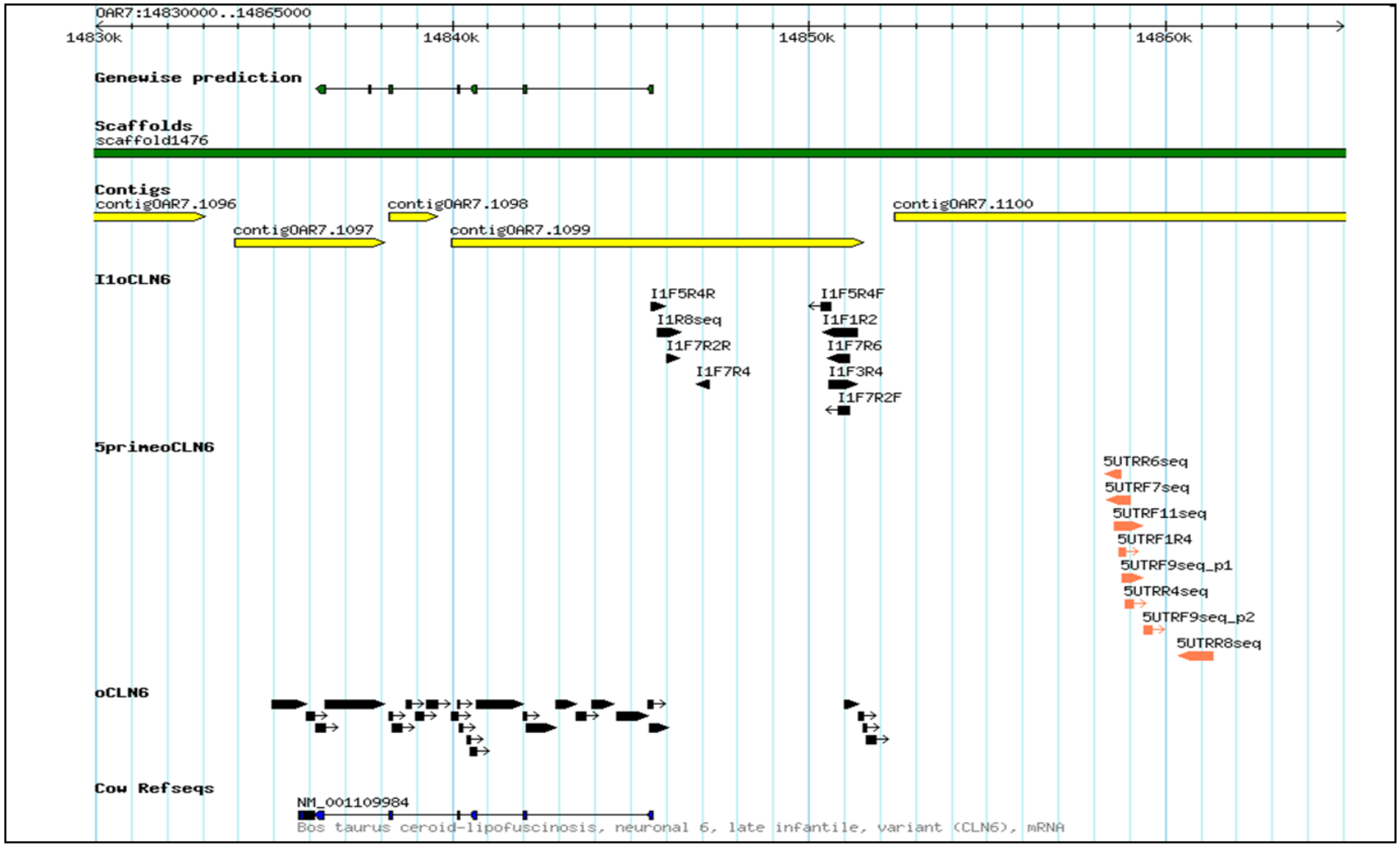


Figure 4.7: Alignment of known ovine sequences (track 'oCLN6') and new non-coding sequences (tracks: '5primeroCLN6' and 'I1oCLN6') against OAR7: 14,830,000-14,865,000 (assembly OARv2.0) using the CSIRO genome browser (http://www.livestockgenomics.csiro.au/cgi-bin/gbrowse/oarv2.0/). 'Scaffold476' (represented by track 'Scaffolds') is the scaffold assembled using four different contigs: 'contigOAR7.1097, 'contigOAR7.1098', contigOAR7.1099 and contigOAR7.1100 (represented by track 'Contigs'). Sequence gaps exist between all contigs in the ovine genome assembly. The predicted location of the seven exons of ovine *CLN6* are shown in the track 'Genewise prediction' and the location of the six exons (exons 2-7) of bovine *CLN6* (NM_001109984) are shown in the track 'Cow RefSeqs'.

The overall positioning of the newly generated sequences in the ovine genome corresponded closely with their alignment to the bovine genome discussed earlier ('Cow Refseqs' track). Sequences for the 5'UTR ('5primeUTR' track**)** encompassed the overlapping sequences of '5UTRR6seq', '5UTRF7seq', '5UTRF11seq', '5UTRF1/ 5UTRR4', '5UTRF9seq_p1' (part 1 of the '5UTRF9' sequence generated), and '5UTRR4seq' aligned approximately to OAR7: 14,858,310 - 14,859,114. Sequence '5UTRF9seq_p2' (part 2 of the '5UTRF9' sequence generated) aligned to 14,859,389 - 14,859,597. As in the bovine alignment, sequence '5UTRR8seq' unexpectedly aligned to a region not in the proximity of the predicted primer location (Figure 4.2) at OAR7: 14,860,356 - 14,861,335 bp which is upstream of the expected 5'-CNCS region.

Sequences generated for intron 1 aligned to three different regions of ovine intron 1 (I1oCLN6): The first sequence encompassing the overlapping sequences of 'I1F5R4_Rseq', 'I1R8seq' and 'I1F7R2_Rseq' aligned to OAR7: 14,845,562 - 14,846,352. The second sequence: 'I1F7R4' aligned to OAR7: 14,846,836 - 14,847,175 and the third sequence encompassing the overlapping sequences of 'I1F5R4', 'I1F1R2seq', 'I1F7R6_seq', 'I1F3R4seq' and 'I1F7R2_Fseq' aligned to OAR7: 14,850,326 - 14,851,340.

A total of 10,035 bp ovine sequence was generated using Sanger sequencing for intron 1 and upstream of ovine *CLN6*. Approximately 1,450 bp was new ovine sequence that did not overlap with previously known ovine sequence, but aligned well with bovine BTA10 (at the time of sequencing in 2008). Furthermore, the sequence generated was used to fill and reduce the gap in the ovine OARv2.0 genome assembly.

4.4 Discussion

From the five conserved non-coding sequences (CNCS) identified in previously described work (Chapter 3), two regions (intron 1 and upstream of ovine *CLN6)* were partially sequenced in this study. Although the sequencing efforts were laborious, time-consuming and generally

unsuccessful, new sequence was generated. Of the sequence generated about 10% was novel and was used to fill and reduce a gap in the ovine OARv2.0 genome assembly.

Most of the non-ovine specific primers aligned to expected regions in OAR7 however three primers (5'UTRF1, 5'UTRR2 and 5'UTRF3) located upstream of ovine *CLN6* showed multiple BLAST hits (Figure 4.7) across numerous ovine chromosomes. It is likely that these primers were designed from regions containing repeat sequences, based on their high E values (expected number of hits) of up to 385. However, short sequences also have relatively high E values because these sequences possess a higher probability of occurring in the database purely by chance (Altschul et al., 1990) as compared to smaller E values which indicate lower probability of false positives (Matsuda and Fukusaki, 2013). Only primer 5'UTRR4 was successful in generating sheep sequence by direct sequencing, which is likely due to it aligning well to ovine chromosome 7 with a score of 42.1 and an E value of 0.002.

The I1F7 and I1R6 primer pair is actually divergent primers and was chosen by accident. However, its PCR product of 700bp (Figure 4.4b) which mapped on to the ovine and bovine genomes (Figure 4.6) fell within the highly masked region as shown by the RepeatMasker track. It was most likely that the primers amplified repetitive DNA that mapped to the extremely enriched repeat sequence within ovine intron 1.

Two computational tools (BLAST and BLAT) were used to analyse the alignment between the CNCS sequence generated, the known ovine sequence and the bovine and ovine genomes. BLAT performance was fast, using index derived from the assembly of the entire genome in memory instead of whole genome data, but was unable to process sequence with less than 95% sequence homology and was limited in relation to batch queries with 25 sequences or less (Kent, 2002). BLAST analysis required longer computational time using whole genome information on the server. There was a high tendency of the analysis to break down but it allowed repeat regions and sequences with low homology to be analysed (Altschul et al., 1990). A combination of both alignment tools successfully mapped sequences to the genome of interest.

Several problems were encountered during this study. Firstly, there was only limited ovine sequence available for generating primer sequences in the CNCS regions of interest when the study commenced in 2007. Thus cross-species and bovine sequence information was used, which meant that only a limited number of primers worked. Secondly, the bovine early genome version was incomplete, thus parts of the regions targeted for amplification contained ambiguous nucleotides. Thirdly, the ovine genome assembly was far from complete during this study, as OARv2.0 was considered a working draft release with the more accurate version OARv3.0 to be released later in 2011 (Archibald et al., 2010).

In hindsight, the primer walking strategy of sequencing short fragments of sequences at a time from a long sequence, thus 'walking' from a known region to an unknown region method would have been an alternative to complete sequencing the non-coding regions (Chinault and Carbon, 1979). Another approach would have been to subclone the BAC containing the region of interest and sequencing the BAC DNA (Quail et al., 2011). However, all of these methods are laborious and costly for a relatively small region.

In conclusion, the generation of non-coding sequence using this approach was laborious and time consuming and not very effective. With the development of access to next generation sequencing (NGS) technology, the Sanger sequencing approach was abandoned and NGS utilised for the following studies (Chapters 5, 6 and 7).

CHAPTER 5: OVINE BACTERIAL ARTIFICIAL CHROMOSOME (BAC) SEQUENCING USING A NEXT-GENERATION SEQUENCING (NGS) 454 PYROSEQUENCING PLATFORM

5.1 Introduction

The previous chapters of this study describe how five conserved non-coding sequence (CNCS) regions were identified using an *in silico* approach, and how traditional Sanger sequencing based methods failed to adequately amplify these regions. This work generated 1.5 kb sequence of CNCS within and upstream to the ovine *CLN6*. However the sequence in this region was still far from complete. The current chapter describes how the ovine BAC clone 270H8 containing the regions of interest was sequenced using a next-generation sequencing (NGS) Roche 454 pyrosequencing platform. The regions of interest were *CLN6* and flanking Calmodulin-like protein 4 gene or CALML4 gene (*CALML4*).

5.2 Materials and methods

5.2.1 Ovine bacterial artificial chromosome (BAC) clones

5.2.1.1 Characterisation of ovine BAC clones

Ovine BAC clones 270H8 and 35C9 were obtained from Dr. Daniel Vaiman (INRA, Jouy-en-Josas, France) after screening of an ovine BAC library with *CLN6* specific primers as previously described (Chapter 2.3) and DNA of these two clones was extracted using methods outlined in

Chapter 2.3.1. Both BAC clones were screened for *CLN6* including a CNCS probe located approximately 2 kb upstream to the gene (Chapter 3) and downstream *CALML4*. PCR primers were designed using sequences from GenBank (GeneID: 539277), International Sheep Genomics Consortium (ISGC) and CNCS sequence generated in an earlier study (Chapter 4).

PCR to amplify the region spanning CNCS to *CLN6* (CNCS-*CLN6)* was performed using the KOD Hot Start DNA Polymerase kit (Merck Biosciences Limited, Australia). PCR conditions for amplification of CNCS-*CLN6* were as follows; a total reaction volume of 20 µl containing 10X PCR Buffer for KOD Hot Start DNA Polymerase (Merck Biosciences Limited, Australia), 1mM MgS04, 0.2 mM KOD Hot start DNA Polymerase supplied dNTP, 10 pmol of primers 5UTRF9rc and I1R6 (Table 5.1), 1U of KOD Hot start DNA Polymerase and 50 ng of BAC DNA. PCR was performed under the following conditions: polymerase activation at 95°C for 2 min, 40 cycles of denaturation at 95°C for 20 sec, annealing at 67° C for 10 sec and extension at 70°C for 2 min, followed by a final additional extension at 72°C for 1 min.

PCR to amplify the region spanning exons 2 to 5 of *CALML4* was designed for standard PCR amplification using the Qiagen Hot Start DNA Polymerase kit (Qiagen, Hilden, Germany). PCR conditions for amplification of exons 2 to 5 of *CALML4* were as follows: a total reaction volume of 20 µl containing 10X Buffer 1.5mM MgCl (Qiagen, Hilden, Germany), 2.5 mM $MgCl_2$, 0.2 mM dNTP (Astral Scientific, Australia), 20 pmol of forward primer CALL2F1 and of the reverse primer 3UTRR8 (Table 5.1), 0.5 U of Qiagen Hot Start DNA Polymerase and 50 ng of BAC DNA. PCR was performed under the following conditions: polymerase activation at 95° for 15 min, 40 cycles of denaturation at 95° for 30 sec, annealing at 55° for 30 sec, and extension at 72° for 4 min followed by a final additional extension at 72° for 10 sec.

Table 5.1: Primer sequences for PCR identification of the *CLN6* and *CALML4* genes from BAC clones 270H8 and 35C9.

Primer	Direction	Sequence 5'>3'	Length (bases)	Region amplified
5UTRF9rc	forward	AGAGAATGAGATTGGGGTCAGAGT	24	*CNCS-CLN6*
I1R6	reverse	AACTAGGGTGAGGCAAGTGAGA	22	*CNCS-CLN6*
CALL2F1	forward	CGGTGGGGGTTTAGAGACA	19	*CALML4*
3UTRR8	reverse	GTGGTGTGTGACGTGCCTAA	20	*CALML4*

5.2.1.2 DNA extraction from cultured BAC clones

Culture and purification of BAC clones 270H8 and 35C9 were performed using methods outlined in Chapter 2.3.1.

5.2.1.3 Quantification and quality assessment of BAC DNA

DNA concentrations were determined using a UV spectrophotometer and gel electrophoresis; with a 2 µl DNA aliquot loaded into the NanoDropTM ND-1000 spectrophotometer (NanoDrop Technologies Inc, Thermo Scientific, USA) and 6 µl DNA run with a Geneworks 1 kb DNA ladder (Geneworks, Australia) on a 1% agarose gel. Quality assessment based on results from these two methods revealed that only DNA from BAC clone 270H8 reached the acceptable concentration and purity suitable for sequencing (Figure 5.2).

5.2.2 Sequencing of BAC clone

Three microgram of purified DNA from BAC clone 270H8 was sent at room temperature to a High Throughput DNA Sequencing Unit at University of Otago, Dunedin, New Zealand for 454 pyrosequencing (Goldberg 2006, Roche 454 sequencing technology website: http://454.com/products-solutions/how-it-works/index.asp). A sequencing library was constructed with the BAC sample nebulised into fragments of 300 to 800 bp in length. 454 sequencing adapters were ligated to the 3' and 5' ends of the fragments to create a template DNA library. A Roche Multiplex Identifier (MID) barcode (sequence: ACGAGTGCGT) was attached to the sample library during library construction to distinguish it from other unrelated samples run in the same sequencing reaction and sequenced with a GS FLX Genomic sequencer on a 1/16th equivalent of a LR70 FLX plate.

5.2.3 Bioinformatics analysis

The Roche GS FLX Genomic sequencer produced outputs in standard flowgram format (SFF) files which include sequence reads and quality scores for each read. Sequence reads from the SFF files were extracted and stored into FASTA using SFF_extract software (http://bioinf.comav.upv.es/sff_extract/index.html) and these sequences pre-processed by screening and automated trimming of the 454 sequence adaptors and primers using SeqClean (http://seqclean.sourceforge.net). Low quality sequence reads with an average quality of <20 for any part of the sequence within a window size of 50 bp and reads shorter than 100 bp were removed using Mothur software (www.mothur.org).

The remaining sequence reads were assembled into sets of sequence contigs using the MIRA whole genome shotgun and EST sequence assembler program (Chevreux et al., 2004, http://www.chevreux.org/projects_mira.html). MIRA assembly was conducted under default settings. Assembly was conducted on an 8 core 2.94 GHz Linux workstation with 96 GB of RAM. File extraction, bioinformatics clean-up prior to sequencing and assembly of the sequence reads conducted above were performed by Dr. Kyall Zenger.

Homology searches for these sequences were performed using BLAST (Altschul et al., 1990) program available on the National Center for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov/) in order to identify *E. coli* and *pBeloBACII* contamination, which was then removed manually. The contaminant-free contigs were aligned to ISGC sheep sequence version 1.5 using GeneDoc (Nicholas et al., 1997) as well as to bovine chromosome 10 (BTA10) corresponding to the region of interest on sheep chromosome 7 (OAR7) using the Ensembl Genome Browser (http://www.ensembl.org/; Hubbard et al., 2002). Detection of repeat elements was carried out using the Repeatmasker program (Smit et al., *RepeatMasker Open-3.0*. 1996 - 2010 http://www.repeatmasker.org) against species Bovidae Btau4.0 as a reference repeat database. Most of the bioinformatic analysis, including BLAST analysis and repeatmasking of

the assembled sequence, was performed with the assistance of Drs Matthew Hobbs and Julie Cavanagh.

5.3 Results

5.3.1 Screening of BAC clones

Ovine BAC clones were screened using PCR amplification, which generated two PCR products for CNCS-*CLN6* and *CALML4* (Exons 2-5) of expected sizes 7 kb and 3 kb, respectively (Figure 5.1).

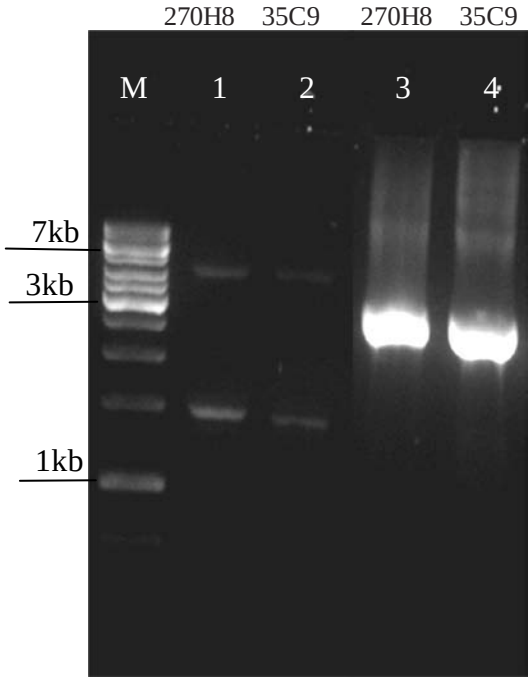


Figure 5.1: PCR amplification of CNCS-*CLN6* and *CALML4* regions using BAC clones 270H8 and 35C9 as DNA templates. (M) Geneworks 1kb DNA ladder, (1,2) CNCS-CLN6 PCR product(3,4) CALML4 PCR product. The so-called 'smiling effect' needs to be considered for band size matching.

Purified DNA yielded concentrations of 80 ng/µl and 54 ng/µl and OD_{260}/OD_{280} ratios of 1.65 and 1.35 for BAC clones 270H8 and 35C9, respectively. Multiple bands were present on the gel signifying the different forms of DNA: the supercoiled BAC DNA (lower band), relaxed BAC

DNA (midband) and a combination of nicked and damaged BAC DNA on the upper band (Figure 5.2).

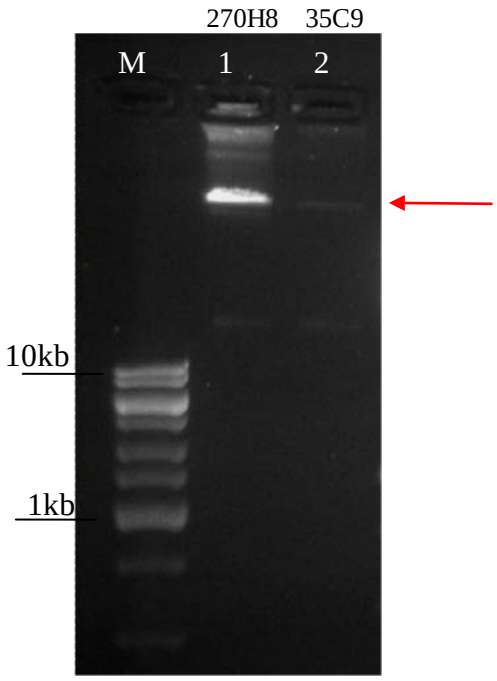


Figure 5.2: Quality assessment of ovine BAC DNAs for pyrosequencing. (M) Geneworks 1 kb Ladder, (1) BAC clone 270H8, (2) BAC clone 35C9. A 5 µl aliquot of purified genomic DNA from a 250 µl eluate was analysed by electrophoresis on a 1% agarose gel. The upper band shows damaged BAC DNA, the midband shows supercoiled relaxed pure BAC DNA for sequencing (pointed by red arrow) and lower band shows the supercoiled BAC DNA.

5.3.2 Sequence assembly

Sequencing of the ovine BAC clone 270H8 using the Roche 454 FLX instrument generated 2 million bases of reads with an average read length of 400 bp. These reads were assembled into 114 contigs at 13.49X fold coverage using MIRA default assembly parameters. This resulted in a cumulative size of 152,641 bases with sequence contig sizes ranging from 40 (contig NCL_c61) to 18863 bases (NCL_c3). There were 9,096 sequence reads used for assembly, 20 singletons and 15 unassembled reads. After assembly, contigs of low quality and singletons were removed before subsequent analysis. Raw data of the 114 contigs are supplied in Appendix 4.

There were 1,248 asterisk symbols (*) found throughout the sequence contigs which represented homopolymer (stretches of DNA of identical bases) errors. These were randomly located with only 20% occurring before or after stretches of A/T nucleotides, and were removed manually prior to sequence assembly. Once assembled, contigs in FASTA format were used as inputs into the BLAST program. BLAST hits identified 29 BAC contigs which were smaller than 200 bp and matched to the *pBeloBACII* BAC vector (Accesion no: U51113) and *E.coli* (GenBank GeneID: 12319) bacterial genomic DNA sequences. These contigs were also removed manually as they represent undesired BAC host vector and bacterial genomic DNA contamination leaving 85 contigs for assembly.

5.3.3 Gene content and representation in BAC sequence

Eighty-five contaminant-free BAC contigs were aligned against cattle BTA10 and visualised using the Ensembl genome browser (Figure 5.3). For each contig, positions of the strongest BLAST hits with more than 90% sequence identity were used for graphical representation on BTA10 spanning 160,000 kb between BTA10:14,800,000-14,960,000 (Figure 5.3). Contigs containing genes were annotated using sequences that matched with published bovine sequences: *CALML4* (GenBank GeneID: 539277), *FEM1b* (GenBank GeneID: 540252) and *PIAS1* (GenBank GeneID: 509231).

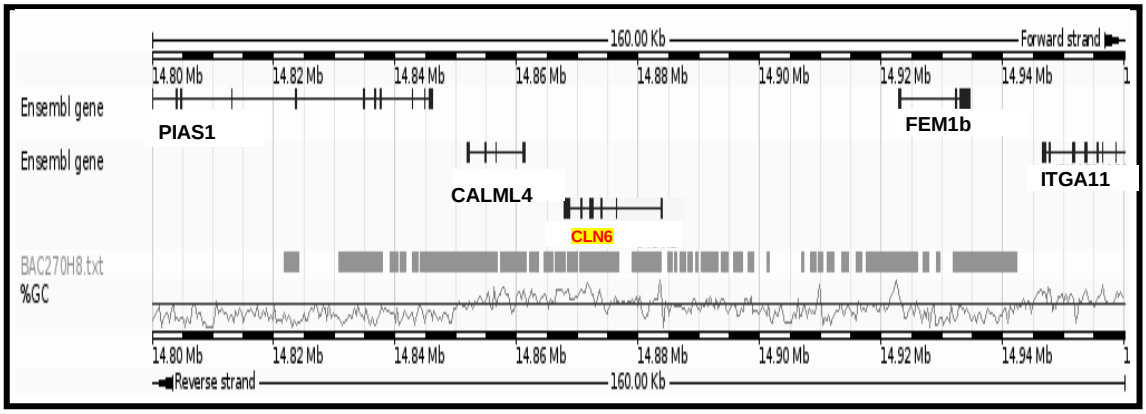


Figure 5.3: Alignment of the 454 sheep BAC 270H8 contigs against BTA10:14,800,000-14,960,000 (assembly Btau 4.0) using the Ensembl browser. The image is in reverse DNA strands. The *CLN6* (in red font and yellow highlight) is flanked by upstream genes *FEM1b* and *ITGA11* and downstream genes *CALML4* and *PIAS1*. The grey shaded regions spanning from *PIAS1* to just downstream of *ITGA11* represent sheep BAC contigs that match well to cattle reference sequence; unshaded areas in between show gaps in the sheep BAC contig sequences.

5.3.4 Multialignment and generation of consensus sequence

Approximately 110 kb of sheep BAC sequence was generated using the 85 contaminant-free BAC contigs with adjacent contigs aligned in multiple GeneDoc files. These were merged to produce a sequence assembly, furthermore referred to as 'initial sequence assembly'. This initial sequence assembly was annotated manually using results from the BLAST analysis against multiple reference sequences, these being the international sheep genomic sequence (ISGC) versions 1 and 2, the published *CLN6* sheep mRNA sequence (GenBank GeneID: 678673), the unpublished sheep sequences obtained at Lincoln University and the *CNCS* sequence generated earlier (Chapter 4). The initial sequence assembly was then used as a template to produce a final consensus sequence assembly referred to as the 'consensus sequence assembly' which represents the region of interest for the two upcoming sequencing approaches in Chapters 7 and 8.

The consensus sequence assembly was used as query in RepeatMasker to identify known repeats. A total of 40,211 bp (36.35%) of the original 110,618 bp sequences were masked, leaving a final sequence of 70,407 bases. Among the repeats identified, 15,624 and 16,317 bp were of short interspersed nuclear elements (SINE) and long interspersed nuclear elements (LINE), respectively, 4,127 bp were of long terminal repeats (LTR) and 3,020 bp were of DNA elements.

5.4 Discussion

The two ovine BAC clones 270H8 and 35C9 used for analysis in this study were screened for the CNCS region and the *CALML4* gene downstream to ovine *CLN6* prior to sequencing. This span of sequences was identified to be of interest due to the hypothesis that the disease causing mutation for ovine NCL in the South Hampshire sheep is in sequences flanking ovine *CLN6* (Chapter 1.3). Compared to the labourious Sanger sequencing method used in sequencing the CNCS (Chapter 4), the Roche 454 pyrosequencing method produced longer sequence reads averaging 400 bp than those generated by Sanger (Mardis, 2008; Zhou et al., 2010), thus allowing generation of a normal sheep genomic reference sequence. This reference sequence was used for mutation screening approaches (Chapter 6 and 7).

Purified DNA from ovine BAC clone 270H8 was chosen as the sequencing template instead of BAC clone 35C9. Clone 35C9 had a OD260/280 reading of 1.35 which suggested a high level of protein contamination. DNA purity is a critical factor for consideration (Liu et al., 2013) as it is likely to impede sequencing.

Agarose gel electrophoresis analysis of both BAC DNAs revealed the presence of different forms of DNA including the supercoiled BAC DNA, the purified BAC vector and an insert, which theoretically should have been removed during purification. The relaxed DNA mid band was the purified contaminant-free genomic DNA required for the sequencing project, whereas the slowest traveling DNA in the agarose gel (upper band) was the combination of nicked/sheared damaged DNA, and a smear of BAC DNA which does not hybridize with probes on the gDNA and migrate easily into the gel. The band closest to the well was likely to be an analysis artifact (H. Zhou pers. comm.). A better method for analysis and estimation of the size of large DNA construct such as BAC clones is to use a standard agarose with a supercoiled DNA ladder or to use pulse field gel electrophoresis (PFGE; Herschleb et al., 2007); neither of which were available in our laboratory at that time.

The DNA from BAC clone 270H8 was sent for sequencing after the high throughput DNA sequencing unit confirmed that the quantity and quality of the DNA was sufficient. In this case, sheared DNA and minor contaminations were not of concern. The DNA was destined to be fragmented and 'over sequenced' such that it had excessive sequence coverage for the region. Further information on the 454 sequencing chemistry has been described in Chapter 1.6.2.1.

Homopolymer length sequencing error is very common in 454 sequencing reads, constituting 39% of error rates, as stated by Huse et al. (2007). These errors occur due to the unique technique of sequencing for the 454 platform as nucleotide bases are not called directly as in Sanger sequencing but rely on the intensity of lumimenscence brightness emitted each time a nucleotide is added to the DNA strand (Mardis, 2008). Manual removal of these homopolymers prevents possible problems with sequence assembly, as the length variation can generate ambiguity when encountered causing long stretches of one or more nucleotides.

In finalising this assembly, there were circumstances when nucleotide bases varied between aligned sequences. To call the correct nucleotide base several conditions were followed to finalise consensus between ISGC, *CLN6* and *CALML4* published mRNA, unpublished genomic DNA and BAC sequencing. Only when these conditions were met was the particular nucleotide base called. In some regions the sequence may not have been fully accurate and in that particular situation the best sequence was called. There may be errors or miscalling of some bases in the final 110 kb consensus sequence as base calling decisions were made based on the resources available at that time. Several conditions ensured that the basecalling method was standardised throughout the sequence assembly. These conditions were as follows:

i. During alignment, if within a region only BAC sequence was present with no other sequence backup then the BAC sequence was called
ii. If a single nucleotide varied between all reference sequences then the sheep BAC nucleotide was called
iii. If the BAC sequence contains ambiguity of 'N' but there are specific nucleotides in another reference sequence (even from one source) the sequence from the other source was called.

The sheep sequence in the publicly available ISGC and GenBank databases was incomplete when the study began in 2010. Thus the Roche 454 sequences generated from this study bridged gaps and enriched sequence information of the genome specifically in the *CLN6* region of interest, which was crucial to provide reference sequence for mutation screening (Chapters 7 and 8).

In conclusion, the Roche 454-pysequencing of ovine BAC was cost effective, efficient and provided approximately 120 kb of ~14X coverage sequence of normal sheep genomic reference sequence.

CHAPTER 6: MUTATION SCREENING APPROACH 1: SEQUENCE CAPTURE FOR TARGETED SEQUENCING

6.1 Introduction

The generation of new ovine sequences (Chapters 4 and 5), supplemented with known ovine sequences from published and unpublished sources greatly enriched sequence information within and flanking ovine *CLN6*. A consensus sequence formed using a combination of these sequences was subjected to two mutation screening approaches, to be described in this and the following chapter (Chapter 7). The mutation screening approach described in the present chapter is based on NGS sequencing of enriched genomic DNA that was captured using sequence capture.

6.2 Materials and methods

6.2.1 Sheep genomic DNA samples

Genomic DNA samples from three affected, two carriers and a normal sheep from the NCL South Hampshire (SH) research flock were prepared by Nadia Mitchell, Lincoln University, New Zealand (Table 6.1). The DNA was isolated from pelleted sheep blood leucocytes with the Qiagen QIAamp DNA mini kit and the DNA eluted in the supplied Qiagen AE buffer (10 mM Tris.Cl; 0.5 mM EDTA; pH 9.0), and stored at -80°C prior to shipment. These DNA pellets were dissolved in 600 µl TE buffer. DNA was quantified using the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies Inc, Thermo Scientific, USA) and produced OD260/280 ratio readings between 1.83 and 1.89. Genomic DNA samples were shipped in 1.5

ml microcentrifuge tubes individually sealed with parafilm and shipped in August 2009 at RT to LC Sciences, Houston, Texas, USA.

Table 6.1: Sheep genomic DNA used for sequence capture.

No	ID	Breed	Relationship between animals	Genotype*	Phenotype
1	CPW156	Coopworth		*GG	Normal (N)
2	SH1022/07	South Hampshire		*AA	Affected (A)
3	SH1032/08	South Hampshire	full sibs	*AG	Carrier (C)
4	SH1033/08	South Hampshire		*AA	Affected (A)
5	SH1038/08	South Hampshire	full sibs	*AG	Carrier (C)
6	SH1039/08	South Hampshire		*AA	Affected (A)

*Genotype identified using the indirect DNA test for NCL in South Hampshire sheep (Tammen et al., 2006)

6.2.2 Targeted high-throughput sequencing

6.2.2.1 Creation of custom designed microarray chip

The sequence capture method used in this study is essentially based on capture of fragmented and amplified genomic DNA via hybridization to customised probes on the µParaflo microarray chip. Sequence capture was conducted by LC Sciences following the service provider's sample preparation protocol and a previously described protocol (http://www.lcsciences.com/applications/genomics/targeted-genome-sequencing/targeted-sequencing/).

The consensus reference sequence was narrowed down to a region of 73,072 bp (Appendix 5) to focus on *CLN6* and its flanking genes upstream and downstream. The 73,072 bp sequence (referred to as 'capture reference sequence' from here onwards) comprised of ovine genomic sequence *CLN6*, *Fem1b* and *CALML4*.

The capture reference sequence was screened for repetitive elements using the RepeatMasker program (http://www.repeatmasker.org/) version open 3.2 run at default mode against bovine Btau4.0 repeat library files available on RepBase Update 20090120 (Jurka et al., 2005; http://www.girinst.org/). No ovine repeat library is available in this software tool. Repetitive elements were masked by replacing the repetitive nucleotides with N's. BLAST searches against

the bovine btau4 assembly identified additional small regions that aligned to multiple regions in the genome and these were masked with the maskedseq program (emboss.sourceforge.net/apps/release/6.1/emboss/apps/maskedseq.html) with assistance from Dr. Matthew Hobbs. The final masked sequence is called 'masked capture reference sequence'. The masked capture reference sequence of 41,313 bp was used by the service provider to design the customised probes for the μParaflo microarray chip.

For quality control (QC) purposes Quantitative PCR (q-PCR) was conducted to allow assessment of effective enrichment and validation of successful probe design. A total of nine regions were used and primers are listed in Table 6.2. Four of the regions and primers were chosen by the service provider (PF02, PF22, PF23 and PF29) and an additional five regions were provided to the service provider (Q12, Q23, Q31, Q45 and Q54).

Table 6.2: Primer sequences for amplification of nine Quantitative PCR (q-PCR) regions.

Q-PCR region	Primer direction	Sequence 5'>3'	Primer length (bp)	PCR product length (bp)
Q12	forward	ACATGAAAGGAGGATTTGAGGCAG	24	31
	reverse	CTACCTCGCCTTCCCTGC	18	
Q23	forward	GCAACAGTGTCGGTAAAGCC	20	38
	reverse	GGGAATGCTAATAGAAGAGACAGCG	25	
Q31	forward	CGGGCTGCTGAACTCTCAAG	20	25
	reverse	TGCCTGTTGAAGTCTTGAGTGG	22	
Q45	forward	CTGCTGTCAAATTTCACCAGTCAC	24	36
	reverse	TCTGGGTGGCATAACATTAACTGC	24	
Q54	forward	TGTTTCTTCTATTGTGAGCAGTGGC	25	26
	reverse	AGAATGTTTCCAAGTCAGCAAGGT	24	
PF02	forward	AAAGGCCCATCATTACCGGAG	21	56
	reverse	AGAGGCTGCAATTTCTGGGTAG	22	
PF22	forward	GGAAACCCACTGTGCCTAGC	20	43
	reverse	CAGGGTGATACTGTCGTGGTT	21	
PF23	forward	CGACATTTCCCAGAACCCTCT	21	36
	reverse	GCACCTGCCTTCGGAATCTC	20	
PF29	forward	CCAGAACAAAGGAGAGGCGTC	21	52
	reverse	TTCCCCACAAACTCTGGCAT	20	

6.2.2.2 Capture and enrichment of the target sequence

At the service provider, capture and enrichment of the genomic DNA samples were conducted according to the sample preparation protocol by LC Sciences. The protocol is summarised below.

Firstly, the gDNA samples (~2 µg each) were fragmented using NEBNext dsDNA Fragmentase (M0348, New England Biolabs), purified with QIAquick PCR spin columns (Qiagen, Australia) and eluted with 30 µl of EB solution. Overhangs of the DNA fragments were converted to blunt ends using *E.coli* DNA ligase, T4 DNA polymerase, and Klenow enzyme (NEB) at 20ºC for 30 minutes, then purified and eluted. An 'A' base was then added to the 3' end of the blunt phosphorylated DNA fragments using NEB's Klenow fragment (3' to 5' exo minus) at 37ºC for 30 minutes, they were then purified and eluted. Fragments were ligated to adapters (Illumina, Australia), which have single 'T' base overhangs at its 3' end, with the molar ratio of adapter to DNA fragments as 10:1. The ligated DNA fragments were then loaded to a 2 % TAE agarose gel and run at 120V for 30 minutes. The major band ranging from ~150 to 500 bp was excised and extracted using a QIAquick gel extraction kit, and was eluted in 30 ml of EB buffer. The DNA fraction of 150-500 bp was amplified by PCR (15 cycles) using the Illumina common PCR primer pair 5'AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT(N)AGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG3' (forward) and 3' TTACTATGCCGCTGGTGGCTCTAGATGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGA(N)TCTAGCCTTCTCGAGCATACGGCAGAAGACGAAC 5' (reverse) (with 'N' refering to where the DNA fragments were inserted), with Phusion polymerase (NEB), followed by purification with a QIAquick PCR spin column (Qiagen). The amplified DNA fragments were used as templates to generate single-stranded DNA (ssDNA) fragments by single strand PCR using only the forward PCR primer from the common PCR primer pair. The ssDNA fragments were then purified.

Capture of targeted DNA fragments was then performed. To sum up, the single-stranded PCR products, three spike-in controls (Table 6.3), and 0.1 mg/ml Bovine serum albumin (BSA, Sigma-Aldrich) were added to a DNA hybridization buffer. The prepared DNA solution (pre-hyb, sample, total ~10 µg DNA inside) was loaded by a circulation pump to a µParaflo microarray chip with DNA probes designed to capture the targeted DNA fragments. The pre-hyb sample was hybridized inside the chip at 30ºC for 24 hours. Before hybridization, a small portion of the same pre-hyb samples was kept for q-PCR analysis.

Table 6.3: Sequences for three spike-in controls added to a DNA hybridization buffer during sequence capture.

Sequence name	Sequence 5'>3'	Length
CtrOM45	GACCACGAGCATAGGATCCGTAACATTAGCAGAGCGAGGTATGTA	45
CtrOM65	ACCACAGTCCATGCCATCACAGCACACTTATAGATCGTCATAACATTAGCAGAGCGAGGTATGTA	65
CtrOM85	CCAGGCATTCCTATCAGTCTCCACTCAAGTATCATCCAGGAAATATGTGCGGTGTACATCTAACATTAGCAGAGCGAGGTATGTA	85

After hybridization, the chip was washed with manufacturers' buffers and RNase-free water in a 1.5 ml centrifuge tube to strip the chip. After water stripping, the chip was eluted with water. When completed, all the water used (~400 ml) was combined and dried in a Speed-Vac at 60ºC for 45 minutes. The dried pellet was hydrated in 100 µl of RNase-free water and vortexed for 2 minutes.

The enrichment phase comprised of further amplification of the eluted sample by PCR (14 cycles) using the common PCR primers. The purified amplified captured DNA (post-hyb sample) is called the final sample prepared for sequencing.

Enrichment of the captured DNAs was verified by q-PCR using Power SYBR Green PCR Master Mix (Applied Biosystems) on the ABI Prism 7000 (Applied Biosystems). The reactions were performed in triplicate. To determine the total DNA amount of the two samples, the common PCR set was used as q-PCR primer to obtain the q-PCR cycle number of the two samples. To determine the amount of internal targeted DNA regions inside the two samples, only seven specific q-PCR primer sets (PF02, PF22, Q12, Q23, Q31, Q45 and Q54; Table 6.2) were used to obtain q-PCR cycle numbers of these seven DNA regions. The enrichment ratio of the

internal targeted DNA fragment of pre-hyb and post-hyb samples was determined using a formula (Figure 6.1).

$$R_{enrichment} = a \times 2^{(CT_{total,post} - CT_{target,post}) - (CT_{total,pre} - CT_{target,pre})}$$

where

a: dilution factor, in this case, a=100;

$CT_{total,post}$: cycle number of post-hyb sample using the common PCR primer set;

$CT_{total,pre}$: cycle number of pre-hyb sample using the common PCR primer set;

$CT_{target,post}$: cycle number of post-hyb sample using a specific q-PCR primer set for one targeted DNA region;

$CT_{target,pre}$: cycle number of pre-hyb sample using a specific q-PCR primer set for one targeted DNA region;

Figure 6.1: Formula to determine enrichment ratio of the internal targeted DNA fragment of pre-hybridization (pre-hyb)and post-hybridization (post-hyb) samples.

Captured DNAs from the genomic samples were sent to the High throughput DNA sequencing unit at the University of Otago, Dunedin, New Zealand for next-generation sequencing with the GS FLX system (454 Roche, USA).

6.2.2.3 Next-generation sequencing of captured DNAs

Following evaluation of the capture DNA fragments by agarose gel electrophoresis, fragments from each of the four genomic samples were processed into individual capture libraries. Each library consisted of a set of single-stranded template DNA fragments representing the entire span of the individual sample sequence. Each library was flanked by appropriate amplification and sequencing Multiplex Identifier (MID) adaptors, then purified and quantitated. These unique Roche rapid library MID adaptors were assigned to each library using a standard blunt-end ligation protocol during the preparation of DNA libraries for sequencing on the GS FLX system (454 Roche, USA). Each of these adaptors contained a unique 10 base sequence that is recognised by the sequencing analysis software, allowing for automated sorting of MID-

containing reads to the correct library. Tagging multiple libraries with unique MIDs allowed the samples to be amplified and sequenced together, in a single region of the PicoTiterplate (PTp) device. The RL 1 to 4 MID adaptor sequences are shown in Table 6.4. Sequencing of the samples was processed using the GS FLX system (454 Roche, USA) Titanium chemistry on a 1/8 PTp device in a single run.

Table 6.4: Sequence of the Roche rapid library MID attached to four capture libraries.

MID codes	Sequence 5'>3'	Sheep genomic library
RL1	ACACGACGACT	SH1033/08
RL2	ACACGTAGTAT	SH1038/08
RL3	ACACTACTCGT	SH1039/08
RL4	ACGACACGTAT	CPW156

6.2.3 Bioinformatic analysis

All resulting sequence data were analysed with the assistance of Dr. Matthew Hobbs. Initial bioinformatic analysis used the GS FLX Reference Mapper (GS Mapper) assembly package (ver. 2.2.22.20). As the MID codes remained in the sequence reads, preferences were set up to specify which MID codes were in the reads so that the software will ignore them and just analyse the sequence after the codes. The spike-in control and common sequences were also filtered prior to assembly of the captured sequences.

6.3 Results

6.3.1 Design of the custom microarray chip

As indicated in the introduction section, a consensus reference sequence of 73,072 bp was chosen to focus on *CLN6* and its flanking genes *Fem1b* and *CALML4*. Prior to designing the microarray probe, repeat sequences needed to be removed to avoid capture and

overrepresentation of repetitive DNA elements not related to the region of interest. RepeatMasker identified a total of 30,652 bp (41.95 % of the total sequence) that corresponded to known bovidae repeats. These included 29,733 bp total interspersed repeats (12588 SINEs, 11684 LINEs, 3257 LTRs and 2204 DNA elements), 75 bp small RNA sequence, 506 bp of simple repeats and 338 bp low complexity sequences. Additional 1,101 bp was removed with maskedseq program. This resulted in a total of 31,753 bp sequences masked from the original sequence using both RepeatMasker and maskedseq programs, leaving 41,319 bp of unmasked sequences used for design of the customised probes for sequence capture. This masked sequence is referred to as masked capture reference sequence.

The q-PCR analysis of the genomic samples was implemented at the service provider to validate the capture reference sequence prior to sequence capture. Initially two (CPW156 and SH1033) of the six genomic DNAs were randomly chosen for this purpose. Four pairs of primers which were designed based on the masked capture reference sequence were used to validate four sections (PF02, PF22, PF23 and PF29; Table 6.2) of the target sequence in the genomic DNA samples. Only two of the four primer pairs resulted in successful amplification (PF02 and PF22). This raised concerns about the quality of the reference sequence. Therefore, an additional five regions were identified, which represented regions in the *CLN6* (n=2), *CALML4* (n=1) as well as *FEM1b* (n=2).

The seven q-PCRs amplified PCR products of expected length. This suggested that the reference sequence was of sufficient quality to proceed. The recent availability of the ovine genome assembly Oar v3.1 has since verified that the failure of the two q-PCR primer pairs PF23 and PF29 was not due to inaccuracies in the masked capture reference sequence. BLAT search of these q-PCR regions against the USCS genome browser ovine ISGC OAR7 ver. 3.1 assembly aligned to the predicted ovine regions at position 14,820,671 - 14,829,744 (https://genome.ucsc.edu/cgi-bin/hgGateway?hgsid=385998333_5bl7eEll7bMPJiSBlz4gaxAR4zP5&clade=mammal&org=Sheep&db=0) verified the sequence was correct and did not include any known repeats.

6.3.2 DNA capture and enrichment

Gel electrophoresis of the captured and enriched DNAs (Figure 6.2) identified the expected smear of fragments as well as a distinct band of less than 200 bp in all the animals.

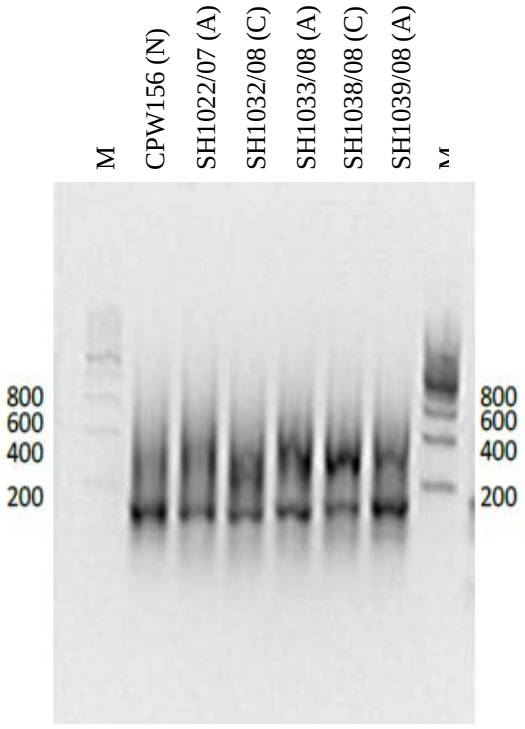


Figure 6.2: Agarose gel image of the captured products including three synthetic spike-in controls. The identification of each animal is shown at the top of the gel image (phenotypes in brackets, N= normal; A= affected; C= carrier). Relevant fragment sizes of the DNA marker (M) are shown on the left-hand side of the 1% TAE-agarose gel and elctrophoresised at 120V for 30 min. The band less than 200 bp are amplified spike-in controls.

The high prevalence of the band of less than 200 bp was deemed problematic by the service provider as it suggested interference likely to be related to the spike in control added to the DNA hybridization buffer. Therefore the band was recommended to be removed using gel-cut purification prior to sequencing. However, as the NGS service provider anticipated that the 454 sequence would be followed up with an additional run for deeper sequencing at high coverage and that these sequences could be removed prior to assembly, the bands were left in.

Verification of enrichment of the target sequences was analysed using the q-PCR to calculate the enrichment ration pre and post hybridization using common PCR primers as well as the seven targeted q-PCR primers (Table 6.2).

Enrichment ratios, which are the relative quantities of the seven specific DNA fragments of the captured fragments after enrichment compared to before enrichment from each of the six genomic samples, are listed in Table 6.5. The enrichment ratios varied greatly across and within animals, with a range between 10 and 26,580,339. Based on the ratios, only four genomic samples comprised of two affected (SH1033/08, SH1039/08), a carrier (SH1038/08) and a normal sheep (CPW156) were chosen for sequencing as their ratios are considered to be within the normal range recommended by the service provider.

Table 6.5: List of enrichment ratios of seven specific DNA fragment controls.

Index	Sample ID	Targeted Fragment	CT (total.post)	CT (target.post)	CT (total.pre)	CT (target.pre)	Dilution Factor	Enrichment Ratio
1	SH1022/07	Q12	10.11	23.35	12.95	26.35	100	112
2	SH1022/07	Q23	10.11	24.65	12.95	31.79	100	1,970
3	SH1022/07	Q31	10.11	24.71	12.95	31.24	100	1,291
4	SH1022/07	Q45	10.11	28.13	12.95	32.76	100	346
5	SH1022/07	Q54	10.11	33.70	12.95	33.15	100	10
6	SH1022/07	PF02	10.11	26.38	12.95	31.70	100	558
7	SH1022/07	PF22	10.11	26.78	12.95	32.81	100	913
8	SH1032/08	Q12	10.59	23.42	14.58	25.95	100	36
9	SH1032/08	Q23	10.59	25.70	14.58	31.85	100	447
10	SH1032/08	Q31	10.59	23.61	14.58	30.69	100	851
11	SH1032/08	Q45	10.59	26.96	14.58	32.40	100	273
12	SH1032/08	Q54	10.59	31.26	14.58	33.35	100	27
13	SH1032/08	PF02	10.59	25.57	14.58	32.25	100	645
14	SH1032/08	PF22	10.59	24.54	14.58	31.79	100	958
15	SH1033/08	Q12	13.51	24.44	14.25	26.72	100	291
16	SH1033/08	Q23	13.51	27.44	14.25	33.04	100	2,904
17	SH1033/08	Q31	13.51	24.99	14.25	31.11	100	4,164
18	SH1033/08	Q45	13.51	31.00	14.25	34.33	100	602
19	SH1033/08	Q54	13.51	35.68	14.25	35.02	100	38
20	SH1033/08	PF02	13.51	27.44	14.25	31.97	100	1,383
21	SH1033/08	PF22	13.51	23.42	14.25	37.05	100	759,061
22	SH1038/08	Q12	13.63	23.00	13.95	30.24	100	12,110
23	SH1038/08	Q23	13.63	25.64	13.95	35.78	100	90,389
24	SH1038/08	Q31	13.63	23.39	13.95	31.83	100	27,820
25	SH1038/08	Q45	13.63	25.12	13.95	38.94	100	1,158,524
26	SH1038/08	Q54	13.63	28.61	13.95	31.59	100	632
27	SH1038/08	PF02	13.63	25.15	13.95	40.00	100	2,365,734
28	SH1038/08	PF22	13.63	22.05	13.95	40.39	100	26,580,339
29	SH1039/08	Q12	14.15	22.07	14.86	31.53	100	43,054
30	SH1039/08	Q23	14.15	23.67	14.86	37.49	100	884,104
31	SH1039/08	Q31	14.15	22.74	14.86	35.98	100	591,433
32	SH1039/08	Q45	14.15	24.49	14.86	39.71	100	2,333,164
33	SH1039/08	Q54	14.15	26.85	14.86	39.34	100	351,668
34	SH1039/08	PF02	14.15	23.61	14.86	38.90	100	2,449,161
35	SH1039/08	PF22	14.15	22.52	14.86	40.00	100	11,175,656
36	CPW156	Q12	13.39	26.10	14.80	27.40	100	93
37	CPW156	Q23	13.39	25.99	14.80	32.69	100	3,912
38	CPW156	Q31	13.39	24.25	14.80	31.77	100	6,907
39	CPW156	Q45	13.39	28.48	14.80	33.23	100	1,013
40	CPW156	Q54	13.39	30.52	14.80	33.74	100	351
41	CPW156	PF02	13.39	25.87	14.80	32.98	100	5,198
42	CPW156	PF22	13.39	24.47	14.80	38.88	100	819,200

6.3.3 Sequencing of the captured DNAs

Four (SH1033, SH1039, SH1038 and CPW156) out of the original six genomic samples were sequenced using the GS FLX system Titanium chemistry. Samples for sequencing were comprised of fragmented DNAs of about 700 bp in length at a total amount of 500 ng per captured genomic sample. The number of reads generated was 27,247 kb and 86,044 kb for affected sheep SH1033 and SH1039 respectively, 59,175 kb for carrier sheep SH1038 and 151,528 kb for normal sheep CPW156.

After filtering the spike-in control sequences, common sequences and MID barcodes, initial assembly using the GS FLX GS Mapper software mapped only 10% of the reads from each genomic sample back to the capture reference sequence. The majority of the sequences were short and mapped to various regions of the reference sequence at low coverage. A preliminary search for *CLN6* sequence identified that not all regions of *CLN6* were present and that those regions were not covered by sufficient sequencing depth to permit mutation screening.

6.4 Discussion

At the time of research, the mutation screening technology was relatively new and was deemed a revolutionary process for enrichment of selected genomic regions from full complexity human genomic DNA. Although it was initially used for the human genome, this technology was considered for our study given the lack of genome information surrounding and within ovine *CLN6*. However, due to limitations discussed below this approach was not successful in this project. Only 10% of reads were mapped back to the capture reference sequence and those regions that were present were not covered in enough depth for mutation screening.

Conventionally, targeted sequence variation discovery is achieved by Sanger sequencing of PCR products (Stephens et al., 2006). This was attempted early in the research with little progress (Chapters 3 - 5). In recent years, new technologies have emerged that capture targeted sequences

within the genome resulting in an enriched pool of target sequences and potentially greater sequence coverage for each targeted region during sequencing (Summerer, 2009; Grover et al., 2012). Technologies used for capture and enrichment of the target sequences have been reviewed (Mamanova et al., 2010; Grover et al., 2012; Summerer, 2009).

Little was known about sequence capture limitations because this technology was new at the time of study in 2009 - 2010. Reviews have shown the technology to produce variation of capture uniformity that results in dropout of difficult regions, which in turns provides insufficient capture to allow for full coverage of the targeted regions at a depth allowing for reliable nucleotide calling (Summerer, 2009). Repetitive sequence targeted regions are also best avoided, which complicates capture design in many cases where flanking regions are low complexity sequence (Summerer, 2009). Application of targeted genomic enrichment technologies as a diagnostic tool has also been found to be challenging because there is lack of definitive parameters for QC of the sequence data (Chou et al., 2010). With no reliable cutoff threshold set for determining true sequence variants, heterozygous-allele call cannot be established confidently in cases when various variants are detected (Chou et al., 2010). There was also very few, if any, established companies with experience doing sequence capture on organisms with more limited genomic resources. A local company (Roche NimbleGen) had the technology but was not equipped for working with non-human species (C. French pers. comm.). The service provider for this study is located overseas and due to inexperience with working with the sheep genome combined with novelty of the technology itself, was unable to provide sufficient consultation throughout the experiment and satisfactory bioinformatics support with data.

In comparison to conventional whole genome sequencing and Sanger sequencing to discover causative mutations, sequence capture is relatively more cost-effective and requires less labour with the resulting data considerably less cumbersome to analyse (Mamanova et al., 2010; D'Ascenzo et al., 2009). This technology has been shown to be more beneficial when working on organisms whose genome sequences are readily available, rather than organisms that possess few genomic resources (Grover et al., 2012). At the time of study, constraint was imposed by limited availability of sequences flanking ovine *CLN6* due to the ovine genome assembly Oar v1.0 being in its first working draft.

The ovine BAC sequence obtained in Chapter 5 covered the region of interest (*CLN6* and flanking genes) and when aligned with known sequence from published and unpublished sources, enabled the generation of a reference sequence for this mutation screening approach. However, as indicated in Chapter 5, the reference sequence had its limitations. Merging of contigs and semi-manual alignment of sequences using the Genedoc program might have introduced errors and some genome areas that are difficult to sequence due to GC richness might not have been accurately represented. However as will be shown in future chapters (Chapter 7 and 8), the overall reference sequence was reasonably good quality.

One of the most important aspects of the sequence capture approach is accuracy of the capture probe sequence to ensure that the custom probe targets specific regions of interest in the genome. Targeted regions that contain highly repetitive sequence complicate sequence capture design and reduce specificity of binding capacity during hybridization (Summerer, 2009) which results in capture of high abundance of non-target regions in the genome. Thus, repeat masking is essential prior to customised capture probe design. However repeatmasking of sequences using RepeatMasker was not available for the ovine genome, instead those of bovidae were used which might possibly not cover all ovine specific repeats. Similarly, additional Blast analysis was conducted against corresponding bovidae repeats. With only 10% of the captured sequence mapping back to the reference sequence, this likely indicates the presence of repeats which were not efficiently masked during filtering of sequence prior to capture array design. With poor capture specificity; it was highly unlikely that the information would be useful for mutation screening. However, reanalysis with RepeatMasker in August 2014 have now identified that not all repeats were masked. A recent run identified additional 3,279 bp of repeat sequence which is likely to explain the high level of undesired sequence reads generated with this technology.

The suggested normal range of enrichments between 1,000 to several tens of thousands by the LC Sciences (Dr. Q. Zhu pers. comm.) led to choosing of only four genomic samples (SH1033/08, SH1038/08, SH1039/08 and CPW156) with relatively high ratios across most or all DNA fragments for sequencing. These samples were considered to have achieved sufficient levels of enrichment, and thus be suitable for sequencing. The extremely high variations of enrichment ratios within and between animals suggest that the procedure might need further

optimisation. Because non-uniformity of sequence capture is a known drawback with this technology, this could likely explain the overcapture of some regions which are potentially redundant reads, as compared to dropout of regions deemed difficult (Summerer, 2009) including those with repetitive sequences and high GC content (Porreca et al., 2007).

Using this technology also meant introduction of long non-informative sequences from the service provider into the NGS reads (Summerer, 2009). This included chemically synthetic spike-in control oligonucleotides, Roche MID tags and the common sequence. Although removal of the additional 200 bp amplified bands was recommended using additional laboratory steps, the alternative option of filtering these sequences prior to captured sequence assembly was taken due to anticipation that the GS FLX system (454 Roche, Australia) would generate sufficient coverage of the targeted region. However, due to already numerous other challenges described above, presence of these additional sequences further complicated bioinformatic analysis and limited data output (Summerer, 2009).

Considerable delays impeded the smooth running of this study at various stages. This included last minute change of the service provider's accessibility to a sequencing service which necessitated finding another company available at short notice. Eventually the captured samples were shipped to New Zealand for sequencing. Overall the study took approximately ten months (between October 2009 and July 2010), from shipping the sheep genomic DNA to obtaining sequence reads.

Despite limitations outlined above, this technology has the potential to be an invaluable tool for detection of disease causing variants if a compromise can be achieved between sequence capture specificity and uniform coverage of the targeted region (Mamanova et al., 2010; Porreca et al., 2007). Efficient discovery of causative mutations have been documented (Guelly et al., 2011; D'Ascenzo et al., 2009).

As the data for the second mutation screening approach (LR-PCR amplification and sequencing described in Chapter 7) became available throughout the initial data analysis, and as serious concerns with the depth and completeness of coverage for this study became obvious early, further data analysis of this incomplete data set was abandoned.

CHAPTER 7: MUTATION SCREENING APPROACH 2: NGS SOLID SEQUENCING OF LONG-RANGE PCR PRODUCTS

7.1 Introduction

The mutation screening approach described in this chapter was conducted in parallel with the mutation screening approach employed using sequence capture and re-sequencing (Chapter 6). The present approach does not involve masking repeat sequences, utilises a different NGS platform, and targets the whole genomic region of the *CLN6*, as well as the flanking sequences which include *CALML4*, using long-range PCR (LR-PCR). Amplicons from NCL affected, carrier and normal (control) sheep were then sequenced using the ABI SOLiD NGS platform (Applied Biosystems, USA) and analysed using bioinformatic tools.

7.2 Materials and methods

7.2.1 Sheep genomic DNA

Genomic DNA from three affected, two carrier and one normal sheep from the NCL South Hampshire (SH) research flock, as well as an affected and a normal sheep from the NCL Merino research flock, were used for LR-PCR amplification and sequencing (Table 7.1). These DNA samples comprised of five South Hampshire and one normal Coopworth sheep (provided by Prof. David Palmer at Lincoln University, Christchurch, New Zealand) and from two Merino sheep from the University of Sydney. Genotyping of these sheep with direct and indirect DNA tests for NCL is described earlier in this work (Chapters 2.1.1 and 2.1.2).

Table 7.1: Animals used for LR-PCR amplification and SOLiD sequencing.

No	Sheep ID	Breed	Relationship between animals	Genotype	Phenotype	Total amount of DNA
1	CPW156	Coopworth		*GG	normal	30 mg
2	SH1022/07	South Hampshire		*AA	affected	32 mg
3	SH1032/08	South Hampshire	full sibs	*AG	carrier	35 mg
4	SH1033/08	South Hampshire		*AA	affected	34 mg
5	SH1038/08	South Hampshire	full sibs	*AG	carrier	38 mg
6	SH1039/08	South Hampshire		*AA	affected	28 mg
7	L06	Merino	full sibs	**CC	normal	20 mg
8	L07	Merino		**TT	affected	20 mg

* Genotype identified using the indirect DNA test for exon 7 polymorphism in South Hampshire sheep (Tammen et al., 2006)
** Genotype identified using the direct DNA test for exon 2 missense mutation in Merino sheep (Tammen et al., 2006)

The South Hampshire and Coopworth sheep DNA samples provided were isolated by Nadia Mitchell (Lincoln University, Christchurch, New Zealand) from pelleted sheep blood leucocytes using the Qiagen QIAamp DNA mini kit (Qiagen, Hilden, Germany) described earlier (Chapter 2). DNA was quantified using the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies Inc, Thermo Scientific, USA) and dried by freezing in a high vacuum (lyophilised) into a DNA pellet before transport to the University of Sydney. Each DNA pellet was resuspended in 1xTE (Tris-EDTA, pH 8.0, Amresco, Ohio, USA) to achieve a final concentration of 50 ng/µl prior to use as DNA template for LR-PCR amplification. DNA samples from Merino sheep, available from a previous study (Houweling, 2009), were quantified using a NanoDrop ND-1000 spectrophotometer and diluted to 50 ng/µl prior to use as a DNA template for LR-PCR amplification.

7.2.2 LR-PCR primers and protocol

7.2.2.1 LR-PCR Primers

Sixty standard primers (average of eight primers for each PCR amplification region) covering the region of interest were designed for initial long-range PCR amplifications and commercially synthesised by Sigma-Aldrich (Australia). These primers were paired for PCR using various combinations and PCR conditions. After optimisation, twenty-eight primers were successful in generating contiguous partially overlapping PCR products (Figure 7.1). As a requirement for SOLiD sequencing, primers were re-synthesised with an amine-modification, where an amine group was incorporated on the 5'-terminus. These modified primers were synthesised (Invitrogen, Life Technologies, USA) and purified by high performance liquid chromatography (HPLC). Modified primer pairs required re-optimisation of the PCR conditions. The final optimised PCR conditions for all PCR reactions are shown in Table 7.2.

7.2.2.2 LR-PCR protocol

The KOD Hot Start DNA Polymerase kit (Merck Biosciences Limited, Australia) and supplied reagents were used for LR-PCR amplification. Each LR-PCR had a total reaction volume of 50 µl of 1X KOD Buffer, 2.5 mM $MgSO_4$, 0.2 µM dNTP, 10 pmol of each primer (using combinations of primers set out in Table 7.2), 0.2 U of KOD Hot Start DNA Polymerase (Merck Biosciences Limited, Australia) and 50 ng of genomic DNA. The PCR reaction for product '2ir' (Table 7.2) contained the addition of dimethyl sulfoxide (DMSO) to a final concentration of 5 % v/v. LR-PCR was performed in multiple 50 µl PCR reactions (five to twenty reactions per animal for each PCR product) to generate a minimum of 70 ng of each LR-PCR product per animal. The standard PCR protocol was initially performed for all the PCR amplifications. For templates that were difficult to amplify, particularly 'C4dr', '1hr' and 'C1cr', the touchdown protocol was additionally used to enhance specificity and product formation.

The standard protocol was 95℃ for 2 min, 29 cycles at 95℃ for 20 sec, various annealing temperatures (Table 7.2) for 10 sec, 70℃ for 2 min, an additional extension at 72℃ for 1 min followed by cooling down at 4℃ for 15 min. The touchdown protocol was 95℃ for 2 min, 10 cycles of 95℃ for 20 sec, annealing temperatures starting at 70℃ for 20 sec and extension at 72℃ for 45 sec. This was followed by 22 cycles (‘C1cr’, ‘C4dr’) or 27 cycles (‘1hr’) of 93℃ for 20 sec, annealing temperatures of 60℃ for 20 sec and extension at 72℃ for 45 sec, with a final extension at 72℃ for 3 min followed by 15℃ ° for 20 sec and cooling down at 4℃ for 1 min.

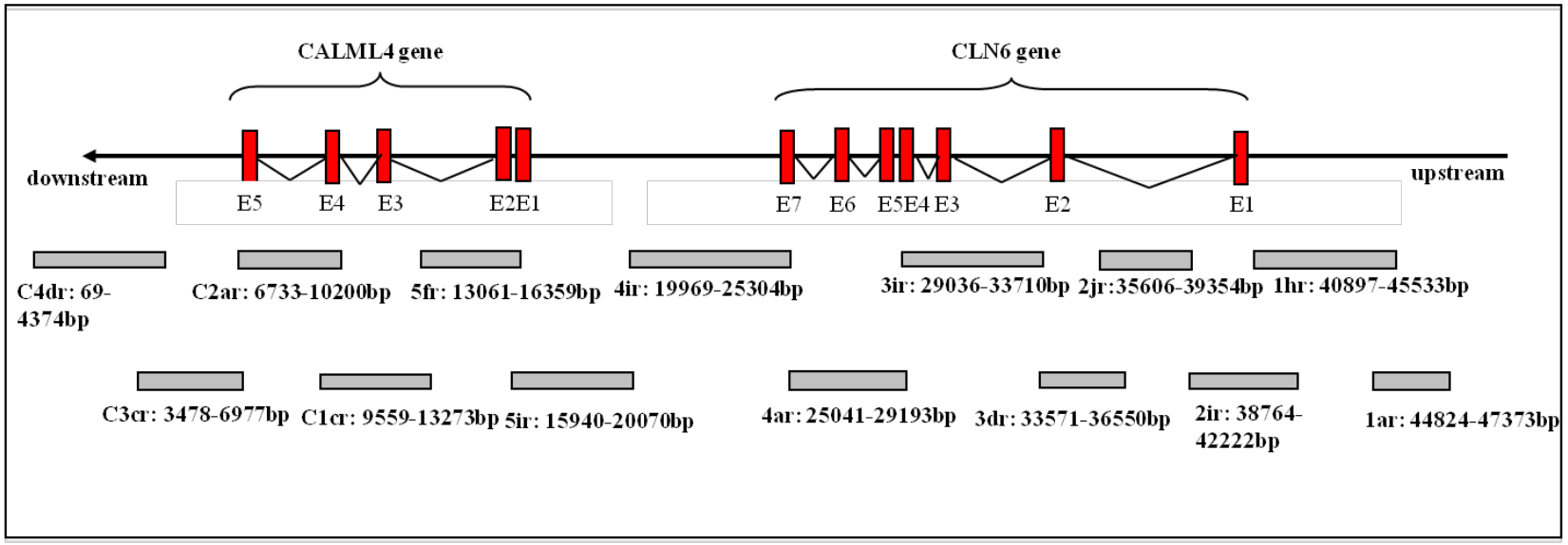


Figure 7.1: Schematic diagram (not to scale) of position of the 14 LR-PCR products for sequencing regions of interest within and surrounding ovine *CLN6*. Approximately 49,123 kb of sequence is shown and the orientation of the *CALML4* and *CLN6* are in accordance with the ovine and bovine genomes. The red vertical boxes represent gene exons and the grey horizontal boxes represent location of the LR-PCR products. Position of the LR-PCR products as well as length of exonic/intronic regions are based on known ovine sequences: published exons from GenBank and ISGC and non-coding sequence obtained from collaborator Nadia Mitchell as well as predicted length of sequences based on bovine genome sequee.

Table 7.2: PCR conditions to generate LR-PCR products covering the regions of interest.

PCR product	Primer	Direction	Sequence 5'>3'	Length (bp)	Region amplified	Expected PCR size (bp)	Annealing temperature
1ar	5UTRF23	forward	GAACAATGAAGTCCAAGCAGAA	22	CLN6_5UTR	2560	62
	5UTRR24	reverse	GGCAAGATCCCTGGAAATG	19	CLN6_5UTR		
1hr	I1F25	forward	TACTTCCCTTCTCTTCCATTCAAAC	25	CLN6_5UTR	4620	65
	5UTRR42	reverse	GCCTAAACCTACTCTAACCCACTTC	25	CLN6_5UTR		
2ir	I1F21	forward	CCACAAACTTCCTAAAACTGACTCC	25	CLN6_Intron1	3820	65
	5UTR38	reverse	CACCCCCTCCAAGTCTCTAAC	21	CLN6_5UTR		
2jr	I1F15	forward	CTGGAGGTGGGAGTGAGAAA	20	CLN6_Intron1	3730	65
	I1R24	reverse	CTTCACTCTGTCTTCATTCCGTTTT	25	CLN6_Intron1		
3dr	E2F5	forward	ATCCACAGCTTCATGCTCAC	20	CLN6_Intron2	2990	60
	I1R18b	reverse	CTGCCCTTTTTCTTTCTGCT	20	CLN6_Intron1		
3ir	E3F7	forward	GACCTCCCACCACCCCTAAT	20	CLN6_Intron3	4680	65
	I1R22	reverse	CTGACTGCTTTCGTTTACTTCTTTC	25	CLN6_Exon2		
4ar	I6F1	forward	TGACTGACGAGGAGAAAGCA	20	CLN6_Intron6	4160	62
	I3R2	reverse	TTCAAAGTAAGGAAGGTGCAGTC	23	CLN6_Intron3		
4ir	4eF1	forward	CCATCAGCAAGCCCTCTC	18	CLN6_3UTR	5340	68
	4eR2	reverse	GAAGCCCTCCAGCAAGTGT	19	CLN6_Intron6		
5ir	CALL1F1	forward	CCTTCCTGTTCCCCTTCTTC	20	CALL4_Intron1	4130	65
	i5aR2	reverse	GAGAGGCTTGGCACACTGATA	21	CALL4_5UTR		
5fr	CALL2F1	forward	CGGTGGGGGTTTAGAGACA	19	CALL4_Intron2	3300	58
	3UTRR8	reverse	GTGGTGTGTGACGTGCCTAA	20	CALL4_5UTR		
C1cr	CALL3F1	forward	CCTTTTTTGGGTCCTCTTGTT	21	CALL4_Exon4	3720	58
	CALL2R4	reverse	CCCTGTCCCTGTCCCTGTC	19	CALL4_Intron1		
C2ar	CALL4F3	forward	CGCCTCCTCGTTCAATAGTC	20	CALL4_Exon5	3460	60
	CALL4R2	reverse	GTTTCCTTTGGTTTCCGTCTC	21	CALL4_Intron3		
C3cr	CALL3UTRF3	forward	GCCACGCTTGGTTGACTAAT	20	CALL4_3UTR	3520	62
	CALL3UTRR2	reverse	CGGGTTGTCCTTTTCTGTTC	20	CALL4_Intron4		
C4dr	CALL3UTRF7	forward	TTCAGAAACACCAGGTACTTA	21	CALL4_3UTR	4200	58
	CALL3UTRR8	reverse	TCAGTAAGAAAGGGGAATGACAG	23	CALL4_3UTR		

LR-PCR products were visualised by agarose gel electrophoresis as described earlier (Chapter 2.7) and concentrations of LR-PCR products were estimated by comparison of the DNA bands to the GeneRuler 1 kb DNA ladder (Thermoscientific, USA), a standard of known concentration.

7.2.3 Preparation of LR-PCR products for next generation sequencing

7.2.3.1 Preparation of LR-PCR products prior to shipment

LR-PCR products generated for each of the eight animals were assessed using the following steps prior to shipment to the service provider:

1. The LR-PCR products were purified to remove primers, excess dNTPs, polymerase or buffer components that might interfere with the sequencing reactions. The ExoSAP-IT clean up kit (USB, Ohio, USA) or Qiagen QIA quick purification kit (Qiagen, Hilden, Germany) was used for purifying LR-PCR products which showed a single band of the expected size on agarose gel electrophoresis, whereas the Qiagen QIAquick gel extraction kit (Qiagen, Hilden, Germany) was used for purification of two PCR products ('C1cr' and 'C4dr') that consistently produced multiple bands, in addition to the band of the expected sizes of 3.7 and 4.2 kb, respectively. These methods were performed following protocols described earlier (Chapters 2.6.1 and 2.7.1).

2. Multiple 50 µl PCR reactions using DNA from the same animal and the same primer combination were pooled (8 animals x 14 regions = 112 pooled LR-PCR products). Pooling was performed with five to twenty PCRs from the same animal and primer combination pooled according to results from estimated concentration after gel electrophoresis.

3. Each pool of LR-PCR products from the same animal and primer combination was concentrated using the Microcon centrifugal filter devices (Milipore, USA) (blue top

filters with Ultracel YM-100 membrane) following the user guide (Milipore, USA). PCR products of 250 to 750 µl total volumes were concentrated to smaller volumes of 15 to 171 µl.

4. Concentrations and the sizes of the pooled LR-PCR products were determined. Each of the 112 LR-PCR products was plated into individual wells of the microfluidic DNA chips (each chip can fill 12 samples), for analysis with the Agilent 2100 Bioanalyzer, according to the manufacturer's instructions. Analysis of 1 µl aliquots from each sample in the Bioanalyzer involved setting up the chip priming station and Bioanalyzer and preparation of the Agilent DNA 7500 Assay protocol (Agilent Technologies, USA). The Bioanalyzer concentration measurements for each sample were assessed, and those that were not of expected size and did not achieve the required minimum concentration of 70 ng were re-amplified, and steps one to four repeated. LR-PCR products of expected size that exceeded this quality control are referred to as amplicons. Amounts of DNA, concentrations, volumes of pooled LR-PCR and PCR band sizes according to the Bioanalyzer analyses of each LR-PCR pool are shown in Table 7.5. Statistical analyses of the Bioanalyzer data were conducted using Excel software and Genstat Release 16 (http://www.vsni.co.uk/) and shown in Table 7.6.

5. These amplicons were shipped in two 96-well plates individually capped and sealed with parafilm, packaged on 5kg dry-ice and delivered overnight to the sequencing service provider at Life Technologies, Melbourne, Mulgrave, Australia.

7.2.3.2 SOLiD4 barcoded fragment library preparation and sequencing

Sequencing libraries were prepared by the service provider (Life Technologies, Melbourne, Australia) as the first step in which samples (amplicons in two 96-well plates) were adapted for sequencing by oligonucleotide ligation and detection (SOLiD, Life Technologies) sequencing. The samples comprised of fourteen amplicons from each of the eight sheep which were pooled in

equimolar ratios into two pools of seven non-overlapping amplicons (Table 7.4). Each of this pool represents a library, which means there are 2 libraries for each animal.

The barcoded fragment library was utilised in this study. Each library is tagged with a barcode, containing unique 5 - 10 base sequences, on one of the adaptors, to enable multiplexed sequencing analysis where multiple samples are run simultaneously in a single sequencing run.

Preparation of the barcoded fragment libraries using the amplicon pools was performed using the SOLiD4 fragment library barcoding kit (Applied Biosystem, USA) and the SOLiD fragment library construction kit (Applied Biosystem, USA). A summary of the workflow description of the preparation of barcoded fragment libraries and a schematic diagram of a typical barcoded fragment are shown below in Figures 7.2 and 7.3, respectively.

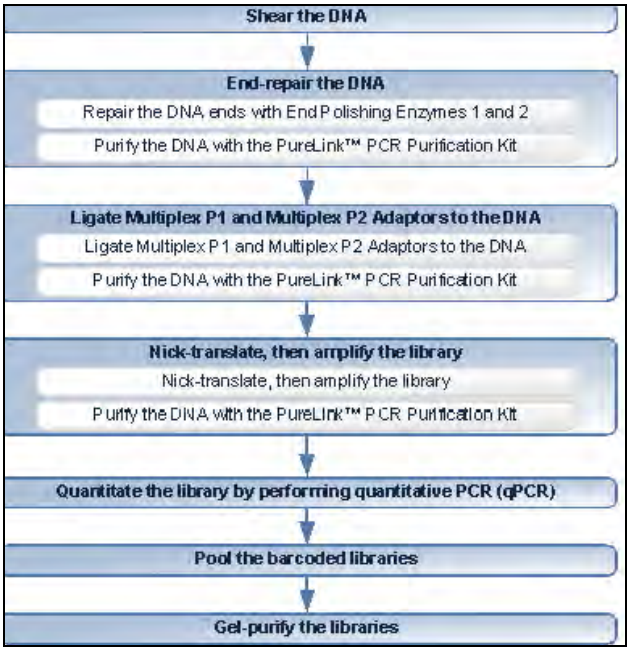


Figure 7.2: Preparation workflow for the barcoded fragment libraries (adapted from Applied Biosystems SOLiD4 system library preparation guide, April 2010).

Summary of construction of the barcoded fragment libraries are as follows:

1. Each pool of seven amplicons was sheared into small DNA fragments with a mean fragment size of 165 bp and a fragment size range of 150 to 180 bp, using the Covaris S2 System.
2. DNA ends of the sheared DNA fragments were end-repaired with End Polishing Enzymes 1 and 2 and purified with the PureLink PCR purification kit (Applied Biosystems, USA).
3. Ends of the end-repaired DNA fragments in each pool of fragmented amplicons were ligated with adaptors (Multiplex P1 Adaptor and a Multiplex P2 Adaptor) and purified with the PureLink purification kit (Applied Biosystems, USA). The Multiplex P2 adaptors included unique barcode sequences (total of 16 barcodes with 1 barcode for each of the two pools of seven amplicons for each of the eight sheep) with each barcode subsequently generating a single fragment library to be multiplexed in the same sequencing run. Sequences for Multiplex P1 and P2 adaptors used for barcoded fragment library construction are listed in Table 7.3. An example of the pool-barcode arrangement is as follows: For animal CPW 156, barcode 1 was ligated to pool 1 (amplicons 1,3,5,7,9,11,13) and barcode 9 was ligated to pool 2 (amplicons 2,4,6,8,10,12,14). Barcode labelings of the remaining seven animals are shown in Table 7.4.
4. After adaptor-ligation and purification, DNA underwent nick translation and was amplified for nine cycles using primers specific to the Multiplex P1 and Multiplex P2 adaptors (Multiplex library PCR-1: CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT and Multiplex library PCR-2: CTGCCCCGGGTTCCTCATTCT primers). After amplification, the library was purified with a PureLink PCR purification kit (Applied Biosystems, USA).
5. The library was quantitated using the SOLiD Library TaqMan quantitation kit (PN 4449639), according to the manufacturers instructions described in ‘SOLiD4 system library quantitation with the SOLiD library TaqMan quantitation kit’.
6. The barcoded fragment libraries were pooled in equal molar amounts.
7. The barcoded fragment libraries were run on a SOLiD library size select selection gel.

Once the barcoded fragment libraries were constructed, the templated bead preparation was performed following the manufacturer's instructions (Applied Biosystems SOLiD system templated bead preparation guide, March 2010). In summary, this involved clonal amplification using an emulsion PCR of each library template on the SOLiD P1 DNA beads, enrichment of the templated beads and deposition of the beads onto a sequencing slide. The beads were then sequenced on the SOLiD Analyzer on a single segment of a four-quadrant slide, with a 50 bp read length expected for the SOLiD4 system. An overview of the sequencing chemistry for the templated bead preparation and sequencing is described at: http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/SOLiD-next-generation-sequencing/next-generation-systems/SOLiD-sequencing-chemistry.html.

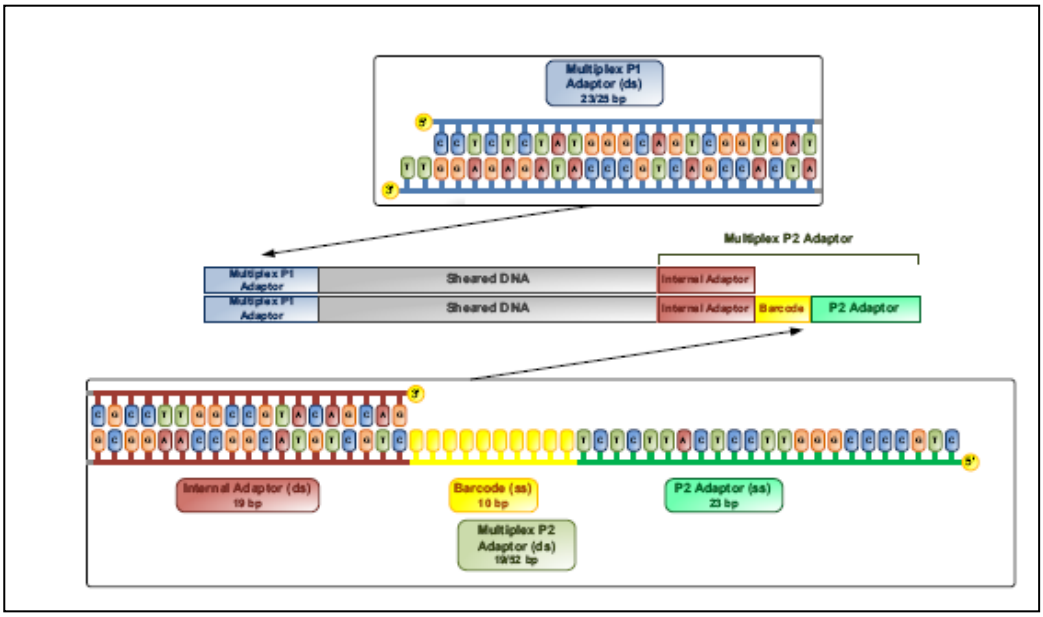


Figure 7.3: Schematic diagram of a typical SOLiD4™ barcoded fragment after library construction (adapted from Applied Biosystems SOLi4 system library preparation guide, April 2010).

Table 7.3: Multiplex P1 and P2 adaptors used for barcoded fragment library construction. Each adaptor consists of two sequences to form a double strand as shown in Fig 7.1. The bold sequence flanked by the internal adaptor sequence on the 5' end and the P2 adaptor sequence on the 3' end is the barcode sequence

Adaptor and barcode	Sequence 5'>3'	Length
Multiplex P1 Adaptor	ATCACCGACTGCCCATAGAGAGGTT	25
	CCTCTCTATGGGCAGTCGGTGAT	23
Multiplex P2 Adaptor with Barcode-001	CGCCTTGGCCGTACAGCAG	19
	CTGCCCCGGGTTCCTCATTCTCT**GTGTAAGAGG**CTGCTGTACGGCCAAGGCG	52
Multiplex P2 Adaptor with Barcode-002	CGCCTTGGCCGTACAGCAG	19
	CTGCCCCGGGTTCCTCATTCTCT**AGGGAGTGGT**CTGCTGTACGGCCAAGGCG	52
Multiplex P2 Adaptor with Barcode-003	CGCCTTGGCCGTACAGCAG	19
	CTGCCCCGGGTTCCTCATTCTCT**ATAGGTTATA**CTGCTGTACGGCCAAGGCG	52
Multiplex P2 Adaptor with Barcode-004	CGCCTTGGCCGTACAGCAG	19
	CTGCCCCGGGTTCCTCATTCTCT**GGATGCGGTC**CTGCTGTACGGCCAAGGCG	52
Multiplex P2 Adaptor with Barcode-005	CGCCTTGGCCGTACAGCAG	19
	CTGCCCCGGGTTCCTCATTCTCT**GTGGTGTAAG**CTGCTGTACGGCCAAGGCG	52
Multiplex P2 Adaptor with Barcode-006	CGCCTTGGCCGTACAGCAG	19
	CTGCCCCGGGTTCCTCATTCTCT**GCGAGGGACA**CTGCTGTACGGCCAAGGCG	52
Multiplex P2 Adaptor with Barcode-007	CGCCTTGGCCGTACAGCAG	19
	CTGCCCCGGGTTCCTCATTCTCT**GGGTTATGCC**CTGCTGTACGGCCAAGGCG	52
Multiplex P2 Adaptor with Barcode-008	CGCCTTGGCCGTACAGCAG	19
	CTGCCCCGGGTTCCTCATTCTCT**GAGCGAGGAT**CTGCTGTACGGCCAAGGCG	52
Multiplex P2 Adaptor with Barcode-009	CGCCTTGGCCGTACAGCAG	19
	CTGCCCCGGGTTCCTCATTCTCT**AGGTTGCGAC**CTGCTGTACGGCCAAGGCG	52
Multiplex P2 Adaptor with Barcode-010	CGCCTTGGCCGTACAGCAG	19
	CTGCCCCGGGTTCCTCATTCTCT**GCGGTAAGCT**CTGCTGTACGGCCAAGGCG	52
Multiplex P2 Adaptor with Barcode-011	CGCCTTGGCCGTACAGCAG	19
	CTGCCCCGGGTTCCTCATTCTCT**GTGCGACACG**CTGCTGTACGGCCAAGGCG	52
Multiplex P2 Adaptor with Barcode-012	CGCCTTGGCCGTACAGCAG	19
	CTGCCCCGGGTTCCTCATTCTCT**AAGAGGAAAA**CTGCTGTACGGCCAAGGCG	52
Multiplex P2 Adaptor with Barcode-013	CGCCTTGGCCGTACAGCAG	19
	CTGCCCCGGGTTCCTCATTCTCT**GCGGTAAGGC**CTGCTGTACGGCCAAGGCG	52
Multiplex P2 Adaptor with Barcode-014	CGCCTTGGCCGTACAGCAG	19
	CTGCCCCGGGTTCCTCATTCTCT**GTGCGGCAGA**CTGCTGTACGGCCAAGGCG	52
Multiplex P2 Adaptor with Barcode-015	CGCCTTGGCCGTACAGCAG	19
	CTGCCCCGGGTTCCTCATTCTCT**GAGTTGAATG**CTGCTGTACGGCCAAGGCG	52
Multiplex P2 Adaptor with Barcode-016	CGCCTTGGCCGTACAGCAG	19
	CTGCCCCGGGTTCCTCATTCTCT**GGGAGACGTT**CTGCTGTACGGCCAAGGCG	52

Table 7.4: Amplicon pool-barcode arrangements used in the SOLiD barcoded fragment library preparations. Each box represents an individual sheep and the two-barcoded pools of seven equimolar non-overlapping amplicons.

CPW156 (SH_N)		**SH1038/08 (SH_C2)**	
Barcode-001 (sheep1_1)*	**Barcode-009 (sheep1_9)***	**Barcode-005 (sheep5_5)***	**Barcode-013 (sheep5_13)***
1ar, 2ir, 3dr, 4ar, 5ir, C1cr, C3cr	1hr, 2jr, 3ir, 4ir, 5fr, C2ar, C4dr	1ar, 2ir, 3dr, 4ar, 5ir, C1cr, C3cr	1hr, 2jr, 3ir, 4ir, 5fr, C2ar, C4dr
SH1022/07 (SH_A1)		**SH1039/08 (SH_A3)**	
Barcode-002 (sheep2_2)*	**Barcode-010 (sheep2_10)***	**Barcode-006 (sheep6_6)***	**Barcode-014 (sheep6_14)***
1ar, 2ir, 3dr, 4ar, 5ir, C1cr, C3cr	1hr, 2jr, 3ir, 4ir, 5fr, C2ar, C4dr	1ar, 2ir, 3dr, 4ar, 5ir, C1cr, C3cr	1hr, 2jr, 3ir, 4ir, 5fr, C2ar, C4dr
SH1032/08 (SH_C1)		**L06 (M_N)**	
Barcode-003 (sheep3_3)*	**Barcode-011 (sheep3_11)***	**Barcode-007 (sheep7_7)***	**Barcode-015 (sheep7_15)***
1ar, 2ir, 3dr, 4ar, 5ir, C1cr, C3cr	1hr, 2jr, 3ir, 4ir, 5fr, C2ar, C4dr	1ar, 2ir, 3dr, 4ar, 5ir, C1cr, C3cr	1hr, 2jr, 3ir, 4ir, 5fr, C2ar, C4dr
SH1033/08 (SH_A2)		**L07 (M_A)**	
Barcode-004 (sheep4_4)*	**Barcode-012 (sheep4_12)***	**Barcode-008 (sheep8_8)***	**Barcode-016 (sheep8_16)***
1ar, 2ir, 3dr, 4ar, 5ir, C1cr, C3cr	1hr, 2jr, 3ir, 4ir, 5fr, C2ar, C4dr	1ar, 2ir, 3dr, 4ar, 5ir, C1cr, C3cr	1hr, 2jr, 3ir, 4ir, 5fr, C2ar, C4dr

Parenthesis following each animal's ID indicates breed and disease status of each sheep. CPW= Coopwoth, M= Merino, SH= South Hampshire sheep, A= affected, C= carrier, N= normal
* ID used in the Gbrowse Genome Browser (see Figure 7.6)

7.2.4 Bioinformatic analysis

Initial bioinformatic analysis and technical support for the SOLiD sequence data was performed by Drs John Davis (Applied Biosystems, Mulgrave, Melbourne) and Matthew Hobbs (University of Sydney). This involved identification of polymorphisms (SNPs and small indels), aligning reads to the provided ovine reference sequence and visualisation of the annotated data on the generic genome browser (GBrowse version 2.38), which is a part of the generic model organism GMOD suite of genome analysis software tools (Stein et al., 2002).

Samples were labeled according to their respective amplicons pools and barcodes (Table 7.4): sheep 1 (CPW156) samples were labelled 'sheep1_1' corresponding to sheep 1 pool with barcode 001, and 'sheep 1_9' corresponding to sheep 1 pool with barcode 009.

Preliminary bioinformatic analysis was performed by separating reads according to their respective barcodes and aligning the generated sequence reads to the ovine reference sequence generated in Chapter 5, using the SOLiD system analysis pipeline tool. The output generated from alignment was in BAM (binary alignment map) file format that stores large numbers of nucleotide sequence alignments. Two BAM files were generated for each of the eight sheep. Also obtained were FASTA format files of nucleotide sequences to which reads had been mapped.

7.2.4.1 Polymorphism discovery: SNP detection

The reads were aligned to the reference sequence and the resulting alignment BAM files were processed by the SNP calling software using SAMtools (Li et al., 2009) on the SOLiD USA cluster. Output files were transferred to the service provider in Australia. SNP calling aims to determine in which positions at least one of the bases differs from a reference sequence. A range of quality control data for each SNP is provided (e.g. SNP occurred in how many reads/ likely to be heterozygous or homozygous). The SNP discovery BAM files representing the output of the SOLiD SNP pipeline without any downstream filtering or analysis were provided to our group for further analysis. The SNP pipeline also generated general feature format (GFF) files, which are text files used to store annotations for visualisation in genome annotation viewers (e.g. IGV, GBrowse).

SNP detection on SOLiD sequencing data was performed using the BioScope software resequencing find SNPs analysis tool (Applied Biosystems, USA). The input files required for SNP detection include the reference sequence in FASTA file format (same file used previously for mapping), the BAM file containing nucleotide sequence alignments, and their corresponding position error and probe error files which record the frequencies of di-color mismatches between reads and the reference at different positions in a read.

Default configurations of the SNP parameter values and the BioScope software were defined by the service provider. The four output files generated are a ‘SNP.gff’ file, which contains the list

of SNPs, a ‘Consensus_Calls.txt’ file, which covers all positions that have coverage and provides general information about each position, a ‘consensus_Basespace2.fasta’ file, which is the updated FASTA file of the chromosome sequence with SNP sites encoded in IUB codes and N for all non-covered positions, and a ‘quartiles.txt’ file, which lists the quartile and percentile information about the coverage and color quality value distribution of all positions of the chromosome.

7.2.4.2 Polymorphism discovery: Small indel detection

For identification of small insertions or deletions (indels) on the SOLiD fragment data, the alignment BAM files generated were processed by the SOLiD indel pipeline using Bioscope 1.3 (Applied Biosystems) on the SOLiD US cluster and transferred back to the service provider in Australia. Ranges of quality control data were provided for each indel. The indel BAM files representing the output of the SOLiD indel pipeline without any downstream filtering or analysis were provided for further analysis. The indel pipeline also generated GFF file formats to be viewed in genome annotation viewers.

Small indel detection was performed on SOLiD sequencing data using the BioScope software small indel caller, using BAM files produced from the barcoded fragment libraries. The small indel pipeline, by default, is able to determine insertions up to 3 bp and deletions up to 11 bp.

The input files required for small indel detection include the reference sequence in FASTA file format (the same file used previously for mapping), the *.bam file containing nucleotide sequence alignments, and the CMAP reference file. The output files generated include a smallIndel.gff file, which contains the list of small indels, the TXT file which provides an alternative format of the data contained in the GFF file, SQL file formats and an UNGAPPED file which contains more details of the reads counted towards the number of ungapped alignments used for the coverage ratio.

7.2.4.3 Polymorphism evaluation

Both the SNP and indel GFF files were transferred to an Excel spreadsheet, sorted by position, sheep and animal attributes and then visually assessed for any SNPs or indels that segregate with the disease phenotype.

BAM and GFF files obtained from SNP and indel detections were uploaded into password protected GBrowse software (http://crcidp.vetsci.usyd.edu.au/cgi-bin/gb2/gbrowse/NCL/) containing various horizontal tracks uploaded by the user, such as coverage, *CLN6* coding sequence and ovine reference sequence. All three (alignment, SNP and indel) BAM files could be visualised in the genomics viewer.

Following the release of Version 3.1 of the ovine genome assembly in 2012 (OARv3.1, http://www.livestockgenomics.csiro.au/sheep/oar3.1.php) sequence of the OARv3.1 (GenBank accession CM001588.1) at positions 14,796,391 - 14,845,049 was also uploaded into the Gbrowse viewer to allow comparison between these two versions, after initial analysis in 2010 with Version 2.0 (OARv2.0; Archibald et al., 2010).

7.3 Results

7.3.1 Generation of LR-PCR products

Fourteen LR-PCR products ranging from approximately 2.6 to 5.3 kb in length were generated with partial overlaps between 110 bp (between products: '4ir' - '5ir') and 1.34 kb (between products: '1hr' - '2ir'). These contiguous products cover a region of 49,123 kb of OAR7 spanning from approximately 9.1 kb upstream of the *CLN6* start codon to approximately 6.7 kb downstream of the *CAMLM4* stop codon (Figure 7.1). Sequences generated from these LR-PCR products are shown in Appendix 5.

7.3.2 Bioanalyzer data analysis

Multiple LR-PCRs generated for the same animal and the same amplification region were pooled and concentrated into smaller volumes using the Microcon centrifugal filter devices to achieve a minimum concentration of 0.5 ng/µl required for analysis on the Agilent 2100 bioanalyzer. A summary of the bioanalyzer data (product size, concentration, volume and total sample amount) was made (Table 7.5). Molarity information enables equimolar representation of each animal for each PCR region prior to SOLiD sequencing. Concentrations of pooled LR-PCR product ranged from 0.51 to 16.98 ng/µl and final amounts of products ranged from 70 to 353.5 ng, with a mean of 225 ± 66 ng (Table 7.5). These values met requirements for quality control purposes for determining a minimum amount of product that must be present for sequencing, except for product 'C4d'.

Table 7.5: Bioanalyzer data of 112 LR-PCR products for SOLiD sequencing. DNA represents eight sheep: CPW156 (SH_N), SH1022 (SH_A1), SH1032 (SH_C1), SH1033 (SH_A2), SH1038 (SH_C2), SH1039 (SH_A3), L06 (M_N) and L07 (M_A).

PCR Product	Expected Size (kb)	DNA	Actual Size (kb)	Concentration (ng/µl)	*Volume (µl)	Total Amount (ng)	Molarity (nmol/l)
1ar	2.6	SH_N	2.8	4.9	50	245	2.6
		SH_A1	2.7	7.83	42	328.86	4.5
		SH_C1	2.6	9.18	30	275.4	5.3
		SH_A2	2.6	9.41	30	282.3	5.4
		SH_C2	2.5	6.79	41	278.39	4.1
		SH_A3	2.5	6.16	41	252.56	3.7
		M_N	2.7	4.13	56	231.28	2.3
		M_A	2.5	4.68	62	290.16	2.8
1hr	4.6	SH_N	5.5	1.52	140	212.8	0.4
		SH_A1	4.9	1.63	110	179.3	0.5
		SH_C1	4.7	0.9	135	121.5	0.3
		SH_A2	4.7	1.65	140	231	0.5
		SH_C2	4.7	0.72	160	115.2	0.2
		SH_A3	4.6	0.89	134	119.26	0.3
		M_N	4.6	0.51	171	87.21	0.2
		M_A	4.7	1.8	110	198	0.6
2ir	3.8	SH_N	4	3.4	70	238	1.3
		SH_A1	3.3	1.9	103	195.7	0.9
		SH_C1	3.4	4.48	57	255.36	2.0
		SH_A2	3.3	1.78	107	190.46	0.8
		SH_C2	3.4	2.75	93	255.75	1.2
		SH_A3	3.5	1.2	127	152.4	0.5
		M_N	3.6	7.13	36	256.68	3.0
		M_A	3.4	1.05	130	136.5	0.5
2jr	3.7	SH_N	3.6	1.35	135	182.25	0.6
		SH_A1	3.7	1.05	145	152.25	0.4
		SH_C1	3.6	1.29	145	187.05	0.6
		SH_A2	3.5	1.64	140	229.6	0.7
		SH_C2	3.7	1.06	160	169.6	0.4
		SH_A3	3.5	1.88	138	259.44	0.8
		M_N	3.7	1.46	90	131.4	0.6
		M_A	3.8	1.25	150	187.5	0.5
3dr	3.0	SH_N	2.9	3.96	57	225.72	2.1
		SH_A1	2.8	1.3	88	114.4	0.7
		SH_C1	2.8	1.93	98	189.14	1.1
		SH_A2	2.8	2.1	97	203.7	1.2
		SH_C2	2.8	2.1	100	210	1.2
		SH_A3	2.7	2.75	71	195.25	1.5
		M_N	2.7	1.56	52	81.12	0.9
		M_A	2.7	1.47	48	70.56	0.8

PCR Product	Expected Size (kb)	DNA	Actual Size (kb)	Concentration (ng/µl)	*Volume (µl)	Total Amount (ng)	Molarity (nmol/l)
3ir	4.7	SH_N	4.8	7.3	37	270.1	2.3
		SH_A1	4.8	3.51	70	245.7	1.1
		SH_C1	4.8	9.11	30	273.3	2.9
		SH_A2	4.7	5.73	60	343.8	1.8
		SH_C2	4.8	5.16	60	309.6	1.6
		SH_A3	4.9	4.18	60	250.8	1.3
		M_N	4.7	7.74	26	201.24	2.5
		M_A	4.7	9.37	26	243.62	3.0
4ar	4.2	SH_N	4.2	2.38	30	71.4	0.8
		SH_A1	5	0.94	115	108.1	0.3
		SH_C1	4.5	1.76	51	89.76	0.6
		SH_A2	4.6	0.79	128	101.12	0.3
		SH_C2	4.7	1	70	70	0.3
		SH_A3	4.1	2.3	68	156.4	0.9
		M_N	4.5	3.35	55	184.25	1.1
		M_A	4.7	1.79	72	128.88	0.6
4ir	5.3	SH_N	5.6	8.35	36	300.6	2.3
		SH_A1	5.3	3.84	67	257.28	1.1
		SH_C1	5.3	6.49	40	259.6	1.8
		SH_A2	5.3	6.99	38	265.62	2.0
		SH_C2	5.4	5.64	50	282	1.6
		SH_A3	5.5	5.49	50	274.5	1.5
		M_N	5.3	4.07	50	203.5	1.2
		M_A	5.3	8.46	33	279.18	2.4
5ir	4.1	SH_N	4.2	2.23	98	218.54	0.8
		SH_A1	4.2	2.39	90	215.1	0.9
		SH_C1	4.2	2.39	96	229.44	0.9
		SH_A2	4.2	2.69	94	252.86	1.0
		SH_C2	4.3	4.23	50	211.5	1.5
		SH_A3	4.2	2.36	93	219.48	0.9
		M_N	4.3	3.11	76	236.36	1.1
		M_A	4.2	5.21	55	286.55	1.9
5fr	3.3	SH_N	3.7	8.2	30	246	3.3
		SH_A1	3.5	7.42	35	259.7	3.2
		SH_C1	3.6	7.15	35	250.25	3.0
		SH_A2	3.5	6.01	36	216.36	2.6
		SH_C2	3.5	5.05	70	353.5	2.2
		SH_A3	3.8	5.76	50	288	2.3
		M_N	3.4	16.24	16	259.84	7.2
		M_A	3.4	14.59	20	291.8	6.5
C1cr	3.7	SH_N	4.5	4.77	46	219.42	1.6
		SH_A1	4.5	7.33	45	329.85	2.5
		SH_C1	4.9	4.81	44	211.64	1.5
		SH_A2	5.1	7.35	45	330.75	2.2
		SH_C2	4.5	6.32	36	227.52	2.1
		SH_A3	5.1	3.35	45	150.75	1.0
		M_N	4.5	14.92	20	298.4	5.1
		M_A	4.7	4.17	60	250.2	1.3

PCR Product	Expected Size (kb)	DNA	Actual Size (kb)	Concentration (ng/µl)	*Volume (µl)	Total Amount (ng)	Molarity (nmol/l)
C2ar	3.5	SH_N	3.6	10.83	25	270.75	4.6
		SH_A1	3.7	13.51	18	243.18	5.5
		SH_C1	3.6	14.32	18	257.76	6.0
		SH_A2	3.6	11.59	22	254.98	4.8
		SH_C2	3.9	12.68	20	253.6	5.0
		SH_A3	3.7	11.17	24	268.08	4.6
		M_N	3.8	16.98	15	254.7	6.8
		M_A	3.8	14.84	20	296.8	6.0
C3cr	3.5	SH_N	3.4	6.98	36	251.28	3.1
		SH_A1	3.4	7.65	40	306	3.4
		SH_C1	3.3	9.82	29	284.78	4.5
		SH_A2	3.4	9.16	28	256.48	4.1
		SH_C2	3.3	7.35	40	294	3.4
		SH_A3	3.3	8.04	30	241.2	3.7
		M_N	3.7	7.08	40	283.2	2.9
		M_A	3.4	5.3	50	265	2.3
C4dr	4.2	SH_N	see Fig. 7.4	<0.5	25	#	#
		SH_A1	see Fig. 7.4	<0.5	25	#	#
		SH_C1	see Fig. 7.4	<0.5	25	#	#
		SH_A2	see Fig. 7.4	<0.5	25	#	#
		SH_C2	see Fig. 7.4	<0.5	25	#	#
		SH_A3	see Fig. 7.4	<0.5	25	#	#
		M_N	see Fig. 7.4	<0.5	25	#	#
		M_A	see Fig. 7.4	<0.5	25	#	#

* total volume of PCR pool to achieve minimum 70 ng of each LR-PCR product per animal
# concentration too low for detection using the Bioanalyzer. See Figure 7.4 for bands visualisation and estimated concentration.

The concentration of pooled product 'C4dr' for all animals remained too low (<0.5 ng/µl) for detection by the bioanalyzer kit used, despite exhaustive repetition and optimisation attempts. As this region was at the margin of the region of interest and did not contain any parts of *CLN6*, further optimisation was abandoned and the samples with low concentrations were included in the sequencing (Figure 7.4). The concentration of the band was calculated to be approximately 0.5 ng using the concentration of the lowest band of the size marker.

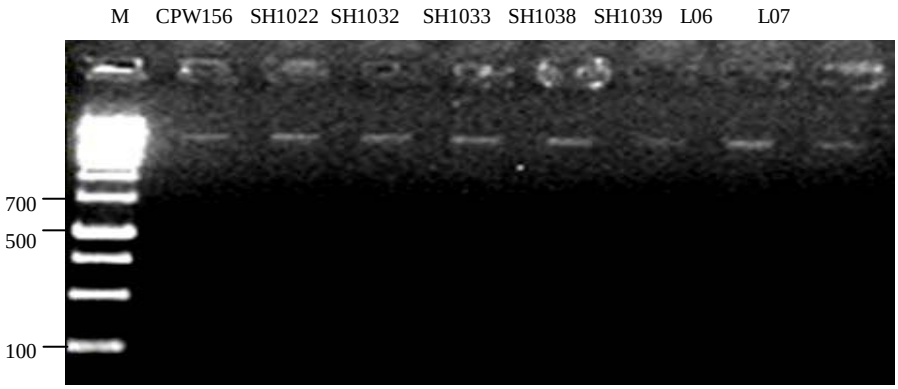


Figure 7.4: Agarose gel visualisation of pooled PCR products C4dr generated with primers CALL3UTRF7 and CALL3UTRR8. The identification for each animal is shown at the top of the gel image. Relevant fragment sizes of the Fermentas GeneRuler 1 kb Plus DNA Ladder (M) is shown on the far left of the 1% agarose gel. The expected length of the products is 4.2 kb in all eight animals.

For most of the LR-PCR products, the bioanalyzer analysis (Table 7.5) confirmed results from previous agarose gel analyses where single bands of expected sizes were identified. However, for all animals, product '2jr' consistently produced a 67 bp band in addition to the expected product of approximately 3.7 kb. Figure 7.5 shows an example of the bioanalyzer electropherogram trace for normal Merino sheep (L06).

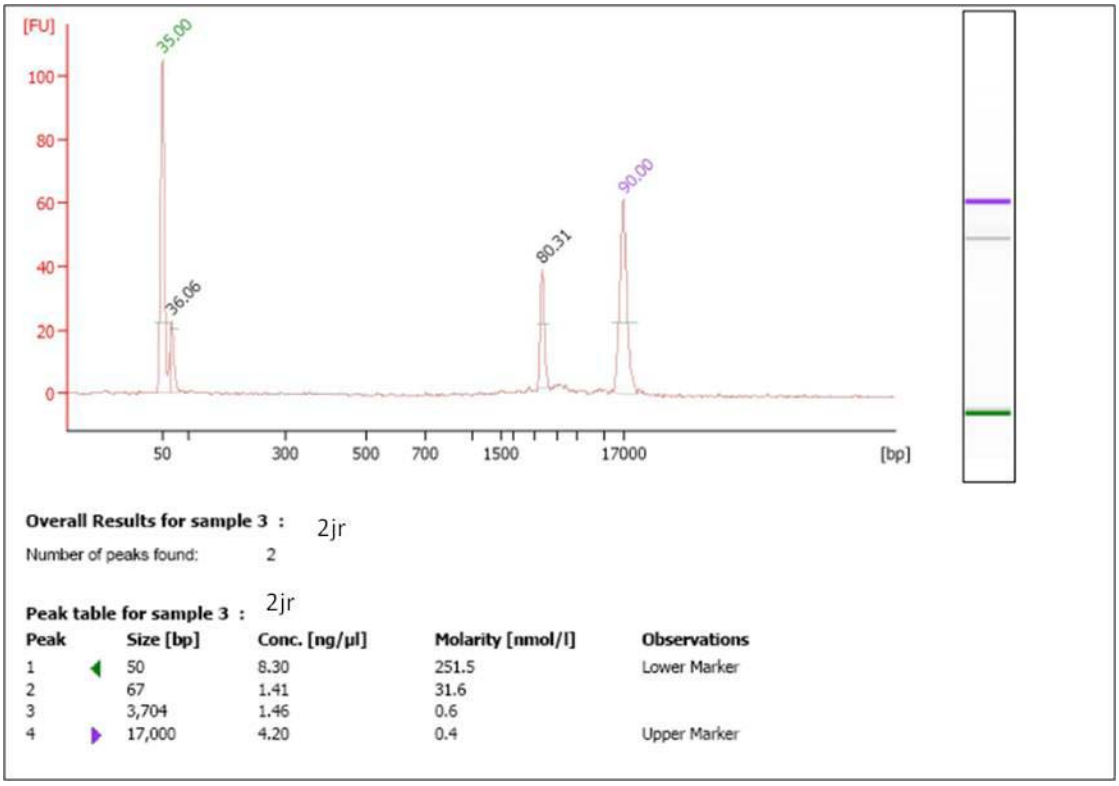


Figure 7.5: Electropherogram shows the results of DNA quantitation of PCR product '2jr' for sheep L06 (Merino normal) using the Agilent 2100 Bioanalyzer at default settings. The graph displays plot of DNA fragment size (bp) versus fluorescence intensity (FU). The 3.704 kb DNA fragment with 80.21 FU represents the expected band and the 67 bp with 36.06 FU represents the additional unexpected band. DNA lower and upper markers, 50 and 17,000 bp in size on the first and last peaks, respectively, are internal standards. The results table provides information of the two detected DNA fragments with their respective concentration. The diagram on the far right is the corresponding *in silico* gel image.

Table 7.6: Statistical analysis of expected and actual band sizes from all 112 PCR products. Analyses were performed using Excel software and Genstat Release 16 (http://www.vsni.co.uk/).

Product	Expected Size (kb)	Actual Size mean (kb) ± SD	P-value	Coefficient of variation	Sizing accuracy (at 10%CV*) 2s.d	Sizing accuracy (at 10%CV*) 1s.d
1ar	2.6	2.613 ± 0.1	0.763	4.310015795	0.261	0.131
1hr	4.6	4.8 ± 0.3	0.099	6.20019841	0.48	0.24
2ir	3.8	3.488 ± 0.2	0.006	6.581164982	0.349	0.174
2jr	3.7	3.638 ± 0.1	0.14	2.915904252	0.364	0.182
3d	3	2.775 ± 0.1	0	2.548132545	0.278	0.139
3ir	4.7	4.775 ± 0.1	0.02	1.480851898	0.478	0.239
4ar	4.2	4.538 ± 0.3	0.013	6.336943631	0.454	0.227
4ir	5.3	5.375 ± 0.1	0.111	2.16737627	0.538	0.269
5ir	4.1	4.225 ± 0.05	0.00012	1.095645088	0.423	0.211
5fr	3.3	3.55 ± 0.1	0.002	3.983700176	0.355	0.178
C1cr	3.7	4.725 ± 0.3	0	5.740540452	0.473	0.236
C2ar	3.5	3.713 ± 0.1	0.001	3.032974078	0.371	0.186
C3cr	3.5	3.4 ± 0.1	0.068	3.850903945	0.34	0.17
C4dr	4.2	nil	nil	nil	nil	nil

* Agilent 2100 Bioanalyzer sizing accuracy
(http://www.genomics.agilent.com/article.jsp?pageId=2423&_requestid=869374)

Sizing by the bioanalyzer identified that the size of PCR pools was often not identical with the expected size, and variations in size between PCR products of the same PCR but different animals were observed. The mean actual sizes in several of the pools was significantly different from the expected sizes even when considering a reported Bioanalyzer sizing reproducibility of 5% CV, as shown by the one sample t-test (Table 7.6). The p-value generated from the student T-test at two-tailed distribution is shown in the 'P-value' column.

This is not unexpected as the reference sequence could include assembly errors and there might be differences between the BAC/reference sequence and sheep analysed here. A large difference in mean value is seen in PCR product 'C1cr' with an actual mean of 4.725 ± 0.3 compared to an expected size of 3.7 kb. This region covered upstream sequence of *CALML4* and did not contain any parts of *CLN6*. However, as indicated in Chapter 5, regulatory elements affecting *CLN6* could be located in this region and *CALML4* could be a gene of interest if no disease causing mutation was found in *CLN6*.

To investigate if variation in size among animals for the same PCR product was within sizing error of the Bioanalyzer 2100 (reported to be 5% CV) or related to disease status or breed, the average (mean) band size for each PCR pool, along with the extent of variability of sizes among animals in relation to this mean was calculated. It was assumed using ANOVA (Analysis of variance, Table 7.6) that the LR-PCR products for the same region are identical among animals (irrespective of disease status or the occurrence of repeats). In Table 7.6, sizing for most of the PCR pools ('1ar', '2jr', '3dr', '3ir', '4ir', '5ir', '5fr', 'C2ar' and 'C3cr') was considered moderately to highly reproducible (CVs of 5% or less). Sizing reproducibility for four PCR pools ('1hr', '2ir', '4ar', 'C1cr'), was appreciably worse at 5.7% CV (C1c) to 6.6% CV (2i). 'C4dr' could not be assessed, as discussed previously.

Standard deviation of the actual PCR band size (in kb) ranged between 0.05 ('5ir') and 0.3 ('1hr', '4ar' and 'C1cr'), and the coefficient of variation was between 1.09 ('5ir') and 6.6 ('2ir'). From these two statistical analyses, the variation between animals for the same PCR products at '1hr', '2ir', '4ar' and 'C1cr' was significantly different in relation to mean of the population, with high values of 6.2, 6.6, 6.3 and 5.7, respectively. This could arise from breed differences or from individual variation in the number and extent of repeat units. The concept that size

differences could be associated with disease status was considered, but none of the PCR pools with size variations showed a significant association of size with disease phenotype.

7.3.3 SOLiD sequencing of LR-PCR products

Sequencing of the 112 LR-PCR products in a single slide SOLiD run resulted in 45,923,380 raw 50-mer reads at approximately 30X coverage. Individual reads ranged between 1,486,412 and 6,061,520 for each barcode containing 7 amplicons, per animal.

7.3.3.1 Identification of SNPs

A total of 238 SNPs were discovered in the eight sheep (Table 7.7). In this table, the final two columns refer to SNP genotypes identified in 8 sheep sequenced using the SOLiD platform. In each column, numbers on the left of the divide symbol refer to the total homozygous and heterozygous SNP identified in the particular animal, whereas numbers on the right of the divide symbol refer to those in the coding region. For example, sheep CPW156 has a total of 116 SNP with 60 being homozygous and 45 heterozygous. Out of these 60 homozygous SNP, one is in the coding region.

One hundred and forty seven of these SNPs were unique, and 91 occurred in more than one animal. All affected South Hampshire and Merino sheep showed a low number of SNPs (6 – 11 SNPs) whereas carrier and normal sheep displayed a higher level of variation from the reference sequence (11 – 51 SNPs).

Table 7.7: Summary of SNP discoveries.

No	Sheep ID	Breed	Phenotype	SNP		
				Total	Homozygous	Heterozygous
					Total /coding#	Total /coding#
1	CPW156	Coopworth	Normal	116	60/1*	56/0
2	SH1022/07	South Hampshire	Affected	9	3/0	6/0
3	SH1032/08	South Hampshire	Carrier	106	15/0	91/1*
4	SH1033/08	South Hampshire	Affected	11	3/0	8/0
5	SH1038/08	South Hampshire	Carrier	51	2/0	49/1*
6	SH1039/08	South Hampshire	Affected	6	3/0	3/0
7	L06	Merino	Normal	115	64/1*	51/0
8	L07	Merino	Affected	7	6/1**	1/0

number of SNP in the coding region of the *CLN6* and *CALML4*
* previously known SNP used as an indirect DNA test for NCL in South Hampshire sheep c.822G>A (Tammen et al., 2006)
** previously known disease causing mutation for NCL in Merino sheep c.184C>T (Tammen et al., 2006)

SNPs that segregated with the disease phenotype or that mapped to the *CLN6* or *CALML4* coding sequence were given the highest priority for further analysis. Two previously known SNPs were found to be in coding regions of *CLN6* (Table 7.7) and described in Table 7.8. No SNPs were identified in *CALML4* coding regions. These SNPs were in reverse complement sequences because they corresponded to the aligned ovine reference sequence, which was also in reverse complement. The c.184C>T (p.Arg62Cys) SNP in exon 2 previously characterised as the disease causing mutation in the Merino sheep (Tammen et al., 2006) was identified, as expected, only in the affected Merino sheep, L07 and this sheep was homozygous for this SNP.

The second known SNP identified was the silent c.822G>A polymorphism in exon 7, which was used as an indirect NCL DNA test for the selection of SH animals in this study (Tammen et al., 2006). It was identified, as expected, in the normal Coopworth sheep CPW156 and Merino sheep L06, and these sheep were homozygous for this SNP (the ovine reference sequence contained an A nucleotide in this position which is associated with the disease in the South Hampshire sheep). The SNP was also identified in both carrier sheep SH1032/08 and SH1038/08, and these sheep were heterozygous for this SNP. The remaining 236 SNPs were in the non-coding regions and did not segregate with the NCL disease phenotype in sheep.

Table 7.8: Summary of previously known SNPs discovered within ovine *CLN6* in Coopworth, South Hampshire and Merino sheep. Sequences and SNPs are in reverse complement (RC) as they correspond to the ovine reference sequence.

Position (bp)	Mutation	Region	Genotype	Ref	No	Sheep ID	Breed	Phenotype	Description	Mutation significance
33,605	c.184C>T	Exon 2	AA	GG	8	LO7	Merino	Affected	Single nucleotide substitution	direct DNA test
	(Rev Com: G>A)									
24, 608	c.822G>A	Exon 7	CC	TT	1	CPW156	Coopworth	Normal	Silent mutation/ Polymorphism	indirect DNA test, LOD score with affected SH: 13.3
	(Rev Com: C>T)		Y=CT	TT	3	SHW1032/ 08	South Hampshire	Carrier	Silent mutation/ Polymorphism	indirect DNA test, LOD score with affected SH: 13.3
			Y=CT	TT	5	SH1038/08	South Hampshire	Carrier	Silent mutation/ Polymorphism	indirect DNA test, LOD score with affected SH: 13.3
			CC	TT	7	L06	Merino	Normal	Silent mutation/ Polymorphism	LOD score with affected Merino: 7.83

7.3.3.2 Discovery of small indels

A total of 425 small indels (73 insertions and 352 deletions) were discovered in the eight sheep (Table 7.9) of which 237 were unique (observed only in one of the eight animals) and 188 occurred in more than one animal. Most of these indels were the size of a single nucleotide and the largest indel was a 9 bp deletion. Overall there were 31 homozygous and 1145 hemizygous indel sites among all the animals. Indels found in the coding sequence of *CLN6* (exons 4, 5, 6, and 7) and *CALML4* (exons 3, 4 and 5) did not segregate with the NCL disease phenotypes.

Table 7.9: Summary of small INDEL discovery

No.	Sheep ID	Breed	Disease status	Insertion	Deletion	Homozygous indels
1	CPW156	Coopworth	Normal	29	144	2
2	SH1022/07	South Hampshire	Affected	21	104	5
3	SH1032/08	South Hampshire	Carrier	23	119	4
4	SH1033/08	South Hampshire	Affected	26	97	3
5	SH1038/08	South Hampshire	Carrier	24	94	4
6	SH1039/08	South Hampshire	Affected	25	131	4
7	L06	Merino	Normal	33	159	5
8	L07	Merino	Affected	26	121	4

7.3.4 Visualisation of sequence coverage on the Gbrowse genome viewer

SOLiD sequences generated for each of the eight sheep were visualised using the GBrowse genome viewer (Stein et al., 2002). This browser consisted of two panels: an overview panel consisting of horizontal tracks for the ovine reference sequence (Ref_NCLseq) and gene transcripts (gene) for both the *CLN6* and *CALML4* genes, and a details panel consisting of various horizontal tracks uploaded by the user, with each track corresponding to the overview panel. User uploaded tracks included SOLiD sequence tracks for which the x-axis depicts the number of reads for each sheep track, as well as other derived tracks, including GC content, predicted PCR position tracks, sequences from the ovine genome version 3.1, conserved non-coding sequence and repeat elements.

A screenshot of the browser (Figure 7.6) in the details section shows 18 tracks consisting of the location of the partially overlapping 14 LR-PCR regions named '1ar' to 'C4dr' (PCR product), a GC content plot (GC content) as well as the 16 sets of sequence reads consisting of two tracks (each containing 7 non-overlapping LR-PCR products) for each of the eight sheep sequenced. These 16 sets of reads are labelled according to their respective unique barcodes (Table 7.4).

The distribution of the reads varied among the different PCR regions and animals (Figure 7.6) showing differences in depth of sequence coverage. Reads within the LR-PCR regions also showed great variability (0 to 7999 reads) across all animals, overall with a good coverage achieved in the key area of interest (PCR regions '1ar' to '5ar') containing *CLN6* and immediate flanking sequences. Twelve small blocks of reads (666 bp) could not be aligned to the reference sequence (see 'unmappable' track in Figure 7.6). There appear to be no specific sequence reads for PCR region 'C4dr' for all sheep, perhaps because of the problems with generating sufficient amounts of PCR product for sequencing as described above (Chapter 7.3.2).

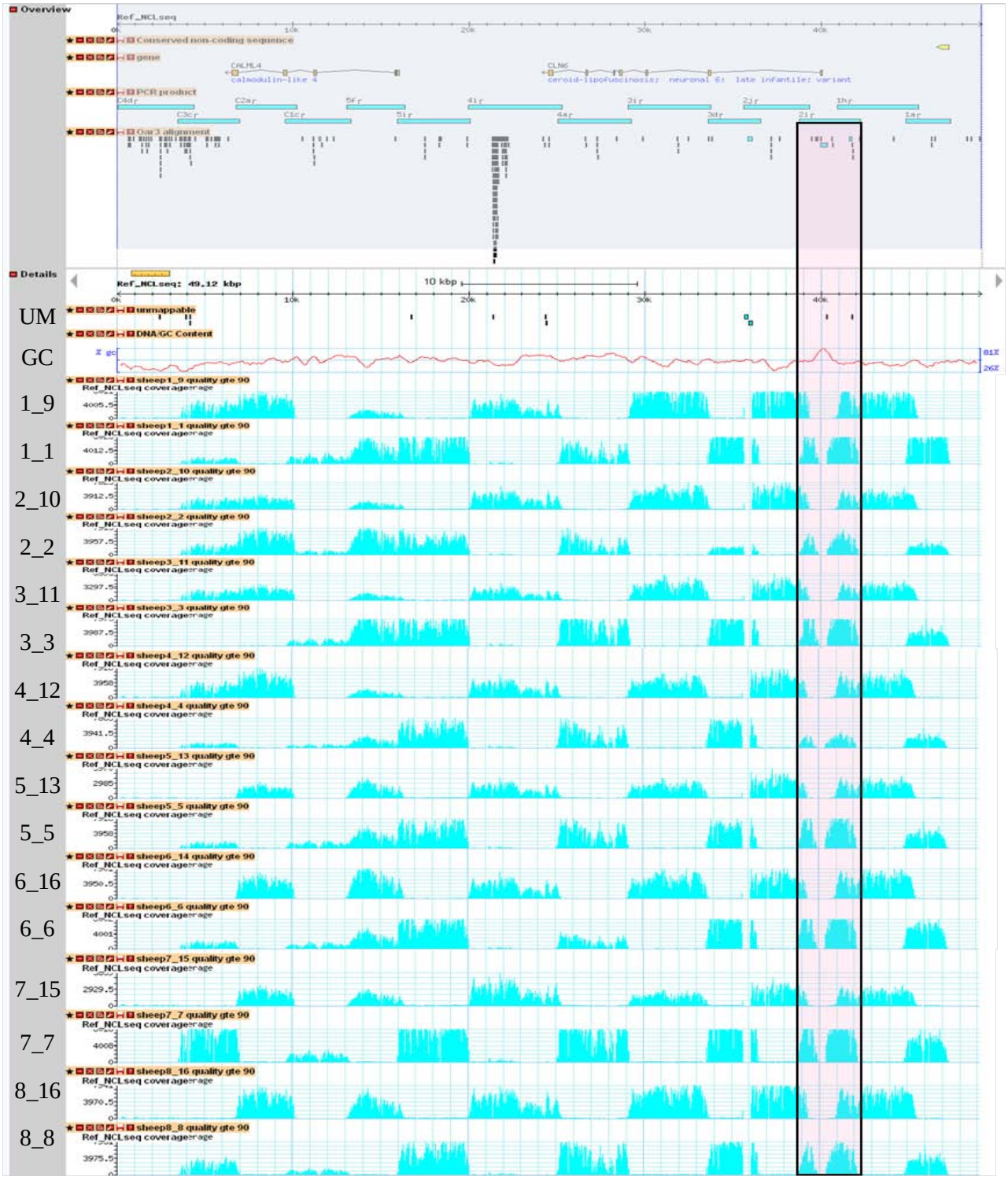


Figure 7.6: A 49,123 kb screenshot of the Gbrowse genome browser depicting data mapped to the ovine reference sequence (Ref_NCLseq): *CLN6* and *CALML4* (gene), 14 LR-PCR products (PCR product) and OARv3.1 genome sequence (Oar3 alignment) in the overview panel. Tracks in the details panel are unmappable sequence (UM), GC content plot (GC) and 16 sequence reads from all eight sheep (1_9 and 1_1 normal Coopworth, 2_10 and 2_2, 4_12 and 4_4, 6_14 and 6_6 for the three affected South Hamsphire sheep, 3_11 and 3_3, 5_13 and 5_5 for the two South Hampshire carriers and 7_15 and 7_7 and 8_16 and 8_8 for the normal and affected Merinos, respectively. See Table 7.1 for corresponding sheep ID). The x-axis represents the number of reads for each sheep track. The pink box represents an area of interest within the region of LR-PCR product '2ir' in all eight sheep where no sequence reads appear to be aligned to the ovine reference sequence.

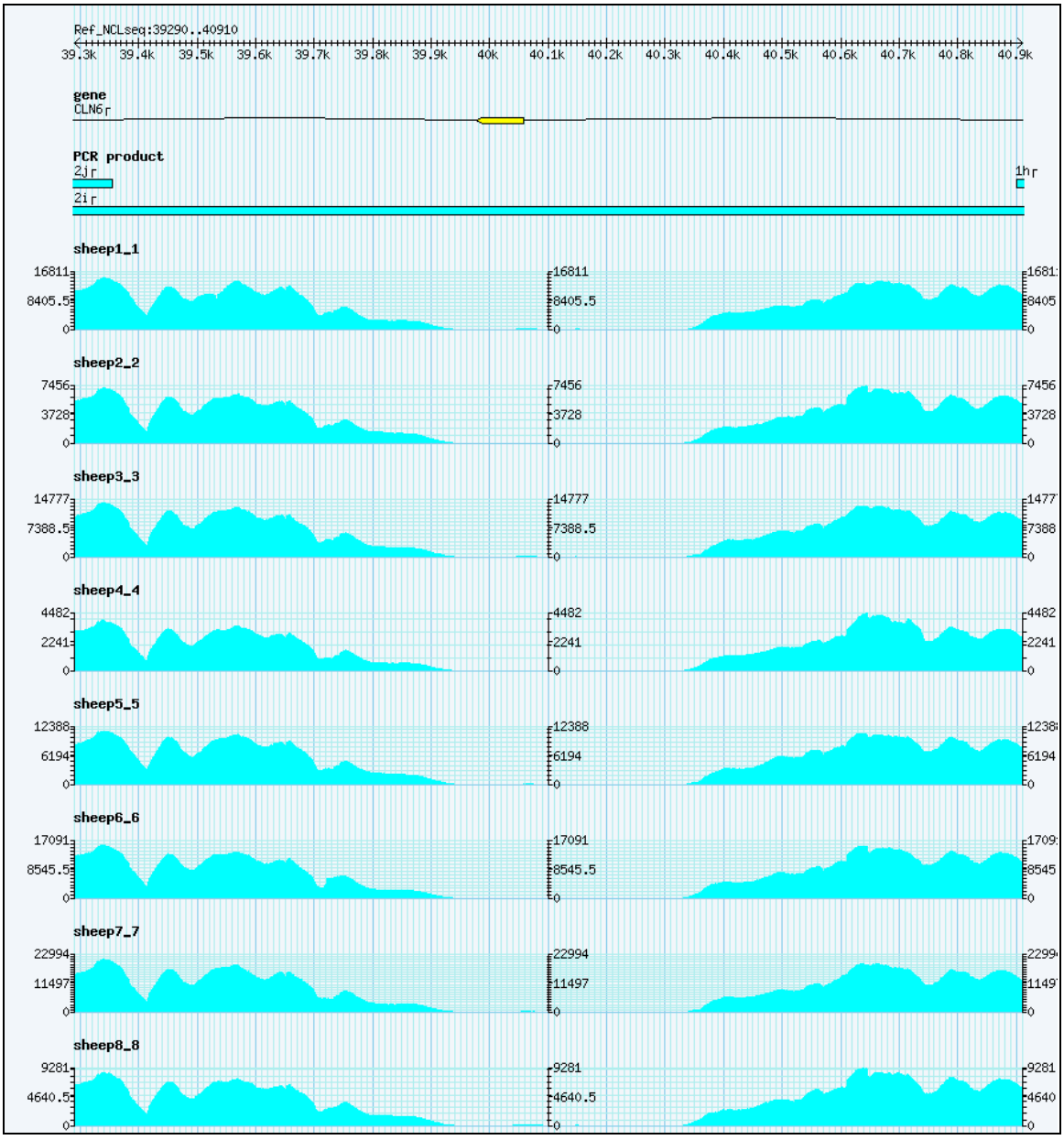


Figure 7.7: Gbrowse genome browser screenshot focused on the '2ir' LR-PCR region for all sheep (region highlighted in pink in (a)). In all affected South Hampshire sheep (sheep2_2, sheep 4_4 and sheep 6_6) no sequence reads align to the reference sequence for approximately 400 bp. Normal and carrier South Hampshire sheep (sheep1_1, sheep 3_3 and sheep 5_5) and both Merino sheep (sheep 7_7, sheep 8_8) have very low coverage reads in the middle of this region.

For the ten LR-PCR regions '1ar', '1hr', '2ir', '2jr', '3dr', '3ir', '4ar', '4ir', '5ir' and 'C1cr', sequence reads aligned to the PCR product track in the overview panel, whereas for regions '5fr', 'C2ar' and 'C3cr' reads aligned unexpectedly. For PCR region '5fr' reads appeared in both tracks of the same sheep in all South Hampshire and the Coopworth sheep (sheep 1 - 6) but aligned as expected in the Merino sheep (sheep 7 and 8). For PCR region 'C2ar' reads appeared to align as expected except for one affected SH sheep (sheep 2), which had reads aligning to both

tracks. For PCR region ‘C3cr’ reads appeared to align as expected except for the Coopworth control (sheep 1) and two affected South Hampshire sheep (sheep 2 and 4), which had reads aligning to both tracks.

A further deviation of the expected alignment of reads occurred in all animals in several regions where reads align at relatively low coverage (< 500 reads) in non LR-PCR regions for that track. This was likely due to the many short repetitive elements or motifs in the sequenced area (see Repeatmasker track).

There was uniformity in coverage across most animals and most PCR regions, suggesting that the experimental design of sequencing two pools of 7 LR-PCR products in equimolar ratios was overall very successful. However there was a relatively low coverage for sheep 4 in PCR region ‘2ir’, sheep 2 in PCR region ‘3dr’, sheep 6 in PCR region ‘4ar’, sheep 7 in PCR region ‘3ir’ and sheep 1 - 4 in PCR region ‘5fr’. Sequence coverage for PCR product ‘C1cr’ was consistently low across all animals (<1600 reads) even though the LR-PCR products performed well in the bioanalyzer quality control step (Table 7.5).

This analysis of sequence coverage using GBrowse identified an area of interest within the region of LR-PCR product ‘2ir’ (pink box, Figure 7.6) in all eight sheep, where no sequence reads aligned to the ovine reference sequence. This area spans approximately 500 bp and corresponds to position 39,900 - 40,399 bp in the ovine reference sequence. Upon zooming in at position 39,920 - 40,335 bp, the gap was observed to exist only in the three affected South Hampshire sheep (sheep 2, 4 and 6) whereas the remaining 5 sheep appear to have some reads partially covering this region (Figure 7.7).

7.3.5 Alignment of the OARv3.1 to the ovine reference sequence

When the ovine genome OARv3.1 became available in 2013, genome sequence from this version was uploaded as an additional track on the GBrowse overview and details panels (‘Oar3’ alignment; Figure 7.6). In general, there was a good similarity between OARv3.1 and the ovine reference sequence, especially considering that one is derived from a single Texel sheep and that the ovine reference sequence is a combination of an ovine BAC and Merino, Coopworth and South Hampshire sheep genomic sequences. However some differences were noted.

There were large gaps in alignment between the Oar3 sequence and LR-PCR regions ‘C2ar’ and ‘5fr’. ‘Oar3’ sequence aligned to the LR-PCR region and ‘4ir’ contained the longest contiguous repeat sequence (see ‘repeatmasker’ track). This sequence was analysed using the Repeatmasker program (Smit et al., 1996) and the repeat was found to match with repeat class/ family long interspersed nuclear element LINE/RTE - Bovb (137) at position 3710 - 2708.

There was a significant sequence variation in LR-PCR region ‘2ir’ not determined in the ‘Oar3’ at position 39,980 - 40,330 in the ovine reference sequence (see Oar3 alignment track). Interestingly, there were eight flanking ‘GGGGCGGA’ octonucleotide sequence repeats immediately upstream to this region. Exon 1 lies within this unsequenced gap region of genomic DNA and thus its exact position could not be ascertained.

There was also a significant sequence variation in LR-PCR region ‘2jr’ due to the presence of an insertion in the ovine reference sequence at position 35,820 - 36,079.

Alignment between the OARv3.1 sequence and the ovine reference sequence for PCR region ‘C4dr’ showed a high sequence similarity, with the only mismatches due to the presence of ambiguous or ‘N’ nucleotide sequence in the ovine reference sequence. Further analysis revealed that the origin of the ovine reference sequence in this region originated from merging of 19 overlapping ovine BAC contigs (see Chapter 5).

7.4 Discussion

A large deletion of approximately 415 bp, which appears to cover the whole of exon 1 of *CLN6*, was only observed in the three-affected South Hampshire and was considered the most likely causative mutation for NCL in the South Hampshire sheep. No other DNA variants appeared to segregate with the disease in South Hampshire sheep but the known Merino disease causing mutation as well as a SNP used as an indirect DNA test for the South Hampshire were correctly identified. Confirmation and definition of break points of the identified deletion of additional sheep with this genetic variant are described in the following Chapter 8.

Identification of a disease causing mutation in a specific genomic region can be a challenging task. The recent reduction in the cost of whole genome sequencing (WGS) and advances in bioinformatics has made WGS a time and cost effective approach, even in species with preliminary draft genome assemblies such as horses (Towers et al., 2013) and dogs (Drögemüller et al., 2014). However, when this study was conducted in August 2010 the sheep genome assembly was incomplete and costs for whole genome sequencing were prohibitive.

The combination of long-range PCR with NGS offered the possibility to perform mutation analysis in a relatively large region of interest in a time-efficient and economical way with the LR-PCR amplification amplifying products much larger (up to 12 kb for genomic DNA and 20 kb for phage/plasmid DNA) than those achieved with conventional *Taq* polymerases (up to 3 kb) (Mullis et al., 1986).

At the start of this study, 60 primers were designed for amplification of the whole *CLN6* genomic region and flanking sequences, including *CALML4*. Extended PCR optimisation of the various primer combinations resulted in 28 primers that generated 14 partially overlapping LR-PCR products of expected sizes (Table 7.2) that covered an estimated 49 kb region of interest. The remaining 32 primers generated either no product or multiple bands. It was discovered later using RepeatMasker (Smit et al., 1996 - 2010) that at least several of the remaining primers were located in repeat regions, which would explain some of the generation of multiple PCR products.

Initially there was a debate about which sequencing platform was to be used (either Illumina, 454 pyrosequencing or SOLiD sequencing-by-ligation) and the labeling approach that needed to be applied to the primers (either biotinylated or amine modification). The decision made was based on efficiencies of cost and resources (sheep DNA, overwhelming raw data and bioinformatic analysis for non-targeted regions of interest) when compared to sequencing the entire genomic DNA using other platforms. Other considerations included assessment of the relative ease of sequencing through challenging templates using the SOLiD platform, and compatibility between the amine labeling approach and chosen NGS platform. A comparison of these NGS platforms was described earlier (Chapter 2).

Once the labeling approach was chosen, primers were re-synthesised to include 5’ amine modifications. There was no literature found to show that adding an amino group at the 5’ end is

able to alter the original properties of the oligonucleotide, however a re-optimisation step in the form of increase in annealing temperatures was required for efficient PCR amplification of most of the LR-PCR products in this study. End modification positioned at the 3'- or 5'-end of the primers greatly reduced or prevented over-representation of amplicon ends (overlapping intervals of the amplicon ends which results in extremely high coverage in this area compared to other areas) in the sequencing libraries, thus improving the overall sequence coverage uniformity (Petermann, pers. comm.).

Analysis of SOLiD sequence data provided by the service provider identified SNPs relative to the ovine reference sequence. Identification of two known SNPs in the *CLN6* (Tammen et al., 2006) using the SOLiD sequence generated from the LR-PCR products are evidence that this approach worked successfully for mutation screening.

SNPs found within a gene coding sequence are often given the highest priority for further analysis because they are likely to affect the amino acid sequence of a protein and could be disease associated (Cargill et al, 1999). The majority of the identified SNPs in this study were in non-coding regions immediately adjacent to the *CLN6* and *CALML4* gene coding regions. SNPs found in these regions can affect biological functions such as splicing and gene regulation (Jaenisch and Bird, 2003; Cargill et al., 1999). However those identified in this study were of low significance, as they did not segregate with the NCL disease phenotype in sheep.

The known SNP identified in exon 7 (c.822C>T in reverse complement) was found to segregate as expected in the normal Coopworth and carrier South Hampshire sheep. However, the reference sequence contained the T nucleotide, which is the allele associated with the disease in the South Hamsphire NCL research flock. This SNP has been identified in unrelated sheep from different breeds and the ovine reference sequence which comprised of consensus between the early ovine genome draft (OARv2.0), ISGC, *CLN6* and *CALML4* published mRNA sequence, unpublished in house genomic DNA and BAC sequencing (Chapter 5).

As indels play an important role in biological processes and human disease (Ley et al., 2003; Strausberg et al., 2003; Pao et al., 2004; Cox et al., 2005), their accurate detection, annotation, and characterisation are critical for high-throughput human resequencing studies. Although

indels were detected in the coding sequence of the *CLN6* and *CALML4* genes, these did not segregate with the NCL phenotype, thus no further analysis were implemented at this stage.

Employment of the LR-PCR and SOLiD sequencing approach worked successfully due to rigourous consideration of possible issues in the design stage of the study. Prior to sequencing, the purified amplicons were pooled in equimolar ratios using molar information obtained from the Bioanalyzer analysis. It is known that accurate equimolar pooling is important for equal distribution of reads, sufficient coverage and successful variant detection (Harakalova et al., 2011). Our experimental design of sequencing two equimolar pools of 7 LR-PCR products each per animal resulted in successful sequencing for all 8 sheep with an overall good uniformity in coverage. However, some inconsistencies in read alignment (read depth and position) were observed.

Poor quality samples usually produce lower depths of sequencing (Ulahannan et al., 2013) which might have contributed to low coverage for PCR region 'C4dr' across all sheep, however, this is not likely to be the reason why sheep 4 in PCR region '2ir', sheep 2 in PCR region '3dr', sheep 6 in PCR region '4ar', sheep 7 in PCR region '3ir' and sheep 1 - 4 in PCR region '5fr' had relatively low read depths (Figure 7.6). The PCR products for those regions in these animals did not have the lowest amount of DNA when compared to other animals in the same PCR region (Table 7.5).

The low sequence read depth in PCR region 'C1cr' is not likely due to DNA quantity issues (Table 7.5). In instances where long-range PCR products are equimolarly pooled sequence coverage drastically drops in fragments smaller than the average length (Knierim et al., 2011) which in our case is 4 kb. However, 'C1cr' is 4.5 kb longer than average length and the reasons for the poor read depth in all animals for this region as well as the lower read depth for specific regions in some animals described above remain unclear. Considering that the read depth in these "lower coverage" regions was at least 1,000 reads, they are still sufficient for mutation screening.

Unexpected read alignments was observed for one or more animals in both tracks of the same sheep in these PCR regions: sheep 1 - 6 in PCR region '5fr', sheep 2 in PCR region 'C2ar' and

sheep 1, 2 and 4 for PCR region 'C3cr' most likely occurred because of human error. This could have taken place during the pooling of multiple PCR products from different reactions to achieve the minimum amount of DNA for SOLiD sequencing. Ideally, this should have been identified in the Bioanalyzer analysis but the three PCR regions are of similar sizes (between 3.4 to 3.7 kb) so this might have been undetected. Review of the Bioanalyzer data did not show any unusual peaks to indicate possible mixing of samples from these different regions. Alternatively, human error could have occurred when creating equimolar pooling of amplicons from individual wells in the 96-well plates into two pools of seven non-overlapping LR-PCR products.

The various SNPs and indels identified by the service provider, when compared to the ovine reference sequence, did not correspond to the disease phenotype in the South Hamphire sheep. Visualisation of sequence reads in GBrowser suggested that a large deletion occurred in all three affected South Hamsphire sheep in the regions of PCR product '2ir', positioned approximately at 39,920 - 40,335 bp in the ovine reference sequence and 14,836,464 - 14,838,151 bp in the OARv3.1 ovine genome sequence. Considering that the LR-PCR region '2ir' contains exon 1 of *CLN6*, the large deletion is highly likely to include the whole of exon 1 as well. Alignment of this region to the OARv3.1 revealed a large gap in sequence, thus confirmation of the sequence could not be established yet.

However, the identified variant was positioned in a region where all sheep appeared to have relatively low coverage in sequence with less than 1000 reads, located within an unsequenced gap region in OAR7v3.1 (39,980 - 40,330 bp) which in addition has been reported to be difficult to sequence in other species (Tammen et al., 2006). Sequencing challenging templates has been shown to decrease the sequencing coverage irrespective of method utilized (Sanger or NGS) (Bachmann et al., 2003; Kieleczawa, 2006; Yu et al., 2013). Bioanalyzer analysis of all the 112 LR-PCR products showed that the length of all the products was not significantly different from expected sizes and that sizes among animals of different phenotypes within a specific region did not show any significant differences. Statistical analysis of the data (Table 7.6) further supports this observation showing that there were no significant findings suggestive of the disease causing such a large insertion or deletion.

The difficulties of sequencing this region with both Sanger and NGS approaches is most likely caused by the extremely high GC content of 81% and the various motifs and repeats (see GC

content and Repeatmasker tracks in Figure 7.6) which required the addition of PCR additive DMSO (Winship, 1989) for effective amplification of the PCR product '2ir'. Approaches to sequence through these templates have been suggested; however they all seem to be quite specific to particular types of difficult templates and not broadly applicable for all templates. Such examples include using a novel method termed 'Slow down PCR' (Bachmann et al., 2003) and a combination of both DMSO and betaine additives in the PCR for sequencing through GC rich regions (Jensen et al., 2010).

Further work arising from these studies could include the verification of the genetic variants that do not segregate with the disease phenotypes. The SNPs and indels identified in this study were not confirmed by independent methods, and could represent sequencing errors or could be due to errors in the reference sequence. After comparison with information about known genetic variants in the ovine *CLN6* and *CALM4* genes (e.g. ENSEMBL variation tables for these genes), any new polymorphisms might be of interest for future research in relation to protein function and/or genetic marker development.

CHAPTER 8: VERIFICATION OF A LARGE DELETION ASSOCIATED WITH THE SOUTH HAMPSHIRE SHEEP NCL DISEASE PHENOTYPE

8.1 Introduction

LR-PCR amplification and next generation sequencing with the ABI SOLiD platform identified a novel deletion of approximately 415 bp, spanning exon 1 of the ovine *CLN6* gene and flanking sequences (Chapter 7). This deletion segregated with the NCL disease phenotype in the South Hampshire sheep. This chapter describes further characterisation of the deletion to determine functional effects. Efforts to develop a diagnostic test for NCL in South Hampshire sheep are also discussed.

8.2 Materials and methods

8.2.1 Sheep genomic DNA

Genomic DNA was collected from eight affected, five carrier and two normal unaffected sheep from the NCL South Hampshire research flock, as well as from two Merino sheep from the University of Sydney (Table 8.1). These genomic DNA samples were used as templates for PCR amplification and sequencing and were prepared using methods described earlier (Chapters 2.5, 2.6 and 2.7). Genotyping of these sheep with direct and indirect DNA tests for NCL is described earlier in this work (Chapters 2.1.1 and 2.1.2). Six of the samples (SH1022/07, SH1032/08, SH1033/08, SH1038/08, L06 and L07) were also used in initial identification of the deletion by LR-PCR amplification and SOLiD sequencing as described in Chapter 7.

Table 8.1: Animals used for PCR amplification and sequencing.

No	ID	Breed	Relationship between animals	Genotype	Phenotype	Total amount of DNA
1	CPW384	Coopworth		*GG	normal	30 mg
2	CPW392	Coopworth		*GG	normal	31 mg
3	SH09/00	South Hampshire		*AA	affected	38 mg
4	SH1007/07	South Hampshire		*AG	carrier	41 mg
5	SH1008/07	South Hampshire		*AG	carrier	42 mg
6	SH1009/07	South Hampshire		*AG	carrier	43 mg
7	SH1021/07	South Hampshire		*AA	affected	32 mg
8	SH1022/07	South Hampshire		*AA	affected	35 mg
9	SH1023/07	South Hampshire		*AA	affected	34 mg
10	SH1024/07	South Hampshire		*AA	affected	38 mg
11	SH1025/07	South Hampshire		*AA	affected	38 mg
12	SH1032/08	South Hampshire	full sibs	*AG	carrier	39 mg
13	SH1033/08	South Hampshire		*AA	affected	38 mg
14	SH1037/08	South Hampshire		*AA	affected	38 mg
15	SH1038/08	South Hampshire		*AG	carrier	40 mg
16	L06	Merino	full sibs	**CC	normal	20 mg
17	L07	Merino		**TT	affected	20 mg

* Genotype identified using the indirect DNA test for exon 7 polymorphism in South Hampshire sheep (Tammen et al., 2006)
** Genotype identified using the direct DNA test for exon 2 missense mutation in Merino sheep (Tammen et al., 2006)

8.2.2 Primers and PCR protocol

Five primers were designed manually from the ovine reference sequence generated in Chapter 5, at position 39,732 to 40,535 bp (Figure 8.1), to amplify the region containing the disease associated deletion between positions 39,920 and 40,335 bp (Figure 7.6). Primers were screened in Netprimer (http://www.premierbiosoft.com/netprimer/index.html) and paired for PCR using a variety of combinations and thermocycler conditions. Primers were synthesised commercially by Sigma. Primer sequences and positions on the ovine reference sequence and ovine chromosome 7 genome version 3.1 (referred to as OARV3.1 from here on) are shown in Table 8.2.

Table 8.2: Primers for amplifying the disease associated deletion

Primer	**Direction**	**Sequence 5'>3'**	**Position** on OARV3.1
F1_E1d	Forward	ATCGCCGTGGCTGAGA	14,835,554 - 14,835,569
R2_E1d	Reverse	ATACAGGTTTCGGGGAGCC	14,836,579 - 14,836,561
F3_E1d	Forward	CGAGTGGGCGAGGAAAC	14,835,619 - 14,835,635
R4_E1d	Reverse	GAGCGCAGCAGATCCCA	14,836,529 - 14,836,513
F5_E1d	Forward	CGCTGGGAAGACCCAC	14,835,650 - 14,835,665

\Qiagen Hot Start DNA Polymerase and supplied reagents were used for amplification of the targeted regions. Polymerase chain reaction with a total reaction volume of 20 µl was performed using 1X Qiagen Buffer, 1.5 mM $MgCl_2$, 0.2 µM dNTP, 20 pmol of each primer (using combinations of primers in Table 8.2), 0.15 U of Qiagen Hot Start DNA Polymerase and 50 ng of genomic DNA.

The following thermocycler conditions were used to optimise the primer combinations: 95° for 15 min for polymerase activation and DNA denaturation, followed by 33 cycles of denaturation at 95° for 30 sec, annealing temperatures (T_a) at 55, 60, or 62°C for 30 sec and extension at 72°C for 60 min, and an additional extension at 72° for 1 minute followed by cooling down at 4°C for 15 min. Following PCR, amplification products were checked using agarose gel electrophoresis.

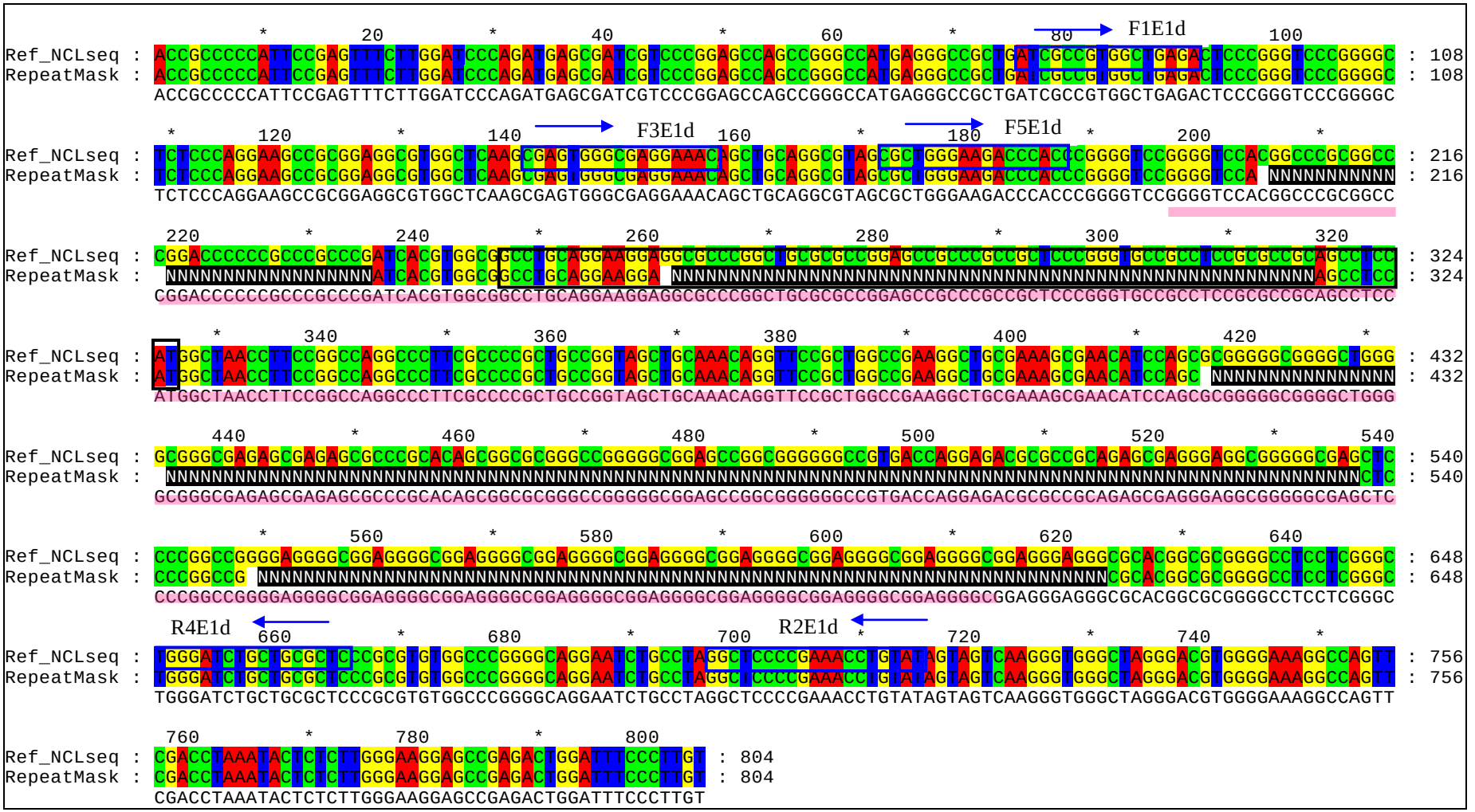


Figure 8.1: Positions of primers for the validation of a large disease associated deletion. *CLN6* exon 1 and 5' flanking sequence from the ovine reference sequence at position 39,732 to 40,535 (Chapter 7) are illustrated with GeneDoc (Nicholas et al, 1997). The ovine reference sequence is shown with (RepeatMask) and without (Ref_NCLseq) repeat masking. The position of exon 1 is indicated by a black box (sequence is reverse complement in accordance with the genome assembly). The positions of primers used for amplification of the disease associated deletion in affected SH sheep is indicated by blue boxes and arrows and the position of the deletion in the ovine reference sequence is marked by pink shading of the consensus sequence.

8.2.3 Sanger sequencing of PCR products

PCR products generated from seven animals (SH1021/07, SH1025/07, SH1007/07, SH1008/07, SH1009/07, L06 and CPW384) using primer combination F3_E1d (forward primer) and R4_E1d (reverse primer) were used as sequencing templates at an annealing temperature of 60°C. PCR products were purified using ExoSAP-IT, then 10 pmol of either the forward or reverse primer was added to approximately 20 ng of the PCR product in a final volume of 12 µl. Due to the low yield of the PCR products from carrier sheep (SH1007/07, SH1008/07, and SH1009/07), these products were only sequenced with the forward primer. Eleven sequencing samples were shipped to the AGRF sequencing service provider (Adelaide, Australia) and sequenced using the Applied Biosystems 3730 capillary sequencers (AB3730) (Figure 8.2).

8.2.4 Bioinformatics analysis

8.2.4.1 Sequence output and multialignment

The first few bases of sequences were trimmed to remove bases with a Phred quality score under 20 (< Q20) and sequence signal intensity less than 700 (Table 8.3). These trimmed sequences were visualised using BioEdit sequence alignment editor (Hall, 1999) manually aligned to one-another and to the ovine reference sequence using GeneDoc (Nicholas et al., 1997).

8.2.4.2 In silico evaluation of predicted effects of the deletion

In silico approaches were used to predict the effects of the identified deletion, as it was beyond the scope of this study to assess functional aspects of the mutant gene in the laboratory.

An additional 7 codons in the published *CLN6* sequence are predicted to code for the amino acid methionine (Tammen et al., 2006). It was considered that some of these codons represent

alternative start codons in the affected South Hampshire sheep. The known exon-intron boundaries were added manually to the diagram in order to visualise the positions of these methionine amino acids in relation to the coding exons.

Visual inspections of transcript variants of both human and mouse *CLN6* entries in the Ensembl genome database (http://asia.ensembl.org/index.html?redirect=no) were used to investigate the possibility of splice variants that do not use the start codon in exon 1. The sequence of the alternative exon 1 was used in BLAST searches against the nucleotide collection (nt), reference RNA sequence (refseq_RNA), and expressed sequence tags (est) to identify if similar sequences exist in sheep or any other species, or if they match to the 369 bp conserved non-coding sequence (CNCS) identified in an earlier study (Chapter 3).

The ovine and bovine cDNAs for *CLN6* obtained from Genbank IDs DQ458790.1 and NM_001109984.1, respectively, were also used in a BLAST search against nucleotide collection (nt), reference RNA sequence (refseq_RNA), comprising the NCBI transcript reference sequence and expressed sequence tags (est) from various species to identify splice variants in sheep and cattle not listed in Ensembl.

The rVISTA software was used to find potential regulatory elements by analysing sequences for the presence of known eukaryotic transcription factor binding sites (TFBS) (Loots et al., 2002). The 540 bp of normal Coopworth sheep sequence (including the region of the proposed disease causing deletion) was used as the input sequence, as well as the human comparative sequence found by BLAST analysis of positions 32422 - 32725 bp of human Genbank ID: NG_008764.2. Transcription factor EB (TFEB) Microphthalmia-associated transcription factor (MITF) and common eukaryotic promoters CAAT, GC and TATA were chosen from the transcription factor database (TRANSFAC; http://www.gene-regulation.com/pub/databases.html) matrices available via the software to refine the analyses.

8.3 Results

The only PCR pairing that amplified a product of expected size in normal animals was that of forward primer F3_E1d and reverse primer R4_E1d. PCR products generated in animals of different phenotypes were approximately 600 bp and 150 bp in length (Figure 8.2). Several animals did not demonstrate visible products, for some animals unspecific bands were noted and not all carrier animals presented with the two expected products.

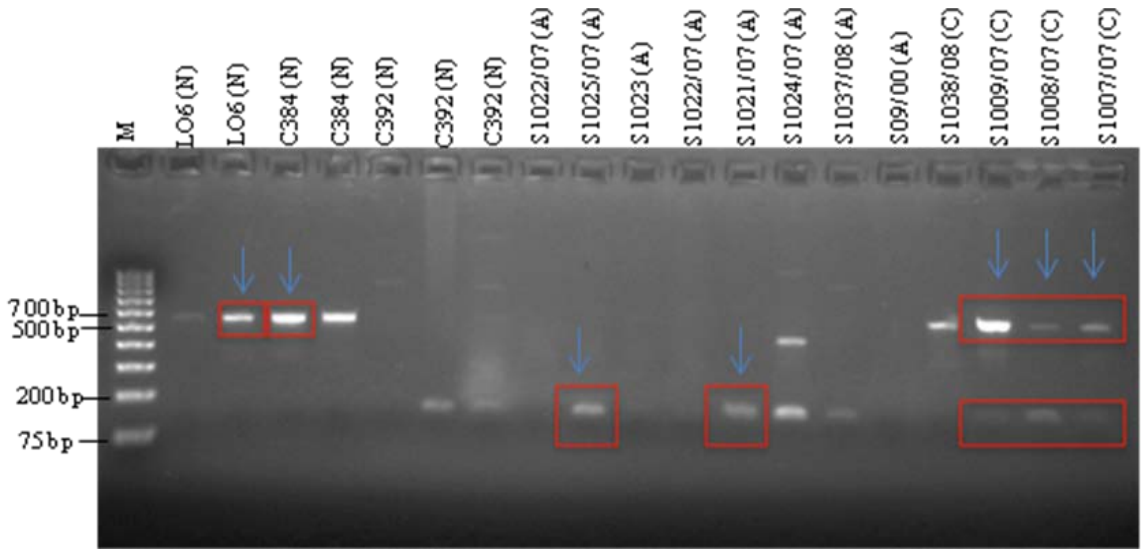


Figure 8.2: Agarose gel visualisation of PCR products generated with primers F3_E1d and R4_E1d. The identification of each animal is shown at the top of the gel image (phenotypes in brackets, N= normal; A= affected; C= carrier). Relevant fragment sizes of the Fermentas GeneRuler 1 kb Plus DNA Ladder (M) are shown on the left-hand side of the 1% agarose gel. Affected animals show a 150 bp product, normal animals a 600 bp product and carriers are expected to show both bands. Red boxes and blue arrows show PCR products from seven animals selected for sequencing.

Eleven sequencing reactions, comprised of three carrier sheep sequenced with only the forward primer, and two normal and affected sheep each sequenced with both forward and reverse primers, generated a 16 to 481 bp of high quality sequence data (Table 8.3).

Table 8.3: Quality bases and signal intensities for each sequencing sample

Animal	Disease phenotype	Sequencing primer	Bases*	Sequence signal intensity**
CPW384	Normal	F3_E1d	472	2011
CPW384	Normal	R4_E1d	474	2187
L06	Normal	F3_E1d	476	2730
L06	Normal	R4_E1d	417	1297
SH1007	Carrier	F3_E1d	481	751
SH1008	Carrier	F3_E1d	428	1336
SH1009	Carrier	F3_E1d	441	3730
SH1021	Affected	F3_E1d	86	2810
SH1021	Affected	R4_E1d	92	1864
SH1025	Affected	F3_E1d	45	2928
SH1025	Affected	R4_E1d	16	1012

* Number of quality bases detected with a Phred quality score higher than 20
****** Signals over 6000 may produce poor read

Sequences generated for affected, carrier and normal animals (Appendix 6) were aligned against the ovine reference sequence (Figure 8.3). Sections of sequence chromatograms that defined break points of the disease associated deletion were determined (Figure 8.4).

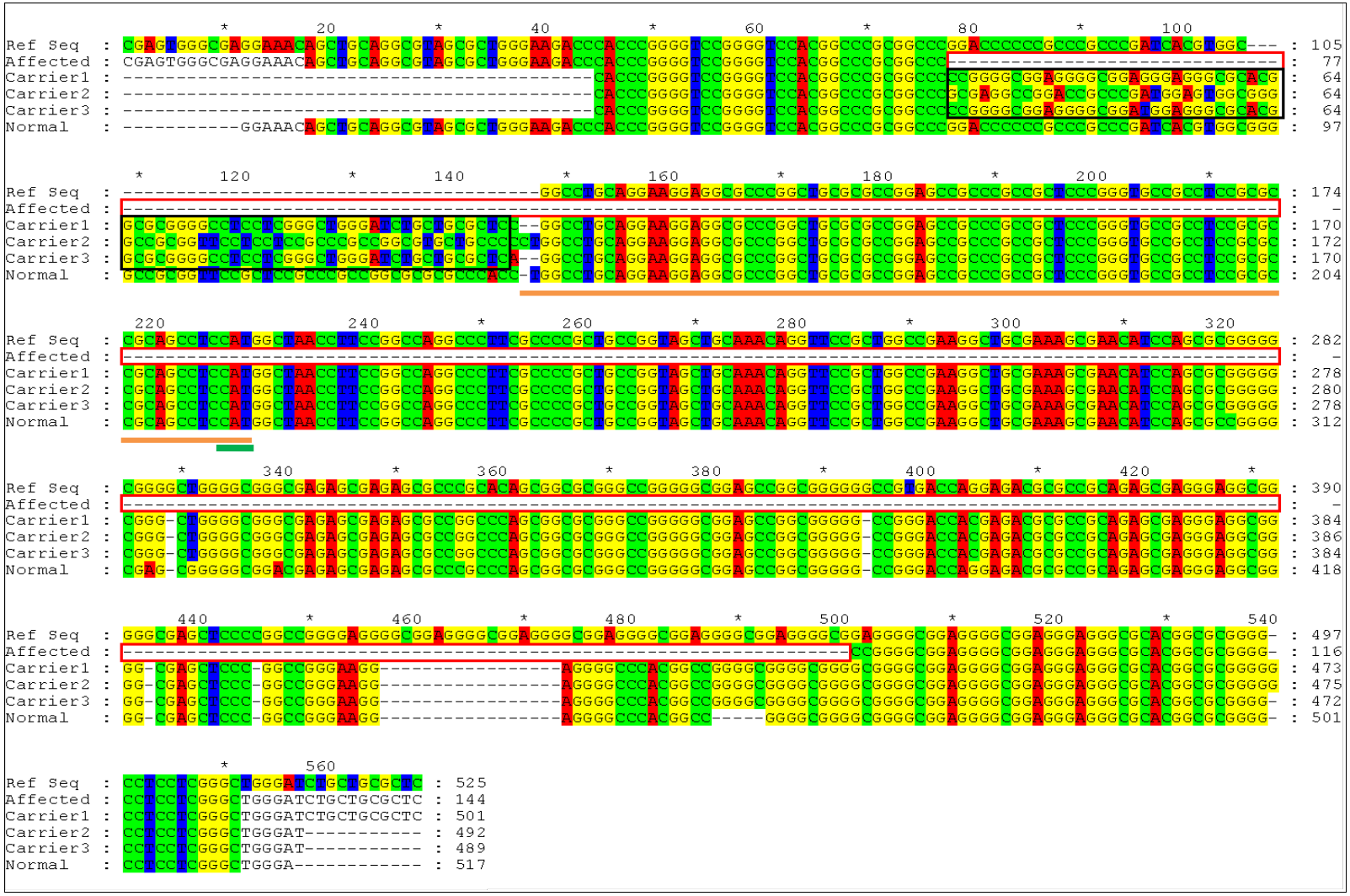


Figure 8.3: Alignment of sequence surrounding the disease associated deletion in animals of various phenotypes. The sequence is of the reverse strand. The 402 bp deletion in affected sheep is indicated by a red box. The mismatched sequence between carrier animals and aligned sequences starts where the deletion begins, and runs the length of the affected 144 bp as indicated by a black box. The published 83 bp exon 1 *CLN6* sequence is underlined in orange and the start codon underlined in green. Sequences and positions of PCR primers are shown as non-shaded sequences at the beginning and end of the alignment.

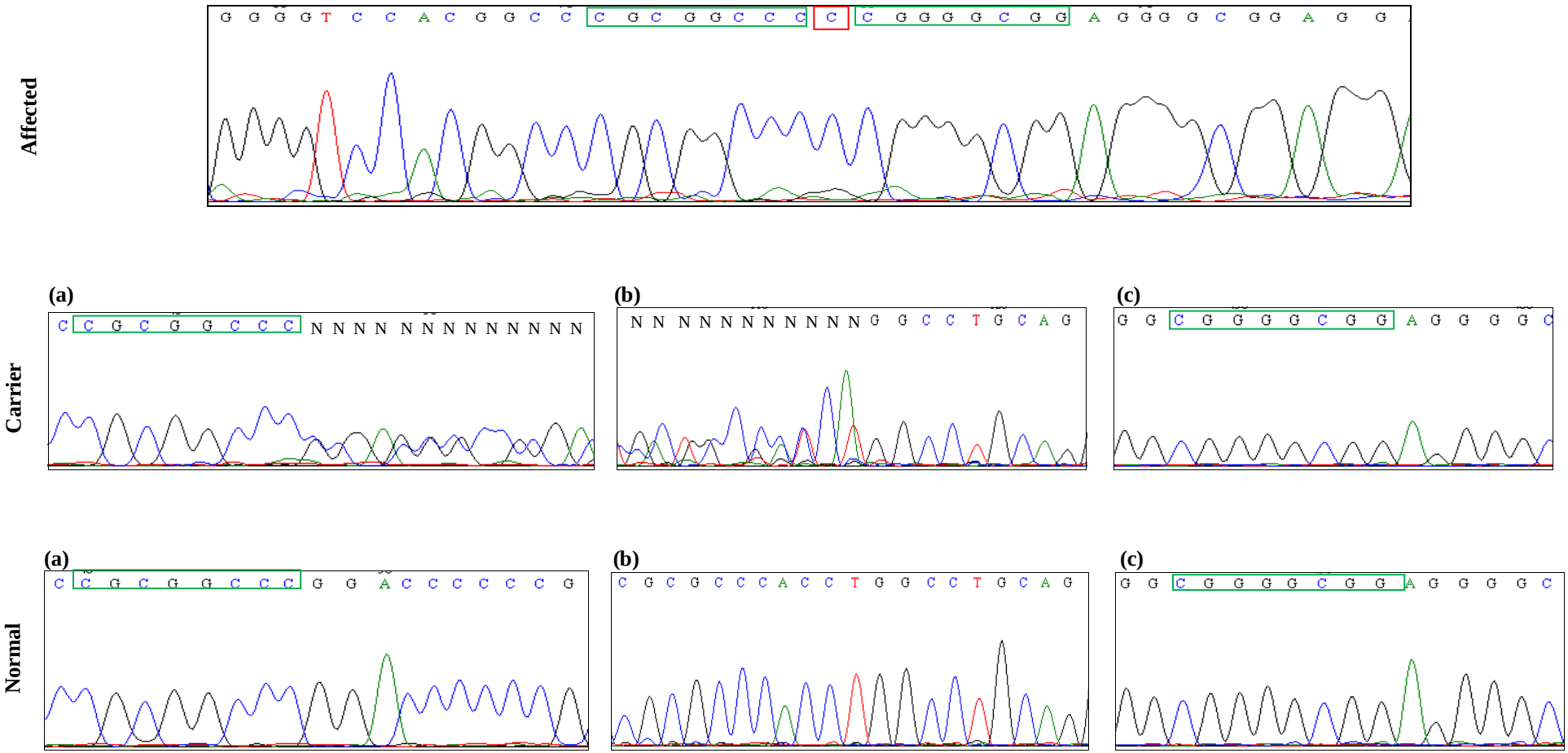


Figure 8.4: Chromatograms illustrating the boundaries of the 402 bp deletion and 1 bp insertion (g. -251_+150del and g.+150_151insC) associated with NCL in South Hampshire sheep. Eight base pairs of the sequence flanking the deletion are highlighted by green boxes in affected, carrier and normal sheep and the 1 bp insertion is highlighted by a red box in the affected sheep. The chromatogram of the affected sheep is shown as a continuous sequence in a single box. Carrier and normal sheep chromatograms are shown in three boxes (a, b and c). Box (a) indicates the sequence prior to the deletion, with double sequence (indicated by stretch of Ns representing the ambiguity in base calling) emerging at the start of the deletion in the carrier sheep. Box (b) shows in the carrier animal the re-emerging of normal sequence at the end of the 144bp PCR product that represents the deletion allele (double sequence prior to the re-emerging of normal sequence is represented by stretch of Ns). Box (c) shows normal sequence for both normal and carrier sheep after the deletion.

The disease associated deletion was identified as a 402 bp deletion and 1 bp insertion, namely g.-251_+150del and g.+150_151insC. In affected sheep this resulted in the deletion of exon 1 of *CLN6*, which includes the start codon. An exact position of the deletion in the current ovine genome assembly is difficult to provide as the deletion partially overlaps with a gap in OARV3.1 (deletion position: 39,949- 40,331 bp and gap position: 39,980- 40,330 bp) (Figure 8.5). The gap in the OARV3.1 is most likely attributable to the presence of GC-rich sequence motifs (see repeatmasker track) in this area. SOLiD sequencing of LR-PCR amplicons (Chapter 7) also identified poor overall coverage of the sequence represented by this gap in OARV3.1.

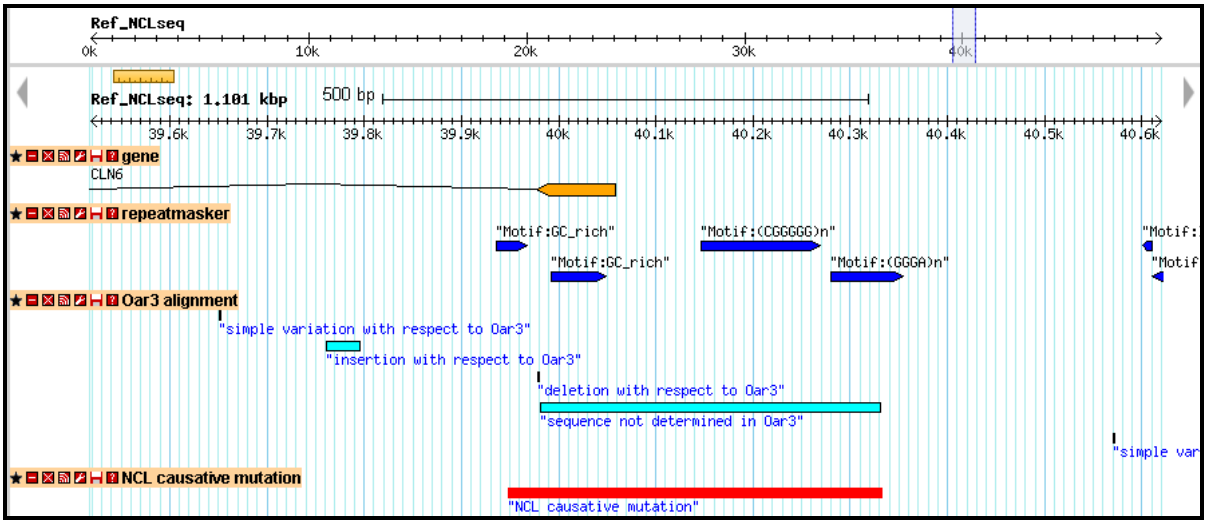


Figure 8.5: GBrowse visualisation of position of the g.-251_+150del and g.+150_151insC mutation in relation to exon 1 of ovine *CLN6*, LR-PCR product 2ir, Oar3 ovine reference sequence for chromosome 7:39,519-40,619 and repetitive sequences as identified by repeatmasker. The ovine reference sequence is used as the reference sequence and the position of the mutation (g.-251_+150del and g.+150_151insC) is indicated by the red bar, which largely overlaps with a gap in the Oar3 reference sequence (indicated by the light blue bar 'sequence not determined in Oar3). Repetitive motives are shown in dark blue (see repeatmasker track) and the positions of exon 1 and intron 1 of *CLN6* are indicated by the orange box and a line named CLN6, respectively (see gene track).

Sequencing identified differences between the ovine reference sequence and the sequenced animals, as well as variations between sequenced animals (Figure 8.3). The PCR product for the normal allele varied between animals, from 540 bp in the normal Coopworth sheep, to either 545 or 546 bp in the carrier South Hampshire sheep. The affected allele in the South Hampshire sheep was shown to be 144 bp. Sequence alignments in Genedoc, as well as the chromatograms, clearly show the double sequence in carrier animals starting where the deletion starts, up to the

144 bp length of the affected allele, after which the normal sequence re-emerges (Figures 8.3 and 8.4).

Following sequencing verification of the deletion in affected and carrier animals, further attempts to optimise the PCR to develop a direct DNA test continued both in Australia and also in New Zealand (Nadia Mitchell, Lincoln University), with the mutation validated in 7 additional animals (1 normal, 3 carrier and 3 affected). However, optimisation for carrier animals failed, and inconsistent amplifications of the two expected amplicons was a persistent problem.

Tammen and colleagues (2006) reported that quantitative PCR (qPCR) with *CLN6* specific primers amplified products from cDNA of affected South Hampshire sheep, however at significantly reduced mRNA quantity when compared to carriers or normal sheep. This spurred an investigation of the effect of the deletion mutation on formation of mRNA which identified seven potential alternative start codons in the published predicted CLN6 protein sequence (Tammen et al., 2006). Five of these were conserved between sheep, human, dog and mouse, with the first additional methionine amino acids positioned at the end of exon 2 (Figure 8.6).The remaining methionine amino acids were found in exons 3 to 7.

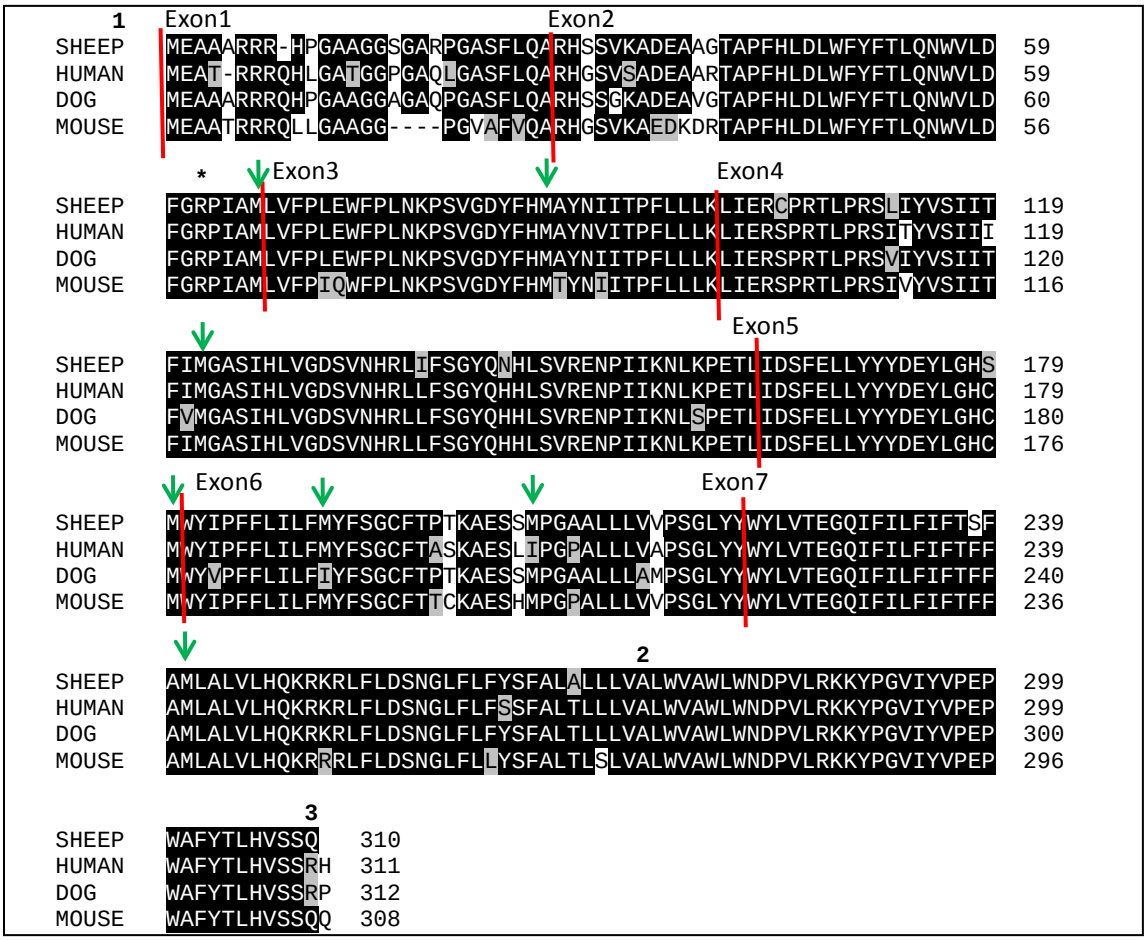
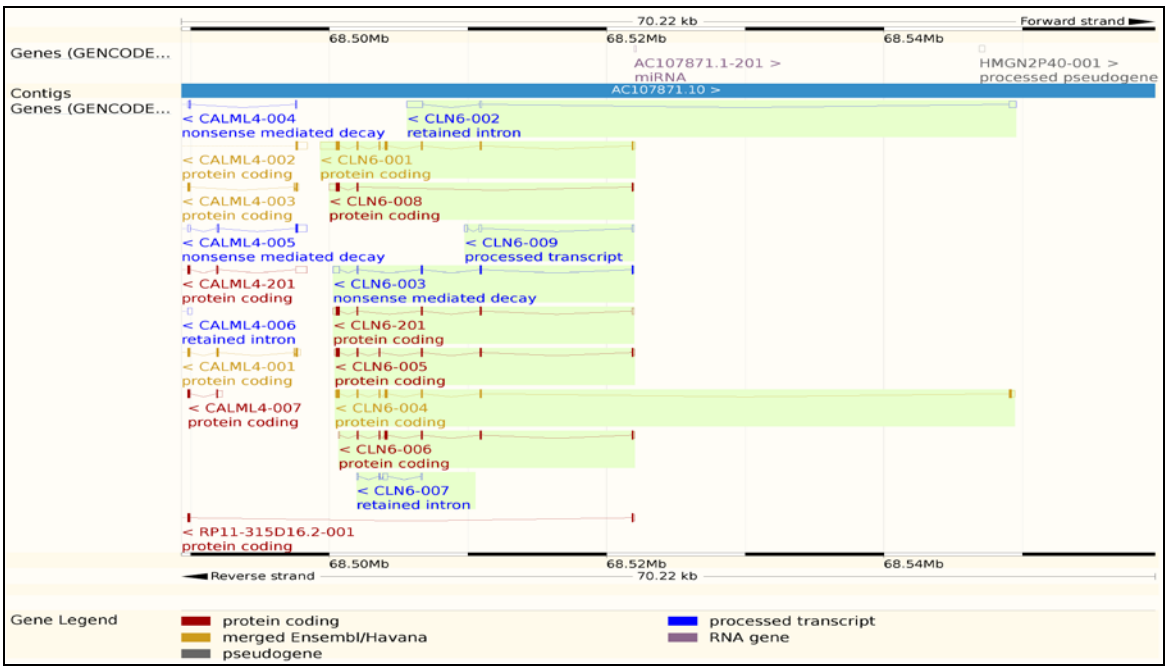


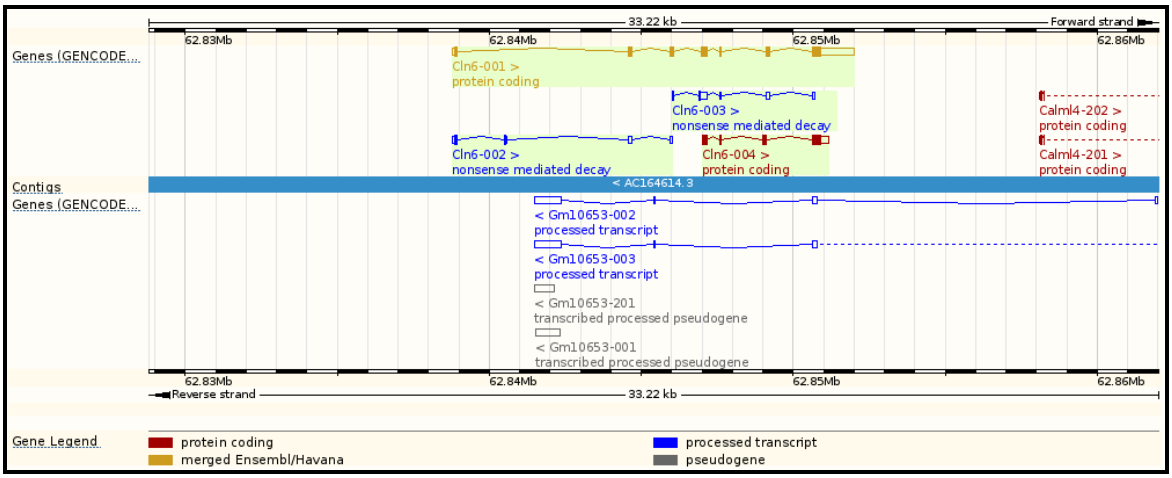
Figure 8.6: The published ovine *CLN6* predicted protein sequence (Tammen et al., 2006). Seven additional methionines present in the coding sequence are indicated with green arrows. These methionines could potentially act as alternative start codons in affected South Hampshire sheep. The known exon-intron boundaries were added manually using red lines and labelled accordingly, to visualise the positions of those methionines relative to the coding exons.

Visual inspection of the human Ensembl transcript variants (Figure 8.7a and b) shows 3 splice variants lacking exon 1 with alternative start codons, namely, CLN6-002, CLN6-004 and CLN6-007. Variants CLN6-002 and CLN6-004 suggest that an alternative exon 1 is positioned approximately 28 kb upstream of the primary *CLN6* exon 1 in humans. BLAST analysis of this sequence resulted in significant matches against predicted gorilla uncharacterized partial mRNA (NCBI reference: XM_004056391.1), and predicted rhesus macaque *CLN6* mRNA (NCBI reference: XM_001082578.2) as well as the human reference sequence (NCBI reference: NG_008764.2) and human cDNA clones (GenBank: AC107871.10, AC021553.14, AC067837.6, AK124013.1, AK293197.1). The sequence did not align with the 369 bp CNCS sequence identified in Chapter 3 and was not identified in any other species.

Human variant CLN6-007 implies the existence of a splice variant that is lacking exons 1, 2 and 7. This variant is described as having a retained intron, which is an alternative splicing, believed to be largely derived from unspliced or partially spliced pre-mRNAs (Galante et al., 2004). In mice, splice variants Cln6-003 and Cln6-004 appear to lack exons 1 and 2 or exons 1, 2 and 3, respectively. Variants Cln6-002 and Cln6-003 are considered to undergo nonsense-mediated decay, a pathway where aberrant mRNAs that encodes incomplete polypeptides due to existence of premature stop codons, are eliminated once they are detected during gene translation (Baker and Parker, 2004). This pathway functions to reduce errors in gene expression.



(a)



(b)

Figure 8.7: The Ensembl images showing *CLN6* gene transcripts for (a) humans and (b) mice in negative and positive strands, respectively. There are 10 transcripts of varying length detected in humans, 6 of these being protein coding; and 4 in mice, of which 2 are protein coding. In humans, variants CLN6-002 and CLN6-004 suggest that an alternative exon 1 is positioned upstream of the published exon 1, and variant CLN6-007 suggests a splice variant that is lacking exons 1, 2 and 7 might exist. In mice, splice variants Cln6-003 and Cln6-004 appear to lack exons 1 and 2 or exons 1, 2 and 3, respectively. (Images were taken from the Ensembl human and mouse *CLN6* entries: http://asia.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000128973;r=15:68499330-68549549 and http://asia.ensembl.org/Mus_musculus/Gene/Summary?db=core;g=ENSMUSG00000032245;r=9:62838785-62852006).

BLAST analysis of the ovine and bovine cDNAs found no evidence of splice variants in sheep and cattle. The human and sheep CNCS sequences retrieved BLAST hits of highly similar regions in human and mouse genome regions, as expected, based on the outcome of VISTA analysis in Chapter 3. Short sequences (<40 bp) with high E-values between $0.52 \times 10^{6.3}$ and $0.024 \times 10^{3.6}$ in regards to 'ref_seq', respectively were also found.

The rVISTA analysis identified 4 MITF and 2 TFEB binding sites in the 5' upstream of ovine *CLN6* intron 1, within the region deleted in affected South Hampshire sheep, which was positioned between 78 and 475 bp of the input sequence from normal Coopworth sheep (Figure 8.8). The CAAT box and GC box were detected throughout the sequence, with 12 GCs conserved between human and sheep and positioned 5' upstream of exon 1. Seven of these GCs are included in the deleted sequence. The aligned sequences did not appear to contain a TATA box.

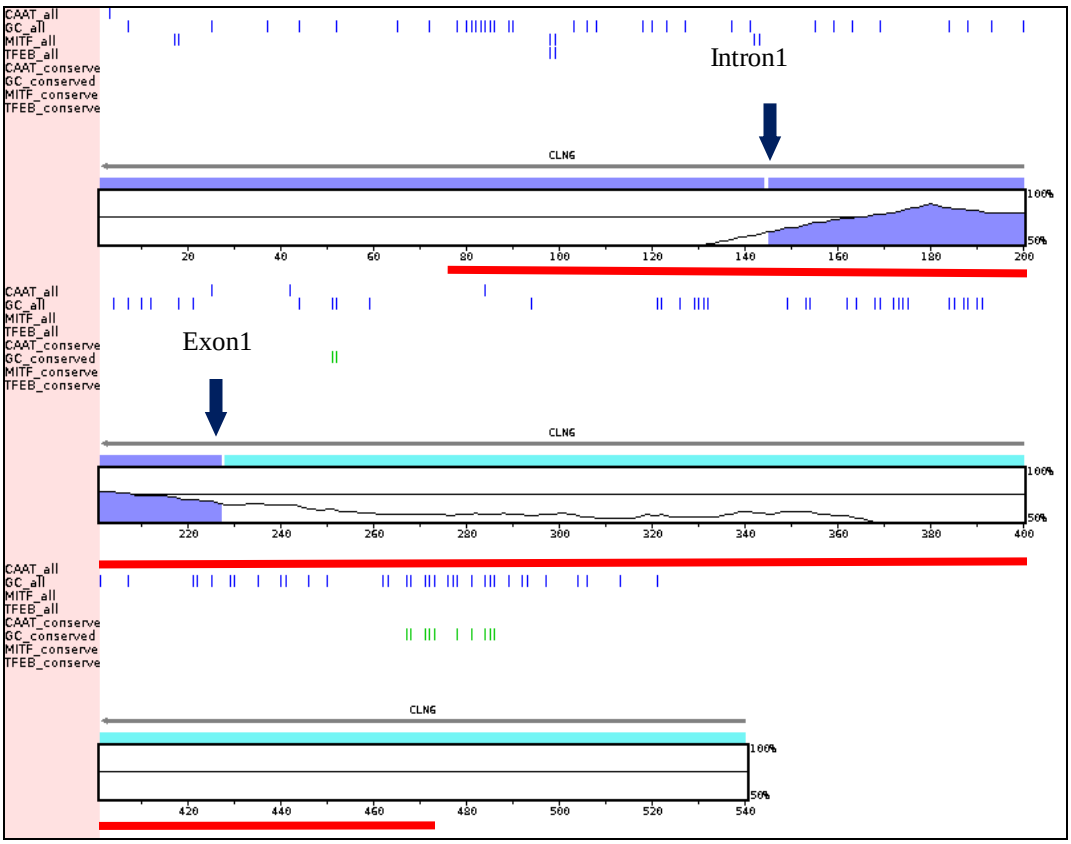


Figure 8.8: Output of the rVISTA analysis showing binding sites for CAAT, GC, MITF and TFEB identified in sheep and humans. The plot of conservation identifies the level of conservation between sheep and humans for this region. The positions of intron 1 and exon 1 are identified by blue bars, and the turquoise bar represents the 5'UTR sequence. The disease-associated deletion is positioned between 78 bp and 475 bp in reverse complement and identified by a red bar. The TRANSFAC matrices are shown on the left with '*_all' indicating accumulative sites found in both species and '*_conserved' indicating sites conserved between human and sheep.

8.4 Discussion

A novel deletion of approximately 415 bp in NCL affected South Hampshire sheep was initially identified using LR-PCR amplification and SOLiD next generation sequencing (NGS) (Chapter 7). This approach successfully identified the disease-associated deletion, but did not determine the boundaries of the deletion accurately and also could not identify if heterozygote animals for the disease allele were heterozygous for the deletion. To verify the deletion, PCR products spanning the deletion were Sanger sequenced in additional affected, carrier and normal animals. The resulting sequence identified the exact site and size of the deletion and confirmed that carriers of the disease were heterozygous for the deletion, whereas homozygous normal animals did not have the deletion.

The NCL disease associated mutation, (g.-251_+150del and g.+150_151insC) which comprises a 402 bp deletion and a 1 bp insertion, includes the 83 bp sequence of the complete first exon of ovine *CLN6*. Alignment of the affected to the normal sheep identified the position of the deletion at 39,949 - 40,330 bp of the ovine reference sequence generated in Chapter 5. This varied slightly from the position predicted from the LR-PCR amplification and sequencing approach, which was 39,920 - 40,335 bp (Chapter 7 and Figure 8.1). This minor discrepancy arose because there was difficulty in indicating the exact breakpoints of the deletion based exclusively on visual inspection of SOLiD reads of normal, carrier and affected animals using Gbrowse.

In a previous study, Tammen et al. (2006) claimed that they sequenced the full coding region of ovine *CLN6* in affected South Hampshire sheep without finding any mutations. This is in apparent disagreement with the findings presented here. However, it was also reported that sequencing of exon 1 was difficult, notably due to the high GC content of 82%, as compared to humans at 77%, and others have also reported difficulties with sequencing the 5' region of the gene (Sharp et al., 2003). High GC-rich contents are known to result in problems with traditional Sanger and NGS sequencing (Bachmann et al., 2003; Kieleczawa, 2006; Yu et al., 2013). Thus, it was not surprising that the SOLiD sequence generated for this region resulted in low number of reads in normal animals as shown in Figure 7.6.

Sequencing the ovine *CLN6* exon 1 was conducted on nested PCR products (Haff, 1994) in the experimental work described by Tammen et al. (2006) (Nadia Mitchell, pers. comm.). It is possible that despite stringent use of negative controls, contamination with either an ovine cDNA *CLN6* clone used in the lab (Tammen et al., 2006) or PCR product representing the normal allele was still present.

Tammen et al. (2006) also reported significantly reduced *CLN6* mRNA transcript levels using qPCR on cDNA from affected South Hampshire sheep and an apparent tendency to intermediate levels in heterozygote sheep (although there was no significant difference between normal and carrier South Hampshire sheep). Primers used spanned the exon 2/3 and exon 3/4 boundaries and thus did not contain the g.-251_+150del and g.+150_151insC deletion. This change in expression was not noted in Merino sheep with the CLN6 variant. It was hypothesised that the disease

causing mutation in South Hampshire sheep was a regulatory mutation. However, the finding of this study, namely that a large deletion including the entire exon 1 is likely to be disease causing, questions whether any transcript would be expected to be found in affected animals. An *in vitro* study investigated *CLN6* transcripts in normal sheep and other species, in order to predict any possible effects of the deletion on mRNA formation, in relation to the findings of Tammen et al. (2006). This low mRNA transcript occurrence was thought to be most likely due to alternative splicing (Tammen et al., 2006), where different combinations of exons are spliced together to produce different mRNA isoforms of a gene (Gilbert, 1978). Cross-species transcript information can predict whether an exon is alternatively spliced with a fairly high degree of accuracy (Philipps et al., 2004).

The disease-associated deletion, which includes exon 1, has led to the loss of the primary ATG start codon of ovine *CLN6* (Tammen et al., 2006). The 7 highly conserved methionines identified in exons 2 - 7 in the predicted gene protein sequence could potentially act as alternative start codons. In many genes, the initiation or the start codon is located in unconventional regions, such as in exon 2 (Mnatzakanian et al., 2004) and 4 (Ren, 1994) of the Methyl-CpG binding protein 2 and human A_1 adenosine receptor genes respectively. This occurrence is likely because gene translation is not necessarily initiated at the first ATG encountered by the ribosome, but rather the triplet codon must be within an appropriate consensus sequence (Latchman, 2005) before translation can begin.

The TATA box, although included in the list of matrices to be identified, did not appear in the deleted sequence. This is not surprising, as the TATA box, which is often within the gene core promoter, is found in only 24% of human genes (Yang et al., 2006). Most genes lack a TATA box and use an initiator element or downstream core promoter instead. Furthermore, the TATA box could be further upstream from the investigated sequence, which only extends for 313 bp upstream of the published exon 1.

The human *CLN6* transcript variants CLN6-002 and CLN6-004, which both contain an alternative exon 1 upstream of the gene, showed significant sequence matches against human and non-human primate *CLN6* sequences, but not high sequence similarity to any other species, nor

with the 369 bp CNCS sequence identified in Chapter 3. This, and the indication that CLN6-004 is protein coding shows that this alternative exon is likely to exist in primates, but might not exist in non-primates such as sheep and cattle. This may also be due to the reference sequence being poor in this region due to high GC content.

Alternative splice variants that start in exon 3 (CLN6-007 and Cln6-003) and 4 (Cln6-004) in humans and mice might also exist in sheep. However, they would not explain the qPCR results in the 2006 study by Tammen and colleagues, as the primers used in the qPCR study in South Hampshire sheep were positioned across the exon 2/3 and exon 3/4 boundaries of ovine *CLN6*.

In sheep, alternative splicing has been documented that excludes exon 5 (Tammen et al., 2006) but this is not likely to be affected by the mutation discovered in this study. Considering the large number of splice variants reported, a further splice variant that uses an alternative start codon in exon 2 to initiate translation into protein is possible, and could explain the very low mRNA transcript detected in the affected SH sheep as compared to normal sheep. Interestingly, there was evidence of alternative splicing affecting the human *CLN6* exon 1 for one of the haplotypes reported in the NCL Mutation Database, however, these patients were later identified to have mutations in *CLN7*, and thus no further investigations in relation to the *CLN6* gene were carried out (Sarah Mole, pers. comm.).

These bioinformatic analyses have given some evidence of the potential existence of splice variants that lack exon 1. Further work should involve RNA sequencing, which is a next generation sequencing of cDNA giving a snapshot of RNA presence and quantity from a genome at a given time (Chu and Corey, 2012). This would result in a more accurate understanding of the existence of varying transcript forms and levels in normal, carrier and disease-affected South Hampshire sheep.

Proving that a mutation is disease causing, and not just a genetic variation, can be challenging. Previous studies have identified *CLN6* as a strong positional candidate gene for NCL in South Hampshire sheep due to linkage analysis (Broom et al., 1998; Tammen et al., 2006) to a region which was homologous to the regions containing NCL causing *CLN6* genes in humans, mice,

Merino sheep and dogs (Wheeler et al., 2002; Sharp et al., 2003, Gao et al., 2002; Tammen et al., 2006; Katz et al., 2011). This is supported by the finding of extraordinarily tight linkage to an A/G allelic variance in exon 7 that has been used to predict disease status in the South Hampshire flock (Tammen et al., 2006). This SNP has been used to predict the disease status of the offspring of 22 matings of GG normal, AG carriers and AA affected animals without any errors (Nadia Mitchell, pers. comm.).

The large deletion identified in Chapter 7 and verified here, is the only variant that appears to segregate with the disease phenotype and is expected to have a large effect on the structure and quantity of the resulting protein. Unfortunately, we were not able to develop a direct DNA test that allowed confirmation that the deletion is segregating in the wider population. Furthermore, any investigation on the effect of protein function, structure, localisation and quantity is compromised by the lack of a reliable antibody that detects CLN6 protein in sheep (D. N. Palmer, pers. comm.).

The function of the CLN6 protein is poorly understood (Mole et al., 2004; Kollmann et al., 2013) which makes it difficult to predict what the exact effects of lack of CLN6 protein on cellular function might be. However, the CLN6 protein has been predicted to encode a transmembrane protein (Sonhammer et al., 1998) with the structure typically modelled as a simple multiple pass protein localised to the endoplasmic reticulum membrane (Kmoch et al., 2013). It is still unclear how mutations in the protein lead to lysosomal dysfunction (Mole et al., 2011).

The identified deletion is predicted to lead to a shortened protein product in affected South Hampshire sheep, and would likely not fulfill the same function as the normal CLN6 protein. Assuming that the deletion does not interfere with the splicing of the other exons, another potential translation initiation site is predicted in exon 2 at cDNA position 195, but function would depend on the proximity of promoter and regulatory elements. If such a protein is produced, it lacks the N-terminal region likely to code for the insertion signal sequence of a transmembrane protein and protein folding editing motifs.

There are currently 53 known mutations in the human *CLN6* gene that cause the childhood form of NCL or variant late infantile NCL -vLINCL (Sharp et al., 1997), and 15 mutations causing adult onset NCL or ANCL as well as Kufs type A and B (Arsov et al., 2011) listed in the NCL Mutation Database: http://www.ucl.ac.uk/ncl/cln6.shtm. Five of these causative mutations have been documented in exon 1 (Wheeler et al., 2002; Arsov et al., 2011; Kousi et al., 2012) in human patients with NCL. These were mostly missense mutations except for one single base pair deletion reported by Wheeler et al. (2002), which leads to a frame shift mutation. Exon 1 is thus likely to code for functionally important sections of the protein. In humans, a pentameric repeat polymorphism, although non-disease associated, has been found upstream of human *CLN6* (Sharp et al., 2003). The specific number of these pentameric repeats varies, with an example of five copies present between –179 and –203 in cDNA FLJ20561 (Genbank ID: AK000568). These repeats are not related to the repeat sequences found flanking the ovine deletion described in Chapter 7. Insertion of a GC rich sequence CTCCGCTCCGCCCCGCCTCC between -195 and -214 in the human *CLN6* gene (Kousi et al., unpublished data; http://www.ucl.ac.uk/ncl/mutation.shtml) was not identified in the comparable ovine region sequenced in Chapter 8.

In other genes, deletions of exon 1 have been described to cause disease. A 674 bp disease associated deletion spanning the start codon, exon 1 and adjacent 5' sequence, was identified in the lamin AC (*LMNA*) gene responsible for inherited myocardial fibrosis in humans (van Tintelen et al., 2007). This deletion resulted in a decrease in protein in the carriers as compared to non-carriers (control). No alternative product was detected from the mutated allele, which is likely because the truncated protein is less stable (van Tintelen et al., 2007). A decrease in *CLN6* expression has been observed in the heterozygous South Hampshire sheep, although, in contrast to *LMNA*, the *CLN6* mutated allele produced some mRNA, albeit in significantly lower amounts (Tammen et al., 2006).

It has been envisaged that the GC rich sequence 5' to *CLN6*, which was difficult to sequence, was likely to contain regulatory elements (Sharp et al., 2003; Tammen et al., 2006). The effect of the deletion on such regulatory elements was investigated using rVISTA, which identified six potential TFEB and MITF transcription factor binding sites within the deleted region. The

coordinated lysosomal expression and regulation (CLEAR) (Sardiello et al., 2009; Palmieri et al., 2011), which is part of the lysosomal gene network, has revealed possible associations with mechanisms causing lyosomal dysfunctions in lysosomal storage diseases or LSDs (Settembre et al., 2013) such as NCL. One of the mechanisms involves the role of transcription factor EB (TFEB) that binds to CLEAR target sites in promoters of lysosomal genes and regulates gene expression (Sardiello et al., 2009).

Several genes causing the various forms of NCL disease including PPT1 (*CLN1*), TPP1 (*CLN2*), *CLCN7* and *CLN3* have been reported to have direct TFEB and MITF targets with a known role in lysosomal function (Palmieri et al., 2011). Thus, deletions of these binding sites (if they are functional) are likely to affect the regulation of potential remaining splice variants of *CLN6*.

To establish support for the claim that the g.-251_+150del and g.+150_151insC mutation is disease causing, development of a DNA test to genotype a large number of additional animals from the SH research flock as well as non-related controls is in progress. Currently, there is only an indirect DNA test available for routine diagnosis of NCL in the South Hampshire sheep (Tammen et al., 2006). However a direct DNA test is more desirable because it examines the presence or absence of a known mutation.

A DNA test employing PCR amplification using primers immediately flanking or further away from the deletion site followed by gel electrophoresis was anticipated to be an efficient disease screening method. However, primer design was challenging, as the region surrounding the deletion contains a high number of repeats, G-C rich in content (74.3%) and primers were predicted to contain secondary structures (as analysed by Netprimer). Prior to primer synthesis, these primers were searched using BLAST against the OARv3.1 genome assembly and BLAST hits returned between 18 to 25 chromosome hits including the targeted chromosome 7. Although this is a high number of hits, which will likely reduce primer specificity, primers were synthesised as no alternative was available.

Due to these challenges it was not surprising that only one primer combination amplified products of the expected size. Unfortunately, despite many attempts and usage of several PCR additives (DMSO, Qiagen Q solution and Betaine at 1M) to optimise the PCR conditions, a

reliable test was not obtained. Consistent amplification of all samples was not achieved and low quality DNA samples would often fail. Furthermore, preferential amplification of one allele relative to another was observed in obligate carriers. Preferential amplification has been described to occur due to several mechanisms, including significant differences in the length of the alleles and mismatches between the primer and a specific allele template, which favours amplification of the allele best match to the template (Walsh et al., 1992). Here, a significant difference in length is the most likely explanation. Considering that the position of the primers is in non-coding regions and we have observed genetic variation in animals sequenced in the non-coding regions (Figure 8.3), mismatches between primers and primer binding sites might also have played a role. Future work could include trying different polymerases.

Other genotyping approaches were considered apart from the classical PCR amplification and gel electrophoresis method for fragment size analysis employed in this study. For example, incorporation of the TaqMan deletion assays (Applied Biosystems, USA) into Real-Time PCR reactions have been shown to genotype mutations of various sizes successfully, including a 276 bp indel (Robledo et al., 2003) and a 308, 769 bp deletion in the GJB6 gene (Fedick et al., 2012). This method incorporates allele-specific probes designed to anneal to the mutation and custom primers that anneal to the sequences flanking the mutation site in a real-time PCR reaction. Fluorescence is measured using specific software and genotypes inferred based on these values (Hui et al., 2008). As the design of the probes and primers is critical to the success of the assay, it important to use a reference sequence that is high quality and well annotated, which is challenging for this study, as the region of interest is extremely GC-rich and contains abundant repetitive elements. Identification of a single nucleotide substitution as the disease causing mutation for NCL in the Merino sheep (Tammen et al., 2006) utilised a combination of PCR amplification, restriction fragment length polymorphism identification or RFLP (Kan and Dozy, 1978; Botstein et al., 1980; Uryu et al., 1990) and agarose gel electrophoresis. However this method is not suitable for this mutation, as the initial PCR fragments would clearly indicate genotype of the animals and thus not provide differentially cleavable sites for RFLP.

Future work could include an mRNA study using the deletion as a target for the probe to more accurately measure levels in heterozygotes, and also confirm using other probes whether there are other alternative transcripts in affected animals. Another future work arising from these

studies includes developing a diagnostic test in order to screen for NCL in additional animals. This test will subsequently replace the existing indirect DNA test (Tammen et al., 2006) as a routine test for future management of the South Hampshire sheep. This involves not only testing more SH research flock animals with known disease status, but also unrelated animals of different breeds, which should not have the deletion.

In conclusion, the 402 bp deletion and 1 bp insertion in the *CLN6*, namely, g.-251_+150del and g.+150_151insC identified in South Hampshire sheep affected with neuronal ceroid lipofuscinosis is proposed to be the disease causing mutation. This mutation is deletion of the whole of exon 1 containing the ATG start codon. Given the deletion and expression of some mRNA in affected South Hampshire sheep (Tammen et al., 2006), the likelihood of alternative splicing was also explored, with possible alternative start codons in the 5' upstream or in exons 2 – 7 of ovine *CLN6*. Although there was no functional validation of the identified deletion, the significance of this finding not only will facilitate the development and implementation of genetic testing for the South Hampshire breed of sheep, but will also improve the understanding of the disease in both sheep and human forms of this variant late-infantile NCL.

CHAPTER 9: GENERAL DISCUSSION AND CONCLUSION

Neuronal ceroid lipofuscinoses (NCL/Batten disease) are a group of diseases that occur in many species and are caused by mutations in at least 17 distinct genes (http://www.ucl.ac.uk/ncl/mutation.shtml; Bond et al., 2013). Although South Hampshire (SH) sheep are one of the best characterised NCL animal models, the disease causing mutation had not previously been identified in the coding region of the positional candidate gene, ovine *CLN6* (Tammen et al., 2006). The main objective of the research described in this thesis was to discover and characterise the mutation responsible for NCL in SH sheep.

The lack of ovine genome sequence was the main hindering factor at the beginning of this research in 2006. Therefore multi-species sequence alignment for the genomic region surrounding and including *CLN6* was initially used to identify possible conserved non-coding regions, as these might represent regulatory elements that should be prioritised for mutation screening. A new ovine sequence for the region of interest was consequently developed. Advances in next generation sequencing (NGS) technologies (Zhang et al., 2011; Liu et al., 2012; Koboldt et al., 2013) and access to increasingly more accurate ovine genome assemblies over the years (Cockett, 2003, 2006; Archibald et al., 2010) resulted in changes to the research plan and allowed this study to succeed and reveal a combined 402 bp deletion and 1 bp insertion in ovine *CLN6*, namely g.-251_+150del and g.+150_151insC as the proposed disease causing mutation. The mutation is predicted to lead to the deletion of the whole of exon 1 and the ATG start codon as well as flanking non-coding sequence. The research presented here supports evidence that ovine *CLN6* is the causative gene for NCL in SH sheep.

Unlike human NCL variants caused by mutations in *CLN1*, *CLN3*, *CLN5*, and *CLN8*, there is no evidence of a major founder mutation in *CLN6* for humans (Sharp et al., 2003) or in sheep (Tammen et al., 2006). The mutation identified in this research makes it a total of 80 reported mutations (including non-disease causing variations) in *CLN6* across all species to date (http://www.ucl.ac.uk/ncl/mutation.shtml). Of these reported mutations, 68 mutations are located

in the coding regions (http://www.ucl.ac.uk/ncl/mutation.shtml); Bond et al., 2013) and 12 in non-coding regions (http://www.ucl.ac.uk/ncl/mutation.shtml; Bond et al., 2013) of the gene.

Two reported mutations in humans are located upstream of exon 1 (Sharp et al., 2003; Kousi et al., unpublished). Sharp and associates (2003) described a polymorphic pentameric repetitive element CTCCG present between –179 and –203 of *CLN6*. Although the repeat element was not associated with the disease in humans (Sharp et al., 2003), and different repetitive elements were identified in the ovine sequence in this region (Chapter 8), the high level of repetitive elements and the richness of GC content upstream of exon 1 and including exon 1 provided challenges in the *de novo* genome assemblies as well as any attempts of mutation screening for NCL at both the DNA and RNA levels (Sharp et al., 2003; Tammen et al., 2006). Insertion of the sequence CTCCGCTCCGCCCCGCCTCC between -195 and -214 in the human *CLN6* (Kousi et al., unpublished data; http://www.ucl.ac.uk/ncl/mutation.shtml) was not present in the comparable ovine region (Chapter 8).

Following formation of a reference sheep sequence (Chapters 3 to 5), two mutation screening approaches were implemented to address the main objective of this research. Both approaches essentially aimed to re-sequence specific regions of the ovine genome in multiple individuals for the identification of a genetic variation that segregates with the disease. The first approach, namely, sequence capture and targeted sequencing with the 454 Pyrosequencing platform (Chapter 6) failed. Sequence capture aimed to isolate and enrich a specific genomic region using hybridization prior to NGS (Porreca et al., 2007). With this technique, genomic DNA samples and reference sequence for capture-probe design were provided to the service provider, who generated the custom designed probes and conducted the capture and enrichment. Challenges likely relating to capture probe design and constraints of this method with regions of excessively high GC content (Porreca et al., 2007), led us to abandon further analysis of the 454 data once it was identified that this approach did not capture DNA representing the region of interest at sufficient depth (Chapter 6). The second approach involving sequencing of LR-PCR products with the SOLiD NGS platform (Chapter 7) successfully identified the proposed disease causing mutation. The LR-PCR amplification method was more challenging and laborious than sequence capture as it involved extensive optimisation of 14 long-range PCR products covering a large

region of 49,123 bp for each of the six genomic DNA samples chosen to represent three affected, two carriers and one normal sheep. This approach was time-consuming and it took six months to optimise PCRs and amplify sufficient amounts of PCR products for all regions.

If the research would commence at the present time, with a sheep genome available as reference sequence and relatively low costs for whole genome sequencing, the most efficient approach would be to sequence the complete genome of an affected and a carrier sheep with sufficient sequencing depth and to then identify possible disease causing mutations in the region of interest. However, considering the problems observed with both Sanger and NGS methods to sequence the GC-rich areas around and including exon 1 (Tammen et al., 2006; Chapters 4 to 7) such an approach would have only been successful if sufficient deep sequencing data could have been generated.

Most known *CLN6* mutations are point mutations or very small deletions or insertions (Sharp et al., 2003; Tammen et al., 2006), which makes this mutation the largest mutation identified in *CLN6* so far. Different mutations in *CLN6* in humans have been known to cause NCL variants described to have late-infantile (Wheeler et al., 2002; Kousi et al., 2012) or adult onset (Arsov et al., 2011). Although SH and Merino sheep share very similar NCL disease aetiology, they have distinct mutations and onset of diseases (Graydon and Jolly, 1984; Mayhew et al., 1985; Jolly, 1995; Cook et al., 2002).

It is interesting that a single nucleotide substitution in position c.184C>T in Merino sheep resulting in an amino acid change (p.Arg62Cys) with normal levels of mRNA transcript (Tammen et al., 2006) appears to have a more severe impact on onset and disease progression than missing all of exon 1 and having substantially less mRNA transcript in SH sheep. It would be expected that larger deletions with a predicted large effect on the resulting protein would cause more severe clinical signs than mutations that cause a replacement of a single amino acid. Such cases are demonstrated in patients with Cri-du-chat syndrome where larger deletions are likely to result in more severe intellectual disability and developmental delay (Rodriguez-Caballero et al., 2010) and Jacobsen syndrome where the degree of neurocognitive deficiency is strongly associated with the size of the deletion (Mattina and Grossfeld, 2009). However, the

lack of understanding of the function of the normal CLN6 protein (Kollmann et al., 2013; Prof. D. N. Palmer pers. comm.) and limited information on splicing variants and their function (Chapter 8; Tammen et al., 2006; Dr. S. Mole pers. comm.) make any discussion on the impact of specific mutations speculative.

Identifying the NCL causative mutation was anticipated to improve diagnostic strategies using DNA testing. The indirect DNA test routinely used in the current experimental research flock is effective because all the animals in the flock are genotypically configured (normal 'GG', heterozygotes 'AG' or affected sheep 'AA') (Tammen et al., 2006). However, recombination between the SNP and the disease causing mutation is always a possibility when using an indirect DNA test (White and Gemmell, 2009). In fact, recombination has been observed in a single sheep in the SH research flock (Tammen et al., 2006). The indirect test is predicted to be ineffective in the wider population because the 'A' allele linked to the disease mutation in the SH research flock occurrs at a high frequency in other sheep breeds including animals sequenced in the publicly available sheep EST sequence and BAC library (Tammen et al., 2006).

A direct DNA test that detects the specific disease causing mutation rather than relying on a polymorphism associated with the disease would be more accurate than an indirect test. However, developing a direct DNA test for this deletion mutation using PCR has been difficult, and as discussed in Chapter 8, has produced inconsistencies in amplification and preferential amplification of one allele relative to the other in carriers. However, the need for a commercial DNA test might be limited. South Hampshire sheep are a unique breed which is not widely used for production, and localised only in New Zealand (Prof. D. N. Palmer pers. comm.). The disease has so far only been diagnosed in one flock by Jolly and West (1976) and animals from this flock were used to establish the SH NCL research flock. Screening for the mutation in other SH flocks and other breeds would allow for the estimation of the allele frequency to see if the disease is more common than expected. Incidence of inherited diseases is often underreported, especially in livestock and in diseases with delayed onset. In New Zealand the ovine NCL cases were reported in the Borderdale breed, and these sheep were consequently identified to represent a model for the human CLN5 forms of late-infantile NCL (Frugier et al., 2008). Identification of the disease causing mutation in Borderdale sheep (c.571+1G>A) has resulted in successful surveillance for

the disease in the research flock (Frugier et al., 2008). Irrespective of an immediate need for the test for the SH NCL mutation in a commercial setting, a direct DNA test to enable a more reliable and effective screening of this mutation in the current SH experimental flock should be developed .

The information gained in this research provides support that ovine *CLN6* is the causative gene for NCL in SH sheep and thus allows for more effective strategies for developing therapeutic approaches. It is well understood that different therapeutic approaches are required for NCL disease caused by defects in soluble lysosomal proteins compared to defects in membrane proteins (Mole et al., 2011; Augustine et al., 2013; Bennett and Rakheja, 2013). Treatment strategies for NCL caused by defects in membrane proteins are limited (Kohan et al., 2011; Augustine et al., 2013) and animal models that represent diseases caused by such defects in membrane proteins are therefore of special interest (Bond et al., 2013). Previous therapies evaluated in SH sheep have included hematopoietic cell transplantation (Westlake et al., 1995), and anti-inflammatory drug therapy with minocycline (Kay and Palmer, 2013), which were not dependent on knowledge of the disease causing gene or mutation. Unfortunately, neither of these approaches resulted in reduction of clinical signs or pathology. A confirmation that NCL in SH sheep is caused by a mutation in *CLN6* would be an essential prerequisite for collaborators in New Zealand to proceed with their gene therapy approach for ovine NCL (Linterman et al., 2011; Hughes et al., 2014).

It is anticipated that a combination of different therapeutic approaches will be needed to develop a cure in humans (Kohan et al., 2011), particularly for those variants caused by defects in membrane proteins. Therapeutic approaches might include combinations of gene therapy with pharmaceutical approaches potentially to be developed, if the disease mechanism is better understood. However, any therapeutic interventions for treatment of human patients will require preclinical evaluation of safety and effectiveness in animal models before human clinical trials can commence (Bond et al., 2013).

SH sheep are considered to be an excellent model for NCL in humans due to their long lifespan, brain size and structure, as well as presentation of clinical progression which closely resembles

that in humans. Moreover, the discovery that subunit c of mitochondrial ATP synthase (SCMAS) is the main storage material in the SH sheep (Palmer et al., 1989a) resulted in the crucial finding that SCMAS is the main storage material in most human and animal NCL variants (Hall et al., 1991; Martinus et al., 1991; Jolly et al., 2002a). The identification of the proposed disease-causing mutation in this breed and confirmation that it is indeed a model for the CLN6 variant in humans has further strengthened the invaluable role of this animal model.

LIST OF REFERENCES

Abitbol, M., Thibaud, J. L., Olby, N. J., Hitte, C., Puech, J. P., Maurer, M., Pilot-Storck, F., Hedan, B., Dreano, S., Brahimi, S., Delattre, D., Andre, C., Gray, F., Delisle, F., Caillaud, C., Bernex, F., Panthier, J. J., Aubin-Houzelstein, G., Blot, S. & Tiret, L. (2010). A canine Arylsulfatase G (ARSG) mutation leading to a sulfatase deficiency is associated with neuronal ceroid lipofuscinosis. *Proc Natl Acad Sci U S A* 107(33): 14775-14780.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* 215(3): 403-410.

Ameres, S. L. & Zamore, P. D. (2013). Diversifying microRNA sequence and function. *Nat Rev Mol Cell Biol* 14(8): 475-488.

Anderson, M., Elliott, E. & Zurynski, Y. (2013). Australian families living with rare disease: experiences of diagnosis, health services use and needs for psychosocial support. *Orphanet Journal of Rare Diseases* 8(1): 22.

Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J. M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A., Gelpke, M. D., Roach, J., Oh, T., Ho, I. Y., Wong, M., Detter, C., Verhoef, F., Predki, P., Tay, A., Lucas, S., Richardson, P., Smith, S. F., Clark, M. S., Edwards, Y. J., Doggett, N., Zharkikh, A., Tavtigian, S. V., Pruss, D., Barnstead, M., Evans, C., Baden, H., Powell, J., Glusman, G., Rowen, L., Hood, L., Tan, Y. H., Elgar, G., Hawkins, T., Venkatesh, B., Rokhsar, D. & Brenner, S. (2002). Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. *Science* 297(5585): 1301-1310.

Appleby, E. C., Longstaffe, J. A. & Bell, F. R. (1982). Ceroid-lipofuscinosis in two Saluki dogs. *J Comp Pathol* 92(3): 375-380.

Archibald, A. L., Cockett, N. E., Dalrymple, B. P., Faraut, T., Kijas, J. W., Maddox, J. F., McEwan, J. C., Hutton Oddy, V., Raadsma, H. W., Wade, C., Wang, J., Wang, W. & Xun, X. (2010). The sheep genome reference sequence: a work in progress. *Anim Genet* 41(5): 449-453.

Armstrong, D., Koppang, N. & Rider, J. A. (1982). *Ceroid-lipofuscinosis (Batten's disease)*. Elsevier Biomedical Press.

Arsov, T., Smith, K. R., Damiano, J., Franceschetti, S., Canafoglia, L., Bromhead, C. J., Andermann, E., Vears, D. F., Cossette, P., Rajagopalan, S., McDougall, A., Sofia, V., Farrell, M., Aguglia, U., Zini, A., Meletti, S., Morbin, M., Mullen, S., Andermann, F., Mole, S. E., Bahlo, M. & Berkovic, S. F. (2011). Kufs disease, the major adult form of neuronal ceroid lipofuscinosis, caused by mutations in CLN6. *Am J Hum Genet* 88(5): 566-573.

Athanasiadis, P., Malousi, A., Kouidou, S. & Maglaveras, N. (2013). GREMET: an integrative tool for the prediction of mutation effects on gene regulation. *Comput Methods Programs Biomed* 111(1): 214-219.

Augustine, E. F., Adams, H. R. & Mink, J. W. (2013). Clinical trials in rare disease: challenges and opportunities. *J Child Neurol* 28(9): 1142-1150.

Awano, T., Katz, M. L., O'Brien, D. P., Sohar, I., Lobel, P., Coates, J. R., Khan, S., Johnson, G. C., Giger, U. & Johnson, G. S. (2006 (a)). A frame shift mutation in canine TPP1 (the

ortholog of human CLN2) in a juvenile Dachshund with neuronal ceroid lipofuscinosis. *Mol Genet Metab* 89(3): 254-260.

Awano, T., Katz, M. L., O'Brien, D. P., Taylor, J. F., Evans, J., Khan, S., Sohar, I., Lobel, P. & Johnson, G. S. (2006 (b)). A mutation in the cathepsin D gene (CTSD) in American Bulldogs with neuronal ceroid lipofuscinosis. *Mol Genet Metab* 87(4): 341-348.

Bachmann, H. S., Siffert, W. & Frey, U. H. (2003). Successful amplification of extremely GC-rich promoter regions using a novel 'slowdown PCR' technique. *Pharmacogenetics* 13(12): 759-766.

Baker, K. E. & Parker, R. (2004). Nonsense-mediated mRNA decay: terminating erroneous gene expression. *Curr Opin Cell Biol* 16(3): 293-299.

Balzer, S., Malde, K. & Jonassen, I. (2011). Systematic exploration of error sources in pyrosequencing flowgram data. *Bioinformatics* 27(13): i304-i309.

Batten, F. E. (1903). Cerebral degeneration with symmetrical changes in the maculae in two members of a family. . *Trans Opthalmol Soc UK* 23: 386-390.

Bennett, M. J. & Rakheja, D. (2013). The neuronal ceroid-lipofuscinoses. *Developmental Disabilities Research Reviews* 17(3): 254-259.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Wheeler, D. L. (2005). GenBank. *Nucleic Acids Res* 33(Database issue): D34-38.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Wheeler, D. L. (2008). GenBank. *Nucleic Acids Res* 36(Database issue): 11.

Berman, B. P., Nibu, Y., Pfeiffer, B. D., Tomancak, P., Celniker, S. E., Levine, M., Rubin, G. M. & Eisen, M. B. (2002). Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the Drosophila genome. *Proc Natl Acad Sci U S A* 99(2): 757-762.

Bildfell, R., Matwichuk, C., Mitchell, S. & Ward, P. (1995). Neuronal ceroid-lipofuscinosis in a cat. *Vet Pathol* 32(5): 485-488.

Blanchette, M. & Tompa, M. (2002). *Discovery of regulatory elements by a computational method for phylogenetic footprinting*. Genome Res. 2002 May;12(5):739-48.

Bond, M., Holthaus, S. M., Tammen, I., Tear, G. & Russell, C. (2013). Use of model organisms for the study of neuronal ceroid lipofuscinosis. *Biochim Biophys Acta* 1832(11): 1842-1865.

Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32(3): 314-331.

Bras, J., Verloes, A., Schneider, S. A., Mole, S. E. & Guerreiro, R. J. (2012). Mutation of the parkinsonism gene ATP13A2 causes neuronal ceroid-lipofuscinosis. *Hum Mol Genet* 21(12): 2646-2650.

Bray, N., Dubchak, I. & Pachter, L. (2003). AVID: A global alignment program. *Genome Res* 13(1): 97-102.

Brean, A. (2004). [An account of a strange instance of disease--Stengel-Batten-Spielmayer-Vogt disease]. *Tidsskr Nor Laegeforen* 124(7): 970-971.

Brogna, S. & Wen, J. (2009). Nonsense-mediated mRNA decay (NMD) mechanisms. *Nat Struct Mol Biol* 16(2): 107-113.

Bronson, R. T., Lake, B. D., Cook, S., Taylor, S. & Davisson, M. T. (1993). Motor neuron degeneration of mice is a model of neuronal ceroid lipofuscinosis (Batten's disease). *Ann Neurol* 33(4): 381-385.

Bronson, R. T., Donahue, L. R., Johnson, K. R., Tanner, A., Lane, P. W. & Faust, J. R. (1998). Neuronal ceroid lipofuscinosis (nclf), a new disorder of the mouse linked to chromosome 9. *Am J Med Genet* 77(4): 289-297.

Broom, M. F., Zhou, C., Broom, J. E., Barwell, K. J., Jolly, R. D. & Hill, D. F. (1998). Ovine neuronal ceroid lipofuscinosis: a large animal model syntenic with the human neuronal ceroid lipofuscinosis variant CLN6. *J Med Genet* 35(9): 717-721.

Broom, M. F. & Zhou, C. (2001). Fine mapping of ovine ceroid lipofuscinosis confirms orthology with CLN6. *Eur J Paediatr Neurol* 5 Suppl A: 33-35.

Brudno, M., Do, C. B., Cooper, G. M., Kim, M. F., Davydov, E., Green, E. D., Sidow, A. & Batzoglou, S. (2003). LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res* 13(4): 721-731.

Calo, E. & Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? *Mol Cell* 49(5): 825-837.

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C. R., Lim, E. P., Kalyanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe, A., Warrington, J., Lipshutz, R., Daley, G. Q. & Lander, E. S. (1999). Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 22(3): 231-238.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C. A., Taylor, M. S., Engstrom, P. G., Frith, M. C., Forrest, A. R., Alkema, W. B., Tan, S. L., Plessy, C., Kodzius, R., Ravasi, T., Kasukawa, T., Fukuda, S., Kanamori-Katayama, M., Kitazume, Y., Kawaji, H., Kai, C., Nakamura, M., Konno, H., Nakano, K., Mottagui-Tabar, S., Arner, P., Chesi, A., Gustincich, S., Persichetti, F., Suzuki, H., Grimmond, S. M., Wells, C. A., Orlando, V., Wahlestedt, C., Liu, E. T., Harbers, M., Kawai, J., Bajic, V. B., Hume, D. A. & Hayashizaki, Y. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* 38(6): 626-635.

Carrilho, E. (2000). DNA sequencing by capillary array electrophoresis and microfabricated array systems. *Electrophoresis* 21(1): 55-65.

Cesta, M. F., Mozzachio, K., Little, P. B., Olby, N. J., Sills, R. C. & Brown, T. T. (2006). Neuronal ceroid lipofuscinosis in a Vietnamese pot-bellied pig (Sus scrofa). *Vet Pathol* 43(4): 556-560.

Chalkley, M. D., Armien, A. G., Gilliam, D. H., Johnson, G. S., Zeng, R., Wunschmann, A., Kovi, R. C. & Katz, M. L. (2013). Characterization of Neuronal Ceroid-Lipofuscinosis in 3 Cats. *Vet Pathol*.

Chan, E. Y. (2005). Advances in sequencing technology. *Mutat Res* 573(1-2): 13-40.

Chevreux, B., Pfisterer, T., Drescher, B., Driesel, A. J., Muller, W. E., Wetter, T. & Suhai, S. (2004). Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res* 14(6): 1147-1159.

Chinault, A. C. & Carbon, J. (1979). Overlap hybridization screening: isolation and characterization of overlapping DNA fragments surrounding the leu2 gene on yeast chromosome III. *Gene* 5(2): 111-126.

Chou, L. S., Liu, C. S., Boese, B., Zhang, X. & Mao, R. (2010). DNA sequence capture and enrichment by microarray followed by next-generation sequencing for targeted resequencing: neurofibromatosis type 1 gene as a model. *Clin Chem* 56(1): 62-72.

Chu, Y. & Corey, D. R. (2012). RNA sequencing: platform selection, experimental design, and data interpretation. *Nucleic Acid Ther* 22(4): 271-274.

Cockett, N. E. (2003). Current status of the ovine genome map. *Cytogenet Genome Res* 102(1-4): 76-78.

Cockett, N. E. (2006). The sheep genome. *Genome Dyn* 2: 79-85.

Collins, F. S., Morgan, M. & Patrinos, A. (2003). The Human Genome Project: lessons from large-scale biology. *Science* 300(5617): 286-290.

Cook, R. W., Jolly, R. D., Palmer, D. N., Tammen, I., Broom, M. F. & McKinnon, R. (2002). Neuronal ceroid lipofuscinosis in Merino sheep. *Aust Vet J* 80(5): 292-297.

Cooper, G. M. (2000). *The Cell: A Molecular Approach.* Sinauer Associates.

Cooper, J. D., Russell, C. & Mitchison, H. M. (2006). Progress towards understanding disease mechanisms in small vertebrate models of neuronal ceroid lipofuscinosis. *Biochim Biophys Acta* 10: 873-889.

Cotman, S. L. & Staropoli, J. F. (2012). The juvenile Batten disease protein, CLN3, and its role in regulating anterograde and retrograde post-Golgi trafficking. *Clin Lipidol* 7(1): 79-91.

Cotman, S. L., Vrbanac, V., Lebel, L. A., Lee, R. L., Johnson, K. A., Donahue, L. R., Teed, A. M., Antonellis, K., Bronson, R. T., Lerner, T. J. & MacDonald, M. E. (2002). Cln3(Deltaex7/8) knock-in mice with the common JNCL mutation exhibit progressive neurologic disease that begins before birth. *Hum Mol Genet* 11(22): 2709-2721.

Cox, C., Bignell, G., Greenman, C., Stabenau, A., Warren, W., Stephens, P., Davies, H., Watt, S., Teague, J., Edkins, S., Birney, E., Easton, D. F., Wooster, R., Futreal, P. A. & Stratton, M. R. (2005). A survey of homozygous deletions in human cancer genomes. *Proc Natl Acad Sci U S A* 102(12): 4542-4547.

Crawford, A. M., Dodds, K. G., Ede, A. J., Pierson, C. A., Montgomery, G. W., Garmonsway, H. G., Beattie, A. E., Davies, K., Maddox, J. F., Kappes, S. W. & et al. (1995). An autosomal genetic linkage map of the sheep genome. *Genetics* 140(2): 703-724.

Cummings, J. F. & de Lahunta, A. (1977). An adult case of canine neuronal ceroid-lipofuscinosis. *Acta Neuropathol* 39(1): 43-51.

Dalrymple, B. P., Kirkness, E. F., Nefedov, M., McWilliam, S., Ratnakumar, A., Barris, W., Zhao, S., Shetty, J., Maddox, J. F., O'Grady, M., Nicholas, F., Crawford, A. M., Smith, T., de Jong, P. J., McEwan, J., Oddy, V. H. & Cockett, N. E. (2007). Using comparative genomics to reorder the human genome sequence into a virtual sheep genome. *Genome Biol* 8(7): R152.

Das, A. M., Jolly, R. D. & Kohlschutter, A. (1999). Anomalies of mitochondrial ATP synthase regulation in four different types of neuronal ceroid lipofuscinosis. *Mol Genet Metab* 66(4): 349-355.

D'Ascenzo, M., Meacham, C., Kitzman, J., Middle, C., Knight, J., Winer, R., Kukricar, M., Richmond, T., Albert, T. J., Czechanski, A., Donahue, L. R., Affourtit, J., Jeddeloh, J. A. & Reinholdt, L. (2009). Mutation discovery in the mouse using genetically guided array capture and resequencing. *Mamm Genome* 20(7): 424-436.

Dawson, G., Schroeder, C. & Dawson, P. E. (2010). Palmitoyl:protein thioesterase (PPT1) inhibitors can act as pharmacological chaperones in infantile Batten disease. *Biochem Biophys Res Commun* 395(1): 66-69.

de Gortari, M. J., Freking, B. A., Cuthbertson, R. P., Kappes, S. M., Keele, J. W., Stone, R. T., Leymaster, K. A., Dodds, K. G., Crawford, A. M. & Beattie, C. W. (1998). A second-generation linkage map of the sheep genome. *Mamm Genome* 9(3): 204-209.

Dickson, L. R., Dopfmer, I., Dalefield, R. R., Graydon, R. J. & Jolly, R. D. (1989). A method of cerebro-cortical biopsy in lambs. *N Z Vet J* 37(1): 21-22.

Drögemüller, M., Jagannathan, V., Howard, J., Bruggmann, R., Drogemuller, C., Ruetten, M., Leeb, T. & Kook, P. H. (2014). A frameshift mutation in the cubilin gene (CUBN) in Beagles with Imerslund-Grasbeck syndrome (selective cobalamin malabsorption). *Anim Genet* 45(1): 148-150.

Edwards, J. F., Storts, R. W., Joyce, J. R., Shelton, J. M. & Menzies, C. S. (1994). Juvenile-onset neuronal ceroid-lipofuscinosis in Rambouillet sheep. *Vet Pathol* 31(1): 48-54.

Elgar, G. & Vavouri, T. (2008). Tuning in to the signals: noncoding sequence conservation in vertebrate genomes. *Trends Genet* 24(7): 344-352.

Eliason, S. L., Stein, C. S., Mao, Q., Tecedor, L., Ding, S. L., Gaines, D. M. & Davidson, B. L. (2007). A knock-in reporter model of Batten disease. *J Neurosci* 27(37): 9826-9834.

Evans, J., Katz, M. L., Levesque, D., Shelton, G. D., de Lahunta, A. & O'Brien, D. (2005). A variant form of neuronal ceroid lipofuscinosis in American bulldogs. *J Vet Intern Med* 19(1): 44-51.

Evans, E. E., Jones, M. P., Crews, A. J. & Newkirk, K. (2012). Neuronal ceroid lipofuscinosis in a mallard duck (Anas platyrhynchos). *J Avian Med Surg* 26(1): 22-28.

Farias, F. H., Zeng, R., Johnson, G. S., Wininger, F. A., Taylor, J. F., Schnabel, R. D., McKay, S. D., Sanders, D. N., Lohi, H., Seppala, E. H., Wade, C. M., Lindblad-Toh, K., O'Brien, D. P. & Katz, M. L. (2011). A truncating mutation in ATP13A2 is responsible for adult-onset neuronal ceroid lipofuscinosis in Tibetan terriers. *Neurobiol Dis* 42(3): 468-474.

Fedick, A., Su, J. & Treff, N. R. (2012). Development of TaqMan allelic discrimination based genotyping of large DNA deletions. *Genomics* 99(3): 127-131.

Fiske, R. A. & Storts, R. W. (1988). Neuronal ceroid-lipofuscinosis in Nubian goats. *Vet Pathol* 25(2): 171-173.

France, M., Geraghty, F. & Taylor, R. (1999).Ceroid lipofuscinosis in ferrets. In *Annual Conference Australian Society Veterinary Pathology*, p. 50.

Franks, J. N., Dewey, C. W., Walker, M. A. & Storts, R. W. (1999). Computed tomographic findings of ceroid lipofuscinosis in a dog. *J Am Anim Hosp Assoc* 35(5): 430-435.

Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. & Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 32(Web Server issue): W273-279.

Frey, K. G., Herrera-Galeano, J. E., Redden, C. L., Luu, T. V., Servetas, S. L., Mateczun, A. J., Mokashi, V. P. & Bishop-Lilly, K. A. (2014). Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. *Bmc Genomics* 15(96): 1471-2164.

Fritchie, K., Siintola, E., Armao, D., Lehesjoki, A. E., Marino, T., Powell, C., Tennison, M., Booker, J. M., Koch, S., Partanen, S., Suzuki, K., Tyynela, J. & Thorne, L. B. (2009). Novel mutation and the first prenatal screening of cathepsin D deficiency (CLN10). *Acta Neuropathol* 117(2): 201-208.

Frugier, T., Mitchell, N. L., Tammen, I., Houweling, P. J., Arthur, D. G., Kay, G. W., van Diggelen, O. P., Jolly, R. D. & Palmer, D. N. (2008). A new large animal model of CLN5 neuronal ceroid lipofuscinosis in Borderdale sheep is caused by a nucleotide substitution at a consensus splice site (c.571+1G>A) leading to excision of exon 3. *Neurobiol Dis* 29(2): 306-315.

Fuller, M., Meikle, P. J. & Hopwood, J. J. (2006).Epidemiology of lysosomal storage diseases: an overview. In *Fabry Disease: Perspectives from 5 Years of FOS* (Eds A. Mehta, M. Beck and G. Sunder-Plassmann). Oxford: Oxford PharmaGenesis.

Furusawa, Y., Mizukami, K., Yabuki, A., Kuwamura, M., Chang, H. S., Hossain, M. A., Rahman, M. M., Uddin, M. M., Mitani, S. & Yamato, O. (2012). Mutational analysis of the feline CLN3 gene and an ultrastructural evaluation of lysosomal storage materials in a cat with neuronal ceroid lipofuscinosis: an investigation into the molecular basis of the disease. *Vet J* 194(3): 425-428.

Futerman, A. H. & van Meer, G. (2004). The cell biology of lysosomal storage disorders. *Nat Rev Mol Cell Biol* 5(7): 554-565.

Gagniuc, P. & Ionescu-Tirgoviste, C. (2012). Eukaryotic genomes may exhibit up to 10 generic classes of gene promoters. *Bmc Genomics* 13: 512.

Galante, P. A., Sakabe, N. J., Kirschbaum-Slager, N. & de Souza, S. J. (2004). Detection and evaluation of intron retention events in the human transcriptome. *Rna* 10(5): 757-765.

Gao, H., Boustany, R. M., Espinola, J. A., Cotman, S. L., Srinidhi, L., Antonellis, K. A., Gillis, T., Qin, X., Liu, S., Donahue, L. R., Bronson, R. T., Faust, J. R., Stout, D., Haines, J. L., Lerner, T. J. & MacDonald, M. E. (2002). Mutations in a novel CLN6-encoded transmembrane protein cause variant neuronal ceroid lipofuscinosis in man and mouse. *Am J Hum Genet* 70(2): 324-335.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D. & Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31(13): 3784-3788.

Gavin, M., Wen, G. Y., Messing, J., Adelman, S., Logush, A., Jenkins, E. C., Brown, W. T. & Velinov, M. (2013). Substrate Reduction Therapy in Four Patients with Milder CLN1 Mutations and Juvenile-Onset Batten Disease Using Cysteamine Bitartrate. *JIMD Rep* 11: 87-92.

Gibbs, R. A., Weinstock, G. M., Metzker, M. L., Muzny, D. M., Sodergren, E. J., Scherer, S., Scott, G., Steffen, D., Worley, K. C., Burch, P. E., Okwuonu, G., Hines, S., Lewis, L., DeRamo, C., Delgado, O., Dugan-Rocha, S., Miner, G., Morgan, M., Hawes, A., Gill, R., Celera, Holt, R. A., Adams, M. D., Amanatides, P. G., Baden-Tillson, H., Barnstead, M., Chin, S., Evans, C. A., Ferriera, S., Fosler, C., Glodek, A., Gu, Z., Jennings, D., Kraft, C. L., Nguyen, T., Pfannkoch, C. M., Sitter, C., Sutton, G. G., Venter, J. C., Woodage, T., Smith, D., Lee, H. M., Gustafson, E., Cahill, P., Kana, A., Doucette-Stamm, L., Weinstock, K., Fechtel, K., Weiss, R. B., Dunn, D. M., Green, E. D., Blakesley, R. W., Bouffard, G. G., De Jong, P. J., Osoegawa, K., Zhu, B., Marra, M., Schein, J., Bosdet, I., Fjell, C., Jones, S., Krzywinski, M., Mathewson, C., Siddiqui, A., Wye, N., McPherson, J., Zhao, S., Fraser, C. M., Shetty, J., Shatsman, S., Geer, K., Chen, Y., Abramzon, S., Nierman, W. C., Havlak, P. H., Chen, R., Durbin, K. J., Egan, A., Ren, Y., Song, X. Z., Li, B., Liu, Y., Qin, X., Cawley, S., Cooney, A. J., D'Souza, L. M., Martin, K., Wu, J. Q., Gonzalez-Garay, M. L., Jackson, A. R., Kalafus, K. J., McLeod, M. P., Milosavljevic, A., Virk, D., Volkov, A., Wheeler, D. A., Zhang, Z., Bailey, J. A., Eichler, E. E., Tuzun, E., Birney, E., Mongin, E., Ureta-Vidal, A., Woodwark, C., Zdobnov, E., Bork, P., Suyama, M., Torrents, D., Alexandersson, M., Trask, B. J., Young, J. M., Huang, H., Wang, H., Xing, H., Daniels, S., Gietzen, D., Schmidt, J., Stevens, K., Vitt, U., Wingrove, J., Camara, F., Mar Alba, M., Abril, J. F., Guigo, R., Smit, A., Dubchak, I., Rubin, E. M., Couronne, O., Poliakov, A., Hubner, N., Ganten, D., Goesele, C., Hummel, O., Kreitler, T., Lee, Y. A., Monti, J., Schulz, H., Zimdahl, H., Himmelbauer, H., Lehrach, H., Jacob, H. J., Bromberg, S., Gullings-Handley, J., Jensen-Seaman, M. I., Kwitek, A. E., Lazar, J., Pasko, D., Tonellato, P. J., Twigger, S., Ponting, C. P., Duarte, J. M., Rice, S., Goodstadt,

L., Beatson, S. A., Emes, R. D., Winter, E. E., Webber, C., Brandt, P., Nyakatura, G., Adetobi, M., Chiaromonte, F., Elnitski, L., Eswara, P., Hardison, R. C., Hou, M., Kolbe, D., Makova, K., Miller, W., Nekrutenko, A., Riemer, C., Schwartz, S., Taylor, J., Yang, S., Zhang, Y., Lindpaintner, K., Andrews, T. D., Caccamo, M., Clamp, M., Clarke, L., Curwen, V., Durbin, R., Eyras, E., Searle, S. M., Cooper, G. M., Batzoglou, S., Brudno, M., Sidow, A., Stone, E. A., Payseur, B. A., Bourque, G., Lopez-Otin, C., Puente, X. S., Chakrabarti, K., Chatterji, S., Dewey, C., Pachter, L., Bray, N., Yap, V. B., Caspi, A., Tesler, G., Pevzner, P. A., Haussler, D., Roskin, K. M., Baertsch, R., Clawson, H., Furey, T. S., Hinrichs, A. S., Karolchik, D., Kent, W. J., Rosenbloom, K. R., Trumbower, H., Weirauch, M., Cooper, D. N., Stenson, P. D., Ma, B., Brent, M., Arumugam, M., Shteynberg, D., Copley, R. R., Taylor, M. S., Riethman, H., Mudunuri, U., Peterson, J., Guyer, M., Felsenfeld, A., Old, S., Mockrin, S. & Collins, F. (2004). Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428(6982): 493-521.

Gibbs, R. A., Rogers, J., Katze, M. G., Bumgarner, R., Weinstock, G. M., Mardis, E. R., Remington, K. A., Strausberg, R. L., Venter, J. C., Wilson, R. K., Batzer, M. A., Bustamante, C. D., Eichler, E. E., Hahn, M. W., Hardison, R. C., Makova, K. D., Miller, W., Milosavljevic, A., Palermo, R. E., Siepel, A., Sikela, J. M., Attaway, T., Bell, S., Bernard, K. E., Buhay, C. J., Chandrabose, M. N., Dao, M., Davis, C., Delehaunty, K. D., Ding, Y., Dinh, H. H., Dugan-Rocha, S., Fulton, L. A., Gabisi, R. A., Garner, T. T., Godfrey, J., Hawes, A. C., Hernandez, J., Hines, S., Holder, M., Hume, J., Jhangiani, S. N., Joshi, V., Khan, Z. M., Kirkness, E. F., Cree, A., Fowler, R. G., Lee, S., Lewis, L. R., Li, Z., Liu, Y. S., Moore, S. M., Muzny, D., Nazareth, L. V., Ngo, D. N., Okwuonu, G. O., Pai, G., Parker, D., Paul, H. A., Pfannkoch, C., Pohl, C. S., Rogers, Y. H., Ruiz, S. J., Sabo, A., Santibanez, J., Schneider, B. W., Smith, S. M., Sodergren, E., Svatek, A. F., Utterback, T. R., Vattathil, S., Warren, W., White, C. S., Chinwalla, A. T., Feng, Y., Halpern, A. L., Hillier, L. W., Huang, X., Minx, P., Nelson, J. O., Pepin, K. H., Qin, X., Sutton, G. G., Venter, E., Walenz, B. P., Wallis, J. W., Worley, K. C., Yang, S. P., Jones, S. M., Marra, M. A., Rocchi, M., Schein, J. E., Baertsch, R., Clarke, L., Csuros, M., Glasscock, J., Harris, R. A., Havlak, P., Jackson, A. R., Jiang, H., Liu, Y., Messina, D. N., Shen, Y., Song, H. X., Wylie, T., Zhang, L., Birney, E., Han, K., Konkel, M. K., Lee, J., Smit, A. F., Ullmer, B., Wang, H., Xing, J., Burhans, R., Cheng, Z., Karro, J. E., Ma, J., Raney, B., She, X., Cox, M. J., Demuth, J. P., Dumas, L. J., Han, S. G., Hopkins, J., Karimpour-Fard, A., Kim, Y. H., Pollack, J. R., Vinar, T., Addo-Quaye, C., Degenhardt, J., Denby, A., Hubisz, M. J., Indap, A., Kosiol, C., Lahn, B. T., Lawson, H. A., Marklein, A., Nielsen, R., Vallender, E. J., Clark, A. G., Ferguson, B., Hernandez, R. D., Hirani, K., Kehrer-Sawatzki, H., Kolb, J., Patil, S., Pu, L. L., Ren, Y., Smith, D. G., Wheeler, D. A., Schenck, I., Ball, E. V., Chen, R., Cooper, D. N., Giardine, B., Hsu, F., Kent, W. J., Lesk, A., Nelson, D. L., O'Brien W, E., Prufer, K., Stenson, P. D., Wallace, J. C., Ke, H., Liu, X. M., Wang, P., Xiang, A. P., Yang, F., Barber, G. P., Haussler, D., Karolchik, D., Kern, A. D., Kuhn, R. M., Smith, K. E. & Zwieg, A. S. (2007). Evolutionary and biomedical insights from the rhesus macaque genome. *Science* 316(5822): 222-234.

Gilbert, W. (1978). Why genes in pieces? *Nature* 271(5645): 501.

Gilles, A., Meglecz, E., Pech, N., Ferreira, S., Malausa, T. & Martin, J. F. (2011). Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. *Bmc Genomics* 12(245): 1471-2164.

Goebel, H. H. & Dahme, E. (1986). Ultrastructure of retinal pigment epithelial and neural cells in the neuronal ceroid-lipofuscinosis affected Dalmatian dog. *Retina* 6(3): 179-187.

Goebel, H. H., Bilzer, T., Dahme, E. & Malkusch, F. (1988). Morphological studies in canine (Dalmatian) neuronal ceroid-lipofuscinosis. *Am J Med Genet Suppl* 5: 127-139.

Goebel, H. H. (1992). Retina in various animal models of neuronal ceroid-lipofuscinosis. *Am J Med Genet* 42(4): 605-608.

Goebel, H. H., Schochet, S. S., Jaynes, M. & Gutmann, L. (1998). Ultrastructure of the retina in adult neuronal ceroid lipofuscinosis. *Acta Anat* 162(2-3): 127-132.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J. & Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res* 38(Web Server issue): 3.

Graydon, R. J. & Jolly, R. D. (1984). Ceroid-lipofuscinosis (Batten's disease). Sequential electrophysiologic and pathologic changes in the retina of the ovine model. *Invest Ophthalmol Vis Sci* 25(3): 294-301.

Green, P. D. & Little, P. B. (1974). Neuronal ceroid-lipofuscin storage in Siamese cats. *Can J Comp Med* 38(2): 207-212.

Grover, C. E., Salmon, A. & Wendel, J. F. (2012). Targeted sequence capture as a powerful tool for evolutionary analysis. *Am J Bot* 99(2): 312-319.

Guelly, C., Zhu, P. P., Leonardis, L., Papic, L., Zidar, J., Schabhuttl, M., Strohmaier, H., Weis, J., Strom, T. M., Baets, J., Willems, J., De Jonghe, P., Reilly, M. M., Frohlich, E., Hatz, M., Trajanoski, S., Pieber, T. R., Janecke, A. R., Blackstone, C. & Auer-Grumbach, M. (2011). Targeted high-throughput sequencing identifies mutations in atlastin-1 as a cause of hereditary sensory neuropathy type I. *Am J Hum Genet* 88(1): 99-105.

Guiretti, D., Sempere, A., Lopez-Atalaya, J. P., Ferrer-Montiel, A., Barco, A. & Valor, L. M. (2016). Specific promoter deacetylation of histone H3 is conserved across mouse models of Huntington's disease in the absence of bulk changes. *Neurobiol Dis* 3(16): 30025-30020.

Gupta, P., Soyombo, A. A., Atashband, A., Wisniewski, K. E., Shelton, J. M., Richardson, J. A., Hammer, R. E. & Hofmann, S. L. (2001). Disruption of PPT1 or PPT2 causes neuronal ceroid lipofuscinosis in knockout mice. *Proc Natl Acad Sci U S A* 98(24): 13566-13571.

Haff, L. A. (1994). Improved quantitative PCR using nested primers. *PCR Methods Appl* 3(6): 332-337.

Hafner, S., Flynn, T. E., Harmon, B. G. & Hill, J. E. (2005). Neuronal ceroid-lipofuscinosis in a Holstein steer. *J Vet Diagn Invest* 17(2): 194-197.

Haines, R. L., Codlin, S. & Mole, S. E. (2009). The fission yeast model for the lysosomal storage disorder Batten disease predicts disease severity caused by mutations in CLN3. *Dis Model Mech* 2(1-2): 84-92.

Hall, N. A., Lake, B. D., Dewji, N. N. & Patrick, A. D. (1991). Lysosomal storage of subunit c of mitochondrial ATP synthase in Batten's disease (ceroid-lipofuscinosis). *Biochem J* 275(Pt 1): 269-272.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95-98.

Haltia, M. (2003). The neuronal ceroid-lipofuscinoses. *J Neuropathol Exp Neurol* 62(1): 1-13.

Haltia, M. (2006). The neuronal ceroid-lipofuscinoses: From past to present. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1762(10): 850-856.

Haltia, M. & Goebel, H. H. (2013). The neuronal ceroid-lipofuscinoses: a historical introduction. *Biochim Biophys Acta* 11(800): 29.

Harakalova, M., Nijman, I. J., Medic, J., Mokry, M., Renkens, I., Blankensteijn, J. D., Kloosterman, W., Baas, A. F. & Cuppen, E. (2011). Genomic DNA pooling strategy for next-generation sequencing-based rare variant discovery in abdominal aortic aneurysm regions of interest-challenges and limitations. *J Cardiovasc Transl Res* 4(3): 271-280.

Hardison, R. C. (2000). Conserved noncoding sequences are reliable guides to regulatory elements. *Trends Genet* 16(9): 369-372.

Harper, P. A., Walker, K. H., Healy, P. J., Hartley, W. J., Gibson, A. J. & Smith, J. S. (1988). Neurovisceral ceroid-lipofuscinosis in blind Devon cattle. *Acta Neuropathol* 75(6): 632-636.

He, S., Liu, S. & Zhu, H. (2011). The sequence, structure and evolutionary features of HOTAIR in mammals. *BMC Evol Biol* 11(102): 1471-2148.

Herschleb, J., Ananiev, G. & Schwartz, D. C. (2007). Pulsed-field gel electrophoresis. *Nat Protoc* 2(3): 677-684.

Hillier, L. W., Miller, W., Birney, E. & Warren, W. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432(7018): 695-716.

Houweling, P., Cavanagh, C., Windsor, P., Mitchell, N., Palmer, D., Raadsma, H. & Tammen, I. (2005).A large animal model for Finnish variant NCL. In *10th International Congress on Neuronal Ceroid Lipofuscinosis*.

Houweling, P. J., Cavanagh, J. A. & Tammen, I. (2006 (a)). Radiation hybrid mapping of three candidate genes for bovine Neuronal Ceroid Lipofuscinosis: CLN3, CLN5 and CLN6. *Cytogenet Genome Res* 115(1): 5-6.

Houweling, P. J., Cavanagh, J. A., Palmer, D. N., Frugier, T., Mitchell, N. L., Windsor, P. A., Raadsma, H. W. & Tammen, I. (2006 (b)). Neuronal ceroid lipofuscinosis in Devon cattle is caused by a single base duplication (c.662dupG) in the bovine CLN5 gene. *Biochim Biophys Acta* 1762(10): 890-897.

Houweling, P. (2009).The molecular characterisation of large animal models for Neuronal Ceroid Lipofuscinosis. University of Sydney.

Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T., Durbin, R., Eyras, E., Gilbert, J., Hammond, M., Huminiecki, L., Kasprzyk, A., Lehvaslaiho, H., Lijnzaad, P., Melsopp, C., Mongin, E., Pettett, R., Pocock, M., Potter, S., Rust, A., Schmidt, E., Searle, S., Slater, G., Smith, J., Spooner, W., Stabenau, A., Stalker, J., Stupka, E., Ureta-Vidal, A., Vastrik, I. & Clamp, M. (2002). *The Ensembl genome database project.* Nucleic Acids Res. 2002 Jan 1;30(1):38-41.

Hughes, S. M., Moroni-Rawson, P., Jolly, R. D. & Jordan, T. W. (2001). Submitochondrial distribution and delayed proteolysis of subunit c of the H+-transporting ATP-synthase in ovine ceroid-lipofuscinosis. *Electrophoresis* 22(9): 1785-1794.

Hughes, S. M., Hope, K. M., Xu, J. B., Mitchell, N. L. & Palmer, D. N. (2014). Inhibition of storage pathology in prenatal CLN5-deficient sheep neural cultures by lentiviral gene therapy. *Neurobiol Dis* 62: 543-550.

Hui, L., DelMonte, T. & Ranade, K. (2008). Genotyping using the TaqMan assay. *Curr Protoc Hum Genet* 2(2).

Huse, S. M., Huber, J. A., Morrison, H. G., Sogin, M. L. & Welch, D. M. (2007). Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol* 8(7): R143.

Jaenisch, R. & Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* 33: 245-254.

Jalanko, A., Vesa, J., Manninen, T., von Schantz, C., Minye, H., Fabritius, A. L., Salonen, T., Rapola, J., Gentile, M., Kopra, O. & Peltonen, L. (2005). Mice with Ppt1Deltaex4 mutation replicate the INCL phenotype and show an inflammation-associated loss of interneurons. *Neurobiol Dis* 18(1): 226-241.

Jansen, G., Hazendonk, E., Thijssen, K. L. & Plasterk, R. H. (1997). Reverse genetics by chemical mutagenesis in Caenorhabditis elegans. *Nat Genet* 17(1): 119-121.

Järplid, B. & Haltia, M. (1993). An animal model of the infantile type of neuronal ceroid-lipofuscinosis. *J Inherit Metab Dis* 16(2): 274-277.

Jasty, V., Kowalski, R. L., Fonseca, E. H., Porter, M. C., Clemens, G. R., Bare, J. J. & Hartnagel, R. E. (1984). An unusual case of generalized ceroid-lipofuscinosis in a cynomolgus monkey. *Vet Pathol* 21(1): 46-50.

Jensen, M. A., Fukushima, M. & Davis, R. W. (2010). DMSO and betaine greatly improve amplification of GC-rich constructs in de novo synthesis. *PLoS One* 5(6): 0011024.

Jiang, Y., Xie, M., Chen, W., Talbot, R., Maddox, J. F., Faraut, T., Wu, C., Muzny, D. M., Li, Y., Zhang, W., Stanton, J. A., Brauning, R., Barris, W. C., Hourlier, T., Aken, B. L., Searle, S. M., Adelson, D. L., Bian, C., Cam, G. R., Chen, Y., Cheng, S., DeSilva, U., Dixen, K., Dong, Y., Fan, G., Franklin, I. R., Fu, S., Fuentes-Utrilla, P., Guan, R., Highland, M. A., Holder, M. E., Huang, G., Ingham, A. B., Jhangiani, S. N., Kalra, D., Kovar, C. L., Lee, S. L., Liu, W., Liu, X., Lu, C., Lv, T., Mathew, T., McWilliam, S., Menzies, M., Pan, S., Robelin, D., Servin, B., Townley, D., Wang, W., Wei, B., White, S. N., Yang, X., Ye, C., Yue, Y., Zeng, P., Zhou, Q., Hansen, J. B., Kristiansen, K., Gibbs, R. A., Flicek, P., Warkup, C. C., Jones, H. E., Oddy, V. H., Nicholas, F. W., McEwan, J. C., Kijas, J. W., Wang, J., Worley, K. C., Archibald, A. L., Cockett, N., Xu, X. & Dalrymple, B. P. (2014). The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* 344(6188): 1168-1173.

Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science* 316(5830): 1497-1502.

Jolly, R. D. & West, D. M. (1976). Blindness in South Hampshire sheep: a neuronal ceroidlipofuscinosis. *N Z Vet J* 24(6): 123.

Jolly, R. D., Janmaat, A., West, D. M. & Morrison, I. (1980). Ovine ceroid-lipofuscinosis: a model of Batten's disease. *Neuropathol Appl Neurobiol* 6(3): 195-209.

Jolly, R. D., Shimada, A., Dopfmer, I., Slack, P. M., Birtles, M. J. & Palmer, D. N. (1989). Ceroid-lipofuscinosis (Batten's disease): pathogenesis and sequential neuropathological changes in the ovine model. *Neuropathol Appl Neurobiol* 15(4): 371-383.

Jolly, R. D., Martinus, R. D., Shimada, A., Fearnley, I. M. & Palmer, D. N. (1990). Ovine ceroid-lipofuscinosis is a proteolipid proteinosis. *Can J Vet Res* 54(1): 15-21.

Jolly, R. D., Gibson, A. J., Healy, P. J., Slack, P. M. & Birtles, M. J. (1992). Bovine ceroid-lipofuscinosis: pathology of blindness. *N Z Vet J* 40(3): 107-111.

Jolly, R. D. (1993). Round-table discussion of animal models of ceroid-lipofuscinosis (Batten disease). *J Inherit Metab Dis* 16(2): 278-279.

Jolly, R. D., Palmer, D. N., Studdert, V. P., Sutton, R. H., Kelly, W. R., Koppang, N., Dahme, G., Hartley, W. J., Patterson, J. S. & Riis, R. C. (1994 (a)). Canine ceroid-lipofuscinoses: A review and classification. *Journal of Small Animal Practice* 35(6): 299-306.

Jolly, R. D., Hartley, W. J., Jones, B. R., Johnstone, A. C., Palmer, A. C. & Blakemore, W. F. (1994 (b)). Generalised ceroid-lipofuscinosis and brown bowel syndrome in Cocker spaniel dogs. *N Z Vet J* 42(6): 236-239.

Jolly, R. D. (1995). Comparative biology of the neuronal ceroid-lipofuscinoses (NCL): an overview. *Am J Med Genet* 57(2): 307-311.

Jolly, R. D. (1997). The ovine model of neuronal ceroid lipofuscinosis (NCL): its contribution to understanding the pathogenesis of Batten disease. *Neuropediatrics* 28(1): 60-62.

Jolly, R. D. & Walkley, S. U. (1999). Ovine ceroid lipofuscinosis (OCL6): postulated mechanism of neurodegeneration. *Mol Genet Metab* 66(4): 376-380.

Jolly, R. D., Bayliss, S. L., Das, A., Cooke, D. G. & Brown, S. (2001). ATP synthase activity in ovine ceroid lipofuscinosis [OCL6]. *Eur J Paediatr Neurol* 5 Suppl A: 131-134.

Jolly, R. D., Arthur, D. G., Kay, G. W. & Palmer, D. N. (2002 (a)). Neuronal ceroid-lipofuscinosis in Borderdale sheep. *N Z Vet J* 50(5): 199-202.

Jolly, R. D., Brown, S., Das, A. M. & Walkley, S. U. (2002 (b)). Mitochondrial dysfunction in the neuronal ceroid-lipofuscinoses (Batten disease). *Neurochem Int* 40(6): 565-571.

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O. & Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 110(1-4): 462-467.

Kan, Y. W. & Dozy, A. M. (1978). Antenatal diagnosis of sickle-cell anaemia by D.N.A. analysis of amniotic-fluid cells. *Lancet* 2(8096): 910-912.

Kanninen, K. M., Grubman, A., Caragounis, A., Duncan, C., Parker, S. J., Lidgerwood, G. E., Volitakis, I., Ganio, G., Crouch, P. J. & White, A. R. (2013 (a)). Altered biometal homeostasis is associated with CLN6 mRNA loss in mouse neuronal ceroid lipofuscinosis. *Biol Open* 2(6): 635-646.

Kanninen, K. M., Grubman, A., Meyerowitz, J., Duncan, C., Tan, J. L., Parker, S. J., Crouch, P. J., Paterson, B. M., Hickey, J. L., Donnelly, P. S., Volitakis, I., Tammen, I., Palmer, D. N. & White, A. R. (2013 (b)). Increased zinc and manganese in parallel with neurodegeneration, synaptic protein changes and activation of Akt/GSK3 signaling in ovine CLN6 neuronal ceroid lipofuscinosis. *PLoS One* 8(3): e58644.

Katz, M. L., Shibuya, H., Liu, P. C., Kaur, S., Gao, C. L. & Johnson, G. S. (1999). A mouse gene knockout model for juvenile ceroid-lipofuscinosis (Batten disease). *J Neurosci Res* 57(4): 551-556.

Katz, M. L., Khan, S., Awano, T., Shahid, S. A., Siakotos, A. N. & Johnson, G. S. (2005 (a)). A mutation in the CLN8 gene in English Setter dogs with neuronal ceroid-lipofuscinosis. *Biochem Biophys Res Commun* 327(2): 541-547.

Katz, M. L., Narfstrom, K., Johnson, G. S. & O'Brien, D. P. (2005 (b)). Assessment of retinal function and characterization of lysosomal storage body accumulation in the retinas and brains of Tibetan Terriers with ceroid-lipofuscinosis. *Am J Vet Res* 66(1): 67-76.

Katz, M. L., Coates, J. R., Cooper, J. J., O'Brien, D. P., Jeong, M. & Narfstrom, K. (2008). Retinal pathology in a canine model of late infantile neuronal ceroid lipofuscinosis. *Invest Ophthalmol Vis Sci* 49(6): 2686-2695.

Katz, M. L., Farias, F. H., Sanders, D. N., Zeng, R., Khan, S., Johnson, G. S. & O'Brien, D. P. (2011). A missense mutation in canine CLN6 in an Australian shepherd with neuronal ceroid lipofuscinosis. *J Biomed Biotechnol* 2011: 198042.

Kay, G. W., Hughes, S. M. & Palmer, D. N. (1999). In vitro culture of neurons from sheep with Batten disease. *Mol Genet Metab* 67(1): 83-88.

Kay, G. W., Jay, N. P. & Palmer, D. N. (2011). The specific loss of GnRH-positive neurons from the hypothalamus of sheep with CLN6 neuronal ceroid lipofuscinosis occurs without glial activation and has only minor effects on reproduction. *Neurobiol Dis* 41(3): 614-623.

Kay, G. W., Palmer, D. N., Rezaie, P. & Cooper, J. D. (2006). Activation of non-neuronal cells within the prenatal developing brain of sheep with neuronal ceroid lipofuscinosis. *Brain Pathol* 16(2): 110-116.

Kay, G. W. & Palmer, D. N. (2013). Chronic oral administration of minocycline to sheep with ovine CLN6 neuronal ceroid lipofuscinosis maintains pharmacological concentrations in the brain but does not suppress neuroinflammation or disease progression. *J Neuroinflammation* 10: 97.

Kent, W. J. (2002). BLAT--the BLAST-like alignment tool. *Genome Res* 12(4): 656-664.

Khurana, V. & Lindquist, S. (2010). Modelling neurodegeneration in Saccharomyces cerevisiae: why cook with baker's yeast? *Nat Rev Neurosci* 11(6): 436-449.

Kieleczawa, J. (2006). Fundamentals of sequencing of difficult templates--an overview. *J Biomol Tech* 17(3): 207-217.

Kleinjan, D. A., Seawright, A., Childs, A. J. & van Heyningen, V. (2004). Conserved elements in Pax6 intron 7 involved in (auto)regulation and alternative transcription. *Dev Biol* 265(2): 462-477.

Kmoch, S., Stranecky, V., Emes, R. D. & Mitchison, H. M. (2013). Bioinformatic perspectives in the neuronal ceroid lipofuscinoses. *Biochim Biophys Acta* 11(41): 26.

Knierim, E., Lucke, B., Schwarz, J. M., Schuelke, M. & Seelow, D. (2011). Systematic comparison of three methods for fragmentation of long-range PCR products for next generation sequencing. *PLoS One* 6(11): 30.

Koboldt, D. C., Steinberg, K. M., Larson, D. E., Wilson, R. K. & Mardis, E. R. (2013). The next-generation sequencing revolution and its impact on genomics. *Cell* 155(1): 27-38.

Kohan, R., Cismondi, I. A., Oller-Ramirez, A. M., Guelbert, N., Anzolini, T. V., Alonso, G., Mole, S. E., de Kremer, D. R. & de Halac, N. I. (2011). Therapeutic approaches to the challenge of neuronal ceroid lipofuscinoses. *Curr Pharm Biotechnol* 12(6): 867-883.

Kohan, R., Carabelos, M. N., Xin, W., Sims, K., Guelbert, N., Cismondi, I. A., Pons, P., Alonso, G. I., Troncoso, M., Witting, S., Pearce, D. A., Dodelson de Kremer, R., Oller-Ramirez, A. M. & Noher de Halac, I. (2013). Neuronal ceroid lipofuscinosis type CLN2: a new rationale for the construction of phenotypic subgroups based on a survey of 25 cases in South America. *Gene* 516(1): 114-121.

Kohlschutter, A. & Schulz, A. (2009). Towards understanding the neuronal ceroid lipofuscinoses. *Brain Dev* 31(7): 499-502.

Koike, M., Nakanishi, H., Saftig, P., Ezaki, J., Isahara, K., Ohsawa, Y., Schulz-Schaeffer, W., Watanabe, T., Waguri, S., Kametaka, S., Shibata, M., Yamamoto, K., Kominami, E., Peters, C., von Figura, K. & Uchiyama, Y. (2000). Cathepsin D deficiency induces lysosomal storage with ceroid lipofuscin in mouse CNS neurons. *J Neurosci* 20(18): 6898-6906.

Kollmann, K., Uusi-Rauva, K., Scifo, E., Tyynela, J., Jalanko, A. & Braulke, T. (2013). Cell biology and function of neuronal ceroid lipofuscinosis-related proteins. *Biochim Biophys Acta* 11(81): 9.

Kolovos, P., Knoch, T. A., Grosveld, F. G., Cook, P. R. & Papantonis, A. (2012). Enhancers and silencers: an integrated and simple model for their function. *Epigenetics Chromatin* 5(1): 1756-8935.

Koppang, N. (1970). Neuronal ceroid-lipofuscinosis in English setters. *J small Anim Pract* 10:639-644: 639-644.

Koppang, N. (1988). The English setter with ceroid-lipofuscinosis: a suitable model for the juvenile type of ceroid-lipofuscinosis in humans. *Am J Med Genet Suppl* 5: 117-125.

Koppang, N. (1992). English setter model and juvenile ceroid-lipofuscinosis in man. *Am J Med Genet* 42(4): 599-604.

Kopra, O., Vesa, J., von Schantz, C., Manninen, T., Minye, H., Fabritius, A. L., Rapola, J., van Diggelen, O. P., Saarela, J., Jalanko, A. & Peltonen, L. (2004). A mouse model for Finnish variant late infantile neuronal ceroid lipofuscinosis, CLN5, reveals neuropathology associated with early aging. *Hum Mol Genet* 13(23): 2893-2906.

Koressaar, T. & Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23(10): 1289-1291.

Kousi, M., Lehesjoki, A. E. & Mole, S. E. (2012). Update of the mutation spectrum and clinical correlations of over 360 mutations in eight genes that underlie the neuronal ceroid lipofuscinoses. *Hum Mutat* 33(1): 42-63.

Kovács, A. D., Saje, A., Wong, A., Ramji, S., Cooper, J. D. & Pearce, D. A. (2012). Age-dependent therapeutic effect of memantine in a mouse model of juvenile Batten disease. *Neuropharmacology*.

Kuwamura, M., Hattori, R., Yamate, J., Kotani, T. & Sasai, K. (2003). Neuronal ceroid-lipofuscinosis and hydrocephalus in a chihuahua. *J small Anim Pract* 44(5): 227-230.

Kuwamura, M., Nakagawa, M., Nabe, M., Yamate, J., Inoue, M., Satoh, H. & Yamato, O. (2009). Neuronal ceroid-lipofuscinosis in a Japanese domestic shorthair cat. *J Vet Med Sci* 71(5): 665-667.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, Y., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen,

L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S. & Chen, Y. J. (2001). Initial sequencing and analysis of the human genome. *Nature* 409(6822): 860-921.

Lane, S. C., Jolly, R. D., Schmechel, D. E., Alroy, J. & Boustany, R. M. (1996). Apoptosis as the mechanism of neurodegeneration in Batten's disease. *J Neurochem* 67(2): 677-683.

Langer-Safer, P. R., Levine, M. & Ward, D. C. (1982). Immunological method for mapping genes on Drosophila polytene chromosomes. *Proc Natl Acad Sci U S A* 79(14): 4381-4385.

Larney, C., Bailey, T. L. & Koopman, P. (2015). Conservation analysis of sequences flanking the testis-determining gene Sry in 17 mammalian species. *BMC Dev Biol* 15(34): 015-0085.

Latchman, D. S. (2005). *Gene Regulation*. Taylor & Francis Group.

Laurent, P., Schibler, L., Vaiman, A., Laubier, J., Delcros, C., Cosseddu, G., Vaiman, D., Cribiu, E. P. & Yerle, M. (2007). A 12 000-rad whole-genome radiation hybrid panel in sheep: application to the study of the ovine chromosome 18 region containing a QTL for scrapie susceptibility. *Anim Genet* 38(4): 358-363.

Lettice, L. A., Heaney, S. J., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., Goode, D., Elgar, G., Hill, R. E. & de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* 12(14): 1725-1735.

Levine, M. & Tjian, R. (2003). Transcription regulation and animal diversity. *Nature* 424(6945): 147-151.

Ley, T. J., Minx, P. J., Walter, M. J., Ries, R. E., Sun, H., McLellan, M., DiPersio, J. F., Link, D. C., Tomasson, M. H., Graubert, T. A., McLeod, H., Khoury, H., Watson, M., Shannon, W., Trinkaus, K., Heath, S., Vardiman, J. W., Caligiuri, M. A., Bloomfield, C. D., Milbrandt, J. D., Mardis, E. R. & Wilson, R. K. (2003). A pilot study of high-throughput, sequence-based mutational profiling of primary human acute myeloid leukemia cell genomes. *Proc Natl Acad Sci U S A* 100(24): 14275-14280.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16): 2078-2079.

Lindblad-Toh, K., Wade, C. M., Mikkelsen, T. S., Karlsson, E. K., Jaffe, D. B., Kamal, M., Clamp, M., Chang, J. L., Kulbokas, E. J., 3rd, Zody, M. C., Mauceli, E., Xie, X., Breen, M., Wayne, R. K., Ostrander, E. A., Ponting, C. P., Galibert, F., Smith, D. R., DeJong, P. J., Kirkness, E., Alvarez, P., Biagi, T., Brockman, W., Butler, J., Chin, C. W., Cook, A., Cuff, J., Daly, M. J., DeCaprio, D., Gnerre, S., Grabherr, M., Kellis, M., Kleber, M., Bardeleben, C., Goodstadt, L., Heger, A., Hitte, C., Kim, L., Koepfli, K. P., Parker, H. G., Pollinger, J. P., Searle, S. M., Sutter, N. B., Thomas, R., Webber, C., Baldwin, J., Abebe, A., Abouelleil, A., Aftuck, L., Ait-Zahra, M., Aldredge, T., Allen, N., An, P., Anderson, S., Antoine, C., Arachchi, H., Aslam, A., Ayotte, L., Bachantsang, P., Barry, A., Bayul, T., Benamara, M., Berlin, A., Bessette, D., Blitshteyn, B., Bloom, T., Blye, J., Boguslavskiy, L., Bonnet, C., Boukhgalter, B., Brown, A., Cahill, P., Calixte, N., Camarata, J., Cheshatsang, Y., Chu, J., Citroen, M., Collymore, A., Cooke, P., Dawoe, T., Daza, R., Decktor, K., DeGray, S., Dhargay, N., Dooley, K., Dorje, P., Dorjee, K., Dorris, L., Duffey, N., Dupes, A., Egbiremolen, O., Elong, R., Falk, J., Farina, A., Faro, S., Ferguson, D., Ferreira, P., Fisher, S., FitzGerald, M., Foley, K., Foley, C., Franke, A., Friedrich, D., Gage, D., Garber, M., Gearin, G., Giannoukos, G., Goode, T., Goyette, A., Graham, J., Grandbois, E., Gyaltsen, K., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Healy, C., Hegarty, R., Honan, T., Horn, A., Houde, N., Hughes, L., Hunnicutt, L., Husby, M., Jester, B., Jones, C., Kamat, A., Kanga, B., Kells, C., Khazanovich, D., Kieu, A. C., Kisner, P., Kumar, M., Lance, K., Landers, T., Lara, M., Lee, W., Leger, J. P., Lennon, N., Leuper, L., LeVine, S., Liu, J., Liu, X., Lokyitsang, Y., Lokyitsang, T., Lui, A., Macdonald, J., Major, J., Marabella, R., Maru, K., Matthews, C., McDonough, S., Mehta, T., Meldrim, J., Melnikov, A., Meneus, L., Mihalev, A., Mihova, T., Miller, K., Mittelman, R., Mlenga, V., Mulrain, L., Munson, G., Navidi, A., Naylor, J., Nguyen, T., Nguyen, N., Nguyen, C., Nicol, R., Norbu, N., Norbu, C., Novod, N., Nyima, T., Olandt, P., O'Neill, B., O'Neill, K., Osman, S., Oyono, L., Patti, C., Perrin, D., Phunkhang, P., Pierre, F., Priest, M., Rachupka, A., Raghuraman, S., Rameau, R., Ray, V., Raymond, C., Rege, F., Rise, C., Rogers, J., Rogov, P., Sahalie, J., Settipalli, S., Sharpe, T., Shea, T., Sheehan, M., Sherpa, N., Shi, J., Shih, D., Sloan, J., Smith, C., Sparrow, T., Stalker, J., Stange-Thomann, N., Stavropoulos, S., Stone, C., Stone, S., Sykes, S., Tchuinga, P., Tenzing, P., Tesfaye, S., Thoulutsang, D., Thoulutsang, Y., Topham, K., Topping, I., Tsamla, T., Vassiliev, H., Venkataraman, V., Vo, A., Wangchuk, T., Wangdi, T., Weiand, M., Wilkinson, J., Wilson, A., Yadav, S., Yang, S., Yang, X., Young, G., Yu, Q., Zainoun, J., Zembek, L., Zimmer, A. & Lander, E. S. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438(7069): 803-819.

Lingaas, F., Aarskaug, T., Sletten, M., Bjerkas, I., Grimholt, U., Moe, L., Juneja, R. K., Wilton, A. N., Galibert, F., Holmes, N. G. & Dolf, G. (1998). Genetic markers linked to neuronal ceroid lipofuscinosis in English setter dogs. *Anim Genet* 29(5): 371-376.

Linterman, K. S., Palmer, D. N., Kay, G. W., Barry, L. A., Mitchell, N. L., McFarlane, R. G., Black, M. A., Sands, M. S. & Hughes, S. M. (2011). Lentiviral-mediated gene transfer to the sheep brain: implications for gene therapy in Batten disease. *Hum Gene Ther* 22(8): 1011-1020.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L. & Law, M. (2012). Comparison of next-generation sequencing systems. *J Biomed Biotechnol* 2012: 251364.

Liu, C., Du, Y., Xie, W. & Gui, C. (2013). Purification of plasmid and BAC transgenic DNA constructs. *Methods Mol Biol* 1027: 203-215.

Lönnqvist, T., Vanhanen, S. L., Vettenranta, K., Autti, T., Rapola, J., Santavuori, P. & Saarinen-Pihkala, U. M. (2001). Hematopoietic stem cell transplantation in infantile neuronal ceroid lipofuscinosis. *Neurology* 57(8): 1411-1416.

Loots, G. G., Locksley, R. M., Blankespoor, C. M., Wang, Z. E., Miller, W., Rubin, E. M. & Frazer, K. A. (2000). Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science* 288(5463): 136-140.

Loots, G. G., Ovcharenko, I., Pachter, L., Dubchak, I. & Rubin, E. M. (2002). rVista for comparative sequence-based discovery of functional transcription factor binding sites. *Genome Res* 12(5): 832-839.

Loots, G. G. (2008). Genomic identification of regulatory elements by evolutionary sequence comparison and functional analysis. *Adv Genet* 61: 269-293.

Ludwig, M. Z. (2002). Functional evolution of noncoding DNA. *Curr Opin Genet Dev* 12(6): 634-639.

Lyly, A., von Schantz, C., Heine, C., Schmiedt, M. L., Sipila, T., Jalanko, A. & Kyttala, A. (2009). Novel interactions of CLN5 support molecular networking between Neuronal Ceroid Lipofuscinosis proteins. *BMC Cell Biol* 10: 83.

Maddox, J. F., Davies, K. P., Crawford, A. M., Hulme, D. J., Vaiman, D., Cribiu, E. P., Freking, B. A., Beh, K. J., Cockett, N. E., Kang, N., Riffkin, C. D., Drinkwater, R., Moore, S. S., Dodds, K. G., Lumsden, J. M., van Stijn, T. C., Phua, S. H., Adelson, D. L., Burkin, H. R., Broom, J. E., Buitkamp, J., Cambridge, L., Cushwa, W. T., Gerard, E., Galloway, S. M., Harrison, B., Hawken, R. J., Hiendleder, S., Henry, H. M., Medrano, J. F., Paterson, K. A., Schibler, L., Stone, R. T. & van Hest, B. (2001). An enhanced linkage map of the sheep genome comprising more than 1000 loci. *Genome Res* 11(7): 1275-1289.

Mahmood, F., Fu, S., Cooke, J., Wilson, S. W., Cooper, J. D. & Russell, C. (2013). A zebrafish model of CLN2 disease is deficient in tripeptidyl peptidase 1 and displays progressive neurodegeneration accompanied by a reduction in proliferation. *Brain* 136(Pt 5): 1488-1507.

Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., Howard, E., Shendure, J. & Turner, D. J. (2010). Target-enrichment strategies for next-generation sequencing. *Nat Methods* 7(2): 111-118.

Mardis, E. R. (2008). Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* 9: 387-402.

Mardis, E. R. (2013). Next-generation sequencing platforms. *Annu Rev Anal Chem (Palo Alto Calif)* 6: 287-303.

Margulies, M., Egholm, M., Altman, W., Attiya, S., Bader, J., Bemben, L., Berka, J., Braverman, M., Chen, Y. & Chen, Z. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376 - 380.

Martinus, R. D., Harper, P. A. W., Jolly, R. D., Bayliss, S. L., Midwinter, G. G., Shaw, G. J. & Palmer, D. N. (1991). Bovine ceroid-lipofuscinosis (Batten's disease): The major component stored is the DCCD-reactive proteolipid, subunit c, of mitochondrial ATP synthase. *Veterinary Research Communications* 15(2): 85-94.

Matsuda, F., Tsugawa, H. & Fukusaki, E. (2013). Method for assessing the statistical significance of mass spectral similarities using basic local alignment search tool statistics. *Anal Chem* 85(17): 8291-8297.

Mattina, T., Perrotta, C. S. & Grossfeld, P. (2009). Jacobsen syndrome. *Orphanet J Rare Dis* 4: 9.

Mayhew, I. G., Jolly, R. D., Pickett, B. T. & Slack, P. M. (1985). Ceroid-lipofuscinosis (Batten's disease): pathogenesis of blindness in the ovine model. *Neuropathol Appl Neurobiol* 11(4): 273-290.

Mayor, C., Brudno, M., Schwartz, J. R., Poliakov, A., Rubin, E. M., Frazer, K. A., Pachter, L. S. & Dubchak, I. (2000). VISTA : visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* 16(11): 1046-1047.

Melville, S. A., Wilson, C. L., Chiang, C. S., Studdert, V. P., Lingaas, F. & Wilton, A. N. (2005). A mutation in canine CLN5 causes neuronal ceroid lipofuscinosis in Border collie dogs. *Genomics* 86(3): 287-294.

Mignone, F., Gissi, C., Liuni, S. & Pesole, G. (2002). Untranslated regions of mRNAs. *Genome Biol* 3(3): 28.

Mikkelsen, T. S., Wakefield, M. J., Aken, B., Amemiya, C. T., Chang, J. L., Duke, S., Garber, M., Gentles, A. J., Goodstadt, L., Heger, A., Jurka, J., Kamal, M., Mauceli, E., Searle, S. M. J., Sharpe, T., Baker, M. L., Batzer, M. A., Benos, P. V., Belov, K., Clamp, M., Cook, A., Cuff, J., Das, R., Davidow, L., Deakin, J. E., Fazzari, M. J., Glass, J. L., Grabherr, M., Greally, J. M., Gu, W., Hore, T. A., Huttley, G. A., Kleber, M., Jirtle, R. L., Koina, E., Lee, J. T., Mahony, S., Marra, M. A., Miller, R. D., Nicholls, R. D., Oda, M., Papenfuss, A. T., Parra, Z. E., Pollock, D. D., Ray, D. A., Schein, J. E., Speed, T. P., Thompson, K., VandeBerg, J. L., Wade, C. M., Walker, J. A., Waters, P. D., Webber, C., Weidman, J. R., Xie, X., Zody, M. C., Graves, J. A. M., Ponting, C. P., Breen, M., Samollow, P. B., Lander, E. S. & Lindblad-Toh, K. (2007). Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences. *Nature* 447(7141): 167-177.

Minatel, L., Underwood, S. C. & Carfagnini, J. C. (2000). Ceroid-lipofuscinosis in a Cocker Spaniel dog. *Vet Pathol* 37(5): 488-490.

Mitchison, H. M., Bernard, D. J., Greene, N. D., Cooper, J. D., Junaid, M. A., Pullarkat, R. K., de Vos, N., Breuning, M. H., Owens, J. W., Mobley, W. C., Gardiner, R. M., Lake, B. D., Taschner, P. E. & Nussbaum, R. L. (1999). Targeted disruption of the Cln3 gene provides a mouse model for Batten disease. The Batten Mouse Model Consortium [corrected]. *Neurobiol Dis* 6(5): 321-334.

Mizukami, K., Chang, H. S., Yabuki, A., Kawamichi, T., Kawahara, N., Hayashi, D., Hossain, M. A., Rahman, M. M., Uddin, M. M. & Yamato, O. (2011). Novel rapid genotyping assays for neuronal ceroid lipofuscinosis in Border Collie dogs and high frequency of the mutant allele in Japan. *J Vet Diagn Invest* 23(6): 1131-1139.

Mnatzakanian, G. N., Lohi, H., Munteanu, I., Alfred, S. E., Yamada, T., MacLeod, P. J., Jones, J. R., Scherer, S. W., Schanen, N. C., Friez, M. J., Vincent, J. B. & Minassian, B. A. (2004). A previously unidentified MECP2 open reading frame defines a new protein isoform relevant to Rett syndrome. *Nat Genet* 36(4): 339-341.

Mole, S. E., Michaux, G., Codlin, S., Wheeler, R. B., Sharp, J. D. & Cutler, D. F. (2004). CLN6, which is associated with a lysosomal storage disease, is an endoplasmic reticulum protein. *Exp Cell Res* 298(2): 399-406.

Mole, S., Williams, R. & Goebel, H. H. (2011). *The Neuronal ceroid lipofuscinoses (Batten disease)*. Oxford University Press.

Morozova, O. & Marra, M. A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92(5): 255-264.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. & Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* 1: 263-273.

Myllykangas, L., Tyynela, J., Page-McCaw, A., Rubin, G. M., Haltia, M. J. & Feany, M. B. (2005). Cathepsin D-deficient Drosophila recapitulate the key features of neuronal ceroid lipofuscinoses. *Neurobiol Dis* 19(1-2): 194-199.

Nakamoto, Y., Yamato, O., Uchida, K., Nibe, K., Tamura, S., Ozawa, T., Ueoka, N., Nukaya, A., Yabuki, A. & Nakaichi, M. (2011). Neuronal ceroid-lipofuscinosis in longhaired Chihuahuas: clinical, pathologic, and MRI findings. *J Am Anim Hosp Assoc* 47(4): 14.

Nakayama, H., Uchida, K., Shouda, T., Uetsuka, K., Sasaki, N. & Goto, N. (1993). Systemic ceroid-lipofuscinosis in a Japanese domestic cat. *J Vet Med Sci* 55(5): 829-831.

Narfström, K. & Wrigstad, A. (1995). Clinical, electrophysiological, and morphological findings in a case of neuronal ceroid lipofuscinosis in the Polish Owczarek Nizinny (PON) dog. *Vet Q* 17(1).

Narfström, K., Wrigstad, A., Ekesten, B. & Berg, A. L. (2007). Neuronal ceroid lipofuscinosis: clinical and morphologic findings in nine affected Polish Owczarek Nizinny (PON) dogs. *Vet Ophthalmol* 10(2): 111-120.

Needleman, S. B. & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* 48(3): 443-453.

Nibe, K., Miwa, Y., Matsunaga, S., Chambers, J. K., Uetsuka, K., Nakayama, H. & Uchida, K. (2011). Clinical and pathologic features of neuronal ceroid-lipofuscinosis in a ferret (Mustela putorius furo). *Vet Pathol* 48(6): 1185-1189.

Nicholas, K. B., Nicholas H.B. Jr. & Deerfield, D. W. (1997). GeneDoc: Analysis and Visualization of Genetic Variation. 4:14.

Nilsson, S. E. & Wrigstad, A. (1997). Electrophysiology in some animal and human hereditary diseases involving the retinal pigment epithelium. *Eye* 11(Pt 5): 698-706.

Nimmo Wilkie, J. S. & Hudson, E. B. (1982). Neuronal and generalized ceroid-lipofuscinosis in a cocker spaniel. *Vet Pathol* 19(6): 623-628.

Nobrega, M. A., Ovcharenko, I., Afzal, V. & Rubin, E. M. (2003). Scanning human gene deserts for long-range enhancers. *Science* 302(5644): 413.

Noskova, L., Stranecky, V., Hartmannova, H., Pristoupilova, A., Baresova, V., Ivanek, R., Hulkova, H., Jahnova, H., van der Zee, J., Staropoli, J. F., Sims, K. B., Tyynela, J., Van Broeckhoven, C., Nijssen, P. C., Mole, S. E., Elleder, M. & Kmoch, S. (2011). Mutations in DNAJC5, encoding cysteine-string protein alpha, cause autosomal-dominant adult-onset neuronal ceroid lipofuscinosis. *Am J Hum Genet* 89(2): 241-252.

O'Brien, D. P. & Katz, M. L. (2008). Neuronal ceroid lipofuscinosis in 3 Australian shepherd littermates. *J Vet Intern Med* 22(2): 472-475.

Ohno, S. (1972). So much "junk" DNA in our genome. *Brookhaven Symp Biol* 23: 366-370.

Okazaki, H., Jonas, E., Thomson, P. C. & Tammen, I. (2013).Bovine neuronal ceroild lipofusinosis in Australian Devon cattle. 203-206 Armidale: Association for the Advancement of Animal Breeding and Genetics.

Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J. J. & de Jong, P. J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 52(1): 1-8.

Oswald, M. J., Palmer, D. N. & Damak, S. (1999). Splicing variants in sheep CLN3, the gene underlying juvenile neuronal ceroid lipofuscinosis. *Mol Genet Metab* 67(2): 169-175.

Oswald, M. J., Palmer, D. N., Kay, G. W., Barwell, K. J. & Cooper, J. D. (2008). Location and connectivity determine GABAergic interneuron survival in the brains of South Hampshire sheep with CLN6 neuronal ceroid lipofuscinosis. *Neurobiol Dis* 32(1): 50-65.

Oswald, M. J., Palmer, D. N., Kay, G. W., Shemilt, S. J., Rezaie, P. & Cooper, J. D. (2005). Glial activation spreads from specific cerebral foci and precedes neurodegeneration in presymptomatic ovine neuronal ceroid lipofuscinosis (CLN6). *Neurobiol Dis* 20(1): 49-63.

Palmer, D. N., Barry, L. A., Tyynela, J. & Cooper, J. D. (2013). NCL disease mechanisms. *Biochim Biophys Acta* 1832(11): 1882-1893.

Palmer, D. N., Fearnley, I. M., Medd, S. M., Walker, J. E., Martinus, R. D., Bayliss, S. L., Hall, N. A., Lake, B. D., Wolfe, L. S. & Jolly, R. D. (1989 (b)). Lysosomal storage of the DCCD reactive proteolipid subunit of mitochondrial ATP synthase in human and ovine ceroid lipofuscinoses. *Adv Exp Med Biol* 266: 211-222.

Palmer, D. N., Jolly, R. D., van Mil, H. C., Tyynela, J. & Westlake, V. J. (1997 (b)). Different patterns of hydrophobic protein storage in different forms of neuronal ceroid lipofuscinosis (NCL, Batten disease). *Neuropediatrics* 28(1): 45-48.

Palmer, D. N., Martinus, R. D., Cooper, S. M., Midwinter, G. G., Reid, J. C. & Jolly, R. D. (1989 (a)). Ovine ceroid lipofuscinosis. The major lipopigment protein and the lipid-binding subunit of mitochondrial ATP synthase have the same NH2-terminal sequence. *J Biol Chem* 264(10): 5736-5740.

Palmer, D. N., Tyynela, J., van Mil, H. C., Westlake, V. J. & Jolly, R. D. (1997 (a)). Accumulation of sphingolipid activator proteins (SAPs) A and D in granular osmiophilic deposits in miniature Schnauzer dogs with ceroid-lipofuscinosis. *J Inherit Metab Dis* 20(1): 74-84.

Palmieri, M., Impey, S., Kang, H., di Ronza, A., Pelz, C., Sardiello, M. & Ballabio, A. (2011). Characterization of the CLEAR network reveals an integrated control of cellular clearance pathways. *Hum Mol Genet* 20(19): 3852-3866.

Pao, W., Miller, V., Zakowski, M., Doherty, J., Politi, K., Sarkaria, I., Singh, B., Heelan, R., Rusch, V., Fulton, L., Mardis, E., Kupfer, D., Wilson, R., Kris, M. & Varmus, H. (2004). EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc Natl Acad Sci U S A* 101(36): 13306-13311.

Parenti, G., Pignata, C., Vajro, P. & Salerno, M. (2013). New strategies for the treatment of lysosomal storage diseases (review). *Int J Mol Med* 31(1): 11-20.

Peterson, M. E., Chen, F., Saven, J. G., Roos, D. S., Babbitt, P. C. & Sali, A. (2009). Evolutionary constraints on structural similarity in orthologs and paralogs. *Protein Sci* 18(6): 1306-1315.

Philipps, D. L., Park, J. W. & Graveley, B. R. (2004). A computational and experimental approach toward a priori identification of alternatively spliced exons. *Rna* 10(12): 1838-1844.

Poet, M., Kornak, U., Schweizer, M., Zdebik, A. A., Scheel, O., Hoelter, S., Wurst, W., Schmitt, A., Fuhrmann, J. C., Planells-Cases, R., Mole, S. E., Hubner, C. A. & Jentsch, T. J. (2006). Lysosomal storage disease upon disruption of the neuronal chloride transport protein ClC-6. *Proc Natl Acad Sci U S A* 103(37): 13854-13859.

Porreca, G. J., Zhang, K., Li, J. B., Xie, B., Austin, D., Vassallo, S. L., LeProust, E. M., Peck, B. J., Emig, C. J., Dahl, F., Gao, Y., Church, G. M. & Shendure, J. (2007). Multiplex amplification of large sets of human exons. *Nat Methods* 4(11): 931-936.

Porto, C., Cardone, M., Fontana, F., Rossi, B., Tuzzi, M. R., Tarallo, A., Barone, M. V., Andria, G. & Parenti, G. (2009). The pharmacological chaperone N-butyldeoxynojirimycin enhances enzyme replacement therapy in Pompe disease fibroblasts. *Mol Ther* 17(6): 964-971.

Prabhakar, S., Poulin, F., Shoukry, M., Afzal, V., Rubin, E. M., Couronne, O. & Pennacchio, L. A. (2006). Close sequence comparisons are sufficient to identify human cis-regulatory elements. *Genome Res* 16(7): 855-863.

Quail, M. A., Matthews, L., Sims, S., Lloyd, C., Beasley, H. & Baxter, S. W. (2011). Genomic libraries: II. Subcloning, sequencing, and assembling large-insert genomic DNA clones. *Methods Mol Biol* 772: 59-81.

Raab, J. R. & Kamakaka, R. T. (2010). Insulators and promoters: closer than we think. *Nat Rev Genet* 11(6): 439-446.

Raben, N., Schreiner, C., Baum, R., Takikita, S., Xu, S., Xie, T., Myerowitz, R., Komatsu, M., Van der Meulen, J. H., Nagaraju, K., Ralston, E. & Plotz, P. H. (2010). Suppression of autophagy permits successful enzyme replacement therapy in a lysosomal storage disorder--murine Pompe disease. *Autophagy* 6(8): 1078-1089.

Rajagopal, N., Srinivasan, S., Kooshesh, K., Guo, Y., Edwards, M. D., Banerjee, B., Syed, T., Emons, B. J., Gifford, D. K. & Sherwood, R. I. (2016). High-throughput mapping of regulatory DNA. *Nat Biotechnol* 25(10).

Ramirez, A., Heimbach, A., Grundemann, J., Stiller, B., Hampshire, D., Cid, L. P., Goebel, I., Mubaidin, A. F., Wriekat, A. L., Roeper, J., Al-Din, A., Hillmer, A. M., Karsak, M., Liss, B., Woods, C. G., Behrens, M. I. & Kubisch, C. (2006). Hereditary parkinsonism with dementia is caused by mutations in ATP13A2, encoding a lysosomal type 5 P-type ATPase. *Nat Genet* 38(10): 1184-1191.

Ranta, S., Zhang, Y., Ross, B., Lonka, L., Takkunen, E., Messer, A., Sharp, J., Wheeler, R., Kusumi, K., Mole, S., Liu, W., Soares, M. B., Bonaldo, M. F., Hirvasniemi, A., de la Chapelle, A., Gilliam, T. C. & Lehesjoki, A. E. (1999). The neuronal ceroid lipofuscinoses in human EPMR and mnd mutant mice are associated with mutations in CLN8. *Nat Genet* 23(2): 233-236.

Read, W. K. & Bridges, C. H. (1969). Neuronal lipodystrophy. Occurrence in an inbred strain of cattle. *Pathol Vet* 6(3): 235-243.

Reece, R. L. & MacWhirter, P. (1988). Neuronal ceroid lipofuscinosis in a lovebird. *Vet Rec* 122(8): 187.

Ren, H. & Stiles, G. L. (1994). Posttranscriptional mRNA processing as a mechanism for regulation of human A1 adenosine receptor expression. *Proc Natl Acad Sci U S A* 91(11): 4864-4866.

Riis, R. C., Cummings, J. F., Loew, E. R. & de Lahunta, A. (1992). Tibetan terrier model of canine ceroid lipofuscinosis. *Am J Med Genet* 42(4): 615-621.

Robledo, R., Beggs, W. & Bender, P. (2003). A simple and cost-effective method for rapid genotyping of insertion/deletion polymorphisms. *Genomics* 82(5): 580-582.

Rodriguez-Caballero, A., Torres-Lagares, D., Rodriguez-Perez, A., Serrera-Figallo, M. A., Hernandez-Guisado, J. M. & Machuca-Portillo, G. (2010). Cri du chat syndrome: a critical review. *Med Oral Patol Oral Cir Bucal* 15(3): e473-478.

Ronaghi, M., Uhlen, M. & Nyren, P. (1998). A sequencing method based on real-time pyrophosphate. *Science* 281(5375): 363, 365.

Rossmeisl, J. H., Jr., Duncan, R., Fox, J., Herring, E. S. & Inzana, K. D. (2003). Neuronal ceroid-lipofuscinosis in a Labrador Retriever. *J Vet Diagn Invest* 15(5): 457-460.

Rozen, S. & Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 132: 365-386.

Sachs, B. (1896). A family form of idiocacy, generally fatal and associated wuth early blindness (amaurotic family idiocacy). *NY Med J* 63: 697-703.

Saftig, P., Hetman, M., Schmahl, W., Weber, K., Heine, L., Mossmann, H., Koster, A., Hess, B., Evers, M., von Figura, K. & et al. (1995). Mice deficient for the lysosomal proteinase cathepsin D exhibit progressive atrophy of the intestinal mucosa and profound destruction of lymphoid cells. *EMBO J* 14(15): 3599-3608.

Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M. & Shiroishi, T. (2005). Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development* 132(4): 797-803.

Sandelin, A., Wasserman, W. W. & Lenhard, B. (2004). ConSite: web-based prediction of regulatory elements using cross-species comparison. *Nucleic Acids Res.*

Sanders, D. N., Farias, F. H., Johnson, G. S., Chiang, V., Cook, J. R., O'Brien, D. P., Hofmann, S. L., Lu, J. Y. & Katz, M. L. (2010). A mutation in canine PPT1 causes early onset neuronal ceroid lipofuscinosis in a Dachshund. *Mol Genet Metab* 100(4): 349-356.

Sanders, D. N., Kanazono, S., Wininger, F. A., Whiting, R. E., Flournoy, C. A., Coates, J. R., Castaner, L. J., O'Brien, D. P. & Katz, M. L. (2011). A reversal learning task detects cognitive deficits in a Dachshund model of late-infantile neuronal ceroid lipofuscinosis. *Genes Brain Behav* 10(7): 798-804.

Sanger, F. & Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* 94(3): 441-448.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A.*

Santavuori, P., Haltia, M., Rapola, J. & Raitta, C. (1973). Infantile type of so-called neuronal ceroid-lipofuscinosis. 1. A clinical study of 15 patients. *J Neurol Sci* 18(3): 257-267.

Santavuori, P. & Moren, R. (1977). Experience of antioxidant treatment in neuronal ceroid-lipofuscinosis of Spielmeyer-Sjogren type. *Neuropadiatrie* 8(4): 333-344.

Santavuori, P., Heiskala, H., Westermarck, T., Sainio, K. & Moren, R. (1988). Experience over 17 years with antioxidant treatment in Spielmeyer-Sjogren disease. *Am J Med Genet Suppl* 5: 265-274.

Santavuori, P., Heiskala, H., Autti, T., Johansson, E. & Westermarck, T. (1989). Comparison of the clinical courses in patients with juvenile neuronal ceroid lipofuscinosis receiving antioxidant treatment and those without antioxidant treatment. *Adv Exp Med Biol* 266: 273-282.

Santorelli, F. M., Garavaglia, B., Cardona, F., Nardocci, N., Bernardina, B. D., Sartori, S., Suppiej, A., Bertini, E., Claps, D., Battini, R., Biancheri, R., Filocamo, M., Pezzini, F. & Simonati, A. (2013). Molecular epidemiology of childhood neuronal ceroid-lipofuscinosis in Italy. *Orphanet Journal of Rare Diseases* 8(1): 19.

Sardiello, M., Palmieri, M., di Ronza, A., Medina, D. L., Valenza, M., Gennarino, V. A., Di Malta, C., Donaudy, F., Embrione, V., Polishchuk, R. S., Banfi, S., Parenti, G., Cattaneo,

E. & Ballabio, A. (2009). A gene network regulating lysosomal biogenesis and function. *Science* 325(5939): 473-477.

Savukoski, M., Kestila, M., Williams, R., Jarvela, I., Sharp, J., Harris, J., Santavuori, P., Gardiner, M. & Peltonen, L. (1994). Defined chromosomal assignment of CLN5 demonstrates that at least four genetic loci are involved in the pathogenesis of human ceroid lipofuscinoses. *Am J Hum Genet* 55(4): 695-701.

Savukoski, M., Klockars, T., Holmberg, V., Santavuori, P., Lander, E. S. & Peltonen, L. (1998). CLN5, a novel gene encoding a putative transmembrane protein mutated in Finnish variant late infantile neuronal ceroid lipofuscinosis. *Nat Genet* 19(3): 286-288.

Sayers, E. W., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., DiCuccio, M., Edgar, R., Federhen, S., Feolo, M., Geer, L. Y., Helmberg, W., Kapustin, Y., Landsman, D., Lipman, D. J., Madden, T. L., Maglott, D. R., Miller, V., Mizrachi, I., Ostell, J., Pruitt, K. D., Schuler, G. D., Sequeira, E., Sherry, S. T., Shumway, M., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusova, T. A., Wagner, L., Yaschenko, E. & Ye, J. (2009). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 37(Database issue): 21.

Schulz, A., Kohlschutter, A., Mink, J., Simonati, A. & Williams, R. (2013). NCL diseases - clinical perspectives. *Biochim Biophys Acta* 11(6): 17.

Seehafer, S. S., Ramirez-Montealegre, D., Wong, A. M., Chan, C. H., Castaneda, J., Horak, M., Ahmadi, S. M., Lim, M. J., Cooper, J. D. & Pearce, D. A. (2011). Immunosuppression alters disease severity in juvenile Batten disease mice. *J Neuroimmunol* 230(1-2): 169-172.

Selden, N. R., Al-Uzri, A., Huhn, S. L., Koch, T. K., Sikora, D. M., Nguyen-Driver, M. D., Guillaume, D. J., Koh, J. L., Gultekin, S. H., Anderson, J. C., Vogel, H., Sutcliffe, T. L., Jacobs, Y. & Steiner, R. D. (2013). Central nervous system stem cell transplantation for children with neuronal ceroid lipofuscinosis. *J Neurosurg Pediatr* 11(6): 643-652.

Settembre, C., Fraldi, A., Medina, D. L. & Ballabio, A. (2013). Signals from the lysosome: a control centre for cellular clearance and energy metabolism. *Nat Rev Mol Cell Biol* 14(5): 283-296.

Shabalina, S. A. & Spiridonov, N. A. (2004). The mammalian transcriptome and the function of non-coding DNA sequences. *Genome Biol* 5(4): 25.

Sharan, R., Ben-Hur, A., Loots, G. G. & Ovcharenko, I. (2004). CREME: Cis-Regulatory Module Explorer for the human genome. *Nucleic Acids Res* 32(Web Server issue): W253-256.

Sharp, J. D., Wheeler, R. B., Lake, B. D., Savukoski, M., Jarvela, I. E., Peltonen, L., Gardiner, R. M. & Williams, R. E. (1997). Loci for classical and a variant late infantile neuronal ceroid lipofuscinosis map to chromosomes 11p15 and 15q21-23. *Hum Mol Genet* 6(4): 591-595.

Sharp, J. D., Wheeler, R. B., Lake, B. D., Fox, M., Gardiner, R. M. & Williams, R. E. (1999). Genetic and physical mapping of the CLN6 gene on chromosome 15q21-23. *Mol Genet Metab* 66(4): 329-331.

Sharp, J. D., Wheeler, R. B., Parker, K. A., Gardiner, R. M., Williams, R. E. & Mole, S. E. (2003). Spectrum of CLN6 mutations in variant late infantile neuronal ceroid lipofuscinosis. *Hum Mutat* 22(1): 35-42.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D. & Higgins, D. G. (2011). Fast, scalable

generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7(539): 75.

Siintola, E., Partanen, S., Stromme, P., Haapanen, A., Haltia, M., Maehlen, J., Lehesjoki, A. E. & Tyynela, J. (2006). Cathepsin D deficiency underlies congenital human neuronal ceroid-lipofuscinosis. *Brain* 129(Pt 6): 1438-1445.

Siintola, E., Topcu, M., Aula, N., Lohi, H., Minassian, B. A., Paterson, A. D., Liu, X. Q., Wilson, C., Lahtinen, U., Anttonen, A. K. & Lehesjoki, A. E. (2007). The novel neuronal ceroid lipofuscinosis gene MFSD8 encodes a putative lysosomal transporter. *Am J Hum Genet* 81(1): 136-146.

Sisk, D. B., Levesque, D. C., Wood, P. A. & Styer, E. L. (1990). Clinical and pathologic features of ceroid lipofuscinosis in two Australian cattle dogs. *J Am Vet Med Assoc* 197(3): 361-364.

Sleat, D. E., Donnelly, R. J., Lackland, H., Liu, C. G., Sohar, I., Pullarkat, R. K. & Lobel, P. (1997). Association of mutations in a lysosomal protein with classical late-infantile neuronal ceroid lipofuscinosis. *Science* 277(5333): 1802-1805.

Sleat, D. E., Sohar, I., Gin, R. M. & Lobel, P. (2001). Aminoglycoside-mediated suppression of nonsense mutations in late infantile neuronal ceroid lipofuscinosis. *Eur J Paediatr Neurol* 5 Suppl A: 57-62.

Sleat, D. E., Wiseman, J. A., El-Banna, M., Kim, K. H., Mao, Q., Price, S., Macauley, S. L., Sidman, R. L., Shen, M. M., Zhao, Q., Passini, M. A., Davidson, B. L., Stewart, G. R. & Lobel, P. (2004). A mouse model of classical late-infantile neuronal ceroid lipofuscinosis based on targeted disruption of the CLN2 gene results in a loss of tripeptidyl-peptidase I activity and progressive neurodegeneration. *J Neurosci* 24(41): 9117-9126.

Sleat, D. E., Ding, L., Wang, S., Zhao, C., Wang, Y., Xin, W., Zheng, H., Moore, D. F., Sims, K. B. & Lobel, P. (2009). Mass spectrometry-based protein profiling to determine the cause of lysosomal storage diseases of unknown etiology. *Mol Cell Proteomics* 8(7): 1708-1718.

Smit, A., Hubley, R. & Green, P. (1996 - 2010). RepeatMasker Open-3.0.

Smith, K. R., Dahl, H. H., Canafoglia, L., Andermann, E., Damiano, J., Morbin, M., Bruni, A. C., Giaccone, G., Cossette, P., Saftig, P., Grotzinger, J., Schwake, M., Andermann, F., Staropoli, J. F., Sims, K. B., Mole, S. E., Franceschetti, S., Alexander, N. A., Cooper, J. D., Chapman, H. A., Carpenter, S., Berkovic, S. F. & Bahlo, M. (2013). Cathepsin F mutations cause Type B Kufs disease, an adult-onset neuronal ceroid lipofuscinosis. *Hum Mol Genet* 22(7): 1417-1423.

Smith, K. R., Damiano, J., Franceschetti, S., Carpenter, S., Canafoglia, L., Morbin, M., Rossi, G., Pareyson, D., Mole, S. E., Staropoli, J. F., Sims, K. B., Lewis, J., Lin, W. L., Dickson, D. W., Dahl, H. H., Bahlo, M. & Berkovic, S. F. (2012). Strikingly different clinicopathological phenotypes determined by progranulin-mutation dosage. *Am J Hum Genet* 90(6): 1102-1107.

Smith, N. G., Brandstrom, M. & Ellegren, H. (2004). Evidence for turnover of functional noncoding DNA in mammalian genome evolution. *Genomics* 84(5): 806-813.

Smith, R., Sutton, R., Jolly, R. & Smith, K. (1996). A retinal degeneration associated with ceroid-lipofuscinosis in adult miniature Schnauzers. *Vet. Comp. Opthalmol.* 6 (1996) 187–191.

Sohal, R. S. (1981). *Age Pigments*. Elsevier Science Ltd.

Sonnhammer, E. L., von Heijne, G. & Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* 6: 175-182.

Staropoli, J. F., Karaa, A., Lim, E. T., Kirby, A., Elbalalesy, N., Romansky, S. G., Leydiker, K. B., Coppel, S. H., Barone, R., Xin, W., MacDonald, M. E., Abdenur, J. E., Daly, M. J., Sims, K. B. & Cotman, S. L. (2012). A homozygous mutation in KCTD7 links neuronal ceroid lipofuscinosis to the ubiquitin-proteasome system. *Am J Hum Genet* 91(1): 202-208.

Stein, L. D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J. E., Harris, T. W., Arva, A. & Lewis, S. (2002). The generic genome browser: a building block for a model organism system database. *Genome Res* 12(10): 1599-1610.

Steinfeld, R., Reinhardt, K., Schreiber, K., Hillebrand, M., Kraetzner, R., Bruck, W., Saftig, P. & Gartner, J. (2006). Cathepsin D deficiency is associated with a human neurodegenerative disorder. *Am J Hum Genet* 78(6): 988-998.

Stephens, M., Sloan, J. S., Robertson, P. D., Scheet, P. & Nickerson, D. A. (2006). Automating sequence-based detection and genotyping of SNPs from diploid samples. *Nat Genet* 38(3): 375-381.

Strausberg, R. L., Simpson, A. J. & Wooster, R. (2003). Sequence-based cancer genomics: progress, lessons and opportunities. *Nat Rev Genet* 4(6): 409-418.

Studdert, V. P. & Mitten, R. W. (1991). Clinical features of ceroid lipofuscinosis in border collie dogs. *Aust Vet J* 68(4): 137-140.

Summerer, D. (2009). Enabling technologies of genomic-scale sequence enrichment for targeted high-throughput sequencing. *Genomics* 94(6): 363-368.

Tabet-Aoul, K., Oustry-Vaiman, A., Vaiman, D., Saidi-Mehtar, N., Cribiu, E. P. & Lantier, F. (2000). Cytogenetical anchoring of sheep linkage map and syntenic groups using a sheep BAC library. *Genet Sel Evol* 32(4): 441-457.

Tammen, I., Cook, R. W., Nicholas, F. W. & Raadsma, H. W. (2001). Neuronal ceroid lipofuscinosis in Australian Merino sheep: a new animal model. *Eur J Paediatr Neurol* 5 Suppl A: 37-41.

Tammen, I., Kurtz, E., Tier, B., Gibson, A. J., Windsor, P. A. & Raadsma, H. (2002).Neuronal ceroid lipofuscinosis in Devon cattle – pedigree analysis and preliminary molecular genetic investigation. In *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*France.

Tammen, I., Houweling, P. J., Frugier, T., Mitchell, N. L., Kay, G. W., Cavanagh, J. A., Cook, R. W., Raadsma, H. W. & Palmer, D. N. (2006). A missense mutation (c.184C>T) in ovine CLN6 causes neuronal ceroid lipofuscinosis in Merino sheep whereas affected South Hampshire sheep have reduced levels of CLN6 mRNA. *Biochim Biophys Acta* 1762(10): 898-905.

Taylor, R. M. & Farrow, B. R. (1988). Ceroid-lipofuscinosis in border collie dogs. *Acta Neuropathol* 75(6): 627-631.

Taylor, R. M. & Farrow, B. R. (1992). Ceroid lipofuscinosis in the border collie dog: retinal lesions in an animal model of juvenile Batten disease. *Am J Med Genet* 42(4): 622-627.

Thomas, J. W., Touchman, J. W., Blakesley, R. W., Bouffard, G. G., Beckstrom-Sternberg, S. M., Margulies, E. H., Blanchette, M., Siepel, A. C., Thomas, P. J., McDowell, J. C., Maskeri, B., Hansen, N. F., Schwartz, M. S., Weber, R. J., Kent, W. J., Karolchik, D., Bruen, T. C., Bevan, R., Cutler, D. J., Schwartz, S., Elnitski, L., Idol, J. R., Prasad, A. B.,

Lee-Lin, S. Q., Maduro, V. V., Summers, T. J., Portnoy, M. E., Dietrich, N. L., Akhter, N., Ayele, K., Benjamin, B., Cariaga, K., Brinkley, C. P., Brooks, S. Y., Granite, S., Guan, X., Gupta, J., Haghighi, P., Ho, S. L., Huang, M. C., Karlins, E., Laric, P. L., Legaspi, R., Lim, M. J., Maduro, Q. L., Masiello, C. A., Mastrian, S. D., McCloskey, J. C., Pearson, R., Stantripop, S., Tiongson, E. E., Tran, J. T., Tsurgeon, C., Vogt, J. L., Walker, M. A., Wetherby, K. D., Wiggins, L. S., Young, A. C., Zhang, L. H., Osoegawa, K., Zhu, B., Zhao, B., Shu, C. L., De Jong, P. J., Lawrence, C. E., Smit, A. F., Chakravarti, A., Haussler, D., Green, P., Miller, W. & Green, E. D. (2003). Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* 424(6950): 788-793.

Thompson, J. D., Plewniak, F. & Poch, O. (1999). A comprehensive comparison of multiple sequence alignment programs. *Nucleic Acids Res* 27(13): 2682-2690.

Towers, R. E., Murgiano, L., Millar, D. S., Glen, E., Topf, A., Jagannathan, V., Drogemuller, C., Goodship, J. A., Clarke, A. J. & Leeb, T. (2013). A nonsense mutation in the IKBKG gene in mares with incontinentia pigmenti. *PLoS One* 8(12).

Tyynelä, J., Palmer, D. N., Baumann, M. & Haltia, M. (1993). Storage of saposins A and D in infantile neuronal ceroid-lipofuscinosis. *FEBS Lett* 330(1): 8-12.

Tyynelä, J., Sohar, I., Sleat, D. E., Gin, R. M., Donnelly, R. J., Baumann, M., Haltia, M. & Lobel, P. (2000). A mutation in the ovine cathepsin D gene causes a congenital lysosomal storage disease with profound neurodegeneration. *EMBO J* 19(12): 2786-2792.

Tyynelä, J., Sohar, I., Sleat, D. E., Gin, R. M., Donnelly, R. J., Baumann, M., Haltia, M. & Lobel, P. (2001). Congenital ovine neuronal ceroid lipofuscinosis--a cathepsin D deficiency with increased levels of the inactive enzyme. *Eur J Paediatr Neurol* 5 Suppl A: 43-45.

Ulahannan, D., Kovac, M. B., Mulholland, P. J., Cazier, J. B. & Tomlinson, I. (2013). Technical and implementation issues in using next-generation sequencing of cancers in clinical practice. *Br J Cancer* 109(4): 827-835.

Umemura, T., Sato, H., Goryo, M. & Itakura, C. (1985). Generalized lipofuscinosis in a dog. *Nihon Juigaku Zasshi* 47(4): 673-677.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M. & Rozen, S. G. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Res* 40(15): 22.

Url, A., Bauder, B., Thalhammer, J., Nowotny, N., Kolodziejek, J., Herout, N., Furst, S. & Weissenbock, H. (2001). Equine neuronal ceroid lipofuscinosis. *Acta Neuropathol* 101(4): 410-414.

Uryu, N., Maeda, M., Ota, M., Tsuji, K. & Inoko, H. (1990). A simple and rapid method for HLA-DRB and -DQB typing by digestion of PCR-amplified DNA with allele specific restriction endonucleases. *Tissue Antigens* 35(1): 20-31.

Uvebrant, P. & Hagberg, B. (1997). Neuronal ceroid lipofuscinoses in Scandinavia. Epidemiology and clinical pictures. *Neuropediatrics* 28(1): 6-8.

Vaiman, D., Billault, A., Tabet-Aoul, K., Schibler, L., Vilette, D., Oustry-Vaiman, A., Soravito, C. & Cribiu, E. P. (1999). Construction and characterization of a sheep BAC library of three genome equivalents. *Mamm Genome* 10(6): 585-587.

Valayannopoulos, V. & Wijburg, F. A. (2011). Therapy for the mucopolysaccharidoses. *Rheumatology* 49(50).

van der Worp, H. B., Howells, D. W., Sena, E. S., Porritt, M. J., Rewell, S., O'Collins, V. & Macleod, M. R. (2010). Can animal models of disease reliably inform human studies? *PLoS Med* 7(3): 1000245.

van Tintelen, J. P., Hofstra, R. M., Katerberg, H., Rossenbacker, T., Wiesfeld, A. C., du Marchie Sarvaas, G. J., Wilde, A. A., van Langen, I. M., Nannenberg, E. A., van der Kooi, A. J., Kraak, M., van Gelder, I. C., van Veldhuisen, D. J., Vos, Y. & van den Berg, M. P. (2007). High yield of LMNA mutations in patients with dilated cardiomyopathy and/or conduction disease referred to cardiogenetics outpatient clinics. *Am Heart J* 154(6): 1130-1139.

Vandevelde, M. & Fatzer, R. (1980). Neuronal ceroid-lipofuscinosis in older dachshunds. *Vet Pathol* 17(6): 686-692.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigo, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J.,

Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A. & Zhu, X. (2001). The sequence of the human genome. *Science* 291(5507): 1304-1351.

Vesa, J., Hellsten, E., Verkruyse, L. A., Camp, L. A., Rapola, J., Santavuori, P., Hofmann, S. L. & Peltonen, L. (1995). Mutations in the palmitoyl protein thioesterase gene causing infantile neuronal ceroid lipofuscinosis. *Nature* 376(6541): 584-587.

Vesa, J., Chin, M. H., Oelgeschlager, K., Isosomppi, J., DellAngelica, E. C., Jalanko, A. & Peltonen, L. (2002). Neuronal ceroid lipofuscinoses are connected at molecular level: interaction of CLN5 protein with CLN2 and CLN3. *Mol Biol Cell* 13(7): 2410-2420.

Vucicevic, D., Corradin, O., Ntini, E., Scacheri, P. C. & Orom, U. A. (2015). Long ncRNA expression associates with tissue-specific enhancers. *Cell Cycle* 14(2): 253-260.

Vuillemenot, B. R., Katz, M. L., Coates, J. R., Kennedy, D., Tiger, P., Kanazono, S., Lobel, P., Sohar, I., Xu, S., Cahayag, R., Keve, S., Koren, E., Bunting, S., Tsuruda, L. S. & O'Neill, C. A. (2011). Intrathecal tripeptidyl-peptidase 1 reduces lysosomal storage in a canine model of late infantile neuronal ceroid lipofuscinosis. *Mol Genet Metab* 104(3): 325-337.

Walsh, P. S., Erlich, H. A. & Higuchi, R. (1992). Preferential PCR amplification of alleles: mechanisms and solutions. *PCR Methods Appl* 1(4): 241-250.

Wang, D., Jao, L. E., Zheng, N., Dolan, K., Ivey, J., Zonies, S., Wu, X., Wu, K., Yang, H., Meng, Q., Zhu, Z., Zhang, B., Lin, S. & Burgess, S. M. (2007). Efficient genome-wide mutagenesis of zebrafish genes by retroviral insertions. *Proc Natl Acad Sci U S A* 104(30): 12428-12433.

Warrier, V., Vieira, M. & Mole, S. E. (2013). Genetic basis and phenotypic correlations of the neuronal ceroid lipofusinoses. *Biochim Biophys Acta* 11(30): 28.

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S. E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., Bork, P., Botcherby, M., Bray, N., Brent, M. R., Brown, D. G., Brown, S. D., Bult, C., Burton, J., Butler, J., Campbell, R. D., Carninci, P., Cawley, S., Chiaromonte, F., Chinwalla, A. T., Church, D. M., Clamp, M., Clee, C., Collins, F. S., Cook, L. L., Copley, R. R., Coulson, A., Couronne, O., Cuff, J., Curwen, V., Cutts, T., Daly, M., David, R., Davies, J., Delehaunty, K. D., Deri, J., Dermitzakis, E. T., Dewey, C., Dickens, N. J., Diekhans, M., Dodge, S., Dubchak, I., Dunn, D. M., Eddy, S. R., Elnitski, L., Emes, R. D., Eswara, P., Eyras, E., Felsenfeld, A., Fewell, G. A., Flicek, P., Foley, K., Frankel, W. N., Fulton, L. A., Fulton, R. S., Furey, T. S., Gage, D., Gibbs, R. A., Glusman, G., Gnerre, S., Goldman, N., Goodstadt, L., Grafham, D., Graves, T. A., Green, E. D., Gregory, S., Guigo, R., Guyer, M., Hardison, R. C., Haussler, D., Hayashizaki, Y., Hillier, L. W., Hinrichs, A., Hlavina, W., Holzer, T., Hsu, F., Hua, A., Hubbard, T., Hunt, A., Jackson, I., Jaffe, D. B., Johnson, L. S., Jones, M., Jones, T. A., Joy, A., Kamal, M., Karlsson, E. K., Karolchik, D., Kasprzyk, A., Kawai, J., Keibler, E., Kells, C., Kent, W. J., Kirby, A., Kolbe, D. L., Korf, I., Kucherlapati, R. S., Kulbokas, E. J., Kulp, D., Landers, T., Leger, J. P., Leonard, S., Letunic, I., Levine, R., Li, J., Li, M., Lloyd, C., Lucas, S., Ma, B., Maglott, D. R., Mardis, E. R., Matthews, L., Mauceli, E., Mayer, J. H., McCarthy, M., McCombie, W. R., McLaren, S., McLay, K., McPherson, J. D., Meldrim, J., Meredith, B., Mesirov, J. P., Miller, W., Miner, T. L., Mongin, E., Montgomery, K. T., Morgan, M.,

Mott, R., Mullikin, J. C., Muzny, D. M., Nash, W. E., Nelson, J. O., Nhan, M. N., Nicol, R., Ning, Z., Nusbaum, C., O'Connor, M. J., Okazaki, Y., Oliver, K., Overton-Larty, E., Pachter, L., Parra, G., Pepin, K. H., Peterson, J., Pevzner, P., Plumb, R., Pohl, C. S., Poliakov, A., Ponce, T. C., Ponting, C. P., Potter, S., Quail, M., Reymond, A., Roe, B. A., Roskin, K. M., Rubin, E. M., Rust, A. G., Santos, R., Sapojnikov, V., Schultz, B., Schultz, J., Schwartz, M. S., Schwartz, S., Scott, C., Seaman, S., Searle, S., Sharpe, T., Sheridan, A., Shownkeen, R., Sims, S., Singer, J. B., Slater, G., Smit, A., Smith, D. R., Spencer, B., Stabenau, A., Stange-Thomann, N., Sugnet, C., Suyama, M., Tesler, G., Thompson, J., Torrents, D., Trevaskis, E., Tromp, J., Ucla, C., Ureta-Vidal, A., Vinson, J. P., Von Niederhausern, A. C., Wade, C. M., Wall, M., Weber, R. J., Weiss, R. B., Wendl, M. C., West, A. P., Wetterstrand, K., Wheeler, R., Whelan, S., Wierzbowski, J., Willey, D., Williams, S., Wilson, R. K., Winter, E., Worley, K. C., Wyman, D., Yang, S., Yang, S. P., Zdobnov, E. M., Zody, M. C. & Lander, E. S. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* 420(6915): 520-562.

Weimer, J. M., Custer, A. W., Benedict, J. W., Alexander, N. A., Kingsley, E., Federoff, H. J., Cooper, J. D. & Pearce, D. A. (2006). Visual deficits in a mouse model of Batten disease are the result of optic nerve degeneration and loss of dorsal lateral geniculate thalamic neurons. *Neurobiol Dis* 22(2): 284-293.

Weissenbock, H. & Rossel, C. (1997). Neuronal ceroid-lipofuscinosis in a domestic cat: clinical, morphological and immunohistochemical findings. *J Comp Pathol* 117(1): 17-24.

Westlake, V. J., Jolly, R. D., Jones, B. R., Mellor, D. J., Machon, R., Zanjani, E. D. & Krivit, W. (1995). Hematopoietic cell transplantation in fetal lambs with ceroid-lipofuscinosis. *Am J Med Genet* 57(2): 365-368.

Wheeler, R. B., Sharp, J. D., Schultz, R. A., Joslin, J. M., Williams, R. E. & Mole, S. E. (2002). The gene mutated in variant late-infantile neuronal ceroid lipofuscinosis (CLN6) and in nclf mutant mice encodes a novel predicted transmembrane protein. *Am J Hum Genet* 70(2): 537-542.

Wheeler, D. A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y. J., Makhijani, V., Roth, G. T., Gomes, X., Tartaro, K., Niazi, F., Turcotte, C. L., Irzyk, G. P., Lupski, J. R., Chinault, C., Song, X. Z., Liu, Y., Yuan, Y., Nazareth, L., Qin, X., Muzny, D. M., Margulies, M., Weinstock, G. M., Gibbs, R. A. & Rothberg, J. M. (2008). The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452(7189): 872-876.

White, D. J. & Gemmell, N. J. (2009). *Can indirect tests detect a known recombination event in human mtDNA?* : Mol Biol Evol. 2009 Jul;26(7):1435-9. doi: 10.1093/molbev/msp073. Epub 2009 Apr 15.

Whitfield, T. W., Wang, J., Collins, P. J., Partridge, E. C., Aldred, S. F., Trinklein, N. D., Myers, R. M. & Weng, Z. (2012). Functional analysis of transcription factor binding sites in human promoters. *Genome Biol* 13(9): 2012-2013.

Williams, R. E. & Mole, S. E. (2012). New nomenclature and classification scheme for the neuronal ceroid lipofuscinoses. *Neurology* 79(2): 183-191.

Wingender, E., Dietze, P., Karas, H. & Knuppel, R. (1996). TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res* 24(1): 238-241.

Winship, P. R. (1989). An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. *Nucleic Acids Res* 17(3).

Wöhlke, A., Philipp, U., Bock, P., Beineke, A., Lichtner, P., Meitinger, T. & Distl, O. (2011). A one base pair deletion in the canine ATP13A2 gene causes exon skipping and late-onset neuronal ceroid lipofuscinosis in the Tibetan terrier. *PLoS Genet* 7(10): e1002304.

Wood, P. A., Sisk, D. B., Styer, E. & Baker, H. J. (1987). Animal model: ceroidosis (ceroid-lipofuscinosis) in Australian cattle dogs. *Am J Med Genet* 26(4): 891-898.

Woods, P. R., Walker, M. A., Weir, V. A., Storts, R. W., Menzies, C. & Shelton, M. (1993). Computed tomography of Rambouillet sheep affected with Neruonal Ceroid Lipofuscinosis. *Veterinary Radiology & Ultrasound* 34(4): 259-262.

Woods, P. R., Storts, R. W., Shelton, M. & Menzies, C. (1994). Neuronal ceroid lipofuscinosis in Rambouillet sheep: characterization of the clinical disease. *J Vet Intern Med* 8(5): 370-375.

Worgall, S., Sondhi, D., Hackett, N. R., Kosofsky, B., Kekatpure, M. V., Neyzi, N., Dyke, J. P., Ballon, D., Heier, L., Greenwald, B. M., Christos, P., Mazumdar, M., Souweidane, M. M., Kaplitt, M. G. & Crystal, R. G. (2008). Treatment of late infantile neuronal ceroid lipofuscinosis by CNS administration of a serotype 2 adeno-associated virus expressing CLN2 cDNA. *Hum Gene Ther* 19(5): 463-474.

Wrigstad, A., Nilsson, S. E., Dubielzig, R. & Narfstrom, K. (1995). Neuronal ceroid lipofuscinosis in the Polish Owczarek Nizinny (PON) dog. A retinal study. *Doc Ophthalmol* 91(1): 33-47.

Yang, C., Bolotin, E., Jiang, T., Sladek, F. M. & Martinez, E. (2006). Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* 389(1): 52-65.

Yu, T. W., Chahrour, M. H., Coulter, M. E., Jiralerspong, S., Okamura-Ikeda, K., Ataman, B., Schmitz-Abe, K., Harmin, D. A., Adli, M., Malik, A. N., D'Gama, A. M., Lim, E. T., Sanders, S. J., Mochida, G. H., Partlow, J. N., Sunu, C. M., Felie, J. M., Rodriguez, J., Nasir, R. H., Ware, J., Joseph, R. M., Hill, R. S., Kwan, B. Y., Al-Saffar, M., Mukaddes, N. M., Hashmi, A., Balkhy, S., Gascon, G. G., Hisama, F. M., LeClair, E., Poduri, A., Oner, O., Al-Saad, S., Al-Awadi, S. A., Bastaki, L., Ben-Omran, T., Teebi, A. S., Al-Gazali, L., Eapen, V., Stevens, C. R., Rappaport, L., Gabriel, S. B., Markianos, K., State, M. W., Greenberg, M. E., Taniguchi, H., Braverman, N. E., Morrow, E. M. & Walsh, C. A. (2013). Using whole-exome sequencing to identify inherited causes of autism. *Neuron* 77(2): 259-273.

Zeman, W. & Dyken, P. (1969). Neuronal ceroid-lipofuscinosis (Batten's disease): relationship to amaurotic family idiocy? *Pediatrics* 44(4): 570-583.

Zeman, W. & Rider, J. A. (1975).The dissection of a degenerative disease. . In *Proceedings of four round table conferences on the pathogenesis of Batten's disease (neuronal ceroid-lipofuscinosis). Amsterdam.*, 393.

Zeman, W. (1976). The neuronal ceroid-lipofuseinoses. . *Progress in Neuropathol Appl Neurobiol* 3: 207-223.

Zhang, J., Chiodini, R., Badr, A. & Zhang, G. (2011). The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics* 38(3): 95-109.

Zhao, J., Akinsanmi, I., Arafat, D., Cradick, T. J., Lee, C. M., Banskota, S., Marigorta, U. M., Bao, G. & Gibson, G. (2016). A Burden of Rare Variants Associated with Extremes of Gene Expression in Human Peripheral Blood. *Am J Hum Genet* 98(2): 299-309.

Zhou, X., Ren, L., Li, Y., Zhang, M., Yu, Y. & Yu, J. (2010). The next-generation sequencing technology: A technology review and future perspective. *Science China-Life Sciences* 53(1): 44 - 57.

Zhu, K., Wang, H., Gul, Y., Zhao, Y., Wang, W., Liu, S. & Wang, M. (2012). Expression characterization and the promoter activity analysis of zebrafish hdac4. *Fish Physiol Biochem* 38(2): 585-593.

Z-Nagy, I. (1988). *Lipofuscin: State of the Art.* Elsevier Science & Technology

Zon, L. I. & Peterson, R. T. (2005). In vivo drug discovery in the zebrafish. *Nat Rev Drug Discov* 4(1): 35-44.

Zuniga, A., Michos, O., Spitz, F., Haramis, A. P., Panman, L., Galli, A., Vintersten, K., Klasen, C., Mansfield, W., Kuc, S., Duboule, D., Dono, R. & Zeller, R. (2004). Mouse limb deformity mutations disrupt a global control region within the large regulatory landscape required for Gremlin expression. *Genes Dev* 18(13): 1553-1564.

APPENDICES

Appendix 1

Unpublished genomic sequence for Ovine *CLN6*.

This genomic sequence was provided by a collaborator (N. Mitchell, Lincoln University, pers. comm.) and used to supplement the ovine 'Ovis aries' *CLN6* mRNA sequence (GenBank: DQ458790.1) for cross-species sequence analysis.

>oCLN6_12601bp

ATGGAGGCTGCGGCGCGGAGGCGGCACCCGGGAGCGGCGGGCGGCTCCGGCGCGCAGCCGGGCGCCTCCTTCCTGCA
GGCCAGGTGGGCGCGCGCCGGCGGGCGGAGCGGAACCGCGGCCCCGCCACGTGATCGGGCGGGCGGGGGGTCCGGGC
CGCGGGCCGTGGACCCCGGACCCCGGGTGGGTCTTCCCAGCGCTACGCCTGCAGCTGTTTCCTCGCCCACTCGCTTG
AGCCACGCCTCCGCGGCTTCCTGGGAGAGCCCCGGGACCCGGGAGTCTCAGCCACGGCGATCAGCGGCCCTCATGGC
CCGGCTGGCTCCGGGACGATCGCTCATCTGGGATCCAAGAAACTCGGAATGGGGGCGGTGGCGGCAAGCCCATACGA
CACTGCCGCTCAAATAGGGCCCCCAAAGCGCCTGCAAGTCCCGGGCACCAACCCAAGTTCGCATAGTCCTTCTAACC
ATGACTAGAATTTGGATTCTGGTGTTAACGGGAGGCTCACTCCTTGCTCCRAGGACCCAWAATGTTGGGAATGCTAM
AAAAGACAACTAAATGTGGAAACTCCCTCAAGAGAAAAACCAGGACGGGCTTTACCGACACTGTTGCTGTTCCTTTC
TCCAGCAAGGTAAAAGTACTGATTTATGTTTTGATCTTGAGGCCGAAACTTGGGGCGAATAAAAGGACAAAAGCTGC
CTTTAATCTTCCACTTCACTCGGTCTCATTCCATGTAACAGGCACCAACATCAATGCAAAGAAAAGAGATTTGTAAA
CATTCCCGTGGGTCCCAGCGTTGAAACAGAGTGGGTAAACTGAGCCCGGAGGGGCCCCAATCCCTAGGG*RGGGGGGG*
TTAAAGAGTTYCMCCCKGGTATTTKKTGKGGGGGCTTTTAATTKGAGGATGGGTGGGCTTGGGTGAMCACTTCKTTA
NTCCTGGCMCMCCARKAAARTGTGGGCAMCTKGGGCMGGAGGGGGGKGGGATTAGGGAAAGGANTTTAGGNTTGACC
TTTGGNTCAAACCCCACTTCCACACCCCCTTTCCATTCTGCCACCAGGAGGGATAATGRGATTCCCCCAACAGGCTT
GSCTTCCCTGAGCTTTTGAGTTGCTCTAGCAACCCCARGTGGGGACCATTCKTTKGGAATTTGGGCTGAAAAAGGAT
TCCCCAGAAAAATTACCGAAAGCAATAGCAGTGCTKACAGAGGGAAACCTTCATGGCACCTCTAAGCTAATGTTTGA
TTAGCCGCTGAAACGAACAGGTAARRWWSMMWRACTAAGTCACTTCAGTCATGTCTGATTTTTTGCCACCTCATGGA
CTGTATAATCCACCAGGCTCCTCCGTCCATGGGATCTCCCAGAATACTGGAGTGGGTTACCGTTTCCTTCTGCAGGG
GATCTTCCTCACCCAGGGATCAAACTCACATCTCGTCTTGGCAGGCAGATTYTTTACCACTGAGCCACCAGGGAAAC
CTATGACTTTAAGACTAGAGGTTAACAGCAGAAGCTCTGAAACGTAATCATTTGAGGCAACCTGCTCCCGTATTCTT
GCCTGGAGAATCCCCATGGACAGAGGAGCCTGGTGGGCTGCAGTCCACGGGGTCGCACAGAGTCAGACACAACCGCT
TACGCACAGCACAACAGCACTTAACGTCTTGGGCTCAGTTGTTTCACTTGTAAAATGAGGGTAATAGTGTGACCTCA
GAGGGTTTCTGTGAGGATTAAGAGACTTAAATGAGATGGTACATGTAAACAGCTTGGAGGAGTGCCTGGAATTGTAG
AAAGTGTTCGGTAAAGCCTAGCTAGCATCATCCATCGCCATGTTTCATGCCTTTCAGCCACGTTTGCATCCTCTGCT
CCATCCGATCTCCCCGTCACCTTATAGACTTATAGGTGAGAGCCCAGTGTGACTGTATCTGTTGAGTATTTTCCTTC
TTGAATCCTAGAGAAGCAACTGTCAGGGCAGACGGAGTCCATCTCCGATCCCTTCCCCCAGGCCCCAGTACTTCGAG
CCAAACCTTTGCCTAGAACGTTAACCATGCTGCTGTTTCCTCCTGCCCACAAATTCCTGTGGCTTGTCCTGGTGGGG
*AATCTTCCCATGGAGGAGAGGAGGGCTGACTGCTTTCGTTTACTTCTTTCCCAG*GCACAGCTCCGTCAAGGCTGACG
AGGCTGCTGGCACGGCTCCCTTCCACCTTGACCTCTGGTTCTACTTCACTCTGCAGAACTGGGTTCTGGACTTTGGC
CGCCCCATAGCCATG*GTGAGCATGAAGCTGTGGATGGAACCCCCCAACTGGCAGTGTGGAGACAGTGTGGTCTGAAT*
GATCAGAGTGACCTCAGACAAGCCCTTGCCCCTACTGGTACACTGTGAGAATGCGGTCCCTTAGTGCTGCGGCCGTA
CCCAGTGTGGCAGCTGCCAGAGAGTGGGGGATCCGAACAGTCGTACTGTGGTGTGACCCCACCACCTGCCCCAGTAT
CTCGCATTCAGGTTGAACCAAAGCACCAATTCAGACAGTGGAGTATGGAAAAGGATCCTACCAGTCTTGCCCACAGC
CTTGGATAGAATTGTGTTTTCTGGAGGTGAAGAGTATGGGCTGAGTGTGCGTGTCATATGTGCCTTGAGCCTTGGGT
AATTTTTAGTGTCGTCTCGTGAAACCGTCCCCTTCTCTCCAAGGCCACCCCATGCCTCAGCGTACAGTGGTGGCCTC
TGTGGAGATTAACACTGGAAGAGCACCAAAGCCCCTCACTTGATGAGCAGCTAGACAGCTAGATGGATAGACAGATA
GGTAACAGAAGGTGAGGTATCACCATGTACTGAGGTGTTAATTTTTGCATATGGGTGGTAAAAACAACAATATCCAG
CTCTGAAGCTGACACCTTGTACATGCGAGTGATATTGCAAAGTAGCAAAGTCCTTTGGGAAAGCAGATGGTCAAGGT

GATAACTGGTACTTAGATAGTGCCTAGCAAGATATATTCCTGTCCTAAAGGAATAACAATAATTACAGTAAGCTAGA
AGTAGCAGTTAAGATCTTTATGTGTTTAAACTTCTTAATTCTCCCATCAGTGCTATTGTTAATCTCGGTTTTGTAGA
TGAGAAGACTTTGGACAGAGAGGGTAAGTAGCTGGCTCGGGGTCACCCAGTCCGCAAGTGGAGGAGCCAGCATCTGC
CACCCCCACCCAGGCACTCTGGCCCCAGAGCCCACCCCCCATTTCCATCATGATTGTATTCAGAACCCTAATATGTT
CATCCCGCCTGGCTCAGTACGACCCTTTCCTAGAATCTGCCCTGAAGATGTAATCACAGGCGAGAGCCTGTGCCAAA
GGTGTTCACTGCAGTGTTGCTCACAATAGCCAAAACCTGGAAGGGACTTGAGTGTCCAGCAGTAGGGACTGGGTAAG
TCAGCTTTGCTCTATGCAGGATGGAGGATGCTGTCTGGGCAGTGAACAGGAATTGTGGGTACTATAGAGCATTCTGG
AAGGTTTAAGGAGAAAACAGATGCCTGGGTGGCATGCACATGAGCTGCCTGCCAGTCAGTAACGGCCCAAACGATCT
TATATACGTGGGATCCAAGACTTCACGGGGATGAAGAAAAGCTGCTGGCAGTGGTTTTGGGGAGGAGGTATGGTGGA
TGCTGAGTCTCCTTTCTGTAGGCTTTTGGTAAGGTTTACAAGCTGTATTTTAAATGTAAGTGAAGAAAAGGGACTGA
CCCAGGAGTCAGGGCACCCAGTCAGGTTTGTAGGTGATTTGGGGCAAGTCTGTGTCATCTGCGGAGAAGGCAATGGC
AACCCACTCCAATACTCTTGCCTGGAAAATCCCATGGGCAGAGGAGCCTGGTAGGCTGCAGTCCATGGGATCGCTAA
GAGTCGGACACAACTGAGCGACTTCACTTTCACTTTGCACTTTCCTGCATTGGAGAAGGAAATGGCAACCCACTCCA
GTGTTCTTCCCTGGAGAATCTCAGGGACGGGGGAGCCTGGTGGGCTGCCGTTATGGGGTCGCACAGAGTCAGACACG
ACTGAAGCGACTTAGCAGCAGCAGCAGCAGCAGTATCATCTGGGCTTTGGTTCTCTCATCTGTAAGAAGACTGGTTG
GACCAGCTGACCTGAGCCCCCTTGCAGCACAAAGGTTCCAGGGGACAAAGGAGCTGCCCATCAAACTGGTGTTCCAG
GGAGAGAGGAGATGAGGCCAGGGTGGCTCGTGGCTCTCCTCACTCACTAAGACTCAGCCTCATCTGGCCTAATAGGG
CTCCCCGGTAGGGCAGATCACCTCCCGCCATCCTGCTTAGAGTCGCTGTGGCCGCTGTAAGCCTAAGTGGGCTCAGG
AGGAGCATACATCTGAAAAGGTCTAGTGTATACATGAGCTTTTCAGATGTATACATGATAGTCTCAAATTTTTAATA
ACTGCAAAATAAAACCTTTTACAAGTTATAAAAGTAGTATGTCATGGTTCTATAACAAGTAAGTTCATCCCCAGTAA
TATAAAGAAGTGTGAGACTCCCTTTGCAGTCCCACTTCTGACTGTTTGGGGCGCATCCTTCCACTCCTTTCTCTGCT
CATGCAACACATGTGTACATCTCTTTTTTCCCCTCTAACCAAAATAAGATTTTACTTTATACCTTAGTCTATAGCTT
ACTTGTTTTACTTTAGACTGTTTCATGGATAGCTCTCCATGTCAGCACATACGAGGTGTCATCTCTGTTCATGGTTG
CTTAACCTTCTGTAATAAAGAGGTGCTGTAGTGTATTCAAGTAGTCCTGTTTGGTTGAAGAGTTGGCCTATTTCTGA
TTCTTTGCTATCACCGGGTGTTTCGGCCAGCTGTATATCTTGCTGTAGATAAAGCTATAGAGATGCCCACACAGCAA
GTAAGAAAGAGATTGAGACTCCTGTYTCTTGGGAGATTTATTTCTGCAGGTTGGTTTTATCCAGCACAACATTTTGT
AATTAAAAGAAATAAACATTTAAAGTTTTGATTGAAATTTATGATCGTTACAGGTGTTCTCCATCAGGAGATTCTAG
AGGTACAGTTTGGCAGAGCAGGGATTGGAAGTGTTGTGATCTTGTATAGGTGTTATCACACTAACCCCGACAAGGCT
GGACTCACTGGGGTAGACCAGTAGGAAGAACATGGGAAGTCCAGACAGAGGAATGGCCCTGGCCTGTCTCAGGGAAG
GTAGGAAAGGAAGCACTGAAACTGCGTTTAGAACTCACTTATTTTCTCCTTTGTTGCAGAAGGTATAGTTCTGACCT
GATTGGGAAGTGGTCCATCTGCTCAGTGGACAGGGCGTTTTTCACACCTGGGTGGTCAGCATTTCTTGTCATCATTC
TAAGGAGCCAGACCACTAAGGGTCCATGTAGACCATCTGGGCCACATTTCGTGTTACAGATGGGGAAACTGAGGCAG
GAAGAAGGGAAGAACTTGCCCTGGGGTCACCCAGAAGGTTTTTGGCATGCTTGAGTCCCAAAGGATCTTCCAACTCC
CAGGCTGGTGGGTACCAAGGGACCTTTACTTAGTATCCACAGGAAGAGGGACACAGGATCAGACTCTGGGTTTGTAC
CTGCTGTGTGACTGGGGACACGTCTACCTTCTCCGGGACTCAGTTTGTCTGGAAACACCTTCACCTCAGTGCTGCCA
GGAGGATTCCAGGAGATCACAGAAAGAGCACAGTGTGTGCAGAGGCAGTTCTNNNNCTGCGTGGGTCAGGCCTTGTA
AACTGCATTCTCCACTTACTCCCCAGCTGGTGTTCCCTCTCGAGTGGTTTCCGCTCAACAAGCCCAGCGTGGGGGAC
TACTTCCACATGGCCTACAACATCATCACGCCCTTCCTCCTGCTCAAGGTACTACCCCAAGCCCCTCCTCTTCCTCC
CTGCCCTCGTTCCTTTCTCTGCTCTGCCGCTCAGCCCTGGGAGTGCCCTGAATCAGGCCGTCCCCTTGCCTGACTCT
GGGGGCCGAAGGCTATGACAGAGGGGATGTGTCAAGGGCCTGAGGGCCCCAGGGGAGGCAGGGCTGCCTCTGCCTGG
GAGGGATGGAGCAGACTTCCCCAGGGAGGGGCATGTGATTTGGGACTCCACCTGATGTGCCTGGGGGTGAACTGGGA
GAGACCTTCAKGCGGGAGTATCTGAAAAGAAGGCTGGGCTTGAGGAGCCGTGTTTGGGGAGGCTGAGGCGGAAGGAG
TGGCTGGCTTGCTCAGGAAGAGGTCAGGCACCAGTYGGGCAGGCAGGTGGSGGCCGGASGAAGGAAGGGCATCAATG
GAGGGTGAACTCTCTCCTTCGGACCGAGAGCCCTGACTTTGAGGCGCCACTGGGAACAGAGAGATGTCTCCTAGAGG
GTTCTCTGATCTGCTGGCTTTGCATTTTTGCCAAGTAAAGGCTCTTAANNCTACTGCTATTTACTATCCCCATCCGT
TGMTGTTTTTAGAGACATTTCCGAACTTCCTTGGTGGCTCAGACGGTAAAGCGTCTGCCTATAATGCGGGAGACCCG
GTTCAGTCCCTGGGTCAGGAAGATCTCCTGGAGAAGGAAATGGCAACCCACTCCAGTATTCTTGCCTGGAAAATCCC
ATGGATGGAAGAGCCTGKTAGGCTACAGTCCATGGGGTCATAAAGAGTCGGACTCAGCTGAGTGACTTCATTTTCTT
TCTTTCAAAGTAAGGAAGGTGCAGTCTCTGAATTTGAAGGGGTCTTTAGTTAAACAACTTAGTTTCAGAAACACCTG
CCCCACATTTTTGTTTGCATGGTTCAGTTGCTCAGAAATGGGAGCAGCCCAGCCTCAACAGTCCAGCCCAGTCACCC
AGAAGTAGAATTAGGGGTGGTGGGAGGTCAGGGCCTGGCTTGCCTCTATGGAAGGCCTAGCTGTCATCTCCAGCACA
GTATATGGGTGTGTTTGACCTTGAGCAATGAGCTGGCCCCGCGGAGTGTGTTCCTTCTTGCGCCAAAGAGGGTGCTC
TTCCCCAGCACCTGCTGGGTTCACAGGGATCGTGTAAGATGATGGAGGCTGGACTGACTCCCCTAGAGGGAACCCGT
GTGAGAGAGGCTTTCCGGCTGGTGGCATCCATCCCACTCCAGGTGTGAGCGGGGCCAGCAGAGGGCGTCTCCGGTGG
TGATGGTGTCAGGGGTCATGCTGTGGGGCCTCCCTCTGACCTCAGCTCATCGAGCGGTGCCCCCGCACCTTGCCACG
CTCCCTGATCTACGTTAGCATCATCACCTTCATCATGGGGGCCAGCATCCACCTGGTGGGCGACTCAGTGAACCACC
GCCTGATCTTCAGTGGCTACCAGAACCATCTGTCGGTCCGTGAGAACCCCATCATCAAGAATCTCAAGCCAGAGACG

CTGGTGAGCCCCCACCTCCGCTGTGCGGTCTGTCCTAGAGGCGAAAGACATTGCCGACCCATCACAGGTGCCCCAGA
GAGAGCAGGTCGCCCCCCCAACCCACCCTCAACCCCCAGCCAGGATGCCCCCACCCAGATCCCCTCCTCCCCATAGG
GAGGGACTGGGGGCAGTTAAGGGAGGCATGGGCAGGCAGGGGGAGGGTAGCCGGGGGCTGCGCGGCCCGACACCCTT
CCTGAAGCCCCACCCTCTCTGCCCCCCGCCTGGCGTCTGCTGCCCCCCAGATCGACTCCTTTGAGCTGCTCTACTAC
TACGATGAGTACCTGGGCCACTCCATGTGGTGAGTGATGGGGCCCATGGCCCCGGCCAGCTCTGCCGCTGGCCATAG
TGACGCTGAACAAGGGGCCCTGGGGTCTCCAAGTGTCCAGCCCAAACGCTGGAATGGCGTGCTCTTCCCCAGCAGGG
GGCGCTGTCGCTCCTCTGGCTCCCGCCTCGCCGCCTTCCGTGCCCACCGGCCCGCAGTGTCCGCGGCGGCCACTARA
GGGCAGCGAGCGCCTGTCCTCCCAGGGCCGGGGCTGCAGCGCTGGTAAACAACTCCCCCCACCCTGGGGATGAKGAT
GAGGACCCGGAGGCAGGGAGGGAGCTGASGAGGGCCGGCTGAACCAGGCTGGGGGCGASGTGGGAGTASGSSTGCTC
TTAGTGCCGCCTCCTCCTCCGAAMTCCCCTTGGRGSTCCTTCCTCCCACCCCTTCAMCCCACASCGGAGGAAACCTC
ACAGATGTGGGCTCAGCCACAGGAGATRCGGGCCGAGMTGCAGGCAGTSGTGTGTTCCAGAGCCCACTTTATCCCAG
GATGAGCTGGGGCCGATGGGAGGGGCTTTGCCCATGGGCTGACAGMTAGTGCCAGAGCCACCCCTGGACATGGTGGC
CCCACTCCCGAGTGAGRTGGGSTCACTCGTGTCACGCGGGCCCYACATWGTCCTTCTGGGTTCTCTTTTGGGGCGTA
GAAAGTCAGGGTGTCYTTGCACACACAGRCCTCTTGCCAGTTCTCCCTGGCCTGTGCCAGGAGACCCTTCCCATGGA
AGASTGGGGAAGGGAGGAGGAGAGATGWACAGGGGCCTGGAGAAATGAAAGGGCCGGACTCGGAGGGAGTGGGGGCC
ACTGAGGGGAAGRTGGGGAAGGAGGTGAGTGGGGGTGGGAACAGAAAGAGGGTGCTGTGCGAGATGGCACAGGCGTG
TGGGGAGCAGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGGAGAGGAGAGAGGGTGGTACTTGGGTGAGCTTTGA
ATGGAAGGGGGTGAGGCTGGGGACTGAGTGTGAGTGAAGGAGCTTCACAGAGCTCCTGGGAGGTCCCCAGCAGGTGC
TGTTCCTCCCCTCACCCCTTTTTTCCCCCCAGACTGGGACCTTGGGCCTCTCCCCCCTCAACCCTGCAGGGTGCACC
CTTTCCCAAGGCAGTGAGGGGCTGGGGGNTGCTGGTCTTGGGGTGAGTGTGTGGGGCAGAGAGAGTGAGAGAGACAT
GCACACACGGGGAGTGAGTGAGGTGGSTGGGACAGTGTGGGAGGAGTTTGGCCCCACGCCCCTGCCCATGGCTSGGA
TCTGCAGGCSTCATCTTCCTTCCTAGACCCCCCGCCCGGGAGAGGCTGGGAGTGCCCTCACCCAAACAAAGTGAGAC
CGCACAGGGTGGTACAGGCAGACAGTCCTAGGACGGTGGGGGGACGGGGGCTGGCCGGCCGGGCGGCCAAGCAGCTG
AGCACCGGCTTCCCTATTGCAGGTACATCCCCTTCTTCCTTATCCTCTTCATGTACTTCAGCGGCTGCTTTACTCCC
ACCAAAGCTGAGAGCTCAATGCCAGGGGCAGCCCTGCTCCTGGTGGTGCCCAGCGGCCTGTACTACTGGTGAGTGGA
TGGCGGACCTGGAGTGGCAGGGAGGGGGCCCCCTAGTGATCCAGTCAGTGACAATATTCTTCCGACCGCAGAGCTGC
TGTGTGAGCCCTGTGAGCCTCCCTGTTGGGGTCCCTGCTCCCTCCTCTTCAGTGGGGACTGGGACAGTGCCTCCTGG
CCGCCACTCAGTGACCAGGAAGGCGGGCCTGAGTGTGGCCTCTCTACACTCAGAGAGTAACCTCAGGGCCCATAGAA
TCCACATCGAAGCATAATTTCTGTGTGTGCTTAGTTGCTGAGTCGTTCCAACTCTTTGTGCGCCCATGGACTGCAGC
CCACCAGGCTCCCCTGTCCATAGGATTCTCCCAGGCAGGAATACTAGGGTGGGTTGCCATTTCCTTCTCCAGGGGAT
TTTCCCAACCCGGGGATCGAACCTGCGTCTCTTGTGTCTTCTGTGTTGCAGGCAGATTCTTCACCCACTGAGCCATC
GGGAACGCCCGGGGAGGGCTGTAGAGCAGAAGTCTGTAGAGCCTGGCAGGGTGGGAGGGGTGGGGGCTGAATGTCCC
CCTCTGTGAAATGAGGATGATAGCAGGCTGTCTCCCTTTTGGGACATGGTGAGAACCCAGAGGGAGAGAGAGGGGCT
CGTAAACCATGAAGCCTCATGGAGACGACAAGGATACTTACGGGGGAATTCCATGGCTTCAGGGTTTCCCACCTGGT
TTCCTCTTCCCATTCCAGAGGACCAGGCCTCAGCCAGGCTTTTCCCTCTCCTSTGTCACTTSTGCTCTGGAGACAGG
CTCCCCAGGGATTGTGGGTGTCTTAGAAAGGGTGGGTGTCCTAGCCCTTCTCTGGGGGGAAGCCTCCAGCCTAATGA
AAAAGGCCTCCTCTCCTGGGGCCTTCTCCCCCACCGCCAGGGGCTGCCCCTGGCCAGCAGTGAGGGTTCATGAAGGT
GAATCCACATGCCCACTGAGGACAGGATTAGAAGGAATGGGTGACGACCACAGCTGGAGGGAAAGAGGTGAGAAGTG
AGAAGAAGCTTAGCCCATTCTTACCCATGGTGTGTCAATGCGGTGACAGCTACAGGGGGCCATTCACCGAACAATGT
GCCTTTACAAGTATAGAGTTCTCATCAGTAACCTCGAGAGCTGGGGCCATTTATGCCCCCTTTATAAGTAAGAAAAC
AAAGGCTCCAAGATGCCATGTAAATTCCCAAGGTGACTCAGCCACTCAGTTGCAGAGCTATGCAGTTGCTTGAGTGG
GATGGGTCAGCATGGAGGGCAGGGGAGTGGATGAGATAACCTCTGGGAAGCCCTCCAGCAAGTGTTACTGCATTCCT
GGCCATATGCCAGCTACTGCCCTGAGAGCTGGAGGGAACCCTGGGGAATGAGACTGACCCAGCCTTCAAGGAGACAG
TCAGGACAACCCCATACACTGAAACTGCGTCCCAAACTCAGATGCCTTTCGGAACTAGACAAGAGATAGAAATGCAT
CAAGTAGGCATCAGGTGAGGCAGGAGGGAGTGGTGGGGACCGTGGCAAGCTGCGGGGCCATGCTGTCTCTGAGCAGC
TGCTTTCTCCTCGTCAGTCATTGTCATGTGGGAATATAAACCTAACGTTAGCAGGGAAGTCAGAACGCCACGTCGGC
AGGTAATCTGTTATTTAAATATTGGCAACAAACTCGACTTTTTGAAAAACAGTAATTCTAGAAGCCAGATCGGAATG
CTGTGTACTGGCCCTGGCTCGGAGCGGAGCCCCTCCCTTGATGCGCTCTTCTGACCGGCTGGGAGGGCAGCCTGGCG
CTTGGTAGAGGGCAGAACGCTGGGCAGGAGGGCCCCCACCCCTCACTGCCCGCCTGCCTCCTCAGGTACCTGGTCAC
CGAGGGCCAGATCTTCATCCTCTTCATCTTCACCTCCTTTGCCATGCTGGCCCTCGTCCTGCACCAGAAGCGGAAGC
GCCTCTTCCTCGACAGCAACGGCCTCTTCCTCTTCTACTCCTTCGCCCTCGCCCTCCTGCTCGTGGCGCTCTGGGTC
GCCTGGCTGTGGAATGACCCCGTCCTCAGGAAGAAGTACCCGGGCGTCATCTATGTCCCCGAGCCCTGGGCCTTCTA
CACCCTCCACGTCAGCAGCCAATACTGAGTCCTGGGCACAGGCCCCTGGTGCTCTGTGGGTGGGAGATGGATGGATG
AGAGTGAGTGTCGGGAGTGTGTGCGTGTGTGCACCCACTGGGGATGGACGGCCGAAGGTGTCTGAGTGTGCGGAGTG
TGTGTGTGTGCACCCACTGGGGATGGACGGCCGAAAGTGTCTGAGTGTGCGGAGTGTGTGCCTGTGTGCACCCACAT
GTGTGTAGCCCGCAGCTGAGAGTGTGGGTGAGAGTACGCCTGGTGCCTGTGGATAGGGGGTGGGGTGGGGGTGGGGT
GTGTGTGTGTGTGTGTGAGGCAGTGTGTCCAGACCAAGAGCTGGCCCCGGCCTTCCCGTGCACGCTGGAGTCGGGCC

TTCTCACTCTCACTGCCCAGGGGCTGTCTCTTCCTCCTGGGTGTCTCCTGGTAGCCGCAGGGGAGAGGCCTGGGTGG
GGCAGAGGTTGGGAGGGAACTTAGCCTGTCCCTACCGCCCGGCTCACCCCTCAGACCACCTGACCCTGAGGCCAGGC
CACAGGGGTTCCGTGAACCTGCTGTGTACTCATTTGCAATGCACCCCCTTCCCTGGGCCATGTTCGCAGCAGACGGG
CCGACGGGTCCCAAGTCGGCTGTGGGCTGGTTTGCCGTCTCCTCTCAGGGCAGCCTTGGGAGGGGTGAATGGTTCCT
CCTTGTCCCGGCCTTCTTCAACCCTGGGAGGGGTATGTGCACCCCGGATCCTCACCCCACTGCAGGCTGCGTGGGGC
TGGGAGCCAGAGCCAGCCCTCCCCCAACTGCCGGCCTGATCTTGGGGTCTGGCAGTGGTCGTTCCCAGCCACCTTCC
CCTCCCGGGAACTCGGTCTCTGTCCCGCCTTCAGGATCGTGGCACCTGCTTCGTTCACCTGCTCAAGGTTTCTGAGT
ACTTGCACTCACTAGAGGCAGACATCGTGGGGCTCTCGTCAGCGTGTCTCAGCCTGAATCCACACCTGCCACCAGCG
GCCCTACTCCCAGGAAAGGGGGCCTTGTTCTGGCAAGATGCCCCCCACCCCAGCACCAGGGCCAGGCACACAGCCTC
GCTTCCTGGCATTGCTGGCCTTCATGCTTGTGAGGGGCTCAGGGGGGCTACCTCGCCTTCCCTGCCCTCCTCTCCTC
TCCTGTGCTGCCAGGGACTCTGCCTCAAATCCTCCTTTCATGTAGGGGTCCCCCAGCCCCGGTCAGTCCCTGTCAGC
TGGCAGCATCAGGGGTGAGCAGTGTGAATAAAGGCAGTGTGGAGCCAAGC

Appendix 2

Conserved non-coding sequence (CNCS) identified using the VISTA program.

This 369 bp human specific sequence was identified to be highly conserved between human, macaque, dog, mouse and rat when analysed using the VISTA multialignment program. This region is located 5' upstream of the *CLN6* and 1,906 bp upstream of the *CLN6* start codon in the human genome.

>CNCS
CGGACAGGGAGAATTCTCATCTTGTTCATTCAGAAAGGCCCAGCCCCCTGCCAGGTTGGACTCCCCTCCCCTCAAGG
GGTAGCAGCGGACTCCAGAAAAAAGGCAGATGTGTTTGTATTTAAGAGATTGGAGGATTAAATGAGCTAGAGAGGCT
CCCTCCACACGGCCTCATCCTCTGATGGCTGGAAGCCCACCTCTGAGTGTTATTAAATCGATGATGACATCATGGGC
ACTGCATGCCCCTCCCTCTGGAGCCCAGCAGGCTTGCAAAGCCTGCAGAGGATAAGAAATTGGGGTCAGAGCGGGAA
GCCTCAATCCTTCTTCTCTGGTGTATACCGAGATCGATGTGTAGATCAGATTCACCAGAAT

Appendix 3

Sheep specific sequences of CNCS regions within and flanking the ovine *CLN6*.

>PCR_5UTRF1/5UTRR4
TCTACCAGAGAAGAAGGATTGAGGCTTCCCACTCTGACCCCAATCTCATTCTCTGCTGCCTTTTTGAGCCTGCTGGG
CTCCGAAGGAAGGGACATACAGTGCCCGNNNNATGTAATTATCATTTAATAATACCCAGAAGTGGGTGGCCAGTCAT
CAGAGGATGAAGTTGTGCAGTGGGGGCCTTTCTGAATGAAC

>DirectSequencing5UTRR4
ACCTCTTCATACCGACAAATCTGCCTCCTGGCTGGGGCTGACAGCTGCCCCTGAGTGGAGGGGAGTCCAAGCTGGCA
GAGAGGCCAGGACTTTCTGAATGAACATAATGAAAACTCCCCTAGTCTGGTGGGATTATCAGCCCCCTCTCCCTTTC
CTGTCCCATCAGGACTTTGAAAACACCATCTGCGTTAGCTCTCAATTTGCATGGAAACGGGGAGTGAGGCTGGGGGT
G

>DirectSequencing5UTRR6
CNCNNAAGGCAGCAGAGAATGAGCTTGGGGTCAGAGTGGGAAGCCTCAATCCTTCTTCTCTGGTGTAGGCAAAGAAC
AATCTGTAGATNAGGAGATCAAACTCAAAATTAAATCTGGGATCCAGTCTGAGTGTTGTTGTTGGCTGACCTCAGAT
ATACTGGCTTTCCCTGGTGCAAAGCAAGGACATCCTCTGCTTGTTGATGACATGGTGAGGAGCAAATGAGAGAAAGG
GAGCTATAAAGCAAATGTGCCCTTTAGTGAGGCCGGGATGGGTGTTGGGGTGTGGCCTGCGGGCTGCTGGAGCCCAG
GATGGAGTTGATGGGAAGTGGCCCAGATGCACGTACCAGCCATTGCCAAGATGGGATATTTGGGCTCTCATCAAAGG
GATAAGGAATAAACCCAACAAAAGTCTCTC
ANGAGAAGAGATAGAAAAGAAAATTNGGNCNGAAGAAGATANTNTGTGGTGGCCCNCCTTCAGGTCCNAAANCCCCA
NTGTTNTACCCGGNGNACCCCCCNAATGCCCNCGGCTNGCNGGGGCCTTTNTTTANNNTAGGTTNNTTNNAGNTTCC
NNTTNNACCCCCCCCCCCGGGGGTTTCCAACAGGCNTNTTTCCAAATGAGTAAACCTGGGNCGNNN

>DirectSequencing5UTRF7
CGGNTGATATCCCCCAGACTAGGGGAGTTTNCATTATGTTCATTCAGAAAGTCCTGGCCTCTCTGCCAGCTTGGAGG
TCCCCTCCACTCAGGGGCAAGNAGTCAGCCCCAGCCAGGAGGCAGATTTGTTGGTATTTAAGAGATTGGAGGATTAA
GTGATCCAGAGAGGCCCCCACTGCACAACTTCATCCTCTGATGACTGGCCACCCACTTCTGGGTATTATTAAATGAT
AATTACATCACAGGGCACTGTATGTCCCTTCCTTCGGAGCCCAGCAGGCTCAAAAAGGCAGCAGAGAATGAGATTGG
GGTCAGAGTGGGNAGCCTCAATCCTTCTTCTCTGGTGTAGGCAAAGAACAATCTGTAGATGGAAGATCAAACTCAAA
ATTAAATCTGGGATCCAGTCTGAGTGTTGTTGTTGGCTGACCTCAGATATACTGGCTTTCCCTGGTGCAAAGCAAGG
ACATCCTCTGCTTGTTGATGACATGGTGAGGAGCAAATGAGAGAAAGGCAGCTATAAAGCAAATGTGCCCCTTTAGT
GANGCCGGGATGGGTGTTGGGGTTGTGGCCTGCGGGCTTGCTGGAGCCCAGGATGGAGTTGATGGGAAGTGGCCCAG
ATGCACGTACCAGCCANTTGCCAAAGATGGNATATTTNGGNCTCTCATCAAAAGGGATAAGAATAAACCCCACCAAA
AGTCTCTCCNNGAGANGNGTTNGNAAAGNAATTNNGGCCGAAGAAGATAACCCNTGGAGGCCCNCCNCAGTNCCAAN
ANAAACTGTTTTTNCNAACAAACCCCCAATNCCNGGNCTATNTAAAGCCTCNAGGNGGGGTTNCANTGAACCCCCNC
CCGGGTTGCCNANGCTNTTNAN

>DirectSequencing_5UTRF9
CTGCTGNNCTCNGANGANGGANATACAGTGCCCTGTGATGTAATTATCATTTAATAATACCCAGAAGTGGGTGGCCA
GTCATCAGAGGATGAAGTTGTGCAGTGGGGGCCTCTCTGGATCACTTAATCCTCCAATCTCTTAAATACCAACAAAT
CTGCCTCCTGGCTGGGGCTGACAGCTGCCCCTGAGTGGAGGGGAGTCCAAGCTGGCAGAGAGGCCAGGACTTTCTGA
ATGAACATAATGAAAACTCCCCTAGTCTGGTGGGATTATCAGCCCCCTCTCCCTTTCCTGTCCCATCAGGACTTTGA
AAACACCATCTGCGTTAGCTCTCAAATTGCATGGAAACGGGGAGTGAGGCTGGGGTGGGGGATAGAGCAAAGCTGGC
ACCAGTTCTTGGGCACTGAGGGGCACACATTTCCAGGGATCTTGCCTACAACTCTCTGGAGCCATGGTTCCTGCCCA
CCANAGAGGCAAATGGTCTCAGCTGCAATTTCAGTGAATAGCATTAAGGAATTTTGCAGCTGGGGAGGTCCTTTGGA
ATGGTCCAGTGGTTTCAACTTTTATTTTAGCANATGTTTTCACTCTTTTCCAACAAGAGCAGTATGAGCTCCTTTTA
ATTTTTACTTATTTTTATTTTGGGGCTGTGCTGCGCGGCATATGGGATCTCAGTTCCCCCAACCAGTGATAGGACCT
GTGCCCCCGTGCATTGNGAGCACAAAGTCTTAATCACCAGTCCACCANGGAAGTCCCATGAGCTCCTTGTTAATGCC
TCACCTAGGACCCTCTCAGCCACACCTGTCTTCCCGGTCTTGCCACTTGTTTCTAGCACAAGCCAATGGCAACCCC

>DirectSequencing_5UTRF11
CCTTGCTTTGCACCNGGGAAAGCCAGTATATCTGAGGTCAGCCAACAACAACACTCAGACTGGATCCCAGATTTAAT
TTTGAGTTTGATCTTCCATCTACAGATTGTTCTTTGCCTACACCAGAGAAGAAGGATTGAGGCTTCCCACTCTGACC
CCAATCTCATTCTCTGCTGCCTTTTTGAGCCTGCTGGGCTCCGAAGGAAGGGACATACAGTGCCCTGTGATGTAATT
ATCATTTAATAATACCCAGAAGTGGGTGGCCAGTCATCANAGGATGAAGTTGTGCAGTGGGGGCCTCTCTGGATCAC
TTAATCCTCCAATCTCTTAAATACCAACAAATCTGCCTCCTGGCTGGGGCTGACAGCTGCCCCTGAGTGGAGGGGAG
TCCAAGCTGGCAGAGAGGCCAGGACTTTCTGAATGAACATAATGAAAACTCCCCTAGTCTGGTGGGATTATCAGCCC
CCTCTCCCTTTCCTGTCCCATCAGGACTTTGAAAACACCATCTGCGTTAGCTCTCAAATTGCATGGAAACGGGGAGT
GAGGCTGGGGTGGGGGATAGAGCAAAGCTGGCACCAGTTCTTGGGCACTGAGGGGCACACATTTCCNGGGATCTTGC
CTACAACTCTCTGGAGCCATGGTTCCTGCCCACCAGAGAGGCAAATGGTCTCAGCTGCAATTTCAGTGAATAGCATT
AAGGAATTTTGCAGCTGGGGAGGTCCTTTGGAATGGTCCAGTGGTTTCAACTTTTATTTTAGCAGATGTTTTCACTC
TTTTCCNACAAGAGCAGTATGAGCTCCTTTTAATTTTTACTTATTTTTATTTTGGGGCTGTGCTGCGCGGCATATGG
GATCTCAGTTCC

>DirectSequencing_5UTRR8
NNNCTGGAATGGGTGGCCATTTCCNCCCCAGGGGANCTTCCCAACNNNNNNNNNGAACCTGCATCTCCCTTCATCTC
CTACATTGGNNNGNNNATTCTTTACCACTGAACCACCTGGGAAGCTCATAGTTAGTCATACAGAGTGAAGTCAATCA
GAAAGAGAAAACGAGATATTGTATTACTAACACATATATGTGGAATCTAGAAAAATGGTGCAGGCGAAACCATTTGC
AGGGCAGGAATAGAGATGCAGACGTAGGGAGGACAGGTGGACACAGTGAGGGAAGGAGAGAGTGGGACGAAGTGGGA
GATTACGATTGACATATATACATTCCCATGTGTAAAACAGGTAACTCGTGGGCATCTGCTATAAAACACAGGAAGCC
CAGCTCAGTGACCTGCGATGATCTAGAGGGGTCAGATGAGGGGGTGGGAGAGAGGTCAAAGAGGTGGGGGATACATA
TGAACATATAGCTGATTCACTTCACTGTACAGCAGAAACTAACACAAAAGTGTAAAGCAATTATATTAAACATTTTT
CAATGAAAATAATTGGAGATAATAACCTCCTTCTGTCTCTTCCCCCAGTGTGGTCACCCTTCAAGCTTCCTTAAGGA
TGACTTCAGCCTGGAGGAGCGTGTCCTACAGGTGACACCAGCTGGTAGAGAAAGCCAGTCTCCATCTCCTGTGGTCC
CTCTTCCCGAGGGGTCTGTCCTCATGCTGTAACCCCTGTCTCCCACGANCAGATGTCNCNTGAATCAGGCCTTTCCT
GACTGNCTTTCTCCCAGCATTACTGCACCTTTCCTGCCCCATTTCTGTCTGCCCTCCTGTGTC
CCAGCTTCTGAGTTTAGAACTCCGTGTGTCATGCCTCCCAGGTGGTGNNTACAGTCGAACAGGGGACTTAAAAGCNN
TCCTGAGAGATGATGAATANTANCTGAANNTANTGNANTCATTTCNNANATGCTTTAATCAANCNATCATGCTNNNN
GNCTGANCTTGCNCANNGATGNNTGTCANTNNTTNNNNNAN

>DirectSequencing_I1R8
GGCTCGAAGTACTGGGGCCTGGGGGAAGGGATCGGAGATGGACTCCGTCTGCCCTGACAGTTGCTTCTCTAGGATTC
AAGAAGGAAAATACTCAACAGATACAATCACACTGGGCTCTCACCTATAAGTCTATAGGTGACGGGGAGATCGGATG
GAGCAGAGGATGCAAACGTGGCTGAAAGGCATGAAACATGGCGATGGATGATGCTAGCTAGGCTTTACCGAACACTT
TCTACATTCCAGGCACTCCTCCAGCTGTTTACATGTACCATCTCATTTAAGTCTCTTAATCCTCACAGAAACCCTCT
GAGGTCACACTATTACCCTCATTTTACAAGTGAAACAACTGAGCCCAAGACGTTAAGTGCTGTTGTGCTGTGCGTAA
GCAGTTGTGTCTGACTCTGTGCGACCCCGTGGACTGCAGCCCACCAGGCTCCTCTGTCCATGGGGATTCTCCAGGCA
AGAATACNGGAGCAGGTTGCCTCAAATGATTACGTTTCANANCTTCTGCTGTTAACCTCTAGTCTTAAAGTCATAGG
TTTCCCTGGTGGCTCANNGGTAAAGAATCTGCCTGCCAAGACNAGATGTGAGTTTGATCCCTGNGTGAGGANNATCC
CCTGCNNANNAANNNGNNNNCCACTCCAGTATTCTGGAGAATCNNNTGGACANANGAGCCTGGNGGANTATNCNNNN
CATGANNNGGNNAANNNTCANACTTGACTGAANTGACTTAGCANNNNCCTACNNGNTCNTTTCANNNNTAATCAACA
TTANCTTANNNNNGCCATGAANGNTTCCNNNCTGNCNNCACTGCTNNTNNNTTNNGTANTTNTCTGGNANCCNNTTT
NNNNNNNTTNNNNNANNNNNCCNCANNNNNTTN

>I1F1/I1R2_F
GGGCGGAGCGCCTGCAGTCCCGGGCACCAACCCAAGTTCTCAGAGTCCTTCTAACCATGACTAGAATTTGGATTCTG
GTGTTAACGGGAGGCTCACTCCTTGCTCCGAGGAACCCAGAATGTTGGGAATGCTAATAGAAGAGACAGCGAAGTGT
GGAAACTCCCCTCCAGAGAAAGAACGAGGAAGGGCTTTACCGACACTGTTGCTGTTCCCTTCTCTCAGCAAGGGAAA
GTTACTGATTTATGTATTGATAGTTGAGGCCAGGCACTTGGGGCGGAGGAAAAGGTACAGAAGCTGCCTTTTAATCT
TCCACTTCACTCTGTCTTCATTCCGTTTTACAGGTCACAGACATTCATTGCATAGAGCAGGCATTCATTAAGCATCT
CTGTGTGCCCCAGCGCTGTACCAGGGGTGGGTAAAGTGAAGCCCGGAGGGCCCCACAGTCCAGTAGGACCGGGGAGG
AAGAGAGGCCAGGAGTTGAAAGTTCGACTTGGAGGGCTTCCTGGAGGAAGTGGCGCTTAAGCTGCATGTCAGAATAC
TCAGTAGGTGACAGAGGAATAAAGGCGATTTCAGCTCCTGGAAAGCTTGGGAAAGGCAAGCAGTGCAGTTTTGCTGG
CACTTTCTGTGTGTACAGTGGCAAGACTGTGAAGGGCCTTTAATGGCAGCAGAACTGGGACCAGAGTGAGAGCGTAG
AGTGCAAAAAAAATTTAAGGGGAAGCTCACTCTTGGGATGCTAGTGCTCAGAAGTCCTCTATGGCCCACCTGTGCCT
CCTTAAATGTGGGTCCTACGCGTCTCACTTGCCTCATCCTAGTTCCTGTCCTAAATGGCCGGTTAAGGGTGCAGCTT

TTGTTCTGTAGGGCACAGGGGGAGCCATGGAAGGGTTTCCAGCACTAGTGTGCTCGGAGTCAGTTTAGGAAGTTTGG
GGTATGATTTGGGAAAAGGAAGGAAT

>I1F3/I1R4_R
NNNNNNNNNNTTTANGAGGCACAGGTGGGCCATAGAGGANTTCTGAGCACTAGCATCCCAAGAGTGAGCTTCCCCTT
AAATTTTTTTTGCACTCTACGCTCTCACTCTGGTCCCAGTTCTGCTGCCATTAAAGGCCCTTCACAGTCTTGCCACT
GTACACACAGAAAGTGCCAGCAAAACTGCACTGCTTGCCTTTCCCAAGCTTTCCAGGAGCTGAAATCGCCTTTATTC
CTCTGTCACCTACTGAGTATTCTGACATGCAGCTTAAGCGCCACTTCCTCCAGGAAGCCCTCCAAGTCGAACTTTCA
ACTCCTGGCCTCTCTTCCTCCCCGGTCCTACTGGACTGTGGGGCCCTCCGGGCTTCACTTTACCCACCCCTGGTACA
GCGCTGGGGCACACAGAGATGCTTAATGAATGCCTGCTCTATGCAATGAATGTCTGTGACCTGTAAAACGGAATGAA
GACAGAGTGAAGTGGAAGATTAAAAGGCAGCTTCTGTACCTTTTCCTCCGCCCCAAGTGCCTGGCCTCAACTATCAA
TACATAAATCAGTAACTTTCCCTTGCTGAGAGAAGGGAACAGCAACAGTGTCGGTAAAGCCCTTCCTCGTTCTTTCT
CTGGAGGGGAGTTTCCACACTTCGCTGTCTCTTCTATTAGCATTCCCAACATTCTGGGTTCCTCGGAGCAAGGAGTG
AGCCTCCCGTTAACACCAGAATCCAAATTCTAGTCATGGTTAGAAGGACTCTGAGAACTTGGGTTGNTGCCCGGGAC
TTGCAGGCGCTTTGGGGGGCCNTATTTNAN

>I1F7/I1R6_R
CCCGTAGACTGATTGCTGTTCCTTCTCTCAGCAGGGAAGTTACTGATTTATGTATTGATAGTTGAGGCCAGGCACTT
GGGGCGGAGGAAAAGGTACAGAAGCTGCCTTTTAATCTTCCACTTCACTCTGTCTTCATTCCGTTTTACAGGTCACA
GACATTCATTGCATAGAGCAGGCATTCATTAAGCATCTCTGTGTGCCCCAGCGCTGTACCAGGGGTGGGTAAAGTGA
AGCCCGGAGGGCCCCACAGTCCAGTAGGACCGGGGAGGAAGAGAGGCCAGGAGTTGAAAGTTCGACTTGGAGGGCTT
CCTGGAGGAAGTGGCGCTTAAGCTGCATGTCAGAATACTCAGTAGGTGACAGAGGAATAAAGGCGATTTCAGCTCCT
GGAAAGCTTGGGAAAGGCAAGCAGTGCAGTTTTGCTGGCACTTTCTGTGTGTACAGTGGCAAGACTGTGAAGGGCCT
TTAATGGCAGCAGAACTGGGACCAGAGTGAGAGCGTAGAGTGCAAAAAAAATTTAAGGGGAAGCTCACTCTTGGGAT
GCTAGTGCTCAGAAGTCCTCTATGGCCCACCTGTGCCTCCTTAAATGTTGGGTCCTACGCGTCTCACTTGCCTACCC
CCTAGTTGAGG

>I1F5/I1R4_F
ACAGTTCAATGTGTAGCTCACTCTTGGGTGCTAGTGCTCAGAAGTCCTCTATGGCCCACCTGTGCCTCCTTAAATGT
TGGGTCCTACGCGTCTCACTTGCCTCACCCTAGTTCCTGTCCTAAATGGCCGGTTAAGGGTGCAGCTTTTGTTCTGT
AGGGCACAGGGGAGCCATGGAAGGTTTTCCAGCACTAGTGTGCTCGGAGTCAGTTTTAGGAAGTTTGTGGTATGATT
TGGGATAAGGGAGGAAATCTCAGAGCCTCCGCTGAGAGTAGCTCTTCCTGCTGCTACAGCACATCTGGGTTCTGGAT
CCAATACAAAGAGC

>I1F5/I1R4_R
GATGACGCAGTCAGGTGGAGGGAGCCGTGCCAGCAGCCTCGTCAGCCTTGACGGAGCTGTGCCTGGGAAAGAAGTAA
ACGAAAGCAGTCAGCCCTCCTCTCCTCCATGGGAAGATTCCCCACCAGGACAAGCCACAGGAATTTGTGGGCAGGAG
GAAACACACATGGTTAAGTTCTAGGCAAAGGTTTGGCTCGAAGTACTGGGGCCTGGGGGAAGGGATCGGAGATGGAC
TCCGTCTGCCCTGACAGTTGCTTCTCTAGGATTCAAGAAGGAAAATACTCAACAGATACAATCTCACTGGGCTCTCA
ACTATAAGTCTATAGGTGACCGGGAGATCGGATGGAGCATACGATACCCACGTGGCTGAAAGGTATGAAACATGGCG
ATGGATGATGCCAGCTCGGCTTTATCTTCCTCTTTCTCCCTCTCAGGGTC

>I1F7/I1R4_R
CCGAGGCAGTGTGCACTGTGGCCGGGAGCCGTGCCAGCAGGCTCGGGCTCTTGTGGGAACTGGGCCTGGTGGGAATG
CCTGGAACTGGGGATGTGACCCTCTCGGGTGTTTTTATCCTCTTTTTCCATGCTCTTCTCTGCAGGGCCGCTTCATT
CACATCTGACTGTCCCTTGAGGCTGGAACTTTGGCCTCAAGCAACCCCGGGCTCTCGGGCTCTCAGGAGGGTGCTCT
TTTCAACTGTAGTTAGAAAAGTCCCACTGGGCCATCCTGGGTTTTTGGCCTTCCCTCTGTCTCAGGGAACATCATCT
TTCACCAGAAGGAGGAGGGAATGGGAAATAGTCTGGCCGAGGAATAATTGATGCCTCTTGTTATTAAGCAAGGCAAC
CTTGGCCAACCCTCAGTTTCCTTTTTCGAAAGATGCAAAGGTACGAGCACCCCGGAGGCTCACCCCCAATCTCTCAT
ATAACCTCC

>I1F7/I1R2_F
CGTTTCTGAAACTGATTGCTGTTCCTTCTCTCAGCAAGGGAAAGTTACTGATTTATGTATTGATAGTTGAGGCCAGG
CACTTGGGGCGGAGGAAAAGGTACAGAAGCTGCCTTTTAATCTTCCACTTCACTCTGTCTTCATTCCGTTTTACAGG
TCACAGACATTCATTGCATAGAGCAGGCATTCATTAAGCATCTCTGTGTGCCCCAGCGCTGTACCAGGGGTGGGTAA
AGTGAAGCCCGGAGGGCCCCACAGTCCAGTAGGACCGGGGAGGAAGAGAGGCCAGGAGTTGAAAGTTCGACATGCAG

GGCTTCCTGGAGGAAGTGGCGCTTACCCTGCTTGTCCTAATCCTCCCCTCGTGCCAAACGACCCTGTCGATTTCTTC
TCTCGGACTACTTGAGC

>I1F7/I1R2_R
CNNNNNNNNNNNNNNTNNNNNCTNNTTTNAGTCTCTTAATCCTCACAGAAACCCTCTGAGGTCACACTATTACCCTC
ATTTTACAAGTGAAACAACTGAGCCCAAGACGTTAAGTGCTGTTGTGCTGTGCGTAAGCAGTTGTGTCTGACTCTGT
GCGACCCCGTGGACTGCAGCCCACCAGGCTCCTCTGTCCATGGGGATTCTCCAGGCAAGAATACGGGAGCAGGTTGC
CTCAAATGATTACGTTTCAGAGCTTCTGCTGTTAACCTCTAGTCTTAAAGTCATAGGTTTCCCTGGTGGCTCAGTGG
TAAAGAATCTGCCTGCCAAGACGAGATGTGAGTTTGATCCCTGGGNGANGAAGATCCCCTGCNNANNGANCNNNNNN
NNNNCTNNANNNTTCTGNAGANNNANNTNNNNNNNNNTANNNTGGTGGATTATACCNCNNNNTGTNTNNNNNNNNN
NNNNCACNCNANNNNNNNNNNN

Appendix 4

Contig sequences generated from ovine BAC clone 270H8 using the 454 pyrosequencing platform.

>NCL_C1
TTGACNCGCCGTGTTTATTTTTTTTTTTTTTTTTTTG*CTAGGTCTGGCCACTCTAGGAGGGGAGGCGGACTGGGA
AGCAGGCAGAAGGAAGGGGAGGGGCAGAAAGTGGAGAGAGGATTGAGAAGTTGGGGGAGGGTGAAAAAGCGGCCCTG
ACTCTCCCTTCTGGGCTGTGGCAGAGTACAAGGAATGCTTCTCCCTGTATGACAAGCAGCAGCGCGGGAAGATTAAA
GCCACGGACCTGCTGATGGTGATGCGGTGCCTGGGGGCCAGCCCAACGCCTGGGGAGGTGCAGCGGCACCTGCAGAC
TCACAGGATAGGTGAGTGGGCCCGGCTGGGCCAGCTGCCACGGAAGGAGGAGTCTGGGTGAGTGGGTGGGCCCTGAG
CCCTATAGCTCTTGTCCCAGGGACGCAGGCAGGCAGGCAGGTCGGGTGTGACCAGCTCAGAGGTTTGAGAGGTCCGT
GTTAGCCAGAGTTCTCTGGCTGGGTGTGGAGGCACCTGGCGAGGTGTGCTCGTAGCAACGGGTGACGCCCCATCCCA
GGGGACCAGGGGCTCTAGACATTTCACCACAGTTGTCTCAAACCTGCAGTCCAGACGGTGCTGAACGATCCCAAGGT
TGTTTTGATGGAAGTGTCCCTTTACTTAGATGTGTCATGTGCTTAGCTCAGCAACACCATGAAGGAGGCGCTATCCT
AATCCAAGTTTTACAGATAAAGATCAAGATTCAGAGCCCAGCTGGTAACTTCCGAGCCCTTAAGCCAATGGTGGAAT
CAGATTGGAACCAGATTCTCATTTCTAAGACTGAGCTCCCCCAAACTGTTCAAAATGGGAATAAAACATAGAGGAAA
ACCCACCC*AAACATTACCAAGTTGTCCTCTGCTGATGTGATTATGGGTGATTCTTCATTCCTCTTGTTTCC*TGTT
TTTGGCGAAGAGAACATATACTAGTTTTACAAG*CAAAAACGAGTTGCTTTTTGTTGGTGGTGGCAAC*TTTTTTTT
TAAAAAG**AGATGGAGGAATGTCAGATGCCAGGATGGCTGTTTGCATCGATAGCTGGACACTCAGGAAGGAAAGAG
GAAGGAACACTTGGTCAACAGCCCAACTGGACCTCCGTTTCCCCATCTCTCAGTCGGGTTCCCACCCTCTGCTGGGG
TGATCCCAAGGCCCCATCAGAGGGCGAGTGCAGCGGCAATGGGAGCTTGTCCTGAGGCTGGGCTGGGGGTGGCCTCT
GAAGGCCCCAAGCCTGGTGACGGGGAGCTCAGCCACCGAGAACCCCTGCTCTCTTCAAAGCGTGCTGGTTTCCTTTG
GTTTCCGTCTCCTTCAGACAGGAAGCTGTGGGCAGGGGCACTTCCGATTTCCTGCAGTTGGAGCCCAGCCTGGGCTG
AGGGCCCCAGGCATGCCTGGAGGAGGGGTGTGGGCTGGGGTGGAGGCTGCAGTGAAGGGAAGGGCAAGTGAGGGAGG
CTTGAGCCCCATCATCCACCTGTCTTTGTAGATGTGTTGCAGCGACTGTCAAAGGGGGAGGATAGACATGTTGGTGC
ACTTCCTCCAGTTCAGGGCAGGAGAGAAGCTTCCAGAAGGGGTACAGAAAGCGGTGCTTCATCCTGTTCTTGTCACT
GCTCTGAGCTTCTGTCTCTTTAGTGGGCAAAATGAGATGTCCAGGCACTGAGGGTTGTTGAAAAGATTAAATGAGGC
ATATCCATGGTGGGAACGCAGCAAGACAGGCTGCAGCCTCTTCCCCATGAACACAGCCACCACCACGGTGCTTCCTG
AGTTGGGTGGCTCTTGAGCTGGCCTTGAAGGCTTGGGCAGGAGGCCAGCTGACATCCCCAAGAGAAGCCTTTTCCTC
CTTGCTCACTCTGCTGTTTGGGGGTTTCCTAGACAGAAACGGAGAGCTGGATTTCTTTACTTTCCTGACCATCATGC
ATATGCAGATAAAACAAGAGGACCCAAAAAAGGAAATT*C*TTTTGG*CCATGTTGATGGCGGACAAGGAGAAGAAA
GGCTACATCATGGCATCTGAATTGCGGTCAAAACT*GATGCAACTGGGGGA*GAAGCTCACCCACAAAGAAGGTAAT
GGTCTGGGTCCCTATTTCCCAACAGTGGGAGGTGACTAAGGGGTT*GGGGGC*ACACCAAAGTCCAAATCATGCCAC
CTCAGTTACGTAGGCATTGCATCTGATGGCATTTGGACGTTGGGACTAGATGCTAAGCTTGGTGAACTCTGGACTGG
AACCTGGAAAACTAGGGAACTGTGAGTGGCAGAAGTTAGCCAGGGTGAAGGGACAGTAGGGAAACCCAGTTAGGAGC
TACTTGGGCATTACTCTATGTCTTCTAACTTTCACTGGGGGCCTTCATGACTGCGACACCTCTAGGATCACCAGTGT
CCCTCTCATACCCATTAAAGAAACAGGTGCAGTGCGCGGCCTGTGACGAGCAGAGCTAGGCGCTAGACACAGGGCTA
GCAGACACCGGGACCCCCACCCTCCGACCTCGCTCAGGCACTCCTCTCACCACAGCTGTCTTCTGAAGTTGCTGCTG
CCACCAAGTTTCTAAATTGAGATGTTACTTTGCTCCCTTTTCTCATGAGTCAGCCAGCTCTGAGCTCTTCCAGAAGG
AATGTCCTGTCTACCTTCCCTGTCTTCTTGCCTCAGAGCCAGAGGA*GGGGCTTGGGCCTTCTGGGAAAGCTGGATG
TTCATGGTGAGGGTGACCCCAATGCTGAGCCCTTTGCAGATGTCCTCCCAGCCGGCCTGACAGGCCCCTGGGTCCAG
CCACAGTGGCCATGCCAGGTTGCCACCCGAGTCCACACT*CCCCAACCCCAGAGGCTGCAATTTCTGGGTAGGCCGC
CGGGCTCACCTGTGATTGGGGCAGGAGGACACGGTCGGTGGACGAAAAAGGCTCCGGTAATGATGGGCCTTTAGCGT
TTGGTGAAATCACAACAGACTTGAATCCTGGCTCCCCCTCCTTCTGGGGGAGCTCAGGTGAGTTTCTTAACCTCTCT
CACCCTCCTCAGTTTCCTCGTTGGGAAATCAGGTTACAATAGGCAATCCCTCAGGACCCTGCAAGAGCTAACATTCT
ATAAAGCAGTCTGCTAAGATTTTATCTGCAGTCACAAAGGCACAGTGC*AAAGGAAGTAAAAATAGCGCTAGAGAAG
GCAGATGGAGCTTCCCCCG*CTCTTCCAGGCTGCCAGCTATCACAATCCCATCTTTCATTGTGGCTGTTTTAACGTG
AACAGGGAGTCACAGGCTATGCTTCCCTCCCGGGCTCCAAGGACACTACTCAGCCTTTTAAGGATCAGGTGGAGTGT
TAGCAGAGGTCAGC*AGGGGCCTTCATTTTTGCGGTTAATGATCCGTCTAAAACAATACCATGGTCCCCAAACTTAA
ACATCTCTCTCATTTGTTTAACCAGTAAGGAAACATTTCAAAGAAACTGGCAGTGTCTGCCAGAGACTGATGTCCTT

TTCCTCTGGCTATGGGAGATAGAAAAGGTAATACTGGAAGTAAGACTCCTTTACCTTCAGTGATAGAAAAGTTGGTT
TACTGCAGAAGCATCCTTGGAGCAGCTGAGACGGCTGTGAGGAGGCC*TCCTTTG*C**AAACCGGTTCTCTGTGGT
TTTTCAGGGAACTGCTATTCTAACCAGAAGGTACCATCAGATTCCAGCTGCCCCTCTTCTGGACAGCTGGAGGCTGC
AGTGACCGGTACTGGACTTGTCCATTGGGACAGTTTTAACCAAGTGGTTGATGGGTAGACACTCCTCCATGAAATAA
GATAAATTACAGTCTCAGAGTTCTACAAGTACTCAGAAGTACTATCAGCTCAAACGCAGTGCATAGCAATTACGATT
TTGAGGAGGATGCTTAAGCAGGAAGTCACAGGATGCTTTTTCACTGTCCAGAGTTGTGTGGGGGCTGTCTGCGGCTG
GCAAAGTTGACACAGAGCAGTGAGAGAGTTTTGTAAGAGGGTGAGGAGGGTCCTGAGAAGCCACCATTGAAAAGTGT
AAACTGAACTGGCTCTTCTTGTCAACTCCTGGGACAAGCTTGTATTTTATCCCAAGAATCCAGAACCAGTAAATG*T
ACG*TTTCTTTTTTTCTGAGCAACGTTAG*AG*AACAA*TGGC*AGTTTTTGTAGGCTTTGAGCTTGGACACAGTTG
AATTTATTTTCTCTTATACTATCCCACCTCCGAAACACTTTCAGAAAGAGATGATAGTGGCAGATACCATTTGTCAC
CAAGCCCCCTCTTGGGGTGGGGGGATGTCATTTTGGAAGGCAGGCCATAGGCTGGGCTAGGGGCCCCAGGATGGAGC
TGCAGGGGTGTCTGTTATGGAGGGCGTCTGAGA*GGGCGGGC*GGTAGAA**GGTAAAATCCATGCCAACACTTTCT
GGAAATACTGGCACAGGAGCCAGATA*TTTTCTGAAG*AAAAAAAAAGGTACAAGGTGGTTTTA*TTGACTGAATCC
TTACCAGTACACCCCCTA*TCCTGGCCCCT*GCAGAAAGCCAGATGAACTCAGAATTGATTCCATGGTCTCCAAGGG
GGA*GGGAGCGGGTTGTCC*TTTTCTGTTCACTGGTTACTTCCTCACCTATGGAAGCAGTCAACTGGAGGTTCAGGG
AAGACCTCTGGGGAGGCTCAACCCCCAGCTGGATGCCTTATCATATGACACTGAAGCCAAACTGAGAAATTACTGTG
ATTTATCGCAGTGGAGGATCTCTTCAGGGAAGCAGATATCGAACCGAATGGCAAAGTGAAGTATGACGAGTTCATCC
AGAAGCTCACCATTCCTGTGCGGGACTATTGAACGAGGAGGCGAGAGAGCCGCCCCTGGGCCTGAGCACTCAGACTG
TGGGTTTTGGAAAAGAAAGTGTTTCTTACACTTGCGGGAGATGGCAAACACAGCAAGATGACAAACCTAAGCCCAGC
AAGAGGACAAGGAAACAAAATTATTCCAGCCACGGTATTAGCATGTCTTTAATAAAACTTTGTACTGTTTTAACAAC
TACCATTAGTCTAGCCCTTTATAGCCTGTGGTTTAGAGATGGGCCAGGGGTCTAGTCCCAACACTGCCATTTCCCAG
CCAGGTGATTCTGGGCAAATCATTTCACCCCTCAAAGTGAGGGATCCTCACTAAAGTGGATTCTTGCTTCTTACCTC
ACAGAGCCTATTGAGAATGGGCTCTGCTTTCGCAAACTGTAGGACACTATGTAATCCTGGGTTGCCATGATTGTGCT
TTTCCTAAAAAGGGCAATTACAGGAAGACCAATATTCTGTATCTGGGTGCTGGAACCATCTTCCACATTTTCAAGAT
AAAGCTAGCTGGGAAACAGCAACAAGCAAAGAGCTCAAACAAACTATGGCAACATTTCTTGCAGCACCATGGGGTCA
TGTGTAAATTCTCATTCTCAGTGCATACTAATGTAGTCCTGGATATTCATTCATAAGCAGAGTATACAACAGGTTGG
AACAAGGCAGCATCAGCCAAATGTTGGGAAGAAGGACCTGACTACAGCCGACCAAAGGGGACCGGGTGTGACAAGCT
GCATTCACTGCAGCCTTGCAGACCCCGAGAAGACTCATGGTGGTGCTTCTCTCAGCATTTTGAGTTCTTCCCTGGCC
CATTATCAGAGGGGTTCATGGGACTAGGATGGGTTTGCTTAAGTGGGTCAGACTGATGGCTTCAACCCGTTTCCTCA
GGACTAGAGGCAAGTAGTAAGGGGATGGAACAGACCACAGCCGTGACACAGCAGGAGCCCTGC*AAAACTGTTCCA*
GGGGTACTGGTCCTGCCAGGAGCTTTTCTAGCGCAAGTGGCTTTTTAAGTTAGTTATGAGTCAGCACAGTCC*AAAA
AAGGGCTGGAGTCACACATCAGCATGTTGAGAGGGGAGCCAGGCAGAACTCAAAGTACAGAGATTTTCATCCTTATA
TGTATCAAAATTGTGCTCTC*TGGGCTGTGGC*TTTTGCTAATGGTACCTTAGTGGCTAACGCAAAAGAAAACCAAA
TTAAGAGGGAGGTGTGTCTCTGGGGA*GATTTACACAGGACACAAGTTTGAGGGCGAGGCACAAGGGAAATTTCAGG
GGACAAAGCAGTCACAGTCAGTGTAACTATCTCAGTAAGGATGCGTGGCTTCAGACCAAATGGAGGTCCTGAGAGAG
AGTCACTTTTGTTTTCCCAACAGTGTTTAACTGCCCATTAAGTGATCAGGAATGCCTAGGAGCTGGAGAGAAGGGTT
ATGT*GGGGGCAGGCAGTTATGTGCCCCCCACCCACCCACCCTAAACACCAAGTAGGTGAACTTGTGATTGTCTAGC
TGGGGATCTCAGGCTCATACAGTACAAGGGCAGATCCAGGTTTTGTGAGATCTGAGGGTTATACGATTCGGCAGCCT
TCC*AGAATAATACAAGAGAAGTACAGAATTACTAAGGGCTGCTCCCAGGACCCTGGAAGAGGCCAGAGCAAATGA*
GGGGCTCTTGTGCTTAAACTTCATTACCTTCATGGTAAATTCATCTGAGGCTGGTTAG*ATGCTGGCACACCATGGA
GAACTGCTATATTTCACTTTCAGGCAAGGCTCAAGGGTCAAGCTCCACATCAACTGTTAACCCCTCATGGACATGAG
GAAGTCAGTCTTGCATAGCCAACCTGGAATGCCAAGGATTAGACAGTAGATACCTGGGATTTAGATGGACAGGAAAA
GGAAGTTCTGGGCTATAATAGTAAATGAGGGTCCAGTCCCTCAAAGCTTGGCCTTTATACGCTACTGGGTGGTACTC
ATATGGCTTCCAGGCAAGTTTTATGAATCTCATTTGTCTAACAGGAATCTGACGTATAACTTGCCAAAGACACCTAA
GTGGTATGTACTGGCCATTAGAACCTGGGGCTCATGTCCACTGACTCTGTGTTCTAAGGGTGTTCAGTGGGTAGGAT
ACCAAGATCCTGGGATGTGTTCTAATTTAAAAATATACCCCAGATAAGGCTCTGACAAGGACCCCATGCCTGGGTTT
CATGCTCATTCCTAACACTGGTTTACCTAGTGCCTAGTATGATTTGCAAAAAGGCCACAGTGCTCTTAACAGTCAAA
TTGAAAATTAATTAAACAGATTTAAAACT**AAACCAGTGCTGTGGTTCACTACAAAGCTAGCCACGTTTCAGTAAG
AAAGGGGAATGACAGCTGTATATTCATGACTTACCTCAAAGCATGTTGCAAGAGAATAAACTACTTGTCATGCTTTG
AAATCTCTGCATCAAAGGGGCTCTGGAAGCACAAAGCATTGTCACTTTTTCTCACAGGGACTGTCCCCTCGTATACC
CAGCAGTGTACTTTACAGAAGACAAATTAACTGAAGGTTTTTC*TTTTATTACATCTAAAGAGCTCTACATAAACAG
GTAACATTCAAT*AGGTAAAC*AATTTTTTTTTCC*AATGCATGTAATAAAT*ATTTTCACTTGGTACTTTTATACA
AACTGACATTGGTCTACTATACATTTTTAAAAG*CCATTTTACTGGTTTGGCATGCGGTAT*GGAAATTCTAAGAGA
GAAAGTTTTAAGGCAATGAATCACAGATTTAAGTTCATGGAATTTATGGTAACTTTTATATCTGTTTATATACA*TT
TTTCCCCTTTGTT*AATT*AAACAGTTAACAGCAGCACACTCTGGAACCACCAGCTAGTCTCCCTCCCTCTTTCTGA
AATCTAAGCTTTGTGTTTTCTTTTAATT*AAAAAAACAGAATCCAACATGTATTGAAAAAAC*AAAATT*CTTACTA
ACATCTTACTAAAATAATTTCAAATGTGCTTTAAAGTCCTAAAGATCTTGAAGTGTTTTATGTGTTTTAAATTGTTT

AAAAACAATGTTCTTTCTACAATTAGTAATT*AAAATTGGGATATATTTCCATTCAGTTTCAGACATTTGACTCAAA
GAAATATCTGCC*AAGAAATCACATTTCTCAGAAGCTTAAGTATA*TCTTTTCTAAGTCAGTAATAGAGAATTGTGA
TAATGTCTAAATTAACAGTAGAAAAATTTTAATGCTCAAAATAAATTAGTCAACCAAGCGTGGCATGGTTAAGGTCC
AGATAGTTTTAAAAACGTGTTCAGTAATGCCCACAAAGTAATTAAAAGCCACTTCAACCCTGCTAGTCAAACAGCTA
TTATACTAAAGGTGTGTGTGCGTATACACACACACCCACAGTTCTGTACAGAAACACAAACATATGCACACCTATTA
TGAGTGACCAAGGCAAGGGGGA*TGGTGGTAGCTCGACTAAAGATTCTGAGAATCGTCATTTGTCCACAGACAGCAT
GTCAGTGTTAGCACTATGCCAGGGTAGGGGGTCGGGGGCGGGGCGCGTGGTAGTGATAGTTGGCTAGGGAAAAAAGA
AACAACTGTGCTTTAAAAAAACACAAAACTTAGAAGATAAAGATGCAGTGCATTACTCCACTAATAATTACTATGGA
TCATAAAAGT*CTGATTGATTAACATCGGCCATGAGACCAACAGAGTAATAGCTACTACTACCGCAACATTTGAGAA
AGTTAACAACTGAAGGGAACAAACTGCTTAGAGTGGAATGAACTAAAAT*AAAAAG*ACAACTGAACGGTTAGTTGC
TCCTGTGGAGACGACCAAAGCTACTAGACAGACATTTATTTCTCCCTGTTACCTGTTCTTCCCTTAGATTGGTATAA
CGCAAGACACTGAGGGAAAATAACTCTCTCCAAAACT*GGGCTGAAGAAAGTAGAGGGAC*GGGTTTTC*AACTGGA
TCATTCAAACAGTTCGTAATTCAGCATCATACACAAAGTTATTGATCAGAATTCTGAGACCACTAACAACG*ACAAA
AAAAAGAAAAAAACAAAACTATTCATTTTATCTTAGGTAGCTGAATTAATAGAAACATTTCCTATTACAATGTGAAA
TAAGAAGCAACTCCCTACAATTTAAGTAATTTCTATCCCAATCTGAATTAGACACAAATTATAATACTATTAGATAT
TAGCCTAAGTTAACATGGCAATGAGAGTTCTAAAGTGTTTACTCACCATAACACCAGATTCGTTT*AAAAACTTAAC
*TATTTCC*TGTTTTAAC*TG*ACAAT*G*AA*T**TTTTTA*TTGCAAGTTGTTTAGTGACTATAACTCAGCCAT
GTATGTGTGGAATACACATGCGTAGGGAAGACTATATAACTCACTTGTTACTCACACACACTTGTCTATAAATACAT
ATATACACACACACATATACATATCAATACCAACTTGTTTTTCAAAATTCAAGTCAACAATTTGGTGGCAGAGAGAT
GCTGTAAAAACAGGTATTTTTAGATGCCACACAGTCATGTGAGCCATTAATTTTATCATACAATTGAAGACATAGCC
ATCTTGAATTCTAACTGGATTAAAGTAGGTCAATTTGTTACCACTGCAAGGTTTAAAATATGACAATTTTATACACA
AAGGTGGTCCTTCATCAACTCTGAGAACAACTCTGAGAATATTAACAGTTGTGATATAAAAATCCAGTTTGGAAAAC
TGGTAATCATTATGCCCCTTCGTTATCTTTTCTGTTAAACCTGCTAAACCTTTCAATCAGATAAAAAATATCTTTGA
CACCTTCAGA*TATTACCTGAAGAACGTTAATACCTTCTCTGACGAGGATTTTGGAACACAAGTAGAATCACTTTTT
CCTTCCCTCCAACCCTCACTCATTTCCTGAAAAACAGTTTTCAATATATTTACATGTTCCATT*ATTAGTAAGGATA
GATCAGACTGATCCCTGAATTCTTGAAAAAATCTGACTTTCCTATGTCATGCTTGTATTTAAAAAGTGAAGAAAAAA
AATAAACAAGATTACTAGGCTAAAATT*AAGTACAGTTAGAACCAGAAAATAGAAAATAGAATAAGGATTAAGCCCA
TTTTGAATTTCAATTATTTATCAGATGATAGATTTCTACAGACGGACATGTATGGTGCACTCTATTTTCTGAAGAGA
ATTTACTTTTAAGGATAATTGACTGTCTGTCTTTC*TTTTTTGAGGTGATGTAGCCTTACCTCTGTCAGCTGCACAT
GATACCGAACAGGGATTTCTGTGCTATTTTTACTTCCTACTTTGGTATGACGATCTTACAGG*AAAATTTGTTACCC
TGTGGCCAAAAT*AGCAGGTGTTCACAGACTTGGTTTAAA*TACATAATACTGGCAATCCAAATTTAGAATAAAGAA
CTAGACTAAGCCTACTCCTTACCCAGACAGCTTTACTCCAATCTCTGAAACATATTCTCAGGGGGC*ATCTAGCAAT
TAAGAGAGAATATGTGATGGAGAACATGCAATTGGGACCGTTGATAAAGACGCTCATATGTGAATTCAAGACCTTCA
AAGCACTTTGAAAGAAGCAGGCAACTGGTTTATCAACTGAATGCAAACTCGGGAGAAGGAACAGTAACGACAGAAAA
ATACACACAGACACACACACTGAACACACACACCTACAAGCTATGTCCTGAAAATACACAGGGAATGCAGAGCGACA
GATGAGATTTATAACTTAAGGCTGAATAAAATATGCACACAAAGAGAATCCAAGTCTGTCATTTTTAGAAAATGGAA
ATTAAAACTATCCCCCCCCCCCGG*CCCCCTCCCCACAACAAAGAAAAGCCCACAATTTAAAACTTCATACATGCAA
TTCCTTTTCTGTTATTGCAAATAAAATAATGCCATTGAACTAAAATAAAACCAAAACCATAGGACAAACGATGAAAA
AAAATCTTCAGAATCCAGAAAAAATGAACCGAAGGTGGTTTGTGGGTACTTG
>NCL_C2
CTCCTGTTTTTTTTTTTTC*TTTTTTCACTTTACCTGCTCCTAATTCTG*AATGTCTGTAACTAAGTTTGCTTTCTG
CCCTCTAACTCTACAGGAGCTAC*AAGTTAC*ATTTTTTTCCTTTGTTGCTGTTCAGTCGCTCAGTCATGTCTGACT
TTG*C*AACCCCATGGATTGATTGCAGCACGTCAGGCTTCTCTGTCCTTCACCATCTCCTGCAGTTTGTTCAAACTC
ATGTCCATTGAGTGAGTGATGTCATCCAACCATCTCGTCCTCCGTCATCCCCTTTTCCTCCTGCCCTCAATCTTTCC
CAGCATCAAGGTCTTTTCAAGTCAGTTCTTTGCATCAGGTGGCCAAAGTATTGCAGCTTCAGCTTCGGCATCAGTCC
TTCCAATGAATATTCAGGGTTGATTTCCTTTAGGATGGACTGGTTTGATTTCTCAAGAGTCTTCTCCAACACTACAG
TTTGAAAGCATCAGTTCTTTGACACTCAGACTTCTTTAGGGTCCAACTCTCACATCCATACGTGACACCTGG*AAAA
ACCATAGCATTGACTATATGGACC*TTTGTCAGAAAAGT*GAAGTCTCTGCTTTTTAATATAGTGTCTATGTTTGTC
AG*CTTTTCTTCCAAAGAGTAAGCATC*TTTTAATTTCATGACTGCAGTCACCATTGACAGTGGTTTTAGAGCCCAA
G*AAAATAAAATCTGTCACTGTTTTCCCATCTATTTGCCATGAAGTGATGGAACCAGATGCCATGA*TCTTAGTTTT
TTGAATGTTGAG*TTTTGACCAGC*TTTTTCACTCTCCTTTTTCACATTCATCAAAAGGCTCTTTAGTTCTTCTTTG
CTTTCTGCCATAAGGGTGGTGCCATCTGCATATCTGAGGTTGCTGATATTTTCCCTGGCAATCTTGATTCCAGCTTG
TGCTTTATCCAGCCCAACAGCTTGCATGATGTACTCTTCACATAAGTTAAATAAGCAGGGTGACAATGTACAGCCTT
GACATACTCCTTTTCCAATTTTGAACCAGTCTGTTGTTCCACGGTCGATTCTAACCGTTACTTCTTGACCTGCACAC
AGGTTTCTCAGGAGGCACGTAAGGTATTCTGGAGTTATTTTTATATCTTAGTTACTC*TTTTAGCTTAGCTTAGTAA
CTCTTGATATTAAACATACTCTGTTTAAATCACTATGTGGTTTCCATTTTCTGACTGGACAAAGGTGGTACACTTGG
AATACATAGAGTATTTGGTTTCTC*TTTTTCCAAAGAGAGCCTGACTGATAGAGGCTAACCATAGCCCTGGGTCTGT

AGGCATCTTGGCACTGCTGCTGCTGCTGCTGCTGCTAAGTCGCTTCAGTCGTGTCTGACTCTGTGTGACCCCATAGA
TGGCAGCCCACCAGGCTCCTCCGTCCCTGGGATTCTCCAGGC*AAAAACACTGGAGTGGGTTGCCATTTCCTTCTCC
AATGCATGAAAGTGAAAAGT*GAAAGTGAAGTCGCTCAGTCATGTGCGACTCTTCTCGACCCCAAGGACTGTAGCCC
ACCAGGCTCCTCCCTCCATGGGACTTTCCAGGCAAGAGTACTGGAGTGGGGTGCCATTGCCTTCTCAGATCTTGGCA
CTAAGCAGAAACAAATGCATTCTTCTCTGGGGTTTACTTTCCTGCTTAGATCCTTAAGACTCATTTGTTCATTTAAC
AGA*TTTGTTGAGCTCCTGTTATGTATCAGTCATTGTTCTAGATGATGGGGATAAACTAATGAACAGAAGAAAAACC
CTACCTTATATTTTCACAGAAATATACCATCAAATATAGAAGGAAATATGCTACCATGAATAGAAGACAAAATAGAA
CACATTTATCAATAGATCTCTACGG*TTTTTAG*ATATTGCAATTTTCTAGATAGAAGAAGATATAAATAAACTAGA
AGGGAAATGTTTAAAGAAATAAAAGGTGGGGACTTCCTTAGTGGTCYTGGTGGTTAAGAAT*CCACCTGCCAATGTA
GGGGACATGGGTTTGATCCCTGGTTCGGGAAGATTCCACATGCTGCAGGGCAACTAAGGCCACGCACCACCAACTAC
TGAAGCCTGTGAGCTCTAGAGCCTGTGCTCCACAACAATGAGAGG*CTGGCGCACCACAAGTTGAGAGTAGCCCATG
ATAACCACAACTAGAGAGAGGCCACATGAGCAAC*AAG*ACCCA*GCACAGCC*AAAAATAGACAATCCCATTGCTA
GTGGGATTGGAAAATTGTGCATCCCTACGGGAAAACCTTCTGGCAATTTCTCACAGTGTTAAATAGAGCTACTACAT
AACTTAGATATATTGGTATATAACCAAGAGAAAAGAAAATATATATTCACGCAAAAAC*AAGTACATAAATGTTCAT
GGCAGCATTATTCA*TAATACTCCCAAAGTAGT*AACAACCCAAATGTCCATCAACTGGTAAATAAATAAACACAAT
ATGGTAGACCCATATAATGGAATATTATTCAGCCGCAAGAAGGAACAAAGTAATGATATTTGCTAAAATATGGATGA
AAATTGGAAACATTTTGCTAG*ATAAAAAAAAGTT*ATAGAAGGTTATATATATGTTGTATGATTCCATTCATATGA
ATTGTCTCCACTAGGCAAATCCAAAGAGAAAGAAAATAGATTAGTGATTTTGGGGAGACAAGGGTAGAAGGAAACTG
GCAGTTATTGCTAATGAGTATAGAGTTTCTTTTC*AGGGA*GATCGAA*TTTTTCAAGAATTCATGGTAATTATTTC
ACAACTTCTGTGAATGTGAATTGTACACTTTCAAAGGGTGAATCTTACTATTTGCAAATTATATCTCAGTAGAGCTG
TTATTTTTTTAAATGAGCAAGGCAAAGAGACTATTG*AAAATGAAGAGACACATCTG*AAAAGGGAGCAAATAGAAC
TTAACAATTTAGTCATTGAGATT*AAAAACT*CAATAGACAGACTAAACAACAGATTATGCATAGCTGGACGATGAA
TTAATGAACCGGAAGATAGATCTGAGGAAATGACCAGAATGCAGTGCAGAGAGTCAAGGAGATGAGGAATATCAGTG
AGAGGTTAAGGAAAGAACAGATGGAATGAGACGGTCTAAAATATATCCAATCAGAATCCCAG*AAGGAGAGAATAGA
GAGAATAAAGAAGAAGCAATGTTTGAAAAGCTAATGGCTAAGAATTTCCAGATCTGATAAAAGACACCAACACACGG
ACAGGAG*AAAAAGAACATGCATCCAGCAGGTTAAAAAAAGACATACACATGGACATATTGCAGAGAAAGCACAGAG
CATTACAGGCAACATTGTTAAAAGCAGCCAGAGAGAAAAGAAAGACTGCCTACAAAGAAACCACAAATCGGACAGTA
ACAGTAACAGTGAATCTAAAGGACAGAAAGGTATCTTCCAGGTACAGAGAGAAAAT*AACTGCTAACCCAGATAACA
GTAAAATAAACTATTTATTTAAATAAACAGTTAAATAGCAGTTAAAATAACTGTAAAACACTGCTAATGAAAGATGG
TGATGACATGTAATGTGTGGGGTTGAAAAGGACAGGCACAAACACTGGACAATAATAATTTATAGGTGGGGAAAGAG
AGGTAAGCCTGTGCACCACAACTACTG*CCCCCCAGCTCTGGAGCCCTTAAGCCAAAACTACTGAAGCCCACCTGCT
CTAGAGCCTGCCTGTGCTCCGCAACAGAAGAAGCCACGGAGAGCCCACCCACTGCAGCTACAGAATAGCCCCTGCTC
TCCACAGCTAGAGAAAGCCCTCGCACAGCAGCGAAGACCCAGCGCAGCCAATCAAAGAAAAAAGGTGGGGA*GGGGG
GAACC*ACAAGTATAAAAGGGACAGACAGTTAAGAGAGCTCAGATCTAGA*TTTTTTAAAAATCTAGACACACCTAA
AATTAAAAGACAGAGAAAGGATGAACGTAAAATGATGAGCAAGAGC*AAAAGCTCATTAAAAGAAAGCTGGTGTGAT
TATATTGATATTAGATAAAATAGACTATAAAGCAAAAAT*GTATTTA*TCAGTTTAA*TGATACAAGGTTTAATCCA
TCGGGAAAATATAATGGTTCTAAA*GTTGTGGGTGCCTACTAAAATAATTTACAAATACACAAGATAAAATGTACAG
AACTATGGATAGAAATTGACTAATCCACTAACATAGTAGGGGATTTCAATACAGAGCTCTCAATTATTAATGGGTAA
AAAAAAG*ATTGGAATAACACAATCAACCAG*CTTGATTTAGTGGATATTTATAAATCACTCTACATCCAACAATCA
AGGAACAGACATTC*TTTTCCAGTGCACATGGAACATTTCCAAAAATTGACTATATAAAAGGCCATAAAGCCGGTTT
CAACAAATTTCAGAGAATTTGATTCACATAAATTATATTTTATGACCTATGATCATGGATATAGCTTATGCAATTAA
GTTAAAAGTCTTTAACAGAAAGATAAATGACAGCGTCCTCTAAGTTTTGAAAATTGGAAAGTACATTTCTATATAAT
TTATGAGTCAAGTAAGATATAATAGAAATTAAAAGTTGTATAGAAACAACTATATGGTACATGATGAAAATTGTAGA
TATCAAAACTGGTGGGTGTAGCAAAAATGGTAATAACAGGGAAATTTATTGCTTTAGATATTTACAA
>NCL_C3
TACTTAGTC*TATTTTN**TTTTTTTTAAGGAATATAAAATATTTTGTTTATATAACAAGGGACATTTATGATTTAA
AGTTGATAAATG*TTTTACC*AAAATTGCCTAG*AAAAGCATTTCTGTTAAGCCTGTGTCAGCGTGTTCCCCTTTGC
AGCAGTTTTGAGGATTTCATGAAGGAAAACAACTAAAAGAAGCATTAAAAGT*AATGAAATATCCAGATGTTCTGCA
CATGCCAAACTGCGATAGGTGGTTTCCACTCTTCCATTATTTATGGGAAATGATACAACACATAATGACATCGGGAG
GATTTTTTAAAAGGATAGCTGCAGATTAATCT*GAAAAATGTGCT*AACTTAATAATAGTTAATAATTCGTAAAGTT
GAATCCATT*GAAATTGTTGAATTATGCTCGGTGATACATACAAAATAGTAATCTTAAGCAAC*TTTTCAGAGAATA
TTGATGGACTGTTTGCCTTTGAGCTTGGATTCTAAAATGCATAGATATCTCTGTTTTGATGTGATTAATCTAAATTC
CAGCAGTCAACATTTGATTCAGCTCACAGACATGTAACAATTATGATTGCTGTGTTTTCTCCTATGAAACAAGTTGT
CAGATTACAGTTATGCATTCTG*TTTTTGTCCCCTTCTCCCTCTCTTCTTGTGTGGTATCTTTTC*AAATTGAAAAC
TGCTTTTCCAGAGGGCATTATTTGTGCCTCCCTAAGGTTTTTATGACAGGTGAAAAGGAACTGGGA*TA*TTTTTTT
CACTTTCTTCATTTAGTCTGTGATAAGAAGTAGAAAAATTACTAGAGTGGTATCGACTCAATGTTTTTGATGATGAA
AATAATTCTCATTGCCATTTTGAAAGTTATTTCTCTGAGGGCAGCAGATTTTGAAACCAGAAAATGTCACTTATTCC

GACAGGTTTTACAGAGGTCTTTATGGTTTTGGTTTTGTTTTAATGGCAACATTTTGACTGTCAGACTAAAAGAATAT
TCTGTGATTATCTTTAAGCATCACAGAAATTCAAGGGTACATACCTATTTCTTTCAATGTT*AAAAATGATAAGT*A
AAATT*AGCTATATCTGTATTTTTTTCTTCTAGTGCCAAATGAATGCCTTAGCTACTCATAGTGCATGGTACTGTAA
GTGAAGACCTGCAG*C*TTTTTTT*TTTAATGAAAAGCATTATAATGGTGTAG*CAACATCAAATATAAAATTAAAC
AACAAGATTTCAGTTAACATTTTATGTATAGATATTGCTTTG**TGAAGTCTAAGAAAAAGGTTGCAGCTGGATCAG
TATAAAAGAGAATGACCAAGCCAAAATCAGCACCTAGGGCCTTAAAGAAAAGAGAGATACCCCATGAAATGAGATGC
TCTGAAGGATGGGAGGGAATGGAT*GGGGGACTTTAAAAACCTATTCCCCCAAGAAT*GCAGGGTACTTTTACTGTT
GCTGTAGTTATGCTTTGCGTGTCTCCTTGCTTCAGTTAGGAA*TGTTTTGATGGTGAATTCTGCCATAGTAATCTGC
TTCTGTTTTCCTGACCTGGGAATTTTACAAGAAAGGTTTTTAAGATGGTCTAAAGGGAAGTATTTGATCCCTAATTT
GAAAATTACTGGCCTTGATCTTCCTGGTGTGAATCAAGGAGCTAGTGTACCTTGCTCTTGGCAACTCTCATGGCCTG
AAAGAGAAAAGGCCCATCCTCCAGGTGGGTGAATCACACCTTTTTCGCAGCTCTGAACAGACCCTGTATTTTTCAGA
CTTCAAGGCCTGCTTACTTGACAGTCTGAGGAATAGCTGTGATACTTGGTATTTACTTGGTTCCATTATTAATGCAT
ATGA*TTTTTAAACATCAGAACTTAAAATATCAGATAATGTTGCATGTTTTATAAAACATTTCATCCCTTTGGCT*A
CCCAGGATGAGAACTATTTTAGGTATTACGTACTTATTAAATGAGAATAATGATTATCTGTACATATTCAGACAAC*
AAAATATAATGGTACTCAGTGTTTTTG*AGAAGGGGGGACCGGAATGCT*GGGGAGAGTAAAACTGCTGATTGAGAT
GCTGACAAAGTTGACTTTGAAGACAGTGGGATAGATGATGAATTTTGCTAATGGTAACAGA*TTTTTTTTTAAAAC*
TATTCTATAGCAGTACATTGTTCACTATATTGCTAAGGATCTAATTCTGTCTTGTAAAACAAATCCAAATATGTATT
GTGAAACACCTGTATAAAACCATTCTTGAAACAAGTGATCTACATGCCTTG*TTTTTTTTTAAATTTTG*TTTTTGA
TAGAATTATTTTAAAATGTGTTTAGCAAAACCCTTCCCAAACTTAAGTTTCTTTTTTTTTAAAGGG*AAATAAATAG
GTAAGAAAGGTGATAGTTCTGTATCTCAGCTTCTCTGAGAACTCTGGGCTATTGC*TGTTTTTTTAGCC*ACTTTTG
TCATTTTGTGTTTATTATTGGTAGAATTGTGTCTTAGATTATAAATCTTCCTCGATAAATTACAAATTCAAT*CAAA
ACTTATTGCTTAATAATTGTCAT*AAAATTGAATTAGAGACCCAGGTGGATGAATGCCTGGGCAAACCATGTGAAAC
CT*GGGTGGGTTTTGCACTTAAGCATTGGT*AAAATTGATCTCTTGAATTTAGGTAACTGTATAATTCAAGGAATGG
GCCAACAAATCCATCCTTAGACTG*ATAAAGGGGAAATTTTTGGCCTAAGAGCTCTGAAACAACATTGTTTTCTGTC
TACCCTGGATTGCTTTCTCACAGGCCACAGTCCATAAGGTTTTTTGCTGTGTGAACTTCAACCAGCCAATG*AAAAA
GCAAGCTTTGGTCTTGTTTATTTCTGGGAAGGGTTTTATTTCATTACACTGATAGTTGATTAGTAACTTGGTAATTA
TGAGAAACTTAGGTCTGCTAGCACTGTTTGAAGTAAAATGTGATTTGGGGTCAGCAGCTGCATAAGTATCAATTCTG
CATTTGGCAAGTGTTTTTG*AAAATTAGATCCCAAGGGAACAACAGCCTGTCACTAGTCTCCACCTCAGCTGAGGCA
GCTACTTTAGGAATAAGGTTGCTTTGGGTTATTTGAGCAGAAACAGTTCTTTCTCGGGGAAAGGGGCAGGGTGAGGA
GTGAGTTACTGGACCTTGGATTACTTTGCTGGCTTTAGAAGCAGCCAGTGATTGGACTAGAGAGGAGATGGAT*AAA
AGTATAAAGGACTTGCTTTTAAGGTAGATTTTCTGTAAAAACCTTGAGTGTTCCTTTCAGGCTAGCCAGATGTGTCA
ATGTACAGATGACTATTCCATTTTCCTCACTGATTATACTGTAACATTAAAATGGGTTAAAACAAAACTCTGCCTTT
TGTTCTATGAAAAGTTATTAGTCCATAAGA*TTTTAAGCTGTGTGTTTACTCCTGCTCTGAAAATGCACAGCCCTGC
TCTGGATACCCTCTCTGGAAGGTAGAGAGCTCCTAATGTGTGTGGCACATTGCCGGGCTTTCTCATACATTGCTGGC
AGATTTGCCCAAATCTTCAGATGGGCTGGGTAGTACAGTATCTTCTCCAAATGTGGGATCTGAAATAAATACTTGTG
GATAGTGGCGATAGATATATTAA*TTTTTAAAACTGTAAAGTAAATGTCTAGTCCTAGG*TTTGTGGTGTGTTCGTT
TGTGTTAATGTGTAGGGAAGGAACACTGACTTGATGGTTACGAGGAGTGTATCTTGATGTGTGTGCAGGGGTGAGAC
TCGCTAAATTTTTAACAGCTTTTCATTTAGGGGTAAGTCACGTGTTGAGTGGCCTAATC*AAAAAAATGAAGGATCG
AAGATAACAAGCTTTG*CCAAGTGATG*AAC*AAGGTTTTGTATC*TGTTTTTTA*TCAGGTGTTGTAGAATTTGTG
CATGGCCTTTTTTG*TTGTTGCTTAGTAATTGGTAGAGAAGAAAAGATGAGTGAGAAAAATCTTAGCTTTGCTAACC
AGTCTTTTCAGACGTCCATAGCTGAGGAGTAAAATCTGACTGGCTTAACATGGTGAAAGAGCCTATCTTTTATAACA
CTCACCTCATCCTGCAATCCTCCACATAGAAGAATTACGTTGTTGTATTCTAGTAATTCACTGTGATTTATAACAAA
CCAGTGATGTCATCCTGTTGTGCACTTTTGTCAAACCATTTACGTGATTTTAATAAACATAGTAAACTTGCTGACTG
CACCGGAGGTCTGTTAGTGATTTATATATTGCATGGCA*TTTTCTATTTGAGTTTGACATGTAGAATCATTTTTAAT
TTCGTAGTAATCGATAGTCCTAATAGCCCAGCTAATTTGAAACTAACAATATTGCTGTGT*AAAAAGAAAAAAT*GG
TGTTTGTGTTCAGTAAATGTTTG*AAAAAAACT*GCCTTGAAGTGTGTGTCTTTATTGTTCAGTGTTGAGATGATAT
TTTG*TCAAATTCA*TTTTTTTAAAAG*TTGTGAAAGTTGTACAT*GGGGG*AAAAATTCAGATAGTTAAG*AAAAA
TTCAAATAGTAAAAAG*AAAAAAATTTACATCTCTCTTTGCCTCCCATTTGGACTCTTCAGAGGATATTCACTAGTT
TTTTGAGTAGCTTCCCAGAATTTTTCTGTGTGTATATATG*TATTTTTTCTCCTG**CTTTTAAAAATGTAAATGG*
ACTTTTAACTTTCCTACAC*TATTCTGCA*TCTTGTCTATTTTCATTGAGTATATCTTA*GAC*TATTTTTTCATTT
TGGTATGTATAGATTTTTTAAAAAGAACAAATTTATAGCATTTAATTCTACGTATG*AACTTAAAATTTATTTAACC
AGGCCTTCCTGTTGGATTTTTAGGTTTTCAGTTCATTGATATTCTAAATACTGTTTCAGACAGTGAACATCCTTGTG
TACTGTTTTTTCATAGAACAAACTCATTTCTAGCCCTGGAACTTCTGGTTTAGTAACTGAGTGTT*AAAAATCCTGA
CGGGTGAGCCAGTTGCCTTCCATC*TCCCTACATCCTTATTGACATCAGAGGTTATCTCATCATTCAGCTTTTGCTA
AGATGATGGAAATGGAAAATATTACTTCATGGTTTTGATTTGAGTTCCGTTATTATGAATGAGGTTGACATATATTT
ACTGACCTTTTAAATTTGTTTG*C*TTTTGGGCTGTGG*TCTTTTTCTTGCTGTGTGTTAGAAATTGG*AAGTTTTC
TCTCATACATCACAGCTTGTAGTTTGTTTTTAAGTATAGTAG*TTTTTACATGTTCTGAATTTTGTATCTTGCTTGC

AGAGTCTTCTCCATCCCAAGCTCATGCATC*TTTTTTTACCCATGTTCTCTCTTAGTGCTTTTCCCTCCCACGTCTC
TCTCT*CTTCCCCCCTCCCCGCCCCTTCCTCCTCTCCTCTTGTTCTTCGGTTTAAATC*TTTTTATCTACCTGGAAT
TTATTTTGATTAAGGATCCAGCTAGCCAAGAGCTCTTTTATTTTCCACTGATTTGAATCAAAAGTTCAAATGCCCAT
AAAGTAATTGGATCTG*TTTTTGGCTTCTCTGTTCTACTGTGTTGCCCGGGCAGTGGTGGGACCCTGAGGAGTATTG
TCTGGGCTCTGTGACCTCAGGCAAGTTGCTTTTCGATGTTGGCTTCAGTTTCTCTACATTGGGACAATAATATCATC
TAAGTTGTAGTGAGCAAATGGGAGAATTAATATTCTTAACGTAGTGCCTAAAACTGAGGTGCCTGGTAAATGTTAAG
TACTAATCTTCGTGGGGCTGATGTTTCTTCATTTGTCTTTTTTTTCTCTAACCAGAGAAG*ATTTTTTAATTTGGCG
GTCAGCTCTCGGTTGCGGCTTGAGCGATCTTTGTTGTGGCATGCAGGATTTTTCGCTGTGGCACATGGACTTCCTAG
TTGTGTGCGTGCTCCAGAGTGTGTGGGCTTCAGAAGTCAGGGCACTCAAATGCTAGTTTTCACACACCAGCCTAGTT
GCTCTGTGGCATTTGGGATCCTAGATCACCCACCAGGGATTGATCCACGTCCCTTGCATTGCAAGGTGGGTTCTTAA
CCAGTGGAACCACTAGGGAAGTCCCCAAGCCAGAAAATTTTTTAGCAGTTTA*TTTTTAGAAGGTCTTG*AAAATAT
TGTTAG*TTTTTTG*TTTATAATGTAGATAGGAGCCTTTCTTCCATTGTATATACTTCTAAATGTTAATGTTTCGAT
ATAGACATTTTAA*TGGATTTCCCTTTTTAAACATTTATTTCAGTTGCTTCATTTGGATGTCCTGGGTATACTAATC
ATAATTTACAAGTAAATGATAATTTTGCCTTCTCATTTCCAGTATTCATACCTCATATTTGTTTCTTTTACCCAATA
TAGTCATTGGGATTACCAGAAGAATATTAAAGGGTGATATGAGAGCACATTTTCTGCAACTGTTTGCTTTCACATTT
TGATTTAGGGTGTCCCTTTATAGCCGTCTGAAAAACCATTTAGGGCAACTCTCTCACTTTACAGTTGAAGATACAGG
CTCACAAAGAAGTGAGTTGTTCATCCAAGGTCTCGCAAGGACCAGGTGTCAGAGCCTGGGTAGGCTTGGAGCTTCTT
GACCCACTGACCAGTGCTCATTCACTGTAAGACTGACACACCAGGACAGAGAAGCTCACCTGAGGGTCTGCCAGATC
TTTGCTGGGCAGCGGCACCGACCGGACGCAGAAATTAGATTCTAAAGCCTC*TC*TTTTTTTTCCCCCTCTGTGTTT
CAAGTAATGTCTGCTGTGTTAAAGTAACAGGTGCCCAGATAAAAGAAAATGCTGGCAGAAATATATTTCATCAGGTA
CTTGATCTGGAAAACTCCAGTATGAGATAAAAACTTCAGGTGACTCACAAAATCTTGGC*TTTTTG*TTAAGCTTAG
ACAAAACT*ATTTGTTCTAATTTGCCTGTGCTTGGTAAATCTGGTGTGCTCAGTTCATCCAGAGAGCAGGACTGGGG
TTTGCTCTGGCCTTAACCAATGGG*AAAAAACTCTCAGAAATTCGGGCTTATCCAGCATCTACTGTGTGTCAGCACT
GTATTAGGGATGTTATAGGTATAAGACGTATAGGGATGTTCTATACCACATACAAGGACTATGGTCCTTAAATGGGC
C*ATATGGACCTTATGGACAATAAGGACTATGGTCTTGGCCTGTGATAATCTTTTTG*TCCAGCTAGATGCCCCTAT
TTGAAATCTGATTTCC*TATTTTTCCT*C*TAAATTTTAATTTTTTTTAA*TCTGGC*AAC*AAATTTTAATTTTAA
AAGTTTAGGATGTTACTTATGAACTCTGGCTTGGGCCACCTTCTGAAAACATTTCAGAAAACTTTACCATCATCCTT
TTTAGTATAGTGCCAGGGAGATAGGGACGAAGAACGTGCGCAGCTGTTGAGTAAGGATGAGTCCTTAAGTAATGATG
CCCCCAGCTTCTTT*AAAAACATTTTTCTGCTCATACCCTGCTTTGAGCCTCATGTGTTTGCTAGGTACATTTTAGC
CTATATCTTCTTGAACAGTTTCAGAGGCCAAGTGCCAGGCTTCT*CCCCCTCCCCACACCAGCAAAGCGCCTCTGTT
CTGGTGTCAGTTTCTGTTTGGAACACGCTGTCCTACTTTAGCACGATGTCAAAGCTAGTGAAACAGACAGCGACACA
AAAGCACTGCAGATTCAACGTGGCCGATTAAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATG
GTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTA
AAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAAC
CTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGAACCCCTTGCGGCCGCCCGGGCCGTCGACCAATTCTCATG
TTTGACAGCTTATCATCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTTAA
GGGCACCAATAACTGCCTTAAAAAAATT*ACGCCCCG*CCCT*GCCACTCATCGCAGTACTGTTGTAATTCATTAAG
CATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTT
GCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTG
AAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACC
GTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAA
ACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATT
GCCATACGGAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGAT*AAAACTT*GTGCTTA
TTTTTC*TTTACGG*TCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGAC
TGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGA*TTTTTTTC*TCC
ATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACT*CAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGT
GAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA
GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTA*TTCGCGATAAGCTCATGGAGCGG
CGTAACCGTCGCACAGGAAGGACAGAGAAAGCGCGGATCTGGGAAGTGACGGAC**AGAACGGTCAGGACCTGGATT
GGGGAGGCGGTTGCCGCCGCTGCTGCTGACGGTGTGACGTTCTCTGTTCCGGTCACACCACATACGTTCCGCCATTC
CTATGCGATGCACATGCTGTATGCCGGTATACCGCTGAAAGTTCTGCAAAGCCTGATGGGACATAAGTCCATCAGTT
CAACGGAAGTCTACACGAAGGTTTTTGCGCTGGATGTGGCTGCCCGGCACCGGGTGCAGTTTGCGATGCCGGAGTCT
GATGCGGTTGCGATGCTGAAACAATTATCCTGAGAATAAATGCCTTGGCCTTTATATGGAAATGTGGAACTGAGTGG
ATATGCTGTTTTTGTCTGTTAAACAGAGAAGCTGGCTGTTATCCACTGAGAAGCGAACGAAACAGTCGGGAAAATCT
CCCATTATCGTAGAGATCCGCATTATTAATCTCAGGAGCCTGTGTAGCGTTTATAGGAAGTAGTGTTCTGTCATGAT
GCCTGCAAGCGGTAACGAAAACGATTTGAATATGCCTTCAGGAACAATAGAAAT*CTTCGTGCGGTGTTACGTTGAA
GTGGAGCGGATTATGTCAGCAATGGAC*AGAACAACCTAATGAACACAGAACCATGATGTGGTCTGTCCTTTTACAG

CCAGTAGTGCTCGCCGCAGTCGAGCGACAGGGCGAAGCCCTCGAGTGAGCGAGGAAGCACCAGGGAACAGCACTTAT
ATATTCTGCTTACACACGATGCCTGAAAAAACTTCCCTTGGGGTTATCCACTTATCCACGGGGATATTTTTATAATT
A*TTTTTTTATTAGTTTTTAG*ATCTTCTTTTTTAG*AGCGCCTTGTAGGCCTTTATCCATGCTGGTTCTAGAGAAG
GTGTTGTGACAAATTGCCCTTTCAGTGTGACAAATCACCCGT*CAAATGAC*AGTCCTGTCTGTGACAAATT*GCCC
TTAACCCTGTGACAAATTGCCCTCAGAAGAAGC*TG*TTTTTTCACAAAGTTATCCCTGCTTATTGACTCTTTTTTA
*TTTAGTGTGACAATCTAAAAACTTGTCACACTTCACATGGATCTGTCATGGCGGAAACAGCGGTTATCAATCACAA
GAAACGTAAAAATAGCCCGCGAA**TCGTCCAGTCAAACGACCTCACTGAGGCGGCATATAGTCTCTCCCGGGATCA
AAAACGTATGCTGTATCTG*TTCG*TTGACCAGATCAGAAAATCTGATGGCACCCTACAGGAACATGACGGTATCTG
CGAGATCCATGTTGCTAAATATGCTGAAATATTCGGATTGACCTCTGCGGAAGCCAGTAAGGATATACGGCAGGCAT
TGAAGAGTTTCGCGGGGAAGGAAGTGGTTTTTTATCGCCCTGAAGAGGATGCCGGCGATG*AAAAAGGCTATGAATC
TTTTCCTTGGTTTATCAAACGTGCGCACAGTCCATCCAGAGGGCTTTACAGTGTACATATCAACCCATATCTCATTC
CCTTCTTTATCGGGTTACAGAACCGGTTTACGCAGTTTCGGCTTAGTGAAACAAAAGAAATCACCAATCCGTATGCC
ATGCGTTTATACGAATCCCTGTGTCAGTAT*CGTAAGCCGGATGGCTCAGGCATCGTCTCTCTGAAAATCGACTGGA
TCATAGAGCGTTACCAGCTGCCTCAAAGTTACCAGCGTATGCCTGACTTCCGCCGCCGCTTCCTGCAGGTCTGTGTT
AATGAGATCAACAGCAGAACTCCAATGCGCCTCTCATACATTGAGAAAAAGAAAGGCCGCCAGACGACTCATATCGT
ATTTTCCTTCCGCGATATCACTTCCATGACGACAGGATAGTCTGAGGGTTATCTGTCACAGATTTGAGGGTGGTTCG
TCACATTTGTTCTGACCTACTGAGGGTAATTTGTCACAGTTTTGCTGTTTCCTTCAGCCTGCATGGATTTTCTCA*T
ACTTTTTGAACTGTAATTTTTAAGGAAGCCAAATTTGAGGGCAGTTTGTCACAGTTGATTTCCTTCTCTTTCCCTTC
GTCATGTGACCTGATATC*GGGGGTTAGTTCGTCATCATTGATGAGGGTTGATTATCACAGTTTATTACTCTGAATT
GGCTATCCGCGTGTGTACCTCTACCTGGAGTTTTTCCCACGGTGGATATTTCTTCTTGCGCTGAGCGTAAGAGCTAT
CTGACAGAACAGTTCTTCTTTGCTTCCTCGCCAGTTCGCTCGCTATGCTCGGTTACACGGCTGCGGCGAGCGCTAGT
GATAATAAGTGACTGAGGTATG*TGCTCTTCTTATCTCCTTTTGTAGTGTTGCTCTTATTTTAAACAACTTTGTCGG
*TTTTTTGATGACTTTGCGATTTTGTTGTTGCTTTGCAGT*AAATTGCAAGATTTAAT*AAAAAAACG*CAAAGCAA
TGATTAAAGG*ATGTTCAGAATGAAACTCATGGAAACACTTAACCAGTGCATAAACGCTGGTCATGAAATGACGAAG
GCTATCGCCATTGC*ACAGTTTAATGATGACAGCCCGGAAGCGAGGAAAATAACCCGGCGCTGGAGAATAGGTGAAG
CAGCGGATTTAGTTGGGGTTTCTTCTCAGGCTATCAGAGATGCCGAGAAAGCAGGGCGACTACCGCACCCGGATATG
GAAATTCGAGGACGGGTTGAGCAACGTGTTGGTTATACAATTGAACAAATTAATCATATGCGTGATGTGTTTGGTAC
GCGATTGCGACGTGCTGAAGACGTATTTCCACCGGTGATCGGGGTTGCTGCCCATAAAGGTGGCGTTTACAAAACCT
CAGTTTCTGTTCATCTTGCTCAGGATCTGGCTCT*GAAGGGGCTACGTGTTTTGCTCGTGGAAGGTAACGACCCCCA
GGGAACAGCCTCAATGTATCACGGATGGGTACCAGATCTTCATATTCATGCAGAAGACACTCTCCTGCCTTTCTATC
TTGGGGAAAAGGACGATGTCACTTATGCAATAAAGCCCACTTGCTGGCCGGGGCTTGACATTATTCCTTCCTGTCTG
GCTCTGCACCGTATTGAAACTGAGTTAATGGGCAAATTTGATGAAGGTAAACTGCCCACCGATCCACACCTGATGCT
CCGACTGGCCATTGAAACTGTTGCTCATGACTATGATGTCATAGTTATTGACAGCGCGCCTAACCTGGGTATCGGCA
CGATTAATGTCGTATGTGCTGCTGATGTGCTGATTGTTCCCACGCCTGCTGAGTTGTTTGACTACACCTCCGCACTG
CAGTTTTTCGATATGCTTCGTGATCTGCTCAAGAACGTTGATCTTAAAGGGTTCGAGCCTGATGTACGTATTTTGCT
TACCAAATACAGCAATAGTAATGGCTCTCAGTCCCCGTGGATGGAGGAGCAAATTCGGGATGCCTGGGGAAGCATGG
TTCT*AAAAAATGTTGTACGTGAAACGGATGAAGTTGGTAAAGGTCAGATCCGGATGAGAACTGTTTTTGAACAGGC
CATTGATCAACGCTCTTCAACTGGTGCCTGGAGAAATGCTCTTTCTATTTGGGAACCTGTCTGCAATGAAATTTTCG
ATCGTCTGATTAAACCACGCTGGGAGATTAGATAATGAAGCGTGCGCCTGTTATTCCAAAACATACGCTCAATACTC
AACCGGTTGAAGATACTTCGTTATCGACACCAGCTGCCCCGATGGTGGATTCGTTAATTGCGCGCGTAGGAGTAATG
GCTCGCGGTAATGCCATTACTTTGCCTGTATGTGGTCGGGATGTGAAGTTTACTCTTGAAGTGCTCCGGGGTGATAG
TGTTGAGAAGACCTCTCGGGTATGGTCAGGTAATGAACGTGACCAGGAGCTGCTTACTGAGGACGCACTGGATGATC
TCATCCCTTCTTTTCTACTGACTGGTCAACAGACACCGGCGTTCGGTCGAAGAGTATCTGGTGTCATAGAAATTGCC
GATGGGAGTCGCCGTCGTAAAGCTGCTGCACTTACCGAAAGTGATTATCGTGTTCTGGTTGGCGAGCTGGATGATGA
GCAGATGGCTGCATTATCCAGATTGGGTAACGATTATCGCCCAACAAGTGCTTATGAACGTGGTCAGCGTTATGCAA
GCCGATTGCAGAATGAATTTGCTGGAAATATTTCTGCGCTGGCTGATGCGGAAAATATTTCACGTAAGATTATTACC
CGCTGTATCAACACCGCCAAATTGCCTAAATCAGTTGTTGCTCTTTTTTCTCACCCCGGTGAACTATCTGCCCGGTC
AGGTGATGCACTTCAAAAAGCCTTTACAGATAAAGAGGAATTACTTAAGCAGCAGGCATCTAACCTTCATGAGCAGA
AAAAAGCTGGGGTGATATTTGAAGCTGAAGAAGTTATCACTCTTTTAACTTCTGTGCTTAAAACGTCATCTGCATCA
AGAACTAGTTTAAGCTCACGACATCAGTTTGCTCCTGGAGCGACAGTATTGTATAAGGGCGATAAAATGGTGCTTAA
CCTGGACAGGTCTCGTGTTCCAACTGAGTGTATAGAGAAAATTGAGGCCATTCTTAAGGAACTTGAAAAGCCAGCAC
CCTGATGCGACCACGTTTTAGTCTACGTTTATCTG*TCTTTACTTAATGTCCTTTGTTACAGGCCAGAAAGCATAAC
TGGCCTGAATATTCTC*TCTGGG*CCCACTGTTCCACTTGTATCGTCGGTCTGATAATCAGACTGGGACCACGGTCC
CACTCGTATCGTCGGTCTGATTATTAGTCT*GGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGG
ACCACGGTCCCACTCGTATCGTCGGTCTGATAATCAGACTGGGACCACGGT*CCCACTCGTATCGTCGGTCTGATTA
TTAGTCTGGGACCACGGTCCCACTTGTATTGTCGATCAGACTATCAG*CGTGAGACTACGATTCCATCAATGCCTGT
CAAGGGCAAGTATTGACATGTCGTCGTAACCTGTAGAACGGAGTAACCTCGGTGTGCGGTTGTATGCCTGCTGTGGA

TTGCTGCTGTGTCCTGCTTATCCACAACATTTTGCGCACGGTTATGTGGAC*AAAATACCTGGTTACCCAGGCCGTG
CCGGCACGTTAACCGGGCTGCATCCGATGCAAGTGTGTCGCTGTCGACGAGCTCGCGAGCTCGGACATGAGGTTGCC
CCGTATTCAGTGTCGCTGATTTGTATTGTCTGAAGTTG*TTTTTACGTTAAGTTGATGCAGATCAATTAATACGATA
CCTGCGTCATAATTGATTATTTGACGTGGTTTGATGGCCTCCACGCACGTTGTGATATGTAGATGATAATCATTATC
ACTTTACGGGTCCTTTCCGGTGATCCGACAGGTTACGGGGCGGCGACCTCGCGGGTTTTCGCTATTTATGAAAATTT
TCCGGTTTAAGGCGTTTCCGTTCTTCTTCGTCATAACTTAATGTTTTTATTT*AAAATACCCTCTG*AAAAGAAAGG
AAACGACAGGTGCT*GAAAGCGAG*CTTTTTGGCCTCTGTCG*TTTCCTTTCTC*TG*TTTTTGTCCGTGGAATGAA
CAATGGAAGTCCGAGCTCATCGCTAATAACTTCGTATAGCATACATTATACGAAGTTATATTCGATGCGGCCGCAAG
GGGTTCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGT
GCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCCATTCGCCATTCAGGC
TGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCA
AGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTGTAATACGA
CTCACTATAGGGCGAATTCGAGCTCGGTACCC*GGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTTGGTCA
CTTTTTCAATATCCTTTCATGATTAAATTCAGCAAAGTAGAGACAAGGGGGTTGTTTAATGTATGAAAGGGTATTTT
ATAAGAAGCCTTGAGCTAATATCATATTAAATGGTTAAAAACTTCACCCCTAAGATTGAAAATAAGGCAAGAATATC
TTTCTTAATATGTCTGTTCAGCATTGCCCTTGAGATCCTAACCAGCACAATAATGC*AAG*AAAAAAG*AAACTGAT
TAGAAATGAATAATAATTATTGTCTTTATTTG*AAGATGACATCATTGTGTAGGG*AAAAAAATCCCAAAGAATCT*
ACCCAAAAAAAAAG*CTACTTGAGCTAATGTGTGAATTTATATGGTCAGTTTATAAACATTGT*AACAT*CAAAAAAA
TGAAGTACTTAGGGATATATTATGCAAAATACGTGCAAGATTTGTATGCTGAAAACTAT*AAAAC*ACTGGTAAGGG
T*AACTAAAACAGGACTTGAATATCGGAGAGATACATCATGTTAATGGGTCAGAACACTCAGTATTGTTACTGGAGA
GATGTCATTTCTCTCCAGTTTGATCTGTAGGTTAAATTCAGTTAGAGATCCTCAAAAAGTATCTTTACATGAAAGGA
TTGATTGAATCTATAATGGCAAAGGTGAAAAAGCAGAGCTTTTCGTAAAAGAAAGTGACAAACTGATTCTAAAATAA
TTTCAGAAAAATCTTTCAGCTCTG*AAAGAAAATGAACATGTAAATCA*TATTTTATTTTGTTGCAATAGCTAATA*
TTTTTTCCCTCCCTCCCACCT*CCCTCCC*CATCCACACAATAGCTAATATTTATTGAGAATAATGTTATTTTAAAG
CATTTGTTTGTATTAACTCATTTAATTCTTAAAACAATGCTATGAGCTATTAGTATTGTTCTCATTTACAACATGAT
GAAACTGAGGCACAGAAAAGTGACTAAGGCTGAATTATAACCTAGGCTGGCTCTAGATTCTTTGCTCTGAACCACTG
CCTCAGGTTGCCTCTCTTTTTCATTTCTTTTCAGAGTACTAACGCTGAACTTCATATAG*ACAAAACCTAAAGTCAG
TTACTGTTCCCATTGCTGTTTGTCTTGCTGTGTGGGTATTAAGAGGGGTGGTTGTAAAATGGTTAGTAATCACATTT
ATCATATGGCAGTTGTCACCTTTAAGCAGATACTATTCATTAAATGTATCTTAAGAAAGGATTAAGATTGGGTCTAT
AATGG*CAAAAGCA*TGTTTTTCATAGTTTACTTATATAATTAATTAGCATAAGTGATGTGATAACTATTTTCCAGA
C*AAAATCC*AT*GGATTATATGCC*TTTTTAGATGTGATAGCCTACAAGTTAAGTCGACA*TTAAATTTTCAGAAT
TTCCAGGCC*ATAAAATTACTTTTTAGTATGTCATCAGAATCACATCTGG*AACTTTTTAATTAAAAAACTACTCAT
TTGTATCCCTCTTCTCCATTTTGATTCATTGGGCAATTATAAATTAGAAAGCTCTGTAGGTAGTTCTGACATGCATC
TTAGTTGAGATCCACTAGTTAGTTTATGGAGACAGTTGTATAGAGTATCTAAGATTTGGACCACTTCAAAAAAAAT*
CCAGACTATATGCTCATCAAG*CTTAATAAAGGCAGGATTCTTAATTCTTATAAATGGCTATTAGTTCAATAGACAT
CTGTTCTGAAATTCTTCTTTATCCAAGTACTTTTCCCTTTTCTTCTCTCTGTGTATTTATGCATATCTGTGCTTCTA
CCCATCTGTCAGTCAACTGTCTGGGATCAGTTGTGCCTCAGGTCTGAAATTCTGTGATTAACTATGTCCACAACTAC
TCTGCTTGGT*CCCCCTTACACAAATAGACTTTAAAGTTACTGTGAAACAGTTTCTAAAATTATTTGAATATTACAT
TTAGGTATATCCTGACTCACTTTGTGTTAAAGTTTAATTGAAGTTCTAGGGTTTTCAGAAGTAACCTGAATGTTGCT
GGTTTGCAGATGAGAAGTATAAAGATCCTTGTTGTATTAGACTGACCAAGAGTAACAGAATAGCTTTAAAAGTAAAA
TTACCAGTGCTGCCTCTCTTAAGAACCACAGGAAATTTGAACTGAGAAGACTTAGAGATTATTTAGTGTGGGTTGTT
ATTTTCCAAGAAGTTAATCTCAGAGAGGAGGATTGACTTGCCCAGGATAGTGTAACTAATCAGGTTATGGGTAAATA
AATATAAAACTCAAATAATTATATGGTTTTT**AAAAAGG*CAGTTGAACTTTTTTC*TTTTTAC*AGTATGAAATT
AG*AATT*ACAATAAA*TATTTTTATTATCTGAGTCTTTAG*TAAAAATTTATACATGTTAG*TTTATAGTTATTAT
TAGTTT*CTTTTAAAATAATATTGTCATAGTTT*AAAA*GGAAGAAGATA*TTTATCCCAGTTTCTTAGAGTTGCAA
AAGAAGGGTCTGCCATCAGTTGACTGTGGTAAACTGCTTGTATAAATTCCCTTTATGTTAGTGCAGTGCAACTTCTG
TTTTGTCGGTGAGAAATAGGAAGCTTACAGGAAGTGTAGTCAGTCTATAGATAGGTATACGGTCATTTTCCTTCCTT
TGCTTATTGTAGAGATTATAAACT*ATAAAGTTTTTTTTTTC*ATTCAGTTGATTGG*AAAATAATTGAGCAGATTA
TTCTTTAGTGCCATTTTTTTTAAG*TTCC*TCATTCTTGTGTTGCATCAATGATACATCTTTAATAAACTTCTTTTG
TTTGTTTTTTCTTTGATGAATGTCAGTTAAAGAGAAGTTGACTGCAGATCCAGACAGTGAAATAGCTACAACCAGCC
TGAGGGTTTCTCTACTATGTCCAGTAAGTGCTAACTAATATTTGCTCTCCAGGAGGTTTAATACACA*TTTTC*TTT
TTTTATTAAGTATTTCTAGTATTTAACAAATTCTGTTACCTTGAAACATACCCACTTGTTAAGAGTTTATG*AAAAG
TAAAAGATATTATGTAGTCTTAGGATTTGAAATGTAAACAAGTTTATGTCTATAAAATTTTTCTACTCTTAGTTATA
TTTAAATATTGATTATGGTTATCAGATGACTGTCTCTCAGACTTAAGTGTAAAATCTCCAAATGAGGCATGTTTATC
ATTTCAATGTATTTATTGAAATTTCATTTTATTAAATAGTATGCTTCTGAATAAAAAT*CTGGCCATGATTACTCAT
ACATTTAACTCAACCACTTTG*TCAGCAATGACTATAACAAGAATTTAAAACTATTATTTAATAATAATTGAATAAT
TATTGAATTATTACTGAATAATTGAATTGGTATGCATATAAAAAT*CAACTAGACATA*TTTTTAAACCAGAAAATG

AGATCAGTTTCATTCATTTGTATAAAAT*AAATATAATGCCTAAGTGAAAACTGGTAGTGAATGTTTCC*AAAAAT*
GCGTTG*TTAAGGTTTG*CTTTTTAAC*TAATTTGATATTAGGTCTACCAGTAGTAGGAATGGTAGTTGAATGATGC
TTTCATACAAAGATACGT*GAAAAACTAAAAGTAAGTTACATTTTGATTGCCTTTTGTAATCATATTTAAA*TTGGG
GAGAAAATATGAAAAATAAGAGGACTCAGCTATTTCCTCATTTGCTGGGATTGTTCAGGCAAGTCTGATATTGAAGG
CTTTTAATACTGAAATCC*TACTAATCTAAGATTAATTTATTCTT*AAACAAAAATTGTTTTATTTGTCTTATATCT
TAGTCTCTTATAATTTAT*AAACCTCAACACAAGTGTGAAAAATACACCTCAACACAAAAAGTAATGAAAAAGTCTG
TAGCTCAAGAATTAGTTTATGCATATGTTGCATAGACTTAAAAAAAAACTTTAGTTAAGAGAATTTCTCTTCACTAT
TAAGCATCTGTAAAAAAAATACAGTGAAATTAAGAAAACCACCAAACCTATTATTTTATATATAACCTTCTATCTTT
TT
>NCL_C4
GCCCAGGCAGCAGCGCGGCCCGGGGAGGGCTGAGACGCAGAGGTGGCCGAATGCCGCGGGGGGGCAGGCCTGCGACC
CAGCCCGGGCAGGGAGGAGGGCGGTCAGGGCCGCCCCCGAGGCCCCAGGCGCCGGCCCTGGGCCGAGAGGATCTCCA
TGGCGGCCGCCGCCCCGGGCCTCAGGCCGGACCCCTCCTTCCTGCGGGCTGCGACAGGCGCGTGCCCGAGCTACGGG
TCGCCG*GTGAAATCCCCACAGGGCCGCGCGCGCCCCGGAGCGCGAAGGACGGCCGCGTTTCACCCGGCCCGGCCCG
GCAGGCCTAGGGCCTGAACTGCCTCGTCCTCCCGCCGGCGCCCGCCGCCCTCCCCAGGAAACCCGCAGGCCCTAGCC
CACTGGAGCTCAGAAACGCCAGACGCAGTCGAAGGTTTCAGAGGTCCACGCCTGGGAAGGCCCTCGGAGCGAGGGGG
AACGCTGCGGCGCCAGCAGCCCTTGGGCTCGGGACAGGGAGGAAGGCGGGCCTCTGCAGACTGGGCCGAGA*GGGGG
AGGGTGCGACCAGGGGACGGTTAAGTGGGGTGTCTCTTCCAAAGCCTACCGGCTGGCAAACAAGGCCCAATGAAGAA
AGGTTCTGCTACTCGAAGGCGGCCCGTACCCAGTTCGGCCCTGATGTCCGCATTGCCCTTGCTGCTCGGAGCTTCCA
ACTCGTCCCCATTAGGCCCTACGGGTCCCTAGCCTGGGAGTGGTACAGAGTGAGACCACTGAACCGTCAC*TACCGG
GGGGTGAGGGGGCGCGACTCCGGGGCCCTTCCC*TTTTTGCCATAGACCTTATGGGACCCTTCACCCGATTGGGGCC
CGGGGAGGCGAGGACGCGGACCTGGCAGGGTCCACTTTGAGACCAGGGCTGTGTCTCCGAGTCCTAGAAATAAGGGA
GGACTTCGGGTGTCGGCACGAGGCTTTGAGAAGGGACTGGAAGTCCGGCCTGTCCTCCCTGGAGAGGAGGATGAAGT
GTGTGAGCAGGGAACAAAAGAGGGCTCACCCTGGGACGGAAGGCAAGGGCTGTGTTCCCACAGACCAGGACTCTAAT
CCTGAGCCTGATGGAGCCTTCCATCAGGAGAAACCTCAGGGAGTTTTCTCCCATATCTCACAAGATTGTTAGGATTT
GAAAAGAGCATTCATCCAACGAATTTTAGAGGTCTGCTGTGTGCCAGGCTCTGCTTGGCTCTAGGGATTCCGTGATA
GACCAGATGAACACAGTGCCTCTTCAGGGACTTTGTTGTTGTGGAGAAAACTGACAATAAAATAAACAAATAATAGG
CAAGACTGGTTGAGTTTATATCACTATGATTTTG*CTGTGAAGCAAGTGAAATAGAGTGATGGTGGGGT*GATGCTA
ATTCACTGCGTGTGGTGGGATAAGACCTCCATGACAGGTAAGGAGAAGGGGTA*TTTTAGGCAGAGGGAACAGTCTT
GGTGAGTTCAGAAAC*AAAGGTAAGAGTGGCTTCAGCAGAGTAGGCCATAGGAGAGCGACAGGAGATGAAAGCTGGG
TGCAGCTCTTGTTAGGGCCCAATTACCCTTCTTCTCAAATCTGATGGGGGCATTTGAGTAGTTTTAAACAGAAAAGT
GACAAGATATGGTTTATATTCTTTAAAAGATCAGTCTCTAGTAGGCAAGAGGCAAGGCAGGAAGGCCACTTAAGAGG
AAAGGTTTGGGACTAGGATGGTAGGAGTGGAGGTGATGAAGCCGGTGGGGCAGTCACAAAAAGTCCAAGTTCATTAT
GGAGGTCACTCCCTCCATACACACAC*AAAGGATAGTATCTACACCACCGCCCCTAAAACGCAAAGGACAATATCTA
ACATATAGGAAGAGAAAATACATTTATTTCGTCCCAAGGGACCAGGCTGGGCTCAGTTCACATAAATTATTGCATGT
GAACCCCACTATTCCAATCCTGGTTTAAAGGCCTCATCTCAGTTGCAGAGCTATTGC**T*GGGGGTTTGGG*CCAC
TGCAGCACAAC*TTTACCTA*TTTTCAGATCTCTA*TAC**TTTTTTTTCCCATATTAGACAAAAACAACATTGGAA
ATCAGAGCTGAAGATAGGGTTCTACTAGTAAAGTCTCTGAAGAAGCTGGTGGAAAAGAAGAGGAAGCTGGTTGTTTT
C*ATTCTGGGGCTTCAAGATTGGAAAAAATTCCAGTTCTTGACAGGCTGACCCAGGC*TTTTTGGCCTCTGGGCCAA
ATGTGGAGTGTGGGACCT*GGGGGCTCCCTTCCAGCTGATTTCATTCCCAGGCTCTGTATAGAAATACCAGACTCCT
CCTATTCGCACCCCCTCTCTCTGTCTCCCTCCCTTCTTCCCTCTCTGTGTAGCCCTTTTCCCTGCATCATTCTTTTC
AAACTC*AACTTTTTGAAGGTGATTTAGATTCTTCCCTTTGGTTTCATTTTCTGAAGGAGACATGATTCATCCAACT
ATATATTCATTGCTGAATCAGAAACTTTCTTCTTCCTCTTCTTGTCTTTTTGAGATGAAGTAGGGTTCTTCCTTGGA
AAAGGAAATGGCAACCCACTCCAGTGTTCTTGCCTGGAGAATCCCAGGGACAGGAGAGCCTGGTGGGCTGCCATCTC
TGGGGTCGCACAGAGTCGGACATGACTGAAGCGACTTAGCAGCAGCAGCAGGGTTCTTCCTGGTGCTAACACCTCAC
TTTGGTGGTATATCTGGATGAGTGCTAGGAGGGGTGACTAAACCGAGAGTCACGAGAAGGGCTCAGAACTTGCACTT
TAGGGCCTCCACCCCGAGATGGGAATATCCTTCTGCTCCAGCAGCCCAAGGTCAGGA*GCCTT*CTTTGTTTATCAC
CATTGCCTAGTATAAAGTTCTGGGCTCGCTGGGTTGCCCAATAGCCATTAACTGACTGGAGGCTATTTGGTTGCCTA
TA*TTTACCTGGGCTTGGAGGATGATGTTTCTGTGGTCAGGTTGTGCAAATTGGAGAGAAGATGTATAAACACAGCA
AGGCTGTTTTTCCAAAGGCCAGTGCCCAGAAATAGCTCCTTGTTTTTCTGACCTATTCCCTTCCACTTTGACAGAAA
ACCACTGAGCAAATGATCAGTAATAAAATATAACTGAAAACAGTTGTATCCAGATGGGAGATTGATTTGCTAATGAC
AGCAAATTGCAGCCTTCTACCTGGACAAAAGACAGGACCAACCTCCAGCGCTTTGCTTTCTGGGGATTGGAGGTGAC
CTGTGAGTGGAATTAGCTGTACTTTCCATAGGAACATTCATTTTTCCAAGCCAGGTGGGCCCAGCCTGCAGCTCCAA
GTCAGTGGTGTTTCTCAGATTTTGTTTAAAGCGC*ATTCAAAATGTCGTTTTTCAAGCACCAGTAAAGATAC*TTGT
TTTCGTCCTATAAGTCTCCCACATACTCTAATGGAAGAGGATGACTTTGGGTATTGAAGAGATCTGGGTTGGACTGC
CTTCTCTGTACATTAGCCATGGCCTTGGGCAAGTTTCTGCACTGTACAATGGAGGTAATCCTTTATTTGTAGGGCCA
GTGTGATGCTGATAAGCAAGCTCCTTCACCCACCCTCATCCTGTGAAGTAGCGCCCGTCAGTACTTGTTTCTCTTCT

CCCTAGCTCCCACCCTGGTCATACTTCTATTGTGCCATTAACTTGTTCAGGAAAGGGCTTCCGGGACTTCCCTGGAA
GTCTAGTGGTTTAGACACTGCACTCCCATGACAGCA*GAGGGCACAGGGTCCATCCCTGGTCACGGAGCTAAGATCC
TGCATGCTGTGTGGCAT**GGCCAAAAAA*AGAAGAAGGAATTTCC*AGCCCA*TCTGCATTTTTCC*TACAGGCAT
GTGTCTGTGTCTGGACACATGGAGAGCCCATATGCAAGGGGCTAGGGAG*AAGGAAATGGCAACCCACTCCAGTGTT
CTTACCTGGAGAATCCCAGGGATGGGGGAGCCTGGTGGGCTGCCATCTAAGGAGTCGCACAGAGTTGGACACGACTG
AAGCGACTTAGCAGCAGCAGCAACAGCAGGGAGAGTAATTGACAAGAAGTTGGAAAGTCCAACAGAACCACCAAAGA
AAGATGTCAGAAATCAATGACAAGCAAGAAACCACTGATGACCAGTCTCTATTAGCCCATTTTCAGAGGAAAATGTT
GACCCACTGGCCTTTGTCTCCTTGTCCTCTCTCTCCCAGCCTTCCCCGAAAAGCAGATACAAAATTGTCAAAGACCT
AATGTGGAGGAGAGAGGGCAGCTGGTCTGTGTCATTTCCCAACCCAGTTCACCCCTGCTGGTGTAAGTTTACCAGGA
ATGTAGACTGAAATAGGGGCTTCCCTGGTGCCTGCAATGTGAGAGATCCAGGTTCAATCCCTGGGTAAGCAAGATTC
CCTGGAAAAGGCAATGGCAACCCACTCTAGTATTCTGGCTAGAAAATTCCACTGGTAGAGTGAGCTTGGCAGGCTAC
AGTCCGTGGGGCTGCAAAGAGTAAGACAGGACGGAGAGACTTGACACTACTGCTACAGACTGCAGTGGAAACAAGTC
TTCTTTCTTCTAACCTCATTCCTCCCTGGATAGTGGTACCTCCCAGTCAGGTACCCTGGGATTATAACAAATAACAA
TGGTGAGCTCATGCTGCGTGCTTACTGTTTGCCAGGTATCACAGTAAATACCATGTGCAGATTAGCATGTTAAAGTT
GTGCAAGCCTCTGAGGGCCAGGCCACTGTTATCATCCTCATTCTTCAGGTGAAGAAACTGAGACTGCGATATTGTGT
AACTTGCTCATGGGCTCGTAACTAGAGTGTGGTAGAATGTCAGGATTTGGATGTCAGCTG*TTTGATCCCCA*CCCC
CTGACCCTGCTACCAGCCTTCTGTTCCCTCTTTATCCCCACTCTGAGTGGCTCTCATCCCGCCTGTTCTCTGCAGCT
AGGATCTCAGCTAGAGGTCCCCAAGGCTCTCTATGGTCCTGATCAATTTGCCTGAGCCTAGGTGCCAATGAATTAAT
CTGATCATCCATTCTTTGCTAAT*GGGGGAATGCTAGCAGAAACTCTTCCCCACCTGGCCACCCCCCATACACCCTG
GACATTCCTCACTCCTGCCAAGAACACAGGTCCATTTCCAAAGCAAGTGCTGTCTCATCTTAGAGACCATCTTTGTT
TGCATC*TTTGTTTACCACTTTATACCTTTCCTGGGACCCTTATCTCTTTCGAATGATAACTCATGTCTGTCTAGCA
CTTAATGGTCTACACCACTTTCCAATTTGCTTAGCCCAGCCCCCACTTCCTAGGCTTCTTAACCATTTTCTGTAGCC
CAGGTGACAAGAGAGCATGCACTTTAGCATCCAAGGTCCACCACTTGCTAGACAAGTTACTGATTGACTTCTTCACT
CCATTTTCTTATTGTGAAATGGGAGAATAATAGTGCTTGTTTCTCGTTGTGAGGGTTGAATGAAGTAAACACAGAGA
GACACATAG*AAAAACACACCAACACACATACACACACAGCAAGTAACTCATTAAGTGTTGGCTGTTGTTATCATTA
TTTCATTGAATCCCCAAACCCTGTGGGGAAAAAATCAAAGCAAATCAGTTCTTGTCACCAGTTTAACTCAGCTCTCT
AAGCTCTCTGCTAGCTCTGTCAAGCTCTGATTCTACCGTGAATCACGGTTATTTATGGACACAACTTAGTTCTATGA
TGTAGCTACTCTGAAGGCCCAGCCCTTATCAGAGGGCCTTGTATATAGGAGGCTGTCCCGCCAGGTTGGGAGTTAGC
ATGTAACAATAGTCCCTACCTCATCCAGGCGGTGGGAGGAACTAATGAGAAATTTCTGTAATATGCTTTTTAGTATA
GTGCCTGGCACTGAGTAAGCTCCTAATAAACCAGAGTCACTATTGTTCCATCTTACAGTGAGGCTCCCAAACATGTT
CCTTAAAAAC*ATAGACTCTGCAACCATACTGTCTGGGTTCAACTTAGTGCTAATTCTATTAGACTCTAAGTTAATT
AACCACCTGGGCCTCAGTTTCCTCATCTGTAAGATGGACATAATAACAGCATGTGCCTAAATGAGGTTGTTCCATAG
ATCAAATGAGTTAATATAATAAACACACTTAGATGTGTGCCTGTCACATGTTGGTATATGTAAGAGCTTTTCGTCAC
CATTTATAGATGAGGAGACCCAGGCTCAGAGACATCAAGCAGTAGTGGAGCTGGGATCCAAATCTAACTGAGGCTGT
CTGAGACCAGAGCCTCACTCTTACTATTAGGATATGTTGCTGCCCCAGATGATCCATGTATATTTTAACAGGAGCCT
TCTCCTGTGGTGGAGATGTCAAGAGTGGAGGAGAGTAAGCTGGCACCAGAGATCTCTGGGGTAAGGGAGATCACAGG
ATGCCCTCAAGTCAAGCCATAGAGGTGGCTTTTCCTCACCTGTGTCTGGCCATATGATTTTAGCGTTCTATCAAGTC
TCAGACCCAGAAGGCCT*AAAAATGGCACCACCCATGCAGACCTCTGACCCGGAATCTGCAGCAAAACTCTCATTAT
TGCACACTGATGTCACACTCAAAG*ATGGGCTCCTGACTTTTAACACAAAGAGGCCCTAGCCTTGAGGCATTTTCCA
TTTCTGAGGCATGATTCAGCCTAGGAGACAAGACCAGACAGACTTGTTCCAGAGGAAGGAGGCCTGGGTGGCGGATA
GAGAGGGAGAAGAAAAACAAACAGGGCTTAGAGCTGAGCAAGTCTCACAGCTGCTGGGCTGCAGGGGGATGGAGCAG
AAGCGAACAGAGGGAGCCTGCCTGCAGAGAACGGGGCGCCCACTGACCAATTTATCTTTCCTGAAGTCAAATGCATT
GATTTAAATAGAAATCCAGCCTATGCTTCTAAAGCCCCAG*AAAAACACTAAT*AACATTTCACAACTGTGAAGCAC
AATATGATTGCCAAAGCATTCCACATAAGTAAGCTCAATTTTAGCTGCTTGTTAAAATAATTTGTGTCTCTTATTAC
CAAGCATCAAGTGCTTATTGAGGAATACTAGGAAATGTTTTTTAAATGTAAATCTTAAAATGTCAAGTATGAATATA
CTGCACCACTAGTAATAATATTTTCAGTAAATTTCTAACTAC*TTTTTTCCG*AGCCTATGTACATACAGCGGTCAT
TTAAACAAAATTGAGAAGTACTACAGTAG*TTTTTTTCTTTTTTTAA*TTTATTCAATCATGCCATGTGGCATGTGA
GACCTTACCAGATCTGATCCCACAGACCAGAGATCAAACCCATGCCCCTTGCACTGGAAACACAGTCTTAACCACTG
GACTGTCAGGGAAATCCCCAGAAAGCTTTAGATCTTTTCTTTCTCATGTAACGTAAAATCACGGGGA*TTTTCTAAT
GGTGCCAAACGTTTAGCTGTAACAATTCGTAATGACCGCATTTATCAATATAGACGTGCCACGTTTGCTGTAGCCAC
ACCCTAATTCCTGGACATTCCAGATATTTACAAAAGTTGAACTCTTTAAATAACACTGCTGTGAACATCTGTTTACT
TACATTTTTGTCTATATATCTGATTATGTCCTAATAGTAAATTTCTAGTAATGAAATTACTAACTTCAAGAATTTTA
ATATTTTTGAGTACATATAGCCAAATTACTCTTGAGAAAAATGTATTCAGTTTACCCTCTGACCAGCACCGTGTGAC
AATTACCTCTCACTCTCTGATTGAGCCAGTGAGTTAGGTAACAAGCATCTCATGTTACAAATGTGGAAGACTGGGAT
TCATACAACTGAAAGTCAAACTCTAGTGCCCTGGATTTGGTGCCAGGGATCTTCCCACTCCGCCTTCTAGACAAGAC
AAACTCTGAAAAAGATTTTCCTGAGTAATGGAGTTGTAGAAAATTTTATTTTTATATTCTC*ATTTA*TTGGT*GAA
AATTTTTTTTTAAA*TCTTCAGTTTCTTGACCTTATTTGTGCCATGGGCCCCTTTGGCAGCCTAGTTAAACCCCT*G

GGGCTTCCCTGGCGGTTCACTTCTAATGCAGGGGGCAAGGGTTTGATCCCTGGTCAAGGAACCAAGATTTCACATGC
CGCAGGGCACAGCCAAAGAGAAACAAAGG*AAAAAGCCTACCAGCCCCTTCTCAGAGTAATGTATGTATATGCATTT
AA*CCCAGTTATATTAAATATAAAAGCAAAATGGTGATGGAGTAATGCTCCTTTATGTGCTCCTTCTAGAACTACAT
GGAAGGAGAGGTAGCCCATCAGGCGTCCTGAGTGTTCCTGTTCATTCTAGCTGAGTTTGCTCTTGGGATATCAGCAG
CTTCAGTAGCGGGACACTTAGACAGCAGAGTAGCTTATTCCCTGCTTTCTCAAGAACTGCTGGGCCTTGGTTCTCAG
TTCTTATGTAAC*ATAACTTACTACCTGTGCTCCAGTCACCTGGCCCTCTACATTGCGCCGTGGGAATTAGCCCTGA
GGGAAACTGCACAGACATGCTGACACCCTCCCTACCCCTTTTAATATGAATAATGAAGTCTTACATCTGGGATCTAC
ACGTTTTGTGTCTTCAGTCCTTATCCAGGAAACTGTAGCAAGCTGGCTGAGCTCATAGGGAGGGTAAAGTCTGAACA
CCTTCACAATTCTGACCCTACCATTCATTTCAAAGCAGAGTCGAGTGTAAATGATATGTCGAGGTGTATGCAACAGC
TAGAAAGCAGATCAAGAATACGTGTAG*TTTTTTACTGGTACAAAAGTCACTGTAGTTTATTGTATACATTCAGCTA
AAGGTTTCGGAAAATGTCAGGTTTTTTTTTTTTTTCCTCATTCAAGTTCACAGA*TTTTTAGGAGAACTCTTGGCT
GCTGCCAACACCTCAAAAGAAGACCAG*AAAATCCTGTTGATCAAGCAAAGCAAGTCTAAGTTAATTTGATATTTAC
TGCAATAAGGGAGAGTAACACCTTGAAAGAGACTTGGTATAGTTACTAAGAGGATTTCAAGAAAGGGTGTTTACAGG
GTGTGGAGGCCTAGGCTCAAGTGGTTTAAGGT*GGTTTTTC*AAAGCAAGAAACTGATGGGGATTGGGCACAGTTTA
TGACATTTAAGATAGCTTGGGATTTGCAGACACAGAGAAGCAGAGGTCTAGAAGCAAGTCTCCTTAAATAAACTTGG
TCATATATGTGAGCTCTTGGCCTGTGTGAGCGGTCTGTGGTCCTGACAGTCATCTGTCCTGTCCTAGATGAGGAAAC
TATTTGCCCAGATAAATTCATTTCTTAAAGCAAATAATTAATCTACTTATCCTCCTGCATAAGAATTTCCCTGGGAT
TTCCCTGGTGGTCCAGTGGTTAAGCCTCAGTCCTTCCAC*AGCAGGAGGCACAGGATCAGTCCCTGGTCAGGTAAGC
TAAGATTCTGCATACAGTGCAGCACAGCGGGGGAAAAAA**AAAAAAAAAAAAAAAAAC**AAACAAGAATTT*CCTG
*ATTTTTGAATTCAGTGATTTTCTAAATCCTTCATGCTGTGTTAGAACTGGAAGAAATATTCCA*TTTGGAGTTAGA
ATTGGAAGAGGATCATCTAATGGAGAAATCTT*GGGGTTTGGAATCAGAAAAGCTTGCCTTTTAATCCCTGATCGAT
CAATCACTACTGGCATTGTCATTCTTTTCATCATTTCCTTCATCAAAGAAATATGTATTAACCTCCTATGATGGGTT
CTAATAAATTGTCTAAGTTCTTGGAAGCTCAGTTTCCTCTTCAGTAAAATAAAAATTAAAGTAAAAAAAT*AAAT*A
TAAAAAAG*AACTT*CCT*GGAGCAGTTGAGTCATGTTGACACGGGTGGTCTCAGTTCTTACTTGATAAATCGTGTC
CAAGTCTTTGCGACCCCATGGATCATACAGTCCATGGAATTCTCCAGGCCAGAATACTGGAGTGGGTAGCCTTTTCC
TGCTCCAGAGGATCTTCCCAACTCAGGGATGGAAGCCAGGTCTCCCGCATTGCAGGGGGA*TTCTTTATCAGCTGAG
CCACAAGGGAAT*CCCCCCTGGATAAGACCAGCTTCTTAACTGAAAT**CCCCTTTCTCATGGAAGTGGAGGCCTAT
TTTTGCCTGTCTGCCCTTTCCATACTGTGAGTTCACTGTGGGT*GGGGGCCAAGTTTTAAACTCCTAGATGTAGGAT
TCATTTGTTTTCTCTATGTTT*ACCAAAAAA*A*AAAAACAAAACAACCCACACAATACCTT
>NCL_C5
GGAAAATAAATATATTATTAAAATTACTTTCATCTGTTCTTTTTTACC*TTTTTATTGGACACTAGAACATTT*AAA
ATTACATAATGTGGCTCACATTTTATTTCTATTGGGCAACACTGCTTTAGATGTCACAATACTTGTGGTCCCTACTT
ACCAGGGCCGGCTTCATGGGTGTGTAGCCAGCGCAGTGGAACAGGGGCGTCCCATAGTGGGAAGTGCCGTGCATTTC
ACTTAATGCTCTGCCGGCACTGCCTTGATAACCTTAATTGTGCCTCACGTGCCGTCACTTCAGTCGTTTCTGACCCT
TAGTGACCCTGTAGCCTGTCAGCGCTCGGTCCTTGGGATTCTCTAGACAAGAATACTGGACTGGCTTGCGTGCGATG
CCCTCCTCTAACTGTATCTTAGAATTTACATTTCAGAAGGAAATCTCCTAGAACAATGGAGCACGCGCCAGGGGCTA
GAAGCCCTGGTTCCTGCGCCATCCTGCCTTCCGCGGCCTCCGTGTCCCTGGCCCTGCTGCCCCCTTCCGACCCCGGC
TGGGTTCTGGGCGCGGCTGTGAGACGGGGCGGGGCTATAACTGTGGCCATCCCTTCCCCTGCCTGGCTGCCCAGGGT
CACCCACTCACCTGGGTCCAGGGGCGAGAGCTTGACCGACGCTAAGGTCTCCGTACCGTTGGGAGTCGCTTGTCTGC
CCCGATGTGGGCGGCGAGCCCTGGGGAAGGAGAGACACCTGGCCTGAAGTCGCTAGACTTCGCCCGGCGAGGGCGCG
GCGCCGCGGTCAGCAGCTGGCAGGAGCTGGACCCAGCAGCCCGGCGCCGGCCTTCTTTGCTGCCTGCCTGCGGGAGG
GATGCGCCCTGCAGGGCCGTCCTGGAGAGAGAGTGTCCGCTTGCCCCTGTAGCGCCGCGTACTCCGGCTCTGAGCTT
GGGGGTGGATTGGGAGGACTCCTTTCCTCTTTCTAATCTTCCTCTGTCTTTCTAGCGTTTGCCTCTTCTGGACAGTT
CAAGACACCCCTGGGGCAAATCATAACTCACAAATCGTGTGATTTCTGTATTGTGTAATTCCGTGTGTGAATACTAT
GATATTTGTATTTAAAACTGGCATTGCACAGTAGAAAGGTGAAAGGTAAAACTCATGTTAATAATTATAACTTTAAA
TTGTAATTTACTTTATATATTTAAATTAAAATTTTTAATTGTGGCAAAATACCCAACACAGAATTTACCATCTTAAT
CATTTTAA*TCAGCTCAGTAGCTTC*AAGAATGTACATTGCTGTGCAACC*AATATATCCATGTTGCAGTGTGTGTC
AGAAATTTGCTTCCTTGTTAAGACTAAAT*AATATTCCATTGTATGTGTATACCAAA*TTTTATTTATCCACTCATC
CAGC*AATAGGCATTTGGG*TTGCTTCCACTTTTTGCCGTTGTAAATAATCCTGCTATTAATTTTTCTGCATAAGTA
TCTCTTTACCTTTG*TTTTTG*CT**AAAAT*AAAC*AAGACTGTCTTTTG*TTTATTTAGACTTGGTACATACATC
*ACTTTGCATACTTGCGAGTGTATTTGTAGGATAAGTTCCTGAAAGTAAAATCATTGTGTCAAAGGAAATGTGCCTT
TTATGATTTGTTAAATACTGTTAAATTGCTGTCCGCAGATGTTGTACCATTTTTCACAGCTTCACCAACTCTGTATG
TTTTCAGAATTTTTGATCTTTCACAATGTGCTAGATTGAAAATTTTCCATTAATAAGAATGAACAGAAACATGTGAA
ATATGTTTAAACCCATGAGTTTGTAGTAATACAATAAACTTCCCCTGAAACCCAAAATGTTTACTTTTGGAGGATAC
AGAAAGTGAAGAGG*AACTAAAAAAACCTCTTAATGAAAGTGAAAGAGGAGAGTG*AAAAGGTTGGCTTAAAGCTCA
ATATTCCAAAAAACGAAGATCATGGCATCCGGTCCCATCACTTCATGGGAAATAGAT*GGGGAAACAGTGGAAAACG
TGTCAGACTTTATTTTTT*GGGGCTCCAAAATCACTGCAGATGGTGACTGCAGCCATGAAATGAAAAGATGCTTACT

CCTTGGAAGAAAAGTTATGACCAACCTAGATAGCATATTCAAAAGCAGAGACATTACTTTGCCAACTAAGGTCCGTC
TAGTCAAGGCTATGGTTTTTCCAGTGGTCATGTATGGATGTGAGAGTTGGACTGTGAAGAAGGCAGAGCGCCGAAGA
ATTGATGCTTTTGAACTGTGGTGTTGGAGAAGACTCTTGAGAGTGCCTTGGACTGCAAGGAGATCCAACCAGTCCAT
TCTGAAGGAGATCAGCCCCGGGATTTCTTTGGAAGGAATGATG*CTAAAG*CTGAAACTCCAGTACTTGGGCCACCT
CATGCGAAGAGTTGACTCATTGGAAAAGACTCTGATGCTGGGAGGGACTAGGGGGCAGGAGAAAAGGGGATGACAGA
GGATGAGATGGCTGGATGGCATCACCAACTCCATGGACATGAGTCTGACTGAACTCCAGGAGATGGTGATGGACAGG
GAGACCTGGCGTGCTGTGATTCATGGGGTCACAAAGAGTCGGACATGACTGAGCGACTAAACTGAACTGAACTGAAC
TGAGGGAAACAATTCATTATTTTGAAACCAACAAATGAAAACAAAAAGAAAGAATCAAGAATGACGTGCCTTGTCTA
TGTAAACTGTATGCCAGGATAGCTAACTAGTGACCCAGAGGAAGTATCTGTTTAT*AAAAAAATATTCTCATGCAAG
CACT*AAGG*AAATTAAAGGCAA*CAATCATCAATTGTTACAAAGAGA*TGAAGATAAATGAGAAAGAGATGACTAG
ACATCCATCTATGCTTTATAACAGGAGACCACCATGCTACCTCTCAA*CTGTTCTTGCCTAAACATTC*AAACCTGA
ATCTGATCAAGTCATCAGATTTAACTGCCAACAAGTTACAGAAAATACAGGGATAGGAAGAATATGTTAACATATCA
AGGAGATGCAACCAGCAAAATTCAGACTGTAGGAAACTCTACAGAAAAAGTGATAATTTCTTCAATGAATAAATTAA
ACAAAGAAAAAGGAAAAAAGGAGAGGAGGGATAGGAAATCTCTAAATTAAATGTAAGGTAGATAGCAACAAAAT*GC
AATATATGGGTCC*TATTCAGAGCCTGTTTCAAAT*GAAAAAAACATGTTCTATCATTTACGAGGCAGTTGGGAATG
TGAACCCTGATT*GGATATTTAATGATG*TTCAGGAATTAATTGTTGATTTTAAGATATTATGCTATTATGTTGTTT
TAGAATATGATATTTTAGATAATATGGGACATTTTTTTTCAGATTTAAAGAAGTCTTATATTTTAAAGCCATATACT
AAAATAATTGTTGATGAAATGATAACAATGTCTGAGATTTGG*TTCAAGATAATACAATGGGAATGTGTGTGGGATG
GGGTA*TGGACGAACCATGATTGATTATGAGCTTATGATCGCTGAATGATTCATACTCTACAATTCTCTCTTTTATG
TTTGAAATTTT*CC*AAAT*AAAAAAGTT*CATTAAAAAAAAAAAAAGT*CAATTGGGTCCCAAAGTGTCGAACACAA
CTGATCAACTGAACTGAACTGAATCAG*AAACAATGAGTGCAAATAGC**TCTTTTTTTTGAA*TTTTGATGT*AAA
AAGG*AAAAACAAATCAACAGCTATAAGGGTCAACAGAA*GGGTATT*ACCAAAAAA*AAAACAACATAYCCT*TAA
TTATTCAGTTCTTACATATATGGAAAAGTGTACAGCACAAGGACTCCTCACTAAGTAAAGCCCCCGTATATTAAAAG
AGGTATTTGAGT
>NCL_C6
GAGGGGCGGAGGGGCGGAGGGGCGGAGGGGCGGAGGGGCGGAGGGGCGGAGGGGCGGAGGGAGGGCGCA
CGGCGCGGGGCCTCCTCGGGCTGGGATCTGCTGCGCTCCCGCGTGTGGCCCGGGGCAGGAATCTGCCTAGGCTCCCC
GAAACCTGTATAGTAGTCAAGGGTGGGCTAGGGACGTGGGGAAAGGCCAGTTCGACCTAAATACTCTCTTGGGAAGG
AGCCGAGACTGGATTTCCCTTGTCCTTGCGGACCTGAAGGCTTTGGGGGA*GGGGCTCATGTCCCTTCCAGCTGTGT
TAGCCACCTTCCTCAGATGGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAAGTCGCTTCAGTTGTGTC
CGACTCTGTGCGACCCCATAGTCGGCAGCCCACAA*GGCTCCCC*AGTTCCTGGGATTCTCCAGGCAAGAACACTGG
AGTAGCTAGCACTCCTTAAAATGTCAGCTCTGTGGCCTTTGACGCTGGGAATTAAAAACCAGTGGGATAAGGGTAGC
ACGCGTGAAAGAGCGAATTCTTTTCAAGATTCCCATTCTGGCTTTGTCACTCTCTTGGACTCTGTGGCTTCATAAAA
GTTACTTCCCTTCTCTTCCATTCAAACATGTACTATAACTCCTGAGCAGTTCCTCATCTTCGAGACTGGATGACCAG
TACCCAGGAAGACTTAGGTGGGGACGAAACGGAGCAGGATCTTCTGGTCCTT*CCCCCCGGCCCCCCT*CCCCCCCA
A*CACACACACCGTATCTTCTGCCTGCCTCTGACCTGTGGAAAACT*GTAGTCTAAAACGAGTTTAATCACAGAAGT
GAGAAATGCTGAAAC*AAAGGAAAACCGT*CAAAAGAGACCAGATAATGTTCTCATTAAGCGTAGTCAAGGACCTTT
AATTTCTTCTACAGGGCTATAGATAATGTTTGGAGCCATATTCTGTGAGCTGTCTTATAGATACTAAAACAGAGGTG
GAGAGGTTAACTACGTGATGACCAGGCTGAACCCAGGACGTGAGCTGCCACAATTCCGAGAACTGACCACAAAGAAA
TGGGAACAAATGGACCCTGGAACTGAAGATTACTTGTACCTAAAACAATCAAGATGATACTGGTCAGACCACTGATG
ACCAATTTGAAGATGATTGTCAGAGATGACTTTGCTATTTCTGCATGTAGGCCCTGCTCCTCAGTCAGCCAGTTCAG
TCACTCAGTCATGTCCGACTCTTTGTGACCCCACGAACTACAGCAGGCCAGGCTTCCCTGTCCATCACCAACTCCTG
GAGCTTGCTCAAACTCATGTCCGTCTCATCCTCTGTCATCCCCTTCTCCTGCCTTCAGTCTTTCCCAGCATCAGGGT
CTTTTTCAATGAGTCAGTTCTTTGCATCAGGTGGCTAAAGTAATGGAGCTCCAGCTTCAGCATCAGTCCTTCCAATG
AATATTCAGGACGGATTTCCTTTAGGATTGACTGGTTGGATCTCCTTGCAGTCCAAGGGACTCTCAAGAGTCTTCTC
CAACACCACAGTTCAAAAGCATCAATTCTTTAGTGCTTAGCTTTCTTTATAGTCCAACTCTCACATCCATACATGCC
TACTGGAAAAGTCATAGCTTTGACTAGATGGACCTTTGTCAGCAAAAGTGCAGTATTCCTCCTCACTCTGTCTATAA
AAACTCTCACCTCCTGCTTATGGTGGGAGGAGGGCGGAGTCGGCCTTTGGATGAATGTCCTCTCCCCTCCCTG*CCC
CAGTTGCCGGCATCTGAAATAAAGAAACTTTCCTTTCCACCAACCTGGCCTGTTTATTGGCTTCTGACTGGCAAGAA
ACCAGACCCCACACACATACCTTTCAGTAACAGGGGGACTGGTGAGT*GGCGGGGGGGA*GGGGGGGGCGCGGTGGA
GGTGGGGAAGATACTGAATACCGTGAGGTTAGAGACTTGGAGGGGGTGGTATCAGACAAGGCTTCTGGAGGAGAGCT
TAGGGAACTGAACTCCATTTGCTGAGGTGGAGCACAGCCAATTTACTGAGGAGTTGTGGTTGCTGAAAAAAAGCACA
A*CCTAAAAGTTGAGAATTATGTTTTATTTGGGGACATTACTGAGTGCTGTTGACCAGTATACAACCTCTCAGATAG
CTCTGACGAATTGTTCCAAAGAGATAAGGGTGAAGCCAGGATGTTCTGTTCAGTTGCTAAGTCATGTCCAACTCTTT
GCAGCCCCAAGGACTGCAGCAATCCATGTTTGCTCAAATTCATGTCCATTGAGTTGGTAATGCTATCTAACCATCTT
ATCCTCTGTTGCCCCCTTCTCCTTTTGCCTTTAATCATCCCCAGCATCAGGGTCTTTTCCAATAAGTTGGCTCTTCA
CATCAGGTGGCCAAAGTACTGAAGCTTCAGCTTCAGCACCAGTCTTTAAAATAATATTCAGGGTTGATTTCCTTCAG

GACTGACTGGTTTGATCTTCTTGCTGTCCAAGGGAGTCTCAACAGTTTTCTGCAGCACCACAATTCAAAAGCATCAA
TTTTTCAGTGCTCAGCCTTCTTTATGGTCCAGCTGATCTTGGCGAGGCAGCAGTGAGTGTGGTGTGAAGTGAGATCT
TAATTCCCTGCCCAGGAACTGAACATGAGTGGCCTGGATGAGAACCAGGAATCCTAGCCACCAGACCTGCAAGGGCT
AGAGCCTAGAAACTATTTTCCCCTGGATCTTTGCCCCCAGTGAAAAAATGCATTTATCACAGAGGCAAGAACTATAA
ATGCAGCCTAAGAGCAAGTGGGAGAGCACACAGAGAAATGGTTTGTTTAGTTAAAACAGAAGACAGGGAGAGATGCA
CACCCAGAGAGAGAGGGTGTGGGTGTCCTCCCTAGTGAGGAGGAATGAAGCAAAGAGGTTATATAGCTCAGTTTTTC
CAGGTCTTTGTTTCCATTCAGTCCAATTATGTGGTTTC*TTTTTCCTCACCTGACCTACCTTGGGACCCTCCCTTGG
GCACACAGGCACCCCTCAGCCAAGATGGATTTCCTGAGAGGAGCAAGGCTCATTATGGCCTGGTGTTATTCTCTGCC
TTTTGGCCCACAAGGAGCCTTTCTGAGGAGGTGTAGTGTCTCCCTTGTCCCAAAAGAGGCGGGGGAGGGAGATCCCT
TAATCCTTTACTGAAACAGGGTTTTGCTCCTCTTTGTCCTTGCCCTGACTATTACCTGAAGGTGCTTACAAGAGAGA
AACACTGGCCTATTTACCCTGTTTATGCTGTTCCTT*GCATTTCAG*AGGGCAAACAGAAGTCTGATGTCTTAACTG
GAGCCCACCCATCTCTTGTCTCAGAAAATGCTAACAGTTCTAAGTACCCAGCCTGAAGCCTACTTCTTGGTGCCCCA
TGAAATGCAAACAGGAGGCCAGCTGTAAATATCTAACCTGGAACCCATCTATCTCCTACATCACAACTCTCACATCC
ATACATGACTACTAGAAAAACCATAGTTTTGACTTTACAG*ACTTTCATCAGCCAAGTTATGTCTCTAC*TTTTTAG
AATGCTGTCTAGATTTCTCATAGCTTTTCCCAAGGGGCAAGCGTCACTGCTGCAGTGATTTGGGAGCCCAAGAAAAT
AAAATCTGCCACCGTTTCCATTCTTTCCCCATCTGTTTGCCATGAAATGATGGGACCAGATGCCATGATCTTAGTTT
TCTAAATGTTGAGTTTTAAAGCCAGCTTTTTCACTCTCCTCTTTCACCCTCAACAAGAGGCTCTTTAGTTCCTCTTC
ACTTTCTGCCATTAGGAGCCAGGATACATAGGAG*TTTTTGCTGCAAAAAAAATGTAGTCAAACATCAGAAGATAAC
TGCTAGTC*ACAAAAAACC*AGATATCTCAAGTTAATGATTTTAG*TGTGTGGGAAGATGCAAGAGTCTGGGCTCAC
TGACCTTTGATATGCATCTTAA*TTATCTAGGGCCAGTATCTTGTTTCTCTCCATCCAGATG*CCCCTCAGGGTGCA
CATGGAGGGCAGCTGCAGTGGCTGATGGCTTGATGGTGGGCAAGAA*TTTTTTTTGTTGTTCACTTGGTGAAGGAAA
ATTCAGCATTTATTGCAGGTTCTAAGCAAGAAAGTGGGAGACAAACCTCAGATCTACCTCACCTTAGTCTATGAGTT
AGGG*TTTTTAAAG*AGGAGGAATGAAGAGGCTGGGATGGACCATTGTCTTGCGACATTCGTTAATTGTTGTTTCAG
GAGCCAAAGTATCTCTGATTATGATTCTTCAACCAGATGGTCCCATAGTAAA*GGGGGGGGGGGAC*GTCTGTTAGC
TCAACTTGCCCTGGAGAAACAACATGAGTTTTTATGTTAATGATGATATCTCTAATAGCAACTTTAGTACATTAACA
ATGTTATCAACAGCAATTTCAATCATCTGACTTTGGCTTACTAGTGTTCAGTTAGCACAGGACAGAAGTCAGAGATG
ATGAGAAAGGGAATAAAGTTTTGGATAGAGATATTAATCATGGACTCGGTTTAGAGGGACTTGGTTGCAGGTGGACA
TCCTAGTTGTGACTTGGAACTTCTGAATGAACAATGAAGTCCAAGCAGAAGTCTACCAAACACATAAGGGAAGAGTG
TTCCAGGCAGAGGGAATAGCCCACCAGGAGGCCTGGCATTCAGACAGACCCTAATGTTTGGACTAACAGAAAGTAGC
TTAGTGAAAACAGCAGACAGCAGGGAGGAAAAGCAAGAGGAGCTGACCTGGGAAAGGCAACAGAATTAAAGAGGAGA
GCCTTTATCCTGGGGACAGTGGGCAGCAGTTCAGAGCTATGAGGCAGGGGAGCACCGTCATATGTGAGTTTTAGACC
TCCCAGCATGCAGGGCAGATGGATGCAAGGAGGCAGGATAAGAGACCAGGAGATGGAGCATGTGGGGCTCAGTTGTT
CTAGTGAGAGTCCTGGACTGGTTACATACTGGAAGGAGAAAATTGATAACAGCAGGAGTTGTGTGGGATGGAACTTC
CCTGGTTGTCCAGCAGCTAAGACTCTGTGCCCTCAATGCAGGGGGCTGGGGTTCAACTCCTGGTCAGCGAACTAGAT
CCCCACGCTGCAACTCAAGATCCAGCATGCTACAATGAAGACTGGAGATCCTGCAGCTAATACCTGGTG*CAGACAC
ACACAAAGAAAGAGTTGTGTGGGAGGGACCAAGGCTGTGACTTGGATGTCAGGGGTGAGGGGGGAAGAGCTCAGGCA
GTGGCCAGGTTCTGCAGCAGGGTGGATGGTGGAGGTGCCCCTTGGGAAGTGGGTTAGAGTAGGTTTAGGCTAAATGA
GAATTGATCCTTGGACACATTAAACTTGAGGTGGGTTAAGAAGGCTGATGGACCCATGGATCTGAACCTTAGGTCAG
ACCCGGGAGCTGGGCAGACAGACTTGTGAGTCACAGGAATATAGATGACAATTAGAACCTTAGGGGCTGCCTGGGGA
GACAGTTTCCCTGTGAGAAAAGAGAACAGAGGGCAAAGCCTGAGAACAGACAACAAGGGCCTACTGTACAGCACAGG
GAACTATATTAACCATAGTAGGAAATAGAAAAGAACACAAAATAGAAC*AACC*ATCTTGTAATAAAT*CACAACGG
AAAAGAATATGAAAAAG*AATATATATATATATGTATATAC*ACATGCATGAGAACTTCCCAGTGGCTCAGCAATAA
AGAATCCGCCTGCAATGCATGAGGCACAAGAGACACGGGTTCAATCG*CTGGGTC*GGGAAGAT*CCCCTGGAAAAG
GAAATAGTAACCTACTCCAGTATTCTTGCCTGGAAAATTCCACGGACAAAGGAGTCTAGCAGGCCTACAGTTTAACA
TGGTCTCAAAGAGTTGGACACGACTGAACATGCTTGCATATATATATATTTTATATCTACACACATACATGTATACA
TAACTGAATCACTTTGCTGTACACCAGAAACTAACACAACATTGTAATCATCATATACTTCAATT*AAAAAAAAGT*
AAGGTAGGAGAA*TAATTTACAAAGCAGACTGCCTGGATTTCAATCCTGGTGTCGTTGCTAACCAGCTATGACCCTT
GGGTAAGCTGCTCAACTTCTCTCAGCCTCAGTTTACTCATCTGTAACAAGCCTGTTGCAACCACTGCATGGATGATT
CAATGAAACCACACACATGAAGCGCTTAACATAGAGCCTGGCACTGGTGGGTTTGCTGTGTACAACAGTCTGTCTTT
GGACCTGAGGTGTGCCTCCATGATGTATCTTCTTCTGTCCTATTTTCTTTTCTATCTCTTCTCTTGAGAGACTTTTG
TTGGGTTTATTCCTTATCCCTTTGATGAGAGCCCAAATATCCCATCTTGGCAATGGCTGGTACGTGCATCTGGGCCA
CTTCCCATCAACTCCATCCTGGGCTCCAGCAGCCCGCAGGCCACACCCCAACACCCATCCCGGCCTCACTAAAGGGC
ACATTTGCTTTATAGCTCCCTTTCTCTCATTTGCTCCTCACCATGTCATCAACAAGCAGAGGATGTCCTTGCTTTGC
ACCAGGGAAAGCCAGTATATCTGAGGTCAGCCAACAACAACACTCAGACTGGATCCCAGATTTAATTTTGAGTTTGA
TCTTCCATCTACAGATTGTT*CTTTGCCTACACCAGAGAAGAAGGATTGAGGCTTCCCACTCTGACCCCAATCTCAT
TCTCTGCTGCCTTTTTGAGCCTGCTGGGCTCCGAAGGAAGGGACATACAGT*GCCCTGTGATGTAATTATCATTTAA
TAATACCCAGAAGTGGGTGGCCAGTCATCAGAGGATGAAGTTGTGCAGTGGGGGCCTCTCTGGATCACTTAATCCTC

CAATCTCTTAAATACCAACAAATCTGCCTCCTGGCTGGGGCTGACAGCTGCCCCTGAGTGGAGGGGAGTCCAAGCTG
GCAGAGAGGCCAGGACTTTCTGAATGAACATAATGAAAACTCCCCTAGTCTGGTGGGATTATCAGCCCCCT*CTCCC
TTTCCTGTCCCATCAGGACTTTGAAAACACCATCTGCGTTAGCTCTCAAATTGCATGGAAACGGGGAGTGAGGCTGG
GGTGGGGGATAGAGCAAAGCTGGCACCAGTTCTTGGGCACTGAGGGGCACACATTTCCAGGGATCTTGCCTACAACT
CTCTGGAGCCATGGTTCCTGCCCACCAGAGAGGCAAATGGTCTCAGCTGCAATTTCAGTGAATAGCATTAAGGAATT
TTGCAGCTGGGGAGGTCCTTTGGAATGGTCCAGTGGTTTCAAC*TTTTATTTTAGCAGATGTTTTCACTCTTTTCCA
ACAAGAGCAGTATGAGC*TCCTTTTAATTTTTACTTATTTTTA*TTTTGGGGCTGTGCTGCGCGGCATATGGGATCT
CAGTTCCCCCAACCAGTGATAGGACCTGTGCCCCCGTGCATT*GGGAGCACAAAGTCTTAATCACCAGTCCACCAGG
GAAGTCCCATGAGCTCCTTGTTAATGCCTCACCTAGGACCCTCTCAGCCACACCTGTCTTCCCGGTCTTGCCACTTG
TTTCTAGCACAAGCCAATGGCAACCCCCGCCCCGCCCCCAAACTCTGGCCTCAGAGTACAGGAGTGTCAGACACTCC
ATGGGGTGAACCTTTGGCCAATACGGGCAGGGTTGGCTGATGAATTCCTTTCTCTTCCTCTTTCTGAGTGGACTGTC
CTAAGCCGCACACAGTTTCTTTGAAGAAGATCCCACAAGATTGAGCATCAGGGGAGATGTGTGGGGGCCAGCCAAGC
CCTCCGGCCTCTCCCACTTCTCTTTCTCACCCGGCTTTTCCCTCAGGTCTGGTTCTTCAGCGTCACACTCCCCCATA
AAGTGTTGGCACGTACACTTCTGCCTCAAGATCTGCTTCTTTGGAAACCGGAGATAAGATAACACTTGAGAAGGATG
TTCAGTGTATAAAACAGATAGAAGTGAGGCCACTCTAAGTGAAGAGAAACAGGGCCCCAGGTGTCATCTGTTTTTCT
TCCCCCTGCCAGTTCCAAGGCGATGATCGAGGTACCCTCACAAGGCTCCCCAAAGCTTTTGATCTCATTCAATTCTG
CCATCCTGCCACTGAGGACGCATAGAGTTAGAAGCAAAGAGGACTAGACTCTAGGCTTCTCACTCCAAATCTGGCGA
GACGGATCCTCATCTTTCTATCAAGTTTACATAAGGACTTCTGAAAATATTTCTTTTCCCAAAACATCTTCCCACAC
TACTCTGCCCCCTACCCCCACCCCGAAGGCTTCCCCCACTCGGTTTTATTGAGAAATAACTGACATACATCACTGCG
CAAGCTCAGGCACACAGCATGATGGTTTGATTAAAGCATATTATGAAATGATTACAATAGGTTCAGCTAATATTCAT
CATCTCTCAGGAAGGCTTTTAAGTCCCCTGTTCGACTGTAGCCACCACCTGGGAGGCATGACACACGGAGTTCTAAA
CTCAGAAGCTGGGACACAGGAGGGCAGACAGAAATGGGGCAGGAAAGGTGCAGTAATGCTGGGAGAAAGGCAGTCAG
GAAAGGCCTGATTCAGGGACATCTGCTCGTGGGAGACAGGGGTTACAGCATGAGGACAGACCCCTCGGGAAGAGGGA
CCACAGGAGATGGAGACTGGCTTTCTCTACCAGCTGGTGTCACCTGTAGGACACGCTCCTCCAGGCTGAAGTCATCC
TTAAGGAAGCTTGAAGGGTGACCACACTGGGGGGAAGAGACAGAAGGAGGTTATTATCTCCAATTATTTTCATTGAA
AAATGTTTAATATAATTGCTTTACACTTTTGTGTTAGTTTCTGCTGTACAGTGAAGTGAATCAGCTATATGTTCATA
TGTATCCCCCACCTCTTTGACCTCTCTCCCACCCCCTCATCTGACCCCTCTAGATCATCGCAGGTCACTGAGCTGGG
CTTCCTGTGTTTTATAGCAGATGCCCACGAGTTACCTGTTTTACACATGGGAATGTATATATGTCAATCGTAATCTC
CCACTTCGTCCCACTCTCTCCTTCCCTCACTGTGTCCACCTGTCCTCCCTACGTCTGCATCTCTATTCCTGCCCTGC
*AAA*TGGTTTCGCCTGCACCATTTTTCTAGATTCCACATATATGTGTTAGTAATACAATATCTCGTTTTCTCTTTC
TGATTGACTTCACTCTGTATGACTAACTATGAGCTTCCCAGGTGGTTCAGTGGTAAAGAATCTGCCTGCCAATGTAG
GAGATGAAGGGAGATGCAGGTTCAATTCCTGGG*TTGGGAAGATCCCCTGGGGTGGAAATGGCCACCCATTCCAGTA
TTTTTGCCTGGAGAATCCCACTGACTGAAGAGCTGAGCGGGCTACAGTCCATGGGGGCGAGACTGCACACACATGAC
GGACGCTAGGGCCATCTCATTAAGATTCGTGTGGCTGTGACTGCTAGAAACACCTGGGGCCAGGTCAAAGGCTAAAA
GTGATTCTATATTATAGGGGCACGGGCATCTTAAGGAATCTAGGTACCAGGGTGCATCTGGGCCTCGGGGAACAGAT
GAGACCAGGACCTGGACTCAGAGAACCCCGGAAGCCACCTCCCTCCAACTCTGCTTCTGTTCCATCCCTTTCTCTCT
CTCTGAAACTGGC*GCTCCCTGTGCCCTAGCGAACACTGGCATCAAATGGCCTCAGCCAGCAACTTCCAAGTTCACC
TCGCTTCCATCCAGGAGGCAAGCAGAGTCAGCCTGGACTCTTCTGGGCCCAATGCTGAGCTCTCGTGGAGGTTCTGG
TTGGACCAGCTTGAGGCAGGTGCCT*CCCCCGGGCCCATCAGTGAGGCCGGATGCTCTGCAGTCACAGCTGCTCCCT
CGCTGGCCAGACAGGCCAGGGAAGAAGCAGGCGGTTCATTACCTCGGGCTTCAACAAAAGCCTTAAAATAAGAGAGT
CTGTGCTCTTCCATGCCTTAATGGAAGAGATCATTCTTTTCTACATTTGTTGTAAACTCAAACCTTGTCTGGAAATC
TGCCTCCCCATTGGAGCCCAGAGCTGAATTATTCTCAGCCTCCCCACTGCCGCACA*CCCCCCACCACCACCCCGCA
TCTGCCACTGAGAGCTGGGCAAAGCTGAGAGGGTACCAAGGCCTGCCACTGTCCACATCCTGGCTCTTATGTGAATA
TTAGTAAAGTTTACTTCAGTTTCACCTTAGACTTATTTTATAAAACAGGTCAGTGGGCTCAAAAGTCTTCCATC*AA
AAGAATTTCTAAGGACATGCTCTTGAGCCTATGGTGG*AAAAGGCTGATGTGAGCTTCT*AAAACCAGAATTTATGT
GGTGCTTCCTGCTTTGGGAGTCCATGGGTTCAATGCCAGAAAGGAACTCAGGTCCTTTTCTTTCACTCTTAACGACT
ACTGATAACATCTATTGTACAATTGATATGGGCTAGGCCCTGGTCTCAGTCCCTCCCTTCTACTATCA*CCAA*CCC
CGACAACCACTGATTGTCTTTATGTCACCATAGTTTTG*TATTTTCTTAAA*TTTCATATAAATGGAACCATCTAGC
*TTTTTCTATCCAGATTTTCCCACTGAGTATGATGCTTTTGAGATTCATTTGTATTTGGGTACATTTCAAGAATTCA
GCTCTTTTTATTAACAAGTAGAATGATATTGTCTGGATGTACCATGGTTTGTTAATCCATTTGAGGGACATTTGAGT
TGTTTCCACTTTGGGGCTACTGTGATAAAGCTGCTAAGAACATTTACATACAAGTCTTTGTGTGGACTGATATCTTA
TTTCTCTTAGGTAAACACCTAGGAATGGGATCGCTGGCTTCAATGGTAATTGTACCTTTAGCTTTATAAGAAACAGC
AGAGTCTTCCCTGGCAGTCCAGCATTGGCCTAGCCTTTCAGTGCAGGGGGTGTGGGTTTGATCCCTGGTCAGGGAGC
TAGGATTCCACATGCCTCGCAGCTAGGAAACCAAACCATAAAACAGAAGCAATACCGTAGCAAATTCAATAAAGACT
GAAAATAGTTGCAAATTTAAAAAAAAAAATCTTAAAAGGAAAAACTGCAAACTGGAAATAGAAATACCATTTCAAAT
GATCTTTAATCTTAAAAAACTTCAGATGCTCAGCTTAGCATGTATACATCCTCAGTCGCTAAGTCATGTTCTGCTCT
TCCTGAATCCATGGACTGTAACCCTGCACGCTCTGTCCATGGAATTCTCCAGGCGAGAATACTAGAAAGGGTTGCCA

TTTGCTACTCCAGGGGATCTTCCTGACCCAGAGATCAAACCCGTGTCTCTTGCGCCCTCCACGCTGACAGGCAGATT
CTGCGCCACTATGCCACCTGGCCGCTTGGGAAACCCACTGTGCCTAGCAGTAGGCCTCAAAAAAGACTACCTTCTGG
ACAACGGCAGAGGTAACCACGACAGTATCACCCTGATGACAAATCAAGTTGTACGCTTCGCTCCCATGGCCTGGAGT
CTCCTCACCCCTGCGATGTGAGAGGGAAGTTCTAAACTGGGTCATTCACCCCCATCCGATGTCTCATCTTCCCCTGC
CTTTCCATCATCTTCCTGTGCAGGGCTCTGGCCTGAGAGCCACATGCAGAGGTAGGTACAGGCAGTCTATACATAAT
GTATATATTGTCTAGGAATTAAGTTTCCACCCAAAGCAAAAAGCCCTAGCTGCCTGACTTTTAATTAACTATTCATT
AGTTTGGCTACAATATCCTCCATCAAACAATGGGCAAGTCATAAGGCTGAATAGACCCAGCAGACAAAGCGGGATCC
CTTTTCTTAGCTGGATGTCTAACCAGAGCCAATTCTCAGTTACCCAATCAGTCAATAAGCATTTGCTGAACTGTACC
GGAATAGTGGACATGGAACCGATATTTCTCATAACCTTACAGTTCTTGCCCACCCACTTTGGGTACTGTCATTACTT
TTCGACAGATGATTCAACAGAAACTTGAAGATGTGATGTGAGGTTACCTAGCAGAAGCCAGGATGAATTCCAGGTCT
CCCAGCTTTGCCTTCCCTTCCCTGCGTAACCCATAAGGTAGTTGGGTTTGGGGTCGGGGTGGGGAGAGGAGGCT*AA
AACACACGAAACAAGACAATCCAATTTTTTCCCCAGGACATTCCTGAAAAAAATTGACCACAGCATTAAAGTGTCTT
TTCCCCATAAAACACAGTTTGAGTTGGGAATTTGTTCCTGATTAAGGCTTGGACATAAATT*AAACGATGAAAATGT
*AAAATGAAAATAAGTAAAGTTATTACAAACAAATCTTTATAATAACTTGTTAGTAAAACCCT*GCTCAGTCCTTAT
GAAACAGCCACAGAGCCACTGTAGACCTTAATTACACAAGGAGCCCTCTTCCACTGCAAATTTTCTTGTGCAACAAG
AAAAACTTTATGTCTTTAATATGATTTAAAACTAATATGCTAGAAATGGTGAGCAGTGAACATATAGTTAGAAATGC
TTTTATAATGTTTAGAACCGCACTGGGAG**AAAAATGCCAACTGCTTTGCCCTCTTTATAAAAGCTGGACTAACAT
TTTGTATGGATCTTATTTCTCTTTTGTCAAGTTCTTACTCTCCATGGACAGGGTCCACAGATGGGCATCAGGGGACT
TTGCACCACCTGAAATTGTATGCAACATGTTTTGAGGATACAAATACATTGTGGAGCCTCTAGCTTTCATCAGATCC
TCACAAGCCTCTGTGATAGAGCCACAAATTTGAATCAC*TTATTTCTTAGTGTTTCTTCTATTGTGAGCAGTGGCAT
TAAAATTGTTCCACTGACTTTAGAACCTTGCTGACTTGGAAACATTCTTTTAAAAAGCAGCCTGGGGACTTCCCTGG
TGGTCCAGTGATTAAGAAACCACTTTCCAATACAGGGGATGTGGGTTCGACCCCTGGGTCGGGAAGATCCTCTGGAG
AAGGGCATGGCAACCCACTCCAGTATTCTTGCCAGGAAAATCCCATGAAGAGAGGAGCTTGGCAGGCTACAGTCCAT
GGGGTTGCCAAGAATCGAACATGACTGAAGTGACTT*AACACCCACATGCCACAAGTCACAATGAAAGACCCTGCTT
GCTGC*AACAAAAACCTGAGGCAGTCAAATAAATACATATTTAAAAATAAAAAGT*AGCCTGGATCCTCTTGGTGGT
GTAATGGGCTTGATTCCTAGAAGGGATCCCTCCCTATAAAACAACTGGACCCTGAAAAAGAAAAAAATTTTT**AAA
*TGCATAGCTAGTGAAGTCAGATAGTGAAGGAAGTCATCGTTTCCTCACCACTAAAATTAAACAACACAC*AAAAAT
*CATGGGCC*TGAGACTGGAGGAAGGGGCACTGGTTAGAAAGCGAAGATTGGGCAGCCCTAAGGGCAAATGCCCATT
TTAGCTGCTTCTGGGAAGCTTGCCCCTGGGCCAGTGGCAACATCAGGGATCTGCCTCAGGATTCCACCTTAGCTCAG
TTTAAGGGATCCTCTTTTGTTACTGGCTGCACCTCTATGGATTTCCCTGGTGGCTCAGATGGCAAAGCATCTGCCTA
CAGTGCGGGAGACCCGCGTTCGATCCCTGGGT*CGGGAAGATTCTCTGGAGAAGGAAATGGCAACCCACTTCAGTAC
TCTTGCCTGGAAAATCCCATGGACGGAGGAGCATGGTAAGCCCGAAAGGAGTCGGACACGACTGAGCGGACTTC*AC
TT*CC*ACTTTCCACCTCGGAGGTTTTCTCTGCCTCACTGAGGAGAAGCGGAAACGACATTTCCCAGAACCCTCTGC
TTCCGCAGCGGCTGCAGGCGGGAGGCCATTGGTAGAGATTCCGAAGGCAGGTGCGAGGGAGAGGGCACAGTTCTCCA
GGGGAGTTGTAGACTCGTGAAAAGGCGTGAGTTCGTGGTAGCTTCAGGATGTGCTTCTTGAGGCTCCCCTGCAACAT
GCTGCAGCCTGAGAGAGCGAGGGGTGGGGGCGAGGGCCGCTTCCCAGAGGCTCTTGAGATTAGGGTGGCCTTCAGGC
AAGCCTGAGAACCGCCCTCGTGGGTGCTACAGACTGAAGCTGCAGTTGAAGGT*CCCCTTACCTTTGCTTCCCCAAG
GCTTCCGAGGGTTATGTAAACCCCTGATTTCCTGTATCCAAAACTGAGACGCTTGGCATGCCAGCTATCACTTTATT
CTCTCTTGTTCCAAATTTAC*TTTTCAATACCTGCTCTG*TGATCATGGGTGAGCCTGACGTTGTTTTTCAGTAGAT
GGCACTAGAAGGACACTGGAGGAGGAAAGGTTTCTCCTTCTGGTACTGGTGCTGTACGTTGGAGGGTGGTATGAACG
GGAGCGATTCGGGTGGTGCGCTGCCCACACACACACCAAGAACACACAATCTCTCAGTGACCTCCCAGCTCCTGCTT
GGTCTGGTGATCACCTCCCACGGCGCTCAGACACAGACAGTGGATTGCAGGCCTCCCTCCTACAGTGATGCTAAATT
CCCTCTGCACACCTGCATGCCTGACAGCCTGCGCCCTGGGCACTGCTCACCTGTGCTTCACTGGGTTGCTTCTGGCC
TTTTCAGACATTGAGGACAGGCTGCCCATCTGGCCCAGGCTACCTGTCAAACTTCTCTGCCATACAGCGGTCTGCAA
CCACATCTTCTCCAGTGCGGTCTGAACCCCAGC*TTTGGGAAGGGAG*CTCCCCTTCATAGTTTGTCCGTCCTTAGG
TACTTTCTCTCAGCACTATACTCCCCCCCTTT*CTAAAATAATGGTCC*TGTTTTATTTTGCTTATATATTGTTAAT
ATTTTATATATTTATTTGGCTGTGCCAGGTCTTAGTTATGGCACGCAGGATCTTTGATTTTCGTTAATGGCATGTGG
GATTGTTGCAGCG*AAAAGGAAAAC*AGGGGAATGAGTTTTTCTCTTTCCCTTTCCTGAGAACAAAGAAGACAGAGA
GAACAGTGAGCAGGCCTGAACATTCCCAGTTTTTTTTTTTTT
>NCL_C7
CCCTCTAAGAGGGCCCAGCGCCAGTGGGGACAGGGACAGGGACAGGGGGCC*ACCCTGCACGCAGGACTTGAACTCA
GCCCCCCGCCCAGCTCCAGGGGACCTGTCTTGGTCACCCGAGCTTACCTTCCGTCTGGGTTTGCCTCTGCCCAAGAG
GTGGGCCCACTACCATCCTGGGACCCCATCCTATGACAGTTGAGCACCTTCCAACAGTAGCCGGATGGGGACAGCAC
ACTCTTGGGACCGGGGATGGACCAGACTCCACAGGGGGCAAACTGCCTCCCCCATACTCTGCACCAGGGGCCCCGTG
GTATCAGCAGGGCTAGAGTCAATGGGGGACCCATCTCCTCCCCAGTTACCTAAAATGTGTCTTCCTGCTTTAAAGAT
CTGGGGGGTCCTCTGGACTTGGAGGCCAGCCCTGAAACAGTAGGGCTCTTCTCTCTCTACTCCTGCCCCGCCTCCAT
GACCTCTACACCTCCTTGCTCTCCCTGCCAGGCCCTTCTCCTGCCAGACCTGCCTGCCATCCAACATCCTGCTCACT

CCTTTCCTTTGTTGTCCCCACCTCCCCACACAGGTGAACCCTGTGAGCACAAGGATTTGGGGCTTCGCTGTGCACTG
CTGTATTTTTAATCCCTAGAACGGGGCTGGCACAAAAAAAGGACTCATAATTGGCAC*AAAAAAAGGACTCATAACT
GTTGAATGAAAGAACAAATTGGGTTAAATAAACTGGATTAAAATTCCCTTTAAGTTTTCCCTTTGAAAATGCAAATG
CCCCAAGGCAAGAAAGAATCCTCCCCAGACCCAGGATAACACCAGCTGTGCTGGGAATGGAA*CCAACTCTGATGAC
ATCTGTACATGGTACAGACTTAAGAATGTATAGCATCCAACTTACAAGTAATAAAAAT*ACTTTGATTATAAATCCC
ATCTAGCCAATTGATTATCACAAGTGAGTTTCTGCCAGCTCTTGTATCCATTGCCAACCTACAATTTTGATTTAACC
ATAAGTGAGGT*AAGCTGCAGGCCAATGAATGCAAC*TATTTTCCCAATGACTACACTGTCATTACATCTGATGAGT
GATCTACTGTTAAACTATTCCTCACCTTTGTGCAAATTACATCCATTAAACT*GAAGTCTTTCAAATCTTCAGCACT
ACTAATTGCAACAGAGCTATCTTTTAAAAAACTTTTGTAGACGAAGGGACATTCTTTCATCTTCCTTTGTTTAATAA
TAGTTCACAAAATTCCTAAAGGTGTACCACGTAGCTCATGCTAGCAAGAGCAAGCTGGCCCCGGCTTACATCTGGGT
GACGTCCAGGCACAGAGTCAGGTGCGTGCCTGCCACCTGCTCTCTTGCTCACAGGAGCCTGATGAGGCAGTGCCATT
ATTCCCACTTCACAGATGAGACCACCGAGGCTCAGAGGGGACTGGAGCCCAGGTCCACTGCCTCCAAGGTCCCTGGT
CTTTCCATTCAAGCAGTCTTTCTTTCCTGAACGTTTGTGGTCCCTTTCTTTGTGTCAGACTTTAGGATACAAAGGGA
AACGTACCTTCTTCCTTTCTGGGAGATTCAGAGGCCCCAGAGGGAGGGTCAGTGAGCCCCAGCACCATGTTTGGT*G
GGGGGCACACTAACAGCTGGATTCCAGCCCTGAGATTCATCAAGCCTTGTGTTAAGACTGGGTAGGAGAGGCCCCTA
CCCTGGGCCTTGGATTTT*AAAAGTCCCTGCTCTGGACCCCGTTGAGGGGTAAGAATATCAAGGGGCCTATGGGGAG
GCGCCTGCAACCCACCTACCCTCTCTCCCTTATAGACGCTGCTGGTAACCAGAACTCTGCAAGTCCTGCCCAGGTAA
CCATCCCTCAGTTCATCTCCCTGCTTGAGCTCAGCTGAGCCTCCCCAGCAGGGTCTGGCCCCTGGCTTCTGTGTAAC
CCAACTCTGTCGGCTTGGCCTTGTGGAAACAGAACCAGGCACTGTGGGAAACTGAAGCTGCCTGGGGCAGGTTTGGG
GATTCACAGAATAAAGTGCTCCAAATGCCAACGCTAGGCTTGCCGTCAGGGCAGAGGGATCTCTGAGGCCACCTCCA
CTGGCCTTCACCCTTTACAGATGGGGAGACTGAAGCCCACAGCCTAAACAAAGCACCCTTTCCGAGAGCACAGCCCT
GGGCCCTGTTCTCTTTCTCAGCCAGTTACAAAGCCCCCTTTTCAGCTGGGGCCCTCAGTCACAACCAGTTGGTGTCT
GTCCGACCAGCCCCTCT*CCCCCTCCG*CCCCCAACTATCAAGTTCTGGAGGAGAACCGAAGCGCTCCCCCTCCTGT
CAGGCCCTCTTGCAAATCACACTGACGACTGTGTACTAGGCACAGTTCTCACCTCCAAGCCTTAT*CCCCCGCCCCT
CCACCAGG*CCCTTCCTGGTAGCAGTCGGCCCCTGGCTTGTGGGTGGGCCCAGCTCTCCCACCAACAAAACATGAGC
AGAGCAAATTACCTAATGAGACCCATATTTTGCTGCGAGTTGCTAATGAAAGAACAAACCAAACCTACCATTAATTT
GGTCCTGGGAAAGAAACTTGGCCTGCAGCAGAGAGAGAACAGTCAGGAAGGGACTCCATGAAAGCCCCCTGCCTCCC
CAGCCCCGGGCAGCCTGGCCAGGCCACACAGATCCTCACCATCGCGGGGCCTCCAGCTGCCACCCGTGGGCTCGCTG
CTCCCAGAACTGCCTTCCTGTTCCCCTTCTTCCCGGCTTCAAGTCGAAAGTCCGCCCAGCCGGGTGGAGG*CCCCGG
TAATAAATCCCTGGCTGCCATGGAGACCAGGCCCGCCGCCAGGGGTCAGATTCCAGCTTCCTGTGCTCTGTGGATGC
CTGGCTGTGGGGTGGTGGCAGGCTGGATTTTGAGAGATGGGAGCTGCCCGAGGGCTTCGCCGGCACTTGGTCCTGCA
CCAGGAAGCTGATATTCTGTAGCCAGAGGACAAGGGCCGCTGTTCCTTCCTGGGAGCCCTCTGTGGACACTAAGAAG
GGGTCACCCAGAGAGGGGATCGCTGGGT*CCCCCAGGCCTGCCTTCAGGGGTATTTCTCTCTGAGGATTGCATTAGG
GGCTAAATAGAGTCTCCTCCTAGGTGTTGGGCACTTTGCACATCCCCTTTAGGCACGTCACACACCACCTCATTAGT
TCAGTTCAGTCGCTCAGTCGTGTCCGACTCTTTGCGACCCCATGAACCACAGCACGCCAGGCCTCCCTGTCCATCAC
CAACTCCTGGAGTCCACCCAAACCCATGTCCATTGAGTTAGTGATGCCATCCAACCATCTCATCCTGTCATCCCCTT
CTCCTCTTGCCCTCAATCTTTCCCAGCATCAGGGTCTTTTCAAACGAGTCAACTCTTTGCATCAGGTGGCCAAAGTA
TTGGAGTTTCAGCTTCAACATCAGTCATTCCAACGAACACCCAGGACTGGTCTCCTTTAAGGTGAACTGGTTGGATC
TCCTTGCAGTCCAAGGGACTCTCAAGAGTCTTCTCCAACACCACAGTTCAAAAGCATCAATTCTTCGGCACTCAGCT
TTCTTCACAGTCCAACTCTCACATCCATACATGACTACT*GGAAAAACCATAGCCTTGACTAGACGGACCTTTGTTG
ACAAAGTAATGTCTCTGCTTTTTAATATGGGAAGGAAACCCTGCGTGACACCACAGCCCAGTGACAGCCACCTCCTC
CGCCAGAACCCTGGATGCCACACACAGGAGAGAAAGGCGCCTCTGAGGAAGACCTCGGTGCCCATTCCACAGCTCTG
TGCCTCAGCATCCTGGCCTTATGTTGAGAGTTAATGGCCAGATGGTCCA*TTTTCAGCTTGAGTCAGTTCCTTCCCA
TTCTTCTGGGGGTGTGGTAGGTGCCCATGCTGGGGTCTCTGGCCCCACCACAGGCAACTCA*GGGGACTCAGAGGGA
GACCTCTTAAGTGACTTTCCTGGTCAGGAAGGCTCCCTCTTTTCTTCCTTCCCCTCCACTCCCCACTGCACTTCCCA
GCCTTTCCAGCAGAAGCTGCTGACTTTCCCCATTTCTCAGTGGTTTGTCCCCAGCCCACCCCTCCCCACACCCACTG
AGGAAGTGAACGGGGGATGGAGCTGCAGCGCTGGGCACAGCTGGGTTCTCCCCCTCCTTCTCCGAGATCCCTGCTGT
ATCCTGTGTCCACGGCTGGTGGCTGCCGTGCCTTTGGGGCCCCGACTCCAGGGCCCACCCAGTCTGTCGGGACTGGC
AGGTACGGCCACACCGGTTGTGCACTGCACTAGGGCTCCTGGTCTGGGGGCCTGGAAAGGGCTGCATGCGTGCAGAG
GGAGCTGGTGGTAATCTGTAGAAGTTACATGGATCAGGGCAGGGGGGCTGTCTTTGTCTCCTGACCATTTCCTTGTG
AAAGGTTCTCTTTTCCTCGACTTCACACCCATTTCTAGACCCAGAAAGTTTTAGCTGGCGGTTTCTGACTCCTCCAA
GTTGGGGCTGAGGAAAGGCTGGTGGGAGCAGGAAAGGAGGGGAGTTTCTATGCGACTTACAGGAAGAAGAGGAGAGG
GGACATGAGGCGGGGCGTTGGCAGGGCCCGGGGCACAGAGCGGATGCTTGAGGGGAAGCCTGAAGAGCCAGGCCCGG
GCTCGCTTGGAATCACCCTGTGTTT*GGAAAGGGGCTCCTGGGGGAGATGTGATCGTTGGGCTGTACCCCGAGGTGC
TGCGAGCCATATAGTAGCCCAGCCTCGAGACCAAAGAGCCGAGAGACAGACAACAACTTATCAGATGGGGGAATCTT
ACGGATCTGAAGCCACATCCTGGAGCAACACCCCCCTGTGGACGGCAGACAGCAGGCAGGACACGGCGCGCCTTCTG
CTGCTCTGGAGGAGGAGGCGGCTACCAGTTATAGAGAGCATTGACTTCAGAAGGCTCATCAGTTACCAGGGAAACCA

GCTGGGGGCAGGGGCAGGCGTGTGGACAGTCACAGATAAAGGGCCTGTGATGAACAGGACTGGACAGTCCTGTGTGT
GAAGTAGGGACTGGTGGAGCAGGAGATACACAGACAGCAAGTGAACGGCCACCTTGAGTGGCCTGGCCATACACACC
CACACATCCTGCATGACTCTCATGGCCCTGCCCCTCTTCTGAGATATCCCTGGGGTCAGGGATGCAGAGGTGCGTTG
CAACCGCAGATCACAAATAAAATGACAGACTGTTGTTTTTGCTGATTAGTCTCTAAGTTGTTTTCGCTGTTTAGTCT
CTAA*GTTGTTTTCGATTCTTTTGCAACCCCGTGATGACAGTAGCTAGCAGAATACCAAGAGTTCCCAGGTGGTGCA
GTGCTAAAGAATCCGTCCGCCAACACAGAAGACCCAAGAGACACCGGTTCGATCCCTGGGTCTGCAA*GGTTCCCTG
CAGTAGGAAGTGGCTACCCACTCCAGTATTCTTGCCTGGGAAATCCCATGGACAGAGGAGCCTGGCGGGCCACAGTC
CA*GGGGGTCACAGAGTCGGACACGACCGAGTGCACACACAATCAAGAGTAAGACACAAAGAATGCTAGATAGTAAC
TCCCTGAGACGCGAGCCAACTTTGGGTGAATTTTTCATGGAAGTCCAGAGGAGGGAGGCCCTGCAGTCAGACTCGCC
AGTGGTGAGGCCAGGACTTCCCTGATGGCTTAATGGTGGGAAGTCCACCTGCCAAGGCAGGGGACACGAGTTTGGTC
CCTGGTTCCAGAAGATCCCACATGTCTCGGAGCAACCAAGCCCCGCACCATAACTACCAAAGCCTGCACACTCCAGA
GCCTGGGAGTCACAACTAATTAGCCCCTGTGCTGCAACTGCTGCAGCCCGTGTGCCTAGAGCCTGTGCTCCACGCCA
AGAGGAGCTGCTGCAGTGGGAAGCCCACGCTCTGCAACTAGAGAGCCGCCCCACTCTGCAGAGAAAACCCGCGCACA
ACGCAGACAGCGCAGCCTTAAACCAACAAAGTGGTGAGGCCACCACTGTCACCTGGTCCCCAGAGAGATTCCCTCTG
CTCAGGCTAGGGACGAAACATAAGATGCTTAACAGGCACAGTACAGGGGAGATCGTGGCCGTGAGGCAACACACAGG
CAGTGACAGCATTCTGAGGTCACACCACCCCAGCCTGTGGTTGGCGTCCCAGCCTGCAGTCCCCACAGCTGGCGCCG
AAACCTGGAGCAGGGGATGTCTAACAGATCGTAGGGGTCATAGAATGGTGACTTTCTCCAGGGGAGGCCTGGATTAG
TGACCTCAGTAGCGGTCTCAGGTAGGGCTGAGTGGAGGACAGGTGAAGACTGTAAGACCCTTTCTAATCCTAT*CCC
CCAAGAAGTGGCAAACCCACACCACCAGAAACTTACCTGCCAGTCCCAGGCCGTGGTATACGGTGTTCCACTGGGAG
CAGATCCTGTGGTCAGGAC*AAAAACAGTCAGACTTGGATGGTGCTGACTAGTTGCTTTGGGGTTACCATGGCGTAG
TCTCTGGGTCCATTGCTCCTGGCATATTCAGTTATAATGTGGAAGGCTCAAGCTAGCCTCCTAGTTCACATGTCAGA
GAGCAGATTTCCTGTCTCACTTGCTGGCTAAGTAAGTAGTCCATCCATCCTCCGCTTAGCCCGGCAGCAGGGGAGGA
CGGTGTAGGGCTGGTGGAAATGCCAGTGTCTGTCATCTTTTACAGGCCAGCTCCTGAGTCTCATGGCCCATAAACTG
GTTTCCTTGGTCGCTATTGCAGCGCTTTCCCCATAGGGGTGTGCAGCCATCAGCAAGCCCTCTCGCGGTTTCCACTG
GATTTATCCAGATGCTGTTGGATACAGGCCAGTCACTGCTGTCCACCAGCGCGATCACCCCAGACCCCCCTATCAGT
GTGCCAAGCCTCTCAAGGCCAGCCTCTCCTG*TATTTCTTGGGAAACATTTTCAGGAAACTTCCTAAAGAGATCTCA
GTCAAGAGGGACAGGAATCGCCAAGGCTCCAA*TAATTTGCTGATCTCTTTTCCATTCTAAATATTCTTTCATATC*
TAATTTTATTCACTTTTTTAGAAGTTTTCTTATCAAGCATACTC*AAAACTGTGTAAGCCTCAGACCCCACAATCAG
ATGTGTGCCCACTTAGGTGATATCCATTAAGTCCTTGTATGGAGAATGGGCCTGGGACAGGGTTTGGAGGAAACACG
GAATCAACGTAGGGGAGAATTGAGGTTACTAGGGACGTAGGGGCCCTTGTGCTTTGAGAGGCCAAGGCTGGAGCTGG
CTTACTGAATTAAGATTGTGCAGATTGTGACCTGCTGAGTGAGGGGGGGCGGCAGGCAGCACCTTCAGCATCCAGCT
CCCACTCTGGGGGAGTGAAACCCATGGAGTCCTTTTGATTGAAGGCAGTTGGGCTTCCCTGCCAATGCAGGAGACGC
AAGAGATGCGGGTTCAATCTCTGGGTCAGGAAGATCACCTGGAGGGTCCCTTGGAGAAGGAAATGGCTACCCACTCC
AGTATTCTTGCCCAGAAAATCCCGTGGACAGAAGAGCCTGACAGGCTACAGTCCATGGGGTTGCAGAGTAGGACACA
ACTTAGCAACTAAATAAGTTTCTA*TACATCCCCACCAAGAACATGAAATCACACAATTCATTATTTCAGATCCCGG
AAGCCA*GGGGGCCCAACAAGCTAGAAGGGCCATCCTCATCCCAGGTCCTGAGCAGGAGCTATTGAGCAGGACAGCA
GGCATCTGTTACTCCCAAGGCGGTGGGGGCAGAGGTCACTAACTTCTCATGGGTGTAGAAGGTGCCTGGGAAAAGCA
AAGCAGGAAATTGTTGCAGGTATCCTGAACTCCAGTCTTCACGTCTACACTCTGGATGTGAGGAGGGTTCAGTTCAG
TTCAGTCACTCAGTCGTGTCCAACTCTCTGCGACGCCATGAATCGCAGCACGCCAGGACTCCCTGTCCATCACCATC
TCCTGGAGTTCACTCAAACTCATGTCCATTGAGTCAGTGATACCATCCAGCCATCTCATCCTCTGTCGTCCCCTTCT
CCTCCTG*CCCCCAATCCCTCCCAGCATCAGAGTCTTTTCCAATGAGTCAATTCTTCGCATGAAGTGGCCCAAGTAC
TGGAGTTTCAGCTTTAGCATCATTCCTTCCAAAGAACATCCAGGACTGATCTCCTTTAGGATGGACTGGTTGGATCT
CCTTGCAGTCCAAGGGACTCTCAAGAGTCTTCTCCAACACCACAGTTCAAAAGCATCAATTCTTTGGTGCTCAGCTT
TCTTCACAGTCCAAATTTCACATCCATACATGACCACTAGAAAAACCTTAGCCTTGACTAGACGGACCTTTGTAGGC
AATATAACGTCTCTGCTTTTCAATATGCTATCTAGGTTGGTCATAACTTTCCTTCCAAGGAGTGTCTTTTTAATTTC
ATGGCTGCAATCACCATCTGCAGTGATTTTGGAGTCCC*AAAAAATAAAGTCTGACACTATTTCCACTGTTTCCCCA
TCTATTTCCCATGAAGTGATGGGACCAGATGCCATGATCTTCGTTTTCTGAATGTTGAGCTTTAAGCCAACTTTTTC
ACTCTCCTCTTTCACTTTCATCAAGAGGCTTTTTAGCTCCTCTTCACCTTCTGCCATAAGGGTGGTGTCATCTGCAT
ATCTGAGGTTATTGATATTTCTCCCGGCAATCTTGATTCCAACTTGTGCTTCTTCCAGCCCAGTGTTTCTCATGATG
TACTCTGCATATAAGTTAAATAAGCAGGGTGACAATATACAGTCTTAACGTACTCCTTTCCCTATTTGGAACCAGTC
TGTTGTTCCATGTCCAGTTCTAACTGTTGCTTCCTGACCTGCATATAGGTTTCTCAAGAGGCAGGTCAGGTGGTCTG
GTAGTCCCATCTCCCTCAGAATTTTCCACAGTTTATTGTGATCCACACAGTCACAGGCTTTGGCATAGTCAATAAAG
CAGAAAGAGATGTTTTTCTGGAACTCTCTTGCTTTTTCCATGATCCAGTGGATGTTGGTAATTTGATCTCTGGTTCC
TCTGTCTTTTC*TAAAACC*AGCTTGAACATCTGGAAGTTCACAGTTCACATATTGCTGAAGCCTGGCTTGCAGAAT
TTTGAGCATAACTTTACTGGCATATGAGATGAGTGCAATTGTGAGGTAGTTTGAGCATTCTTTGGCATTGCCTTTCT
TTGGGATTGGAATGAACACTGACCTTTTCCAGTCCTGCGGCCACTGCTGAGTTTTCCAAATTTGCTGGCATATTGAG
TGCAGCACTTTCACAGCATCATCTTTCAGGATTTGAAATAGCTCCACTGGAATTCCATCACCTCCACTAGCTTTGTT

CATCGTGATGCTCACTTGATTTCACATTCCAGGATGTCTGGCTCTAGGTGAGTGATCACACCATCATGG*TTA*TCT
TGGTCATGAAGA*TC*TTTTTTGTATAGTC*TTTTTGTGCAGTGAGGAGGGTTACCACACTGTAAAATGCCTCAGCA
GCAACTCTGAAGAACAAAGCTGCTTTCCTT*AAAAGCTGTACAGCCTACGATTGTGTGTCTCCTGTTCAGTGCAGAT
GTCTTGTGGGTTCTAGGAACGGGAGACTATAAAGCCTCTGGTGGTTGGGTGGTGAGGCAGAGACCCTCAAAGATTAA
GGGCTGAGGCCTCCAGGAGTCATGCGACTTGCCCAAGGCTAAGTGCCAGCCCCCCAAAAACCTGCCCCCACCTCCAC
CCAAAGGACACCCTCTCTGCTTCAGTTTCTAGAAGGCGCTAGCAGCATGGCTT*GGGGGCTCCATCTGTAGACGCCC
TCTGCCGGCCCCAGCCCCACAACCCTCAGCCTCCTCCAAAGAGAGCTGGGACCCCCAAACTCAGGATCCTGAGTGGC
ACAGATGGCCAGGGGGCAGTGGCAGCACCTGCCTTCAGCCAGACAGGGCCGAGCAGGCCCTTGGGGTCAGCCCAGGG
CTCCTGCCAGGCCCCAGACCAGCCACTTCCCTGCTCCAGGGTCCCTCCTCCTCCTGCCTCCCAAATTCACACCCCAA
GGAAGCAGCAGCAGCCGCCAGTGGGGCTTGGCTCCACACTGCCTTTATTCACACTGCTCACCCCTGATGCTGCCAGC
TGACAGGGACTGACCGGGGCTGGGGGACCCCTACATGAAAGGAGGATTTGAGGCAGAGTCCCTGGCAGCACAGGAGA
GGAGAGGAGGGCAGGGAAGGCGAGGTAGCCCCCCTGAGCCCCTCACAAGCATGAAGGCCAGCAATGCCAGGAAGCGA
GGCTGTGTGCCTGGCCCTGGTGCTGGGGTGGGGGGCATCTTGCCAGAACAAGGCCCCCTTTCCTGGGAGTAGGGCCG
CTGGTGGCAGGTGTGGATTCAGGCTGAGACACGCTGACGAGAGCCCCACGATGTCTGCCTCTAGTGAGTGCAAGTAC
TCAGAAACCTTGAGCAGGTGAACGAAGCAGGTGCCACGATCCTGAAGGCGGGACAGAGACCGAGTTCCCGGGAGGGG
AAGGTGGCTGGGAACGACCACTGCCA*GACCCCAAGATCAGGCCGGCAGTTGGGGGAGGGCTGGCTCTGGCTCCCAG
CCCCACGCAGCCTGCAGTGGGGTGAGGATCCGGGGTGCACATACCCCTCCCAGGGTTGAAGAAGGCAGGGACAAGGA
GGAACCATTCACCCCTCCCAAGGCTGCCCTGAGAGGAGACGGCAACCAGCCCACAGCCGACTTGGGACCCGTCGGCC
CGTCTGCTGCGAACATGGCCCAGGGAAGGGGGTGCATTGCAAATGAGTACACAGCAGGTTCACGGAACCCCTGTGGC
CTGGCCTCAGGGTCAGGTGGTCTGAGGGGT*GAGCCGGGCGGTAGGGACAGGCTAAGTTCCCTCCCAACCTCTGCCC
CACCCAGGCCTCTCCCCTGCGGCTACCAGGAGACACCCAGGAGGAAGAGACAGCCCTGGGCAGTGAGAGTGAGAAGG
CCCGACTCCAGCGTGCACGGGAAGGCCGGGGCCAGCTCTTGGTCTGGACACACTGCCTCACACACACACACACA*CC
CCCCACCCCCA*CCCCACCCCCTATCCACAGGCACCAGGCATACTCTCACCCACACTCTCAGCTGCGGGCTACACAC
ATGTGGGTGCACACAGGCACACACTCCGCACACTCAGACACTTTCGGCCGTCCATCCCCAGTGGGTGCACACACACA
CACTCCGCACACTCAGACACCTTCGGCCGTCCATCCCCAGTGGGTGCACACACGCACACACTCCCGACACTCACTCT
CATCCATCCATCTCCCACCCACAGAGCACCAGGGGCCTGTGCCCAGGACTCAGTATTGGCTGCTGACGTGGAGGGTG
TAGAAGGCCCAGGGCTCGGGGACATAGATGACGCCCGGGTACTTCTTCCTGAGGACGGGGTCATTCCACAGCCAGGC
GACCCAGAGTGCCACGAGCAGGAGGGCGAGGGCGAAGGAGTAGAAGAGGAAGAGGCCGTTGCTGTCGAGGAAGAGGC
GCTTCCGCTTCTGGTGCAGGACGAGGGCCAGCATGGCAAAGGAGGTGAAGATGAAGAGGATGAAGATCTGGCCCTCG
GTGACCAGGTACCTGAGGAGGCAGGCGGGCAGTGAGGGGTGGGGGCCCTCCTGCCCAGCGTTCTGCCCTCTACCAAG
CGCCAGGCTGCCCT*CCCAGCCGGTCAGAAGAGCGCATCAAGGGAGGGGCTCCGCTCCGAGCCAGGGCCAGTACACA
GCATTCCGATCTGGCTTCTAGAATTACTG*TTTTTCAAAAAGTCGAGTTTGTTGCCAATATTTAAATAACAGATTAC
CTGCCGACGTGGCGTTCTGATTTCCCTGCTAACGTTAGGTTTATATTCCCACATGACAATGACTGACGAGGAGAAAG
CAGCTGCTCAGAGACAGCATGGCCCCGCAGCTTGCCACGGTCCCCACCACTCCCTCCTGCCTCACCTGATGCCTACT
TGATGCATTTCTATCTCTTGTCTAGTTCCGAAAGGCATCTGAGTTTGGGACGCAGTTTCAGTGTATGGGGTTGTCCT
GACTGTCTCCTTGAAGGCTGGGTCAGTCTCATTCCCCAGGGTTCCCTCCAGCTCTCAGGGCAGTAGCTGGCATATGG
CCAGGAATGCAGTAACACTTGCTGGAGGGCTTCCCAGAGGTTATCTCATCCACTCCCCTGCCTCCATGCTGACCCAT
CCCACTCAAGCAACTGCATAGCTCTGCAACTGAGTGGCTGAGTCACCTTGGGAATTTACATGGCATCTTGGAGCCTT
TGTTTTCTTACTTATAAAGGGGGCATAAATGGCCCCAGCTCTCGAGGTTACTGATGAGAACTCTATACTTGTAAAGG
CACATTGTTCGGTGAATGGCCCCCTGTAGCTGTCACCGCATTGACACACCATGGGTAAGAATGGGCTAAGCTTCTTC
TCACTTCTCACCTCTTTCCCTCCAGCTGTGGTCGTCACCCATTCCTTCTAATCCTGTCCTCAGTGGGCATGTGGATT
CACCTTCATGAACCCTCACTGCTGGCCAGGGGCAGCCCCTGGCGGTGGGGGAGAAGGCCCCAGGAGAGGAGGCCTTT
TTCATTAGGCTGGAGGCTT*CCCCCCAGAGAAGGGCTAGGACACCCACCCTTTCTAAGACACCCACAATCCCTGGGG
AGCCTGTCTCCAGAGCAGAAGTGACAGAGGAGAGGGAAAAGCCTGGCTGAGGCCTGGTCCTCTGGAATGGGAAGAGG
AAACCAGGTGGGAAACCCTGAAGCCATGGAATTCCCCCGTAAGTATCCTTGTCGTCTCCATGAGGCTTCATGGTTTA
CGAGCCCCTCTCTCTCCCTCTGGGTTCTCACCATGTCCCAAAAGGGAGACAGCCTGCTATCATCCTCATTTCACAGA
GGGGGACATTCAGCCCCCACCCCTCCCACCCTGCCAGGCTCTACAGACTTCTGCTCTACAGCCCTCCCCGGGC*GTT
CCCGATGGCTCAGTGGGTGAAGAATCTGCCTGCAACACAGAAGACACAAGAGACGCAGGTTCGATCCCCGGGTTGGG
AAAATCCCCTGGAGAAGGAAATGGCAACCCACCCTAGTATTCCTGCCTGGGAGAATCCTATGGACAGGGGAGCCTGG
TGGGCTGCAGTCCATGGGCGCACAAAGAGTTGGAACGACTCAGCAACTAAGCACAC*ACAGAAATTATGCTTCGATG
TGGATTCTATGGGCCCTGAGGTTACTCTCTGAGTGTAGAGAGGCCACACTCAGGCCCGCCTTCCTGGTCACTGAGTG
GCGGCCAGGAGGCACTGTCCCAGTCCCCACTGAAGAGGAGGGAGCAGGGACCCCAACAGGGAGGCTCACAGGGCTCA
CACAGCAGCTCTGCGGTCGGAAGAATATTGTCACTGACTGGATCACTAGGGGG*CCCCCT*CCCTGCCACTCCAGGT
CCGCCATCCACTCACCAGTAGTACAGGCCGCTGGGCACCACCAGGAGCAGGGCTGCCCCTGGCATTGAGCTCTCAG*
CTTTGGTGGGAGTAAAGCAGCCGCTGAAGTACATGAAGAGGATAAGGAAGAAGGGGATGTACCTGCAATAGGGAAGC
CGGTGCTCAGCTGCTTGGCCGCCCGGCCGGCCAGCCCCCGTCCCCCCACCGTCCTAGGACTGTCTGCCTGTACCACC
CTGTGCGGTCTCACTTTGTTTGGGTGAGGGCACTCCCAGCCTCTCCCGGGC*GGGGGGTCTAGGAAGGAAGATGAGG

CCTGCAGATCCGAGCCATGGGCAGGGGCGTGGGGCCAACCTCCTCCCACACTGTCCCAGCCACCTCACTCACTCCCC
GTGTGTGCATGTCTCTCTCACTCTCTCTGCCCCACACACTCACCCCAAGACCAGCAGCCCCCAGCCCCTCACTGCCT
TGGGAAAGGGTGCACCCTGCAGGGTTGA*GGGGGGAGAGGCCCAAGGTCCCAGTCT*GGGGGGAAAAAAGGGGTGAG
GGGAGGAACAGCACCTGCTGGGGACCTCCCAGGAGCTCTGTGAAGCTCCTTCACTCACACTCAGTCCCCAGCCTCAC
CCCCTTCCATTCAAAGCTCACCCAAGTACCACCCTCTCTCCT*CTCC*ACACACACACACACACACACACACACACA
CACGCT*GCT*CCCCACACGCCTGT*G*CCA*GCT*CGCACAGCACCCTCTTTCTGTTCCCACCCCCACTCACCTCC
TTCCCCACCTTCCCCTCAGTGGCCCCCACTCCCTCCGAGTCCGGCCCTTTCATTTCTCCAGGCCCCTGTTCATCTCT
CCTCCTCCCTTCCCCAGCCTTCCATGGGAAGGGTCTCCTGGCACAGGCCAGGGAGAACTGGCAAGAGGCCTGTGTGT
GCAAGGACACCCTGACTTTCTAGGCCCCAAAAGAGAACCCAGAAGGACAATGTAGGGCCCGCGTGACACGAGTGAGC
CCACCTCACTCGGGAGTGGGGCCACCATGTCCAGGGGTGGCTCTGGCACTAGCTGTCAGCCCATGGGCAAAGCCCCT
CCCATCGGCCCCAGCTCATCCTGGGATAAAGTGGGCTCTGGAACAGACGACTGCCTGCAGCTCGGCCCGCATCTCCT
GTGGCTGAGCCCACATCTGTGAGGTTTCCTCCGCTGTGGGGTGAAGGGGTGGGAGGAAGGAGCTCCAAGGGGAGTTC
GGAGGAGGAGGCGGCACTAAGAGCACCCCTACTCCCACCTCGCCCCCAGCCTGGTTCAGCCGGCCCTCCT*CAGCTC
CCT*CCCTGCCTCCGGGTCCTCATCCTCATCCCCA*GGTGGGGGGAGTTGTTTACCAGCGCTGCAGCCCCGGCCCTG
GGAGGACAGGCGCTCGCTGCCCTCTAGTGGCCGCCGCGGACACTGCGGGCCGGTGGGCACGGAAGGCGGCGAGGCGG
GAGCCAGAGGAGCGACAGCGCCCCCTGCTGGGGAAGAGCACGCCATTCCAGCGTTTGGGCTGGACACTTGGAGACCC
CAGGGCCCCTTGTTCAGCGTCACTATGGCCAGCGGCAGAGCTGGCCGGGGCCATGGGCCCCATCACTCACCACATGG
AGTGGCCCAGGTACTCATCGTAGTAGTAGAGCAGCTCAAAGGAGTCGATCT*GGGGGGCAGCAGACGCCAGGC*GGG
GGGCAGAGAGGGTGGGGCTTCAGGAAGGGTGTCGGGCCGCGCAGCCCCCGGCTACCCT*CCCCCTGCCTGCCCATGC
CTCCCTTAACT*GCCCCCAGTCCCTCCCTATGGGGAGGAGGGGATC*TGGGTGGGGGCATCCTGGC*TGGGGGTTGA
GGGTGGGTTGGGGGGGC*GACCTGCTCTCTCTGGGGCACCTGTGATGGGTCGGCAATGTCTTTCGCCTCTAGGACAG
ACCGCACAGCGGAGGTGGGGGCTCACCAGCGTCTCTGGCTTGAGATTCTTGATGATGGGGTTCTCACGGACCGACAG
ATGGTTCTGGTAGCCACTGAAGATCAGGCGGTGGTTCACTGAGTCGCCCACCAGGTGGATGCTGGCCCCCATGATGA
AGGTGATGATGCTAACGTAGATCAGGGAGCGTGGCAAGGTGCGGGGGCACCGCTCGATGAGCTGAGGTCAGAGGGAG
GCCCCACAGCATGACCCCTGACACCATCACCACCGGAGACGCCCTCTGCTGGCCC*CGCTCACACCTGGAGTGGGAT
GGATGCCACCAGCCGGAAAGCCTCTCTCACACGGGTTCCCTCTAGGGGAGTCAGTCCAGCCTCCATCATCTTACACG
ATCCCTGTGAACCCAGCAGGTGCTGGGGAAGAGCACCCTCTTTGGCGCAAGAAGGAACACACTCCGCGGGGCCAGCT
CATTGCTCAAGGTCAAACACACCCATATACTGTGCTGGAGATGACAGCTAGGCCTTCCATAGAGGCAAGCCAGGCCC
TGACCTCCCACCACCCCTAATTCTACTTCTGGGTGACTGGGCTGGACTGTTGAGGCTGGGCTGCTCCCATTTCTGAG
CAACTGAACCATGCAAACAAAAATGTGGGGCAGGTGTTTCTGAAACTAAGTTGTTTAACTAAAGACCCCTTCAAATT
CAGAGACTGCACCTTCCTTACTTTGAAAGAAAGAAAAT*GAAGTCACTCAGCTGAGTCCGACTCTTTA*TGACCCCA
*TGGACTGTAGCCTACCAGGCTCTTCCATCCATGGGA*TTTTCC*AGGCAAGAATACTGGA*GTGGGTTGCCATTTC
CTTCTCCAGGAGATCTTCCTGACCCAGGGACTGAACCGGGTCTCCCGCATTATAGGCAGACGCTTTACCGTCTGAGC
CACCAAGGAAGTTCGGAAATGTCTCTAAAAACACAAGGATGGGGATAGTAAATAGCAGTAG*TTTTAAGAGCCTTTA
CTTGGCAAAA*TGC*AAAGCC*A*GGACCTACGTAGGCGAGAGAGCCC
>NCL_C8
CTTCCCTTCCTTGCCTGCTTCCCACGTCCGCCTCCCCTCCCTACGAGTGGCCAGACCTAGCAAAAAAAAAAAAAAAA
AAAAAAAAA**CCAGGGCTGTCAAGGTAAATTGAGTTTCAGATAAACCTCAGTCCATGAGGTCACAAAGCGTTGGAC
AGGACTTAACAACAACATGCTTACAAAGGTATTTGTCGTTTATCTGAAATGTATGTTTAACTGCATTTCTTGCCTTT
CATCCTGTCAACCCCATCCCCTTTCCCTCCTGACCGCCCTTCTGCCGCCTCCCTCCCTGCTCCTGTCCTCAACTGAC
TAAAACCACAGTGGTCAAGAACCCGCCTGCCAGTGCAGCAGACATGAGACGTGGATTCGATCCCTGGGTC*GGGAGG
ATCCCCTGGAGAAGAGCATGGCAACCCACTCCAGTATCCTTGCTGGAGAGTCCCATGGACAGAGGAGCCTGGTGGGC
TGCAGTCCATGGGGTGGCAGAGTTGGATGTGACTGAGGCAACTTAGCACACACAGGAAAACTTGCTCGCTCTCTTGT
TCCAGGTCCAGCCCAGGGAGACGAGGAGCCCACAGCAGGGCACCAAATTCCTAGGGGATCTCAGCAAATGACAACTT
AGGTGGGAAACAGAATAATAATGCCTTGGGGCTTGCTGGAGGACCAGAAACCCGCTCTGGGAAGGCCTGGCAGAGAG
GCCATCATACGTGGTGACAGATATCACAACTGGGAGCTTCTGAGTCCACTGCTGCAGCCGCCGGTCTCCGGGCTGCT
GAACTCTCAAGCCGCGCTCTCACACA*GGCCCAGGCCCCACTCAAGACTTCAACAGGCACTGCCTGGACGTCCTCCC
AGCTCCGCGGAGATCTTCCTTACTGCTCTCTTTGAACCCTCTCTCCCCTCTGGAGGTGAAGTGGCCTGCAGAGCCAC
AGCTCCCCAGCCCAGTGGAACCAGCCCCCACCTGCATCTGCCCTGACCCCCCAGCCCTGGAGGCACAGGGAGGCCAG
CCAGGAGCTTGGGAAGATAAGGGGTCCTCACAGAGGCCCTGACCACAGGGTGGGGGTGGAACTGGCCAAATGCCCCT
CTCGTGCAGCCCCCGGGACCCATGGGGTAGAGCTCCAGCTTGAGCCCCACTGGGAGCATGCCGGGCCCTGCACTGTT
CCCTAGCACCCTGCACACTTCCCACAGCCAGCTGGTTACTCAGAGGCCTGGACCAGCAGCTTCTGAGCGGATAGTAA
CAAAACCAGCAAAAGCAGCTCCCGGTACTAGCGCTTTCCCCCA*GGCCCTTGTGTCATCCTCACTGTGGCCCTGTGG
GGAGGGACCCTAGTTGTTCCCATGGTGCAGACGAGCCAGCGAGGAAGGCTGGAGCAGCCGAGGGCTCCTGGCAGTGG
GGGCAGTGGGCAGCTAGCTCTGCCCAGGCTACCCTGGAGCTCCACCAGAGGGCGGGAGAGGTGGCAGAGGACGCCCA
GGCCGAGGGCCCCTTCATGCTACCTCATTAGGGCCACTCCCAGCAGCCTAGGGTTTCAGGGGCATTGCTAATCAAAG
GCAGGCCAGGCGGTGGGGGTTTAGAGACAATTACCTGCTTTTATGCCTGTGGAGAGGTTGAAGGGGAACCTC

>NCL_C9
TGTAGATAAAGCTATAGAGATGCCCACACAGCAAGTAAGAAAGAGATTGAGACTCCTGTCTCCTGGGAGATTTATTT
CTGCAGGTTGGTTTTATCCAGCACAACATTTTGTAATTAAAAGAAATAAACATTTAAAGTTTTGATTGAAATTTATG
ATCGTTACAGGTGTTCTCCATCAGGAGATTCTAGAGGTACAGTTTGGCAGAGCAGGGATTGGAAGTGTTGTGATCTT
GTATAGGTGTTAT
>NCL_C10
GAGCTGGGATGCCTGGCTTACTAGTGGAAGGAAGAGAAGGAGGGGCTGCCGGGCCTTCAGACTGGGCAGATACTCAG
AGACAACTAACCGTGAGGACAGTCTAGATAGAGTACGACATGAGTGACTGGTGCACAAGTCCCCCTGTGCCCTTCCG
CTTCATGACCATCAAGGCAAGGCCTCCAAAAGCAGGCAGGGACAACCTCCAGGCCTGAAGAGCTTAGATGAGGAAGA
AGGAAATAGTCTTCACTGTTGGGACTTCCCTGGTGGTCCACTGGCT**AAGACTCTGAGCTCCCAATGCAAGGGGCC
CAGGTTCAATCCCTGGTCAGAGAACGAGATCCCATATGCCACAACT*AAAGAGCCCACATGCCGTGATGAAGATCCA
GTGTGCCTCGATGAAGACCTGGTGCAACCAAATAAATAAATAGATAAACATTTAAAACAAAAAAAAAAGGT*CCTCA
GTGT*TCTGTTCTGGAAACCTTCAGAGCACGATGTTTGGTAGAGAAAATAAAAGGCGTTTTGCAAGTCCGGGAAGCA
CTTAGTGTGAGTCTCTGGAAGCTGTGGATACCAAGCATGTAGTCAGGGCTCGATGACTGTTTCCTCCTCCACCTCAG
GAGGCACCTCAGTCAACAGGGATCCCTGCCATGAGTCCAGCCTGGCTGAGAAGAGCAGAAAGAAAAAGGGCAGGTGC
AAGGCTGACACCCAACTTGGGAATCAAGACTGGTGAGCGTCTGTCTTCTTTGCTGGGACTCTGAGCACACGAGGCAG
ATCACGGACAGGGAAGGCATCCTCT*CCCCCAGCCCTGTCCTGGGGATGAGCATCCTCATGCTAACCCAAGTACCTG
ACTCCTCAGACACAGAGCCCCCAGACTGCCAGCTCTAGGCCCCAATCATCAGAAACTCCTGCTGAGAACGATGGAAG
CACAAGGGCAGGCAACCCAAGACACGGAGGGCCTTTTCTCTCCTCCACAGCTTTACAAGCCTACCATCTGGGAAAAC
AGACACCACAACGGTTGGAGGGTACACACCATACAGCGCATAATATGCAAATGAAACATTCAAAGCAGTCAGCAGTA
CCTGCAGGACTGCCTGGCAGAAAAGGAAGGTGTCTCAGCCACCGTGCAGTCAAGCAGGGACCTATGAATGGGGACTA
TAGAGGGACTGA*TTTTGGTTCTATAGGCACAATGGTGAAATAACCTGACATGATGAGAATTTTAAACAAAAAATAG
AGAAAACAAAAGACAAAAAAAAAGAGGAAAAAAAGGAAAGGCAATTTGAAACTCCAGG*AAAAACAAAAGCAGTATA
AGAAAGTAAATGTAACTTTTAAACTTACTGGACTCTGTAGGGAACAATATTTACATGGCTTAACGAATAAAAACATT
GTATATTGATTTTC*AAAGTTTAGAGTCAATTCACAAACAAAACATGGAAGAAAATTATGGTTAAGAAACAGCAGGT
AAATACTGTCAACCTTGACAACTTGAAAGTAAAACTATAGATGACAGGAGCTAGGGAAAGAGCAAAGAAAGCTAAAA
CTAAGGGTAGGAGTCTAACAGTATCCTTTTG*AAAAGGAATCAAGAGATACTGCCCTTATTTTAAATTAACTATACT
TCAATAAATAAATAAACTTTATTTAAAAAG*ATACTGCTTTATAGTTGACCAAAAC*AGAAATAATGGTATCACGTA
TTTCC*TAAGGCTACAGAGGTAACAAAGAGCAAAACTAAAAAT*GGTC*AT*AAAAAG*AAACT*AT*CTTGTGAAG
ATGGAAGAGGTATGAGTAAACC*AAATCATAACAGGATAATTAGTAGGTCAACATTAGCTTCAGAGGCTAAAAATCG
ATAAACC*AAGAAAT*AGCAGTTGAGTGTAGCTCTTGTGTGTGGACTCTGGAGCCTGCCTGGGTTCATATCCCAGCT
CTATCACTAACCAAATCATGAGCAAGTTACCTCCTGGGCCTCGACGTTCTCATCT*GTAAAA*TGGGGG*ATAATAG
TATATCCTACTTCAAGAATTGTGGT*CGGGAAAAAAAAAGTTGTGGTCGAATTAAATTATGC*AAAAACATTTTTC*
AATGCAAAAACTT*AAG*AGTAGGGCC*TAGTCCACAGAAAGTGGTATGTATGTAGTGTTTGTTACTATTATTTAGA
GATATGAATTTAATTCTGATAGAAACAGCTAAAAGAAGCTAACATGGTTGTCTCTGGTACTAGGGAGAGGTG*AAAA
AGAACCACTGCTTTTTACTCTAAGCCTGTCTGTGCAA*TTTTATTTAATCATGGGCATGTACTACTTTGATTATAAT
ACAAGAAACTGTTAACAGAACACATGAAAACTGGAAAAGAGAGTGAGCACTTTGTCACTAGAGGTGTGTAAGCAAAG
CATGGATGGAGCAGAGGAAATCCAAACATCCTAGGAGGTGGACAGGGTGATTCTGCCTACTGTCTTGTGCAGCAATG
AGACTGAGAACAGGGGGGCAATAAGGCAGGGAGGTGGGCAGAGACATCCTGGAAGAGAGCCAGCCCCTTATGCAATC
CCCAGGAGCATCCTGGTTGGTGGGGA*GCTCATGCAGGAGTAGAGAGGGACATCACACCTGAGCTGCCAGAGTGCAG
CAATAAGAACAGAAACCACAAGCCCTAGGTTTAAGCCCTGGTTCTTGCCAGCAAGTGATGTAACCTCTCCAGGTCTC
AGTTTGCACATAACAAAATATGAAAAG*CCTGAGTAAGACACTGCGCAAGTAAAG*CGGTGACCACATCGGATGCTC
ACAGACGGCTCTAACTGGACTGCCACGGTTCAAGGCAAGGACTCCCACCTTGTAGGGGATCCAGAACCCAGATGTGC
TTTAGCAGCAGGAAGAGCTACTCTCAGCCGAGTCTCTCAGATTTCCTCCCTTATCCCAAATCATACCACAAACTTCC
TAAAACTGACTCCGAGCACACTAGTGCTGGAAAACCTTCCATGGCTCCCCTGTGCCCTACAGAACAAAAG*CTGCAC
CCTTAACCGGCCATTTAGGACAGGAACTAGGGTGAGGCAAGTGAGACGCGTAGGACCCAACATTTAAGGAGGCACAG
GTGGGCCATAGAGGACTTCTGAGCACTAGCATCCCAAGAGTGAGCTTCCCCTTAAATT*TTTTTGCACTCTACGCT
CTCACTCTGGTCCCAGTTCTGCTGCCATTAAAGGCCCTTCACAGTCTTGCCACTGTACACACAGAAAGTGCCAGCAA
AACTGCACTGCTTGCCTTTCCCAAGCTTTCCAGGAGCTGAAATCGCCTTTATTCCTCTGTCACCTACTGAGTATTCT
GACATGCAGCTTAAGCGCCACTTCCTCCAGGAAGCCCTCCAAGTCGAACTTTCAACTCCTGGCCTCTCTTCCTCCCC
GGTCCTACTGGACTGTGGGGCCCTCCGGGCTTCACTTTACCCACCCCTGGTACAGCGCTGGGGCACACAGAGATGCT
TAATGAATGCCTGCTCTATGCAATGAATGTCTGTGACCTGTAAAACGGAATGAAGACAGAGTGAAGTGGAAGATTAA
AAGGCAGCTTCTGTACCTTTTCCTCCGCCCCAAGTGCCTGGCCTCAACTATCAATACATAAATCAGTAACTTTCCCT
TGCTGAGAGAAGGGAACAGCAACAGTGTCGGTAAAGCCCTTCCTCGTTCTTTCTCTGGAGGGGAGTTTCCACACTTC
GCTGTCTCTTCTATTAGCATTCCCAACATTCTGGGTTCCTCGGAGCAAGGAGTGAGCCTCCCGTTAACACCAGAATC
CAAATTCTAGTCATGGTTAGAAGGACTCTGAGAACTTGGGTTGGTGCCCGGGACTTGCAGGCGCTTTGGGGGCCCTA
TTTGAGCGGCAGTGTCCTCTGGGCTTGCCGCCACCGCCCCCATTCCGAGTTTCTTGGATCCCAGATGAGCGATCGTC

CCGGAGCCAGCCGGGCCATGAGGGCCGCTGATCGCCGTGGCTGAGACTCCCGGGTCCCGGGGCTCTCCCAGGAAGCC
GCGGAGGCGTGGCTCAAGCGAGTGGGCGAGGAAACAGCTGCAGGCGTAGCGCTGGGAAGACCCACCCGGGGTCCGGG
GTCCACGGCCCGCGGCCCGGACCCCCCGCCCGCCCGATCACGTGGC
>NCL_C11
TTACAAGTTGTTTTTTTTTCAGACAGGCAGTCTTAGATTTATA*TACAAAAGTAAAACT*GAAAATATAGTTTTGTT
CTCTGTACATTTCTTTTAA*TTTTTCCCTAAAAAAGGAAAAAAAAAC*ACGCTTGTGTACTTTTCTAAACAGAATAA
GGT*AAAACATAGCACAAAGTT*CTCTTCTTTCTGCCAAGTTCCTTTAATCTGGGGTGGGGATGGGAGCAGCAGGGG
CGTGGGGATCAGTCCAATGAAATAATGTCTGGTATGATGCCAAAGATGGCTGCTGTGTCCGTGCTGCTCCTATTTGC
GACAGTGTGGCCATGGCTCTCCCTTAGGCTGTTGGAGGATACAAGGCTGCTGTTACTGCCACTACTGCTTCCGTTGG
T*GGTTGGCAGAGAAGTACTGCCTCCTGCACTTAACTGATCGAG*AAACATCTGTGAGGAGGCATACGG*AAAAAAC
CGAGACGAGTGTAAGAGGTCTTGATCATCTGAAACTGCTGCTGCTGCAGCAGCAAGGAGGGAGGTGTTATAATGCTA
GGGAGG*AGGAAAAAAAAGAC*AGAAAATAAAGTTAATTATTGTCATAAATTAATAAACTCTCTTCTCTCAGCAGAAC
AAAGT*CCTAAATTTAACATATTCTAAGTACCTGGTGTTTCTGAAATATTCAGTAAAACATTTTGCAAAAT*CAAAT
GGTGTTATATAATGATCACATCATTG*ACC*TTTTTTATTTTAATTTTTATTAGCATATCACTGAGGCA*TTTTTG*
TATCTGAAACT*GGGGAC*GGGGGTCAAAATATTGGTTTTTGTTTTAA*TCTGTACTGCAAGCAATTCTCATAATGC
TATTCTTAAAGAGGCATATGAAGGAAAGACGACAGACTACGTCCCCTCCAGGAGATATCTGGCTAAGAAATTTGTCA
TTCAATATGACCCACTGCCTGTTTCTTTATAAAAAGTACTGGGAAACAGACACTGTAC*TTAATATATTCAGGCACA
GTTGAAAGAAAGGGACTTTTATTTAAATTTTTAAAACCTTAAGAATAATGAGTATAACAACTCTGAATTTTCTAAGG
GAATTTATGCTCTTTCTCTTAGCTACATATTATACTTTTAAGC*ATTTTCATGTAGAATTAAAAGATCAA*TTTTTT
TTTCATATAACATACCTGATTGTCTCCTGATAAGAAAGGAAAGAAATCTAACCCTGCAG*AAGAAAAAATTA*TTTT
AGTT*GGGGGGGGGAAAAAAAAAAGAAGAGCCTCCCAAATGCCCCAACACTACAAAAATAAAGACAATATCTTAAAA
GCGAATTCTAAATTCATTTAGCATTTTATTTAAACATGAAGAGTTGAGCTGAATATTTTAGATCTCCCCCATCTCTA
GATTATAAAGCCTGGAGTCCAAGTTTCAGCAATGGTGTGCACTCCCGGTGATGCCATCTATGGGACTTGGGCCACTG
CTGCCCCATTCAGATCCAAACACCAACCCCTTAAACTGACTGAGCACTTTGCACCAAATGCTCTCACTGGCTCCCAT
GAAGTTACCTTTCAATCTAACCTACTCCACACAAGGACCTCACGAGGTATGCAGTTCTATTCGGTCTCATTTTGCGG
ATGAGAAAGTTTGAAGTTAAAAGACGTTAAGCAAGTTGGAAGACAGATTGGCTTCCTCAGATTCCTTCACTTCCCAC
TGTCTGTACTGGATTCTAACTCAGAGACTAGTGACACAAGAGGGGAAGAGACCCGAGGGAGTGTCTCGCTGGCGTGG
AGGACCAGCTCTCATCCCTGGCTGCAGGGAAGTCACTGTAAGCTGGCCCCTTCCAATTCACATTCTGGTCCATACTG
CCCAGTGTGTAATGTGATGAAGAGCAAATAAACCAATGTGAATAAAAT*CCCTCAGTATGACATCTAAACTTGAAGA
AAAGTACAACACACAGATGTAGATAAAAATATGACAAT*ACAAAT*ACTTTGGGGA*TACCTCCAAAGTATGGCAGG
TATCCCACTCCATTATTCTTGCCTGG*AAAATCCCATGGACAGAGGAGCCTGGCAGGCTATATAGTCCATGGGGTCA
CAAAAGTCTAAACTACCACCACTACCACTAACTACACTAGAACAGATTATACCATAAGGATCTAAAAATGTATAGAA
CATTGGTTCTTGACAGTGAGTTTGGCAATGGGCAGTGAGAAACAGGGTTACACTGACTAATAGAGTAGCTACTAACC
ACCTAATGTGGCTATTTAATTTACTAAGATTAAAT*AAAAT*GAAAAACT*CAG*TTCTTTAGTCACACTAGCCACA
TTTATATGCTCAGTAACCACATGTAGCTCAGTAGCTACTGTACTGGCCAGATTACATTCCCCTA*GTTGCAGAAAGT
TCTATTGGACAGAGCTGAAACACGGTTTTCACAATATTCACTAAGAATTCACTCAAATAATCTCAGCTCCCAAGAAT
CCTTGCCAGTGACCCTGTAATTTTACTGTAGAGAAGGGATG*GAGAGAAACTTCCACTCTGTG*TTTTTGTATTATA
TTTTCTACAATGAGAAAATATGGCTTTCTGTAATTTAAAAAATGTTCTTAGATTATATGGCTAATATACAAAATTAT
TAAATTGCAAAGATACATGACAACAGTAATTC*TTTTTGATAATATAAATGACTATTATGAACTCAAAATACATTGT
AGAAATGATTATAATAACATTTCAAATTACTTTCAAAAAGAAGAAACCGGGTTCACCAGAACAAAAATTAAAACTTT
CCAATGTTTCAGTTCTGCATTGTCTAAAAGATAATA
>NCL_C12
CTTAAATGAGATGGTACATGTAAACAGCTGGAGGAGTGCCTGGAATGTAGAAAGTGTTCGGTAAAGCCTAGCTAGCA
TCATCCATCGCCATGTTTCATGCCTTTCAGCCACGTTTGCATCCTCTGCTCCATCCGATCTCCCCGTCACCTATAGA
CTTATAGGTGAGAGCCCAGTGTGATTGTATCTGTTGAGTATTTTCCTTCTTGAATCCTAGAGAAGCAACTGTCAGGG
CAGACGGAGTCCATCTCCGATCCCTTCCCCCAGGCCCCAGTACTTCGAGCCAAACCTTTGCCTAGAACTTAACCATG
TGTGTTTCCTCCTGCCCACAAATTCCTGTGGCTTGTCCTGGTGGGGAATCTTCCCATGGAGGAGAGGAGGGCTGACT
GCTTTCGTTTACTTCTTTCCCAGGCACAGCTCCGTCAAGGCTGACGAGGCTGCTGGCACGGCTCCCTTCCACCTTGA
CCTCTGGTTCTACTTCACTCTGCAGAACTGGGTTCTGGACTTTGGCCGCCCCATAGCCATGGTGAGCATGAAGCTGT
GGATGGAA*CCCCCCAACTGGCAGTGTGGAGACAGTGTGGTCTGAATGATCAGAGTGACCTCAGACAAGCCCTTGCC
CCTACTGGTACACTGTGAGAATGCGGTCCCTTAGTGCTGCGGCCGTACCCAGTGTGGCAGCTGCCAGAGAGTGGGGG
ATCCGAACAGTCGTACTGTGGTGTGACCCCACCACCTGCCCCAGTATCTCGCATTCAGGTTGAACCAAAGCACCAAT
TCAGACAGTGGAGTATGGAAAAGGATCCTACCAGTCTTGCCCACAGCCTTGGATAGAATTG*TGTTTTCTGGAGGTG
AAGAGTATGGGCTGAGTGTGCGTGTCATATGTGCCTTGAGCCTTGGGTAA*TTTTTAGTGTCGTCTCGTGAAACCGT
CCCCTTCTCTCCAAGGCCACCCCATGCCTCAGCGTACAGTGGTGGCCTCTGTGGAGATTAACACTGGAAGAGCACCA
AAGCCCCTCACTTGATGAGCAGCTAGACAGCTAGATGGATAGACAGATAGGTAACAGAAGGTGAGGTATCACCATGT
ACTGAGGTGTTAA*TTTTTGCATATGGGTGGT*AAAAACAACAATATCCAGCT*CTGAAGCTGACACCTTGTACATG

CGAGTGATATTGCAAAGTAGCAAAGTCCTTTGGGAAAGCAGATGGTCAAGGTGATAACTGGTACTTAGATAGTGCCT
AGCAAGATATATTCCTGT*CCTAAAGGAATAACAATAATTACAGTAAGCTAGAAGTAGCAGTTAAGATCTTTATGTG
TTTAAACTTCTTAATT*CTCCCATCAGTGCTATTGTTAATCTCGGTTTTGTAGATGAGAAGACTTTGGACAGAGAGG
GTAAGTAGCTGGCTCGGGGTCACCCAGTCCGCAAGTGGAGGAGCCAGCATCTGCCACCCCCACCCAGGCACTCTGGC
CCCAGAG*CCCA*CCCCCCATTTCCATCATGATTGTATTCAGAACCCTAATATGTTCATCCCGCCTGGCTCAGTACG
ACCCTTTCCTAGAATCTGCCCTGAAGATGTAATCACAGGCGAGAGCCTGTGCC*AAAGGTGTTCACTGCAGTGTTGC
TCACAATAG*CCAAAACCTGGAAGGGACTTGAGTGTCCAGCAGTAGGGACTGGGTAAGTCAGCTTTGCTCTATGCAG
GATGGAGGATGCTGTCTGGGCAGTGAACAGGAATTGTGGGTACTATAGAGCATTCTGGAAGGTTTAAGGAGAAAACA
GATGCCTGGGTGGCATGCACATGAGCTGCCTGCCAGTCAGTAACGGCCCAAACGATCTTATATACGTGGGATCCAAG
ACTTCACGGGGATGAAGAAAAGCTGCTGGCAGTGGTTTTGGGGAGGAGGTATGGTGGATGCTGAGTCTCCTTTCTGT
AGGC*TTTTGGTAAGGTTTACAAGCTGTATTTTAAATGTAAGTGAAGAAAAGGGACTGACCCAGGAGTCAGGGCACC
CAGTCAGGTTTGTAGGTGATTTGGGGCAAGTCTGTGTCATCTGCGGAGAAGGCAATGGCAACCCACTCCAATACTCT
TGCCTGGAAAATCCCATGGGCAGAGGAGCCTGGTAGGCTGCAGTCCATGGGATCGCTAAGAGTCGGACACAACTGAG
CGACTTCACTTT*C*ACTTTGCACTTTCCTGCATTGGAGAAGGAAATGGCAA*CCCA*CTCCAGTGTTCTTCCCTGG
AGAATCTCAGGGACGGGGGAGCCTGGTGGGCTGCCGT*TATGGGGTCGCACAGAGTCAGACACGACTGAAGCGACTT
AGCAGCAGCAGCAGCAGCAGTATCATCTGGGCTTTGGTTCTCTCATCTGTAAGAAGACTGGTTGGACCAGCTGACCT
GAGCCCCCTTGCAGCACAAAGGTTCCAGGGGACAAAGGAGCTGCCCATCAAACTGGTGTTCCAGGGAGAGAGGAGAT
GAGGCCAGGGTGGCTCGTGGCTCTCCTCACTCACTAAGACTCAGCCTCATCTGGCCTAATAGGGCTCCCCGGTAGGG
CAGATCACCTCCCGCCATCCTGCTTAGAGTCGCTGTGGCCGCTGTAAGCCTAAGTGGGCTCAGGAGGAGCATACATC
TGAAAAGGTCTAGTGTATACATGAGC*TTTTCAGATGTATACATGATAGTCTCAAATTTTTAATAACTGCAAAATAA
AACCTTTTACAAGTTATAAAAGTAGTATGTCATGGTTCTATAACAAGTAAGTTCATCCCCAGTAATATAAAGAAGTG
TGAGACTCCCTTTGCAGTCCCACTTCTGACTGTTTGGGGCGCATCCTTCCACTCCTTTCTCTGCTCATGCAACACAT
GTGTACATCTC*TTTTTTCCCCTCTAACCAAAATAAGATTTTACTTTATACCTTAGTCTATAGCTTACTTGTTTTAC
TTTAGACTGTTTCATGGATAGCTCTCCATGTCAGCACATACGAGGTGTCATCTCTGTTCATGGTTGCTTAACCTTCT
GTAATAAAGAGGTGCTGTAGTGTATTCAAGTAGTCCTGTTTGGTTGAAGAGTTGGCCTATTTCTGATTCTTTGCTAT
CACCGGGTGTTTCGGCCAGCTGTATATCTTGCTGTAGATAAAGCTATAGAGATGCCCACACAGCAAGTAAGAAAGAG
ATTGAGACTCCTGTCTCCTGGGAGATTTATTTCTGCAGGTTGGTTTTATCCAGCACAACATTTTGTAATTAAAAGAA
ATAAACATTTAAAGTTTTGATTGAAATTTATGATCGTTACAGGTGTTCTCCATCAGGAGATTCTAGAGGTACAGTTT
GGCAGAGCAGGGATTGGAAGTGTTGTGATCTTGTATAGGTGTTATCACACTAACCCCGACAAGGCTGGACTCACTGG
GGTAGACCAGTAGGAAGAACATGGGAAGTCCAGACAGAGGAATGGCCCTGGCCTGTCTCAGGGAAGGTAGGAAAGGA
AGCACTGAAACTGCGTTTAGAACTCACTTATTTTCTCCTTTGTTGCAGAAGGTATAGTTCTGACCTGATTGGGAAGT
GGTCCATCTGCTCAGTGGACAGGGCGTTTTTCACACCTGGGTGGTCAGCATTTCTTGTCATCATTCTAAGGAGCCAG
ACCACTAAGGGTCCATGTAGACCATCTGGGCCACATTTCGTGTTACAGATGGGGAAACTGAGGCAGGAAGAAGGGAA
GAACTTGCCCTGGGGTCACCCAGAAGGTTTTTGGCATGCTTGAGTCCCAAAGGATCTTCCAACTCCCAGGCTGGTGG
GTACCAAGGGACCTTTACTTAGTATCCACAGGAAGAGGGACACAGGATCAGACTCTGGGTTTGTACCTGCTGTGTGA
CTGGGGACACGTCTACCTTCTCCGGGACTCAGTTTGTCTGGAAACACCTTCACCTCAGTGCTGCCAGGAGGATTCCA
GGAGATCACAGAAAGAGCACAGTGTGTGCAGAGGCAGTTCTGAGGGCTGCGTGGGTCAGGCCTTGTAAACTGCATTC
TCCACTTACTCCCCAGCTGGTGTTCCCTCTCGAGTGGTTTCCGCTCAACAAGCCCAGCGTGGGGGACTACTTCCACA
TGGCCTACAACATCATCACGCCCTTCCTCCTGCTCAAGGTACTACCCCAAGCCCCTCCTCTTCCTCCCTGCCCTCGT
TCC*TTTCTCTGCTCTGCCGCTCAGCCCTGGGAGTGCCCTGAATCAGGCCGTCCCCTTGCCTGACTCTGGGGGCCGA
AGGCTATGACAGAGGGGATGTGTCAAGGGCCTGAGGGCCCCAGGGGAGGCAGGGCTGCCTCTGCCTGGGAGGGATGG
AGCAGACTTCCCCAGGGAGGGGCATGTGATTTGGGACTCCACCTGATGTGCCTGGGGGTGAACTGGGAGAGACCTTC
AGGCGGGAGTATCTGAAAAGAAGGCTGGGCTTGAGGAGCCGTGTTTGGGGAGGCTGAGGCGGAAGGAGTGGCTGGCT
TGCTCAGGAAGAGGTCAGGGACCAGTTGGGCAGGCAGGTGGGGGCCGGAGGAAGGAAGGGCATCAATGGAGGGTGAA
CTCTCTCCTTCGGACCGAGAGCCCTGACTTTGAGGCGCCACTGGGAACAGAGAGATGTCTCCTAGAGGGCTCTCTGC
TCTGCTGGCTTTGCATTTTGCCAAGTAAAGGCTCTTAAAACTACTGCTATTTACTATCCCCATCCTTGTGTTTTTAG
*AGACATTTCCGAACTTCCTTGGTGGCTCAGACGGTAAAGCGTCTGCCTATAATGCGGGAGACCCGGTTCAGTCCC
>NCL_C13
TTTCATTCAGCCTCCATGACCTCCCCTGAGTTCCAGTCAGATTCAAGCAGCTGTTATTTAGGGAAGAGAAGGGATGC
CAGAGGGGGAGAAACAGTCAAGAGCACCCTTTGGGCAAGGTCCTAGTTTCCGATCAAGGAATACATACAAACAGTTC
ATATAAAACTTCTCACTGTCTTCTCC*AAGTGAGGACACATGGTTTGAGGGCTTTAGCCCACTGTGTCCCCCTTTGC
CTGGCAAAGCAATAGAGTTATCCTTTTCTGCTTCACCCGAAACTGTCTGCAAGATTTGATTTGGCACAAGTGTACAG
AGAAGCTGAGCTTTTGGCATC*AAAATGTATTATCATCTCAACAAGTAATCATTATTAAAATTACTGAGGTTTTTTT
CA*TATGAAGTCCTCAAAATTCAGTGTATATGTTACAC*TTAGGGCAC*ATCTCGCTGCTGCTGCTGCTGCTGCTGC
TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAAGTCGCTTCAGTCGTGTCCGACTCTGTGCCATAGAC
AGCAGCCCACCAGGCACCCCCCGTCCCTGGGATTCTCCAGGCAAGAACACTGGAGTGGGTTGCCATTTCCTTCTCCA
ATGCATGAAGGTGAAAAGTGAAAGTGAAGTCACTCAGTTGTGTCGACTTTTAGCGACCCCATGGACTGCAGCCTACC

AGGCTCCTCCGTCCATGGGATTTTCCAGGCAAGAGTACTGGAGTGGGGTGCCATTGCCTTCTCCAGCACATGTCGTT
ACAAGACCTCAATCGCCACATGTGTCTAGTGGCTGATAGAGTGCGCAGTGCAGGTCTAGAGCATGAAAAAGGAAGTG
GCTGGCTGAAGACTATGTGCAGTTGCCTCTGGGTGGCATAACATTAACTGCTAAAACAGAAATACCAACTTGTGTCA
ATCACTTCCTGTGACTGGTGAAATTTGACAGCAGAATTCCTGTGAGTG*ATTTTGATGACTGTCATGAATATTGAGT
CGCTATTTTAACACTGTTACTG*AATTTAAAATATTCCTGTATATTACAGGACAGACATTCATTAAAACAGTGGTTC
TAATGCTAATGCTACGAAACAACTCCTATTCCATAAATATTAGAAGGGGAAGTTAGATGTGTCAATTACCAAGTTGT
CGTTTTAGTCAATAAGTGTGTCTGACTCTTGCGACCCCATGGACTGTAGCCCACCAGGCTCCCCTGTCCATGGGATT
GCCCAGGCAAGAATACTGGAGTGTTGCCATTTTCCTCTCCAGAGGCTCTTCCGTTAACCCAGGGATTGAACTGCATC
TCCTGTATTAGCAAGCAAATTCCTTACCACTCAGCCACCAGGGAAGCCCCATTGCCAAGAGGGAAGCCAAATAAACA
AGCTTATGTTCCATTTGTTAATTTAGAAATGAATACCTTGGGAAATATTGGAAAGTTCTGAGAAATAAAAATTAACA
TTTAAGAAGGTACACTTTTAAGTTTGCAAAAATGGCTACACAGGCAAAAAGGAAGG*AAAAAC*AAAACCT*GGACA
CATAGTTGCAAGCTCCAGGTCAAAGTACTGTGCACCTGAGCTGCTCTGGTTGTGCCTGTAAACACTCAG*AACAAAA
AACCCTGTTTTAC*ATTTTG*CATGTGTTTCTATAGAGGAAATTGCGTGCTGCGAAAACTGTCATTAAGTAGGGTTT
TATTTTTACTAGCCTTTCGAATTCAGGTAAAGAATCT*CCCTGCAATGCAGGAGACCCAGGTTTGATCCCTGGGTCG
GGAACACCCCCCTGGAGAAGGGAATGGCTAACCACTCCAGTATTC*TTGCCTGGAGAATTCCCCAGACAGAGGAACC
TGGTGGGCTATAGTCCATGGGGTCGCAAAAAGTCAGGACACCAGTGAGTGACAAACACTTTCACTTTCCCATTCAGA
AGTT*ACAAAAGTTT*ACTGGCAGACTAGCCTGCCATATAGTGGTAAGAATAGGAAACTACAGGCTGT*GTTAAAAA
TATTTAATACGAAGTATTTGCTATATCAAAATAATAAGTGCAACAGAAAATAAACACCTGGAAATGCACCATCCAAC
TTAAGAACTAGAACATTACCTATAACGTATCCATCCATCTGTATGCCTTTTTCCTATCTTATCCTGCCTCACCAGAG
ATAATACTGGCTTGAATTTTATGTTGACCATTTTC*TTGGTTTCTTTAAAAGGTTGTTTTATCATTTCTGTATGCAT
CCCTAAATGAAACATTGTCTAATTTG*CTCATTTGAGCTTTATAAAAGTGG*CATCATGATTTTATGTAGTTTTCAG
CAACTTGCCTTTTTCATACATTTACTTGACCATGTTACATGTATACCTGGAGTTGATTCACTGTGAAAAGAGACAGC
TATTTATTTATCCATTTACCCATTCTCCTGTCAATGGACATTTCATAGCCTCCAATGTTTTACCATCTTAACAATAC
AGCAATGAATATTTTTATGCATTTTTAAGAATTTCTTTGTAATATGTCTCGTTCAGTTTTACAAAATACATATATAT
AACCAAAGTGTCTTTCAGAGTTGTGCCAATTTACACTCCATTTGTTGTATATAAAAGGTTCCACTGATTTTCATCTT
CAGCCGTACTTGTTGTTACTAAATGTCTTAAATGTTGCCGATGGAGTTATTGGCTGTGCTGGATCTTTGTTGCTGTG
CAGGCTTTGCGTTAGTTCCCACCAGTGGGGGCTACTCTCTAGTCACGGCTGGGCCTCTCTTGTAGGGCACAGGATCC
AGGACGCGTGGCTTCAGTAGTTGTGGTTCCCGGGTTCTCAGTCGCTGTGGCCCGCAGGCTTAGTTGCTCTGTGGCCT
GTGGGACCTTCCCAGATCAGGGATCAAACCCATGTCTTCTGCATTGGCAGGCGGACTGTTCGCCACTGAGCCATCAG
GGAAGCCTGGAAAATCAGTTATATCATGGCGTTGACTTGTATTTTCCCGGCTGTCAGTGAGGTCAAGCATAGTTTAA
TTTAAATGGGCGATTTATGTTTCCTTTTTCATAAAGTGCCTATTCACATCTTGCCCATTTTTAACTGGTTTGTCTTT
TTCTTATTTTAAGCTTAAGATCTTAATCTGGAAATATATGTGTGTGTGTGTGTGTGTGTATATATATATATATATAT
CTCTAGTTTCTAGTATTTAGCTAACTTTGGTGCTATTGTTAATTTAACTAAGTTTGGTTTAGGTAAACTCTCTGATC
AATTGTCGTAATTTTATCCACAGATTTCTTTGGAGGGCTTTCTATATATACAATTATGTTAGCTG
>NCL_C14
AAAAATTAAAAAAAANAAAAGAAAAAAAAAAAAAAGTAATAAAATAATAACGTG*TTTTTGGTCATACTGGAGGGCTA
TGGTCCATGCAGTCTCAAAGAGTCGGGTATGACTGAGCAACTAAGCACATGTTAAATATAATATAGTCACACCATTT
AAAAATATGTATTTGAGTATAAATTAGCCCACTAGGCTCCTGTGTCCATGGAATTCT*CCAGGCAAG*AATACTGGA
GTAAGCTGCCATT*CCCTTC*TCCAGGGGATCTTCCTGATCGAGGGATTGAACCCTGGTCTCCTGCACTGCAGGCAT
CTGAGCCACCAGGGAAGCCCCCGTAAGTATATATATTAAATATAAAATCTAAAATAAAATCTTAGAATAAAAATT*C
GATATTTGTTCTAACTATGCTGCTTTCGATAATCTGACTCCCCACCAACA*CCCCCA*GATTTTAAAAAG*AAGCTG
GTGTAATAACAAGGTCAGGCTTGTTGGCTAACAGCTGTCAAGGTAGGCTGCATCAATGGTGAATGTTTAC*AAAGTA
GAATATAAAGTAGAATTAGTGTTTAAATACCCATTTCATTTAGAATAATGATACTGCTTTTAATATCTTTATTCCCA
CCAGTATGACGAGATAAGCTACTCTAGAGTGAAACATCAACCTGTGTAACATGTTTAGCTGTATTATAAAGATTCCT
GAAATAGCCAAAGTCCTTTAACAAGAGAGGAGCTACATTTATCACAACTGTACTCTTTTAATGTTGAAATTTCAACT
TTCCTGTAACTACAAATATACTCAAATAGAAGATCTTTTCATTTCATGAAAC*AATATCAAGCCT*GAAAATAATCT
GACTGCAAGCATCAAAGTACCTGTTCTTGGTCTGGTGTTAACTTTCCCTCTTAAACCTTGAAATCTGTATCAACGCT
TTCTGCAGACTTTTCTCCTCTCACTTAAATTCTTCCTTCAGGCACTAGTGTGTTGGAATATTTTCTAAATAGCTCCT
ATTCAATTACATCAGGCCACATAACTCTCCTGATATATACCTAAAAGGGGTCACATCTATTTCTAGATTTACCTTTT
GTGTGCTACAACTGTGCCTCAGCAACACAAGACGTCCTTT*GCTGAGATTCTCTGAAGTGTCTCAAGAGA*CCAACA
GTCTGTTTAAGAGTATGGCTATAGAGTCCGCTAC*TTATTTTACCATCCTCTTCAAAATGTTAAACTCTTTAAGGGT
GGGACTGTACTAACTCACTGTATTACCATTTATGTTTAGCACAGGGTGTTTTACATAACGCAGATAACAACGTTTCA
TGAAGTGAACTGAGGCATAACCAAGGGACATTATTATTGACGTGAGACTAATGGGAAGTATGGGAAATGAGCCTGCA
AAATAGAAAGCACTGTGTACTGCAGCGTAAGTCTTGGACCACTTTCATTGCCAGTCTTGCACATCCTGAGGCAGGTT
TAAAGCATTTACCTGGTATTTCTCTGCTTTTTGCTTAACACACTCAACTTAGGGTTAATTTCACAGCATACACTTTC
TAGTTTGAATAAATATTTAAAATCCTATGACCAGCATTAGGATCGCCTGAGAACTCATTAGGGCTCAGACGGTAAAA
AAAAAAATCTGCCT*GCAAT*G*TAGGAGACCCAAGATCAATCCCAAACATTTTCTGTTCCTATTCCTGAATCAGAC

ATGGAGGTACAGGAGAACAGTGTGTTTTAATAAGCCCTTCAGGTGATTCCGATGCATATTCAAGTTTGAGAAACACT
GTACTAGAACCATCTTTTGGTATGATGAATATCATATCCAAGTGACCCCTTTAATGCATTATTCCAGCCCCTTAAGA
AAAGCAAGAGTTCCAGAAAAACATCTGTTTCTGCTTTATTGACTATGCCAAAG*CCTTTGACTGTGTGGATCACAAT
AAAC*TGT*GGAAAATTCTGAGAGAGATGGGAATACCAGACCACCTGACCTGCCTCTTGAGAAACCTATATGCAGGT
CAGGAAGCAACAGTTAGAACTGGACATGGAACAACAGACTGGTTCCAAATAGGAAAAGGAGTACGTCAAGGCTGTAT
ATTGTTGCTCTACTTATTTAGCTTATATGCAGAGTACATCATAAGAAACGCTGGGCTGGAGGAAGCACAAGCTGCAA
TTAAGATTGCCGGGGAAATATCAATAACCTCAGATATGCAGATGACACCACCCTTATGGCAGAAAGTGAAGAAAAAC
TAAAGAGCCTCTTGATGAAAGTGGAAGAGGAGAGTGAAAAAGTTGGCTTAAAGCTCAACATTCAGAAAACGAAGATC
ATGGCATCTGGTCCCATCACTTCATGGGAAATAGATGGGCAAACAGTGGAAACAGTGGCTA*TTTTTTTGGGCTCC*
AAAATCACTGCAGATGGTGATTGCAGCCACGAAATTAAAAGACACTTACTCCTTGGAAGGAAAGTTATGACCAACCT
AGACAGCATATTAAAAAAC*AGAGACATTACTTTG*TCAACAAAAGT*CCATCTAGTCAAGGCTATGGTTTTT*CCA
GTGGTCACGTATGGATGTGAGAGTTGAACTATAAAGAAAGCTGAGCGCGGAAGAATTGATGCTTTTGAACGGTGGTG
TTGGAGAACTTGAGAGTCCCCTGGACTGCAAAGAGATCCAACCAGTCCATCCTAAAGGAGATCAGTCCTGGGTCTTC
ATTGGTAGGACTGATGTTGAGGCTAAAACTCCAATA*TTTTGGCCACCTGATGCAAAGAGCTGACTCATTGGAAAAG
ACCCCGATGCTGGGAAAGATTGAGGGCAGGAGAAGGGGATGACAGAGGATGAGATGGTTAGATGGCATCACTGACTC
AATGGACATGGGTTTGAGTGGACTCCGGGAGTTGGTGATGGACAGGGAGGCCTGGCGTGCTGCAGTTCATGGGGTCA
CAAAGAGTCGGACATGACTTAGCAATTGAACTGAACTGAAGAAATAAATACACTGAGAACAATCCAACTCATTCTCT
GTACTGTTTACAATCGTGCATAACCATTTTATTTGGCAGATCTTTTAATCCTATTTTCAAAATTAACTGTCACAATG
ACTTTTTTTA*TAAAAAG*ATCTCAGCACTTACCCTTTATTATTTAGTGGCGATGTTGGAGATAGGGAAGGACAGGT
TCTCTTGGCAGATGGCTCTTCCTCTTCTTCATCTGATGAACTGTCTATAGTTAGATCAATCACCTCTACTTTCTTAC
TTTTATTTGAGGACTGGTGATGAGAAGCCACTTGGTGCTCTAATGTGGAGCTTAAGCATCCTAGGAGGGAATAATTG
AAACACAAGGCAAAAAAAAAGGCCTGTGATATCTCTTA*TTTTTCCTCAGCCTATGGATTAGTCTATCAGTTATAGT
TTAGACTTTACCAGTTATATTGCTGGGGCCTAAGAAG*AAAAAAGCAATAGTGGCTCTGTTCCTCTAAGCTCAAGAC
TACTAGTATAGATTGTCAAGGA**TTTTTTTTTAAACAATTCTGAGTTAGAAGAATCAAAATGTTACATCCGCCTCA
CAATAAAAACAGTCTGAGAGCAAATTCACATATTACACATTTACCTGTACATTACCTGATACACAGAGAAATGTTAT
GGGAACACTTTCGCTTCCTTACAGCAAAAATATACTAGAAGCTAGAGAAAAATGCAAATGAAAAAAGAACATTCACT
AACCTATATCAAGTCTGTTAAAAATTTAAAAGGATACCATTTCAAAAGCATGACTTTTAC*AAAAAAAAAG*CTTAT
TAAGTTATGTTACTCTGAAACCTCATAC*ACAAAAATATAAAACCTTG*CAAAAGGTG*AATTT*AAAAAG*AGCTT
CACTACTTCTGAACATTTAATCAA*TAATTTAATAAGAGTTACTATGCTATCCATAAATT*ACTGAATATTCTTTAA
AAAAGTTGAGTTTTAA*TATTGAAAACCAGGCTTTAAAAGAGCAAGGAGATTGCACCATCC*TTTTGG*AAAACT*A
ACACTCACCATCAACTCCATTGTAAGAGGCAGAAACTTCCTGTACTTCCTTTTTTGATCTCATAGGTGCCCAAGAGC
CGTCCTCCTTAAACTGTATTTCATCACAATCTGTACAGTACTTTAGGATTTCCATAAACAAGCTAT*AAAAAACAAT
TTG*TCCTTTG*AATATCACATT*CTAAATTAAGA*CCTATTAGAATTACTGTTCAAAAACMATTCAGACCTATTTT
TTG**TACAAGTATATAAATAAGAAAGTACAAGGTATTATTTGTTCTTAAAAT*CTAGGGCTAAATAGGAAAATA*G
TTACTTATTATTATGGTAGG*TTTATGCAAGAAAATATTACATAACTGAAAAAT*GAGGTTTCCAGAGATTT*AAAA
AAAAGT*AACAT*GGGGGGGT*GTATGTTCACAGCAAGTGGGAAGAGGACAAAGAAAG*AAAACT*CTGTATTTCAA
AGACAAAATT*CTGTGTGTGAGTGGTATATAAATAGTGGTTATCTCTGAGAGGTAACATGGAAGATATTTATTCTTT
TTCTTTAAACTTTTCATCATAATCCAAATTGCCTACAATAAGCACATCTTATTCGATAACTGTAGAAATACATGT*A
AAAATATGTACTGAAAGATAGAACTCCTTGATAAATACCATGCAGATACTAAAAAACGAGATTTGCCAAAATAGTGT
AACATGAATGATCAAATTCTTAGGTTATGTTAAACGAAACC*AAAGG*ATAAAAAAATA*ACTGATATTATATAATC
ACAATTATGGG*AAAATAAAAATC*AGAAAATAATACATTC*TTTTTATTTTCCTGCGTTTTCCAATTTTTCTTTAG
GAGTATGTGTGTATGTGTGGATATTATAATAAGAAAGAAAAATCCCAATTGATGGGAGCATCTAAAGTTATGTAATA
TACAGAGTTGGTAAAATATTATTCAATTTATTAGTGATCACTGACATCTCTGAAAAATTAAACATAAAACATTTAAG
TTACACATTTACTCTCCCTCATATGATGGGTTTGGCATCTTTCAAGCTCTGGGTTACACCTGTATTGCTAAATGTAT
TCCATTCCTTATGTTATACCTTCTGGCTCCTTTGTTTCCTCAAATACTATAGATAAGATACATTAAGCATTTTGCTA
GAGCGCACAGGAAATGTCAAGAATATATTTTTTTTCTTTTCACAATTTA*TTTTTG*TTCAGAGGG*AAAAAGGGAA
AAAAAT*GGAGGCTAGAGAATAGGAGAAAATTTT*AAAAGATATTTGAAGATGTGTGGAAGGACAGATAAAACACTG
GAATGTCAGGAAAAAAAGAAATAACTGTTGTGATGCTGTAGATGGAGAAGAAAGACAAAGACAAGTATTTT*AGGGG
AAACAGAAACAAAGACAGCTGCTATTCAGTGTGTTGATCACTTTTAGGAGCCCACACAGCAGGGAGTAGACCCATAT
AGTAAATGAAGCACCAG*AAAATCTTCAAAAACAGCCCAACTTCTGGTCAGCTTAGGTGAACCAGAGTTTCACAATT
CTTTCATTCTTCCCCTAGAAAGGTCTGCACGTCTAGGGACAGATGGATTCATGTTAGGAAAAGATGTAAGCTATACC
CTGCTCAGAACCACCGGTCCTTTAAAATGGTAAATGTCAGGAGGCACTGACAGTGAGGCATGTATCTTTAAAACAAG
AAGACATAGTTAAAAAGGGAAAATATTCACAATATTAATTTC*TTAG*AAAAAAT*ATAATTTCTACAGTCATGAAT
ATCCTACACATATATAAC*TCAGAAACCACTATCTATAGAATTTTAATCCC*AAACCTGATAGGTAACAAAACTTTA
AT*AGATGGTTACTATGAAGTAC*AAGGAAAAACTTT*AAAG*TGACATACCCATCAATAATAAGGTGTTCATATGG
GGCCTTCTTATCACAGACAGGAC*AAACCCAAGTTGG*TTTTTTC*TCATTCATCTGAATATAAAGAGTTGCATCAA
AACACTGCAGATGAGAACATGTAAGGGCCCGACATGGAATTGTCAGCCGCATTTTACCAAGCTTTAAAAGAGGGAGA

CAAAAGTTAAATATTAGACAACCGTTGTTAAAAG*TTTTTCCTGAAACAAACACAATGTGCATGTATTCTGATTTAA
AGCATATTGTGTCTACTTTCAACACACTTCCTGTGAAACATCTACAGGTTAAAAATTACTATTATTTCTTTGGATAT
TAAAATACAGAGACTATGGAAAATCTACAGCAGAGGTTATTTACCCTCAATTCACAGCTTACAGATCCACTTTTACG
GGA
>NCL_C15
GGCCGCGCTGCTGCCTGGGCGCGGCGGCGGCGGCGGCGGCAGCGGCGGCGGCGCCCTGTTGAATGGGCTGTGAG
GGCCCAG*TTTTCAAG*CGCTGGCGAACGCGGCCTCCGGGGGCGCAAGGCAGCAGCAGCAGTGGCGACCAAACGGGT
GTTGGAGTTGGCGGCGGCCATGGAGGGCCTGGCTGGCTATGTATACAAGGCGGCCAGCGAGGGCAAGGTTCTGACTC
TGGCCGCCTTGCTTCTCAACCGGTCTGAAAGCGACATCCGCTATTTGCTTGGCTATGTCAGCCAGCAGGGAGGGCAG
CGCTCCACGCCCCTCATCATCGCAGCTCGCAATGGGCACGCTAAGGTGGTGCGCTTGCTTTTAGAACATTACCGGGT
GCAGACTCAGCAGACTGGCACCGTCCGCTTCGACGGGTAGGTACATCCCAGCTCCTG*CCTCCTCTAACGCGTGCGA
ACCTGTCACCTCATGGATCTTCCCGTCCCACTCTTTCGCTTCCTGGAGGTATC*ACTTCTCCCATTTTTGTATCTCC
TGACCTGCCGCATTAATAACCACTTCATCTTCTAGGGCGGATACTCCATTCCA*TGTTTTGCTGCCCATCCTTATCG
CCTCTACCCCTTATTCCCTCTATAGGTAAATAGCCACACCCTCTGTCTTAATATCTAATAGGTATTCTCCAGTATTC
TTGCTTGGGAAATCCCATAGACAGAGGAGCCTGGCGGGCTTCAGTGAATGGGGTCGCAAATAGTCGGACACGACTGA
GCACGCAGGCAAAACCAATGCTTCTCTGGATAGTGTATGAACAAGCAGTATGGAGGACGGAGCGGCGTTTCAGGTTT
TCCCTGTGTCAGGGCTGGACTTGTAACCGAAATCAGAAATTAACAAGGTCAGCAATTAACATTGGAAACCTCCGATG
AATTATTTGACCGCCCAATTTTAGCTCTTTGAGTGTCTGATAAAGCCTTTAGTGGGTTATGGGACCTTAATGTCAAA
AGAAGCCCCTAGAACCAGCAGCAGCTTCAGCCTCTTGCGGGATTGTCATGCAGCCAGACATCCCCAGACACGATTCT
GACTTACAC*ATTGCTGGGAAGGAGACACGTTGGCTGCG*AAAATGGCATAGGACCTGGTGTGGGAGAAGATGATGC
CATGTAAAATTAGTTATCCCATCATCTGGGGAAGGCTTGATAATTTGCTGTTATTCCACGTGGGGAGTTCAAAAAAT
GCATATTGTCATAGTTAAAATTTGCCATTTAGGTTGAATTGTACATATTTTGAGGTGTCATATTGTTGAATGCAGAA
TTGTCAGGTTTGACATTATCTTTTCTGTATGGATGGTTATAACCTGTTTTAC*AGATTACAGCTTCA*TGAAGATTT
TGCTTCATTGATTGCATTTG**TTTTTGTTGACTTGTGCCAGTAGGTAAGCTCTTCTAGCGTTTTCTTCTTTTGAGA
TAGATGTGATCCAAGTATAAAGATATGTATCTTTAGTTTTGATAATTGAATACTGCTTGGATAAAACTTGATTGAAA
AAAAC*AATT*CTTTGAGTTGACTAGGGCATTGTTGAATTTTGATTGATTTTTTTTTTTTAAG*TTGAAACAAGAAGA
GCTTTTTC*TTCTGTTTCAGATTAGAGGGAAATATTGCAATTAGAGTATTGCATTTTACTGAAATGATAGCCCATGA
GTGGTGGTTAATGTCTAGTACCACCTCTGTAGGCATTAAGGAGTTGGTCTTTCTCATATTAACCACAGATGGCAACC
TGATTCTTGTTTGTTTTCCTTTACTGGCTTCTGTGTTCTTTTTGTAACTCTCTGCATTCCCACAGCCTTTCCTATAC
ACAGAGGAGGAGTGAAGTAGAATGACCTCTGAACTTCGAGCACCATTCCAGTTGTATAAGCACAAGCATCCACTGAG
CTATGAATCTCTGCAGTAACAGCTGCTGGATCCACTGCTGCTAAGGAGATTGCTGAGGCC*TGTTTTGATGAGAGAA
TAGTGGC*AAG*AAAAAATAG*TTTTTTTTAAAGCATTTGAAAAGGCTTTGCTTATCTTTCCTAACTGTCAAAGAGA
GGTGTATTTTGAGGTGGCAATTAAGAGAACCTGAAATGATAAGGAATGTAGCATTGAGCCTTCATAGTGGCTTGCCC
CATATTTTGCACTGACCAGTGAAATTAAGTTTAGTAAGCAAACATATTTTGAACACCTACTGTGTGCCAGGCTCTAT
AGTAGGGATTGGGTGTGCAGAGATGAGTAACATGCGGTTCCCAGGAAGACAGTCAGGTGGCCCTGGGTGGGAGGGAT
TAAGGTGCAGTGGGGAGGGGAAAGATTGGCGAAAATGAGTAATAGCTGCCATGCCATTTAGGG*TTACAGGAATCTT
TTGGACTACTGAGGGCATGCAGGAAGAGGAAGTGCTCAGGTTTGCCTGGGAGAAGCAGGGAAGTTTTCACAGCTGAG
GTGATGTTTGGGCTCAGTCTTAAAGGATAGAAAGAAAGTAAAGAAGGGGAGCAGCAACATGGTGCACTCGCAATGAC
AAGTGTTTCAGTGCTGCTGCAGGGAGGGGTTCAAGCCTGGGAAATTAAGAAAGAGATTGTGGGAGACTGGAGTGAGG
TGGAGGCCAG*A*TTTTGGTGATTGTCATGTAAAATATCAAGAGACTGGTGAGACTGGAATTGCAAACATTTCTGTC
*ATTCGAC*ATTTTGTCTGAG*CCTTCTTCAATCTG*TCTTTTTTTGTTTATGTATTATTACAATTATGGATTCATT
AGCTGCTCTGTCAAGGTAGCTTAG*TTTTTTAAAACTTACTAAATGACTGTGATCGGCCTGGCACTGTGCCCCACAC
ATACTGTGGAAGATAC*AGAAAAAAATACAGTTTCTACTTTTG*AGGAGCTTACAAACTAATTAATAGAAGAATTGA
TATTCGTGGCCACCTTTCTGCAGACTCACAGCATTCTGAGATTAGTTTGCAGTTTGTTCCAGTCATTTAGGCATTTC
AGAAGTAGAGGAAGAAAGTGTTATTTGCAAATGTCATGATTTCACTTTACACTCATTCCAGACATTGTGAAAATGTT
AGTCTGAATGAGGTGTTAATATGAAGCTCACTCACCCCTCA*GGGACTGGTTTTTGAGAGGTACTTAGCCAACCCTT
CACTTGAGACCCATGGCCAGATAGGGACAGACAGGTTCTCTGTGGGCTTTAGCTTAGTGTTTCCTAA*TTTTAGTAA
TTTATGTATCATCTTCACAGTTGTTTGCCAGATTGTTCATCACTTATCCTGTTGTTTTCCTAATAWTTAAAAAAAAA
ACAAACAACAACTCACTTCTTAGCTTCGTCTTAAGCAACAGCACCTGTATGATAATGGGTTTGATACAA*TCCTTTG
TGTGTGTGTGAGAGAGACATGCACAGATATTACAATAACTTCTTTTAAAAAGTCTATAATCTGCCT*AAAAACATCT
ATAAAAGCTTGAAAACTAAAGTGAATTCATAATAGAGCCTCAAGGGAGAAAGAGCCACCCACCTAGGTTAAGCCTGG
TATTTGAGTCCTTGTACTAATTTTTAGGAGTAACATTTGAGATTATTTTACTTATGGCTTACTCTGTTAAGATTAAA
GTTGACATAGATGAGGGGATAAGAAGAAAAGTTACATGAATAGAAATGAAAGAAGAACCTTTAGTTGTAGAATGGGA
AGCTGTTAAGAGTGTTACAAAAATGGACCAGAACAAAGGAGAGGCGTCATGTGGTCTTGGAATATTGTGCTGGTTTG
CGTTCTGAGGGCCTGACTGCGCAATGCCAGAGTTTGTGGGGAACTCTTGTCCTAGAAGAAGTTTGCTGACATGGATA
GGAAGGAGAACAGGACCTCTGGGAATTTTGTCGAGTGCTCTTTTAAAATCAGTGCTAGTACAAATCCTTAAATGTCC
AGATAGAGGAAAGGCGTAGGGGCACAGTGTCACTTGACTTCATTTAGAAAGTCATTTCCAGTGTATATTTCAAGGAA

GTACATATGATCTAATAAATGATTTTAGGACTTTGGTTAAAATAAAGTTAA*TCATTTACACTGAATGAATTTTTAT
TTTCAGTTCAAATGCTTGAACCAGACAATAGATGAGTTTAGGTTTGTGATCAGAAATGTTATTTTTA*TAC*ATTTT
A*TTTTAAAAGGCGAATGCCAGATTAATGTTGCTTG*ATTTTTC*AG*ATGAAATACATATGTGTGTGTGTGTGTGT
GTGTGTGTGTGTGTGTGTGTGCGCGCGCACGCGCGTGCGCGCGCGCGCGCGTGTGCTTGGTCACTTAGTCGTATCTG
ACGGTTGAGACCCCATAGACCATAGCCTGCCAGGCTCCTCTGTCCATGGGATTCTCCAGGCAAGAATACTGGAGTGG
GTTGCCATTTCCTTCTCCAGCGGATCATCCCCACCTAGGAATCGAACCTGGGTCTCCTGCATTGTAGGCAGATTCTT
TACTGACTGAGCTGTGAGGGA*CCCCCT*AAAATTTGTAATTTTAGACCAGAAAGATCAGTGTTTTATAGTGCTTCT
CCCATACTCATTAAGTAAATATTGCATTTGATTATATTTTTATTTACTTTCTAGAAATACTTATAGAAAGTAAATAT
TTTTGG*TATTTTG*TTTTG*AGGCAGTATGCAGCTATAACTTTAGCTGTATTTTA*TGGAC*AAAAAATGTCATAA
GTATTAGGATTATTAGTCCCTTTAATCTTTTTAGTGCTGTATTAAGGTGTAGAAATTTAAAAATTACTAAGGGTAAT
GTTCCTGAAGGACATTCTGCAGTGAATATTTTTGTCATTAAGTA*TTTCAAAA**TTTTTATTGCCTCAAT*AAAAA
TAGATAATTGAAACTTTTTCTTAGGCTTCTGTCATATCACTGT*AGGGAAAATGGAAATTAATAGATTACATTGGAA
AAAAGT*ATTACATGATTATAAAAACC*AAGTGACGTTGTTGTTTAGTTGCTAAGTCATGTTCAACTCTTTTGCGAC
TCCATGGGCTGAAGCCCACCAGGCTCCTCTGTTCATGGGATTTCCCAGGCAAGAATAGTGGAGTGGGTTGCCATTTC
CTTCTCCAGGGGATCTTCCTGACCCAAGGATGGAAACTGCGTTTTCCGCACTGCAGGCAGATTCTTTACCACTGAAC
CACCTGGGAAGCCCAAGTGAAGTATATACTTGTAATGTAACTTTTAATCCTATGTCAGAGTGTGAGTGTGTGTGTGT
GTGTGAGTTGCTTAGTTGTATCTGACGCTTTGCTA*CCCTATAGACTGTAGCCTACCAGGCAAGAACACTGGAGTGG
GTT*GCCATTTCCTTCTCCAAAAGGAACTATAGAAAGAAAGAAAGAAAGTGAAGTTTCTCAGTTGTGTCTGACTCTT
TGCGACCTCATGGATGGTAGCCTACCAGGCTCCTCCATCCATGGAATTTTCCAGGC*AAGAGTACTGGAGTGGGGTG
CCATTTCCTTCTCCCTTATGTCATTAGTAGTGGGGAATTTGAGGCCAGAAAAAAAATTTTAAATCACTTTTTTTTTT
CCTATTTTAAAATGGGGGTCCTATTTATCTTATATGGTTATTGCAAAAGATAATACATGTGAATGC*TTATTTTAGT
ACGACAGTTCTT*GAGTTTTATTTAGTTTTGGTTCCTATTCCTGGCATGTAAC*AATTTTTTA*TATTTAAGAAGTG
CAAGGAAACTTTTGTAACTAT*GTAAATATAACTCTTTTAACTATGTATCAATTTTTGGGTAGATTGCATTAACAAG
CAGATTGGGG*AAAAAAAGC*TTTTTAAATGGAAACATGTTTGAACAATGGCATGTACTGTTTCCTCACATTTGATG
CATTATATAGTCGGACACGACTGAGCGACTGAACTGAACTGATAGTTTCCAAAGA*TTTTTGTG*TTTTTAATCTCA
TATAAGCCTTTTAGTTATATGTCTTTAAAATGGGCTTAATAATGGTTATCTACCTTATAGGATTATTGGGAGAAGGT
TATAATGAGAATT*AAMTVTAKTGCTCTATTGCTTACTACAGTGCCT
>NCL_C16
CAGTGAAGACTATTTCCTTCTTCCTCATCTAAGCTCTTCAGGCCTGGAGGTTGTCCCTGCCTGCTTTTGGAGGCCTT
GCCTTGATGGTCATGAAGCGGAAGGGCACAGGGGGACTTGTGCACCAGTCACTCATGTCGTACTCTATCTAGACTGT
CCTCACGGTTAGTTGTCTCTGAGTATCTGCCCAGTCTGAAGGCCCGGCAGCCCCTCCTTCTCTTCCTTCCACTAGTA
AGCCAGGCATCCCAGCTCCCTGCAGCGCCTTTCTCACTCCCACCTCCAGGCC*TTTGATGCTGCACTCTCGCCGCTT
TTCTGCCAACAGGAATCCACCCATCCATGTTTCCAGGCTGCATTCCTTCCTGCAGGCTTTGCATGCTGCTCCAGGCC
ACACTAGTCCTTCTGTATCTTGATCTCCTGATATATTCTATAGCTGATAGTAATCAGCAGCAGAGTAGCAAACAACT
TAGAGTATACGTCTTGAATCCATATAGACCCCGGTTTGAATTCTGTGTACTTCAGGACTCCTGGGTTGCAGAGTGAC
AGAAAGCCAGTGTGAACTGCAGCCCCCTTAGCAAAAGAGAATGTCTGGGCTCTTGTGGCCAAACTGTTGCCTTGGGG
ATGCCTGGAACTGGGGATGTGAACCCTCTCGGGCCTTTTTCTCCTCTTTTTCCCTGCTCTTCTCTGCAGGGCCGCTT
CATTCTCATCTGACTGTCCCTTGAGGCTGGTACTTTGGCCTCAGGCAGCCCCGGGCTCTCGGGCTCTCAGGAGGGTG
C*TCTTTTCAACTGTAGTTAGAAAAGTCCCACTGGGCCATCCTGGGTTTTTGGCCTTCCCTCTGTCTCAGGGAACAT
CATCTTTCACCAGTAGGAGGAGGGAATGGGAAATAGTCTGGCAGAGGAATAGTTGATGCAGCTTCTTATTAGGGCAG
GCAACCTTGGCCAAGCCTCAGTTTCCTTGTTTGTAAAATGGGAGGGTAAGAGTCTCCTGGTATTGCTGTGGGGCTTC
ATTGAGATGGTGCTGGTGACACTTCTAGCATGGCACACAGTAAGTGGGCACTGAGCAGAGGGGGTGGATAGGAAGAC
TTAGCTGACCTGGACAAACCCACTCCACCCATCCATTCTGCCACCAGAGGATAATGAGATCCCCAACAGCTTGCCTT
CCTGAGTCTGAGTTGCTCTAGCAACCCAGTGCGGACCATTCTTGGAATTGGCTGAAAAGGATCCCAGAAAATTACCG
AAGCAATAGCAGTGCTGACAGAGGGAAACCTTCATGGCACCTCTAAGCTAATGTTTGATTAGCCGCTGAAATGAACA
GGTAGGTGCCTGCTAAGTCACTTCAGTCATGTCTGACTCTTTGCCACCTCATGGACTGTATAATCCACCAGGCTCCT
CTGTCCATGGGATTCTCCAGAATACTGGAGTGGGTTACCGTTTCCTTCTGCAGGGGATCTTCCTCACCCAGGGATCA
AACTCACATCTCGTCTTGGCAGGCAGATTCTTTACCACTGAGCCACCAGGGAAACCTATGACTTTAAGACTAGAGGT
TAACAGCAGAAGCTCTGAAACGTAATCATTTGAGGCAACCTGCTCCCGTATTCTTGCCTGGAGAATCCCCATGGACA
GAGGAGCCTGGTGGGCTGCAGTCCACGGGGTCGCACAGAGTCAGACACAACTGCTTACGCACAGCACAACAGCACTT
AACGTCTTGGGCTCAGTTGTTTCACTTGTAAAATGAGGGTAATAGTGTGACCTCAGAGGGTTTCTGTGAGGATTAAG
AGACTTAAATGAGATGGTACATGTAAACAGCTGGAGGAGTGCCTGGAATGTAGAAAGTGTTCGGTAAAGCCTAGCTA
GCATCATCCATCGCCATGTTTCATGCCTTTCAGCCACGTTTGCATCCTCTGCTCCATCCGATCTCCCCGTCACCTAT
AGACTTATAGGTGAGAGCCCAGTGTGATTGTATCTGTTGAGTATTTTCCTTCTTGAATCCTAGAGAAGCAACTGTCA
GGGCAGACGGAGTCCATCTCCGATCCCTTCCCCCAGGCCCCAGTACTTCGAGCCAAACCTTTGCCTAGAACTTAACC
ATGTGTGTTTCCTCCTGCCCACAAATTCCTGTGGCTTGTCCTGGTGGGGAATCTTCCCATGGAG
>NCL_C17

CACACAAAAATCTTACAAGTGTGTGCATTTTTTCCAGAGAATGCTAAGTTGCTTCAGTCGTGTCCGACTCTGTGCGAC
CCCA*CAGACGGCAGCCCATCAGGCTCCACTGTCCCTGGGATTCTCCAGGCAAGAGCATTGGAGTGGGTTGCCATTT
CCTTCTCCAATGCATCAAAGTGAAAAGTGACAGTGAAGTCACTCAGTCGTGTCCAACTCTTAGCGACCTCATGGACT
GCAGCCTACCAGGCTCCTCCATCCATGGGATTTTCCAGGCAAGAGTACTGGAGTGGGGTGCCATTGCCTTCTCCTTT
CCAGAGAATGCTTTTACCAAATTCTTAGGGTAGTCCCGAATCCAGAAAAGAACCACTTTACCAGGTACTGGGAAAGG
AAAAATGAATGCCTTACTAAATGCTCAAGATACAATCAACCTGATCTCTGAATGAGAGTAATTCACAGGTCAAAACA
GAATAAGCACTCCCGTGTTTGGCATAACTGCCAAATGGCAGCTCTGAGCAAAGCCAGTTGTTCACACTCTCTTCCCT
CACTGGGTACCATCACAGGCAATTCAGCAAACACCTCTGGACCTCTTGCATCTCTGTTATGTTATGAGACTTGCTCT
TCTTGCCCTGCAAACAGTGCTTCTGCTGACACAGGTCGAGTCTGTCCTATGAACAATGAGATCTGCTTCATTTGGTC
TGAACTGCTAGCTTTGGCCCAAAGCAGAAAAGCTGTCCCAA*GACCTGCCCAGGGAGCTACCATATCACCACCTCCT
TGCATGGAGGACACTTCATACAGTTCAGCCAGATAACCACTGCAGGGTCCCCAGAGGTCATAGTGCCCATATAGTAT
AGAAGTCACCTACTTTCAATTACCATA*TTTTTTTTCACTAAATATTAGGAAAAGATTTTCACACATGTGAGCTTGA
AAAAT*ATGAAGGTTTGAAAAGTAT*AAAAGTT*AAATAGG*AAAAAAAAAAG*ATTCTTGGCTTTG*AATTTCAGG
GAAACGTGCAATTTCCTAGTCTACCAAGGAAACTGTCATTTGGGGTCCTAAATTATTGTTCTGGCACCAGTGACTTA
CTGTGCCTCTAGGAAAGGCACTTAACCCTTGCTGGGCCTCTCTTTCTATCTTTGATAGAACATGATTTCCCCCTTAC
TTCAACAGGATTTAGTGATGAAGAACAAACAAGGACCCAGAATGTGCCATCATCTTGAGAAAGGAGAAATGCCAGTG
TATAGTATCAGAAGTATGAACAGGATACAGAGAAGGATACAGAGAAAACTGACAGTTTCCTCCTCTGTGCTCCCAAC
TCCACCCTGCATATGATCACGCTTATTTATTTATGTGTTTCTCCTCCAAGAAACACATCTGGCTCATCTCTGTATCT
TCAGCACTAGGAAATGTACCTGCTACATATACAGCAGGAAGATGGTCCTGTGTACTGATCAGGCTTCAGCTGTGCAT
GCTGGATAGTTTCTAGGAGGCACCATTCACTTAGGCTGTGAATTTTGGATGTTATTTAGTCTCTCTGAGCCTTAGTT
TTGTAAAATGGGGATAAAAATAGCATCTATGCCATAAGACTGAGGGTCTAAATTATATTCATTGTACAGGGTCTAAA
ACATAGTAAGCACTCAATATATG*CCCCCCATTATTAATTAAAGTACAGAAACACATTTACTTGAAATCACTTTATG
AGCTAGTAGGAGACCCACTCAAACCTCAGCCTAGTCAGTCTTCTCTTCTATGCTCACACAGTATTAAACAATTCCGT
TACTTCCCTTCTACACTTCTATTTAATTTTTATGATTTGATTTACCTCAGTACCTGATGAAAAGGCATACACATAAA
CAGAACTAGCTCTCAT*AAC*AAAAAA*AA*CACCTATGCCAAGACATACCACATTAATCTAGAAATA*TTTTCCTA
AAACAATCTAGTTGGTGGTGCTAAATTAAATTTTTATGTCTAAAGTTCAGGCCTTAACTTATCTGGGAGGTGCAAAT
TTCTACTGAGTAGGTATATTTACCACTGCCTTAAAACTAAGTAGAGAGATCCTAGATGAATACAGCGATCACAGACG
TGGTTCAGGTAACTGCATAACTAGTAGGCTACTTCTGACAGTTTATGAAGACAGTGGATAAGAAGATAAGATACAAA
GACTGTTTCAGAATCCAAGCAAAAGCTGTAGAACTTATACTCTGAAAAGCAATTATTCTCCAAAGACCATCATCTCA
CATACCTTTAGCCCTAAATTCCCAGCTGTCAGTTCATCTTTTCATTGGGAGAGCAACTCTCAAGCCCTTATGATCAG
CCACACTGACAAGTTTACCTACAGAATTTTTAAGGAACACTGCTACATGAATGGCACAGGAAGGACATGTATAAACA
GGTACAAGAAAGAAGTTCAATTTGAAATGCAATTTATTAGCTGTCACACAAATGTCGTTTCAAAGAATTACAGAGTA
ATTGTTACGCGATAACTCTCATATTCCTGACTTTT*AAAAACAATCACTAAATTCTAAATACTTATCTAAGAGTGGA
CACTGAATTCTCTAACTTCAGAGAGGAAGCTGAGCTTTGAACATTTTCATCTTGTAATTATAAACCCCTTTTCCCTA
TATGGGATGGGATGGATATTTGCATCATTTGCAAATAAACACTTTGAAAACAACTCCCAAATAAGTTTTCTAAAGCC
ACTAATCTTATTATAACATAATAGTAATTGAGAATACTGCAAGAGAAAAAGAAATACCCTACTTAGATATAAGTTCA
AACCAGCTTATACCAAAAGGACTCTAAAAAACATCAACAAT*ATATGTACTTTTCCCTTGT*AAAAAG*AAAACAAC
AGTGTGAAATTGGGACTTTTACAACTTTTTCATCATTATCATAAGTATTTCATGAATAAATCTTCAAAGAGATTCTT
TTTCCTTTTCTTCTATCAACTCCAAGTGATTTGCATTTGAGTTTACAGGTCTAGGGAGTACCTCACAAAGCAGGTTC
ACAAATATTCTTCCAGTTCTATTTTTAA*TAAAT*CAGAACT*AAAAAATT*CCCAAAG*AAC*AAAAG*CATTTTT
AA*TTATTTTCATTGCCTTTAGCTAAACAGCACACAAACAGAAACTCATTTATTTTG*AGCTGTGTATATCACAACA
GGGCAATTCAAATGCACAGTCAGACATTATATTTACATATTTCAAAGAGAACAAATC*TCTTTTATTATTCTATAAA
TTTGATATATAGATGAGTTGTCA**TTAAAAAG*TTTTAC*TTTTTAG*AGTAAAAACCTACCTAAAATATAAGCAT
TACTGTTCCAGAATAATTGTAAAGGCATTGTTTCTTAGGTCAAATAGAAATCCATAAAAGTATTATAATTATGGATT
ATGTATAGAGAGAGTTGTTTTTCCTACAATCTGAATTTTGAAAGTGTAAGTCAATTTAAAATGAATTGCTATAGTTG
CTG*AAAATAAAAAAGT*GGATATGGAAAGAAAAACTTAAAGTTGGCCCATTAAATTAAGAGGTACAAGGAGCAAAA
CATGATTAGCA*TTTTG*TAATAATACAAAGATCACCTGAACTATGAAATTCAAAACTAGTTCTATAAATAAGAACA
TCACAATTTATATTGTAGTGCTTTTTTA*TACAAATAAACGCTTCATCTAAATGTTAAATATAAACATCATATATAT
TAGCACAACTGTAAACACTTTTAAAATGAGAATGC*AAAAAAACAAG*ACCCTTTCCAGTGATTTAAATGGAAAATG
AAGCTTATGCACATTCCTAAATTTCTTATATTCTCTAAATTGGTATTTTATAAT*GGAACGGGGTGGC*AATAAATC
ATGTCAAAGGGGAGGGTGGATAAATAGGTGTTTGGAAGAAAGAGGGTATGATAGGTATACAACTT*AAAAAAACTCT
TCTTACTCTTCACTAACCTGGACTGTGAACC*ATTTTTCTAAAT*GTTAAAAAAT*ACCACATTCCTTTACACCATA
GTCTATTATAATGTCTGATTAACTTAATCCATCCCAGGTCCCTCCAGTCAATAACTTTTATTGTACAAAAAAATTTA
ACATTAAATCTACCATCTTACCCATCTTAAAGTGTACAATTCAATAACACTATGTTCACACCACTGTACGACACATC
TCTAGAACTTTTTCATCTAGCACTACTGAAACCCTCCCCCCA*GCCCATAGTAACTGTCTTTCTGTTTTCTGTTTCC
ATGATTTTC*ACTACTTTAG*ATACTTC*ATGA*GTGGAATCATATGGTATTTATG*CTTTTGTGATTAG*TTTATT
TTGCTTATTGTAATATTATCAAGGTTTAT*CCACCTTGTAGCATATGACAGG*A*TTTCCTTTTTTTC*ATGTTGCA

TACATAATATATCATATTTTATTTGTTCACCTCTTGATGGTCATTTGGGTTGCCTCCACCTCTTGGCTACTGTGAAT
AACGTTGTGATGAATGTGGTTATGCAAATTATTTCTTCGAGATCTTACTTTGAATTCTTTTAGATATATGAATTGAT
GGATCACATATTCATTTTAAC*TTTTTGAGGTACCTCCATACTGTTTTCCATAATGGCTACAGCATTTTATATTCCC
ATCAAGAGTGCACAAGGGTTCCAGTATCTGTGGCACACTTACTGACACTTGTTATCTTATGTTTTTTTCGATTGGAG
CTATCCTGATGGGCATGAAGTGATTATCTCATCTGTGC*TTTTGATATGCATTTCCTTAATGATTAGCAGTGTTGAA
CATCTTTCCCTGTGCTGTTGGCCATCTATGTATCTTCTTTGGAGAAAGGTCTATTCAAGTCCTTTGCTTTATTTTTA
ATCTAAGTTTTTTTTTTTTTTTCTGTTGGGTTGCTTGCCAAATTCAACTGAAGCTTTT*CCCCCATGTTTCTTTTCG
GAATGGAACCATTAAGGGTATTACACTTTAGATCTTAAATCCATTGTGAG*TTAATTTTGAGGGTTCAACTTCATTA
TTCTGCATATGAATATCCAGTTTTCCCAATGCCA*TTTCTTTAGAG*ACTTTTTCCTATTATGCAGTATTGGCATTC
CTAGGTATTTGATCATATATGCTATAAGAATTTG*TTTTTGGGCCACTGGTCTATATTATCTATTAATATCATTATG
TCAATACTATACCATCTTAATTACTACTGCTTTC*TAGCATGTTTTGAAATTAGGGAATGGAAGAGCTCAAACTCTG
TTGTCATTAAAGTTGCTTTGGTTACTCAAAGTCCCTTGAGATTCCAT*AG*AAATTTTAGGA*TTTTTTTTTTCTTA
ATTTCTGTAAAACATGCCAATGGGATTTGATAGG*AAAAACAATGAATCTGTAGATCACTTTGGTAGTATGGACATT
TTAACAATGTTAAGTTTTCTAGTCCATAATTACAGGATGTTTGTTTATTTCTATTTTC*TCTGAG*TTTTTGCAATG
TTTTACAGTTTTCAATCTCCTTGGTTGAGTTTATTCCTAAATATTTTACCCTTTTTGATGGTATAGTAAATGGGAGG
CTCAGACGGTAAAGCGTCTGCCTACAATGCGGGAGACCTAGGTTTGATTCCTGGGTTGGGAAGATCCCCTGGAGAAG
GAAATGGCAACCCACTCCAGCACTCTTGCCTGGAAAATCCCATGGATGGCAGAGCCTGGTGGGCTACAGTCATGGGG
TCACAAACAGTCGGATATGACTGAGCGACTTCACTTTCTTTCTTTCTTTGTTTCCTTTTC*AGTTTGGTTACCTAG*
TTCCTTTTCAGAAAGACAACTAACTTTTTAAAG*TTCTAGTTATTTA*TTTTTGGCTGTGCTGGATCTTCGCTGCTG
CATGGGCTTTCTCTAGCTGGGGTGAGT*GGGGGCTGCTCTTTGCTGCAGCATGCAGGGTTTCCATTGCTGTTGCTTC
TCTTGTGGAGCACTCGCTCTCGGGCACACAGGCTTCAGTAGTTGTGGCTCCAGGCGCTAGAGCATGGGCTCGACAGC
TGTGGCACACAGGCCGAGTTGCTCTGCAGCATGTAGGATCTTCCTGGATCAGGAATCCACCTGTGTCTCCTACACTG
GCAGGTGGATTCTTTCCACTGAGCCACCAGGAAG*CCCAAAAGGCCACTAACATTTGACTGCTACTTTGTGTTGATT
TTGACTATAACTTTGTTAGTCCTAACACGTTTTTTGTTGGTTTGAGGCATCTTAAGGGGAG*TTTTCTACATAAAAG
ATCATGTCATCTGAAAACAGAAGTTATTTCACTGTTCCTTTCCAATCTACCACCAGCCCAATCGTTCTGGCTAGAAC
TTCCAGTACTATATTGAACAGCAGTGGTGCAAATGGGCATCCTTGTCTTGTTCTTTACCTTGGAGGAACAGCTTTCA
AGAAGTCTTTCCACACTGAGTATAATGCTCATTGTGGGCTTTTCATATGTTGAGGAAGCATCTTTCTATTTCTAGTT
TGCTGAGTATTTTTATCATGAAAGGGTGTTGAATTTTATTAAATG*CTTTTG*TAATTTTTTTCTATACTGAACTGA
GATGTGGGTTTTGTTCTCCATTTTGTTAATGTGCTCTATTACAATGATTTA*TTTTTGTATGCTGAACCAGTCTTAC
AATTCCAGCTATGAATTCCACTTGGTCTGGTGTATGATTATTCTAATGCACTGGTGAATTCAATTTGCTAGAATTTT
GTTGAAAGTTTTTGTTCCATCA*GGGATACTGGTATGCAGTTTTCTAGTTTGTTTGTTTCATTTCGTCTTGTCTTGT
GTCTTTGTGTGGTTTT*GGGGATCGGAAGGAAATGGCAACCCACTCCAGTGTTCTTGCCTGGAGAATCCCAGGGACG
GGGGAGCCTGGTGGGCTGCCATCTATGGGGTCGCACAGAGTCGGACACAACTGAAGTGACTTAGCAGCAGCAACACT
ATCCTCAAAGAATGAGTTTGGGAAAAGGCTTGGTGTTAATTCTTTAAACGTTTGGTAGGATTCTCCACTGAAGTCAA
GGGGTCCCAGGC*TTTTCTTTGTTGGGAGGTTTTGCATTGTTCCTTCAATCTCCTCACTAGTTACAAGTCTGTTCAG
ATTTCTCCACGATTCAGTCTTGGTAAGTCATATGCTTGCAGGAATTTATCTACTTCTTCTAGGTTATCTAATTTGTT
GGTGTTAACAGTCTCATAATCCTTTTAATTTCTGTGACATCTGTTATAATGTCTCTTCTTTCATTTCTAATTTGAGC
TATCTGAATCTTTCCTCTTTTTTCTTAGTTTAGCTAAGGGCTTGTCAA*TTTTGTTGATCTTTCAAAAACCAACTCT
TAGTCTCGTGATTTTTTTTGGAC*TCTTTTTCTCGTCTCTATTTCATTTATCTCCTTTCTAATCTGTATTTCTTTCT
GTGTGCTGACTTCGGCTTTAGTTTGTTTTTCTAGTTTCTTGAGGTACAAGATTAGGTGGTTGATTTGAGTTCTTTC*
TTTTTCAATGTAAGCATTTACTGCTGTAAACTTTCCTCCTCCTACTCTTCTGCTACA*TCTT*G*TAAATTTTGGTA
TACCGTATTTTTATTTTCGTCTTCAAATATTTTCT*AAATT*CCCTGAGACCTCTTCTTTTGATTCA*TCGGTTGAG
AATGTATTATTAAATTTCCACATGTTTGTGGATTTTCC**TGTTTTTTTTTCCTGTTACTGATTCCTAGTTTCATCA
CATTGTGATCAAAGAAGATACTGTATATGATTTTAATCTTAGATTTGTTAGGACTTGTTTTGTAGCCTAACCCATGC
TCTATCTTACAGAGTGTTCTGTGTGCCCTTGAGAAGAATGTATATTCTTCTTTGCTGGGTGCAGCATATCTGTTAGG
TCCAACTGGTATACAGTGTTGATTAAGCTCTTTGCTTACTTACTGATCTTCTGTCTGGTTGTTCTATCAATAACCAA
AAGTAGGGTATCAAAGTCTACTACTATTAATATTATTGTGTTGCTGTCTATTTCTCCCTTCAGTTCTCTCAACTACT
GAATCATATACGTGGGAAAGAATTTAAAATAATTATACAGGGACTTCCCTGGCAGTCCAGGGGTTAAAACTCTGCCT
TTTCATTGCAGGGGGGCACAGGTTTGATCCCTGGTTGGGGAACCATATGGCATGGCAAAAAAAAA*AAAAAAAC*AC
AACCAAACCAAACCAACAACAGA
>NCL_C18
ACTTTCTTTAAAAAAATTTTTTTTTTAATTTTTTTGAAGTCTAATCAGTTTTTTTCTTTTATGG*TTTTTTTTTTT
GCATTATGTTTAAGAAATCATCAACCCCCAAGGTCATAAAGATAGTTCTTTTTTTTTTTTTTAAGTAAGCTTTTATT
TGTTTTG*TTTC*ACATATTATAGGTCTTTG*CCA*CTTTG*AAATTTTT*G*TGAGTAGGGTCAAGTTTAC*TTAT
TTTCC*AAGCAGATGTCTAGTGGCTGCTTCTTCACTATTTATTAAACAGACCATACTTTATACCTGAATTTCAGTGG
ACCTTTGGTCATTAATCTGTGTTCCTGTATATGTGAATCTATTCC*TA*GGG*TGGGGTTTTTGGTTGTTTTCTTTT
GATTAGTATTTATCTATCCTTGTGCCAGTGACACACCATCTTACCATATTTTGTCCTCCAAC*TTTATTTTTCTTTC

AAGATTATCTTGGTTGTTCTGGGTCCTTCCTTTGTGTCACTCTATACATTTTAGAGTTGGCTTTTCAGTATCTACAC
ACATGTGGAGGCATGCACACACATATGCCTACTGGGGTTTTGATTTGGATTGCATTGAATCTGTAGATTATTTTGGC
GGAGAATTGACATCTTAACATTCTTGAATCT*CCAAAACAATAAACATGGACTATTTCTTCATTTCTTTAAGTCTTT
GATTTATCTCAGGAATGTTTTA*TAGTTTTCAGTATAGAGTAGAGGTCTTGTACA*TTTTTCATTAGATTTACTATT
AGACATTGGTA*TTTTCTGGTGCTTTTGTAAATTGTATTC*TTTTTTGACTGAAAGAAAAC*TTTTTTC*TTTTTTA
AG*AAAACA*TTTTTGAATTTTGTGATTATAATCATGACAATAAAGATGTAATAAGTGAATTTTAATTAAACT**AA
ATTGTCTGCAAAAGCTACAGTCGGCAACCACTGAATCATTTCAACTGATGCATGTCTCATTCTCTTGATTCA*TTCT
TTTTTAAG*ATTTGTTTGTGTATGTTTATTCTTTTGGCTCCGGTGGGTCTTCGTGGCTGTAGTGGGCTTTCTCTGGT
TGTGGTGGGCGGGCCTCTCATTGCAGTGGCTTCTCTTGTTGGGAGTGTGGGCTCCAGGTAGGCTTCCCAGGCTTCGG
TCGCTGTGATCCGGGGCTCAGCAGTCACGACTCTTGGGCTCTAG*AGCATAGGCTTAGCAGTTGTGGTGCGCAGGCT
TTGTTGCTCTGCAGCTTGTGGGATCTAACCAGTGTCCCCTCCATTGCAGGGTGCATTCTTTACCACTAGATGACCAG
GGAACCCCCTCTTGATTGATTCTTTAGTTTAAAAAATTAATTAACAATTAAATGACATTTTAAAAATGAAAAATTTA
AAAGGAAACTTCAGAAGTATTCAAATTTATTTTGAGGACTGTTTTCCATACAGTAAAATGTACAGA*TTTTTCTTAT
GCAATTTGATAAATTTTTTT**CCCCTAAATATCCACCCGTGTCGCCACTACCCAGTTCAAAATATAGGACATTGCT
AGTAGCTTTAAG*CCCCTTCCGAATCAATCCTACACT*CTCCCCCGTAGATAACCATATAGGTAGCTGTTACTGAGA
ATGCTTTTACTTATAGAAATTTTGCTTCATTATTTGAAACTAAAGAAGTAATTGCTTGGATGTTTATAAAGTATGGC
TTTTA*TTTTCAGTAGAAGTGATAGACAGTATTGATAGTTGGGGTTGAACTATGA*TTTTTTTTTTTACC*AAGTAG
ATTG*TAC*TTTTTTTAC**TTCATAGGATATATAATTTGACCTGTGTTGTATCATTCATGTTGTAAGGAGCAGTAA
TTATTTCATTTTCATTGCTTTAGTAGAGTATGTACATACTTACTAGTTTAGATAATCCGAGGTAAAATTATCTGCTC
AAATACTTTTCTTGACGCAGTGTCTAACATTAATACAGAGATTGTGGTGGCTTTCTTACTGATATACTGAGAGAAAC
TGTGGTACCTGAGGGTGTTGGGAGGGAGGAACGTTCTGTGCTCCTCTAAGAGCCAGGGAATCAGCTTGGTGGTAAAG
AAATTACAAAACTTTGTACATTTCCTCTGTTTTCTGAAATTTATGTACTTCTGCTAGTAGGAAATCAGTGACTC*TT
TA*TTTTTTCTTTAAGTTTTTATACTGTGGAAATACTAATAATCGCTCACATTTATTGAGTTCATACTAGGTACAAG
TACTATGCTGAGTGCTTTACTTAAATAACCTGATTTTATTTTCACCGAAACTAATGTTTAACTAATGCT*CTGTAGA
CTGAGGTTT*GGGGTCAGTGTAAATGATTCCATCCTTTAAATTTATGCTAGGGCAGGTGACCCCCATTTTCTAACTG
TTGGCTCTTTTC*ATGAACAGGTATGTCATCGATGGGGCCACTGCTCTTTGGTGTGCAGCAGGAGCGGGACATTTTG
AAGTTGTTAAGCTTCTAGTCAGTCATGGAGCCAACGTGAACCACACCACAGTAACTAACTCAACCCCTCTGCGGGCA
GCATGCTTTGATGGCAGATTGGACATTGTGAAATACTTGGTCG*AAAATAATGCCAACATCAGCATTGCCAACAAGT
ATGACAACACTTGCCTAATGA*TTGCGGCCTATAAGGGACACACTGACGTGGTCAGATATCTTTTAGAACAACGTGC
TGA*TCCGAATGCTAAAGCACATTGTGGAGCCACAGCATTGCATTTTGCAGCTGAAGCTGGGCACATAGATATTGTG
AAGGAGCTGATAAAATGGCGTGCTGCTATAGTGGTGAACGGCCATGGGATGACGCCATTAAAAGTAGCTGCTGAAAG
CTGTAAAGCTGATGTTGTCGAACTGTTGCTCTCTCATGCTGATTGTGACCGGAGAAGTCGGATTGAAGCTTTGGAGC
TCTTGGGTGCCTCCTTTGCAAATGATCGCGAGAACTACGACATCATGAAGACATATCACTATTTATATTTAGCTATG
TTGGAGAGGTTTCAAGATGGTGATAACATTCTTGAGAAAGAGGTTCTCCCACCAATCCATGCTTATGGGAATAGAAC
TGAATGTAGAAATCCTCAGGAACTGGAATCCATTCGGCAAGACAGAGATGCTCTTCATATGGAAGGCCTTATAGTTC
GGGAACGGATTTTAGGTGCCGACAACATTGATGTTTCCCATCCCATCATTTACAGGGGAGCTGTTTATGCAGATAAC
ATGGAATTCGAACAGTGTATCAAGTTGTGGCTTCATGCCCTGCATCTCAGACAGAAAGGCAACAGAAATACCCATAA
GGATCTTCTTCGATTTGCCCAAG*TTTTTTCACAGATGATACATTTAAATGAAACTGTGAAGGCCCCAGACATAGAA
TGTGTTTTGAGATGCAGTGTGTTGGAAATAGAACAGAGTATGAACAGAGTAAAAAATATTCCAGATGCTGATGTCCA
CAGTGCTATGGACAATTATGAATGTAACCTCTATACCTTTCTGTATTTAGTGTGCATCTCCACCAAAACACAGTGCA
GTGAAGAAGATCAGTGCAAAATTAAC*AAGCAGATCTACAACCTGATTCACCTGGACCCCAGAACTCGTGAAGGTTT
CACCTTGCTGCATCTAGCTGTCAACTCAAATACCCCGGTTGATGATTTCCACACCAATGACGTTTGCAGCTTTCCAA
ACGCGCTTGTCACAAAGCTCCTGCTGGACTGTGGTGCTGAGGTGAATGCTGTGGACAATGAAGGGAACAGCGCCCTT
CACATTATCGTGCAGTACAACAGGCCCATCAGTGA*TTTTTTGACCTTGCACTCTATCATCATTAGTCTAGTTGAAG
CTGGCGCTCACACTGACATG**ACAAATAAACAGAATAAAACTCCGCTAGACAAAAGTACAACTGGGGTATCAGAAA
TACTACTTAAAACTCAAATGAAGATGAGTCTCAAGTGCCTGGCTGCCCGGGCAGTTCGGGCTAATGACATTAACTAC
CAAGACCAGATCCCCAGAACTCTTGAAGAGTTTGTTGGATTTCATTAAGTGACTGGATATGTAAAATCGTTTAATGT
GGTGCTAAAAAGTAAAGGACTTTAATCACAGACAATAGAATTATGTGTTCATAAATTCTACTTTTCTTTCCACTACC
CCTCCTCCCTACGTATCCTTCCTTAGTTTTGTATTTGGTCTTCTTGTCTCATGTGGTTATTGATTTCAAATACACTT
TAAACAAAGCCACATTGTTTAGTGTAACTATAATCTTAAGGTGTTTGGATATTGGTCACTTAGCTATTTTTTTTTTT
TT
>NCL_C19
CGACTGAGCGACTGAACTGAACTGATAGTTTCCAAAGATTTTTTGTG*TTTTTAATCTCATATAAGCCTTTTAGTTA
TATGTCTTTAAAATGGGCTTAATAATGGTTATCTACCTTATAGGATTATTGGGAGAAGGTTATAATGAGAATTAAAT
GTGCTCTATTGCTTACTACAGTGCCTGATACATTGACATTATTAACTCTGTGTTCAAACTCTTTCTTTCAATTGACT
TGTGTCTCTTTTGTTTTTAAA*TTATATCTATTTTTAG*TTAAGGTATAATATCAAGTAAGGTGCATAAATCTTAAG
TGTATAAATGTTATAAAACTTTCATTTGCTTACATCCATGTAATCACTTTCTAGATCAATAAGCAGAACATTGATGT

C*AC*TTTAAAAGTTTCCCTTGTGCCCTTTCTCAGTCATTTCCTCTCCCTT*CCCCTCCCAGAAGCAATGATTATTC
TGAGTTCTATCTCTGTAGATTAGTTCTGCCTACACTTGAACTTCCTATAAATGGAGTCATACAGTGTACTCTTTTTT
TTTCTGTCTGGCTTCTTTTTGTTTAAGACAGTATCTATGAGAGTCACCCACTGAGTTGCATGTATAAGTATTTTTAA
AATGTTTTATAATATTATGTTATATGAATGTATGGCAGTTTATCCTGTTGACAGATATTTTGGTTGTTTTTAGTTTT
AGTATTTGACTATTATAAATTGCGTTGGTATGAACATGTTTGTATATGTCTTCTGGTGGAATTTGTGTTCATTTCTC
TTAAGTATATACCAAGGGGTGGGACTGTTGGCTCATAGGGTTGGTGTATGTTTAATACTTTCTCAAGCAGTCAACCA
CCAAAGTGGTTGAACCAGATTATATTCCCACTAGAAATTTCCATCAACAGGAGGGGTTTCAGTTACTTCACATCCTG
GCCAATACTTGGTATTGTCATGTCAGTATTATTTATTTCATCTATTCTAGTGAATGTGTATTGGTAACTCCTGGTTT
TAATTTGCATTTCCTTGCTGAGTAATTATCTTGAACATCTTTTCATATGTTTATTGGTTGTTTGAATATATGCTTTT
GAAGTATACTTCTTCAGTCTTTTC*ACATTCTTTAAAAAAAATGCAGTTGTTTGCTTTTTTAG*TGATTTA*TCAGA
TAATCGATTGTAGATATCTCCAGTGAGTGGATTGCTGGTCACTTTCTTTAAAAAATTTTTTTTTAAATTTTTTTGTA
AGTCTAATCAGTTTTTTTCTTTT
>NCL_C20
TTTCCAATGTTTCAGTTCTGCATTGTCTAAAAGATAATACATCATAGCTACCCACAATGTGTGCTGCCTT*ATAACC
AGA*TT*CGGGG*AACAAAAATAATTTTCATTAATAT*AAAAATT**CAAAGCAACATAAGCAACATCCTGGAAACC
ATTCAAAAGCCAATTGCCGACTACTTTGAAGT*ACAAAAAACCCA**GAGATCATTTCTTTATTTTTACC*ACTGGT
AGTGAAACACTGGCTATTGATGAAGGTACTAATAGGAAACGGAGTCAGTATTTATTTACGTCCGAAGTATATATTTC
TCCCACCTTCCTAGATTCCAACCTCCTCTATCCCTGATCTTTCAATCTTTCTTAAACTTACCCAGAAGAAGACTGTA
TAGATCAAAGACATTTATAATGCTATATGGCAACACAGAATGGGTAACTCACCTTGTAAGTCATAAGGCATGGGTGT
CATGTGGAAAGGATGTCTGTAGTCTTGGATGAGGGATGTGTTAATATAGCTTGTGTCTACAGCAGGGAGGCTTGGGG
TACGGGACACTGGAGATGCTTGATGTGGAAGACTCAGTATGCTAGAAGAAAAAAAGGTTTGCAAAATTTT*C*ATGA
GTCAGCTGAACTT*AAGTTTAAT*AAAAGGGCAGAATCTCACAATGAGTCCCAG*TTTTTG*TGCTGAAAATAAAAA
CAAAAG*CAAAACAATATATAATTTAGACTACCATGATTATTAGG*AAATAATTT*G*TATACTTAA*TAGCAGCTT
TC*AGAGGCTATAAGGAAAAGAGATGGATGCAATGGTATCCAACATTCATAAATCACATGTTGTCTTCGGAACACCA
TCTACACTCTTAGAATGCAGGCCAAAGGCATAGTTGAGCCACTTCAATGAAATCTAAGTGAAAAGTATAGCAGACAA
AATTCTAAAACCTTTAAAAAGTAAGTACTAGGGTGTTGGCAAGTAAAAGAATAGGTAAATCCTTGCAAAGAAAACAC
ATCTAACAAGACTAGGGGAATGAAGCAATCTGAGCAGTGAAGAGTTAACAGCTGGCCCTCAGACTTGTAAGATTTAT
GAGAATAAAAACTAGGTACTGTTTATCTTGGAATACCTAATGCCCAGCACAGTCACGAGGAATGCTGTTTGCATAAA
AGTCTGCTGAATGAACGAAGGCAGCAAGACAAGCAGTAGAGATCCTCCAAAATGAACAGGTGATGGAGCGTTTTTAA
AGGGAGGACAGTAGAAATGGAAAAAGAAAAAAGAACAAAGAGGGATTAGGGGAGAATAGAGGGAAATGTTTAAGTAG
CTGAGTAGTTAAGTAC*TTTTTTACCAAAAATTAATGAAAAGCAAAAAGGGGGATGTAATAAAGGAAAAAGAAAAGG
GAGAATTAAAATGATTTTGAAAAGAAAAAGG*ATTCTCATAGTCCTGTGTGAAATCATAAGGGCCATATGCTATCCC
AAACATATAAGAGGCCATCAGATATGTTATAGAAAACAGAGAGACAGACAGAGGCTGTC*TTTTGCATTTAAAAGTT
CATGTAAAGTGCTGGGGGGTATGGAGATGCAATTTCAATGGTAGAGAGTATCCCTAACGATGCCAAGCAAGGCTGGA
TTTCCAGGAACAG*CGTTCTAGTTATGAAGCGCTACC*TTGTTTTTATGAATGGGATATGAAGC*ATAAAATACACT
CTCATCCATCTC*ATAAAATTGTTTTGTCTAATGG*AAACTT*AATTAAACTTCTTACATCTACAGAAAGTACCATT
TTCCATTAGCTATTTCCTCTGACAATACTCCAGGGATTATAAAGGATTACAAGAGAGACAAAAGCAGTTGAAACGCT
AGTCTTTGAAAAAC*ATCTTGGTATACCAAAGAGCAAAGGAAATTTCTGGAGGAATACTAAACACTCAGTGGCTGAA
GACGGCAAAATATTTTACATTTACTAATGTTTGCCTCTTTCTACTTCCTGGCTGGC*TCTTTTTCTTCCAATTTTGT
GACTGTATTCTAAAATGATTTCCACCTTCACACTGCACAATAACATATCCAATATGCTGGGTTTAGGTTTCTATTTA
TAGGCTAGTCATTACTGAAAAATGGGCACTGGTATTGAATATGGTGTTATCCT*ACTAAAAAA**AAAAATCATAAG
*CTTTCAGGTAGC*TTGTTTTTAAATCTGAAAATAGAGAAAGGTTTGGCATTTGCAAATACTCTATCTGGGTTCAAT
TAGTCACTGGGAAGCTCAAGAAGTAGTTAATGCATCTAATGTAATTTCCCAAACATAATAAACAAGCTATTTTTGGA
GAAACAGCACAGCAAATATTCAAAGAAGAAATATTAGATTTCTAAACACTCAAAAAACATTTCCTTTAAGGAAAAAT
GTTTACCTATATATGACTACTATCAATATCCTTAAGAGTAGAACTTATGTATATATAATGATGCACTTAATCCATGA
AATATTTATACATCCAGATTTATGAAATAAATTTTAATAGTGCTAAAATGGCATTTAAAGCATGTAAGAATTAAAGA
TTTTAAAAATTTAAAAAAAAAAA
>NCL_C21
GTACTACTGTCTTC*TTTCTTTTCTTTT**TTTTTTTTTTGCTTTTTTGGCCATACTATGCAGTATGTGGGATCTTA
GTTCCCTGCTGCTGCTATGTCACTTCAGTCGTGTCCTACTCTGTGCGACCCCATAGACGGCAGCCCACCAGGCTCCT
CTGTCCCTGGGATTCTCCAGGGAAGAATACTGGAGTGGGTTGCCATTATCCTTCTCCAATGCATGAAAGTG*AAAAC
TGAAAGGGAAGTCACTCAGTCACGCCCAACTCAATGACCTCATGGACTGCAGCCCACCAGGCTCCGCAGTCCATGGG
A*TTTTCCAGGACCA*GGGGTCAAACCTGTGTCCCCTGCAGTGGTGGCTTGGAGTCCCAACCACTGGACTGTCAGGG
ACGTCCCTCCTGTAGGCTTTTTTTTTATTAANA
>NCL_C22
GGWAGTTTTGTTTTATTTCCTGTCTTTTTT*TTTTTTTTTTTATCAAGAATGGATGTTAACGTTTACTGACTGCCTT
TACTACATCTATTTAGATAATTATACAACTTCTCC*TTTAACCCATTAATGTGATGAATTAAAGTTATGGCC*TTTA

ATGCTAAATGATCC*AAATATTCTTGGGATAAATCCAGCTTAGTTATGATGAATTGCCCTTAATATAAGCTGTTGGA
CTTGATTTGCTAGTAGTTTGTTTATAATATTTCCGTCTATGTTTGTGAGGGATTCTGGTCTGAATTTTCTTTTTCAT
ACAGTTCTTGATTGATTTTATTTCAAGGTTTTACTAACCTCATATGAGTTGAGGAGCAGTGCTTTATTTGATATTCT
GTATGTGAAATTAGAATCATTTGTTCTGTGAAAATCTGC*TAGGAATTGCCTGGAAAATGGCAAGGCTTGTTTTTTT
TGTTGTTGTTGTTGTTTTTTAAGACATTGAAACTGATTTAATTTCCTTAATACTGTTGGAAACTCAAGCTTTCTTGT
AACAGCTTTGATAAGTTGCATTTTTCTAGAAATGTTTCCCTTTTTAATAAGTTTTT*AAATTCATTGGTGTAATGTC
CTTTATAGTATGCTCATTTTATCTGTTTAATCTCTTCTGTATTTGCAGCTTTCTTCAGTTCTCTGTTCCTATTATTA
CTTGTTTATGCGTGCTGCCTACTACCCTCTTTTGTAATGAATCTTACCAAAAGTTTGTC*TATTTTATCTGGCCTTG
GCTTGTTGATCATCTCTATTGTATATTTGTTTCATTTC
>NCL_C23
TACCGCTGGCTGGGGAAAAATTACTGGCTACGCGGGTACCCAACAATATATGGAAGCAATGGGCGTCCCGGGTTTTA
*TGCTGCCACTGGTGATTCTGCTTGAGTTTGGTGGTGGTCTGGCAATCCTGTTCGGTTTCCTGACTCGCACCACAGC
CCTGTTTACTGCGGGCTTTACGCTGCTGACGGCATTTTTATTTCACAGCAACTTTGCTGAA
>NCL_C24
TTCTCCGAGTGGTATGCGCTGGATCTGGTGAAAAACCAGGATGGCGCGGTGGTGGGTTGTACCGCACTGTGCATCGA
AACCGGTGAAGTGGTTTATTTCAAAGCCCGCGCTACCGTGCTGGCGACTGGCGGAGCAGGGCGTATTTATCAGTCCA
CCACCAACGCCCACATTAACACCGGCGACGGTGTCGGCATGGCTATCCGTGCCGGCGTACCGGTGCAGGATATGGAA
ATGTGGCAGTTCCACCCGACCGG
>NCL_C25
TTCGAGGCCGCGGCTCAGGCGCACACCGACAAAGAT*GCAGGCG*TTTTTGGCATCGGCAAAACGGTAGTTGAACTC
GG*TGACCGAACGCCCGCCAAGCAGTTGGCAGAATTTGAGGAAGCTGCCTTTTTC*TTCCGGAATGGTCACCGCCAA
CAACGCTTCACGCTGTTCGCCCAGTTCGCAGCGTTCTGAGACGTAGCGCAGGCCGTGGAAGTTCACGTTGGCACCGG
AAAGAATATGCGCCAGCCGTTCGCCG
>NCL_C26
AGAGTAGGACACGACTGAAGTGACATAGCAGCAGCAGGGAACTAAGATCCCACATACTGCATAGTATGGCCAAAAAA
GCAAAAAAAAAAAAAAAGAAAAGAAAAAGAAAG*ACAGTAGTCATCATTTATAATGGTGAGTCATTGAAAGCATCAT
TCCTTTAAGAGTAGAATGAAACAAAAGTAGTCTGCTATCACCTCTTCTGCTCCACAAGACAGGGGCAAGGCAGATAG
GAGAAAGAAGGAAACAACACTGTCATTATTTGCAGCTAACATAATTGTATATATAGAAAGCCCTCAAAGAATCTGTG
GATAAATTACGACAATTGATCAGAGAGTTTACCAAACCAAACTTGTTAAATACAATGCACCAAAGTTG
>NCL_C27
TCTCCAAAGAGATTTCAAGGCTACGCGGCCAAACGAGAAGTGGGTTACCGATGTTACTGAATTTGCAGTCAATGGGC
GCAAGCTGTATTTGTCTCCAGTAATAGATCTCTTCAACAACGAAGTTATTTCTTACAGCCTTTCGGAAAGACCAGTG
ATGAACATGGTTGAGAATATGCTCGATCAGGCATTC*AAAAAGCTTAATCCTCACGAGCATCCTGTTCTG
>NCL_C28
GGAGTTCAAATGCACCTGCTCGCGTGAACGTTGCGCCGATGCGCTGAAAACGCTGCCTGATGAAGAAGTTGATAGCA
TCCTGGCGGAAGATGGCGAAATTGACATGCATTGTGATTACTGCGGTAACCACTATCTGTTCAATGCGATGGATATT
GCTGAAATCCGCAACAACGCGTCTCCGGCAGATCCGCAAGTTCATTAATGAGTATGTCCGG
>NCL_C29
AGCCCGCACCGTTACCTGTGGTAATGGTGATGGTGGTGGTAATGGTGGTGCTAATGCGTTTCATGGATGTTGTGTAC
TCTGTAATTTTTATCTGTCTGTGCGCTATGCCTATATTGGTTAAAGTATTTAGTGACCTAAGTCAATAAAA*TTTTA
ATTTACTCACGGCAGGTAACCAGTTCAGAAGCTGCTATCAGACACTC*TTTTTTTAATCCACACAGAGACATATTGC
CCGTTGCAGTCAGAA
>NCL_C30
GGCCTGCAGGAAGGAGGCGCCCGGCTGCGCGCCGGAGCCGCCCGCCGCTCCCGGGTGCCGCCTCCGCGCCGCAGCCT
CCATGGCTAACCTTCCGGCCAGGCCCTTCGCCCCGCTGCCGGTAGCTGCAAACAGGTTCCGCTGGCCGAAGGCTGCG
AAAGCGAACATCCA*GCGCGGGGGCGGGGCTGGGGCGGGCGAGAGCGAGAGCGCCCGCACAGCGGCGCGGGCCGGGG
GCGGAGCCGGCGGGGGGCCGTGACCAGGAGACGCGCCGCAGAGCGAGGGAGGCGGGGGCGAGCTCCCCGGCCGGG
>NCL_C31
GAAAGTGTGTATACGGCAACGTATCGGGGGTTCGAATCCCCCCCTCACCGCCATATTTAAAGAAGAGCTCGTACGAA
AGTACGGGCTTTTTTTTCGTATATTGCACACACCGGGGGGATGAGAAGCCCCGACCGGGGTTCGACAACTGGCGACA
GCCAGTTGGACAGACCGTGAACGCAGTGAGCGGGCTGCCCGCAGGGCGAGCGAAGCGAGTCAATCCCCCCCT
>NCL_C32
TGTGCGCAATCACTTGCGAACGCACGCGGAAGGTCTGGCGCACGTCTTCGTTGACGATCAGGTCGACGTAGCGCTGG
CGATAGCGCTGCTCGGTGTCGGTCAGGCCGTGGTGCTT
>NCL_C33
TGGCTGCGATCATGTCGACAATTAACGCCCAACTGCTGCAAAGTTCCGCTACGATCATTAAAGATCTCTATCTGAAT
ATCCGTCCGGATCAAATGCAAAACGAGACGCGTCTGAAGCGGATGTCGGCGGTAATTACGTTAGTTCTCGGCGCGTT

GCTGCTGCTTGCCGCCTGGAAGCCGCCAGAAATGATCATCTGGCTGAATTTGTTGGCCTTCGGTGGGCTGGAAGTCC
G
>NCL_C34
GAGGTTACTCCGTTCTACAGGTTACGACGACATGTCAATACTTGCCCTTGACAGGCATTGATGGAATCGTAGTCTCA
CGCTGATAGTCTGATCGACAATACAAGTGGGACCGTGGTCCCAGACCGATAATCAGACCGACAACACGAGTGGGATC
GTGGTCCCAGACTAATAATCAGACCGACGATACGAGTGGGACCGTGGTCCCAGACTAATAATCAGACCGACGATACG
AGTGGGACCGTGGTTCCAGACTAATAATCAGACCGACGATACGAGTGGGACCGTGGTCCCAGACTAAT
>NCL_C35
GAATGTGCGCACTTTTACGGCTAATCACCACATAACCTGATCCGGTGGAGATCCCCAGATGCGGCTGTTTGTTCTGC
CGTGAGGAGAGAAGCACTGCATCCAGTTGTTGCGCCTTAAGCCAGTCGCGCAGCGAAGCGAGTAATGTCATCTGGAT
TTTCCTTCTTACAGGCTATCGATTAACAATTTGCCTTTACGGAACATCATCAG
>NCL_C36
ACGCTTTTTAACGTCCGCATACTTGTCAGGCTTGCTGAGAGCCTTTAGATGATAATAAAACGTACTGCGCGGTATCT
CCGCAGCCCTGAG*AAGCTCATCAAGAGGATAAAACT*GCCTTAGCTCGTTGAGTACTTTCACTTTTTCGTGGGATG
AGGCTAAGGCTTTCAGCTTTTTTAGATACATAAGCCGCGTTTCAAGAAATCGAACTTGCCTTTCAAGATCCTCAATG
CGTCGGTCTTTTGACAGCT
>NCL_C37
GGTCGTGGTCGGCGGTACGGCACTTTCCGGCGGTCGCGGTTCCCTGTTCGGTACATTGCTTGGTGTGCTGGTGATTA
CGCTAATCGGTAACGGTCTGGTGCTGCTCGGTATTAACTCCTTTTTCCAGCAGGTGGTACGCGGCGTCATCATCGTG
GTGGCGGTGCTGGCGAATATCTTGCTGACCCAGCGAAGCAGTAAAGCGAAACGCTAAACCTTAACGCCCCTGGCGCG
GGTGCGGCAGGGATGAAAATTAATCGTATT
>NCL_C38
TTTAGTGACCATTACAAAACCTGTTGACAGAAAGTTAAAACAGTTTTGTAATGCATGTTACATAATAAATCAAGGAG
TCCTTATGGGCGCTTTTACAGGTAAGACAGTTCTCATCCTCGGTGGCAGTCGTGGTATCGGTGCCGCTATCGTACGT
CGTTTCGTCACCGATGGGGCCAATGTACGATTCACCTATGCGGGGTCGAAAGATGCCGCTAAACGCCTGGCACAAGA
GACTGGAGCGACAGCAGTATTC
>NCL_C39
CGCCGTTGACGTTGGTATTGCCTTGATAGTGAGCGTTGTCTGGTGCTGCTCCCGGTTTTAGTTCCCATGTCATGGAG
TCTTTGCTGAGTTGATAGCGTTCACTCCAGGTAAAATCGATCTCACCTTCAGCGGGCAGGTCATTAACGACAGGAAA
ATTCGTGGCAACAGCTTTAAAATAGCTGCTCAGTTTTTTACCTGACTTAACGATCAGGTAGTCCAGAGTGGCACAGG
TCGATTCAAAATCGTTG
>NCL_C40
GAGAATGTAATCATCCGATAATCAGTTCCCCGACCTTTTCAGGCCGGACTGATTATCAATGCGCCGAAATCGAATGC
GGACACCGCGGTGTGTTTGCACGTTTTGCACATGATGATTAAATAGGTTCTTCAGTATGAATTTATCCCGTCAGGAA
CAACATACCTTACACGTTCTCGCTAAAGGTAGACGTATTGCGCACGTCCGCGATTCTTCAGGCCGCGTCACTTCCGT
TGAATGCTACAGCC
>NCL_C41
GAAGGTATCGCGCTGGGCTTTATCTCCTACTGCGTGATGAAAATCGGTACCGGACGTCTGCGTGATCTAAGCCCGTG
CGTAATCATCGTTGCGCTGCTGTTTATCCTGAAGATTGTCTTTATCGACGCTCACTAAAATCGATAGCGCGTTAAAA
ACCCGCCATGCATGATTGCTGGCGGGTTTTTTATTTAGATCTTAACTCGCTGAATAAACTCACCAAATGCGGTCAAT
TGCCCGGTCAGGT
>NCL_C42
AAGTTGATAGGTCAGGTTGCCAAGTGTGTCGCAACGATGAGCGCGAATTAGCGCCAGGTCGGCGCGCAGTGGGCGTT
CGAGCAGCCAGGTTTTACCGTCGAGTGTCAGTGTCTGTTTGCCTTCCTCTACGACGGTGCCGACACCCGTTGGGGTG
AGAAAACCACCAAGTCCAGCTCCACCACAGCGAATTTGCTCGATTAGCGTACCTTGCGGCACCAGAACGACGTCCAT
CTCACCAGATAT
>NCL_C43
TAAACCGGGCGTGGATCTGACTGGTAACGCGGCAGCAGGATCTCCAGCTCCCCACGATTGATCTCGTTGATCACCCA
CATCAGCGGCACGGTGGCCGTTACACCATCGCCAATCGCGACCGCCAGCGGATAAGGCCACGCCTCGCGCAAGCCGT
AACCCAGCGTCATCCACAGCACCGGTAACGGTGACGCGACGAGAATCGACCAGAAAAGAGTGCGCAACGTCAGCCAG
AAGTG
>NCL_C44
ACAGCATGATGAACTTAATGCTGCGTTTGCATCCGGGAAAATAAATGGTTCTCAGTTTGAGAATTATAGCCGAAAAA
TACAGGAAACACGGCGAGAGCTTACCGGAGAGGCTCAGGCAGAGCGAGAAGCAGCAAAAGCGCATGATGAACAGGTT
GTTGCTTTGCAACGTCTGATTGCTCAACTTGATCCTGTCGGAACTGCTTTTAATCGTCTGGTAGAACAACAGAAACA
G
>NCL_C45

TATTGCGCGTAACCAGCGTTTCTGGCGTTAAGTCAGAGAACCAGGCGGTAAACAGTTTCTCGTTACAGTCGCGCAGG
CTCTGCGGCTTGTTAACGATCAGCGTCCCTTTCTCTTCGGCACGTTCCAGAATATAGGTCGCGTAGATAAACTCGGT
ATCAAACGGCGGGTCTTTACGCATCAGGATCACATCGAGATCGGCCAGCGGCAGATCCTGTTCACCGACGAACGAAA
A
>NCL_C46
CATGACCGAGGGCGCTACCTGGCTCAAGTCCGAGCTGTGAACACTGTACGATCGCACTGACAAAACTCATAGTGTCA
CAGTTTCCTAACGCCGGAACTTTACGAATTTCTGTGGTGGCGATACGGATCATACGTTCAGCCGTCATATGGCGTGG
AAGAGCTGCTGCCAGTTGCTCTTTCATTGATGGCTGGTTAATAAAACTAATCACGTCGCTATTTTTAACTGCTG
>NCL_C47
GCCCTTATTGCCCGGGGATGGGGAGTAAGCTTAGTCAGCCGTTGTCTCCGGGTGTCGCGTGCGCAGTTGCACGTCAT
TCTCAGACGAACCGATGACTGGATGGATGGCCGCCGCAGTCGTCACACTGATGATACGGATGTGCTTCTCCGTATAC
ACCATGTTATCGGAGAGCTGCCAACGTATGGTTATCGTCGGGTATGGGCGCTGCTTCGCAGACAGGCAG
>NCL_C48
ATTTCCCTGGTGGTCCAGTGGTTAAGCCTCAGTCCTTCCACAGCAGGAGGCACAGGATCAGT*CCCTGGTCAGGTAA
GCTAAGATTCTGCATACAGTGCAGCACAGCGGGGGGAAAAAAAAAAAAAAAAAAAA*AAACAACAAGT
>NCL_C49
AATCATACACCAGACCAAGTGGAATTCATAGCTGGAATTGTAAGACTGGTTCAGCATACAAAAATAAATCATTGTAA
TAGAGCACATTAAC*AAAATGGA*GAACAAAAACCCCAACC*ATCTCAGTTCAGTATAGAAAAAAATTAAACAAAAG
ACATTTAATTAAAATTACAACAACCCTTTCCTATGATAAAAAT
>NCL_C50
GTAGAATGACCTCTGAACTTCGAGCACCATTCCAGTTGTATAAGCACAAGCATCCACTGAGCTATGAATCTCTGCAG
TAACAGCTGCTGGATCCACTGCTGCTAAGGAGATTGCTGAGGCCTGTTTTGATGAGAGAATAGTGGCAAGAAAAAAT
AGTTTTTTTTAAAGTTACATTTGTAAAAGGACTTTGTCTTATTCTTTCC
>NCL_C51
ACTGCGGGCCGGTGGGCACGGAAGGCGGCGAGGCGGGAGCCAGAGGAGCGACAGCGCCCCCTGCTGGGGAAGAGCAC
GCCATTCCAGCGTTTGGGCTGGACACTTGGAGACCCCAGGGCCCCTTGTTCAGCGTCACTATGGCCAGCGGCAGAGC
TGGCCGGGGCCATGGGCCCCATCACTCACCACATGGAGTGGCCCAGGTACTCATCGTAGTAGTAGAGCAGCTCAAAG
GAGTCGATCTGGGGGGCAGCAGACGCCAGGCGGGGGGCAGAGAGGGTGGGGCTTCAGGAAGGGTGTCGGGCCGCGCA
GCCCCCGGCTACCCT*CCCCCTGCCTGCCCATGCCTCCCTTAACTGCCCCCAGTCCCTCCCTATGGGGAGGAGGGGA
TCTGGGTGGGGGCATCCTGGCTGGGGGGGTTTGGAGGGGTGGGGTTTGGGGGGGGGCGGACCTGGCTCTCTCTGGGG
CACCTGTGATGGGGTCGGCAATGTCTTTCGCCTCTAGGACAGACCGCACAGCGGAGGTGGGGGCTCACCAGCGTCTC
>NCL_C52
GGTTAATCCTGCCGTCAGATACCCCAGACCGCTCAAAAAGCA
>NCL_C53
CCTGTCCTAGATGAGGAAACTATTTGCCCAGATAAATTCATTTCTTAAAGCAAATAATTAATCTACTTATCCTCCTG
CATAAGAATTTCCCTGGGATTTCCCTGGTGGTCCAGTGGTTAAGCCTCAGTCCTTCCACAGCAGGAGGCACAGGATC
AGTCCCTGGTCAGGTAAGCTAAGATTCTGCATACAGTGCAGCACAGCGGGGGAAAAAAAAAAAAAAAAAAAACAACAA
ASAATTTACCTTGATTT
>NCL_C54
GATAAGCAAGCTCCTTCACCCACCCTCATCCTGTGAAGTAGCGCCCGTCAGTACTTGTTTCTCTTCTCCCTAGCTCC
CACCCTGGTCATACTTCTATTGTGCCATTAACTTGTTCAGGAAGGCYSGGACTTCCCTGGAAGTCTAGTGGTTTAGA
CACTGCACTCCCATGACAGCA*GAGGGCACAGGGTCCATCCCTGGTCACGGAGCTAAGATCCTGCATGCTGTGTGGC
AT**GGCCAAAAAAAGAAGAAGGAATTTCCAGCCCATCTGCATTTTTCCTACAGGCATGTGTCTGTGTCTGGACACA
TGGAGAGCCCATATGCAAGGGGCTAGGGAGAAGGAAATGGCAACCCACTCCAGTGTTCTTGCCTGGAGAATCCCAGG
GACGGGGGGTGCCTGGTGGGCTGCTGTCTATGGCACAGAGTCGGACACGACTGAAGCGACTTAGCAGCAGCAGCAGC
AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCGAGATCGT*GCCCT*AAGTGTAACAT
>NCL_C55
TATATTTTATATCTACACACATACATGTATACATAACTGAATCACTTTGCTGTACACCAGAAACTAACACAACATTG
TAATCATCATATACTTCAATTAAAAAAAAAGTAAAGGTAGGAAGAAWTTAACTTTAAAACGAAACGACGAACGTAAC
GTCCGTCCGGTAGGTTTACTTTAACTAACCTTCC
>NCL_C56
GGATCAGGTCCGCGGTATTGGTCAGCTTTCCGCGGCTGGAAAAAATACATCGGCAGCCAGAAACCGGAATAGAACAA
AAAAN
>NCL_C57
CTGGCAGTCCAGGGGTTAAAACTCTGCCTTTTCATTGCAGGGGGGCACAGGTTTGATCCCTGGTTGGGGGAACCATA
TGGCATGGCAAAAAAAAAAAAAAAAAAC*ACAACCAAACCCAAACCMAMAC
>NCL_C58

AACACCGGCGAGGGTCGCTGCCCGGGCTGCGAAGGTCAGGGCATGAAGACCATCGAGATGAGC
>NCL_C59
TWAAAAAAGGTAAAAAAAGAACAGATGAAAGTAATTTTAATTAATATATTTATTTTCC
>NCL_C60
AGTGATTGATCGATCAGGGATTAAAAGGCAAGCTTTTCTGATTCCAAACCCCAAGATTTCTCCATTAGATGATCCTC
TTCCAATTCTAACTCCAAAATGGAATATTTCTTCCAGTTCTAACACAGCATGAAGGATTTAGAAAATCACTGAATTC
AAAAATCAGGAAATTCTTGTTTTGTTGTTTTTTTTTTTTTTTTTWCTTTTTTTYCCCCGCTGTGCTGCACTGTATGC
AGAATCTTAG
>NCL_C61
CCTGCGTGCGCAACCATGCGGACCGCGCTTCGGGGCATAC
>NCL_C62
AGAGGGACTGATTTTGGTTCTATAGGCACAATGGTGAAATAACCTGACATGATGAGAATTTTTAAACAAAAAAATGA
GAAAACAAAAGCAAAAAAAANAAAGGAAAAAAGGGAAAAAGGGAAACGAA
>NCL_LRC63
ATTGCCCTTTTTAGGAAAAGCACAATCATGGCAACCCAGGATTACATAGTGTCCTACAGTTTGCGAAAGCAGAGCCC
ATTCTCAATAGGCTCTGTGAGGTAAGAAGCAAGAATCCACTTTAGTGAGGATCCCTCACTTTGAGGGGTGAAATGAT
TTGCCCAGAATCACCTGGCTGGGAAATGGCAGTGTTGGGACTAGACCCCT
>NCL_LRC64
ATTGCCCTTTTTAGGAAAAGCACAATCATGGCAACCCAGGATTACATAGTGTCCTACAGTTTGCGAAAGCAGAGCCC
ATTCTCAATAGGCTCTGTGAGGTAAGAAGCAAGAATCCACTTTAGTGAGGATCCCTCACTTTGAGGGGTGAAATGAT
TTGCCCAGAATCACCTGGCTGGGAAATGGCAGTGTTGGGACTAGACCCCT
>NCL_LRC65
ATTGCCCTTTTTAGGAAAAGCACAATCATGGCAACCCAGGATTACATAGTGTCCTACAGTTTGCGAAAGCAGAGCCC
ATTCTCAATAGGCTCTGTGAGGTAAGAAGCAAGAATCCACTTTAGTGAGGATCCCTCACTTTGAGGGGTGAAATGAT
TTGCCCAGAATCACCTGGCTGGGAAATGGCAGTGTTGGGACTAGACCCCT
>NCL_LRC66
TTTAGACTTTACCAGTTATATTGCTGGGGCCTAAGAAGAAAAAAGCAATAGTGGCTCTGTTCCTCTAAGCTCAAGAC
TACTAGTATAGATTGTCAAGGA*TTTTTTTTTAAACAATTCTGAGTTAGAAGAATCAAAATGTTACATCCGCCTCAC
AATAAAAACAGTCTGAGAGCAAATTCACATATTACACATTTACCTGTACATTACCTGATACACAGAGAAATGTTATG
GGAA
>NCL_LRC67
CGTAAATGGTTTGACAAAAGTGCACAACAGGATGACATCACTGGTTTGTTATAAATCACAGT*GAATTACTAGAATA
CAACAACGTAATTCTTCTATGTGGAGGATTGCAGGATGAGGTGAGTGTTATAAAAGATAGGCTCTTTCACCATGTTA
AGCCAGTCAGATTTTACTCCTCAGCTATGGACGTCTGAAAAGACTGGTTAGCAAAGCTAAGATTTTTCTCACTCATC
TTTTCTTCTCT
>NCL_LRC68
CCTGCGGGCTGCGACAGGCGCGTGCCCGAGCTACGGGTCGCCGGTGAAATCCCCACAGGGCCGCGCGCGCCCCGGAG
CGCGAAGGACGGCCGCGTTTCACCCGGCCCGGCCCGGCAGGCCTAGGGCCTGAACTGCCTCGTCCTCCCGCCGGCGC
CCGCCGCCCTCCCCAGGAAACCCGCAGGCCCTAGCCCACTGGAGCTCAGAAACGCCAGACGCAGTCGAAGGTTTCAG
AGGTCCACGCCTGGGAAGGCCCTCGGAGCGAGGGGGAACGCTGCGGCGCCAGCAGCCCTTGGGCTCGGGACAGGGAG
GAAGGC*GGGCCTCTGCAGACTGGGCCGAG*GGGGGAGGGTG
>NCL_LRC69
ATTGCCCTTTTTAGGAAAAGCACAATCATGGCAACCCAGGATTACATAGTGTCCTACAGTTTGCGAAAGCAGAGCCC
ATTCTCAATAGGCTCTGTGAGGTAAGAAGCAAGAATCCACTTTAGTGAGGATCCCTCACTTTGAGGGGTGAAATGAT
TTGCCCAGAATCACCTGGCTGGGAAATGGCAGTGTTGGGACTAGACCCCT
>NCL_LRC70
GGGCCGGGGCTGCAGCGCTGGTAAACAACTCCCCCCACCCTGGGGATGAGGATGAGGACCCGGAGGCAGGGAGGGAG
CTGAGGAGGGCCGGCTGAACCAGGCTGGGGGCGAGGTGGGAGTAGGGGTGCTCTTAGTGCCGCCTCCTCCTCCGAAC
TCCCCTTGGAGCTCCTTCCTCCCACCCCTTCACCCCACAGCGGAGGAAACCTCACAGATGTGGGCTCAGCCACAGGA
GATGCGGGCCGAGCTGCAGGCAGTCGTCTGTTCCAGAGCCCACTTTATCCCAGGATGAGCTGGGGCCGATGGGAGGG
GCTTTGCCCATGGGCTGACAGCTAGTG
>NCL_LRC71
TCTTATTTTTCCTCAGCCTATGGATTAGTCTATCAGTTATAKTTTAGACTTTACCAGTTATATTGCTGGGGCCTAAG
AAGAAAAAAGCAATAGTGGCTCTGTTCCTCTAAGCTCAAGACTACTAGTATAGATTGTCAAGGA*TTTTTTTTTAAA
CAATTCTGAGTTAGAAGAATCAAAATGTTACATCCGCCTCACAATAAAAACAGTCTGAGAGCAAATTCACATATTAC
ACATTTACCTGTACATTACCTGATACACAGAGAAATGTTATGGGAACACTTTCGCTTCCTTACAGCAAAAATATACT
AGAAGCTAGAGAAAAATGCAAATGAAAAAAGAACATTCACTAACCTATATCAAGTCTGTTAAAAATTTAAAAGGATA

CCATTTCAAAAGCATGACTTTAACAAAAAAAAAGCTTATTAAGTTATGTTACTCTGAAACCTCATACACAAAAATAT
AAAACCTTGCA
>NCL_LRC72
CTGGGAAGCTTGCCCCTGGGCCAGTGGCAACATCAGGGATCTGCCTCAGGATTCCACCTTAGCTCAGTTTAAGGGAT
CCTCTTTTGTTACTGGCTGCACCTCTATGGATTTCCCTGGTGGCTCAGATGGCAAAGCATCTGCCTACAGTGCGGGA
GACCCGCGTTCGATCCCTGGGTCGGGAAGATTCTCTGGAGAAGGAAATGGCAACCCACTTCAGTACTCTTGCCTGGA
AAATCCCATGGACGGAGGAGCATGGTAAGCCCGAAAGGAGTCGGACACGACTGAGC*GACTTCACTTCC*ACTTTCC
ACCTCGGAGGTTTTCTCTGCCTCACTGAGGAGAAGCGGAAACGACATTTCCCAGAACCCTCTGCTTCCGCAGCGGCT
GCAGGCGGGAGGCCATTGGTAGAGATTCCGAAGGCAGGTGCGAGGGAGTTTAGGGCACAGTTCTCCAGGGGAGTTGT
AGACTCGTGAAAA
>NCL_LRC73
TGATA*GGGGGGTCTGGGGTGATCGCGCTGGTGGACAGCAGTGACTGGCCTGTATCCAACAGCATCTGGATAAATCC
AGTGGAAACCGCGAGAGGGCTTGCTGATGGCTGCACACCCCTATGGGGAAAGCGCTGCAATAGCGACCAAGGAAACC
AGTTTATGGGCCATGAGACTCAGGAGCTGGCCTGTAAAAGATGACAGACACTGGCATTTCCACCAGCCCTACA*CCG
TCCTCCCCTGCTGCCGGGCTAAGCGGAGGATGGATGGACTACTTACTTAGCCAGCAAGTGAGACAGGAAATCTGCTC
TCTGACATGTGAACTAGGAGGCTAGCTTGAGCCTTCCACATTATAACTGAATATGCCAGGAGCAATGGACCCAGAGA
CTACGCCATGG
>NCL_LRC74
CCTCCAAAGAGAGCTGGGACCCCCAAACTCAGGATCCTGAGTGGCACAGATGGCCAGGGGGCAGTGGCAGCACCTGC
CTTCAGCCAGACAGGGCCGAGCAGGCCCTTGGGGTCAGCCCAGGGCTCCTGCCAGGCCCCAGACCAGCCACTTCCCT
GCTCCAGGGTCCCTCCTCCTCCTGCCTCCCAAATTCACACCCCAAGGAAGCAGCAGCAGCCGCCAGTGGGGCTTGGC
TCCACACTGCCTTTATTCACACTGCTCACCCCTGATGCTGCCAGCTGACAGGGACTGACCGGGGCTGGGGGACCCCT
ACATGAAAGGAGGATTTGAGGCAGAGTCCCTGGCAGCACAGGAGAGGAGAGGAGGGCAGGGAAGGCGAGGTAGCCCC
CCTGAGCCCCTCACAAGCATGAAGGCCAGCAATGCCAGGAAGCGAGGCTGTGTGCCTGGCCCTGGTGCTGGGGT*GG
GGGGCATCTTGCCAGAACAAGGCCCCCTTTCCTGGGAGTAGGGCCGCTGGTGGCAGGTGTGGATTCAGGCTGAGACA
CGCTGACGAGAGCCCCACGATGTCTGCCTCTAGTGAGTGCAAGTACTCAGAAACCTTGAGCAGGTG
>NCL_LRC75
AGGTTTTTTTAGTTCCTCTTCACTTTCTGTATCCT*CCAAAAGTAAAC*ATTTTGGGTTTCAGGGGAAGTTTATTGT
ATTACTACAAACTCATGGGTTTAAACATATTTCACATGTTTCTGTTCATTCTTATTAATGGAAAATTTTCAATCTAG
CACATTGTGAAAGATCAAAAATTCTGAAAACATACAGAGTTGGTGAAGCTGTGAAAAATGGTACAACATCTGCGGAC
AGCAATTTAACAGTATTTAACAAATCATAAAAGGCACATTTCCTTTGACACAATGATTTTACTTTCAGGAACTTATC
CTACAAATACACTCGCAAGTATGCAAAGTGATGTATGTACCAAGTCTAAATAAACAAAAGCAGTCTTG*TTTATTTT
*AGCAAAAACAAGGTAAAGAGATACTTATGCAGAAAAATTAATAGCAGGATT
>NCL_LRC76
GTGCATCACCTGACCGGGCAGATAGTTCACCGGGGTGAGAAAAAAGAGCAACAACTGATTTAGGCAATTTGGCGGTG
TTGATACAGCGGGTAATAATCTTACGTGAAATATTTTCCGCATCAGCCAGCGCAGAAATATTTCCAGCAAATTCATT
CTGCAATCGGCTTGCATAACGCTGACCACGTTCATAAGCACTTGTTGGGCGATAATCGTTACCCAATCTGGATAATG
CAGCCATCTGCTCATCATCCAGCTCGCCAACCAGAACACGATAATCACTTTCGGTAAGTGCAGCAGCTTTACGACGG
CGACTCCCATCGGCAATTTCTATGACACCAGATACTCTTCGACCGAACGCCGGTGTCTGTTGACCAGTCAGTAGAAA
AGAAG
>NCL_LRC77
GACTTAAATGAGATGGTACATGTAAACAGCTGGAGGAGTGCCTGGAATGTAGAAAGTGTTCGGTAAAGCCTAGCTAG
CATCATCCATCGCCATGTTTCATGCCTTTCAGCCACGTTTGCATCCTCTGCTCCATCCGATCTCCCCGTCACCTATA
GACTTATAGGTGAGAGCCCAGTGTGATTGTATCTGTTGAGTA*TTTTCCTTCTTGAATCCTAGAGAAGCAACTGTC
>NCL_LRC78
AGAGAAGAAAAGATGAGTGNGAAAAATCTTAGCTTTGCTAACCAGTCTTTTCAGACGTCCATAGCTGAGGAGTAAAA
TCTGACTGGCTTAACATGGTGAAAGAGCCTATCTTTTATAACACTCACCTCATCCTGCAATCCTCCACATAGAAGAA
TTACGTTGTTGTATTCTAGTAATTCACTGTGATTTATAACAAACCAGTGATGTCATCCTGTTGTGCACTTTTGTCAA
ACCATTTACGTGATTTTAATAAACATAGTAAACTTGCTGACTGCACCGGAGGTCTGTTAGTGATTTA
>NCL_LRC79
AAGGAATGCTTCTCCCTGTATGACAAGCAGCAGCGCGGGAAGATTAAAGCCACGGACCTGCTGATGGTGATGCGGTG
CCTGGGGGCCAGCCCAACGCCTGGGGAGGTGCAGCGGCACCTGCAGACTCACAGGATAGGTGAGTGGGCCCGGCTGG
GCCAGCTGCCACGGAAGGAGGAGTCTGGGTGAGTGGGTGGGCCCTGAGCCCTATAGCTCTTGTCCCAGGGACGCAGG
CAGGCAGGCAGGTCGGGTGTGACCAGCTCAGAGGTTTGAGAGGTCCGTGTTAG
>NCL_LRC80
AGCTCCTGCCAGCTGCTGACCGCGGCGCCGCGCCCTCGCCGGGCGAAGTCTAGCGACTTCAGGCCAGGTGTCTCTCC
TTCCCCAGGGCTCGCCGCCCACATCGGGGCAGACAAGCGACTCCCAACGGTACGGAGACCTTAGCGTCGGTCAAGCT

CTCGCCCCTGGACCCAGGTGAGTGGGTGACCCTGGGCAGCCAGGCAGGGGAAGGGATGGCCACAGTTATAGCCCCGC
CCCGTCTCACAGCCGCGCCCAGAACCCAGCCGGGGTCGGAAGGGGGCAGCAGGGCCAGGGACACGGAGGCCGCGGAA
GGCAGGATGGCGCAGGAACCAGGGCTTCTAGCCCCTGGCGCGTGCTCCATTGTTCTAGGAGATTTCCT
>NCL_LRC81
TGTAGATAAAGCTATAGAGATGCCCACACAGCAAGTAAGAAAGAGATTGAGACTCCTGTCTCCTGGGAGATTTATTT
CTGCAGGTTGGTTTTATCCAGCACAACATTTTGTAATTAAAAGAAATAAACATTTAAAGTTTTG*ATTGAAATTT*A
TGATCGTTACAGGTGTTCTCCATCAGGAGATTCTAGAGGTACAGTTTGGCAGAGCAGGGATTGGAAGTGTTGTGATC
TTGTATAGGTGTTATCACACTAACCCCGACAAGGCTGGACTCACTGGGGTAGACCAGTAGGAAGAACATGGGAAGTC
CAGACAGAGGAATGGCCCTGGCCTGTCTCAGGGAAGGTAGGAAAGGAAGCACTGAAACTGCGTTTAGAACTCACTTA
TTTTCTCCTTTGTTGCAGAAGGTATAGTTCTGACC
>NCL_LRC82
GCTAAGTCGCTTCAGTCATGTCCGACTCTGTGCGACCCCAGAGATGGCAGCCCACCAGGCTCTCCTGTCCCTGGGAT
TCTCCAGGCAAGAACACTGGAGTGGGTTGCCATTTCCTTTTCCAAGGAAGAACCCTACTTCATCTCAAAAAGACAAG
AAGAGGAAGAAGAAAGTTTCTGATTCAGCAATGAATATATAGTTGGATGAATCATGTCTCCTTCAGAAAATGAAACC
AAAGGGAAGAATCTAAATCACCTTCAAAAAGTT*GAGTTTGAAAAGAATGATGCAGGGAAAAGGGCTACACAGAGAG
GGAAGAAGGGAGGGAGACAGAGAGAGGGGGTGCGAATAGGAGGA
>NCL_LRC83
CTCTCTTTGAACCCTCTCTCCCCTCTGGAGGTGAAGTGGCCTGCAGAGCCACAGCTCCCCAGCCCAGTGGAACCAGC
CCCCACCTGCATCTGCCCTGACCCCCCAGCCCTGGAGGCACAGGGAGGCCAGCCAGGAGCTTGGGAAGATAAGGGGT
CCTCACAGAGGCCCTGACCACAGGGTGGGGGTGGAACTGGCCAAATGCCCCTCTCGTGCAGCCCCCGGGACCCATGG
GGTAGAGCTCCAGCTTGAGCCCCACTGGGAGCATGCCGGGCCCTGCACTGTTCCCTAGCACCCTGCACACTTCCCAC
AGCCAGCTGGTTACTCAGAGGCCTGGACCAGCAGCTTCTGAGCGGATAGTAACAAAACCAGCAAAAGCAGCTCCCGG
TACTAGCGCTTTCCCCCAGGCCCTTGTGTCATCCTCACTGTGGCCCTGTGGGGAGGGACCCTAGTTGTTCCCATGGT
GCAGACGAGCCAGCGAGGAAGGCTGGAGCAGCCGAGGGCTCCTGGCAGTGGGGGCAGTGGGCAGCTAGCTCTGCCCA
GGCTACCCTGGAGCTCCACCAGAGGGCGGGAGAGGTGGCAGAGGACGCCCAGGCCGAGGGCCCCTTCATGCTACCTC
ATTAGGGCCACTCCCAGCAGCCTAGGGTTTCAGGGGCATTGCTAATCAAAGGCAGGCCAGGCGGTGGGGGTTTAGAG
ACAATTACCTGCTTTTATGCCTGTGGAGAGGTTGAAGGGGAACCTCCTGCCCTTTCAGAACTTTCCAGAGCCTGGGC
TTTGGAGGGGCAGACCACCAACAGGACCCAGCGGCAGGAACCAGAGTGGCCCAGGGAAGCAAGGCTTCACTGGTTCC
TGTGAAACCCAATTCCCTCTAAGAGGGCCCAGCGCCAGTGGGGACAGGGACA
>NCL_LRC84
GCCTAATGAAAAAGGCCTCCTCTCCTGGGGCCTTCTCCCCCACCGCCAGGGGCTGCCCCTGGCCAGCAGTGAGGGTT
CATGAAGGTGAATCCACATGCCCACTGAGGACAGGATTAGAAGGAATGGGTGACGACCACAGCTGGAGGGAAAGAGG
TGAGAAGTGAGAAGAAGCTTAGCCCATTCTTACCCATGGTGTGTCAATGCGGTGACAGCTACAGGGGGCCATTCACC
GAACAATGTGCCTTTACAAGTATAGAGTTCTCATCAGTAACCTCGAGAGCTGGGGCCATTTATGCCCCCTTTATAAG
TAAGAAAACAAAGGCTCCAAGATGCCATGTAAATTCCCAAGGTGACTCAGCCACTCAGTTGCAGAGCTATGCAGTTG
CTTGAGTGGGATGGGTCAGCATGGAGGCAGGGGAGTGGATGAGATAACCTCTGGG
>NCL_LRC85
GCCGGACCCCTCCTTCCTGCGGGCTGCGACAGGCGCGTGCCCGAGCTACGGGTCGCCGGTGAAATCCCCACAGGGCC
GCGCGCGCCCCGGAGCGCGAAGGACGGCCGCGTTTCACCCGGCCCGGCCCGGCAGGCCTAGGGCCTGAACTGCCTCG
TCCTCCCGCCGGCGCCCGCCGCCCTCCCCAGGAAACCCGCAGGCCCTAGCCCACTGGAGCTCAGAAACGCCAGACGC
AGTCGAAGGTTTCAGAGGTCCACGCCTGGGAAGGCCCTCGGAGCGAGGGGGAACGCTGCGGCGCCAGCAGCCCTTGG
GCTCGGGACAGGGAGGAAGGCGGGGCCTCTGCAGACTGGGCCGA*GAGGGGGAGGGTGCGACCAGGGGACGGTTAAG
TGGGGTGTCTCTTCCAAAGCCTA
>NCL_LRC86
CAGAGACTACGCCATGGTAACCCCAAAGCAACTAGTCAGCACCATCCAAGTCTGACTGTTTTTGTCCTGACCACAGG
ATCTGCTCCCAGTGGAACACCGTATACCACGGCCTGGGACTGGCAGGTAAGTTTCTGGTGGTGTGGGTTTGCCACTT
CTTGGGGGGATAGGATTAGAAAGGGTCTTACAGTCTTCACCTGTCCTCCACTCAGCCCTACCTGAGACCGCTACTGA
GGTCACTAATCCAGGCCTCCCCTGGAGAAAGTCACCATTCTATGACCCCTACGATCTGTTAGACATCCCCTGCTCCA
GGTTTCGGCGCCAGCTGTGGGGACTGCAGGCT
>NCL_LRC87
GTACTCTTGCCTGGAAAATCCCATGGACGGAGGAGCATGGTAAGCCCGAAAGGAGTCGGACACGACTGAGC*GACTT
CACTTCCACTTTCCACCTCGGAGGTTTTCTCTGCCTCACTGAGGAGAAGCGGAAACGACATTTCCCAGAACCCTCTG
>NCL_LRC88
CTTCCCCAGCAGGGGGCGCTGTCGCTCCTCTGGCTCCCGCCTCGCCGCCTTCCGTGCCCACCGGCCCGCAGTGTCCG
CGGCGGCCACTAGAGGGCAGCGAGCGCCTGTCCTCCCAGGGCCGGGGCTGCAGCGCTGGTAAACAACTCCCCCCACC
CTGGGGATGAGGATGAGGACCCGGAGGCAGGGAGGGAGCTGAGGAGGGCCGGCTGAACCAGGCTGGGGGCGAGGTGG
GAGTAGGGGTGCTCTTAGTGCCGCCTCCTCCTCCGAACTCCCCTTGGAGCTCCTTCCTCCCACCCCTTCACCCCACA

GCGGAGGAAACCTCACAGATGTGGGCTCAGCCACAGGAGATGCGGGCCGAGCTGCAGGCAGTCGTCTGTTCCAGAGC
CCACTTTATCCCAGGATGAGCTGGGGCCGATGGGAGGGGCTTTGCCCATGGGCTGACAGCTAGTGCCAGAGCCACCC
CTGGACATGGTGGCCCCACTCCCGAGTGAGGTGGGCTCACTCGTGTCACGCGGGCCCTACATTGTCCTTCTGGGTTC
TCTTTTGGGCCTAGAAAGTC
>NCL_LRC89
TAATTGCCCTTTTTAGGAAAAGCACAATCATGGCAACCCAGGATTACATAGTGTCCTACAGTTTGCGAAAGCAGAGC
CCATTCTCAATAGGCTCTGTGAGGTAAGAAGCAAGAATCCACTTTAGTGAGGATCCCTCACTTTGAGGGGTGAAATG
ATTT**GCCCAGAATCACCTGGCTGGGAAATGGCAGTGTTGGGACTAGACCCCTGGCCCATCTCTAAACCACAGGCT
ATAAAGGGCTAGACTAATGGTAGTTGTTAAAACAGTACAAAGTTTTATTAAAGACATGCTAATACCGTGGCTGGAAT
AATTTTGTTTCCTTGTCCTCTTGCTGGGCTTAGGTTTGTCATCTTGCTGTGTTTGCCATCTCCCGCAAGTGTAAGAA
ACACTTTCTTTTCCAAAACCCACAGT
>NCL_LRC90
CGTGTCCAACTCTCTGCGACGCCATGAATCGCAGCACGCCAGGACTCCCTGTCCATCACCATCTCCTGGAGTTCACT
CAAACTCATGTCCATTGAGTCAGTGATACCATCCAGCCATCTCATCCTCTGTCGTCCCCTTCTCCTCCTG*CCCCCA
ATCCCTCCCAGCATCAGAGTCTTTTCCAATGAGTCAATTCTTCGCATGAAGTGGCCCAAGTACTGGAGTTTCAGCTT
TAGCATCATTCCTTCCAAAGAACATCCAGGACTGATCTCCTTTAGGATGGACTGGTTGGATCTCCTTGCAGTCCAAG
GGACTCTCAAGAGTCTTCTCCAACACCACAGTTCAAAAGCATCAATTCTTTGGTGCTCAGCTTTCT
>NCL_LRC91
GGTGACCCTGGGCAGCCAGGCAGGGGAAGGGATGGCCACAGTTATAGCCCCGCCCCGTCTCACAGCCGCGCCCAGAA
CCCAGCCGGGGTCGGAAGGGGGCAGCAGGGCCAGGGACACGGAGGCCGCGGAAGGCAGGATGGCGCAGGAACCAGGG
CTTCTAGCCCCTGGCGCGTGCTCCATTGTTCTAGGAGATTTCCTTCTGAAATGTAAATTCTAAGATACAGTTAGAGG
AGGGCATCGCACGCAAGCCAGTCCAGTATTC
>NCL_LRC92
CTGACTAGTTGCTTTGGGGTTACCATGGCGTAGTCTCTGGGTCCATTGCTCCTGGCATATTCAGTTATAATGTGGAA
GGCTCAAGCTAGCCTCCTAGTTCACATGTCAGAGAGCAGATTTCCTGTCTCACTTGCTGGCTAAGTAAGTAGTCCAT
CCATCCTCCGCTTAGCCCGGCAGCAGGGGAGGAGGGCTGTAGGGCTGGTGGAAATGCCAGTGTCTGTCATCTTTTAC
AGGCCAGCTCCTGAGTCTCATGGCCCATAAACTGGTTTCCTTGGTCGCTATTGCAGCGCTTTCCCCATAGGGGTGTG
CAGCCATCAGCAAGCCCTCTCGCGGTTTCCACTGGATTTATCCAGATGCTGTT
>NCL_LRC93
CGCCTCCTTCATGGTGTTGCTGAGCTAAGCACATGACACATCTAAGTAAAGGGACACTTCCATCAAAACAACCTTGG
GATCGTTCAGCACCGTCTGGACTGCAGGTTTGAGACAACTGTGGTGAAATGTCTAGAGCCCCTGGTCCCCTGGGATG
GGGCGTCACCCGTTGCTACGAGCACACCTCGCCAGGTGCCTCCACACCCAGCCAGAGAACTCTGGCTAACACGGACC
TCTCAAACCTCTGAGCTGGTCACACCCGACCTGCCTGCCTGCCTGCGTCCCTGGGACAAGAGCTATAGGGCTCAGGG
CCCACCCACTCACCCA
>NCL_LRC94
GGCAGGGAGAGCAAGGAGGTGTAGAGGTCATGGAGGCGGGGCAGGAGTAGAGAGAGAAGAGCCCTACTGTTTCAGGG
CTGGCCTCCAAGTCCAGAGGACCCCCCAGATCTTTAAAGCAGGAAGACACATTTTAGGTAACTGGGGAGGAGATGGG
TCCCCCATTGACTCTAGCCCTGCTGATACCACGGGGCCCCTGGTGCAGAGTATGGGGGAGGCAGTTTGCCCCCTGTG
GAGTCTGGTCCATCCCCGGTCCCAAGAGTGTGCTGTCCCCATCCGGCTACTGTTGGAAGGTG
>NCL_LRC95
CACAGATGGCCAGGGGGCAGTGGCAGCACCTGCCTTCAGCCAGACAGGGCCGAGCAGGCCCTTGGGGTCAGCCCAGG
GCTCCTGCCAGGCCCCAGACCAGCCACTTCCCTGCTCCAGGGTCCCTCCTCCTCCTGCCTCCCAAATTCACACCCCA
AGGAAGCAGCAGCAGCCGCCAGT*GGGCTTGGCTCCACACTGCCTTTATTCACACTGCTCACCCCTGATGCTGCCAG
>NCL_LRC96
TTTTTCTGCATAAGTATCTCTTTACCTTTGTTTTGCTAAAAATAAACAAGACTGTCTTTTG*TTTATTTAGACTTGG
TACATACATCACTTTGCATACTTGCGAGTGTATTTGTAGGATAAGTTCCTGAAAGTAAAATCATTGTGTCAAAGGAA
ATGTGCCTTTTATGATTTGTTAAATACTGTTAAATTGCTGTCCGCAGATGTTGTACCATTTTTCACAGCTTCACCAA
CTCTGTATGTTTTCAGAATTTTTGATCTTTCACAATGTGCTAGATTGAAAATTTTCCATTAATAAGAATGAACAGAA
ACATGTGAAATATGTTTAAACCCATGAGTTTGTAGTAATACAATAAACTTCCCCTGAAACCCAAAA
>NCL_LRC97
GCTCAGAGCTGCCATTTGGCAGTTATGCCAAACACGGGAGTGCTTATTCTGTTTTGACCTGTGAATTACTCTCATTC
AGAGATCAGGTTGATTGTATCTTGAGCATTTAGTAAGGCATTCATTTTTCCTTTCCCAGTACCTGGTAAAGTGGTTC
TTTTCTGGATTCGGGACTACCCTAAGAATTTGGTAAAAGCATTCTCTGGAAAGGAGAAGGCAATGGCACCCCACTCC
AGTACTCTTGCCTGGAAAATCCCATGGATGGAGGAGCCTGGTAGGCTGCAGTCCATGAGGTCGCTAAGAGTTGGACA
CGACTGAGTGACTTCACTGTCACTTTTCACTTTGATGCATTGGAGAAGGAAATGGCAACCCACTCCAA
>NCL_LRC98
TTTAGACTTTACCAGTTATATTGCTGGGGCCTAAGAAG*AAAAAAGCAATAGTGGCTCTGTTCCTCTAAGCTCAAGA

CTACTAGTATAGATTGTCAAGGATTTTTTTTTTAAACAATTCT
>NCL_LRC99
ACTGCTCACCCCTGATGCTGCCAGCTGACAGGGACTGACCGGGGCTGGGGGACCCCTACATGAAGGAGGATTTGAGG
CAGAGTCCCTGGCAGCACAGGAGAGGAGAGGAGGGCAGGGAAGGCGAGGTAGCCCCCCTGAGCCCCTCACAAGCATG
AAGGCCAGCAATGCCAGGAAGCGAGGCTGTGTGCCTGGCCCTGGTGCTGGGGTGGGGGGCATCTTGCCAGAACAAGG
CCCCCTTTCCTGGGAGTAGGGCCGCTGGTGGCAGGTGTGGATTCAGGCTGAGACACGCTGACGAGAGCCCCACGATG
TCTGCCTCTAGTGAGTGCAAGTACTCAGAAACCTTGAGCAGGTG
>NCL_LRC100
TCGATTACTACGAAATTAAAAATGATTCTACATGTCAAACTCAAATAGAAAATGCCATGCAATATATAAATCACTAA
CAGACCTCCGGTGCAGTCAGCAAGTTTACTATGTTTATTAAAATCACGTAAATGGTTTGACAAAAGTGCACAACAGG
ATGACATCACTGGTTTGTTATAAATCACAGTGAATTACTAGAATACAACAACGTAATTCTTCTATGTGGAGGATTGC
AGGATGAGGTGAGTGTTATAAAAGATAGGCTCTTTCACCATGTTAAGCCAGTCAGATTTTACTCCTCAGCTATGGAC
GTCTGAAAAGACTGGTTAGCAAAGCTAAGATTTTTCNCACTCATCTTTTCTTCTCTACCAATTACTAAGCAACAA
>NCL_LRC101
AGTATATTTTTGCTGTAAGGAAGCGAAAGTGTTCCCATAACATTTCTCTGTGTATCAGGTAATGTACAGGTAAATGT
GTAATATGTGAATTTGCTCTCAGACTGTTTTTATTGTGAGGCGGATGTAACATTTTGATTCTTCTAACTCAGAATTG
TTTAAAAAAAAAATMCTTGACAAT
>NCL_LRC102
ACCGAACACTTTCTACATTCCAGGCACTCCTCCAGCTGTTTACATGTACCATCTCATTTAAGTCTCTTAATCCTCAC
AGAAACCCTCTGAGGTCACACTATTACCCTCATTTTACAAGTGAAACAACTGAGCCCAAGA
>NCL_LRC103
GTACTCTTGCCTGGAAAATCCCATGGACGGAGGAGCATGGTAAGCCCGAAAGGAGTCGGACACGACTGAGCGGACTT
CACTT
>NCL_LRC104
ACTCCCAACGGTACGGAGACCTTAGCGTCGGTCAAGCTCTCGCCCCTGGACCCAGGTGAGTGGGTGACCCTGGGCAG
CCAGGCAGGGGGAAGGGATGGCCACAGTTATAGCCCCGCCCCGTCTCACAGCCGCGCCCAGAACCCAGCCGGGGTCG
GAAGGGGGCAGCAGGGCCAGGGACACGGAGGCCGCGGAAGGCAGGATGGCGCAGGAACCAGGGCTTCTAGCCCCTGG
CGCGTGCTCCATTGTTCTAGGAGATTTCCTTCTG
>NCL_LRC105
AGTGCAGCAGCTTTACGACGGCGACTCCCATCGGCAATTTCTATGACACCAGATACTCTTCGACCGAACGCCGGTGT
CTGTTGACCAGTCAGTAGAAAAGAAGGGATGAGATCATCCAGTGCGTCCTCAGTAAGCAGCTCCTGGTCACGTTCAT
TACCTGACCATACCCGAGAGGTCTTCTCAACACTATCACCCCGGAGCACTTCAAGAGTAAACTTCACATCCCGACCA
CATACAGG
>NCL_LRC106
TACTTATTTAGCTTATATGCAGAGTACATCATAAGAAACGCTGGGCTGGAGGAAGCACAAGCTGCAATTAAGATTGC
CGGGGAAATATCAATAACCTCAGATATGCAGATGACACCACCCTTATGGCAGAAAGTGAAGAAAAACTAAAGAGCCT
CTTGATGAAAGTGGAAGAGGAGAGTGAAAAAGTTGGCTTAAAGCTCAACATTCAGAAAACGAAGATCATGGCATCTG
GT
>NCL_LRC107
TTTGAGAAGGGGGGACCGAATGCTGGGGAGAGTAAAACTGCTGATTGAGATGCTGACAAAGTTGACTTTGAAGACAG
TGGGATAGATGATGAATTTTGCTAATGGTAACAGATTTTTTTTTAAAACTATTCTATAGCAGTACATTGTTCACTAT
ATTGCTAAGGATCTAATTCTGTCTTGTAAAACAAATCCAAATATGTATTGTGAAACACCTGTATAAAACCATTCTTG
AAACAAGTGATCTACATGCCTT
>NCL_LRC108
CACTGGGCTCTCACCTATAAGTCTATAGGTGACGGGGAGATCGGATGGAGCAGAGGATGCAAACGTGGCTGAAAGGC
ATGAAACATGGCGATGGATGATGCTAGCTAGGCTTTACCGAACACTTTCTACATTCCAGGCACTCCTCCAGCTGTTT
ACATGTACCATCTCATTTAAGTCTCTTAATCCTCACAGAAACCCTCTGAGGTCACACTATTACCCTCATTTTACAAG
TGAA
>NCL_LRC109
GGCAGGCCAGGCGGTGGGGGTTTAGAGACAATTACCTGCTTTTATGCCTGTGGAGAGGTTGAAGGGGAACCTCCTGC
CCTTTCAGAACTTTCCAGAGCCTGGGCTTTGGAGGGGCAGACCACCAACAGGACCCAGCGGCAGGACA
>NCL_LRC110
CCTCCTGCCCTTTCAGAACTTTCCAGAGCCTGGGCTTTGGAGGGGCAGACCACCAACAGGACCCAGCGGC
>NCL_S111
TCGTGGTCCCAGACTAATAATCAGACCGACGATACGAGTGGGACCGTGGTCCCAGACTAATAATCAGACCGACGATA
CGAGTGGGACCGTGGTTCCAGACTAATAATCAGACCGACGATACGAGTGGGACCGT
>NCL_LRC112

GAATTGATGCTTTTGAACTGTGGTGTTGGAGAAGACTCTTGAGAGTGCCTTGGACTGCAAGGAGATCCAACCAGTCC
ATTCTGAAGGAGATCAGCCCCGGGATTTCTTTGGAAGGAATGATGCTAAAGCTGAAACTCCAGTACTTGGGCCACCT
CATGCGAAGAGTTGACTCATTGGAAAAGACTCTGATGCTGGGAGGGACTAGGGGGCAGGAGAAAAGGGGATGACAGA
GGATGAGATGGCTGGATGGCATCACCAACTCCATGGACATGAGTCTGACTGAACTCCAGGAGATGGTGATGGACA
>NCL_C113
CGTGTGCTTGGTCACTTAGTCGTATCTGACGGTTGAGACCCCATAGACCATAGCCTGCCAGGCTCCTCTGTCCATGG
GATTCTCCAGTCAAGAATACTGGAGTGGGTTGCCATTTCCTTCTCCAGCGGATCATCCCCACCTAGGAATCGAACCT
GGGTCTCCTGCATTGTAGGCAGATTCTTTACTGACTGAGCTGTGAGGGACCCCCTAAAATTTGTAATTTTAGACCAG
AAAGATCAGTTG
>NCL_S114
GTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATAATCAGACTGGGACCACGGTCCC
ACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCAT

Appendix 5

Sequence used to design the sequence capture probe.

This 73,072 bp sequence comprised of *CLN6* and flanking genes *Fem1b* and *CALML4*.

GTTGTTAAAACAGTACAAAGTTTTATTAAAGACATGCTAATACCGTGGCTGGAATAATTTTGTTTCCTTGTCCTCTT
GCTGGGCTTAGGTTTGTCATCTTGCTGTGTTTGCCATCTCCCGCAAGTGTAAGAAACACTTTCTTTTCCAAAACCCA
CAGTCTGAGTGCTCAGGCCCAGGGGCGGCTCTCTCGCCTCCTCGTTCAATAGTCCCGCACAGGAATGGTGAGCTTCT
GGATGAACTCGTCATACTTCACTTTGCCATTCGGTTCGATATCTGCTTCCCTGAAGAGATCCTCCACTGCGATAAAT
CACAGTAATTTCTCAGTTTGGCTTCAGTGTCATATGATAAGGCATCCAGCTGGGGGTTGAGCCTCCCCAGAGGTCTT
CCCTGAACCTCCAGTTGACTGCTTCCATAGGTGAGGAAGTAACCAGTGAACAGAAAAGGACAACCCGCTCCCTCCCC
CTTGGAGACCATGGAATCAATTCTGAGTTCATCTGGCTTTCTGCAGGGGCCAGGATAGGGGGTGTACTGGTAAGGAT
TCAGTCAATAAAACCACCTTGTACCTTTTTTTTTCTTCAGAAAATATCTGGCTCCTGTGCCAGTATTTCCAGAAAGT
GTTGGCATGGATTTTACCTTCTACCGCCCGCCCTCTCAGACGCCCTCCATAACAGACACCCCTGCAGCTCCATCCTG
GGGCCCCTAGCCCAGCCTATGGCCTGCCTTCCAAAATGACATCCCCCCACCCCAAGAGGGGGCTTGGTGACAAATGG
TATCTGCCACTATCATCTCTTTCTGAAAGTGTTTCGGAGGTGGGATAGTATAAGAGAAAATAAATTCAACTGTGTCC
AAGCTCAAAGCCTACAAAAACTGCCATTGTTCTCTAACGTTGCTCAGAAAAAAAGAAACGTACATTTACTGGTTCTG
GATTCTTGGGATAAAATACAAGCTTGTCCCAGGAGTTGACAAGAAGAGCCAGTTCAGTTTACACTTTTCAATGGTGG
CTTCTCAGGACCCTCCTCACCCTCTTACAAAACTCTCTCACTGCTCTGTGTCAACTTTGCCAGCCGCAGACAGCCCC
CACACAACTCTGGACAGTGAAAAAGCATCCTGTGACTTCCTGCTTAAGCATCCTCCTCAAAATCGTAATTGCTATGC
ACTGCGTTTGAGCTGATAGTACTTCTGAGTACTTGTAGAACTCTGAGACTGTAATTTATCTTATTTCATGGAGGAGT
GTCTACCCATCAACCACTTGGTTAAAACTGTCCCAATGGACAAGTCCAGTACCGGTCACTGCAGCCTCCAGCTGTCC
AGAAGAGGGGCAGCTGGAATCTGATGGTACCTTCTGGTTAGAATAGCAGTTCCCTGAAAAACCACAGAGAACCGGTT
TGCAAAGGAGGCCTCCTCACAGCCGTCTCAGCTGCTCCAAGGATGCTTCTGCAGTAAACCAACTTTTCTATCACTGA
AGGTAAAGGAGTCTTACTTCCAGTATTACCTTTTCTATCTCCCATAGCCAGAGGAAAAGGACATCAGTCTCTGGCAG
ACACTGCCAGTTTCTTTGAAATGTTTCCTTACTGGTTAAACAAATGAGAGAGATGTTTAAGTTTGGGGACCATGGTA
TTGTTTTAGACGGATCATTAACCGCAAAAATGAAGGCCCCTGCTGACCTCTGCTAACACTCCACCTGATCCTTAAAA
GGCTGAGTAGTGTCCTTGGAGCCCGGGAGGGAAGCATAGCCTGTGACTCCCTGTTCACGTTAAAACAGCCACAATGA
AAGATGGGATTGTGATAGCTGGCAGCCTGGAAGAGCGGGGGAAGCTCCATCTGCCTTCTCTAGCGCTATTTTTACTT
CCTTTGCACTGTGCCTTTGTGACTGCAGATAAAATCTTAGCAGACTGCTTTATAGAATGTTAGCTCTTGCAGGGTCC
TGAGGGATTGCCTATTGTAACCTGATTTCCCAACGAGGAAACTGAGGAGGGTGAGAGAGGTTAAGAAACTCACCTGA
GCTCCCCCAGAAGGAGGGGGAGCCAGGATTCAAGTCTGTTGTGATTTCACCAAACGCTAAAGGCCCATCATTACCGG
AGCCTTTTTCGTCCACCGACCGTGTCCTCCTGCCCCAATCACAGGTGAGCCCGGCGGCCTACCCAGAAATTGCAGCC
TCTGGGGTTGGGGAGTGTGGACTCGGGTGGCAACCTGGCATGGCCACTGTGGCTGGACCCAGGGGCCTGTCAGGCCG
GCTGGGAGGACATCTGCAAAGGGCTCAGCATTGGGGTCACCCTCACCATGAACATCCAGCTTTCCCAGAAGGCCCAA
GCCCCTCCTCTGGCTCTGAGGCAAGAAGACAGGGAAGGTAGACAGGACATTCCTTCTGGAAGAGCTCAGAGCTGGCT
GACTCATGAGAAAAGGGAGCAAAGTAACATCTCAATTTAGAAACTTGGTGGCAGCAGCAACTTCAGAAGACAGCTGT
GGTGAGAGGAGTGCCTGAGCGAGGTCGGAGGGTGGGGGTCCCGGTGTCTGCTAGCCCTGTGTCTAGCGCCTAGCTCT
GCTCGTCACAGGCCGCGCACTGCACCTGTTTCTTTAATGGGTATGAGAGGGACACTGGTGATCCTAGAGGTGTCGCA
GTCATGAAGGCCCCCAGTGAAAGTTAGAAGACATAGAGTAATGCCCAAGTAGCTCCTAACTGGGTTTCCCTACTGTC
CCTTCACCCTGGCTAACTTCTGCCACTCACAGTTCCCTAGTTTTCCAGGTTCCAGTCCAGAGTTCACCAAGCTTAGC
ATCTAGTCCCAACGTCCAAATGCCATCAGATGCAATGCCTACGTAACTGAGGTGGCATGATTTGGACTTTGGTGTGC
CCCCAACCCCTTAGTCACCTCCCACTGTTGGGAAATAGGGACCCAGACCATTACCTTCTTTGTGGGTGAGCTTCTCC
CCCAGTTGCATCAGTTTTGACCGCAATTCAGATGCCATGATGTAGCCTTTCTTCTCCTTGTCCGCCATCAACATGGC
CAAAAGAATTTCCTTTTTTGGGTCCTCTTGTTTTATCTGCATATGCATGATGGTCAGGAAAGTAAAGAAATCCAGCT
CTCCGTTTCTGTCTAGGAAACCCCCAAACAGCAGAGTGAGCAAGGAGGAAAAGGCTTCTCTTGGGGATGTCAGCTGG
CCTCCTGCCCAAGCCTTCAAGGCCAGCTCAAGAGCCACCCAACTCAGGAAGCACCGTGGTGGTGGCTGTGTTCATGG
GGAAGAGGCTGCAGCCTGTCTTGCTGCGTTCCCACCATGGATATGCCTCATTTAATCTTTTCAACAACCCTCAGTGC
CTGGACATCTCATTTTGCCCACTAAAGAGACAGAAGCTCAGAGCAGTGACAAGAACAGGATGAAGCACCGCTTTCTG
TACCCCTTCTGGAAGCTTCTCTCCTGCCCTGAACTGGAGGAAGTGCACCAACATGTCTATCCTCCCCCTTTGACAGT

CGCTGCAACACATCTACAAAGACAGGTGGATGATGGGGCTCAAGCCTCCCTCACTTGCCCTTCCCTTCACTGCAGCC
TCCACCCCAGCCCACACCCCTCCTCCAGGCATGCCTGGGGCCCTCAGCCCAGGCTGGGCTCCAACTGCAGGAAATCG
GAAGTGCCCCTGCCCACAGCTTCCTGTCTGAAGGAGACGGAAACCAAAGGAAACCAGCACGCTTTGAAGAGAGCAGG
GGTTCTCGGTGGCTGAGCTCCCCGTCACCAGGCTTGGGGCCTTCAGAGGCCACCCCCAGCCCAGCCTCAGGACAAGC
TCCCATTGCCGCTGCACTCGCCCTCTGATGGGGCCTTGGGATCACCCCAGCAGAGGGTGGGAACCCGACTGAGAGAT
GGGGAAACGGAGGTCCAGTTGGGCTGTTGACCAAGTGTTCCTTCCTCTTTCCTTCCTGAGTGTCCAGCTATCGATGC
AAACAGCCATCCTGGCATCTGACATTCCTCCATCTCTTTTTAAAAAAAAAAGTTGCCACCACCAACAAAAAGCAACT
CGTTTTTGCTTGTAAAACTAGTATATGTTCTCTTCGCCAAAAACAGGAAACAAGAGGAATGAAGAATCACCCATAAT
CACATCAGCAGAGGACAACTTGGTAATGTTTTGGGTGGGTTTTCCTCTATGTTTTATTCCCATTTTGAACAGTTTGG
GGGAGCTCAGTCTTAGAAATGAGAATCTGGTTCCAATCTGATTCCACCATTGGCTTAAGGGCTCGGAAGTTACCAGC
TGGGCTCTGAATCTTGATCTTTATCTGTAAAACTTGGATTAGGATAGCGCCTCCTTCATGGTGTTGCTGAGCTAAGC
ACATGACACATCTAAGTAAAGGGACACTTCCATCAAAACAACCTTGGGATCGTTCAGCACCGTCTGGACTGCAGGTT
TGAGACAACTGTGGTGAAATGTCTAGAGCCCCTGGTCCCCTGGGATGGGGCGTCACCCGTTGCTACGAGCACACCTC
GCCAGGTGCCTCCACACCCAGCCAGAGAACTCTGGCTAACACGGACCTCTCAAACCTCTGAGCTGGTCACACCCGAC
CTGCCTGCCTGCCTGCGTCCCTGGGACAAGAGCTATAGGGCTCAGGGCCCACCCACTCACCCAGACTCCTCCTTCCG
TGGCAGCTGGCCCAGCCGGGCCCACTCACCTATCCTGTGAGTCTGCAGGTGCCGCTGCACCTCCCCAGGCGTTGGGC
TGGCCCCCAGGCACCGCATCACCATCAGCAGGTCCGTGGCTTTAATCTTCCCGCGCTGCTGCTTGTCATACAGGGAG
AAGCATTCCTTGTACTCTGCCACAGCCCAGAAGGGAGAGTCAGGGCCGCTTTTTCACCCTCCCCCAACTTCTCAATC
CTCTCTCCACTTTCTGCCCCTCCCCTTCCTTCTGCCTGCTTCCCAGTCCGCCTCCCCTCCTAGAGTGGCCAGACCTA
GNCAAAAAAAAAAAAAAAAAAAATAAACACGGCGNGTCAACTTCCCTTCCTTGCCTGCTTCCCACGTCCGCCTCCCC
TCCCTACGAGTGGCCAGACCTAGCAAAAAAAAAAAAAAAAAAAAAAAAAANNCCAGGGCTGTCAAGGTAAATTGAGTT
TCAGATAAACCTCAGTCCATGAGGTCACAAAGCGTTGGACAGGACTTAACAACAACATGCTTACAAAGGTATTTGTC
GTTTATCTGAAATGTATGTTTAACTGCATTTCTTGCCTTTCATCCTGTCAACCCCATCCCCTTTCCCTCCTGACCGC
CCTTCTGCCGCCTCCCTCCCTGCTCCTGTCCTCAACTGACTAAAACCACAGTGGTCAAGAACCCGCCTGCCAGTGCA
GCAGACATGAGACGTGGATTCGATCCCTGGGTCNGGGAGGATCCCCTGGAGAAGAGCATGGCAACCCACTCCAGTAT
CCTTGCTGGAGAGTCCCATGGACAGAGGAGCCTGGTGGGCTGCAGTCCATGGGGTGGCAGAGTTGGATGTGACTGAG
GCAACTTAGCACACACAGGAAAACTTGCTCGCTCTCTTGTTCCAGGTCCAGCCCAGGGAGACGAGGAGCCCACAGCA
GGGCACCAAATTCCTAGGGGATCTCAGCAAATGACAACTTAGGTGGGAAACAGAATAATAATGCCTTGGGGCTTGCT
GGAGGACCAGAAACCCGCTCTGGGAAGGCCTGGCAGAGAGGCCATCATACGTGGTGACAGATATCACAACTGGGAGC
TTCTGAGTCCACTGCTGCAGCCGCCGGTCTCCGGGCTGCTGAACTCTCAAGCCGCGCTCTCACACAGGCCCAGGCCC
CACTCAAGACTTCAACAGGCACTGCCTGGACGTCCTCCCAGCTCCGCGGAGATCTTCCTTACTGCTCTCTTTGAACC
CTCTCTCCCCTCTGGAGGTGAAGTGGCCTGCAGAGCCACAGCTCCCCAGCCCAGTGGAACCAGCCCCCACCTGCATC
TGCCCTGACCCCCCAGCCCTGGAGGCACAGGGAGGCCAGCCAGGAGCTTGGGAAGATAAGGGGTCCTCACAGAGGCC
CTGACCACAGGGTGGGGGTGGAACTGGCCAAATGCCCCTCTCGTGCAGCCCCCGGGACCCATGGGGTAGAGCTCCAG
CTTGAGCCCCACTGGGAGCATGCCGGGCCCTGCACTGTTCCCTAGCACCCTGCACACTTCCCACAGCCAGCTGGTTA
CTCAGAGGCCTGGACCAGCAGCTTCTGAGCGGATAGTAACAAAACCAGCAAAAGCAGCTCCCGGTACTAGCGCTTTC
CCCCAGGCCCTTGTGTCATCCTCACTGTGGCCCTGTGGGGAGGGACCCTAGTTGTTCCCATGGTGCAGACGAGCCAG
CGAGGAAGGCTGGAGCAGCCGAGGGCTCCTGGCAGTGGGGGCAGTGGGCAGCTAGCTCTGCCCAGGCTACCCTGGAG
CTCCACCAGAGGGCGGGAGAGGTGGCAGAGGACGCCCAGGCCGAGGGCCCCTTCATGCTACCTCATTAGGGCCACTC
CCAGCAGCCTAGGGTTTCAGGGGCATTGCTAATCAAAGGCAGGCCAGGCGGTGGGGGTTTAGAGACAATTACCTGCT
TTTATGCCTGTGGAGAGGTTGAAGGGGAACCTCCTGCCCTTTCAGAACTTTCCAGAGCCTGGGCTTTGGAGGGGCAG
ACCACCAACAGGACCCAGCGGCAGGAACCAGAGTGGCCCAGGGAAGCAAGGCTTCACTGGTTCCTGTGAAACCCAAT
TCCCTCTAAGAGGGCCCAGCGCCAGTGGGGACAGGGACAGGGACAGGGGGCCACCCTGCACGCAGGACTTGAACTCA
GCCCCCCGCCCAGCTCCAGGGGACCTGTCTTGGTCACCCGAGCTTACCTTCCGTCTGGGTTTGCCTCTGCCCAAGAG
GTGGGCCCACTACCATCCTGGGACCCCATCCTATGACAGTTGAGCACCTTCCAACAGTAGCCGGATGGGGACAGCAC
ACTCTTGGGACCGGGGATGGACCAGACTCCACAGGGGGCAAACTGCCTCCCCCATACTCTGCACCAGGGGCCCCGTG
GTATCAGCAGGGCTAGAGTCAATGGGGGACCCATCTCCTCCCCAGTTACCTAAAATGTGTCTTCCTGCTTTAAAGAT
CTGGGGGGTCCTCTGGACTTGGAGGCCAGCCCTGAAACAGTAGGGCTCTTCTCTCTCTACTCCTGCCCCGCCTCCAT
GACCTCTACACCTCCTTGCTCTCCCTGCCAGGCCCTTCTCCTGCCAGACCTGCCTGCCATCCAACATCCTGCTCACT
CCTTTCCTTTGTTGTCCCCACCTCCCCACACAGGTGAACCCTGTGAGCACAAGGATTTGGGGCTTCGCTGTGCACTG
CTGTATTTTTAATCCCTAGAACGGGGCTGGCACAAAAAAAGGACTCATAATTGGCACAAAAAAAGGACTCATAACTG
TTGAATGAAAGAACAAATTGGGTTAAATAAACTGGATTAAAATTCCCTTTAAGTTTTCCCTTTGAAAATGCAAATGC
CCCAAGGCAAGAAAGAATCCTCCCCAGACCCAGGATAACACCAGCTGTGCTGGGAATGGAACCAACTCTGATGACAT
CTGTACATGGTACAGACTTAAGAATGTATAGCATCCAACTTACAAGTAATAAAAATACTTTGATTATAAATCCCATC
TAGCCAATTGATTATCACAAGTGAGTTTCTGCCAGCTCTTGTATCCATTGCCAACCTACAATTTTGATTTAACCATA
AGTGAGGTAAGCTGCAGGCCAATGAATGCAACTATTTTCCCAATGACTACACTGTCATTACATCTGATGAGTGATCT
ACTGTTAAACTATTCCTCACCTTTGTGCAAATTACATCCATTAAACTGAAGTCTTTCAAATCTTCAGCACTACTAAT

TGCAACAGAGCTATCTTTTAAAAAACTTTTGTAGACGAAGGGACATTCTTTCATCTTCCTTTGTTTAATAATAGTTC
ACAAAATTCCTAAAGGTGTACCACGTAGCTCATGCTAGCAAGAGCAAGCTGGCCCCGGCTTACATCTGGGTGACGTC
CAGGCACAGAGTCAGGTGCGTGCCTGCCACCTGCTCTCTTGCTCACAGGAGCCTGATGAGGCAGTGCCATTATTCCC
ACTTCACAGATGAGACCACCGAGGCTCAGAGGGGACTGGAGCCCAGGTCCACTGCCTCCAAGGTCCCTGGTCTTTCC
ATTCAAGCAGTCTTTCTTTCCTGAACGTTTGTGGTCCCTTTCTTTGTGTCAGACTTTAGGATACAAAGGGAAACGTA
CCTTCTTCCTTTCTGGGAGATTCAGAGGCCCCAGAGGGAGGGTCAGTGAGCCCCAGCACCATGTTTGGTGGGGGGCA
CACTAACAGCTGGATTCCAGCCCTGAGATTCATCAAGCCTTGTGTTAAGACTGGGTAGGAGAGGCCCCTACCCTGGG
CCTTGGATTTTAAAAGTCCCTGCTCTGGACCCCGTTGAGGGGTAAGAATATCAAGGGGCCTATGGGGAGGCGCCTGC
AACCCACCTACCCTCTCTCCCTTATAGACGCTGCTGGTAACCAGAACTCTGCAAGTCCTGCCCAGGTAACCATCCCT
CAGTTCATCTCCCTGCTTGAGCTCAGCTGAGCCTCCCCAGCAGGGTCTGGCCCCTGGCTTCTGTGTAACCCAACTCT
GTCGGCTTGGCCTTGTGGAAACAGAACCAGGCACTGTGGGAAACTGAAGCTGCCTGGGGCAGGTTTGGGGATTCACA
GAATAAAGTGCTCCAAATGCCAACGCTAGGCTTGCCGTCAGGGCAGAGGGATCTCTGAGGCCACCTCCACTGGCCTT
CACCCTTTACAGATGGGGAGACTGAAGCCCACAGCCTAAACAAAGCACCCTTTCCGAGAGCACAGCCCTGGGCCCTG
TTCTCTTTCTCAGCCAGTTACAAAGCCCCCTTTTCAGCTGGGGCCCTCAGTCACAACCAGTTGGTGTCTGTCCGACC
AGCCCCTCTCCCCCTCCGCCCCCAACTATCAAGTTCTGGAGGAGAACCGAAGCGCTCCCCCTCCTGTCAGGCCCTCT
TGCAAATCACACTGACGACTGTGTACTAGGCACAGTTCTCACCTCCAAGCCTTATCCCCCGCCCCTCCACCAGGCCC
TTCCTGGTAGCAGTCGGCCCCTGGCTTGTGGGTGGGCCCAGCTCTCCCACCAACAAAACATGAGCAGAGCAAATTAC
CTAATGAGACCCATATTTTGCTGCGAGTTGCTAATGAAAGAACAAACCAAACCTACCATTAATTTGGTCCTGGGAAA
GANAACTTGGCCTGCAGCAGAGAGAGAACAGTCAGGAAGGGACTCCATGAAAGCCCCCTGCCTCCCCAGCCCCGGGC
AGCCTGGCCAGGCCACACAGATCCTCACCATCGCGGGGCCTCCAGCTGCCACCCGTGGGCTCGCTGCTCCCAGAACT
GCTTCCTGTTCCCCTTCTTCCCGGCTTCAAGTCGAAAGTCCGCCCAGCCGGGTGGAGGCCCCGGTAATAAATCCCTG
GCTGCCATGGAGACCAGGCCCGCCGCCAGGGGTCAGATTCCAGCTTCCTGTGCTCTGTGGATGCCTGGCTGTGGGGT
GGTGGCAGGCTGGATTTTGAGAGATGGGAGCTGCCCGAGGGCTTCGCCGGCACTTGGTCCTGCACCAGGAAGCTGAT
ATTCTGTAGCCAGAGGACAAGGGCCGCTGTTCCTTCCTGGGAGCCCTCTGTGGACACTAAGAAGGGGTCACCCAGAG
AGGGGATCGCTGGGTCCCCCAGGCCTGCCTTCAGGGGTATTTCTCTCTGAGGATTGCATTAGGGGCTAAATAGAGTC
TCCTCCTAGGTGTTGGGCACTTTGCACATCCCCTTTAGGCACGTCACACACCACCTCATTAGTTCAGTTCAGTCGCT
CAGTCGTGTCCGACTCTTTGCGACCCCATGAACCACAGCACGCCAGGCCTCCCTGTCCATCACCAACTCCTGGAGTC
CACCCAAACCCATGTCCATTGAGTTAGTGATGCCATCCAACCATCTCATCCTGTCATCCCCTTCTCCTCTTGCCCTC
AATCTTTCCCAGCATCAGGGTCTTTTCAAACGAGTCAACTCTTTGCATCAGGTGGCCAAAGTATTGGAGTTTCAGCT
TCAACATCAGTCATTCCAACGAACACCCAGGACTGGTCTCCTTTAAGGTGAACTGGTTGGATCTCCTTGCAGTCCAA
GGGACTCTCAAGAGTCTTCTCCAACACCACAGTTCAAAAGCATCAATTCTTCGGCACTCAGCTTTCTTCACAGTCCA
ACTCTCACATCCATACATGACTACTGGAAAAACCATAGCCTTGACTAGACGGACCTTTGTTGACAAAGTAATGTCTC
TGCTTTTTAATATGGGAAGGAAACCCTGCGTGACACCACAGCCCAGTGACAGCCACCTCCTCCGCCAGAACCCTGGA
TGCCACACACAGGAGAGAAAGGCGCCTCTGAGGAAGACCTCGGTGCCCATTCCACAGCTCTGTGCCTCAGCATCCTG
GCCTTATGTTGAGAGTTAATGGCCAGATGGTCCATTTTCAGCTTGAGTCAGTTCCTTCCCATTCTTCTGGGGGTGTG
GTAGGTGCCCATGCTGGGGTCTCTGGCCCCACCACAGGCAACTCAGGGGACTCAGAGGGAGACCTCTTAAGTGACTT
TCCTGGTCAGGAAGGCTCCCTCTTTTCTTCCTTCCCCTCCACTCCCCACTGCACTTCCCAGCCTTTCCAGCAGAAGC
TGCTGACTTTCCCCATTTCTCAGTGGTTTGTCCCCAGCCCACCCCTCCCCACACCCACTGAGGAAGTGAACGGGGGA
TGGAGCTGCAGCGCTGGGCACAGCTGGGTTCTCCCCCTCCTTCTCCGAGATCCCTGCTGTATCCTGTGTCCACGGCT
GGTGGCTGCCGTGCCTTTGGGGCCCCGACTCCAGGGCCCACCCAGTCTGTCGGGACTGGCAGGTACGGCCACACCGG
TTGTGCACTGCACTAGGGCTCCTGGTCTGGGGGCCTGGAAAGGGCTGCATGCGTGCAGAGGGAGCTGGTGGTAATCT
GTAGAAGTTACATGGATCAGGGCAGGGGGGCTGTCTTTGTCTCCTGACCATTTCCTTGTGAAAGGTTCTCTTTTCCT
CGACTTCACACCCATTTCTAGACCCAGAAAGTTTTAGCTGGCGGTTTCTGACTCCTCCAAGTTGGGGCTGAGGAAAG
GCTGGTGGGAGCAGGAAAGGAGGGGAGTTTCTATGCGACTTACAGGAAGAAGAGGAGAGGGGACATGAGGCGGGGCG
TTGGCAGGGCCCGGGGCACAGAGCGGATGCTTGAGGGGAAGCCTGAAGAGCCAGGCCCGGGCTCGCTTGGAATCACC
CTGTGTTTGGAAAGGGGCTCCTGGGGGAGATGTGATCGTTGGGCTGTACCCCGAGGTGCTGCGAGCCATATAGTAGC
CCAGCCTCGAGACCAAAGAGCCGAGAGACAGACAACAACTTATCAGATGGGGGAATCTTACGGATCTGAAGCCACAT
CCTGGAGCAACACCCCCCTGTGGACGGCAGACAGCAGGCAGGACACGGCGCGCCTTCTGCTGCTCTGGAGGAGGAGG
CGGCTACCAGTTATAGAGAGCATTGACTTCAGAAGGCTCATCAGTTACCAGGGAAACCAGCTGGGGGCAGGGGCAGG
CGTGTGGACAGTCACAGATAAAGGGCCTGTGATGAACAGGACTGGACAGTCCTGTGTGTGAAGTAGGGACTGGTGGA
GCAGGAGATACACAGACAGCAAGTGAACGGCCACCTTGAGTGGCCTGGCCATACACACCCACACATCCTGCATGACT
CTCATGGCCCTGCCCCTCTTCTGAGATATCCCTGGGGTCAGGGATGCAGAGGTGCGTTGCAACCGCAGANNTCACAA
ATAAAATGACAGACTGTTGTTTTTGCTGATTAGTCTCTAAGTTGTTTTCGCTGTTTAGTCTCTAAGTTGTTTTCGAT
TCTTTTGCAACCCCGTGATGACAGTAGCTAGCAGAATACCAAGAGTTCCCAGGTGGTGCAGTGCTAAAGAATCCGTC
CGCCAACACAGAAGACCCAAGAGACACCGGTTCGATCCCTGGGTCTGCAAGGTTCCCTGCAGTAGGAAGTGGCTACC
CACTCCAGTATTCTTGCCTGGGAAATCCCATGGACAGAGGAGCCTGGCGGGCCACAGTCCAGGGGGTCACAGAGTCG
GACACGACCGAGTGCACACACAATCAAGAGTAAGACACAAAGAATGCTAGATAGTAACTCCCTGAGACGCGAGCCAA

CTTTGGGTGAATTTTTCATGGAAGTCCAGAGGAGGGAGGCCCTGCAGTCAGACTCGCCAGTGGTGAGGCCAGGACTT
CCCTGATGGCTTAATGGTGGGAAGTCCACCTGCCAAGGCAGGGGACACGAGTTTGGTCCCTGGTTCCAGAAGATCCC
ACATGTCTCGGAGCAACCAAGCCCCGCACCATAACTACCAAAGCCTGCACACTCCAGAGCCTGGGAGTCACAACTAA
TTAGCCCCTGTGCTGCAACTGCTGCAGCCCGTGTGCCTAGAGCCTGTGCTCCACGCCAAGAGGAGCTGCTGCAGTGG
GAAGCCCACGCTCTGCAACTAGAGAGCCGCCCCACTCTGCAGAGAAAACCCGCGCACAACGCAGACAGCGCAGCCTT
AAACCAACAAAGTGGTGAGGCCACCACTGTCACCTGGTCCCCAGAGAGATTCCCTCTGCTCAGGCTAGGGACGAAAC
ATAAGATGCTTAACAGGCACAGTACAGGGGAGATCGTGGCCGTGAGGCAACACACAGGCAGTGACAGCATTCTGAGG
TCACACCACCCCAGCCTGTGGTTGGCGTCCCAGCCTGCAGTCCCCACAGCTGGCGCCGAAACCTGGAGCAGGGGATG
TCTAACAGATCGTAGGGGTCATAGAATGGTGACTTTCTCCAGGGGAGGCCTGGATTAGTGACCTCAGTAGCGGTCTC
AGGTAGGGCTGAGTGGAGGACAGGTGAAGACTGTAAGACCCTTTCTAATCCTATCCCCCAAGAAGTGGCAAACCCAC
ACCACCAGAAACTTACCTGCCAGTCCCAGGCCGTGGTATACGGTGTTCCACTGGGAGCAGATCCTGTGGTCAGGACA
AAAACAGTCAGACTTGGATGGTGCTGACTAGTTGCTTTGGGGTTACCATGGCGTAGTCTCTGGGTCCATTGCTCCTG
GCATATTCAGTTATAATGTGGAAGGCTCAAGCTAGCCTCCTAGTTCACATGTCAGAGAGCAGATTTCCTGTCTCACT
TGCTGGCTAAGTAAGTAGTCCATCCATCCTCCGCTTAGCCCGGCAGCAGGGGAGGACGGTGTAGGGCTGGTGGAAAT
GCCAGTGTCTGTCATCTTTTACAGGCCAGCTCCTGAGTCTCATGGCCCATAAACTGGTTTCCTTGGTCGCTATTGCA
GCGCTTTCCCCATAGGGGTGTGCAGCCATCAGCAAGCCCTCTCGCGGTTTCCACTGGATTTATCCAGATGCTGTTGG
ATACAGGCCAGTCACTGCTGTCCACCAGCGCGATCACCCCAGACCCCCCTATCAGTGTGCCAAGCCTCTCAAGGCCA
GCCTCTCCTGTATTTCTTGGGAAACATTTTCAGGAAACTTCCTAAAGAGATCTCAGTCAAGAGGGACAGGAATCGCC
AAGGCTCCAATAATTTGCTGATCTCTTTTCCATTCTAAATATTCTTTCATATCTAATTTTATTCACTTTTTTAGAAG
TTTTCTTATCAAGCATACTCAAAACTGTGTAAGCCTCAGACCCCACAATCAGATGTGTGCCCACTTAGGTGATATCC
ATTAAGTCCTTGTATGGAGAATGGGCCTGGGACAGGGTTTGGAGGAAACACGGAATCAACGTAGGGGAGAATTGAGG
TTACTAGGGACGTAGGGGCCCTTGTGCTTTGAGAGGCCAAGGCTGGAGCTGGCTTACTGAATTAAGATTGTGCAGAT
TGTGACCTGCTGAGTGAGGGGGGGCGGCAGGCAGCACCTTCAGCATCCAGCTCCCACTCTGGGGGAGTGAAACCCAT
GGAGTCCTTTTGATTGAAGGCAGTTGGGCTTCCCTGCCAATGCAGGAGACGCAAGAGATGCGGGTTCAATCTCTGGG
TCAGGAAGATCACCTGGAGGGTCCCTTGGAGAAGGAAATGGCTACCCACTCCAGTATTCTTGCCCAGAAAATCCCGT
GGACAGAAGAGCCTGACAGGCTACAGTCCATGGGGTTGCAGAGTAGGACACAACTTAGCAACTAAATAAGTTTCTAT
ACATCCCCACCAAGAACATGAAATCACACAATTCATTATTTCAGATCCCGGAAGCCAGGGGGCCCAACAAGCTAGAA
GGGCCATCCTCATCCCAGGTCCTGAGCAGGAGCTATTGAGCAGGACAGCAGGCATCTGTTACTCCCAAGGCGGTGGG
GGCAGAGGTCACTAACTTCTCATGGGTGTAGAAGGTGCCTGGGAAAAGCAAAGCAGGAAATTGTTGCAGGTATCCTG
AACTCCAGTCTTCACGTCTACACTCTGGATGTGAGGAGGGTTCAGTTCAGTTCAGTCACTCAGTCGTGTCCAACTCT
CTGCGACGCCATGAATCGCAGCACGCCAGGACTCCCTGTCCATCACCATCTCCTGGAGTTCACTCAAACTCATGTCC
ATTGAGTCAGTGATACCATCCAGCCATCTCATCCTCTGTCGTCCCCTTCTCCTCCTGCCCCCAATCCCTCCCAGCAT
CAGAGTCTTTTCCAATGAGTCAATTCTTCGCATGAAGTGGCCCAAGTACTGGAGTTTCAGCTTTAGCATCATTCCTT
CCAAAGAACATCCAGGACTGATCTCCTTTAGGATGGACTGGTTGGATCTCCTTGCAGTCCAAGGGACTCTCAAGAGT
CTTCTCCAACACCACAGTTCAAAAGCATCAATTCTTTGGTGCTCAGCTTTCTTCACAGTCCAAATTTCACATCCATA
CATGACCACTAGAAAAACCTTAGCCTTGACTAGACGGACCTTTGTAGGCAATATAACGTCTCTGCTTTTCAATATGC
TATCTAGGTTGGTCATAACTTTCCTTCCAAGGAGTGTCTTTTTAATTTCATGGCTGCAATCACCATCTGCAGTGATT
TTGGAGTCCCAAAAAATAAAGTCTGACACTATTTCCACTGTTTCCCCATCTATTTCCCATGAAGTGATGGGACCAGA
TGCCATGATCTTCGTTTTCTGAATGTTGAGCTTTAAGCCAACTTTTTCACTCTCCTCTTTCACTTTCATCAAGAGGC
TTTTTAGCTCCTCTTCACCTTCTGCCATAAGGGTGGTGTCATCTGCATATCTGAGGTTATTGATATTTCTCCCGGCA
ATCTTGATTCCAACTTGTGCTTCTTCCAGCCCAGTGTTTCTCATGATGTACTCTGCATATAAGTTAAATAAGCAGGG
TGACAATATACAGTCTTAACGTACTCCTTTCCCTATTTGGAACCAGTCTGTTGTTCCATGTCCAGTTCTAACTGTTG
CTTCCTGACCTGCATATAGGTTTCTCAAGAGGCAGGTCAGGTGGTCTGGTAGTCCCATCTCCCTCAGAATTTTCCAC
AGTTTATTGTGATCCACACAGTCACAGGCTTTGGCATAGTCAATAAAGCAGAAAGAGATGTTTTTCTGGAACTCTCT
TGCTTTTTCCATGATCCAGTGGATGTTGGTAATTTGATCTCTGGTTCCTCTGTCTTTTCTAAAACCAGCTTGAACAT
CTGGAAGTTCACAGTTCACATATTGCTGAAGCCTGGCTTGCAGAATTTTGAGCATAACTTTACTGGCATATGAGATG
AGTGCAATTGTGAGGTAGTTTGAGCATTCTTTGGCATTGCCTTTCTTTGGGATTGGAATGAACACTGACCTTTTCCA
GTCCTGCGGCCACTGCTGAGTTTTCCAAATTTGCTGGCATATTGAGTGCAGCACTTTCACAGCATCATCTTTCAGGA
TTTGAAATAGCTCCACTGGAATTCCATCACCTCCACTAGCTTTGTTCATCGTGATGCTCACTTGATTTCACATTCCA
GGATGTCTGGCTCTAGGTGAGTGATCACACCATCATGGTTATCTTGGTCATGAAGATCTTTTTTGTATAGTCTTTTT
GTGCAGTGAGGAGGGTTACCACACTGTAAAATGCCTCAGCAGCAACTCTGAAGAACAAAGCTGCTTTCCTTAAAAGC
TGTACAGCCTACGATTGTGTGTCTCCTGTTCAGTGCAGATGTCTTGTGGGTTCTAGGAACGGGAGACTATAAAGCCT
CTGGTGGTTGGGTGGTGAGGCAGAGACCCTCAAAGATTAAGGGCTGAGGCCTCCAGGAGTCATGCGACTTGCCCAAG
GCTAAGTGCCAGCCCCCCAAAAACCTGCCCCCACCTCCACCCAAAGGACACCCTCTCTGCTTCAGTTTCTAGAAGGC
GCTAGCAGCATGGCTTGGGGGCTCCATCTGTAGACGCCCTCTGCCGGCCCCAGCCCCACAACCCTCAGCCTCCTCCA
AAGAGAGCTGGGACCCCCAAACTCAGGATCCTGAGTGGCACAGATGGCCAGGGGGCAGTGGCAGCACCTGCCTTCAG
CCAGACAGGGCCGAGCAGGCCCTTGGGGTCAGCCCAGGGCTCCTGCCAGGCCCCAGACCAGCCACTTCCCTGCTCCA

GGGTCCCTCCTCCTCCTGCCTCCCAAATTCACACCCCAAGGAAGCAGCAGCAGCCGCCAGTGGGGCTTGGCTCCACA
CTGCCTTTATTCACACTGCTCACCCCTGATGCTGCCAGCTGACAGGGACTGACCGGGGCTGGGGGACCCCTACATGA
AAGGAGGATTTGAGGCAGAGTCCCTGGCAGCACAGGAGAGGAGAGGAGGGCAGGGAAGGCGAGGTAGCCCCCCTGAG
CCCCTCACAAGCATGAAGGCCAGCAATGCCAGGAAGCGAGGCTGTGTGCCTGGCCCTGGTGCTGGGGTGGGGGGCAT
CTTGCCAGAACAAGGCCCCCTTTCCTGGGAGTAGGGCCGCTGGTGGCAGGTGTGGATTCAGGCTGAGACACGCTGAC
GAGAGCCCCACGATGTCTGCCTCTAGTGAGTGCAAGTACTCAGAAACCTTGAGCAGGTGAACGAAGCAGGTGCCACG
ATCCTGAAGGCGGGACAGAGACCGAGTTCCCGGGAGGGGAAGGTGGCTGGGAACGACCACTGCCAGACCCCAAGATC
AGGCCGGCAGTTGGGGGAGGGCTGGCTCTGGCTCCCAGCCCCACGCAGCCTGCAGTGGGGTGAGGATCCGGGGTGCA
CATACCCCTCCCAGGGTTGAAGAAGGCAGGGACAAGGAGGAACCATTCACCCCTCCCAAGGCTGCCCTGAGAGGAGA
CGGCAACCAGCCCACAGCCGACTTGGGACCCGTCGGCCCGTCTGCTGCGAACATGGCCCAGGGAAGGGGGTGCATTG
CAAATGAGTACACAGCAGGTTCACGGAACCCCTGTGGCCTGGCCTCAGGGTCAGGTGGTCTGAGGGGTGAGCCGGGC
GGTAGGGACAGGCTAAGTTCCCTCCCAACCTCTGCCCCACCCAGGCCTCTCCCCTGCGGCTACCAGGAGACACCCAG
GAGGAAGAGACAGCCCTGGGCAGTGAGAGTGAGAAGGCCCGACTCCAGCGTGCACGGGAAGGCCGGGGCCAGCTCTT
GGTCTGGACACACTGCCTCACACACACACACACACCCCCCACCCCCACCCCACCCCCTATCCACAGGCACCAGGCRT
ACTCTCACCCACACTCTCAGCTGCGGGCTACACACATGTGGGTGCACACAGGCACACACTCCGCACACTCAGACACT
TTCGGCCGTCCATCCCCAGTGGGTGCACACACACACACTCCGCACACTCAGACACCTTCGGCCGTCCATCCCCAGTG
GGTGCACACACGCACACACTCCCGACACTCACTCTCATCCATCCATCTCCCACCCACAGAGCACCAGGGGCCTGTGC
CCAGGACTCAGTATTGGCTGCTGACGTGGAGGGTGTAGAAGGCCCAGGGCTCGGGGACATAGATGACGCCCGGGTAC
TTCTTCCTGAGGACGGGGTCATTCCACAGCCAGGCGACCCAGAGTGCCACGAGCAGGAGGGCGAGGGCGAAGGAGTA
GAAGAGGAAGAGGCCGTTGCTGTCGAGGAAGAGGCGCTTCCGCTTCTGGTGCAGGACGAGGGCCAGCATGGCAAAGG
AGGTGAAGATGAAGAGGATGAAGATCTGGCCCTCGGTGACCAGGTACCTGAGGAGGCAGGCGGGCAGTGAGGGGTGG
GGGCCCTCCTGCCCAGCGTTCTGCCCTCTACCAAGCGCCAGGCTGCCCTCCCAGCCGGTCAGAAGAGCGCATCAAGG
GAGGGGCTCCGCTCCGAGCCAGGGCCAGTACACAGCATTCCGATCTGGCTTCTAGAATTACTGTTTTTCAAAAAGTC
GAGTTTGTTGCCAATATTTAAATAACAGATTACCTGCCGACGTGGCGTTCTGATTTCCCTGCTAACGTTAGGTTTAT
ATTCCCACATGACAATGACTGACGAGGAGAAAGCAGCTGCTCAGAGACAGCATGGCCCCGCAGCTTGCCACGGTCCC
CACCACTCCCTCCTGCCTCACCTGATGCCTACTTGATGCATTTCTATCTCTTGTCTAGTTCCGAAAGGCATCTGAGT
TTGGGACGCAGTTTCAGTGTATGGGGTTGTCCTGACTGTCTCCTTGAAGGCTGGGTCAGTCTCATTCCCCAGGGTTC
CCTCCAGCTCTCAGGGCAGTAGCTGGCATATGGCCAGGAATGCAGTAACACTTGCTGGAGGGCTTCCCAGAGGTTAT
CTCATCCACTCCCCTGCCTCCATGCTGACCCATCCCACTCAAGCAACTGCATAGCTCTGCAACTGAGTGGCTGAGTC
ACCTTGGGAATTTACATGGCATCTTGGAGCCTTTGTTTTCTTACTTATAAAGGGGGCATAAATGGCCCCAGCTCTCG
AGGTTACTGATGAGAACTCTATACTTGTAAAGGCACATTGTTCGGTGAATGGCCCCCTGTAGCTGTCACCGCATTGA
CACACCATGGGTAAGAATGGGCTAAGCTTCTTCTCACTTCTCACCTCTTTCCCTCCAGCTGTGGTCGTCACCCATTC
CTTCTAATCCTGTCCTCAGTGGGCATGTGGATTCACCTTCATGAACCCTCACTGCTGGCCAGGGGCAGCCCCTGGCG
GTGGGGGAGAAGGCCCCAGGAGAGGAGGCCTTTTTCATTAGGCTGGAGGCTTCCCCCCAGAGAAGGGCTAGGACACC
CACCCTTTCTAAGACACCCACAATCCCTGGGGAGCCTGTCTCCAGAGCAGAAGTGACAGAGGAGAGGGAAAAGCCTG
GCTGAGGCCTGGTCCTCTGGAATGGGAAGAGGAAACCAGGTGGGAAACCCTGAAGCCATGGAATTCCCCCGTAAGTA
TCCTTGTCGTCTCCATGAGGCTTCATGGTTTACGAGCCCCTCTCTCTCCCTCTGGGTTCTCACCATGTCCCAAAAGG
GAGACAGCCTGCTATCATCCTCATTTCACAGAGGGGGACATTCAGCCCCCACCCCTCCCACCCTGCCAGGCTCTACA
GACTTCTGCTCTACAGCCCTCCCCGGGCGTTCCCGATGGCTCAGTGGGTGAAGAATCTGCCTGCAACACAGAAGACA
CAAGAGACGCAGGTTCGATCCCCGGGTTGGGAAAATCCCCTGGAGAAGGAAATGGCAACCCACCCTAGTATTCCTGC
CTGGGAGAATCCTATGGACAGGGGAGCCTGGTGGGCTGCAGTCCATGGGCGCACAAAGAGTTGGAACGACTCAGCAA
CTAAGCACACACAGAAATTATGCTTCGATGTGGATTCTATGGGCCCTGAGGTTACTCTCTGAGTGTAGAGAGGCCAC
ACTCAGGCCCGCCTTCCTGGTCACTGAGTGGCGGCCAGGAGGCACTGTCCCAGTCCCCACTGAAGAGGAGGGAGCAG
GGACCCCAACAGGGAGGCTCACAGGGCTCACACAGCAGCTCTGCGGTCGGAAGAATATTGTCACTGACTGGATCACT
AGGGGGCCCCCTCCCTGCCACTCCAGGTCCGCCATCCACTCACCAGTAGTACAGGCCGCTGGGCACCACCAGGAGCA
GGGCTGCCCCTGGCATTGAGCTCTCAGCTTTGGTGGGAGTAAAGCAGCCGCTGAAGTACATGAAGAGGATAAGGAAG
AAGGGGATGTACCTGCAATAGGGAAGCCGGTGCTCAGCTGCTTGGCCGCCCGGCCGGCCAGCCCCCGTCCCCCCACC
GTCCTAGGACTGTCTGCCTGTACCACCCTGTGCGGTCTCACTTTGTTTGGGTGAGGGCACTCCCAGCCTCTCCCGGG
CGGGGGGTCTAGGAAGGAAGATGAGGCCTGCAGATCCGAGCCATGGGCAGGGGCGTGGGGCCAACCTCCTCCCACAC
TGTCCCAGCCACCTCACTCACTCCCCGTGTGTGCATGTCTCTCTCACTCTCTCTGCCCCACACACTCACCCCAAGAC
CAGCAGCCCCCAGCCCCTCACTGCCTTGGGAAAGGGTGCACCCTGCAGGGTTGAGGGGGGAGAGGCCCAAGGTCCCA
GTCTGGGGGGAAAAAAGGGGTGAGGGGAGGAACAGCACCTGCTGGGGACCTCCCAGGAGCTCTGTGAAGCTCCTTCA
CTCACACTCAGTCCCCAGCCTCACCCCCTTCCATTCAAAGCTCACCCAAGTACCACCCTCTCTCCTCTCCACACACA
CACACACACACACACACACACACGCTGCCTCCCCACACGCCTGTGCCAGCTCGCACAGCACCCTCTTTCTGTTCCCA
CCCCCACTCACCTCCTTCCCCACCTTCCCCTCAGTGGCCCCCACTCCCTCCGAGTCCGGCCCTTTCATTTCTCCAGG
CCCCTGTTCATCTCTCCTCCTCCCTTCCCCAGCCTTCCATGGGAAGGGTCTCCTGGCACAGGCCAGGGAGAACTGGC
AAGAGGCCTGTGTGTGCAAGGACACCCTGACTTTCTAGGCCCCAAAAGAGAACCCAGAAGGACAATGTAGGGCCCGC

GTGACACGAGTGAGCCCACCTCACTCGGGAGTGGGGCCACCATGTCCAGGGGTGGCTCTGGCACTAGCTGTCAGCCC
ATGGGCAAAGCCCCTCCCATCGGCCCCAGCTCATCCTGGGATAAAGTGGGCTCTGGAACAGACGACTGCCTGCAGCT
CGGCCCGCATCTCCTGTGGCTGAGCCCACATCTGTGAGGTTTCCTCCGCTGTGGGGTGAAGGGGTGGGAGGAAGGAG
CTCCAAGGGGAGTTCGGAGGAGGAGGCGGCACTAAGAGCACCCCTACTCCCACCTCGCCCCCAGCCTGGTTCAGCCG
GCCCTCCTCAGCTCCCTCCCTGCCTCCGGGTCCTCATCCTCATCCCCAGGGTGGGGGGAGTTGTTTACCAGCGCTGC
AGCCCCGGCCCTGGGAGGACAGGCGCTCGCTGCCCTCTAGTGGCCGCCGCGGACACTGCGGGCCGGTGGGCACGGAA
GGCGGCGAGGCGGGAGCCAGAGGAGCGACAGCGCCCCCTGCTGGGGAAGAGCACGCCATTCCAGCGTTTGGGCTGGA
CACTTGGAGACCCCAGGGCCCCTTGTTCAGCGTCACTATGGCCAGCGGCAGAGCTGGCCGGGGCCATGGGCCCCATC
ACTCACCACATGGAGTGGCCCAGGTACTCATCGTAGTAGTAGAGCAGCTCAAAGGAGTCGATCTGGGGGGCAGCAGA
CGCCAGGCGGGGGGCAGAGAGGGTGGGGCTTCAGGAAGGGTGTCGGGCCGCGCAGCCCCCGGCTACCCTCCCCCTGC
CTGCCCATGCCTCCCTTAACTGCCCCCAGTCCCTCCCTATGGGGAGGAGGGGATCTGGGTCATCCTGGCTGGGGGTT
GAGGGTGGGTTGGGGGGGCGACCTGCTCTCTCTGGGGCACCTGTGATGGGTCGGCAATGTCTTTCGCCTCTAGGACA
GACCGCACAGCGGAGGTGGGGGCTCACCAGCGTCTCTGGCTTGAGATTCTTGATGATGGGGTTCTCACGGACCGACA
GATGGTTCTGGTAGCCACTGAAGATCAGGCGGTGGTTCACTGAGTCGCCCACCAGGTGGATGCTGGCCCCCATGATG
AAGGTGATGATGCTAACGTAGATCAGGGAGCGTGGCAAGGTGCGGGGGCACCGCTCGATGAGCTGAGGTCAGAGGGA
GGCCCCACAGCATGACCCCTGACACCATCACCACCGGAGACGCCCTCTGCTGGCCCCGCTCACACCTGGAGTGGGAT
GGATGCCACCAGCCGGAAAGCCTCTCTCACACGGGTTCCCTCTAGGGGAGTCAGTCCAGCCTCCATCATCTTACACG
ATCCCTGTGAACCCAGCAGGTGCTGGGGAAGAGCACCCTCTTTGGCGCAAGAAGGAACACACTCCGCGGGGCCAGCT
CATTGCTCAAGGTCAAACACACCCATATACTGTGCTGGAGATGACAGCTAGGCCTTCCATAGAGGCAAGCCAGGCCC
TGACCTCCCACCACCCCTAATTCTACTTCTGGGTGACTGGGCTGGACTGTTGAGGCTGGGCTGCTCCCATTTCTGAG
CAACTGAACCATGCAAACAAAAATGTGGGGCAGGTGTTTCTGAAACTAAGTTGTTTAACTAAAGACCCCTTCAAATT
CAGAGACTGCACCTTCCTTACTTTGAAAGAAAGAAAATGAAGTCACTCAGCTGAGTCCGACTCTTTATGACCCCATG
GACTGTAGCCTACCAGGCTCTTCCATCCATGGGATTTTCCAGGCAAGAATACTGGAGTGGGTTGCCATTTCCTTCTC
CAGGAGATCTTCCTGACCCAGGGACTGAACCGGGTCTCCCGCATTATAGGCAGACGCTTTACCGTCTGAGCCACCAA
GGAAGTTCGGAAATGTCTCTAAAAACACAAGGATGGGGATAGTAAATAGCAGTAGTTTTAAGAGCCTTTACTTGGCA
AAATGCAAAGCCAGCAGAGCAGAGAGCCCTCTAGGAGACATCTCTCTGTTCCCAGTGGCGCCTCAAAGTCAGGGCTC
TCGGTCCGAAGGAGAGAGTTCACCCTCCATTGATGCCCTTCCTTCCTCCGGCCCCCACCTGCCTGCCCAACTGGTCC
CTGACCTCTTCCTGAGCAAGCCAGCCACTCCTTCCGCCTCAGCCTCCCCAAACACGGCTCCTCAAGCCCAGCCTTCT
TTTCAGATACTCCCGCCTGAAGGTCTCTCCCAGTTCACCCCCAGGCACATCAGGTGGAGTCCCAAATCACATGCCCC
TCCCTGGGGAAGYCTGCTCCATCCCTCCCAGGCAGAGGCAGCCCTGCCTCCCCTGGGGCCCTCAGGCCCTTGACACA
TCCCCTCTGTCATAGCCTTCGGCCCCCAGAGTCAGGCAAGGGGACGGCCTGATTCAGGGCACTCCCAGGGCTGAGCG
GCAGAGCAGAGAAAGGAACGAGGGCAGGGAGGAAGAGGAGGGGCTTGGGGTAGTACCTTGAGCAGGAGGAAGGGCGT
GATGATGTTGTAGGCCATGTGGAAGTAGTCCCCCACGCTGGGCTTGTTGAGCGGAAACCACTCGAGAGGGAACACCA
GCTGGGGAGTAAGTGGAGAATGCAGTTTACAAGGCCTGACCCACGCAGCCCTCAGAACTGCCTCTGCACACACTGTG
CTCTTTCTGTGATCTCCTGGAATCCTCCTGGCAGCACTGAGGTGAAGGTGTTTCCAGACAAACTGAGTCCCGGAGAA
GGTAGACGTGTCCCCAGTCACACAGCAGGTACAAACCCAGAGTCTGATCCTGTGTCCCTCTTCCTGTGGATACTAAG
TAAAGGTCCCTTGGTACCCACCAGCCTGGGAGTTGGAAGATCCTTTGGGACTCAAGCATGCCAAAAACCTTCTGGGT
GACCCCAGGGCAAGTTCTTCCCTTCTTCCTGCCTCAGTTTCCCCATCTGTAACACGAAATGTGGCCCAGATGGTCTA
CATGGACCCTTAGTGGTCTGGCTCCTTAGAATGATGACAAGAAATGCTGACCACCCAGGTGTGAAAAACGCCCTGTC
CACTGAGCAGATGGACCACTTCCCAATCAGGTCAGAACTATACCTTCTGCAACAAAGGAGAAAATAAGTGAGTTCTA
AACGCAGTTTCAGTGCTTCCTTTCCTACCTTCCCTGAGACAGGCCAGGGCCATTCCTCTGTCTGGACTTCCCATGTT
CTTCCTACTGGTCTACCCCAGTGAGTCCAGCCTTGTCGGGGTTAGTGTGATAACACCTATACAAGATCACAACACTT
CCAATCCCTGCTCTGCCAAACTGTACCTCTAGAATCTCCTGATGGAGAACACCTGTAACGATCATAAATTTCAATCA
AAACTTTAAATGTTTATTTCTTTTAATTACAAAATGTTGTGCTGGATAAAACCAACCTGCAGAAATAAATCTCCCAG
GAGACAGGAGTCTCAATCTCTTTCTTACTTGCTGTGTGGGCATCTCTATAGCTTTATCTACAGCAAGATATACAGCT
GGCCGAAACACCCGGTGATAGCAAAGAATCAGAAATAGGCCAACTCTTCAACCAAACAGGACTACTTGAATACACTA
CAGCACCTCTTTATTACAGAAGGTTAAGCAACCATGAACAGAGATGACACCTCGTATGTGCTGACATGGAGAGCTAT
CCATGAAACAGTCTAAAGTAAAACAAGTAAGCTATAGACTAAGGTATAAAGTAAAATCTTATTTTGGTTAGAGGGGA
AAAAAGAGATGTACACATGTGTTGCATGAGCAGAGAAAGGAGTGGAAGGATGCGCCCCAAACAGTCAGAAGTGGGAC
TGCAAAGGGAGTCTCACACTTCTTTATATTACTGGGGATGAACTTACTTGTTATAGAACCATGACATACTACTTTTA
TAACTTGTAAAAGGTTTTATTTTGCAGTTATTAAAAATTTGAGACTATCATGTATACATCTGAAAAGCTCATGTATA
CACTAGACCTTTTCAGATGTATGCTCCTCCTGAGCCCACTTAGGCTTACAGCGGCCACAGCGACTCTAAGCAGGATG
GCGGGAGGTGATCTGCCCTACCGGGGAGCCCTATTAGGCCAGATGAGGCTGAGTCTTAGTGAGTGAGGAGAGCCACG
AGCCACCCTGGCCTCATCTCCTCTCTCCCTGGAACACCAGTTTGATGGGCAGCTCCTTTGTCCCCTGGAACCTTTGT
GCTGCAAGGGGGCTCAGGTCAGCTGGTCCAACCAGTCTTCTTACAGATGAGAGAACCAAAGCCCAGATGATACTGCT
GCTGCTGCTGCTGCTAAGTCGCTTCAGTCGTGTCTGACTCTGTGCGACCCCATAACGGCAGCCCACCAGGCTCCCCC
GTCCCTGAGATTCTCCAGGGAAGAACACTGGAGTGGGTTGCCATTTCCTTCTCCAATGCAGGAAAGTGCAAAGTGAA

AGTGAAGTCGCTCAGTTGTGTCCGACTCTTAGCGATCCCATGGACTGCAGCCTACCAGGCTCCTCTGCCCATGGGAT
TTTCCAGGCAAGAGTATTGGAGTGGGTTGCCATTGCCTTCTCCGCAGATGACACAGACTTGCCCCAAATCACCTACA
AACCTGACTGGGTGCCCTGACTCCTGGGTCAGTCCCTTTTCTTCACTTACATTTAAAATACAGCTTGTAAACCTTAC
CAAAAGCCTACAGAAAGGAGACTCAGCATCCACCATACCTCCTCCCCAAAACCACTGCCAGCAGCTTTTCTTCATCC
CCGTGAAGTCTTGGATCCCACGTATATAAGATCGTTTGGGCCGTTACTGACTGGCAGGCAGCTCATGTGCATGCCAC
CCAGGCATCTGTTTTCTCCTTAAACCTTCCAGAATGCTCTATAGTACCCACAATTCCTGTTYACTGCCCAGACAGCA
TCCTCCATCCTGCATAGAGCAAAGCTGACTTACCCAGTCCCTACTGCTGGACACTCAAGTCCCTTCCAGGTTTTGGC
TATTGTGAGCAACACTGCAGTGAACACCTTTGGCACAGGCTCTCGCCTGTGATTACATCTTCAGGGCAGATTCTAGG
AAAGGGTCGTACTGAGCCAGGCGGGATGAACATATTAGGGTTCTGAATACAATCATGATGGAAATGGGGGGTGGGCT
CTGGGGCCAGAGTGCCTGGGTGGGGGTGGCAGATGCTGGCTCCTCCACTTGCGGACTGGGTGACCCCGAGCCAGCTA
CTTACCCTCTCTGTCCAAAGTCTTCTCATCTACAAAACCGAGATTAACAATAGCACTGATGGGAGAATTAAGAAGTT
TAAACACATAAAGATCTTAACTGCTACTTCTAGCTTACTGTAATTATTGTTATTCCTTTAGGACAGGAATATATCTT
GCTAGGCACTATCTAAGTACCAGTTATCACCTTGACCATCTGCTTTCCCAAAGGACTTTGCTACTTTGCAATATCAC
TCGCATGTACAAGGTGTCAGCTTCAGAGCTGGATATTGTTGTTTTTACCACCCATATGCAAAAATTAACACCTCAGT
ACATGGTGATACCTCACCTTCTGTTACCTATCTGTCTATCCATCTAGCTGTCTAGCTGCTCATCAAGTGAGGGGCTT
TGGTGCTCTTCCAGTGTTAATCTCCACAGAGGCCACCACTGTACGCTGAGGCATGGGGTGGCCTTGGAGAGAAGGGG
ACGGTTTCACGAGACGACACTAAAAATTACCCAAGGCTCAAGGCACATATGACACGCACACTCAGCCCATACTCTTC
ACCTCCAGAAAACACAATTCTATCCAAGGCTGTGGGCAAGACTGGTAGGATCCTTTTCCATACTCCACTGTCTGAAT
TGGTGCTTTGGTTCAACCTGAATGCGAGATACTGGGGCAGGTGGTGGGGTCACACCACAGTACGACTGTTCGGATCC
CCCACTCTCTGGCAGCTGCCACACTGGGTACGGCCGCAGCACTAAGGGACCGCATTCTCACAGTGTACCAGTAGGGG
CAAGGGCTTGTCTGAGGTCACTCTGATCATTCAGACCACACTGTCTCCACACTGCCAGTYGGGGGGTTCCATCCACA
GCTTCATGCTCACCATGGCTATGGGGCGGCCAAAGTCCAGAACCCAGTTCTGCAGAGTGAAGTAGAACCAGAGGTCA
AGGTGGAAGGGAGCCGTGCCAGCAGCCTCGTCAGCCTTGACGGAGCTGTGCCTGGGAAAGAAGTAAACGAAAGCAGT
CAGCCCTCCTCTCCTCCATGGGAAGATTCCCCACCAGRACAAGCCACAGGAATTTGTGGGCAGGAGGAAACACACAT
GGTTAAGTTCTAGGCAAAGGTTTGGCTCGAAGTACTGGGGCCTGGGGGAAGGGATCGGAGATGGACTCCGTCTGCCC
TGACAGTTGCTTCTCTAGGATTCAAGAAGGAAAATACTCAACAGATACAATCACACTGGGCTCTCACCTATAAGTCT
ATAGGTGACGGGGAGATCGGATGGAGCAGAGGATGCAAACGTGGCTGAAAGGCATGAAACATGGCGATGGATGATGC
TAGCTAGGCTTTACCGAACACTTTCTACATTCCAGGCACTCCTCCAGCTGTTTACATGTACCATCTCATTTAAGTCT
CTTAATCCTCACAGAAACCCTCTGAGGTCACACTATTACCCTCATTTTACAAGTGAAACAACTGAGCCCAAGACGTT
AAGTGCTGTTGTGCTGTGCGTAAGCAGTTGTGTCTGACTCTGTGCGACCCCGTGGACTGCAGCCCACCAGGCTCCTC
TGTCCATGGGGATTCTCCAGGCAAGAATACGGGAGCAGGTTGCCTCAAATGATTACGTTTCAGAGCTTCTGCTGTTA
ACCTCTAGTCTTAAAGTCATAGGTTTCCCTGGTGGCTCAGTGGTAAAGAATCTGCCTGCCAAGACGAGATGTGAGTT
TGATCCCTGGGTGAGGAAGATCCCCTGCAGAAGGAAACGGTAACCCACTCCAGTATTCTGGAGAATCCCATGGACAG
AGGAGCCTGGTGGATTATACAGTCCATGAGGTGGCAAAGAGTCAGACATGACTGAAGTGACTTAGCAGGCACCTACC
TGTTCATTTCAGCGGCTAATCAAACATTAGCTTAGAGGTGCCATGAAGGTTTCCCTCTGTCAGCACTGCTATTGCTT
CGGTAATTTTCTGGGATCCTTTTCAGCCAATTCCAAGAATGGTCCGCACTGGGTTGCTAGAGCAACTCAGACTCAGG
AAGGCAAGCTGTTGGGGATCTCATTATCCTCTGGTGGCAGAATGGATGGGTGGAGTGGGTTTGTCCAGGTCAGCTAA
GTCTTCCTATCCACCCCCTCTGCTCAGTGCCCACTTACTGTGTGCCATGCTAGAAGTGTCACCAGCACCATCTCAAT
GAAGCCCCACAGCAATACCAGGAGACTCTTACCCTCCCATTTTACAAACAAGGAAACTGAGGCTTGGCCAAGGTTGC
CTGCCCTAATAAGAAGCTGCATCAACTATTCCTCTGCCAGACTATTTCCCATTCCCTCCTCCTACTGGTGAAAGATG
ATGTTCCCTGAGACAGAGGGAAGGCCAAAAACCCAGGATGGCCCAGTGGGACTTTTCTAACTACAGTTGAAAAGAGC
ACCCTCCTGAGAGCCCGAGAGCCCGGGGCTGCCTGAGGCCAAAGTACCAGCCTCAAGGGACAGTCAGATGAGAATGA
AGCGGCCCTGCAGAGAAGAGCAGGGAAAAAGAGGAGAAAAAGGCCCGAGAGGGTTCACATCCCCAGTTCCAGGCATC
CCCAAGGCAACAGTTTGGCCACAAGAGCCCAGACATTCTCTTTTGCTAAGGGGGCTGCAGTTCACACTGGCTTTCTG
TCACTCTGCAACCCAGGAGTCCTGAAGTACACAGAATTCAAACCGGGGTCTATATGGATTCAAGACGTATACTCTAA
GTTGTTTGCTACTCTGCTGCTGATTACTATCAGCTATAGAATATATCAGGAGATCAAGATACAGAAGGACTAGTGTG
GCCTGGAGCAGCATGCAAAGCCTGCAGGAAGGAATGCAGCCTGGAAACATGGATGGGTGGATTCCTGTTGGCAGAAA
AGCGGCGAGAGTGCAGCATCAAAGGCCTGGAGGTGGGAGTGAGAAAGGCGCTGCAGGGAGCTGGGATGCCTGGCTTA
CTAGTGGAAGGAAGAGAAGGAGGGGCTGCCGGGCCTTCAGACTGGGCAGATACTCAGAGACAACTAACCGTGAGGAC
AGTCTAGATAGAGTACGACATGAGTGACTGGTGCACAAGTCCCCCTGTGCCCTTCCGCTTCATGACCATCAAGGCAA
GGCCTCCAAAAGCAGGCAGGGACAACCTCCAGGCCTGAAGAGCTTAGATGAGGAAGAAGGAAATAGTCTTCACTGGA
GCTGGGATGCCTGGCTTACTAGTGGAAGGAAGAGAAGGAGGGGCTGCCGGGCCTTCAGACTGGGCAGATACTCAGAG
ACAACTAACCGTGAGGACAGTCTAGATAGAGTACGACATGAGTGACTGGTGCACAAGTCCCCCTGTGCCCTTCCGCT
TCATGACCATCAAGGCAAGGCCTCCAAAAGCAGGCAGGGACAACCTCCAGGCCTGAAGAGCTTAGATGAGGAAGAAG
GAAATAGTCTTCACTGTTGGGACTTCCCTGGTGGTCCACTGGCTAAGACTCTGAGCTCCCAATGCAAGGGGCCCAGG
TTCAATCCCTGGTCAGAGAACGAGATCCCATATGCCACAACTAAAGAGCCCACATGCCGTGATGAAGATCCAGTGTG
CCTCGATGAAGACCTGGTGCAACCAAATAAATAAATAGATAAACATTTAAAACAAAAAAAAAAGGTCCTCAGTGTTC

TGTTCTGGAAACCTTCAGAGCACGATGTTTGGTAGAGAAAATAAAAGGCGTTTTGCAAGTCCGGGAAGCACTTAGTG
TGAGTCTCTGGAAGCTGTGGATACCAAGCATGTAGTCAGGGCTCGATGACTGTTTCCTCCTCCACCTCAGGAGGCAC
CTCAGTCAACAGGGATCCCTGCCATGAGTCCAGCCTGGCTGAGAAGAGCAGAAAGAAAAAGGGCAGGTGCAAGGCTG
ACACCCAACTTGGGAATCAAGACTGGTGAGCGTCTGTCTTCTTTGCTGGGACTCTGAGCACACGAGGCAGATCACGG
ACAGGGAAGGCATCCTCTCCCCCAGCCCTGTCCTGGGGATGAGCATCCTCATGCTAACCCAAGTACCTGACTCCTCA
GACACAGAGCCCCCAGACTGCCAGCTCTAGGCCCCAATCATCAGAAACTCCTGCTGAGAACGATGGAAGCACAAGGG
CAGGCAACCCAAGACACGGAGGGCCTTTTCTCTCCTCCACAGCTTTACAAGCCTACCATCTGGGAAAACAGACACCA
CAACGGTTGGAGGGTACACACCATACAGCGCATAATATGCAAATGAAACATTCAAAGCAGTCAGCAGTACCTGCAGG
ACTGCCTGGCAGAAAAGGAAGGTGTCTCAGCCACCGTGCAGTCAAGCAGGGACCTATGAATGGGGACTATAGAGGGA
CTGATTTTGGTTCTATAGGCACAATGGTGAAATAACCTGACATGATGAGAATTTTAAACAAAAAATAGAGAAAACAA
AAGACAAAAAAAAARAGGAAAAAAAGGAAAGGCAATTTGAAACTCCAGGAAAAACAAAAGCAGTATAAGAAAGTAAA
TGTAACTTTTAAACTTACTGGACTCTGTAGGGAACAATATTTACATGGCTTAACGAATAAAAACATTGTATATTGAT
TTTCAAAGTTTAGAGTCAATTCACAAACAAAACATGGAAGAAAATTATGGTTAAGAAACAGCAGGTAAATACTGTCA
ACCTTGACAACTTGAAAGTAAAACTATAGATGACAGGAGCTAGGGAAAGAGCAAAGAAAGCTAAAACTAAGGGTAGG
AGTCTAACAGTATCCTTTTGAAAAGGAATCAAGAGATACTGCCCTTATTTTAAATTAACTATACTTCAATAAATAAA
TAAACTTTATTTAAAAAGATACTGCTTTATAGTTGACCAAAACAGAAATAATGGTATCACGTATTTCCTAAGGCTAC
AGAGGTAACAAAGAGCAAAACTAAAAATGGTCATAAAAAGAAACTATCTTGTGAAGATGGAAGAGGTATGAGTAAAC
CAAATCATAACAGGATAATTAGTAGGTCAACATTAGCTTCAGAGGCTAAAAATCGATAAACCAAGAAATAGCAGTTG
AGTGTAGCTCTTGTGTGTGGACTCTGGAGCCTGCCTGGGTTCATATCCCAGCTCTATCACTAACCAAATCATGAGCA
AGTTACCTCCTGGGCCTCGACGTTCTCATCTGTAAAATGGGGGATAATAGTATATCCTACTTCAAGAATTGTGGTCG
GGAAAAAAAAAAGTTGTGGTCGAATTAAATTATGCAAAAACATTTTTCAATGCAAAAACTTAAGAGTAGGGCCTAGTC
CACAGAAAGTGGTATGTATGTAGTGTTTGTTACTATTATTTAGAGATATGAATTTAATTCTGATAGAAACAGCTAAA
AGAAGCTAACATGGTTGTCTCTGGTACTAGGGAGAGGTGAAAAAGAACCACTGCTTTTTACTCTAAGCCTGTCTGTG
CAATTTTATTTAATCATGGGCATGTACTACTTTGATTATAATACAAGAAACTGTTAACAGAACACATGAAAACTGGA
AAAGAGAGTGAGCACTTTGTCACTAGAGGTGTGTAAGCAAAGCATGGATGGAGCAGAGGAAATCCAAACATCCTAGG
AGGTGGACAGGGTGATTCTGCCTACTGTCTTGTGCAGCAATGAGACTGAGAACAGGGGGGCAATAAGGCAGGGAGGT
GGGCAGAGACATCCTGGAAGAGAGCCAGCCCCTTATGCAATCCCCAGGAGCATCCTGGTTGGTGGGGAGCTCATGCA
GGAGTAGAGAGGGACATCACACCTGAGCTGCCAGAGTGCAGCAATAAGAACAGAAACCACAAGCCCTAGGTTTAAGC
CCTGGTTCTTGCCAGCAAGTGATGTAACCTCTCCAGGTCTCAGTTTGCACATAACAAAATATGAAAAGCCTGAGTAA
GACACTGCGCAAGTAAAGCGGTGACCACATCGGATGCTCACAGACGGCTCTAACTGGACTGCCACGGTTCAAGGCAA
GGACTCCCACCTTGTAGGGGATCCAGAACCCAGATGTGCTTTAGCAGCAGGAAGAGCTACTCTCAGCCGAGTCTCTC
AGATTTCCTCCCTTATCCCAAATCATACCACAAACTTCCTAAAACTGACTCCGAGCACACTAGTGCTGGAAAACCTT
CCATGGCTCCCCTGTGCCCTACAGAACAAAAGCTGCACCCTTAACCGGCCATTTAGGACAGGAACTAGGGTGAGGCA
AGTGAGACGCGTAGGACCCAACATTTAAGGAGGCACAGGTGGGCCATAGAGGACTTCTGAGCACTAGCATCCCAAGA
GTGAGCTTCCCCTTAAATTTTTTTTGCACTCTACGCTCTCACTCTGGTCCCAGTTCTGCTGCCATTAAAGGCCCTTC
ACAGTCTTGCCACTGTACACACAGAAAGTGCCAGCAAAACTGCACTGCTTGCCTTTCCCAAGCTTTCCAGGAGCTGA
AATCGCCTTTATTCCTCTGTCACCTACTGAGTATTCTGACATGCAGCTTAAGCGCCACTTCCTCCAGGAAGCCCTCC
AAGTCGAACTTTCAACTCCTGGCCTCTCTTCCTCCCCGGTCCTACTGGACTGTGGGGCCCTCCGGGCTTCACTTTAC
CCACCCCTGGTACAGCGCTGGGGCACACAGAGATGCTTAATGAATGCCTGCTCTATGCAATGAATGTCTGTGACCTG
TAAAACGGAATGAAGACAGAGTGAAGTGGAAGATTAAAAGGCAGCTTCTGTACCTTTTCCTCCCCGCCCCAAGTGCC
TGGCCTCAACTATCAATACATAAATCAGTAACTTTCCCTTGCTGAGAGAAGGGAACAGCAACAGTGTCGGTAAAGCC
CTTCCTCGTTCTTTCTCTGGAGGGGAGTTTCCACACTTCGCTGTCTCTTCTATTAGCATTCCCAACATTCTGGGTTC
CTCGGAGCAAGGAGTGAGCCTCCCGTTAACACCAGAATCCAAATTCTAGTCATGGTTAGAAGGACTCYGAGAACTTG
GGTTGGTGCCCGGGACTTGCAGGCGCTTTGGGGGCCCTATTTGAGCGGCAGTGTCCTCTGGGCTTGCCGCCACCGCC
CCCATTCCGAGTTTCTTGGATCCCAGATGAGCGATCGTCCCGGAGCCAGCCGGGCCATGAGGGCCGCTGATCGCCGT
GGCTGAGACTCCCGGGTCCCGGGGCTCTCCCAGGAAGCCGCGGAGGCGTGGCTCAAGCGAGTGGGCGAGGAAACAGC
TGCAGGCGTAGCGCTGGGAAGACCCACCCGGGGTCCGGGGTCCACGGCCCGCGGCCCGGACCCCCCGCCCGCCCGAT
CACGTGGCGGCCTGCAGGAAGGAGGCGCCCGGCTGCGCGCCGGAGCCGCCCGCCGCTCCCGGGTGCCGCCTCCGCGC
CGCAGCCTCCATGGCTAACCTTCCGGCCAGGCCCTTCGCCCCGCTGCCGGTAGCTGCAAACAGGTTCCGCTGGCCGA
AGGCTGCGAAAGCGAACATCCAGCGCGGGGGCGGGGCTGGGGCGGGCGAGAGCGAGAGCGCCCGCACAGCGGCGCGG
GCCGGGGGCGGAGCCGGCGGGGGGCCGTGACCAGGAGACGCGCCGCAGAGCGAGGGAGGCGGGGGCGAGCTCCCCGG
CCGGGGAGGGGCGGAGGGGCGGAGGGGCGGAGGGGCGGAGGGGCGGAGGGGCGGAGGGGCGGAGGGAGG
GCGCACGGCGCGGGGCCTCCTCGGGCTGGGATCTGCTGCGCTCCCGCGTGTGGCCCGGGGCAGGAATCTGCCTAGGC
TCCCCGAAACCTGTATAGTAGTCAAGGGTGGGCTAGGGACGTGGGGAAAGGCCAGTTCGACCTAAATACTCTCTTGG
GAAGGAGCCGAGACTGGATTTCCCTTGTCCTTGCGGACCTGAAGGCTTTGGGGGAGGGGCTCAKGTCCCTTCCAGCT
GTGTTAGCCACCTTCCTCAGATGGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAAGTCGCTTCAGTTG
TGTCCGACTCTGTGCGACCCCATAGTCGGCAGCCCACAAGGCTCCCCAGTTCCTGGGATTCTCCAGGCAAGAACACT

GGAGTAGCTAGCACTCCTTAAAATGTCAGCTCTGTGGCCTTTGACGCTGGGAATTAAAAACCAGTGGGATAAGGGTA
GCACGCGTGAAAGAGCGAATTCTTTTCAAGATTCCCATTCTGGCTTTGTCACTCTCTTGGACTCTGTGGCTTCATAA
AAGTTACTTCCCTTCTCTTCCATTCAAACATGTACTATAACTCCTGAGCAGTTCCTCATCTTCGAGACTGGATGACC
AGTACCCAGGAAGACTTAGGTGGGGACGAAACGGAGCAGGATCTTCTGGTCCTTCCCCCCGGCCCCCCTCCCCCCCA
ACACACACACCGTATCTTCTGCCTGCCTCTGACCTGTGGAAAACTGTAGTCTAAAACGAGTTTAATCACAGAAGTGA
GAAATGCTGAAACAAAGGAAAACCGTCAAAAGAGACCAGATAATGTTCTCATTAAGCGTAGTCAAGGACCTTTAATT
TCTTCTACAGGGCTATAGATAATGTTTGGAGCCATATTCTGTGAGCTGTCTTATAGATACTAAAACAGAGGTGGAGA
GGTTAACTACGTGATGACCAGGCTGAACCCAGGACGTGAGCTGCCACAATTCCGAGAACTGACCACAAAGAAATGGG
AACAAATGGACCCTGGAACTGAAGATTACTTGTACCTAAAACAATCAAGATGATACTGGTCAGACCACTGATGACCA
ATTTGAAGATGATTGTCAGAGATGACTTTGCTATTTCTGCATGTAGGCCCTGCTCCTCAGTCAGCCAGTTCAGTCAC
TCAGTCATGTCCGACTCTTTGTGACCCCACGAACTACAGCAGGCCAGGCTTCCCTGTCCATCACCAACTCCTGGAGC
TTGCTCAAACTCATGTCCGTCTCATCCTCTGTCATCCCCTTCTCCTGCCTTCAGTCTTTCCCAGCATCAGGGTCTTT
TTCAATGAGTCAGTTCTTTGCATCAGGTGGCTAAAGTAATGGAGCTCCAGCTTCAGCATCAGTCCTTCCAATGAATA
TTCAGGACGGATTTCCTTTAGGATTGACTGGTTGGATCTCCTTGCAGTCCAAGGGACTCTCAAGAGTCTTCTCCAAC
ACCACAGTTCAAAAGCATCAATTCTTTAGTGCTTAGCTTTCTTTATAGTCCAACTCTCACATCCATACATGCCTACT
GGAAAAGTCATAGCTTTGACTAGATGGACCTTTGTCAGCAAAAGTGCAGTATTCCTCCTCACTCTGTCTATAAAAAC
TCTCACCTCCTGCTTATGGTGGGAGGAGGGCGGAGTCGGCCTTTGGATGAATGTCCTCTCCCCTCCCTGCCCCAGTT
GCCGGCATCTGAAATAAAGAAACTTTCCTTTCCACCAACCTGGCCTGTTTATTGGCTTCTGACTGGCAAGAAACCAG
ACCCCACACACATACCTTTCAGTAACAGGGGGACTGGTGAGTGGCGGGGGGGAGGGGGGGGCGCGGTGGAGGTGGGG
AAGATACTGAATACCGTGAGGTTAGAGACTTGGAGGGGGTGGTATCAGACAAGGCTTCTGGAGGAGAGCTTAGGGAA
CTGAACTCCATTTGCTGAGGTGGAGCACAGCCAATTTACTGAGGAGTTGTGGTTGCTGAAAAAAAGCACAACCTAAA
AGTTGAGAATTATGTTTTATTTGGGGACATTACTGAGTGCTGTTGACCAGTATACAACCTCTCAGATAGCTCTGACG
AATTGTTCCAAAGAGATAAGGGTGAAGCCAGGATGTTCTGTTCAGTTGCTAAGTCATGTCCAACTCTTTGCAGCCCC
AAGGACTGCAGCAATCCATGTTTGCTCAAATTCATGTCCATTGAGTTGGTAATGCTATCTAACCATCTTATCCTCTG
TTGCCCCCTTCTCCTTTTGCCTTTAATCATCCCCAGCATCAGGGTCTTTTCCAATAAGTTGGCTCTTCACATCAGGT
GGCCAAAGTACTGAAGCTTCAGCTTCAGCACCAGTCTTTAAAATAATATTCAGGGTTGATTTCCTTCAGGACTGACT
GGTTTGATCTTCTTGCTGTCCAAGGGAGTCTCAACAGTTTTCTGCAGCACCACAATTCAAAAGCATCAATTTTTCAG
TGCTCAGCCTTCTTTATGGTCCAGCTGATCTTGGCGAGGCAGCAGTGAGTGTGGTGTGAAGTGAGATCTTAATTCCC
TGCCCAGGAACTGAACATGAGTGGCCTGGATGAGAACCAGGAATCCTAGCCACCAGACCTGCAAGGGCTAGAGCCTA
GAAACTATTTTCCCCTGGATCTTTGCCCCCAGTGAAAAAATGCATTTATCACAGAGGCAAGAACTATAAATGCAGCC
TAAGAGCAAGTGGGAGAGCACACAGAGAAATGGTTTGTTTAGTTAAAACAGAAGACAGGGAGAGATGCACACCCAGA
GAGAGAGGGTGTGGGTGTCCTCCCTAGTGAGGAGGAATGAAGCAAAGAGGTTATATAGCTCAGTTTTTCCAGGTCTT
TGTTTCCATTCAGTCCAATTATGTGGTTTCTTTTTCCTCACCTGACCTACCTTGGGACCCTCCCTTGGGCACACAGG
CACCCCTCAGCCAAGATGGATTTCCTGAGAGGAGCAAGGCTCATTATGGCCTGGTGTTATTCTCTGCCTTTTGGCCC
ACAAGGAGCCTTTCTGAGGAGGTGTAGTGTCTCCCTTGTCCCAAAAGAGGCGGGGGAGGGAGATCCCTTAATCCTTT
ACTGAAACAGGGTTTTGCTCCTCTTTGTCCTTGCCCTGACTATTACCTGAAGGTGCTTACAAGAGAGAAACACTGGC
CTATTTACCCTGTTTATGCTGTTCCTTGCATTTCAGAGGGCAAACAGAAGTCTGATGTCTTAACTGGAGCCCACCCA
TCTCTTGTCTCAGAAAATGCTAACAGTTCTAAGTACCCAGCCTGAAGCCTACTTCTTGGTGCCCCATGAAATGCAAA
CAGGAGGCCAGCTGTAAATATCTAACCTGGAACCCATCTATCTCCTACATCACAACTCTCACATCCATACATGACTA
CTAGAAAAACCATAGTTTTGACTTTACAGACTTTCATCAGCCAAGTTATGTCTCTACTTTTTAGAATGCTGTCTAGA
TTTCTCATAGCTTTTCCCAAGGGGCAAGCGTCACTGCTGCAGTGATTTGGGAGCCCAAGAAAATAAAATCTGCCACC
GTTTCCATTCTTTCCCCATCTGTTTGCCATGAAATGATGGGACCAGATGCCATGATCTTAGTTTTCTAAATGTTGAG
TTTTAAAGCCAGCTTTTTCACTCTCCTCTTTCACCCTCAACAAGAGGCTCTTTAGTTCCTCTTCACTTTCTGCCATT
AGGAGCCAGGATACATAGGAGTTTTTGCTGCAAAAAAAATGTAGTCAAACATCAGAAGATAACTGCTAGTCACAAAA
AACCAGATATCTCAAGTTAATGATTTTAGTGTGTGGGAAGATGCAAGAGTCTGGGCTCACTGACCTTTGATATGCAT
CTTAATTATCTAGGGCCAGTATCTTGTTTCTCTCCATCCAGATGCCCCTCAGGGTGCACATGGAGGGCAGCTGCAGT
GGCTGATGGCTTGATGGTGGGCAAGAATTTTTTTTGTTGTTCACTTGGTGAAGGAAAATTCAGCATTTATTGCAGGT
TCTAAGCAAGAAAGTGGGAGACAAACCTCAGATCTACCTCACCTTAGTCTATGAGTTAGGGTTTTTAAAGAGGAGGA
ATGAAGAGGCTGGGATGGACCATTGTCTTGCGACATTCGTTAATTGTTGTTTCAGGAGCCAAAGTATCTCTGATTAT
GATTCTTCAACCAGATGGTCCCATAGTAAAGGGGGGGGGGGACGTCTGTTAGCTCAACTTGCCCTGGAGAAACAACA
TGAGTTTTTATGTTAATGATGATATCTCTAATAGCAACTTTAGTACATTAACAATGTTATCAACAGCAATTTCAATC
ATCTGACTTTGGCTTACTAGTGTTCAGTTAGCACAGGACAGAAGTCAGAGATGATGAGAAAGGGAATAAAGTTTTGG
ATAGAGATATTAATCATGGACTCGGTTTAGAGGGACTTGGTTGCAGGTGGACATCCTAGTTGTGACTTGGAACTTCT
GAATGAACAATGAAGTCCAAGCAGAAGTCTACCAAACACATAAGGGAAGAGTGTTCCAGGCAGAGGGAATAGCCCAC
CAGGAGGCCTGGCATTCAGACAGACCCTAATGTTTGGACTAACAGAAAGTAGCTTAGTGAAAACAGCAGACAGCAGG
GAGGAAAAGCAAGAGGAGCTGACCTGGGAAAGGCAACAGAATTAAAGAGGAGAGCCTTTATCCTGGGGACAGTGGGC
AGCAGTTCAGAGCTATGAGGCAGGGGAGCACCGTCATATGTGAGTTTTAGACCTCCCAGCATGCAGGGCAGATGGAT

GCAAGGAGGCAGGATAAGAGACCAGGAGATGGAGCATGTGGGGCTCAGTTGTTCTAGTGAGAGTCCTGGACTGGTTA
CATACTGGAAGGAGAAAATTGATAACAGCAGGAGTTGTGTGGGATGGAACTTCCCTGGTTGTCCAGCAGCTAAGACT
CTGTGCCCTCAATGCAGGGGGCTGGGGTTCAACTCCTGGTCAGCGAACTAGATCCCCACGCTGCAACTCAAGATCCA
GCATGCTACAATGAAGACTGGAGATCCTGCAGCTAATACCTGGTGCAGACACACACAAAGAAAGAGTTGTGTGGGAG
GGACCAAGGCTGTGACTTGGATGTCAGGGGTGAGGGGGGAAGAGCTCAGGCAGTGGCCAGGTTCTGCAGCAGGGTGG
ATGGTGGAGGTGCCCCTTGGGAAGTGGGTTAGAGTAGGTTTAGGCTAAATGAGAATTGATCCTTGGACACATTAAAC
TTGAGGTGGGTTAAGAAGGCTGATGGACCCATGGATCTGAACCTTAGGTCAGACCCGGGAGCTGGGCAGACAGACTT
GTGAGTCACAGGAATATAGATGACAATTAGAACCTTAGGGGCTGCCTGGGGAGACAGTTTCCCTGTGAGAAAAGAGA
ACAGAGGGCAAAGCCTGAGAACAGACAACAAGGGCCTACTGTACAGCACAGGGAACTATATTAACCATAGTAGGAAA
TAGAAAAGAACACAAAATAGAACAACCATCTTGTAATAAATCACAACGGAAAAGAATATGAAAAAGAATATATATAT
ATATGTATATACACATGCATGAGAACTTCCCAGTGGCTCAGCAATAAAGAATCCGCCTGCAATGCATGAGGCACAAG
AGACACGGGTTCAATCGCTGGGTCGGGAAGATCCCCTGGAAAAGGAAATAGTAACCTACTCCAGTATTCTTGCCTGG
AAAATTCCACGGACAAAGGAGTCTAGCAGGCCTACAGTTTAACATGGTCTCAAAGAGTTGGACACGACTGAACATGC
TTGCATATATATATATTTTATATCTACACACATACATGTATACATAACTGAATCACTTTGCTGTACACCAGAAACTA
ACACAACATTGTAATCATCATATACTTCAATTAAAAAAAAGTAAGGTAGGAGAATAATTTACAAAGCAGACTGCCTG
GATTTCAATCCTGGTGTCGTTGCTAACCAGCTATGACCCTTGGGTAAGCTGCTCAACTTCTCTCAGCCTCAGTTTAC
TCATCTGTAACAAGCCTGTTGCAACCACTGCATGGATGATTCAATGAAACCACACACAYGAAGCGCTTAACATAGAG
CCTGGCACTGGTGGGTTTGCTGTGTACAACAGTCTGTCTTTGGACCTGAGGTGTGCCTCCATGATGTATCTTCTTCT
GTCCTATTTTCTTTTCTATCTCTTCTCTTGAGAGACTTTTGTTGGGTTTATTCCTTATCCCTTTGATGAGAGCCCAA
ATATCCCATCTTGGCAATGGCTGGTACGTGCATCTGGGCCACTTCCCATCAACTCCATCCTGGGCTCCAGCAGCCCG
CAGGCCACACCCCAACACCCATCCCGGCCTCACTAAAGGGCACATTTGCTTTATAGCTCCCTTTCTCTCATTTGCTC
CTCACCATGTCATCAACAAGCAGAGGATGTCCTTGCTTTGCACCAGGGAAAGCCAGTATATCTGAGGTCAGCCAACA
ACAACACTCAGACTGGATCCCAGATTTAATTTTGAGTTTGATCTTCCATCTACAGATTGTTCTTTGCCTACACCAGA
GAAGAAGGATTGAGGCTTCCCACTCTGACCCCAATCTCATTCTCTGCTGCCTTTTTGAGCCTGCTGGGCTCCGAAGG
AAGGGACATACAGTGCCCTGTGATGTAATTATCATTTAATAATACCCAGAAGTGGGTGGCCAGTCATCAGAGGATGA
AGTTGTGCAGTGGGGGCCTCTCTGGATCACTTAATCCTCCAATCTCTTAAATACCAACAAATCTGCCTCCTGGCTGG
GGCTGACAGCTGCCCCTGAGTGGAGGGGAGTCCAAGCTGGCAGAGAGGCCAGGACTTTCTGAATGAACATAATGAAA
ACTCCCCTAGTCTGGTGGGATTATCAGCCCCCTCTCCCTTTCCTGTCCCATCAGGACTTTGAAAACACCATCTGCGT
TAGCTCTCAAATTGCATGGAAACGGGGAGTGAGGCTGGGGTGGGGGATAGAGCAAAGCTGGCACCAGTTCTTGGGCA
CTGAGGGGCACACATTTCCAGGGATCTTGCCTACAACTCTCTGGAGCCATGGTTCCTGCCCACCAGAGAGGCAAATG
GTCTCAGCTGCAATTTCAGTGAATAGCATTAAGGAATTTTGCAGCTGGGGAGGTCCTTTGGAATGGTCCAGTGGTTT
CAACTTTTATTTTAGCAGATGTTTTCACTCTTTTCCAACAAGAGCAGTATGAGCTCCTTTTAATTTTTACTTATTTT
TATTTTGGGGCTGTGCTGCGCGGCATATGGGATCTCAGTTCCCCCAACCAGTGATAGGACCTGTGCCCCCGTGCATT
GGGAGCACAAAGTCTTAATCACCAGTCCACCAGGGAAGTCCCATGAGCTCCTTGTTAATGCCTCACCTAGGACCCTC
TCAGCCACACCTGTCTTCCCGGTCTTGCCACTTGTTTCTAGCACAAGCCAATGGCAACCCCCGCCCCGCCCCCAAAC
TCTGGCCTCAGAGTACAGGAGTGTCAGACACTCCATGGGGTGAACCTTTGGCCAATACGGGCAGGGTTGGCTGATGA
ATTCCTTTCTCTTCCTCTTTCTGAGTGGACTGTCCTAAGCCGCACACAGTTTCTTTGAAGAAGATCCCACAAGATTG
AGCATCAGGGGAGATGTGTGGGGGCCAGCCAAGCCCTCCGGCCTCTCCCACTTCTCTTTCTCACCCGGCTTTTCCCT
CAGGTCTGGTTCTTCAGCGTCACACTCCCCCATAAAGTGTTGGCACGTACACTTCTGCCTCAAGATCTGCTTCTTTG
GAAACCGGAGATAAGATAACACTTGAGAAGGATGTTCAGTGTATAAAACAGATAGAAGTGAGGCCACTCTAAGTGAA
GAGAAACAGGGCCCCAGGTGTCATCTGTTTTTCTTCCCCCTGCCAGTTCCAAGGCGATGATCGAGGYACCCTCACAA
GGCTCCCCAAAGCTTTTGATCTCATTCAATTCTGCCATCCTGCCACTGAGGACGCATAGAGTTAGAAGCAAAGAGGA
CTAGACTCTAGGCTTCTCACTCCAAATCTGGCGAGACGGATCCTCATCTTTCTATCAAGTTTACATAAGGACTTCTG
AAAATATTTCTTTTCCCAAAACATCTTCCCACACTACTCTGCCCCCTACCCCCACCCCGAAGGCTTCCCCCACYCGG
TTTTATTGAGAAATAACTGACATACATCACTGCGCAAGCTCAGGCACACAGCATGATGGTTTGATTAAAGCATATTA
TGAAATGATTACAATAGGTTCAGCTAATATTCATCATCTCTCAGGAAGGCTTTTAAGTCCCCTGTTCGACTGTAGCC
ACCACCTGGGAGGCATGACACACGGAGTTCTAAACTCAGAAGCTGGGACACAGGAGGGCAGACAGAAATGGGGCAGG
AAAGGTGCAGTAATGCTGGGAGAAAGGCAGTCAGGAAAGGCCTGATTCAGGGACATCTGCTCGTGGGAGACAGGGGT
TACAGCATGAGGACAGACCCCTCGGGAAGAGGGACCACAGGAGATGGAGACTGGCTTTCTCTACCAGCTGGTGTCAC
CTGTAGGACACGCTCCTCCAGGCTGAAGTCATCCTTAAGGAAGCTTGAAGGGTGACCACACTGGGGGGAAGAGACAG
AAGGAGGTTATTATCTCCAATTATTTTCATTGAAAAATGTTTAATATAATTGCTTTACACTTTTGTGTTAGTTTCTG
CTGTACAGTGAAGTGAATCAGCTATATGTTCATATGTATCCCCCACCTCTTTGACCTCTCTCCCACCCCCTCATCTG
ACCCCTCTAGATCATCGCAGGTCACTGAGCTGGGCTTCCTGTGTTTTATAGCAGATGCCCACGAGTTACCTGTTTTA
CACATGGGAATGTATATATGTCAATCGTAATCTCCCACTTCGTCCCACTCTCTCCTTCCCTCACTGTGTCCACCTGT
CCTCCCTACGTCTGCATCTCTATTCCTGCCCTGCAAATGGTTTCGCCTGCACCATTTTTCTAGATTCCACATATATG
TGTTAGTAATACAATATCTCGTTTTCTCTTTCTGATTGACTTCACTCTGTATGACTAACTATGAGCTTCCCAGGTGG
TTCAGTGGTAAAGAATCTGCCTGCCAATGTAGGAGATGAAGGGAGATGCAGGTTCAATTCCTGGGTTGGGAAGATCC

CCTGGGGTGGAAATGGCCACCCATTCCAGTATTTTTGCCTGGAGAATCCCACTGACTGAAGAGCTGAGCGGGCTACA
GTCCATGGGGGCGAGACTGCACACACATGACGGACGCTAGGGCCATCTCATTAAGATTCGTGTGGCTGTGACTGCTA
GAAACACCTGGGGCCAGGTCAAAGGCTAAAAGTGATTCTATATTATAGGGGCACGGGCATCTTAAGGAATCTAGGTA
CCAGGGTGCATCTGGGCCTCGGGGAACAGATGAGACCAGGACCTGGACTCAGAGAACCCCGGAAGCCACCTCCCTCC
AACTCTGCTTCTGTTCCATCCCTTTCTCTCTCTCTGAAACTGGCGCTCCCTGTGCCCTAGCGAACACTGGCATCAAA
TGGCCTCAGCCAGCAACTTCCAAGTTCACCTCGCTTCCATCCAGGAGGCAAGCAGAGTCAGCCTGGACTCTTCTGGG
CCCAATGCTGAGCTCTCGTGGAGGTTCTGGTTGGACCAGCTTGAGGCAGGTGCCTCCCCCGGGCCCATCAGTGAGGC
CGGATGCTCTGCAGTCACAGCTGCTCCCTCGCTGGCCAGACAGGCCAGGGAAGAAGCAGGCGGTTCATTACCTCGGG
CTTCAACAAAAGCCTTAAAATAAGAGAGTCTGTGCTCTTCCATGCCTTAATGGAAGAGATCATTCTTTTCTACATTT
GTTGTAAACTCAAACCTTGTCTGGAARTCTGCCTCCCCAYTGGAGCCCAGAGCTGAATTATTCTCAGCCTCCCCACT
GCCGCACACCCCCCACCRCCACCCCGCATCTGCCACTGAGAGCTGGGCAAAGCTGAGAGGGTACCAAGGCCTGCCAC
TGTCCACATCCTGGCTCTTATGTGAATATTAGTAAAGTTTACTTCAGTTTCACCTTAGACTTATTTTATAAAACAGG
TCAGTGGGCTCAAAAGTCTTCCATCAAAAGAATTTCTAAGGACATGCTCTTGAGCCTATGGTGGAAAAGGCTGATGT
GAGCTTCTAAAACCAGAATTTATGTGGTGCTTCCTGCTTTGGGAGTCCATGGGTTCAATGCCAGAAAGGAACTCAGG
TCCTTTTCTTTCACTCTTAACSACTACTGATAACATCTATTGTACAATTGATATGGGCTAGGCCCTGGTCTCAGTCC
CTCCCTTCTACYATCACCAACCCCGACAACCACTGATTGTCTTTATGTCACCATAGTTTTGTATTTTCTTAAATTTC
ATATAAATGGAACCATCTAGCTTTTTCTATCCAGATTTTCCCACTGAGTATGATGCTTTTGAGATTCATTTGTATTT
GGGTACATTTCAAGAATTCAGCTCTTTTTATTAACAAGTAGAATGATATTGTCTGGATGTACCATGGTTTGTTAATC
CATTTGAGGGACATTTGAGTTGTTTCCACTTTGGGGCTACTGTGATAAAGCTGCTAAGAACATTTACATACAAGTCT
TTGTGTGGACTGATATCTTATTTCTCTTAGGTAAACACCTAGGAATGGGATCGCTGGCTTCAATGGTAATTGTACCT
TTAGCTTTATAAGAAACAGCAGAGTCTTCCCTGGCAGTCCAGCATTGGCCTAGCCTTTCAGTGCAGGGGGTGTGGGT
TTGATCCCTGGTCAGGGAGCTAGGATTCCACATGCCTCGCAGCTAGGAAACCAAACCATAAAACAGAAGCAATACCG
TAGCAAATTCAATAAAGACTGAAAATAGTTGCAAATTTAAAAAAAAAAATCTTAAAAGGAAAAACTGCAAACTGGAA
ATAGAAATACCATTTCAAATGATCTTTAATCTTAAAAAACTTCAGATGCTCAGCTTAGCATGTATACATCCTCAGTC
GCTAAGTCATGTTCTGCTCTTCCTGAATCCATGGACTGTAACCCTGCACGCTCTGTCCATGGAATTCTCCAGGCGAG
AATACTAGAAAGGGTTGCCATTTGCTACTCCAGGGGATCTTCCTGACCCAGAGATCAAACCCGTGTCTCTTGCGCCC
TCCACGCTGACAGGCAGATTCTGCGCCACTATGCCACCTGGCCGCTTGGGAAACCCACTGTGCCTAGCAGTAGGCCT
CAAAAAAGACTACCTTCTGGACAACGGCAGAGGTAACCACGACAGTATCACCCTGATGACAAATCAAGTTGTACGCT
TCGCTCCCATGGCCTGGAGTCTCCTCACCCCTGCGATGTGAGAGGGAAGTTCTAAACTGGGTCATTCACCCCCATCC
GATGTCTCATCTTCCCCTGCCTTTCCATCATCTTCCTGTGCAGGGCTCTGGCCTGAGAGCCACATGCAGAGGTAGGT
ACAGGCAGTCTATACATAATGTATATATTGTCTAGGAATTAAGTTTCCACCCAAAGCAAAAAGCCCTAGCTGCCTGA
CTTTTAATTAACTATTCATTAGTTTGGCTACAATATCCTCCATCAAACAATGGGCAAGTCATAAGGCTGAATAGACC
CAGCAGACAAAGCGGGATCCCTTTTCTTAGCTGGATGTCTAACCAGAGCCAATTCTCAGTTACCCAATCAGTCAATA
AGCATTTGCTGAACTGTACCGGAATAGTGGACATGGAACCGATATTTCTCATAACCTTACAGTTCTTGCCCACCCAC
TTTGGGTACTGTCATTACTTTTCGACAGATGATTCAACAGAAACTTGAAGATGTGATGTGAGGTTACCTAGCAGAAG
CCAGGATGAATTCCAGGTCTCCCAGCTTTGCCTTCCCTTCCCTGCGTAACCCATAAGGTAGTTGGGTTTGGGGTCGG
GGTGGGGAGAGGAGGCTAAAACACACGAAACAAGACAATCCAATTTTTTCCCCAGGACATTCCTGAAAAAAATTGAC
CACAGCATTAAAGTGTCTTTTCCCCATAAAACACAGTTTGAGTTGGGAATTTGTTCCTGATTAAGGCTTGGACATAA
ATTAAACGATGAAAATGTAAAATGAAAATAAGTAAAGTTATTACAAACAAATCTTTATAATAACTTGTTAGTAAAAC
CCTGCTCAGTCCTTATGAAACAGCCACAGAGCCACTGTAGACCTTAATTACACAAGGAGCCCTCTTCCACTGCAAAT
TTTCTTGTGCAACAAGAAAAACTTTATGTCTTTAATATGATTTAAAACTAATATGCTAGAAATGGTGAGCAGTGAAC
ATATAGTTAGAAATGCTTTTATAATGTTTAGAACCGCACTGGGAGAAAAATGCCAACTGCTTTGCCCTCTTTATAAA
AGCTGGACTAACATTTTGTATGGATCTTATTTCTCTTTTGTCAAGTTCTTACTCTCCATGGACAGGGTCCACAGATG
GGCATCAGGGGACTTTGCACCACCTGAAATTGTATGCAACATGTTTTGAGGATACAAATACATTGTGGAGCCTCTAG
CTTTCATCAGATCCTCACAAGCCTCTGTGATAGAGCCACAAATTTGAATCACTTATTTCTTAGTGTTTCTTCTATTG
TGAGCAGTGGCATTAAAATTGTTCCACTGACTTTAGAACCTTGCTGACTTGGAAACATTCTTTTAAAAAGCAGCCTG
GGGACTTCCCTGGTGGTCCAGTGATTAAGAAACCACTTTCCAATACAGGGGATGTGGGTTCGACCCCTGGGTCGGGA
AGATCCTCTGGAGAAGGGCATGGCAACCCACTCCAGTATTCTTGCCAGGAAAATCCCATGAAGAGAGGAGCTTGGCA
GGCTACAGTCCATGGGGTTGCCAAGAATCGAACATGACTGAAGTGACTTAACACCCACATGCCACAAGTCACAATGA
AAGACCCTGCTTGCTGCAACAAAAACCTGAGGCAGTCAAATAAATACATATTTAAAAATAAAAAGTAGCCTGGATCC
TCTTGGTGGTGTAATGGGCTTGATTCCTAGAAGGGATCCCTCCCTATAAAACAACTGGACCCTGAAAAAGAAAAAAA
TTTTTAAATGCATAGCTAGTGAAGTCAGATAGTGAAGGAAGTCATCGTTTCCTCACCACTAAAATTAAACAACACAC
AAAAATCATGGGCCTGAGACTGGAGGAAGGGGCACTGGTTAGAAAGCGAAGATTGGGCAGCCCTAAGGGCAAATGCC
CATTTTAGCTGCTTCTGGGAAGCTTGCCCCTGGGCCAGTGGCAACATCAGGGATCTGCCTCAGGATTCCACCTTAGC
TCAGTTTAAGGGATCCTCTTTTGTTACTGGCTGCACCTCTATGGATTTCCCTGGTGGCTCAGATGGCAAAGCATCTG
CCTACAGTGCGGGAGACCCGCGTTCGATCCCTGGGTCGGGAAGATTCTCTGGAGAAGGAAATGGCAACCCACTTCAG
TACTCTTGCCTGGAAAATCCCATGGACGGAGGAGCATGGTAAGCCCGAAAGGAGTCGGACACGACTGAGCGGACTTC

ACTTCCACTTTCCACCTCGGAGGTTTTCTCTGCCTCACTGAGGAGAAGCGGAAACGACATTTCCCAGAACCCTCTGC
TTCCGCAGCGGCTGCAGGCGGGAGGCCATTGGTAGAGATTCCGAAGGCAGGTGCGAGGGAGAGGGCACAGTTCTCCA
GGGGAGTTGTAGACTCGTGAAAAGGCGTGAGTTCGTGGTAGCTTCAGGATGTGCTTCTTGAGGCTCCCCTGCAACAT
GCTGCAGCCTGAGAGAGCGAGGGGTGGGGGCGAGGGCCGCTTCCCAGAGGCTCTTGAGATTAGGGTGGCCTTCAGGC
AAGCCTGAGAACCGCCCTCGTGGGTGCTACAGACTGAAGCTGCAGTTGAAGGTCCCCTTACCTTTGCTTCCCCAAGG
CTTCCGAGGGTTATGTAAACCCCTGATTTCCTGTATCCAAAACTGAGACGCTTGGCATGCCAGCTATCACTTTATTC
TCTCTTGTTCCAAATTTACTTTTCAATACCTGCTCTGTGATCATGGGTGAGCCTGACGTTGTTTTTCAGTAGATGGC
ACTAGAAGGACACTGGAGGAGGAAAGGTTTCTCCTTCTGGTACTGGTGCTGTACGTTGGAGGGTGGTATGAACGGGA
GCGATTCGGGTGGTGCGCTGCCCACACACACACCAAGAACACACAATCTCTCAGTGACCTCCCAGCTCCTGCTTGGT
CTGGTGATCACCTCCCACGGCGCTCAGACACAGACAGTGGATTGCAGGCCTCCCTCCTACAGTGATGCTAAATTCCC
TCTGCACACCTGCATGCCTGACAGCCTGCGCCCTGGGCACTGCTCACCTGTGCTTCACTGGGTTGCTTCTGGCCTTT
TCAGACATTGAGGACAGGCTGCCCATCTGGCCCAGGCTACCTGTCAAACTTCTCTGCCATACAGCGGTCTGCAACCA
CATCTTCTCCAGTGCGGTCTGAACCCCAGCTTTGGGAAGGGAGCTCCCCTTCATAGTTTGTCCGTCCTTAGGTACTT
TCTCTCAGCACTATACTCCCCCCCTTTCTAAAATAATGGTCCTGTTTTATTTTGCTTATATATTGTTAATATTTTAT
ATATTTATTTGGCTGTGCCAGGTCTTAGTTATGGCACGCAGGATCTTTGATTTTCGTTAATGGCATGTGGGATTGTT
GCAGCGAAAAGGAAAACAGGGGAATGAGTTTTTCTCTTTCCCTTTCCTGAGAACAAAGAAGACAGAGAGAACAGTGA
GCAGGCCTGAACATTCCCTAGTTTTTTTTTTTTTCTCCTGTTTTTTTTTTTTCTTTTTTCACTTTACCTGCTCCTAA
TTCTGAATGTCTGTAACTAAGTTTGCTTTCTGCCCTCTAACTCTACAGGAGCTACAAGTTACATTTTTTTCCTTTGT
TGCTGTTCAGTCGCTCAGTCATGTCTGACTTTGCAACCCCATGGATTGATTGCAGCACGTCAGGCTTCTCTGTCCTT
CACCATCTCCTGCAGTTTGTTCAAACTCATGTCCATTGAGTGAGTGATGTCATCCAACCATCTCGTCCTCCGTCATC
CCCTTTTCCTCCTGCCCTCAATCTTTCCCAGCATCAAGGTCTTTTCAAGTCAGTTCTTTGCATCAGGTGGCCAAAGT
ATTGCAGCTTCAGCTTCGGCATCAGTCCTTCCAATGAATATTCAGGGTTGATTTCCTTTAGGATGGACTGGTTTGAT
TTCTCAAGAGTCTTCTCCAACACTACAGTTTGAAAGCATCAGTTCTTTGACACTCAGACTTCTTTAGGGTCCAACTC
TCACATCCATACGTGACACCTGGAAAAACCATAGCATTGACTATATGGACCTTTGTCAGAAAAGTGAAGTCTCTGCT
TTTTAATATAGTGTCTATGTTTGTCAGCTTTTCTTCCAAAGAGTAAGCATCTTTTAATTTCATGACTGCAGTCACCA
TTGACAGTGGTTTTAGAGCCCAAGAAAATAAAATCTGTCACTGTTTTCCCATCTATTTGCCATGAAGTGATGGAACC
AGATGCCATGATCTTAGTTTTTTGAATGTTGAGTTTTGACCAGCTTTTTCACTCTCCTTTTTCACATTCATCAAAAG
GCTCTTTAGTTCTTCTTTGCTTTCTGCCATAAGGGTGGTGCCATCTGCATATCTGAGGTTGCTGATATTTTCCCTGG
CAATCTTGATTCCAGCTTGTGCTTTATCCAGCCCAACAGCTTGCATGATGTACTCTTCACATAAGTTAAATAAGCAG
GGTGACAATGTACAGCCTTGACATACTCCTTTTCCAATTTTGAACCAGTCTGTTGTTCCACGGTCGATTCTAACCGT
TACTTCTTGACCTGCACACAGGTTTCTCAGGAGGCACGTAAGGTATTCTGGAGTTATTTTTATATCTTAGTTACTCT
TTTAGCTTAGCTTAGTAACTCTTGATATTAAACATACTCTGTTTAAATCACTATGTGGTTTCCATTTTCTGACTGGA
CAAAGGTGGTACACTTGGAATACATAGAGTATTTGGTTTCTCTTTTTCCAAAGAGAGCCTGACTGATAGAGGCTAAC
CATAGCCCTGGGTCTGTAGGCATCTTGGCACTGCTGCTGCTGCTGCTGCTAAGTCGCTTCAGTCGTGTCTGACT
CTGTGTGACCCCATAGATGGCAGCCCACCAGGCTCCTCCGTCCCTGGGATTCTCCAGGCAAAAACACTGGAGTGGGT
TGCCATTTCCTTCTCCAATGCATGAAAGTGAAAAGTGAAAGTGAAGTCGCTCAGTCATGTGCGACTCTTCTCGACCC
CAAGGACTGTAGCCCACCAGGCTCCTCCCTCCATGGGACTTTCCAGGCAAGAGTACTGGAGTGGGGTGCCATTGCCT
TCTCAGATCTTGGCACTAAGCAGAAACAAATGCATTCTTCTCTGGGGTTTACTTTCCTGCTTAGATCCTTAAGACTC
ATTTGTTCATTTAACAGATTTGTTGAGCTCCTGTTATGTATCAGTCATTGTTCTAGATGATGGGGATAAACTAATGA
ACAGAAGAAAAACCCTACCTTATATTTTCACAGAAATATACCATCAAATATAGAAGGAAATATGCTACCATGAATAG
AAGACAAAATAGAACACATTTATCAATAGATCTCTACGGTTTTTAGATATTGCAATTTTCTAGATAGAAGAAGATAT
AAATAAACTAGAAGGGAAATGTTTAAAGAAATAAAAGGTGGGGACTTCCTTAGTGGTCYTGGTGGTTAAGAATCCAC
CTGCCAATGTAGGGGACATGGGTTTGATCCCTGGTTCGGGAAGATTCCACATGCTGCAGGGCAACTAAGGCCACGCA
CCACCAACTACTGAAGCCTGTGAGCTCTAGAGCCTGTGCTCCACAACAATGAGAGGCTGGCGCACCACAAGTTGAGA
GTAGCCCATGATAACCACAACTAGAGAGAGGCCACATGAGCAACAAGACCCAGCACAGCCAAAAATAGACAATCCCA
TTGCTAGTGGGATTGGAAAATTGTGCATCCCTACGGGAAAACCTTCTGGCAATTTCTCACAGTGTTAAATAGAGCTA
CTACATAACTTAGATATATTGGTATATAACCAAGAGAAAAGAAAATATATATTCACGCAAAAACAAGTACATAAATG
TTCATGGCAGCATTATTCATAATACTCCCAAAGTAGTAACAACCCAAATGTCCATCAACTGGTAAATAAATAAACAC
AATATGGTAGACCCATATAATGGAATATTATTCAGCCGCAAGAAGGAACAAAGTAATGATATTTGCTAAAATATGGA
TGAAAATTGGAAACATTTTGCTAGATAAAAAAAAGTTATAGAAGGTTATATATATGTTGTATGATTCCATTCATATG
AATTGTCTCCACTAGGCAAATCCAAAGAGAAAGAAAATAGATTAGTGATTTTGGGGAGACAAGGGTAGAAGGAAACT
GGCAGTTATTGCTAATGAGTATAGAGTTTCTTTTCAGGGAGATCGAATTTTTCAAGAATTCATGGTAATTATTTCAC
AACTTCTGTGAATGTGAATTGTACACTTTCAAAGGGTGAATCTTACTATTTGCAAATTATATCTCAGTAGAGCTGTT
ATTTTTTTAAATGAGCAAGGCAAAGAGACTATTGAAAATGAAGAGACACATCTGAAAAGGGAGCAAATAGAACTTAA
CAATTTAGTCATTGAGATTAAAAACTCAATAGACAGACTAAACAACAGATTATGCATAGCTGGACGATGAATTAATG
AACCGGAAGATAGATCTGAGGAAATGACCAGAATGCAGTGCAGAGAGTCAAGGAGATGAGGAATATCAGTGAGAGGT
TAAGGAAAGAACAGATGGAATGAGACGGTCTAAAATATATCCAATCAGAATCCCAGAAGGAGAGAATAGAGAGAATA

AAGAAGAAGCAATGTTTGAAAAGCTAATGGCTAAGAATTTCCAGATCTGATAAAAGACACCAACACACGGACAGGAG
AAAAAGAACATGCATCCAGCAGGTTAAAAAAAGACATACACATGGACATATTGCAGAGAAAGCACAGAGCATTACAG
GCAACATTGTTAAAAGCAGCCAGAGAGAAAAGAAAGACTGCCTACAAAGAAACCACAAATCGGACAGTAACAGTAAC
AGTGAATCTAAAGGACAGAAAGGTATCTTCCAGGTACAGAGAGAAAATAACTGCTAACCCAGATAACAGTAAAATAA
ACTATTTATTTAAATAAACAGTTAAATAGCAGTTAAAATAACTGTAAAACACTGCTAATGAAAGATGGTGATGACAT
GTAATGTGTGGGGTTGAAAAGGACAGGCACAAACACTGGACAATAATAATTTATAGGTGGGGAAAGAGAGGTAAGCC
TGTGCACCACAACTACTGCCCCCCAGCTCTGGAGCCCTTAAGCCAAAACTACTGAAGCCCACCTGCTCTAGAGCCTG
CCTGTGCTCCGCAACAGAAGAAGCCACGGAGAGCCCACCCACTGCAGCTACAGAATAGCCCCTGCTCTCCACAGCTA
GAGAAAGCCCTCGCACAGCAGCGAAGACCCAGCGCAGCCAATCAAAGAAAAAAGGTGGGGAGGGGGGAACCACAAGT
ATAAAAGGGACAGACAGTTAAGAGAGCTCAGATCTAGATTTTTTAAAAATCTAGACACACCTAAAATTAAAAGACAG
AGAAAGGATGAACGTAAAATGATGAGCAAGAGCAAAAGCTCATTAAAAGAAAGCTGGTGTGATTATATTGATATTAG
ATAAAATAGACTATAAAGCAAAAATGTATTTATCAGTTTAATGATACAAGGTTTAATCCATCGGGAAAATATAATGG
TTCTAAAGTTGTGGGTGCCTACTAAAATAATTTACAAATACACAAGATAAAATGTACAGAACTATGGATAGAAATTG
ACTAATCCACTAACATAGTAGGGGATTTCAATACAGAGCTCTCAATTATTAATGGGTAAAAAAAAGATTGGAATAAC
ACAATCAACCAGCTTGATTTAGTGGATATTTATAAATCACTCTACATCCAACAATCAAGGAACAGACATTCTTTTCC
AGTGCACATGGAACATTTCCAAAAATTGACTATATAAAAGGCCATAAAGCCGGTTTCAACAAATTTCAGAGAATTTG
ATTCACATAAATTATATTTTATGACCTATGATCATGGATATAGCTTATGCAATTAAGTTAAAAGTCTTTAACAGAAA
GATAAATGACAGCGTCCTCTAAGTTTTGAAAATTGGAAAGTACATTTCTATATAATTTATGAGTCAAGTAAGATATA
ATAGAAATTAAAAGTTGTATAGAAACAACTATATGGTACATGATGAAAATTGTAGATATCAAAACTGGTGGGTGTAG
CAAAAATGGTAATAACAGGGAAATTTATTGCTTTAGATATTTACAAGAAATGAAACAAATATACAATAGAGATGATC
AACAAGCCAAGGCCAGATAAAATAGACAAACTTTTGGTAAGATTCATTACAAAAGAGGGTAGTAGGCAGCACGCATA
AACAAGTAATAATAGGAACAGAGAACTGAAGAAAGCTGCAAATACAGAAGAGATTAAACAGATAAAATGAGCATACT
ATAAAGGACATTACACCAATGAATTTAAAAACTTATTAAAAAGGGAAACATTTCTAGAAAAATGCAACTTATCAAAG
CTGTTACAAGAAAGCTTGAGTTTCCAACAGTATTAAGGAAATTAAATCAGTTTCAATGTCTTAAAAAACAACAACAA
CAACAAAAAAAACAAGCCTTGCCATTTTCCAGGCAATTCCTAGCAGATTTTCACAGAACAAATGATTCTAATTTCAC
ATACAGAATATCAAATAAAGCACTGCTCCTCAACTCATATGAGGTTAGTAAAACCTTGAAATAAAATCAATCAAGAA
CTGTATGAAAAAGAAAATTCAGACCAGAATCCCTCACAAACATAGACGGAAATATTATAAACAAACTACTAGCAAAT
CAAGTCCAACAGCTTATATTAAGGGCAATTCATCATAACTAAGCTGGATTTATCCCAAGAATATTTGGATCATTTAG
CATTAAAGGCCATAACTTTAATTCATCACATTAATGGGTTAAAGGAGAAGTTGTATAATTATCTAAATAGATGTAGT
AAAGGCAGTCAGTAAACGTTAACATCCATTCTTGATAAAAAAAAAAAAAAAAAGACAGGAAATAAAACAAAACTTCC
CAGCTAACATAATTGTATATATAGAAAGCCCTCCAAAGAAATCTGTGGATAAAATTACGACAATTGATCAGAGAGTT
TACCTAAACCAAACTTAGTTAAATTAACAATAGCACCAAAGTTAGCTAAATACTAGAAACTAGAGATATATATATAT
ATATACACACACACACACACACACACATATATTTCCAGATTAAGATCTTAAGCTTAAAATAAGAAAAAGACAAACCA
GTTAAAAATGGGCAAGATGTGAATAGGCACTTTATGAAAAAGGAAACATAAATCGCCCATTTAAATTAAACTATGCT
TGACCTCACTGACAGCCGGGAAAATACAAGTCAACGCCATGATATAACTGATTTTCCAGGCTTCCCTGATGGCTCAG
TGGCGAACAGTCCGCCTGCCAATGCAGAAGACATGGGTTTGATCCCTGATCTGGGAAGGTCCCACAGGCCACAGAGC
AACTAAGCCTGCGGGCCACAGCGACTGAGAACCCGGGAACCACAACTACTGAAGCCACGCGTCCTGGATCCTGTGCC
CTACAAGAGAGGCCCAGCCGTGACTAGAGAGTAGCCCCCACTGGTGGGAACTAACGCAAAGCCTGCACAGCAACAAA
GATCCAGCACAGCCAATAACTCCATCGGCAACATTTAAGACATTTAGTAACAACAAGTACGGCTGAAGATGAAAATC
AGTGGAACCTTTTATATACAACAAATGGAGTGTAAATTGGCACAACTCTGAAAGACACTTTGGTTATATATATGTAT
TTTGTAAAACTGAACGAGACATATTACAAAGAAATTCTTAAAAATGCATAAAAATATTCATTGCTGTATTGTTAAGA
TGGTAAAACATTGGAGGCTATGAAATGTCCATTGACAGGAGAATGGGTAAATGGATAAATAAATAGCTGTCTCTTTT
CACAGTGAATCAACTCCAGGTATACATGTAACATGGTCAAGTAAATGTATGAAAAAGGCAAGTTGCTGAAAACTACA
TAAAATCATGATGCCACTTTTATAAAGCTCAAATGAGCAAATTAGACAATGTTTCATTTAGGGATGCATACAGAAAT
GATAAAACAACCTTTTAAAGAAACCAAGAAAATGGTCAACATAAAATTCAAGCCAGTATTATCTCTGGTGAGGCAGG
ATAAGATAGGAAAAAGGCATACAGATGGATGGATACGTTATAGGTAATGTTCTAGTTCTTAAGTTGGATGGTGCATT
TCCAGGTGTTTATTTTCTGTTGCACTTATTATTTTGATATAGCAAATACTTCGTATTAAATATTTTTAACACAGCCT
GTAGTTTCCTATTCTTACCACTATATGGCAGGCTAGTCTGCCAGTAAACTTTTGTAACTTCTGAATGGGAAAGTGAA
AGTGTTTGTCACTCACTGGTGTCCTGACTTTTTGCGACCCCATGGACTATAGCCCACCAGGTTCCTCTGTCTGGGGA
ATTCTCCAGGCAAGAATACTGGAGTGGTTAGCCATTCCCTTCTCCAGGGGGGTGTTCCCGACCCAGGGATCAAACCT
GGGTCTCCTGCATTGCAGGGAGATTCTTTACCTGAATTCGAAAGGCTAGTAAAAATAAAACCCTACTTAATGACAGT
TTTCGCAGCACGCAATTTCCTCTATAGAAACACATGCAAAATGTAAAACAGGGTTTTTTGTTCTGAGTGTTTACAGG
CACAACCAGAGCAGCTCAGGTGCACAGTACTTTGACCTGGAGCTTGCAACTATGTGTCCAGGTTTTGTTTTTCCTTC
CTTTTTGCCTGTGTAGCCATTTTTGCAAACTTAAAAGTGTACCTTCTTAAATGTTAATTTTTATTTCTCAGAACTTT
CCAATATTTCCCAAGGTATTCATTTCTAAATTAACAAATGGAACATAAGCTTGTTTATTTGGCTTCCCTCTTGGCAA
TGGGGCTTCCCTGGTGGCTGAGTGGTAAGGAATTTGCTTGCTAATACAGGAGATGCAGTTCAATCCCTGGGTTAACG
GAAGAGCCTCTGGAGAGGAAAATGGCAACACTCCAGTATTCTTGCCTGGGCAATCCCATGGACAGGGGAGCCTGGTG

GGCTACAGTCCATGGGGTCGCAAGAGTCAGACACACTTATTGACTAAAACGACAACTTGGTAATTGACACATCTAAC
TTCCCCTTCTAATATTTATGGAATAGGAGTTGTTTCGTAGCATTAGCATTAGAACCACTGTTTTAATGAATGTCTGT
CCTGTAATATACAGGAATATTTTAAATTCAGTAACAGTGTTAAAATAGCGACTCAATATTCATGACAGTCATCAAAA
TCACTCACAGGAATTCTGCTGTCAAATTTCACCAGTCACAGGAAGTGATTGACACAAGTTGGTATTTCTGTTTTAGC
AGTTAATGTTATGCCACCCAGAGGCAACTGCACATAGTCTTCAGCCAGCCACTTCCTTTTTCATGCTCTAGACCTGC
ACTGCGCACTCTATCAGCCACTAGACACATGTGGCGATTGAGGTCTTGTAACGACATGTGCTGGAGAAGGCAATGGC
ACCCCACTCCAGTACTCTTGCCTGGAAAATCCCATGGACGGAGGAGCCTGGTAGGCTGCAGTCCATGGGGTCGCTAA
AAGTCGACACAACTGAGTGACTTCACTTTCACTTTTCACCTTCATGCATTGGAGAAGGAAATGGCAACCCACTCCAG
TGTTCTTGCCTGGAGAATCCCAGGGACGGGGGGTGCCTGGTGGGCTGCTGTCTATGGCACAGAGTCGGACACGACTG
AAGCGACTTAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCGAGAT
GTGCCCTAAGTGTAACATATACACTGAATTTTGAGGACTTCATATGAAAAAAACCTCAGTAATTTTAATAATGATTA
CTTGTTGAGATGATAATACATTTTGATGCCAAAAGCTCAGCTTCTCTGTACACTTGTGCCAAATCAAATCTTGCAGA
CAGTTTCGGGTGAAGCAGAAAAGGATAACTCTATTGCTTTGCCAGGCAAAGGGGGACACAGTGGGCTAAAGCCCTCA
AACCATGTGTCCTCACTTGGAGAAGACAGTGAGAAGTTTTATATGAACTGTTTGTATGTATTCCTTGATCGGAAACT
AGGACCTTGCCCAAAGGGTGCTCTTGACTGTTTCTCCCCCTCTGGCATCCCTTCTCTTCCCTAAATAACAGCTGCTT
GAATCTGACTGGAACTCAGGGGAGGTCATGGAGGCTGAATGAAAGGAAAATAAATATATTATTAAAATTACTTTCAT
CTGTTCTTTTTTACCTTTTTATTGGACACTAGAACATTTAAAATTACATAATGTGGCTCACATTTTATTTCTATTGG
GCAACACTGCTTTAGATGTCACAATACTTGTGGTCCCTACTTACCAGGGCCGGCTTCATGGGTGTGTAGCCAGCGCA
GTGGAACAGGGGCGTCCCATAGTGGGAAGTGCCGTGCATTTCACTTAATGCTCTGCCGGCACTGCCTTGATAACCTT
AATTGTGCCTCACGTGCCGTCACTTCAGTCGTTTCTGACCCTTAGTGACCCTGTAGCCTGTCAGCGCTCGGTCCTTG
GGATTCTCTAGACAAGAATACTGGACTGGCTTGCGTGCGATGCCCTCCTCTAACTGTATCTTAGAATTTACATTTCA
GAAGGAAATCTCCTAGAACAATGGAGCACGCGCCAGGGGCTAGAAGCCCTGGTTCCTGCGCCATCCTGCCTTCCGCG
GCCTCCGTGTCCCTGGCCCTGCTGCCCCCTTCCGACCCCGGCTGGGTTCTGGGCGCGGCTGTGAGACGGGGCGGGGC
TATAACTGTGGCCATCCCTTCCCCTGCCTGGCTGCCCAGGGTCACCCACTCACCTGGGTCCAGGGGCGAGAGCTTGA
CCGACGCTAAGGTCTCCGTACCGTTGGGAGTCGCTTGTCTGCCCCGATGTGGGCGGCGAGCCCTGGGGAAGGAGAGA
CACCTGGCCTGAAGTCGCTAGACTTCGCCCGGCGAGGGCGCGGCGCCGCGGTCAGCAGCTGGCAGGAGCTGGACCCA
GCAGCCCGGCGCCGGCCTTCTTTGCTGCCTGCCTGCGGGAGGGATGCGCCCTGCAGGGCCGTCCTGGAGAGAGAGTG
TCCGCTTGCCCCTGTAGCGCCGCGTACTCCGGCTCTGAGCTTGGGGGTGGATTGGGAGGACTCCTTTCCTCTTTCTA
ATCTTCCTCTGTCTTTCTAGCGTTTGCCTCTTCTGGACAGTTCAAGACACCCCTGGGGCAAATCATAACTCACAAAT
CGTGTGATTTCTGTATTGTGTAATTCCGTGTGTGAATACTATGATATTTGTATTTAAAACTGGCATTGCACAGTAGA
AAGGTGAAAGGTAAAACTCATGTTAATAATTATAACTTTAAATTGTAATTTACTTTATATATTTAAATTAAAATTTT
TAATTGTGGCAAAATACCCAACACAGAATTTACCATCTTAATCATTTTAATCAGCTCAGTAGCTTCAAGAATGTACA
TTGCTGTGCAACCAATATATCCATGTTGCAGTGTGTGTCAGAAATTTGCTTCCTTGTTAAGACTAAATAATATTCCA
TTGTATGTGTATACCAAATTTTATTTATCCACTCATCCAGCAATAGGCATTTGGGTTGCTTCCACTTTTTGCCGTTG
TAAATAATCCTGCTATTAATTTTTCTGCATAAGTATCTCTTTACCTTTGTTTTTGCTAAAATAAACAAGACTGTCTT
TTGTTTATTTAGACTTGGTACATACATCACTTTGCATACTTGCGAGTGTATTTGTAGGATAAGTTCCTGAAAGTAAA
ATCATTGTGTCAAAGGAAATGTGCCTTTTATGATTTGTTAAATACTGTTAAATTGCTGTCCGCAGATGTTGTACCAT
TTTTCACAGCTTCACCAACTCTGTATGTTTTCAGAATTTTTGATCTTTCACAATGTGCTAGATTGAAAATTTTCCAT
TAATAAGAATGAACAGAAACATGTGAAATATGTTTAAACCCATGAGTTTGTAGTAATACAATAAACTTCCCCTGAAA
CCCAAAATGTTTACTTTTGGAGGATACAGAAAGTGAAGAGGAACTAAAAAAAACCTCTTAATGAAAGTGAAAGAGGAG
AGTGAAAAGGTTGGCTTAAAGCTCAATATTCCAAAAAACGAAGATCATGGCATCCGGTCCCATCACTTCATGGGAAA
TAGATGGGGAAACAGTGGAAAACGTGTCAGACTTTATTTTTTGGGGCTCCAAAATCACTGCAGATGGTGACTGCAGC
CATGAAATGAAAAGATGCTTACTCCTTGGAAGAAAAGTTATGACCAACCTAGATAGCATATTCAAAAGCAGAGACAT
TACTTTGCCAACTAAGGTCCGTCTAGTCAAGGCTATGGTTTTTCCAGTGGTCATGTATGGATGTGAGAGTTGGACTG
TGAAGAAGGCAGAGCGCCGAAGAATTGATGCTTTTGAACTGTGGTGTTGGAGAAGACTCTTGAGAGTGCCTTGGACT
GCAAGGAGATCCAACCAGTCCATTCTGAAGGAGATCAGCCCCGGGATTTCTTTGGAAGGAATGATGCTAAAGCTGAA
ACTCCAGTACTTGGGCCACCTCATGCGAAGAGTTGACTCATTGGAAAAGACTCTGATGCTGGGAGGGACTAGGGGGC
AGGAGAAAAGGGGATGACAGAGGATGAGATGGCTGGATGGCATCACCAACTCCATGGACATGAGTCTGACTGAACTC
CAGGAGATGGTGATGGACAGGGAGACCTGGCGTGCTGTGATTCATGGGGTCACAAAGAGTCGGACATGACTGAGCGA
CTAAACTGAACTGAACTGAACTGAGGGAAACAATTCATTATTTTGAAACCAACAAATGAAAACAAAAAGAAAGAATC
AAGAATGACGTGCCTTGTCTATGTAAACTGTATGCCAGGATAGCTAACTAGTGACCCAGAGGAAGTATCTGTTTATA
AAAAAATATTCTCATGCAAGCACTAAGGAAATTAAAGGCAACAATCATCAATTGTTACAAAGAGATGAAGATAAATG
AGAAAGAGATGACTAGACATCCATCTATGCTTTATAACAGGAGACCACCATGCTACCTCTCAACTGTTCTTGCCTAA
ACATTCAAACCTGAATCTGATCAAGTCATCAGATTTAACTGCCAACAAGTTACAGAAAATACAGGGATAGGAAGAAT
ATGTTAACATATCAAGGAGATGCAACCAGCAAAATTCAGACTGTAGGAAACTCTACAGAAAAAGTGATAATTTCTTC
AATGAATAAATTAAACAAAGAAAAAGGAAAAAAGGAGAGGAGGGATAGGAAATCTCTAAATTAAATGTAAGGTAGAT
AGCAACAAAATGCAATATATGGGTCCTATTCAGAGCCTGTTTCAAATGAAAAAAACATGTTCTATCATTTACGAGGC

AGTTGGGAATGTGAACCCTGATTGGATATTTAATGATGTTCAGGAATTAATTGTTGATTTTAAGATATTATGCTATT
ATGTTGTTTTAGAATATGATATTTTAGATAATATGGGACATTTTTTTTCAGATTTAAAGAAGTCTTATATTTTAAAG
CCATATACTAAAATAATTGTTGATGAAATGATAACAATGTCTGAGATTTGGTTCAAGATAATACAATGGGAATGTGT
GTGGGATGGGGTATGGACGAACCATGATTGATTATGAGCTTATGATCGCTGAATGATTCATACTCTACAATTCTCTC
TTTTATGTTTGAAATTTTCCAAATAAAAAAGTTCATTAAAAAAAAAAAAAGTCAATTGGGTCCCAAAGTGTCGAACAC
AACTGATCAACTGAACTGAACTGAATCAGAAACAATGAGTGCAAATAGCTCTTTTTTTTGAATTTTGATGTAAAAAG
GAAAAACAAATCAACAGCTATAAGGGTCAACAGAAGGGTATTACCAAAAAAAAAACAACATAYCCTTAATTATTCAG
TTCTTACATATATGGAAAAGTGTACAGCACAAGGACTCCTCACTAAGTAAAGCCCCCGTATATTAAAAGAGGTATTT
GAGTATGTTACACTTAGGGCACGATCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
CTGCTGCTGCTGCTAAGTCGCTTCAGTCGTGTCCGACTCTGTGCCATAGACAGCAGCCCACCAGGCACCCCCCGTCC
CTGGGATTCTCCAGGCAAGAACACTGGAGTGGGTTGCCATTTCCTTCTCCCTAGCCCCTTGCATATGGGCTCTCCAT
GTGTCCAGACACAGACACATGCCTGTAGGAAAAATGCAGATGGGCTGGAAATTCCTTCTTCTTTTTTTGGCCATGCC
ACACAGCATGCAGGATCTTAGCTCCGTGACCAGGGATGGACCCTGTGCCCTCTGCTGTCATGGGAGTGCAGTGTCTA
AACCACTAGACTTCCAGGGAAGTCCSRGCCTTCCTGAACAAGTTAATGGCACAATAGAAGTATGACCAGGGTGGGAG
CTAGGGAGAAGAGAAACAAGTACTGACGGGCGCTACTTCACAGGATGAGGGTGGGTGAAGGAGCTTGCTTATCTAGG
CTTTGGAAGAGACACCCCACTTAACCGTCCCCTGGTCGCACCCTCCCCCTCTCGGCCCAGTCTGCAGAGGCCCCGCC
TTCCTCCCTGTCCCGAGCCCAAGGGCTGCTGGCGCCGCAGCGTTCCCCCTCGCTCCGAGGGCCTTCCCAGGCGTGGA
CCTCTGAAACCTTCGACTGCGTCTGGCGTTTCTGAGCTCCAGTGGGCTAGGGCCTGCGGGTTTCCTGGGGAGGGCGG
CGGGCGCCGGCGGGAGGACGAGGCAGTTCAGGCCCTAGGCCTGCCGGGCCGGGCCGGGTGAAACGCGGCCGTCCTTC
GCGCTCCGGGGCGCGCGCGGCCCTGTGGGGATTTCACCGGCGACCCGTAGCTCGGGCACGCGCCTGTCGCAGCCCGC
AGGAAGGAGGGGTCCGGCGGCCGCGCTGCTGCCTGGGCGCGGCGGCGGCGGCGGCGGCAGCGGCGGCGGCGCCC
TGTTGAATGGGCTGTGAGGGCCCAGTTTTCAAGCGCTGGCGAACGCGGCCTCCGGGGGCGCAAGGCAGCAGCAGCAG
TGGCGACCAAACGGGTGTTGGAGTTGGCGGCGGCCATGGAGGGCCTGGCTGGCTATGTATACAAGGCGGCCAGCGAG
GGCAAGGTTCTGACTCTGGCCGCCTTGCTTCTCAACCGGTCTGAAAGCGACATCCGCTATTTGCTTGGCTATGTCAG
CCAGCAGGGAGGGCAGCGCTCCACGCCCCTCATCATCGCAGCTCGCAATGGGCACGCTAAGGTGGTGCGCTTGCTTT
TAGAACATTACCGGGTGCAGACTCAGCAGACTGGCACCGTCCGCTTCGACGGGTAGGTACATCCCAGCTCCTGCCTC
CTCTAACGCGTGCGAACCTGTCACCTCATGGATCTTCCCGTCCCACTCTTTCGCTTCCTGGAGGTATCACTTCTCCC
ATTTTTGTATCTCCTGACCTGCCGCATTAATAACCACTTCATCTTCTAGGGCGGATACTCCATTCCATGTTTTGCTG
CCCATCCTTATCGCCTCTACCCCTTATTCCCTCTATAGGTAAATAGCCACACCCTCTGTCTTAATATCTAATAGGTA
TTCTCCAGTATTCTTGCTTGGGAAATCCCATAGACAGAGGAGCCTGGCGGGCTTCAGTGAATGGGGTCGCAAATAGT
CGGACACGACTGAGCACGCAGGCAAAACCAATGCTTCTCTGGATAGTGTATGAACAAGCAGTATGGAGGACGGAGCG
GCGTTTCAGGTTTTCCCTGTGTCAGGGCTGGACTTGTAACCGAAATCAGAAATTAACAAGGTCAGCAATTAACATTG
GAAACCTCCGATGAATTATTTGACCGCCCAATTTTAGCTCTTTGAGTGTCTGATAAAGCCTTTAGTGGGTTATGGGA
CCTTAATGTCAAAAGAAGCCCCTAGAACCAGCAGCAGCTTCAGCCTCTTGCGGGATTGTCATGCAGCCAGACATCCC
CAGACACGATTCTGACTTACACATTGCTGGGAAGGAGACACGTTGGCTGCGAAAATGGCATAGGACCTGGTGTGGGA
GAAGATGATGCCATGTAAAATTAGTTATCCCATCATCTGGGGAAGGCTTGATAATTTGCTGTTATTCCACGTGGGGA
GTTCAAAAAATGCATATTGTCATAGTTAAAATTTGCCATTTAGGTTGAATTGTACATATTTTGAGGTGTCATATTGT
TGAATGCAGAATTGTCAGGTTTGACATTATCTTTTCTGTATGGATGGTTATAACCTGTTTTACAGATTACAGCTTCA
TGAAGATTTTGCTTCATTGATTGCATTTGTTTTTGTTGACTTGTGCCAGTAGGTAAGCTCTTCTAGCGTTTTCTTCT
TTTGAGATAGATGTGATCCAAGTATAAAGATATGTATCTTTAGTTTTGATAATTGAATACTGCTTGGATAAAACTTG
ATTGAAAAAAACAATTCTTTGAGTTGACTAGGGCATTGTTGAATTTTGATTGATTTTTTTTTTTAAGTTGAAACAAG
AAGAGCTTTTTCTTCTGTTTCAGATTAGAGGGAAATATTGCAATTAGAGTATTGCATTTTACTGAAATGATAGCCCA
TGAGTGGTGGTTAATGTCTAGTACCACCTCTGTAGGCATTAAGGAGTTGGTCTTTCTCATATTAACCACAGATGGCA
ACCTGATTCTTGTTTGTTTTCCTTTACTGGCTTCTGTGTTCTTTTTGTAACTCTCTGCATTCCCACAGCCTTTCCTA
TACACAGAGGAGGAGTGAAGTAGAATGACCTCTGAACTTCGAGCACCATTCCAGTTGTATAAGCACAAGCATCCACT
GAGCTATGAATCTCTGCAGTAACAGCTGCTGGATCCACTGCTGCTAAGGAGATTGCTGAGGCCTGTTTTGATGAGAG
AATAGTGGCAAGAAAAAATAGTTTTTTTTAAAGCATTTGAAAAGGCTTTGCTTATCTTTCCTAACTGTCAAAGAGAG
GTGTATTTTGAGGTGGCAATTAAGAGAACCTGAAATGATAAGGAATGTAGCATTGAGCCTTCATAGTGGCTTGCCCC
ATATTTTGCACTGACCAGTGAAATTAAGTTTAGTAAGCAAACATATTTTGAACACCTACTGTGTGCCAGGCTCTATA
GTAGGGATTGGGTGTGCAGAGATGAGTAACATGCGGTTCCCAGGAAGACAGTCAGGTGGCCCTGGGTGGGAGGGATT
AAGGTGCAGTGGGGAGGGGAAAGATTGGCGAAAATGAGTAATAGCTGCCATGCCATTTAGGGTTACAGGAATCTTTT
GGACTACTGAGGGCATGCAGGAAGAGGAAGTGCTCAGGTTTGCCTGGGAGAAGCAGGGAAGTTTTCACAGCTGAGGT
GATGTTTGGGCTCAGTCTTAAAGGATAGAAAGAAAGTAAAGAAGGGGAGCAGCAACATGGTGCACTCGCAATGACAA
GTGTTTCAGTGCTGCTGCAGGGAGGGGTTCAAGCCTGGGAAATTAAGAAAGAGATTGTGGGAGACTGGAGTGAGGTG
GAGGCCAGATTTTGGTGATTGTCATGTAAAATATCAAGAGACTGGTGAGACTGGAATTGCAAACATTTCTGTCATTC
GACATTTTGTCTGAGCCTTCTTCAATCTGTCTTTTTTTGTTTATGTATTATTACAATTATGGATTCATTAGCTGCTC
TGTCAAGGTAGCTTAGTTTTTTAAAACTTACTAAATGACTGTGATCGGCCTGGCACTGTGCCCCACACATACTGTGG

AAGATACAGAAAAAAATACAGTTTCTACTTTTGAGGAGCTTACAAACTAATTAATAGAAGAATTGATATTCGTGGCC
ACCTTTCTGCAGACTCACAGCATTCTGAGATTAGTTTGCAGTTTGTTCCAGTCATTTAGGCATTTCAGAAGTAGAGG
AAGAAAGTGTTATTTGCAAATGTCATGATTTCACTTTACACTCATTCCAGACATTGTGAAAATGTTAGTCTGAATGA
GGTGTTAATATGAAGCTCACTCACCCCTCAGGGACTGGTTTTTGAGAGGTACTTAGCCAACCCTTCACTTGAGACCC
ATGGCCAGATAGGGACAGACAGGTTCTCTGTGGGCTTTAGCTTAGTGTTTCCTAATTTTAGTAATTTATGTATCATC
TTCACAGTTGTTTGCCAGATTGTTCATCACTTATCCTGTTGTTTTCCTAATAWTTAAAAAAAAAACAAACAACAACT
CACTTCTTAGCTTCGTCTTAAGCAACAGCACCTGTATGATAATGGGTTTGATACAATCCTTTGTGTGTGTGTGAGAG
AGACATGCACAGATATTACAATAACTTCTTTTAAAAAGTCTATAATCTGCCTAAAAACATCTATAAAAGCTTGAAAA
CTAAAGTGAATTCATAATAGAGCCTCAAGGGAGAAAGAGCCACCCACCTAGGTTAAGCCTGGTATTTGAGTCCTTGT
ACTAATTTTTAGGAGTAACATTTGAGATTATTTTACTTATGGCTTACTCTGTTAAGATTAAAGTTGACATAGATGAG
GGGATAAGAAGAAAAGTTACATGAATAGAAATGAAAGAAGAACCTTTAGTTGTAGAATGGGAAGCTGTTAAGAGTGT
TACAAAAATGGACCAGAACAAAGGAGAGGCGTCATGTGGTCTTGGAATATTGTGCTGGTTTGCGTTCTGAGGGCCTG
ACTGCGCAATGCCAGAGTTTGTGGGGAACTCTTGTCCTAGAAGAAGTTTGCTGACATGGATAGGAAGGAGAACAGGA
CCTCTGGGAATTTTGTCGAGTGCTCTTTTAAAATCAGTGCTAGTACAAATCCTTAAATGTCCAGATAGAGGAAAGGC
GTAGGGGCACAGTGTCACTTGACTTCATTTAGAAAGTCATTTCCAGTGTATATTTCAAGGAAGTACATATGATCTAA
TAAATGATTTTAGGACTTTGGTTAAAATAAAGTTAATCATTTACACTGAATGAATTTTTATTTTCAGTTCAAATGCT
TGAACCAGACAATAGATGAGTTTAGGTTTGTGATCAGAAATGTTATTTTTATACATTTTATTTTAAAAGGCGAATGC
CAGATTAATGTTGCTTGATTTTTCAGATGAAATACATATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGC
GCGCGCACGCGCGTGCGCGCGCGCGCGCGTGTGCTTGGTCACTTAGTCGTATCTGACGGTTGAGACCCCATAGACCA
TAGCCTGCCAGGCTCCTCTGTCCATGGGATTCTCCAGGCAAGAATACTGGAGTGGGTTGCCATTTCCTTCTCCAGCG
GATCATCCCCACCTAGGAATCGAACCTGGGTCTCCTGCATTGTAGGCAGATTCTTTACTGACTGAGCTGTGAGGGAC
CCCCTAAAATTTGTAATTTTAGACCAGAAAGATCAGTGTTTTATAGTGCTTCTCCCATACTCATTAAGTAAATATTG
CATTTGATTATATTTTTATTTACTTTCTAGAAATACTTATAGAAAGTAAATATTTTTGGTATTTTGTTTTGAGGCAG
TATGCAGCTATAACTTTAGCTGTATTTTATGGACAAAAAATGTCATAAGTATTAGGATTATTAGTCCCTTTAATCTT
TTTAGTGCTGTATTAAGGTGTAGAAATTTAAAAATTACTAAGGGTAATGTTCCTGAAGGACATTCTGCAGTGAATAT
TTTTGTCATTAAGTATTTCAAAATTTTTATTGCCTCAATAAAAATAGATAATTGAAACTTTTTCTTAGGCTTCTGTC
ATATCACTGTAGGGAAAATGGAAATTAATAGATTACATTGGAAAAAAGTATTACATGATTATAAAAACCAAGTGACG
TTGTTGTTTAGTTGCTAAGTCATGTTCAACTCTTTTGCGACTCCATGGGCTGAAGCCCACCAGGCTCCTCTGTTCAT
GGGATTTCCCAGGCAAGAATAGTGGAGTGGGTTGCCATTTCCTTCTCCAGGGGATCTTCCTGACCCAAGGATGGAAA
CTGCGTTTTCCGCACTGCAGGCAGATTCTTTACCACTGAACCACCTGGGAAGCCCAAGTGAAGTATATACTTGTAAT
GTAACTTTTAATCCTATGTCAGAGTGTGAGTGTGTGTGTGTGTGAGTTGCTTAGTTGTATCTGACGCTTTGCTAC
CCTATAGACTGTAGCCTACCAGGCAAGAACACTGGAGTGGGTTGCCATTTCCTTCTCCAAAAGGAACTATAGAAAGA
AAGAAAGAAAGTGAAGTTTCTCAGTTGTGTCTGACTCTTTGCGACCTCATGGATGGTAGCCTACCAGGCTCCTCCAT
CCATGGAATTTTCCAGGCAAGAGTACTGGAGTGGGGTGCCATTTCCTTCTCCCTTATGTCATTAGTAGTGGGGAATT
TGAGGCCAGAAAAAAAATTTTAAATCACTTTTTTTTTTCCTATTTTAAAATGGGGGTCCTATTTATCTTATATGGTT
ATTGCAAAAGATAATACATGTGAATGCTTATTTTAGTACGACAGTTCTTGAGTTTTATTTAGTTTTGGTTCCTATTC
CTGGCATGTAACAATTTTTTATATTTAAGAAGTGCAAGGAAACTTTTGTAACTATGTAAATATAACTCTTTTAACTA
TGTATCAATTTTTGGGTAGATTGCATTAACAAGCAGATTGGGGAAAAAAAGCTTTTTAAATGGAAACATGTTTGAAC
AATGGCATGTACTGTTTCCTCACATTTGATGCATTATATAGTCGGACACGACTGAGCGACTGAACTGAACTGATAGT
TTCCAAAGATTTTTGTGTTTTTAATCTCATATAAGCCTTTTAGTTATATGTCTTTAAAATGGGCTTAATAATGGTTA
TCTACCTTATAGGATTATTGGGAGAAGGTTATAATGAGAATTAAMTVTAMTGCTCTATTGCTTACTACAGTGCCTGA
TACATTGACATTATTAACTCTGTGTTCAAACTCTTTCTTTCAATTGACTTGTGTCTCTTTTGTTTTTAAATTATATC
TATTTTTAGTTAAGGTATAATATCAAGTAAGGTGCATAAATCTTAAGTGTATAAATGTTATAAAACTTTCATTTGCT
TACATCCATGTAATCACTTTCTAGATCAATAAGCAGAACATTGATGTCACTTTAAAAGTTTCCCTTGTGCCCTTTCT
CAGTCATTTCCTCTCCCTTCCCCTCCCAGAAGCAATGATTATTCTGAGTTCTATCTCTGTAGATTAGTTCTGCCTAC
ACTTGAACTTCCTATAAATGGAGTCATACAGTGTACTCTTTTTTTTTCTGTCTGGCTTCTTTTTGTTTAAGACAGTA
TCTATGAGAGTCACCCACTGAGTTGCATGTATAAGTATTTTTAAAATGTTTTATAATATTATGTTATATGAATGTAT
GGCAGTTTATCCTGTTGACAGATATTTTGGTTGTTTTTAGTTTTAGTATTTGACTATTATAAATTGCGTTGGTATGA
ACATGTTTGTATATGTCTTCTGGTGGAATTTGTGTTCATTTCTCTTAAGTATATACCAAGGGGTGGGACTGTTGGCT
CATAGGGTTGGTGTATGTTTAATACTTTCTCAAGCAGTCAACCACCAAAGTGGTTGAACCAGATTATATTCCCACTA
GAAATTTCCATCAACAGGAGGGGTTTCAGTTACTTCACATCCTGGCCAATACTTGGTATTGTCATGTCAGTATTATT
TATTTCATCTATTCTAGTGAATGTGTATTGGTAACTCCTGGTTTTAATTTGCATTTCCTTGCTGAGTAATTATCTTG
AACATCTTTTCATATGTTTATTGGTTGTTTGAATATATGCTTTTGAAGTATACTTCTTCAGTCTTTTCACATTCTTT
AAAAAAAATGCAGTTGTTTGCTTTTTTAGTGATTTATCAGATAATCGATTGTAGATATCTCCAGTGAGTGGATTGCT
GGTCACTTTCTTTAAAAAAATTTTTTTTTTAATTTTTTTGAAGTCTAATCAGTTTTTTTCTTTTATGGTTTTTTTTT
TTTGCATTATGTTTAAGAAATCATCAACCCCCAAGGTCATAAAGATAGTTCTTTTTTTTTTTTTTAAGTAAGCTTTT
ATTTGTTTTGTTTCACATATTATAGGTCTTTGCCACTTTGAAATTTTTGTGAGTAGGGTCAAGTTTACTTATTTTCC

AAGCAGATGTCTAGTGGCTGCTTCTTCACTATTTATTAAACAGACCATACTTTATACCTGAATTTCAGTGGACCTTT
GGTCATTAATCTGTGTTCCTGTATATGTGAATCTATTCCTAGGGTGGGGTTTTTGGTTGTTTTCTTTTGATTAGTAT
TTATCTATCCTTGTGCCAGTGACACACCATCTTACCATATTTTGTCCTCCAACTTTATTTTTCTTTCAAGATTATCT
TGGTTGTTCTGGGTCCTTCCTTTGTGTCACTCTATACATTTTAGAGTTGGCTTTTCAGTATCTACACACATGTGGAG
GCATGCACACACATATGCCTACTGGGGTTTTGATTTGGATTGCATTGAATCTGTAGATTATTTTGGCGGAGAATTGA
CATCTTAACATTCTTGAATCTCCAAAACAATAAACATGGACTATTTCTTCATTTCTTTAAGTCTTTGATTTATCTCA
GGAATGTTTTATAGTTTTCAGTATAGAGTAGAGGTCTTGTACATTTTTCATTAGATTTACTATTAGACATTGGTATT
TTCTGGTGCTTTTGTAAATTGTATTCTTTTTTGACTGAAAGAAAACTTTTTTCTTTTTTAAGAAAACATTTTTGAAT
TTTGTGATTATAATCATGACAATAAAGATGTAATAAGTGAATTTTAATTAAACTAAATTGTCTGCAAAAGCTACAGT
CGGCAACCACTGAATCATTTCAACTGATGCATGTCTCATTCTCTTGATTCATTCTTTTTTAAGATTTGTTTGTGTAT
GTTTATTCTTTTGGCTCCGGTGGGTCTTCGTGGCTGTAGTGGGCTTTCTCTGGTTGTGGTGGGCGGGCCTCTCATTG
CAGTGGCTTCTCTTGTTGGGAGTGTGGGCTCCAGGTAGGCTTCCCAGGCTTCGGTCGCTGTGATCCGGGGCTCAGCA
GTCACGACTCTTGGGCTCTAGAGCATAGGCTTAGCAGTTGTGGTGCGCAGGCTTTGTTGCTCTGCAGCTTGTGGGAT
CTAACCAGTGTCCCCTCCATTGCAGGGTGCATTCTTTACCACTAGATGACCAGGGAACCCCCTCTTGATTGATTCTT
TAGTTTAAAAAATTAATTAACAATTAAATGACATTTTAAAAATGAAAAATTTAAAAGGAAACTTCAGAAGTATTCAA
ATTTATTTTGAGGACTGTTTTCCATACAGTAAAATGTACAGATTTTTCTTATGCAATTTGATAAATTTTTTTCCCCT
AAATATCCACCCGTGTCGCCACTACCCAGTTCAAAATATAGGACATTGCTAGTAGCTTTAAGCCCCTTCCGAATCAA
TCCTACACTCTCCCCCGTAGATAACCATATAGGTAGCTGTTACTGAGAATGCTTTTACTTATAGAAATTTTGCTTCA
TTATTTGAAACTAAAGAAGTAATTGCTTGGATGTTTATAAAGTATGGCTTTTATTTTCAGTAGAAGTGATAGACAGT
ATTGATAGTTGGGGTTGAACTATGATTTTTTTTTTTACCAAGTAGATTGTACTTTTTTTACTTCATAGGATATATAA
TTTGACCTGTGTTGTATCATTCATGTTGTAAGGAGCAGTAATTATTTCATTTTCATTGCTTTAGTAGAGTATGTACA
TACTTACTAGTTTAGATAATCCGAGGTAAAATTATCTGCTCAAATACTTTTCTTGACGCAGTGTCTAACATTAATAC
AGAGATTGTGGTGGCTTTCTTACTGATATACTGAGAGAAACTGTGGTACCTGAGGGTGTTGGGAGGGAGGAACGTTC
TGTGCTCCTCTAAGAGCCAGGGAATCAGCTTGGTGGTAAAGAAATTACAAAACTTTGTACATTTCCTCTGTTTTCTG
AAATTTATGTACTTCTGCTAGTAGGAAATCAGTGACTCTTTATTTTTTCTTTAAGTTTTTATACTGTGGAAATACTA
ATAATCGCTCACATTTATTGAGTTCATACTAGGTACAAGTACTATGCTGAGTGCTTTACTTAAATAACCTGATTTTA
TTTTCACCGAAACTAATGTTTAACTAATGCTCTGTAGACTGAGGTTTGGGGTCAGTGTAAATGATTCCATCCTTTAA
ATTTATGCTAGGGCAGGTGACCCCCATTTTCTAACTGTTGGCTCTTTTCATGAACAGGTATGTCATCGATGGGGCCA
CTGCTCTTTGGTGTGCAGCAGGAGCGGGACATTTTGAAGTTGTTAAGCTTCTAGTCAGTCATGGAGCCAACGTGAAC
CACACCACAGTAACTAACTCAACCCCTCTGCGGGCAGCATGCTTTGATGGCAGATTGGACATTGTGAAATACTTGGT
CGAAAATAATGCCAACATCAGCATTGCCAACAAGTATGACAACACTTGCCTAATGATTGCGGCCTATAAGGGACACA
CTGACGTGGTCAGATATCTTTTAGAACAACGTGCTGATCCGAATGCTAAAGCACATTGTGGAGCCACAGCATTGCAT
TTTGCAGCTGAAGCTGGGCACATAGATATTGTGAAGGAGCTGATAAAATGGCGTGCTGCTATAGTGGTGAACGGCCA
TGGGATGACGCCATTAAAAGTAGCTGCTGAAAGCTGTAAAGCTGATGTTGTCGAACTGTTGCTCTCTCATGCTGATT
GTGACCGGAGAAGTCGGATTGAAGCTTTGGAGCTCTTGGGTGCCTCCTTTGCAAATGATCGCGAGAACTACGACATC
ATGAAGACATATCACTATTTATATTTAGCTATGTTGGAGAGGTTTCAAGATGGTGATAACATTCTTGAGAAAGAGGT
TCTCCCACCAATCCATGCTTATGGGAATAGAACTGAATGTAGAAATCCTCAGGAACTGGAATCCATTCGGCAAGACA
GAGATGCTCTTCATATGGAAGGCCTTATAGTTCGGGAACGGATTTTAGGTGCCGACAACATTGATGTTTCCCATCCC
ATCATTTACAGGGGAGCTGTTTATGCAGATAACATGGAATTCGAACAGTGTATCAAGTTGTGGCTTCATGCCCTGCA
TCTCAGACAGAAAGGCAACAGAAATACCCATAAGGATCTTCTTCGATTTGCCCAAGTTTTTTCACAGATGATACATT
TAAATGAAACTGTGAAGGCCCCAGACATAGAATGTGTTTTGAGATGCAGTGTGTTGGAAATAGAACAGAGTATGAAC
AGAGTAAAAAATATTCCAGATGCTGATGTCCACAGTGCTATGGACAATTATGAATGTAACCTCTATACCTTTCTGTA
TTTAGTGTGCATCTCCACCAAAACACAGTGCAGTGAAGAAGATCAGTGCAAAATTAACAAGCAGATCTACAACCTGA
TTCACCTGGACCCCAGAACTCGTGAAGGTTTCACCTTGCTGCATCTAGCTGTCAACTCAAATACCCCGGTTGATGAT
TTCCACACCAATGACGTTTGCAGCTTTCCAAACGCGCTTGTCACAAAGCTCCTGCTGGACTGTGGTGCTGAGGTGAA
TGCTGTGGACAATGAAGGGAACAGCGCCCTTCACATTATCGTGCAGTACAACAGGCCCATCAGTGATTTTTTGACCT
TGCACTCTATCATCATTAGTCTAGTTGAAGCTGGCGCTCACACTGACATGACAAATAAACAGAATAAAACTCCGCTA
GACAAAAGTACAACTGGGGTATCAGAAATACTACTTAAAACTCAAATGAAGATGAGTCTCAAGTGCCTGGCTGCCCG
GGCAGTTCGGGCTAATGACATTAACTACCAAGACCAGATCCCCAGAACTCTTGAAGAGTTTGTTGGATTTCATTAA

Appendix 6

Sequences generated for validation of the identified disease associated deletion.

1. Normal sheep CPW384 using forward primer

```
>Ref_seqforward
GTGCTGCTGGAGACCACCCGGGGTCCGGGGTCCACGGCCCGCGGCCCGGACCCCCCGCCCGCCCGATCACGTGGCGG
GGCCGCGGTTCCGCTCCGCCCGCCGGCGCGCGCCCACCTGGCCTGCAGGAAGGAGGCGCCCGGCTGCGCGCCGGAGC
CGCCCGCCGCTCCCGGGTGCCGCCTCCGCGCCGCAGCCTCCATGGCTAACCTTCCGGCCAGGCCCTTCGCCCCGCTG
CCGGTAGCTGCAAACAGGTTCCGCTGGCCGAAGGCTGCGAAAGCGAACATCCAGCGCCGGGGCGAGCGGGGGCGGAC
GAGAGCGAGAGCGCCCGCCCAGCGGCGCGGGCCGGGGGCGGAGCCGGCGGGGGCCGGGACCAGGAGACGCGCCGCAG
AGCGAGGGAGGCGGGGCGAGCTCCCGGCCGGGAAGGAGGGGCCCACGGCCGGGGCGGGGCGGGGCGGAGGGGCGGAG
GGAGGGCGCACGGCGCGGGGCCTCCTCGGGCTGGGACGCC
```

2. Normal sheep CPW384 using reverse primer

```
>Ref_seqreverse
GGGGGAACGCGCGTGCGCCTCCTCCGCCCTCCGCCCCGCCCCGCCCCGGCCGTGGGCCCCTCCTTCCCGGCCGGGAG
CTCGCCCCGCCTCCCTCGCTCTGCGGCGCGTCTCCTGGTCCCGGCCCCCGCCGGCTCCGCCCCCGGCCCGCGCCGCT
GGGCGGGCGCTCTCGCTCTCGTCCGCCCCCGCTCGCCCCGGCGCTGGATGTTCGCTTTCGCAGCCTTCGGCCAGCGG
AACCTGTTTGCAGCTACCGGCAGCGGGGCGAAGGGCCTGGCCGGAAGGTTAGCCATGGAGGCTGCGGCGCGGAGGCG
GCACCCGGGAGCGGCGGGCGGCTCCGGCGCGCAGCCGGGCGCCTCCTTCCTGCAGGCCAGGTGGGCGCGCGCCGGCG
GGCGGAGCGGAACCGCGGCCCCGCCACGTGATCGGGCGGGCGGGGGGTCCGGGCCGCGGGCCGTGGACCCCGGACCC
CGGGTGGGTCTTCCCAGCGCTACGCCTGCAGCTGTTTCCCCCCC
```

3. Carrier sheep SH1008 using forward primer

```
>Carrier1
TTGCTCCTGTAGACCACCCGGGGTCCGGGGTCCACGGCCCGCGGCCCCCGGGGCGGAGGGGCGGAGGGAGGGCGCAC
GGCGCGGGGCCTCCTCGGGCTGGGATCTGCTGCGCTCCGGCCTGCAGGAAGGAGGCGCCCGGCTGCGCGCCGGAGCC
GCCCGCCGCTCCCGGGTGCCGCCTCCGCGCCGCAGCCTCCATGGCTAACCTTCCGGCCAGGCCCTTCGCCCCGCTGC
CGGTAGCTGCAAACAGGTTCCGCTGGCCGAAGGCTGCGAAAGCGAACATCCAGCGCGGGGGCGGGCTGGGGCGGGCG
AGAGCGAGAGCGCCGGCCCAGCGGCGCGGGCCGGGGGCGGAGCCGGCGGGGGCCGGGACCACGAGACGCGCCGCAGA
GCGAGGGAGGCGGGGCGAGCTCCCGGCCGGGAAGGAGGGGCCCACGGCCGGGGCGGGGCGGGGCGGGGCGGAGGGGC
GGAGGGAGGGCGCACGGCGCGGGGGCCTCCTCGGGCTGGGATCTGCTGCGCTC
```

4. Carrier sheep ID SH1010 using forward primer

```
>Carrier2
```

TAGGTATGCTGTAGCCACCCGGGGTCCGGGGTCCACGGCCCGCGGCCCCCGGGGCGGAGGGGCGGATGGAGGGCGCA
CGGCGCGGGGCCTCCTCGGGCTGGGATCTGCTGCGCTCAGGCCTGCAGGAAGGAGGCGCCCGGCTGCGCGCCGGAGC
CGCCCGCCGCTCCCGGGTGCCGCCTCCGCGCCGCAGCCTCCATGGCTAACCTTCCGGCCAGGCCCTTCGCCCCGCTG
CCGGTAGCTGCAAACAGGTTCCGCTGGCCGAAGGCTGCGAAAGCGAACATCCAGCGCGGGGGCGGGCTGGGGCGGGC
GAGAGCGAGAGCGCCGGCCCAGCGGCGCGGGCCGGGGGCGGAGCCGGCGGGGGCCGGGACCACGAGACGCGCCGCAG
AGCGAGGGAGGCGGGGCGAGCTCCCGGCCGGGAAGGAGGGGCCCACGGCCGGGGCGGGGCGGGGCGGGGCGGAGGGG
CGGAGGGAGGGCGCACGGCGCGGGGCCTCCTCGGGCTGGGATTGCCT

6. Affected sheep ID SH1021 using forward primer

>Affected1_F
GTGCTAGCTGGTAGACACCCGGGGTCCGGGGTCACGGCCCGCGGCCCCCGGGGCGGAGGGGCGGAGGGAGGGCGCAC
GGCGCGGGGCCTCCTCGGGCTGGGATCTGCTGCGCTCA

7. Affected sheep ID SH1021 using reverse primer

>Affected1_R
GTGGGAACGCGCGTGCGCCTCCTCCGCCCCTCCGCCCCGGGGGCCGCGGGCCGTGGACCCCGGACCCCGGGTGGGTC
TTCCCAGCGCTACGCCTGCAGCTGTTTCCTCGCCCACTCGA

8. Affected sheep ID SH1025 using forward primer

>Affected2_F
GCCCGCGGCCCCCGGGGCGGAGGGGCGGAGGGAGGGAGCACGGCGCGGGGCCTCCTC

9. Affected sheep ID SH1025 using reverse primer

> Affected2_R
CTCCGCCCCTCCGCCCCGGGGGCCGCGGGCCGTGGACCCCGGACCCCGGGTGGGTCTTCCCAGCGCTACGCCTGCAG
CTGTTTCCTCGCCCACTCG