Modelling Retinitis Pigmentosa in Sheep

Zoe Hancox-Monk

Supervised by Dr Jessie Jacobsen, Dr Russell Snell, Dr Klaus Lehnert and Dr Natasha Mckean

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Biological Sciences, The University of Auckland, 2024.

Abstract

Retinitis Pigmentosa (RP) is one of the most common genetically inherited diseases that is known to cause retinopathy, with a global prevalence of approximately 1 in 4000. The disease affects the cells of the retina, gradually reducing the visual field until complete blindness occurs around middle-age. Autosomal dominant forms of the disease pose a particular challenge, with multiple members of the same family being afflicted by the disease. With no current effective treatment, investigations into generating RP models via gene editing are of particular interest for researchers who work in retinopathies.

We set out to create a cell model of a common mutation in the *RP1* gene called R677X that is known to cause Autosomal Dominant Retinitis Pigmentosa. This occurs through the introduction of a premature stop codon that causes the protein coded by *RP1* to be significantly shorter than a healthy copy and thereby impede its native function. Editing experiments using CRISPR/Cas9 were undertaken in sheep fibroblasts in order to lay the groundwork for the development of a large animal model in sheep. Using CRISPR/Cas9 guides 55 bases downstream from the mutation site, we successfully induced deletions in sheep fibroblasts using nucleofection with an adjusted editing frequency of 47%. The mRNA from the cells carrying these deletions indicates that the frameshifts mutations that were generated in this region would code for a premature stop codon and subsequently delete the majority of the protein coded for by this gene, as is seen in known pathogenic genetic variants of the disease.

Future directions include the generation of an isogenic cell line by replicating the results of this experiment in immortalised fibroblasts. This could then be used to test the ability of CRISPR/Cas9 to knock down the edited copy of the gene and revert the cell to using the healthy copy of *RP1,* creating cells that are hemizygous for the *RP1* gene. This research is incredibly encouraging for the development of a sheep model to not only develop and test potential treatments, but to learn more about the pathology of Retinitis Pigmentosa and how the disease progresses.

Acknowledgements

I would like to start this acknowledgements section somewhat jovially by thanking the late Eddie Van Halen for his contribution to the global catalogue of music – 'Right Now' has been the keystone song in the soundtrack of this thesis. Other honourable mentions include Aretha Franklin, Blue Oyster Cult, The Who, Rush, The Doobie Brothers, Etta James, Stevie Wonder, Earth, Wind & Fire, The Foo Fighters, Bonnie Raitt, Buddy Guy, Nina Simone, The Red Hot Chili Peppers, Couch, Chris Stapleton (I'm as surprised as you are), Pearl Jam, The Rolling Stones, The Tedeschi Trucks Band and, of course, Led Zeppelin.

The first in this inexhaustive list of acknowledgements for people who might actually read this, or tragically (for their sake) already have, are my supervisors. Russ, Jessie, Klaus and Tarsha: I can truly say it has been a pleasure working with you on this project and in this lab. The confidence this experience has given me is immeasurable and I credit a good part of that to your confidence in me, especially when I had none. To my wider community of fellow postgraduate students: Annelies, Ben, Jess, Ash, Ashley, Vic. Your support, editing, practical assistance and belief in me has kept me trudging on through and with a significantly larger smile on my face than I otherwise could have sported. Corey: you are deservedly separated from others. Thank you for your time, energy, and last-minute formatting. The shocking frequency with which you have put up with my rants and somehow brought me back to reality is appreciated more than I can say.

My parents, Pete and Shel: thank you for your unfaltering support and love, particularly when none of my explanations of what I actually do in my thesis made any sense. Your position has always been that you knew I could do this, and that buttress has kept me from collapsing from exhaustion, especially in the past few months.

My poor, poor husband Will. I have absolutely no idea how you managed to live with me, support me and ground me in the past two years, especially with anything I earn going directly to this master's degree, but you did it. You inspired me to study biology after years of being classed as "the arts kid" and have always been my most fervent supporter, even when I do something stupid. With my love and apologies, you're stuck with me now.

A final acknowledgement in memory of my Nain who taught me so much and shaped the person I have become. She had the kindest soul, a steadfast Welsh pride and a wicked sense of humour. In her memory, I say: addysgedig yn well na chyfoethog.

Table of Contents

List of Figures

List of Tables

List of Appendices

Glossary of Abbreviations

RP	Retinitis Pigmentosa
RPE	retinal pigment epithelium
ROS	reactive oxidative species
ORP1	oxygen-regulated protein 1
NCBI	National Centre for Biotechnology Information
AAV	Adeno-Associated Virus
FDA	U.S. Federal Drug Administration
EMA	European Medicines Agency
CRISPR-Cas9	Clustered Regularly Interspersed Short Palindromic Repeats, Cas9
PAM	Protospacer Adjacent Motif
PRA	progressive retinal atrophy
HDR	Homology-Directed Repair
mRNA	messenger ribonucleic acid
PCR	Polymerase Chain Reaction
SNPs	Single Nucleotide Polymorphisms
DMSO	Dimethylsulfoxide
IDT	Integrated DNA Technologies
ATG	Applied Translational Genetics
GFP	Green Fluorescent Protein
SARDI	South Australian Research and Development Institute
HT-PAMDA	PAM determination assay
ClinVar	Clinical Variant
NHEJ	Non-Homologous End Joining

1 Introduction

1.1 Retinal Dystrophies and Retinitis Pigmentosa

Retinitis Pigmentosa (RP) is a disease that primarily affects the cells of the retina, gradually reducing the ability of photoreceptor cells to receive and translate light signals to the brain (OMIM 180100; Liu et al., 2022). Initially described in 1853 by Dutch ophthalmologist F. C. Donders, the term 'Retinitis Pigmentosa' was coined in 1857 in a publication the physician made detailing the process for examination of the retina (Donders, 1857). Although the suffix -itis, meaning inflammation, is now largely deemed a misnomer in relation to Retinitis Pigmentosa, it is well known that inflammation is a component of RP and other retinal dystrophies (Kaur & Singh, 2021). Retinitis Pigmentosa belongs to a group of degenerative disorders of the retina known as hereditary retinal dystrophies, of which it is the most common (Sakai et al., 2022). People presenting with retinal dystrophies suffer from colour and night blindness, peripheral vision abnormalities and progressive reduction of vision leading to partial or complete blindness (Chawla & Vohra, 2023). There are multiple genes and mutations that are known to cause Retinitis Pigmentosa, all of which cause photoreceptor cells to degrade, leading to night blindness and difficulties navigating in low light situations (Sakai et al., 2022).

1.2 Prevalence

Retinitis Pigmentosa is a leading cause of bilateral blindness and visual disability in people under the age of 60 and is known to affect at least 1.5 million people globally (approximately 1 in 4000), though the rate is reported to be higher in South Indian and Chinese populations at 1 in 930 and 1 in 1000 respectively (Wu et al., 2023; Ayuso & Millan, 2010). Other researchers have also estimated the prevalence to be as high as 1 in 750 in rural Central India (Nangia et al., 2012). The high prevalence in these populations is theorised to be due to consanguineous couplings which increases the relative proportion of recessive alleles in the population (Ayuso & Millan, 2010). Retinitis Pigmentosa is a progressive disease that presents within the first few decades of life, with many individuals developing blind spots in their peripheral vision by their thirties and full blindness by their forties (Chivers et al., 2021). This rate of presentation represents the majority of cases; however, there are other variants of the disease, such as X-linked Retinitis Pigmentosa,

where symptoms present earlier, with patients becoming legally blind in their twenties (Chivers et al., 2021).

1.3 Pathology

1.3.1 Anatomy of the Retina

The human retina, encasing the back of the eye, (Figure 1.1B) can be best described as consisting of 10 layers, each playing a distinct role in the translation and transmission of signals to the brain (Sarna et al., 2017). Each layer consists of different cell types that work collectively to translate and transduce signals to the nerves in the nerve fibre layer, which conveys those signals to the brain (Figure 1.1A) (Wen et al., 2022).

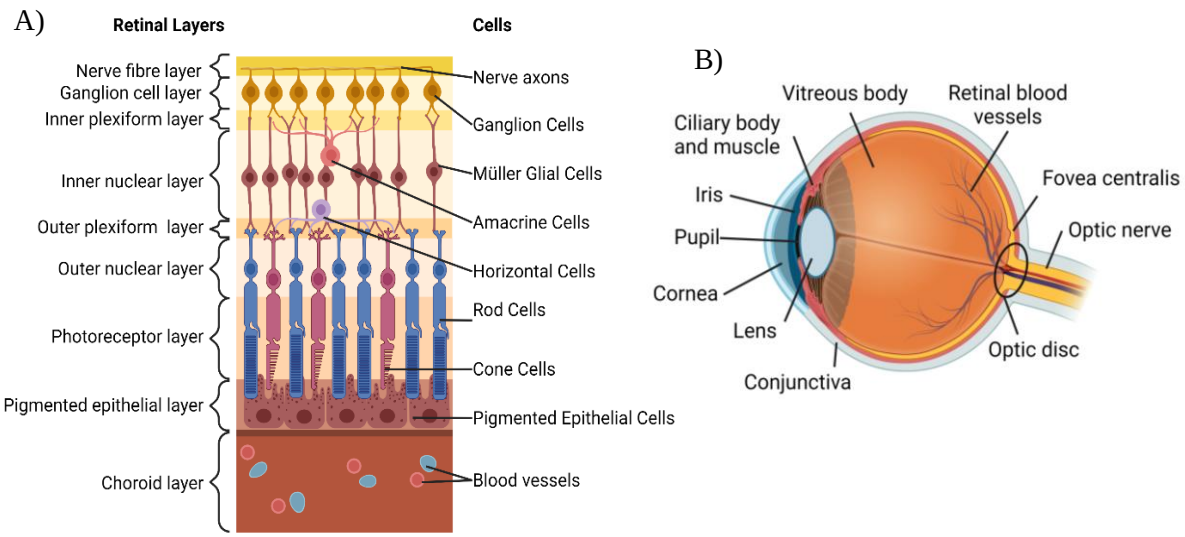


Figure 1.1: Anatomy of the retina. (A) the outermost portion called the choroid layer contains the vasculature for the retina, while a channel inside the optic nerve allows blood vessels to supply the internal portion of the retina and the vitreous body. The photoreceptor layer contains rod and cone cells, which translate different types of light signals; rod cells are responsible for perceiving black and white and cone cells are responsible for the colour spectrum. Light signals are translated into electrical signals and transmitted through the inner nuclear layer and into Ganglion cells, which can then be transmitted to the brain along the axons of these cells. **(**B) The anatomy of the human eye showing the outermost layer called the conjunctiva, the iris, the opening of the iris (pupil) and the muscles that control the dilation of the pupil. The eye is made up of a large vitreous body with blood vessels at the back of the eye near the fovea centralis where the central field of vision is located. The optic disc, also at the posterior portion of the eye is where axis of the retinal ganglion cells is located as they move back into the optic nerve. **Figure generated using BioRender.**

1.3.1.1 Retinal Pigment Epithelium

The retinal pigment epithelium (RPE) is made of a layer of 4 to 6 million cuboidal epithelial cells in each eye (Wen et al., 2022). Across most of the retina, rod cells outnumber cone cells by a factor of 100, but that ratio of 100:1 reverses in the *fovea centralis*, the pocket at the back of the retina just above the optic nerve that is responsible for central vision and colour perception from cone cells (Nguyen et al., 2023). The interior portion of the retina is comprised of melanin-containing RPEs, which have a multitude of regulatory functions such as nutrient and metabolite transportation, phagocytosis, neutralisation of free radicals, and vitamin A metabolism (Wen et al., 2022). Retinal Pigment Epithelial cells typically contain melanosomes which help to manage light absorption (Sarna et al., 2017). These melanosomes, which produce melanin, provide photoprotection by reducing free radicals in the eye and neutralising reactive oxidative species (ROS) within the melanosome, thereby reducing the potential DNA damage that can be inflicted (Istrate et al., 2020).

1.3.2 Disease pathology and progression

In people diagnosed with RP, the disease will typically present with what is called the classic triad of Retinitis Pigmentosa; bone spicules, waxy pallor of the optic disc and ocular vascular attenuation (Bhardwaj et al., 2022). Dark spots of melanin appear in the peripheral field of vision, gradually shifting toward and obscuring the central field of vision due to the migration of RPEs into the perivascular regions of the retina, which obscures the *fovea centralis* (O'Neal & Luther, 2023; Verbakel et al., 2018). As the disease progresses, the death of rod cells affects cone photoreceptors by creating a cytotoxic environment, leading to dyschromatopsia, or the inability to distinguish between different colours (O'Neal & Luther, 2023). The bone spicule shape is formed by melanin-containing pigmented epithelial cells after they have migrated throughout the retina; the waxy pallor of the optic disc is most likely due to the formation of glial cells which cover and increase reflectivity of the disc, and the vascular attenuation is caused by cellular stress in the endoplasmic reticulum which damages endothelial cells (O'Neal & Luther, 2023). Retinitis Pigmentosa is typically a genetically inherited disease with multiple genes associated with disease pathology.

1.4 The genetics of Retinitis Pigmentosa

There are around 3100 pathogenic mutations in more than 80 genes identified as causative of Retinitis Pigmentosa (Daiger et al., 2013; Fahim et al., 2023; O'Neal & Luther, 2023). These genes, tabulated in Appendix 1, have a range of functions, including involvement in phototransduction, morphogenesis of photoreceptors, RNA splicing, ciliary development and transport, and retinal homeostasis. Fahim et al. (2023) found that only half of all cases have a known genetic cause via autosomal dominant, autosomal recessive, and X-linked modes of inheritance (Appendix 3). The remaining 40-50% of occurrences had no known genetic cause or had occurred spontaneously in families with no history of the disease. However, of those familial cases, most pathogenic variants are inherited in an autosomal dominant manner. While the proportions of autosomal recessive to autosomal dominant occurrences are relatively similar, dominant inheritance is slightly higher at 15-25%. The modes of inheritance and estimated proportion of all RP cases are detailed in Table 1.1.

Table 1.1: Modes of inheritance for Retinitis Pigmentosa

Mode of Inheritance (MOI)	**Proportion of All RP Cases**
Autosomal dominant RP (adRP)	15-25%
Autosomal recessive (arRP)	5-20%
X-linked (xlRP)	5-15%
Unidentified genetic cause/single occurrence (Simplex)	40-50%
Digenic RP (induced by two genes)	Very Rare

Retinitis Pigmentosa is classified as syndromic, where certain pathogenic variants affect multiple organ systems, or non-syndromic, where variants will only cause symptoms of RP. The three most common causative genes for both syndromic and non-syndromic RP are *RHO* (OMIM: 108380, MOI: adRP), *USH2A* (OMIM: 608400, MOI: arRP and the most common cause for Type II Usher syndrome); and *RPGR* (OMIM: 312610, MOI: xlRP). Some autosomal dominant variants are known to cause disease through the gain of a toxic function. These variants were of particular interest in this thesis as those with the disease could be treated using an allele-specific gene knock-out method that results in hemizygosity. A hemizygous person only carries one copy of a gene and in certain circumstances, some genes have the capacity to function in a hemizygous state when one copy is damaged (Richard & Hawley, 2011). Of the many genes associated

with autosomal dominant RP, a gene of particular interest is *RP1* (OMIM 603937) because mutations in this gene cause both recessive and dominant RP. Mutations in *RP1* make up approximately 3-5% of all adRP cases, thereby making *RP1* a candidate for the induction of hemizygosity using allele-specific gene knock-out (Fahim et al., 2023).

Yamashita et al. demonstrated in their 2009 study that the protein encoded by the *RP1* gene has two doublecortin binding domains (DCX) which bind to the cytoskeleton of rod cells and facilitates stabilisation of the cell. These two domains are binding sites for the microtubule-associated protein doublecortin, which provides stability to microtubules and the cytoskeletal structure by creating scaffolding within the cell (Liu et al., 2004). When the doublecortin binding domains are mutated, doublecortin is unable to form a strong cytoskeletal structure or regulate polymerization within the cell. This results in the death of rod cells within the eye, which leads to the pathogenesis of RP. Further research has shown that the RP1 protein is expressed in both rod and cone cells and plays a role in the correct orientation and stacking of discs in the outer segment of those cells (Silva et al., 2020). Retinitis Pigmentosa is classified as a rod-cone dystrophy because the mutations that cause the disease affect the rod cells of the eye before the cone cells.

1.4.1 Cross-species conservation and the structure of RP1

The RP1 gene is comprised of 4 exons, the first of which is a non-coding exon. Exons 2 and 3 are significantly shorter than exon 4, which contains the adRP hotspot – mutations in this region will typically cause autosomal dominant forms of RP (Chen et al., 2010; Nanda et al., 2019). A multi-species alignment demonstrated a high level of similarity in the RP1 protein in animals that are more closely related to humans – *Pan troglodytes*, for example, had the same protein length and 98.7% similarity – while more phylogenetically distant species, such as *Salmo trutta* had significantly lower identity at 16.9% (Table 1.2). These annotations were the predicted orthologs for *RP1* and were named “oxygen-regulated protein” or “ORP1” due to this name being an alias of *RP1*.

Table 1.2: Cross-species alignment percentage of protein identity in predicted *RP1* orthologs

Species	NCBI Reference Sequence	Percentage of identity to Human sequence (EAW86749.1)
Pan troglodytes	PNI42337.1	98.7%
Puma concolor	XP_025778786.1	69.7%
Urocitellus parryii	XP_026250174.1	67.5%
Ficedula albicollis	XP_016151595.1	21.9%
Salmo trutta	XP_029556954.1	16.9%

A mutation of interest in the *RP1* gene in humans is p.R677X, which causes a premature stop codon at amino acid 677 in the 4th exon of the gene. The subsequent protein that is translated from this gene is severely truncated, with approximately 68.5% of the protein not being translated. This mutation will be the focus of the work in this thesis.

Figure 1.2 shows the position of this truncating mutation across the species examined in table 1.2, demonstrating that this position has varying levels of amino acid conservation across multiple species.

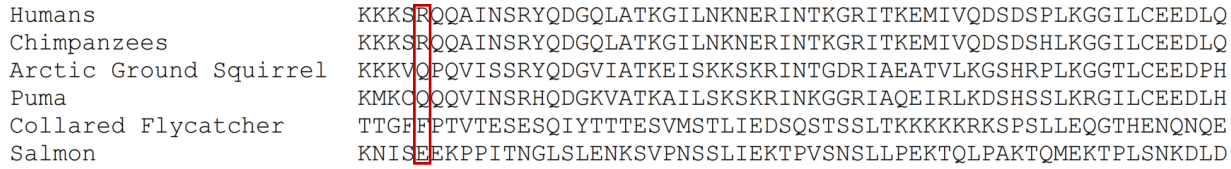


Figure 1.2: Cross-species CLUSTALW alignment of the protein encoded by the RP1 gene across multiple animals. This alignment shows a portion of exon 4 of the protein encoded by *RP1* in six different species from multiple taxa with varying levels of similarity. Outlined in red is the codon affected by the p.R677X mutation in humans and the corresponding position in the orthologs of *RP1* in Chimpanzees, Arctic Ground Squirrels, Pumas, Collared Flycatchers and Salmon.

Figure 1.3 shows a schematic of the *RP1* gene and the structure across the species used in the cross-species alignment. This figure demonstrates the level of mRNA conservation across multiple species, which the nucleotide in the position of the p.R677X mutation highlighted in red.

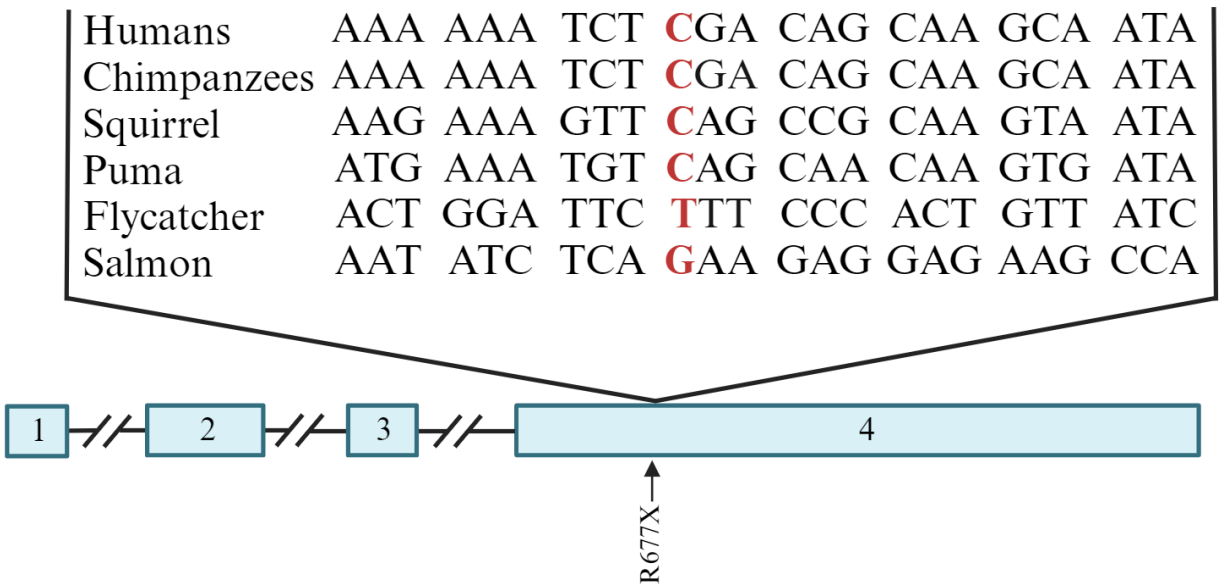


Figure 1.3: Gene schematic of RP1 including codons of cross-species alignment across 24bp of the exon 4 hotspot. This gene schematic shows the exons in the *RP1* gene and the location of the R677X mutation in humans. The mRNA of the species used in the cross-species alignment is separated by codons above and indicates the nucleotide that is affected in humans with the variant of interest. **Figure generated using BioRender.**

The likely orthologs of the human *RP1* gene used in this cross-species alignment (detailed in Table 1.2) all had the same structure and function according to their entries in the National Centre for Biotechnology Information (NCBI) database.

1.5 Current Treatments

The most common treatment for Retinitis Pigmentosa is limited to the management of symptoms with very few approved methods for cessation of symptoms. Many people diagnosed with RP wear Ultraviolet protectant glasses at all times, as exposure to UV light has been shown to increase the rate of deterioration (Parmeggiani et al., 2011). There are multiple treatments that are currently being trialled or have been recommended for many years, including Vitamin A supplementation, artificial retinas, and gene therapies.

1.5.1 RP therapy by Vitamin A Supplementation

Vitamin A was suggested as a potential treatment for RP after mouse models demonstrated that it helped to reduce the levels of small, pigmented granules called lipofuscin in the retina (Berson et al., 2018). This was theorised to reduce the damage that lipofuscin can inflict on retinal cells (Berson et al., 2018). Lipofuscin is known to accumulate in RPE's and has been shown to cause cell death in those with Age-

Related Macular Degeneration and retinopathies (Pan et al., 2021). Lipofuscin is currently believed to convert to cytotoxic derivatives with exposure to light (Pan et al., 2021). Molecularly, lipofuscin is indigestible by lysosomes, a typical cell protection organelle, thereby impeding the cell protective mechanisms in RPEs (Wen et al., 2022). The supplementation of Vitamin A is theorised to reduce these lipofuscin levels; however, multiple literature reviews have failed to find conclusive evidence that supports this theory (as reviewed in Schwartz et al., 2020). Although it has not been proven to slow the progression of RP or improve visual acuity, Vitamin A is not deemed to be harmful when supplemented, so it is still prescribed.

1.5.2 Artificial Retinas

Artificial retinas have been available to people with blindness for at least a decade now, primarily targeting those with age-related macular degeneration and Retinitis Pigmentosa (Humayun & Lee, 2022). The implant is made up of 60 electrodes which provide rudimentary light vision by receiving information from a pair of reading glasses which stimulate the remaining healthy retinal cells to transmit that light information to the optic nerve (Miller, 2013). This means that the degradation of rod cells causing night blindness and reduction of light recognition can be counterbalanced by the prosthesis and provide an improvement in low light environments (Miller, 2013). It is important to note that the prosthesis is still in its developing stages and has a long way to go in being able to provide full restoration of vision, let alone improved visual acuity above recognition of light (Farvadin et al., 2018). Equally, it still holds a significant price tag of ~100,000 USD, something that is unrealistic for many people affected by RP (Miller, 2013). Other issues may also occur in those who have this implant – some studies have found that over the span of 2 years, the prosthesis can move from its implantation point with as-yet-unknown consequences (Ghani et al, 2023).

1.5.3 Gene therapies

With these complications in mind, it then naturally follows that the research has returned to therapies and treatments that could halt disease progression and return the functionality of remaining rod cells (Piri et al., 2021). There has therefore been significant interest in the application of gene therapies for RP, especially given the accessibility of the eye for the delivery of treatments. Most retinal gene therapies currently under

investigation use a viral vector to enter cells and are administered to patients through subretinal or intravitreal injections (Hu et al., 2021). This delivery technique is primarily being developed for treatment of autosomal recessive and X-linked recessive disorders, utilising a gene augmentation approach that adds a functional copy of the gene to supplement expression (Hu et al., 2021). The most recent gene therapy for Retinitis Pigmentosa utilises this gene augmentation approach utilises an Adeno-Associated Virus (AAV) as a delivery mechanism to target the *RPE65* gene (Ducloyer et al., 2020). Known as Luxturna®, the treatment is the first gene therapy to be approved by the U.S. Federal Drug Administration (FDA) and European Medicines Agency (EMA) (Ducloyer et al., 2020). The results from the Luxturna® clinical trials demonstrated a statistically significant improvement in a person's ability to navigate low-light settings, with continuing improvements over a 2-year period (Ducloyer et al., 2020). However, this gene therapy is only applicable to recessive mutations in the *RPE65* gene and thereby restricts the pool of people with RP who can be treated using this therapy.

1.6 Gene editing as a candidate treatment

To date, no therapy has yet been found to be effective in fully restoring or slowing the progression of rod cell death. CRISPR-Cas9 is a gene editing system that has been proposed for difficult-to-treat hereditary diseases since its development in 2012. Its name is an acronym that stands for Clustered Regularly Interspersed Short Palindromic Repeats, the Cas9 standing for the enzyme that acts as the molecular "scissors". It was discovered as a gene editing tool in 2012 by Jennifer Doudna and Emmanuelle Charpentier in conjunction with multiple scientists from across the globe who worked on optimising it for use in human cells. In 1987 a team of Japanese researchers described a series of short, direct repeats that were interspersed with many short sequences in the genome of *Escherichia coli* (Doudna & Charpentier, 2014). Eventually, these sequences were identified as originating from different plasmids and viruses, which caused researchers such as Francisco Mojica to hypothesise that the enzyme that these sequences coded for was a rudimentary immune system capable of cutting the genetic sequences of invading pathogens and splicing it into its own genome for future immune-type regulation (Mojica et al., 1995). Its now proven process, reminiscent of the adaptive immune system in humans, operates as follows:

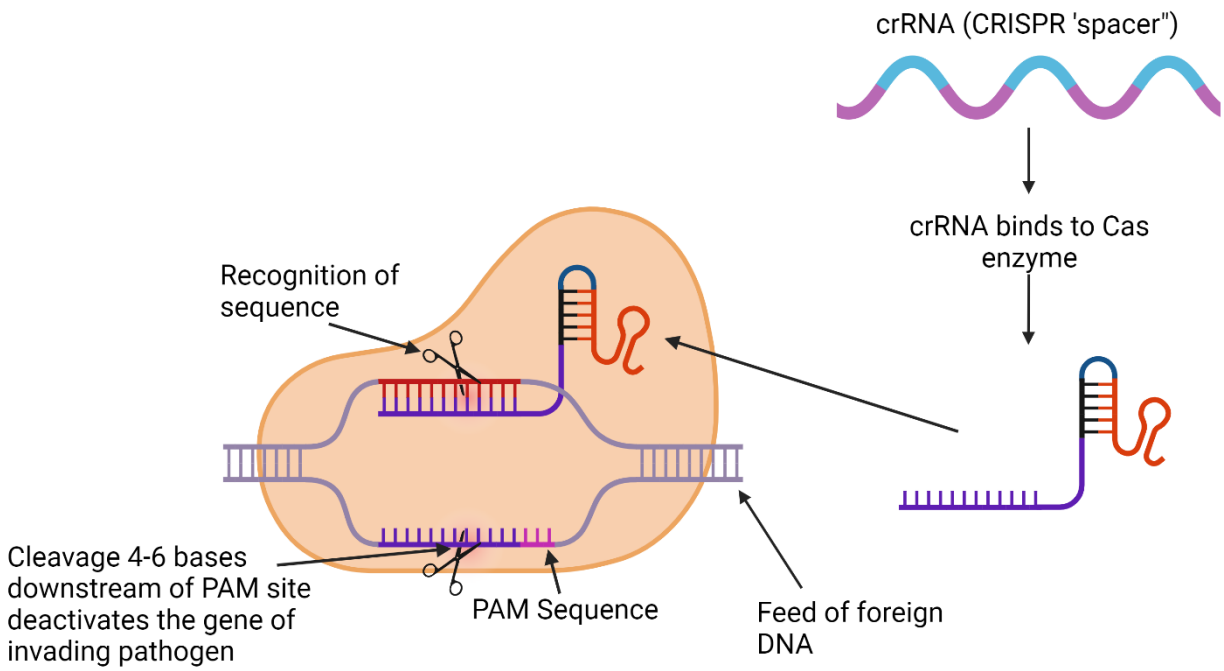


Figure 1.4: The native CRISPR Cas system. The CRISPR "spacer" array comprised of small fragments of foreign DNA can be transcribed into crRNA, which will then bind to a Cas enzyme. This then allows the catalytic centre of the complex to recognise the sequence by the PAM site, which should only be present on non-self DNA. The cleavage of the sequence at this site deactivates the gene of the invading pathogen, thereby protecting the host cell. Figure generated using BioRender.

This native process was then adapted for use in other cells by altering the "spacer" sequence that binds to the Cas enzyme (Doudna & Charpentier, 2014). To edit a gene, the CRISPR-Cas9 system requires what is known as a Protospacer Adjacent Motif (PAM) site – this is a short sequence of nucleotides that is located 4-6 nucleotides downstream of the gene of interest (Kleistiver et al., 2015). This motif operates as a marker to the CRISPR complex: once it has scanned through the DNA strands to find this sequence, it will cleave the DNA using the Cas9 enzyme (Doudna & Charpentier, 2014).

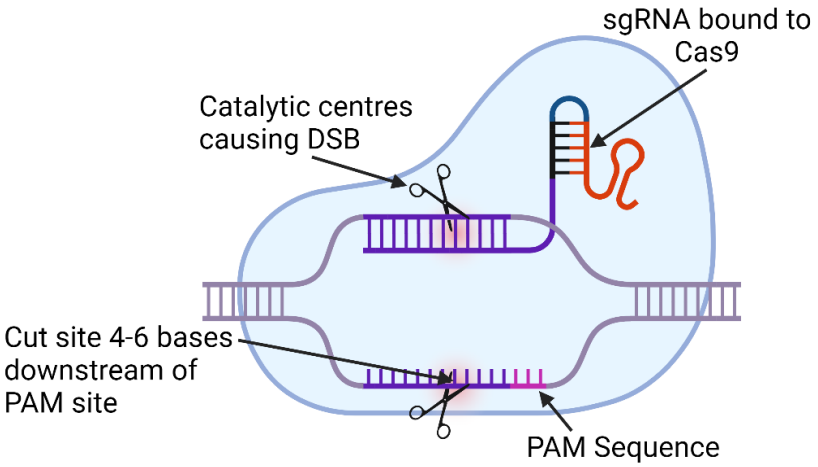


Figure 1.5: Adapting CRISRP-Cas to edit genes. Comparing the native system to the image above, the section of crRNA designed to target the desired section of DNA binds to tracrRNA to create sgRNA. This can then form a complex with the Cas9 enzyme called ribonucleoprotein (RNP) complex. When the CRISPR-Cas9 complex scans through DNA in cells, it will cleave 3-4 bases upstream of the PAM site within the guide sequence. **Figure generated using BioRender.**

1.6.1 Alternative species of Cas

There are a multitude of Cas enzymes which exist in a variety of bacterial and archaea species. Those that have been successfully adapted or utilised in mammalian systems are Cas9, Cas12a nuclease, and Cas3, all of which recognise slightly different PAM motifs (Pickar-Oliver & Gersbach, 2019). There are also a variety of Cas enzymes that have been manipulated to differ from the wild type or from different bacterial systems that will target different PAM sites to increase the level of flexibility in the system and thereby increase the specificity of the enzyme to its gene and corresponding PAM sequence (Walton et al., 2020). Research into alternate Cas9 species has produced variants such as SpGCas9 which recognises an NGN site, and SpRY which recognises an NRN site where R is either A or G (Liang et al., 2022). Collectively, these alternatives to the standard Cas9 NGG PAM site increase the variety and number of genomic sites that could be edited using this gene editing technology.

1.7 Model systems

Establishment of a biological or model system of a disease process or target can facilitate the development and testing of therapeutics. Murine models are often used because of low cost and rapid turnaround; however, they have their limitations, namely gene conservation, differences in body size compared to humans, and differences in brain complexity and physiological makeup (Jacobsen et al., 2010). Larger

animal models, such as sheep, provide a more accurate understanding of how a potential treatment will affect an organism that is similarly sized, and are physiologically and genetically closer to humans than mice (Pinnapureddy et al., 2015). The longevity of the animal can also provide a more accurate representation of how a disease progresses over time, particularly for disorders with late onset of symptoms (Jacobsen et al., 2010; Mckean, 2022).

1.7.1 Naturally occurring animal "models" of Retinitis Pigmentosa

Due to centuries of selective breeding by humans for specific traits, canines are one of the best examples of selective pressure and therefore carry a fascinating collection of genetic diseases caused by low genetic variation. There is compelling evidence in support of naturally occurring Retinitis Pigmentosa in certain dog breeds. Known as progressive retinal atrophy (PRA), the condition presents with similar clinical signs and pathogenesis to human RP, and requires similar diagnostic methods (Bunel et al., 2019). Like Retinitis Pigmentosa, there are multiple genetic mutations that are known to cause PRA in over 100 dog breeds, many of which have a human ortholog that causes Retinitis Pigmentosa or other retinopathies. Bunel et al. listed 16 genes that they had identified canine orthologs of (detailed in Appendix 2). For the vast majority of these genes, the human orthologs were estimated to cause less than 1% of RP cases with the notable exception of *RHO*, mutations in which are estimated to cause approximately 5.25% of RP cases (Appendix 1; Daiger et al., 2013).

1.7.2 Night blindness in sheep

Similar symptoms of Retinitis Pigmentosa have been found in a small flock of polled Wiltshire sheep in New Zealand, with 12 related animals of the 130 flock showing signs of night blindness between 2 and 3 years of age (Hunt et al., 2022). These sheep developed complete blindness by 4 to 5 years old and further examination of their retinas demonstrated loss of rod photoreceptor cells and attenuation of retinal blood vessels, similar to that seen in humans with Retinitis Pigmentosa (Hunt et al., 2022). Further to this, fundal examination revealed similar degeneration with bone-spicule-like formations in the superior aspect of the fundal images, similar in shape to retinal pigmented deposits seen in RP (Hunt et al., 2022). Breeding of two affected individuals resulted in the same disease phenotype among all progeny, indicative of an

autosomal recessive mode of inheritance (Hunt et al., 2022). Although the symptoms found in these 12 sheep cannot be conclusively proven to be the ovine form of Retinitis Pigmentosa, this paper demonstrates that mutations in sheep that cause shortening of the outer segments of photoreceptor cells can result in a similar disease phenotype in this species (Hunt et al., 2022). With this and other aspects of larger animal models in mind, modelling Retinitis Pigmentosa in sheep may be a productive and realistic aspect of treatment development.

1.7.3 Model Systems for Retinitis Pigmentosa

Most model systems of RP have focused on the RHO family of mutations, which have multiple modes of inheritance – very few systems have been developed for mutations in the RP1 gene and none are in large animal models. Mouse models generated by Gao et al. in 2002 demonstrated that disruption of the mouse ortholog of *RP1* (*Rp1*) truncates and disrupts the function of rod cells in the murine eye. Following homozygous knock out of the *Rp1* gene, researchers observed a progressive decrease in rod photoreceptor cells over a period of one year, with disruption in cone photoreceptors occurring after 10 months (Gao et al., 2002). They also observed shortened and morphologically abnormal rod photoreceptor cells that altered the structural integrity of outer segments that helped to illustrate the effects of mutations in this gene *in vivo* (Gao et al., 2002).

Liu et al. (2012) examined the effect of the Q662X nonsense mutation in exon 4 of *Rp1* in mice. Their research succeeded in producing a truncated RP1 protein. The resulting mice that were homozygous for this truncating mutation demonstrated signs of photoreceptor degeneration and cell death, demonstrating that truncating mutations in this region will cause disease pathology when mice are homozygous for the Q662X mutation.

Together, these models have helped to shed light on the role and function *RP1* in disease pathology, but as of yet, no group of researchers have generated a large animal model of this disease to test potential treatments. As mentioned in section 1.7, murine models can be limited due to morphological differences that make treatment development difficult. Notably, the size of a murine eye is significantly smaller than that of any larger animal models that might be created, complicating the testing of any treatment that may

be developed. Furthermore, the life span of a mouse is significantly smaller than that of a human, creating difficulties in understanding the molecular mechanisms of disease pathology from an animal model.

Due to the treatment gaps that exist for Retinitis Pigmentosa, this research project aimed to design and optimise a CRISPR-based approach in a fibroblast cell-model to ultimately generate a sheep model of *RP1*.

1.8 Proposed use of sheep fibroblasts

Immortalised sheep fibroblasts were used in this thesis to trial editing reagents for knocking-in the p.R677X *RP1* mutation by introducing a Homology-Directed Repair (HDR) template carrying the mutation. Fibroblasts were used due to the ease with which they can be cultured and sustained, as well as their availability in the lab. Fibroblasts proliferate quickly, which means that optimisation of editing experiments using CRISPR Cas9 can be expedited.

This thesis is primarily focused on demonstrating that the R677X mutation in the *RP1* gene can be induced in sheep fibroblasts which are able to survive in a hemizygous state following knock-out of the disease allele. Demonstrating that sheep fibroblasts can survive when they are hemizygous for *RP1* facilitates the development of a sheep model in the future by introducing the mutation into single-celled sheep embryos. It also investigates a possible treatment for some autosomal dominant mutations in *RP1*.

1.9 Gene annotation in South Australian Merino Sheep

Upon comparing the annotations of the RP1 genes from sheep and human (NCBI Gene IDs 106991348 and 6101 respectively), it became clear that the annotation of the sheep gene excluded coding regions for several hundred amino acids. This required completion of the sheep gene annotation as a first step in this project.

1.10 Research Aims

The research aims of this project are threefold:

Characterise the *Ovis aries* ortholog of the *RP1* gene. The identification of a full-length transcript of *RP1* in *O. aries* will facilitate accurate modelling of the desired mutation in fibroblasts by identifying available targets using genome editing tools.

Design and optimise CRISPR editing reagents for the *RP1* gene in immortalised ovine fibroblasts. This will optimise the delivery of CRISPR/Cas9 reagents and facilitate the development of an isogenic cell.

Introduction of *RP1* mutation in an *Ovis aries* cell line using HDR. Generating an isogenic cell line of an autosomal dominant truncating mutation like R677X will facilitate the testing of allele-specific knock-out of the mutated copy of the *RP1* gene.

2 Methods

General Materials

Table 2.1: Chemicals and reagents used in this research

Chemical/reagent	Supplier
Polymerase Chain Reaction and general lab work	
KAPA2G™ Enhancer (5X)	KAPA Biosystems
KAPA2G™ Buffer A (5X)	KAPA Biosystems
KAPA2G™ Robust DNA Polymerase (5U/µL)	KAPA Biosystems
Deoxynucleotide Triphosphates (dNTPs) 10mM	KAPA Biosystems
UltraPure™ Distilled Water	Invitrogen™
Gel Electrophoresis	
Molecular Biology Multi-Purpose Grade Agarose	Fisher BioReagents
10x Tris buffered EDTA	Invitrogen™
RedSafe Nucleic Acid Staining Solution	iNtRON Biotechnology
1kb Plus Ladder	New England BioLabs
Bacterial cell culture	
LB Agar (Lennox L Agar)	Invitrogen™
LB (Luria Broth Base)	Invitrogen™
Ampicillin	Thermo Scientific Chemicals
Cell Culture	
Foetal Bovine Serum, qualified, US origin	Gibco™ by ThermoFisher Scientific™
Dulbecco's Modified Eagle Medium (DMEM) with GlutaMAX™	Gibco™ by ThermoFisher Scientific™
Opti-MEM™ Reduced Serum Medium	Gibco™ by ThermoFisher Scientific™
Phosphate Buffered Saline (PBS), pH 7.4	Gibco™ by ThermoFisher Scientific™
Pierce™ Dimethylsufoxide (DMSO)	ThermoFisher Scientific™
Penicillin-Streptomycin (10,000 U/mL)	Gibco™ by ThermoFisher Scientific™
Ciprofloxacin	Sigma-Aldrich
Zeocin™ Selection Antibiotic	InvivoGen
Puromycin	Gibco™ by ThermoFisher Scientific™
Trypsin-EDTA (0.25%)	Gibco™ by ThermoFisher Scientific™
Trypan Blue Stain 0.4%	Invitrogen™
pmaxGFP™ (Green Fluorescent Protein) (1 µg/ µL)	Lonza Bioscience
Alt-R™ CRISPR-Cas9 crRNA	Integrated DNA Technologies
Kits	Supplier
QIAprep Spin Miniprep Kit	QIAGEN
Nucleo-Spin Gel and PCR Clean-Up Kit	Macherey-Nagel
DNeasy® Blood & Tissue Kit	QIAGEN
Lipofectamine™ 3000	Invitrogen™
Lipofectamine™ CRISPRMAX™ Cas9 transfection reagent	Invitrogen™
Ligation Sequencing Kit SQK-LSK114	Oxford Nanopore Technologies
Native Barcoding Kit 24 V14 (SQK-NBD114.24)	Oxford Nanopore Technologies
AMPure XP beads	Beckman Coulter

2.1 Identification of RP1 in Ovis aries

To establish the genomic structure and nucleotide sequence of the *RP1* ortholog in *O. aries* (ARS-UI_Ramb_v3.0 (GCF_016772045.2)), the messenger ribonucleic acid (mRNA) of each individual exon in *RP1* (6101) in humans (GRCh38.p14 (GCF_000001405.40)) was compared to the sheep genome using the blastx function in the Nucleotide Blast Tool on the NCBI website. This was compared exon by exon to identify which exons had homology to the sheep genome. These translated sequences were then compiled into one protein sequence and compared directly with the protein sequence in humans. The sequences that were compared were NP_006260.1 in humans and two separate loci from the sheep annotations in NCBI, LOC101114620 and LOC106991348 for exons 2 and 3, and 4 respectively.

2.2 Primer Design

The primers for this project were designed using the NCBI tool Primer-BLAST. Primers were designed to amplify the NC_056062.1: chr9:35316231-35316710 region of the *RP1* gene containing the intended mutation site at amino acid 676 (Appendix 5) and two different PAM sites for editing; the Cas9 PAM site and the SpG PAM site at amino acid The primers produce a 480bp product spanning NC_056062.1: chr9:35316231-35316710. These primers were also specifically designed to amplify a region of the genome that produces a product with an uneven distribution of nucleotides either side of the Cas9 PAM site so that if cleavage had occurred during an *in vitro* cleavage assay, two distinct bands could be visualised on a gel, with bands at 86 bp and 394 bp.

Table 2.2: Primers designed for the intended mutation site *RP1* in sheep

Primer pair 1	**Sequence (5'->3')**	**Template strand**	**Length (bp)**	**GC%**	**Self-complementarity**	**Self-3' complementarity**
Forward primer	TGGCTTGCCA CAGACTACAT	Plus	2	50.0	6.00	2.00
Reverse primer	GTCCCTTTGA GGGGACTACG	Minus	20	60.0	8.00	2.00
Product length (bp)	480					

Figure 2.1 shows the full length of the PCR product that is amplified by this primer set and the location of these primers in relation to the intended mutation site and the PAM sites of the two species of Cas9, the guides for which are detailed in section 2.

NC_056062.1:chr9:35316231-35316710
TGGCTTGCCACAGACTACATCAGAAAACTCCATTGTGGAGGAAGGAATAGTTGATAATGTCACAGCAGACAACAAAGCTA
GGGTCAGGAATTTAAGAACTTATGGTAACACCGATGATAGATCCAGCCCTTTCTTAGCAGATGCAGCTCATTCTTCAAGT
AACAACTCTGGAACTGACAAAACTATTTCCAAGACCCCAGCTTCAGTAGGATCCTCTACTGTCACTACAAGAATCGACCA
ACTGATTCATGAATTTTCTCAGTGTGGTTTAACAAAACTTCCAGAAAATGAAAAGCAGATTTCATCTTCAGTTGATAGCA
AAAAAAAGATGAAATCTCAGCAGCATGTGATAAATTCTCAGCATCAGGCTGGAGAGATGGCAACTAAAAGAATCCCCAGG
AAGAATAAGAGAATGAACACAAGAGGTAGAATTGCACAGGAAACCATATTGCGAGATTCACGTAGTCCCCTCAAAGGGAC

Key:
Purple is a forward or reverse primer
Orange is the site of the mutation
Turquoise is the SpG PAM site
Green is the Cas9 PAM site

Figure 2.1: Annotated nucleotide sequence of the region being amplified by the designed primers. This figure shows the nucleotide sequence of the 480-base region of the sheep genome that is amplified by the primers that were designed. The forward primer (highlighted in purple at the beginning of sequence) is located at chr9:35316231-35316250, and the reverse primer (end of sequence) is located at chr9:3531661-35316710. The mutation site that would replicate the R677X human mutation is highlighted in orange (chr9:35316569), the SpGCas9 PAM site (chr9:35316569-35316571), which runs on the antisense strand, is highlighted in Turquoise and the Cas9 PAM site (chr9:35316629-35316631) is in green.

2.3 Polymerase Chain Reaction

Polymerase chain reactions (PCR) optimised for the designed primers were performed using the following protocol and amounts for single reactions:

Table 2.3: Single PCR reactions

KAPA 2G™ Buffer A	5 µL
dNTPs (10mM)	0.5 µL
Fwd primer (10µM/ µL)	1.25 µL
Rvs primer (10µM/µL)	1.25 µL
KAPA2G™ Taq Polymerase	0.1 µL
H_2O	6.9 µL
DNA (10 ng/µL)	5 µL
KAPA 2G™ Enhancer	5 µL

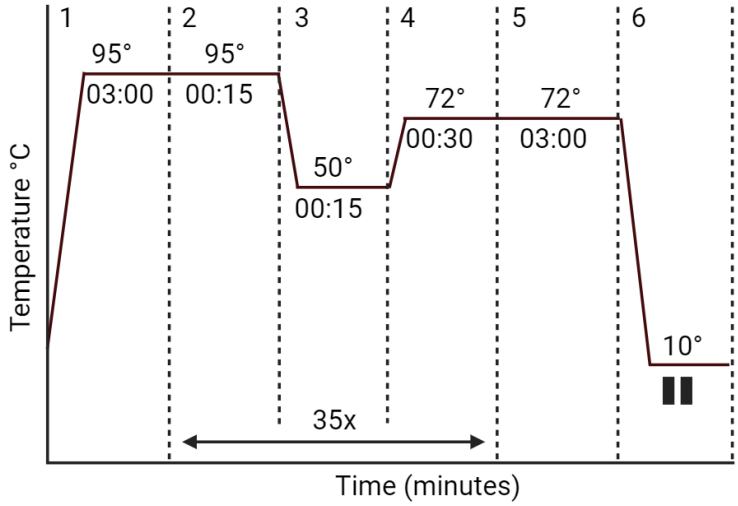


Figure 2.2: PCR conditions This figure shows the PCR conditions optimised for the primers designed in section 2.2. Steps 2-4 are the cycles that were optimal for high yield of these specific amplicons. Figure generated using BioRender.

PCRs were typically run for 35 cycles using the Kapa2G Robust PCR system. A corresponding KAPA enhancer at 20% of the reaction volume was used to improve the quality and yield of PCR products.

2.4 DNA Gel Electrophoresis

Gels were used to check for the presence of DNA in samples, determining the length of fragments generated by enzyme digests and to identify whether amplification of DNA had occurred prior to sample sequencing. They were also used for an *in vitro* cleavage assay that confirmed whether the standard Cas9 PAM site could be cleaved by the guide that had been designed. Gels were made using 10x TBE buffer containing 1.0M Tris, 0.9M Boric Acid and 0.01M EDTA (Invitrogen™) diluted to 1x with Ultrapure H_2O and adding a weight of Molecular Biology Multi-Purpose Grade Agarose to create the desired % density gel (Fisher BioReagents). The amount of Agarose used varied to create gels between 1 and 2%, with 1 gram corresponding to 1% of the total volume. These were combined with RedSafe Nucleic Acid Staining Solution (iNtRON Biotechnology, South Korea) at a concentration of 2 µL per 100 mL in order to visualise

the DNA using a BioRad Molecular Imager (Gel Doc XR+). Kapa loading dye (Kapa Biosystems) was used for loading samples at 20% of the total sample volume in each well and 1 µL of the 1kb Plus DNA Ladder (New England BioLabs) was used as the standard DNA ladder. All gels were run at 120 V in Bio Rad gel tanks using 1x TBE buffer.

2.5 PCR Product Purification

PCR products were purified prior to sequencing using the Nucleo-Spin Gel and PCR Clean-up (Macherey-Nagel) as per manufacturer's instructions with a few minor changes. To purify samples, 1 volume of PCR product was added to 2 volumes of Buffer NTI and mixed thoroughly (for example, 50 µL of sample was added to 100 µL of Binding Buffer) then placed in a spin column for centrifugation. The spin columns contain a silica membrane to which the DNA will adhere, allowing other particulates to be removed in the supernatant. The column was placed in its collection tube and was spun at 11,000 x *g* for 30 seconds and the flow-through discarded. To reduce chaotropic salt carry-over, the silica membrane was washed with 700 µL of Buffer NT3 and spun at 11,000 x *g* for 1 minute. This wash step was repeated, and the spin column was added to a new collection tube and spun for 1 minute at 11,000 x *g* to dry the silica membrane, ensuring there was no buffer that had adhered to the sides of the spin column. Once the membrane is dry, the spin column was placed in a 1.5 mL Eppendorf tube and 20 µL of UltraPure™ Distilled Water (Invitrogen™) was added directly to the membrane. The addition of UltraPure™ Water causes the DNA to pull away from the column, as the backbone structure of DNA is hydrophilic. The columns were spun at 11,000 x *g* for 1 minute, then the elute was collected and placed on the silica membrane to spun down a second time on the same settings. The increase in spinning times for NT3 steps and spinning down the flow-through for the final elution step significantly reduced contaminants and improved the DNA yield. For samples with concentrations too low for sequencing, these were concentrated using a Savant DNA120 SpeedVac Concentrator.

2.6 DNA Quantification

All DNA and purified PCR products in this project were quantified on a NanoDrop™ 1000 Spectrophotometer (NanoDrop™ Technologies, Wilmington, USA). Samples were measured against 1µL of their diluent and then 1 µL of sample was measured. Samples were measured by "blanking" the

NanoDrop™ 1000 Spectrophotometer with 1 µL of their diluent to set the baseline, then 1 µL of sample was measured. The NanoDrop™ 1000 Spectrophotometer also provided information about contamination; low 260/280 absorbance ratios indicated that samples were contaminated with common reagents in DNA cleanup kits, such as phenol and guanidine (ThermoFisher Scientific™, 2012). Samples were also quantified for sequencing using the Qubit 4 Fluorometer (ThermoFisher Scientific™) and the High Sensitivity double-stranded DNA (dsDNA) kit. The kit uses two standards for dsDNA samples and 1 µL of sample is quantified in 200 µL of solution. This provided a more accurate DNA quantification than NanoDrop™ would prior to MinION sequencing to ensure sample concentrations were accurate prior to sequencing. Due to the cost of reagents associated with Qubit quantification, the NanoDrop™ was a cost-effective method for quantifying and identifying contaminants in most samples that were generated in the course of this thesis while Qubit quantification was reserved for the final step of MinION sequencing.

2.7 Sequencing

2.7.1 Sanger Sequencing

Sanger Sequencing was conducted by Kristine Boxen at The University of Auckland DNA Sequencing Facility, Genomics Centre. Sanger Sequencing uses multiple copies of DNA samples of different lengths that are fragmented and tagged with fluorescently labelled di-deoxynucleotides. Samples are amplified via PCR to produce millions of samples with di-deoxynucleotides that cause the chain of DNA to be terminated at different lengths within the same sample. The fragments are separated by length using gel electrophoresis while a computer detects the fluorescence of each nucleotide at the end of each DNA fragment and compiles the signals into a sequence. When these fragments are compared to a reference sequence, such as the region of the DNA that the primers span, any differences between the sample and reference sequence will be identified. This facilitates the identification of Single Nucleotide Polymorphisms (SNPs).

2.7.2 MinION Sequencing

Oxford Nanopore Technology sequencing was undertaken for a more in-depth sequence using the MinION SQK-LSK114 chemistry. The library construction was conducted according to manufacturer's instructions, which comprises of cleaning, barcoding, and reading PCR product from the desired region of the DNA sample. Once prepared, the adapter known as the Flongle, was primed, and flushed, then the library was

loaded as per manufacturer's instructions. The process of library preparation involved samples that have been cleaned via a PCR cleanup kit (Macherey-Nagel), then undergo the process of A-tailing, which facilitates barcoding in future steps. Samples are then cleaned using AMPure XP beads (Beckman Coulter), barcoded via DNA ligation, cleaned again using AMPure XP beads before the samples were pooled and quantified using the Qubit 4 Fluorometer (ThermoFisher Scientific™). The prepared samples were then loaded into the Flongle and sequenced using the programme MinKNOW. The Flongle recognises adapters attached to the end of the DNA strands within the sample, which facilitates pulling that strand through the pore. This process causes the disruption of an electrical current specific to each base. If samples have been successfully barcoded, the barcoded section at the beginning of the sample's sequence allows sequences to be categorised to differentiate between different samples in the same pool. All sequences were aligned to the reference sequence by Dr Andrew Jiang and analysed in Geneious Prime to determine the level of editing that may have occurred.

2.8 Cell Culture

2.8.1 Media

Media was comprised of 10% Foetal Bovine Serum (Moregate Biotech) and Dulbecco's Modified Eagle Medium (DMEM) with GlutaMAX™ from Gibco™. For culturing the South Australian Merino Immortalised Foetal Fibroblasts, the cells required specific antibiotics (Zeocin™ – InvivoGen) and Puromycin (Gibco™) in the media for selection. Penicillin-Streptomycin (Gibco™) was added to make up 1% of the media, while Puromycin (Gibco™) was at a concentration of 0.02% and Zeocin (InvivoGen) at 0.1%. Cell media was changed every second day. All media contained a pH indicator that would turn yellow when the media became too acidic. Cells were incubated at 37°C with 5%CO_2 injected constantly to create an acid-base buffer. This combination of DMEM, FBS and antibiotics will be referred to as Full Media.

2.8.1.1 Antibiotic-containing media for Mycoplasma treatment

For media that was used to treat infection with Mycoplasma, Ciprofloxacin (Sigma-Aldrich) was added to 0.1% of media volume and changed every 48 hours. After 14 days of this treatment, the cells were returned to normal media and tested for mycoplasma after a week using the MycoStrip™ (InvivoGen) dip test.

2.8.1.2 Freezing Media

Appropriate media was required for storing cells in a long-term Liquid Nitrogen Dewar. This freezing media was comprised of 50% FBS, 40% DMEM and 10% Dimethylsulfoxide (DMSO) to prevent the formation of ice crystals that could perforate the cells during freezing.

2.8.2 Cell Passaging and Plating

Cells were passaged when they had reached approximately 90% confluency by detaching them from the surface of a cell culture flask by washing them with PBS and adding up to 5 mL of Trypsin-EDTA (0.25%), phenol red (Gibco™) dependent on the size of the flask or plate. The Trypsin was deactivated using an equal volume of full media, then the number of cells that were required for seeding the size of plate or flask were removed. They were then spun down in 15mL Falcon tubes for larger volumes or 1.5mL Eppendorf tubes for smaller volumes at approximately 200 x *g* for 5 minutes. The supernatant was removed then the cells were washed again with up to 5mL of PBS, spun down and resuspended in warmed Full Media for replating. Cells were generally passaged as a ratio of 1:20 of 1:10 depending on the time that larger volumes of cells were needed for transfections. Cells would be washed and centrifuged to remove any traces of Trypsin in the media before resuspension in warmed Full Media and plating in 12-well, 24-well, 96-well and T25 or T75 flasks.

2.8.3 Cell Counting

Cells were counted on a Countess II FL (Life Technologies) using Countess™ chamber slides (Invitrogen™) using 10 µL of cells suspended in Full Media and 10 µL of Trypan Blue stain 0.4% (Invitrogen™).

2.8.4 Cell Freezing

After detaching the cells from the plate as described in section 2.6.2, and determining their concentration per mL, approximately 1 million suspended in the deactivated Trypsin/media was transferred to a 15 mL Falcon tube and centrifuged at 200 g for 5 minutes to form a cell pellet. The pellet was washed with PBS (Gibco™), the supernatant removed, and the pellet resuspended in freezing media at a temperature of 37°C. Vials were immediately transferred to a Mr Frosty™ container (ThermoFisher Scientific™, Waltham,

USA) and stored in a -80°C freezer for gradual cooling. After 24 hours the cell vials were transferred to the LN_2 Dewar for long-term storage.

2.9 CRISPR/Cas9 system

2.9.1 Cas9 enzyme

Alt-R® CRISPR-Cas9 crRNA was manufactured by Integrated DNA Technologies (IDT) and was used for guides with an -NGG PAM site. This Cas9 was used in varying concentrations, detailed in section 2.12 to create a gradient for determining optimal concentrations for editing.

2.9.1.1 Guides and guide design

Guides were designed using CRISPOR (Tefor Infrastructure) and were selected to be both as close to the mutation site as possible and as efficient as possible with fewest off-target effects. They were also selected to deliberately avoid two PAM sites that had not been successful in prior experiments for creating an *RP1* cell line in the Snell Lab. A section of 1000 bases of the genomic sequence (NC_056062.1: 35316011-35317010) was analysed by CRISPOR. This involved the input of a portion of the genomic sequence CRISPOR to identify appropriate guide sites. CRISPOR also scans for and identifies potential off-target effects based on the portion of the DNA that has been input and included information of what off-target effects would occur in each frame (Concordet & Haeussler, 2018). Guide sites are ranked via colour using green for recommended sites and red for sites the programme suggests avoiding. In Table 2.4, the guide that was used in this research was selected due to its reduced off-targets and closeness to the intended mutation site, which is illustrated in the CRISPOR output detailed in Figure 2.3.

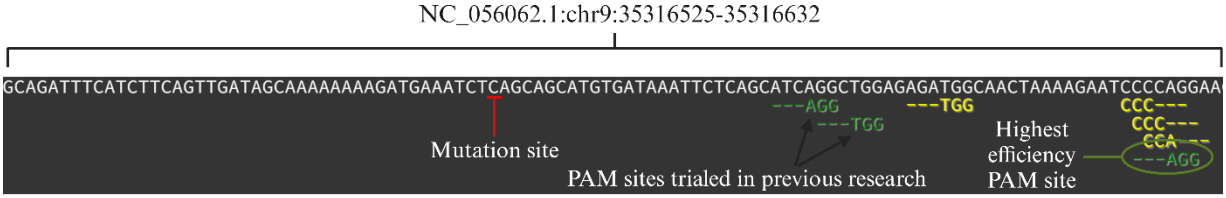


Figure 2.3: Output from CRISPOR showing available PAM sites in the 1000 base sequence that was used for identifying Cas9 guides. The intended mutation site, indicated with a red line, has a limited selection of PAM sites available nearby for cutting nearby. Two of the closest PAM site, H685 and Q686, have already been trialled in editing experiments and resulted in no deletions or base changes being detected. The closest PAM site with high efficiency and a low number of off target positions in the sheep genome is circled in green and is 55 bases downstream of the site of interest.

Table 2.4: Cas9 guide output from CRISPOR

Position	*Guide sequence + PAM + restriction enzymes*	*Off-targets for 0-1-2-3-4 mismatches*
586 / fw	TGGCAACTAAAAGAATCCCC *AGG* Enzymes: *StyD4I, LpnPI, BseDI, BstNI, MboII*	0-0-1-17-137 155 off-target positions

2.9.1.2 HDR Template design

Homology-Directed Repair (HDR) templates were designed using the Alt-R HDR design tool on the Integrated DNA Technologies (IDT) website. Templates included two premature stop codons and 5 silent SNPs, which do not change the amino acid sequence, are spread across the length of the template to aid in identifying edited reads after sequencing. There is an asymmetrical distribution of nucleotides either side of the desired mutation site as recommended by Schubert et al. (2021) in their optimsation paper on HDR repair for Cas9. If incorporated in cells, this template would generate the intended human R677X mutation.

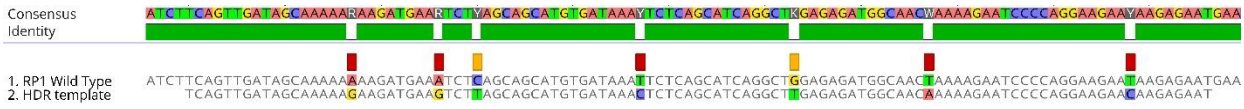


Figure 2.4: HDR Template for RP1 in sheep. This template spans codons 665 to 700 and is 107 bases long and corresponds to chr9:35316539-35316645 in the sheep reference genome. Silent SNPs, which do not change the amino acid sequence, have been introduced and are marked red, while premature stop codons, the leftmost of which is the stop codon we wish to replicate, are labelled in yellow. The first sequence is the wildtype sequence and below it is the HDR template with the base changes that have been made. **Generated by Geneious version 2022.1 created by Biomatters.**

2.9.1.3 *In vitro* cleavage assay

An *in vitro* cleavage assay was performed to determine whether the guide that had been designed would cleave at the intended PAM site in a PCR product amplifying NC_056062.1: chr9:35316231-35316710. This also helped to confirm that the correct section of the genome had been amplified with the primers that were designed. In this assay, 1.6 µL of Cas9 stock (62 µM/L) was diluted to 6.67 µM/L by adding 13.4 µL of Duplex Buffer (5 µM/L). The sgRNA was generated by combining 1 µL of the desired guide (at a concentration 100 µM/L), 1 µL of tracrRNA and 18 µL Duplex Buffer. This solution was heated to 95°C for 5 minutes, then cooled on the benchtop. A gradient of RNP complex concentrations was used, ranging from 175 nM to 2000 nM in the final volume of solution, as detailed in Table 2.5. Samples were added to

200 ng of PCR product and incubated in a Thermocycler at 37°C for 60 minutes, then heated to 80°C for 10 minutes. Once these incubations were complete, 2.74 µL of Proteinase K (QIAGEN) was added to each sample reaction which were then incubated for a further 10 minutes at 65°C. Samples were then run on a 1.5% agarose gel using the same loading dye and 1kb Plus ladder as described in section 2.2 and visualised on a BioRad Molecular Imager.

Table 2.5: volumes of reagents for *In vitro* cleavage assay

Reagent	175nM	250nM	500nM	1000nM	2000nM	Negative control
UltraPura H_2O	Up to 20 µL	Up to 20 µL	Up to 20 µL	Up to 20 µL	Up to 20 µL	Up to 20 µL
Duplex Buffer	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL
SgRNA	0.6 µL	1.2 µL	2.4 µL	4.8 µL	9.6 µL	-
Cas9 (at 6.7 µM/L)	0.38 µL	0.75 µL	1.5 µL	3 µL	6 µL	-
DNA	200ng	200ng	200ng	200ng	200ng	200ng

2.9.1.4 Cas enzyme with altered PAM site

A variant species of Cas9, called SpG was also trialled in this research to get a closer cleavage site to the locus of interest. SpGCas9 was developed by Walton et al. (2020) and was designed to recognise an -NGN site. This guide site was also designed in CRISPOR shown in table 2.6. For this project, SpGCas9 was contained in a plasmid manufactured by Genscript and was delivered as a plasmid within *E. coli*.

Table 2.6: SpGCas9 guide output from CRISPOR

Position	*Guide sequence + PAM + restriction enzymes*	*Off-targets for 0-1-2-3-4 mismatches*
560 / rev	GCTGAGAATTTATCACATGC *TGC* Enzymes: *BstDEI, NIaIII, ApeKI, Fsp4HI, XceI*	0-1-5-43-597 646 off-targets

2.10 Bacterial cell culture for plasmid harvesting

The gene encoding the SpGCas9 enzyme was contained within the SpG-HF1 Cas9 LentiCas9-Blast plasmid comprised of a lentivirus backbone with multiple promoter and antibiotic resistance genes that allow for

selection (Fig. 2.5). The *E. coli* containing the plasmid was sampled and streaked on a 90mm petri dish containing solid LB Agar (Lennox L Agar) (Invitrogen™) and Ampicillin (Gibco™) at a concentration of 100 µg per mL. Working under a Bunsen burner to create a sterile environment, the bacterial cells were sampled using a flame loop, passed briefly through the flame of a Bunsen burner, and streaked on the plate, then left in an incubator (Infors AG-CH-4103 Bottmingen) at 37°C for 24 hours. A single colony was then harvested and placed in a 50 mL falcon tube containing 20 mL of sterilised LB Broth (Invitrogen™) containing 20 µL of ampicillin was placed on a shaking platform in the same bacterial incubator at 150 rpm for a further 24 hours.

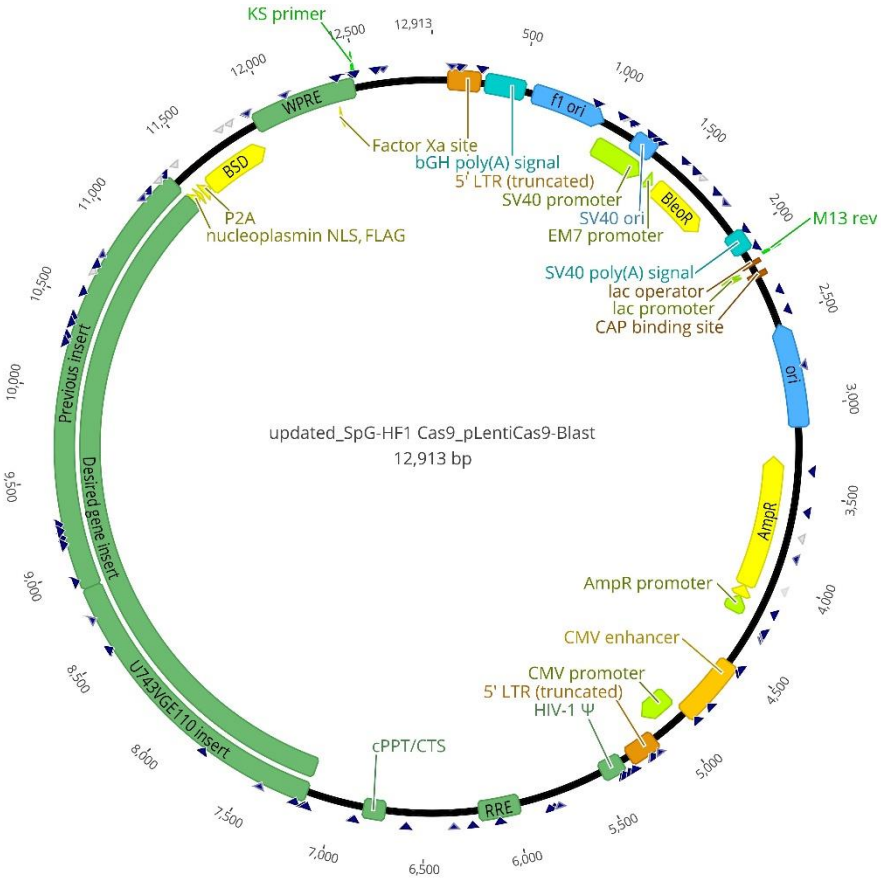


Figure 2.5: SpG-HF1 Cas9 pLentiCas9-Blast plasmid map. This map shows the plasmid that was manufactured for the Snell lab by GenScript with the site of the gene insert labelled "Desired gene insert". The gene is contained on a lentivirus backbone with an Ampicillin resistance gene for selection. **Generated by Geneious version 2022.1 created by Biomatters.**

2.10.1 Plasmid Purification

The following day the plasmids within the bacterial cells were harvested using the QIAprep Spin Miniprep Kit (QIAGEN) as per manufacturer's instructions. Briefly, the entirety of the overnight culture was pelleted at 8,000 rpm (6800 x *g*) for 3 minutes, then the bacterial cells were resuspended in 250 µL of Buffer P1 and transferred to a microcentrifuge tube. 250 µL of Buffer P2 was added then the tube was inverted until the solution turned from cloudy to clear. Once the cells had lysed, the sample was centrifuged for 10 minutes at 13,000 rpm (~ 17,900 x *g*). 800 µL of supernatant was then transferred to a spin column and was centrifuge for 30-60 seconds until the flow-through could be discarded. The plasmids, which were bound to the silica membrane of the microcentrifuge tube, were then washed using 500 µL of Buffer PB, centrifuged to discard the flow through, then were washed with 750 µL of Buffer PE. The spin column was then transferred to a new collection tube and was centrifuged at 13,000 rpm (~17,900 x *g*) for 1 minute to remove any residual wash buffer. The plasmid samples were then eluted in 50 µL UltraPure™ Water. They were run on a 2% agarose gel after a plasmid digest using the *Hin*dIII enzyme to positively identify the plasmid.

2.11 Plasmid Digest and Identification

In order to confirm that the correct plasmid had been purified, approximately 500ng of purified plasmid was digested using the *Hin*dIII enzyme (Roche) and compared to an existing elution of SpGCas9 that had been purified by Dr Alex Trevarton in the Snell lab. 500 ng of each sample of plasmid DNA was incubated with 5 units of the *Hin*dIII enzyme in an Eppendorf® Mastercycler® Nexus Thermal Cycler. Samples were then loaded onto a 2% agarose gel (as described in section 3.3.1). Digestion of the SpG-HF1 Cas9 pLentiCas9-Blast plasmid with the *Hin*dIII enzyme produces five fragments of 556, 584, 1020, 5030 and 5541 base pairs long.

2.12 Transfection of sheep fibroblasts for editing

2.12.1 Transfection with Cas9

Transfection protocols were provided by Dr Natasha Mckean and Dr Victoria Hawkins and were adapted from previously successful editing experiments for this project. The first protocol, provided by Dr Natasha Mckean, was for transfecting cells in a 24-well cell culture plate, while the second protocol, provided by

Dr Victoria Hawkins, used a reverse transfection method in a 96-well plate then cells were grown out to larger plates and flasks for analysis and storage. These protocols will be called A and B for Dr Mckean's and Dr Hawkins' recommendations, respectively.

2.12.1.1 Protocol A

In protocol A, the RNP complex was prepared on the bench top by making up an sgRNA complex to 3.33 pmol/µL by adding the following volumes of reagent to a 1.5 mL Eppendorf tube: 1 µL of guide (crRNA), 1 µL tracrRNA and 28uL of Duplex Buffer. This was placed in a heat block at 95°C for 5 minutes then allowed to cool the benchtop for 10 minutes. This was spun down briefly, then stored on ice until use.

Using a fume hood, 1 µL of Cas9 stock solution was diluted to 2.48 pmol/µL by adding 24 µL of Duplex Buffer, then was placed on ice until ready to use. The volumes of reagents detailed in table 2.6 were added to individual Eppendorf tubes for each experimental condition and total to 25 µL for each condition.

Table 2.7: Experimental summary for protocol A using Cas9

Reagent	*1.9 pmol*	*3.9 pmol*	*7.8 pmol*	*15.6 pmol*	*31 pmol*	*Negative control*
OptiMem	21.1 µL	19.7 µL	17.1 µL	11.5 µL	0.7 µL	25 µL
Cas9 (2.48 pmol)	0.8 µL	1.6 µL	3.1 µL	6.3 µL	12.5 µL	-
sgRNA (3.33pmol/ µL)	0.6 µL	1.2 µL	2.3 µL	4.7 µL	9.3 µL	-
Cas9 Plus Reagent	2.5 µL	2.5 µL	2.5 µL	2.5 µL	2.5 µL	2.5 µL

In a separate tube, 175 µL of OptiMem was added to 10.5 µL of the CRISPRMAX™ reagent and the 26.5 µL was immediately added to the tubes for each condition and incubated for 10 minutes at room temperature. Once the incubation step is complete, 50 µL of solution was added to the corresponding experimental well. After 24 hours, the full media on each well was removed, the cells were washed with 1 mL of warmed PBS and then 1 mL of fresh media was added. The cells were closely observed until they could be grown out into a T25 flask.

2.12.1.2 Protocol B

In protocol B, higher concentrations of the RNP complex and Cas9 were used in a 96-well plate. The fluorophore (ThermoFisher Scientific) was used as a transfection control for the RNP complex. The transfection was performed when enough cells had been grown for the population to be split. Each well required approximately 60,000 cells to reach a confluence of approximately 80% at transfection.

RNP complexes were prepared at a 1:1 ratio between crRNA and tracrRNA. 0.9 µL of crRNA was added to 0.9 µL of tracrRNA in an Eppendorf tube, then was heated to 95°C for 5 minutes. The solution was then cooled on the bench or 10 minutes and spun down. This was split into two concentrations of guide complex by adding 1 µL of this solution to 9 µL of Duplex buffer to create a 5nM solution. The remaining 0.8 µL of solution remained undiluted at 50nM. These two tubes were then placed on ice until ready for use. 0.5 µL of stock Cas9 was added to 4.5 µL of Duplex Buffer to create a working solution of 6.2uM. The volumes of the reagents detailed in table 2.7 were added to 4 individual Eppendorf tubes with a total volume of 25 µL.

Table 2.8: Experimental summary for protocol B for reverse transfection using Cas9

Reagent	*Transfection control*	*100nM*	*500nM*	*ATTO control*
sgRNA	-	1.2 µL (5 µM)	0.6 µL (50 µM)	-
ATTO-tracrRNA (100 µM)	-	-	-	1 µL (100 µM)
Cas9 Plus Reagent	-	0.5 µL	0.5 µL	0.5 µL
Cas9	-	1.62 µL (at 6.2uM)	0.807 µL (at 62.uM)	-
OptiMEM	25 µL	20.48 µL	22.49 µL	23.5 µL

In a separate tube, 98 µL of OptiMEM and 2 µL of CRISPRMAX™ was combined and then 25 µL was added immediately to each experimental condition. These complexes were incubated at room temperature for 20 minutes.

While the RNP complexes are assembling in the experimental tubes, a cell split was performed on the fibroblasts as detailed in section 2.8.2 and counted as detailed in section 2.8.3. The volume of solution to obtain the desired number of cells to seed 4 wells (240,000 cells) was calculated and that volume was

pelleted via centrifugation at 200 x *g* for 5 minutes. The cells were then washed and resuspended in 200 µL of warmed full media. Once the incubation period was complete, 50 µL of both the RNP complex solution and the resuspended cells was pipetted into the corresponding well. The media was changed the following day in the same manner as in Protocol A using 100 µL of warmed PBS and full media.

2.12.2 Transfection with SpGCas9

Due to the SpGCas9 enzyme requiring replication time in the cells, cells were transfected in two separate instances between 24 and 48 hours apart. In preparation for transfection, cells were plated according to the instructions in section 2.8.2 and seeded in a 24-well plate at a confluence of ~30%, or approximately 40,000 cells per well – the total volume for this experiment was 320,000 cells for 5 experimental conditions and 3 controls. Once the cells had reached approximately 40% confluence, 500ng SpG plasmid was transfected using the Lipofectamine 3000 protocol.

Three controls were used; one well contained GFP to confirm transfection of SpGCas9, one was transfected with lipofectamine twice but had no guide added, and one well contained wild-type cells with nothing but media changes at the same time as the experimental wells. For the GFP control in this experiment, 0.5 µL of GFP was added to 25 µL of OptiMEM and 1 µL of the P3000 in a 1.5 µL Eppendorf tube. In a separate Eppendorf tube, 5.1 µL of the SpG plasmid elution, which had a concentration of 587.5 ng/µL, was added to 144.9 µL of OptiMEM and 6 µL of the P3000 reagent to a total volume of 156 µL. In a third Eppendorf tube, 200 µL of OptiMEM and 12 µL of Lipofectamine were combined, then 159 µL were combined with the plasmid solution, leaving 53 µL of lipofectamine remaining for two controls. 26.5 µL of the remaining lipofectamine mix was added to the GFP control and then 23.5 µL of OptiMEM was added to make the final lipofectamine-only control up to a volume of 50 µL. All solutions were incubated for 10 minutes at room temperature. The controls were then added to their corresponding wells and 52.5 µL of the SpG and lipofectamine solution was pipetted onto the cells in a dropwise fashion.

After 24 hours had elapsed, the media was discarded, the cells were washed with 1 mL of warmed PBS and then 1 mL of full media was added. The cells were transfected again within a 48-hour period depending on

how healthy they were and if they were growing. When they reached a stage of approximately 50% confluence and were dividing well, they were transfected again with the sgRNA.

Like Protocol A, the sgRNA was formed by adding 1 µL of crRNA, 1 µL of tracrRNA and 28 µL of Duplex Buffer, was heated to 95°C for 5 minutes, then cooled on the benchtop for 10 minutes before being spun down and placed on ice until use.

The reagents detailed in table 2.7 were compiled in individual tubes for each experimental condition until the solutions were ready to be applied to the cells.

Table 2.9: Transfection of the SpGCas9 guide

Reagent	*1.9 pmol*	*3.9 pmol*	*7.8 pmol*	*15.6 pmol*	*31 pmol*	*Negative control*
OptiMEM	21.9 µL	21.3 µL	20.2 µL	17.8 µL	13.2 µL	25 µL
sgRNA	0.6 µL	1.2 µL	2.3 µL	4.7 µL	9.3 µL	-
Cas9 Plus Reagent	2.5 µL	2.5 µL	2.5 µL	2.5 µL	2.5 µL	2.5 µL

In a separate tube, 175 µL of OptiMEM was added to 10.5 µL of the CRISPRMAX™ Reagent, then 25 µL was added to the corresponding tube for each experimental condition. These were incubated at room temperature for 10 minutes, then all 50 µL in each tube was added to the corresponding well.

In all transfection experiments, cells were grown out over time in cell culture plates and flasks of increasing growth area until the population had reached at least 1.5 million cells in a T25 flask. Cells were then trypsinised and counted, as detailed in section 2.8.2, in preparation for freezing down aliquots of ~1 million cells for long-term storage in LN2 (section 2.8.4). All remaining cells at this stage underwent DNA extraction.

2.13 DNA extraction

DNA was extracted according to the DNeasy Blood and Tissue Kit (Invitrogen™) protocol. The portion of cells for extraction was pipetted into a 15 mL Falcon tube and was washed with 5 mL of warmed PBS before being pelleted at 300 x *g* for 5 minutes to remove any remaining Trypsin. They were then resuspended in 200 µL of PBS and 20 µL of Proteinase K was mixed by gently pipetting up and down.

Once the Proteinase K had be thoroughly mixed in, 200 µL of Buffer AL was added using the same method of mixing. The Falcon tubes were then incubated at 56°C in a water bath with a shaker for 10 minutes. Samples then had 200 µL of 100% ethanol added and were vortexed to create a homogenous solution. The entire mixture was placed in a DNeasy Mini spin column and placed inside a collection tube. Samples were then centrifuged at 6000 x *g* (8000 rpm) for 1 minute. The flow-through was discarded and the spin column placed in a new collection tube. 500 µL of Buffer AW1 was then added and the samples were centrifuged on the same settings for the same length of time and the flow-through was discarded. In a new collection tube, 500 µL of Buffer AW2 was added to the spin column and the sample was centrifuged at 20,000 x *g* (14,000 rpm) for 3 minutes. Carefully removing the column from the collection tube to prevent contact with the flow-through, the column was then placed in a clean centrifuge tube and centrifuged at the same speed for 1 minute to remove any residual ethanol. The spin column was placed in a 1.5 mL microcentrifuge tube, 100 µL of Buffer AE was added directly to the membrane of the spin column and was incubated at room temperature for 1 minute. The spin column was then centrifuged at 6,000 x *g* (8000 rpm) for 1 minute to elute the DNA. This incubation then centrifugation step was repeated after adding the flow-through from the microcentrifuge tube to the membrane of the spin column again. This final re-elution step, as mentioned in section 2.5 for PCR product purification, was undertaken as it consistently produced higher DNA yields.

3 Results

3.1 Annotation of *RP1* in *Ovis aries*

An aim of this thesis was to characterise the *RP1* gene in *Ovis aries.* To ensure we were targeting the correct position in the *O. aries* genome, the full-length ortholog of the *RP1* gene needed to be identified. An initial search in the National Centre for Biotechnology Information (NCBI) 'Gene' database for the *RP1* ortholog identified a gene called 'RP1 axonemal microtubule associated (ID: 106991348) with an annotated transcript: XM_015097752.4. This transcript had 61% similarity to exon 4 of the human *RP1* gene (ID: 6101) when aligned using the nucleotide blast function (blastn) of the NCBI BLAST web page. The predicted protein for the sheep gene (XP_014953238.3) contained the same open reading frame, but coded for a protein that was 1860 amino acids long and did not contain exons 1, 2 or 3.

It was clear that, if the gene (ID: 106991348) was the sheep ortholog of *RP1*, the annotation was missing exons 1-3 which needed to be identified in the *O. aries RP1* sequence. The genomic structure of *RP1* is highly similar across multiple taxa, with the orthologs of the species examined in the cross-species alignment all being comprised of one non-coding exon and three coding exons. It is therefore reasonable to assume that the structure of the *RP1* gene in sheep will be the same as humans. Individual searches using the human *RP1* mRNA sequence (NM_006269.2) for exons 1, 2 and 3 using the translated nucleotide function (blastx) identified an uncharacterised gene (ID: 101114620) that had similarity to exons 2 and 3. Two of the exons in this uncharacterised gene had a percent identity to the human *RP1* exons 2 and 3 of 79.5% and 96.5%, respectively.

Table 3.1: Human and ovine *RP1* mRNA and protein NCBI accession numbers

Species	*NCBI gene ID*	*Genomic Location*	*mRNA accession number*	*Protein accession number*	*Protein length (aa)*
Human	6101	NC_000008.11 (chr8:54559185-54871234)	NM_006269.2	NP_006260.1	2156
Sheep	101114620 (exons 2 and 3)	NC_056062.1 (chr9:35213176-35418176)	XM_060393411.1	XP_060249394.1	846
	106991348 (exon 4)	NC_056062.1 (chr9:35315208-35322281)	XM_015097752.4	XP_014953238.3	1860

In all species that were analysed in the cross-species alignment, *RP1* exon 1 is untranslated (5'-UTR), and different species have varying levels of similarity to the human sequence for exon 1. A sequence with similarity to human exon 1 was initially unable to be located in sheep when the entirety of the human mRNA for this exon (107 bp) was used. Therefore, the search was refined by attempting to identify subregions with high similarity between humans and sheep using a small section of 42 nucleotides extending from the 5' nucleotide of human *RP1* exon 1. This sequence was input using the low stringency setting against the latest assembly of the ovine genome (NC_056062.1 Chromosome 9 Reference ARS-UI_Ramb_v3.0 Primary Assembly). This identified a small section of 17 nucleotides of the sheep genome upstream of both genes (ID: 101114620 and ID:106991348) with 100% identity (as seen in Figure 3.1). Viewing this section of nucleotides in NCBI's Genome Data Viewer, the 100 nucleotides of genomic DNA immediately 3' were copied and the resulting 117 nucleotide sequence was compared with exon 1 using CLUSTALW. This comparison returned a region of DNA 96 bases in length (NC_056062.1 chr9:35307169-35307264) with a similarity score to the human exon 1 of 61.4% (Figure 3.1).

```
Sequence 1: Ovis aries    96 bp
Sequence 2: Homo sapiens  107 bp

Sequences (1:2) Aligned. Score: 61.4583%
                     5'
1. O. aries          AGTTTCTTTGCACGAAACAAACAGGTTACATATCCAGTAACATTTGTGAGGATTATTGAA
2. H. sapiens        AGTTTCTTTGCACGAAATGA---GGTTACATATCCAGTGACATTTATTTGAGCTATTTAA
                     ****************  *   *************** ****** *  *   **** **
                                                                    3'
1. O. aries          CT---TGAGCCATCTTTTT-----------AAAGGAAGTTTCAAGTTCTG
2. H. sapiens        ACAACTTAAACATCTTTTTCTTTTCTTAATAAGGGACGTTTCAAGTTGTG
                         * *  *********           ** *** ********** **
```

Figure 3.1: CLUSTALW alignment of the human RP1 exon 1 with 117 nucleotides extracted from the sheep genomic sequence that showed homology to the 5' region of the exon. The output of the alignment is shown above, with 96 bases having identity to the human exon 1. Exon 1 in sheep is 14 nucleotides shorter than exon 1 in humans. These deletions (outlined in red) are the likely cause of why this exon was unable to be located by searching for the entirety of the human exon 1 in the sheep genome.

This approach was successful in identifying the ovine exon 1 sequence completing the annotation of the gDNA of the ovine *RP1* gene. An alignment of the human cDNA and inferred ovine cDNA is included in Appendix 6. We propose that the compilation of the genomic region identified in figure 3.1 and the genes 101114620 and 106991348, shown in Figure 3.2 is the most likely sheep ortholog for *RP1*. Table 3.1 contains the accession numbers of the sheep genes that were used to generate Figure 3.2.

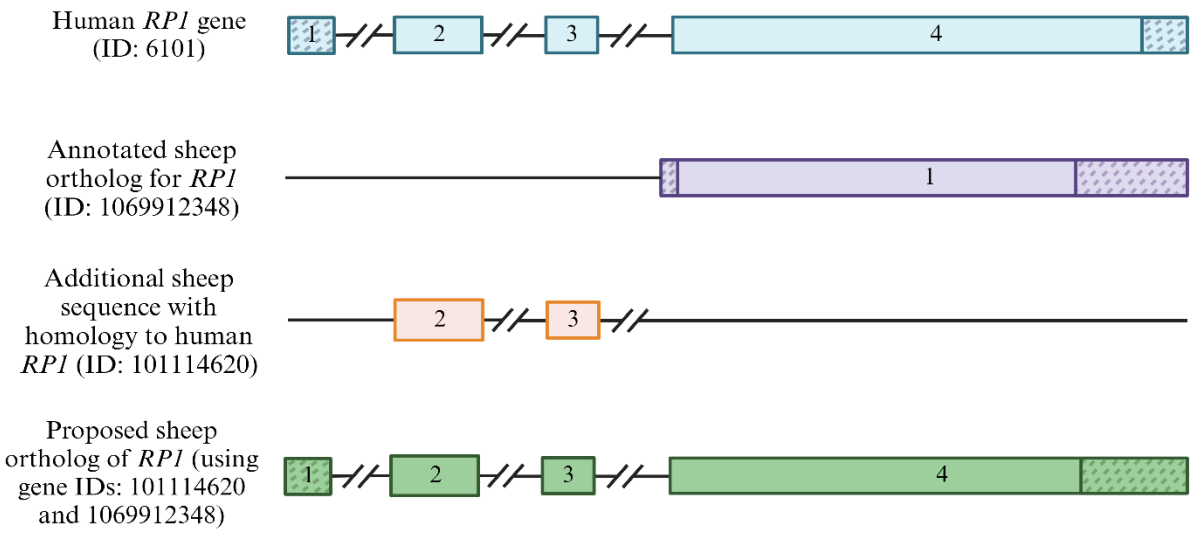


Figure 3.2: Schematic comparison of Human RP1 gene, two potential sheep orthologs and the proposed compiled ortholog. The human *RP1* gene (blue) compared to the NCBI suggested annotation (purple), and the additional sequence that had homology to the human *RP1* exons 2 and 3 determined via a translated nucleotide blast using the NCBI blastx tool (orange). The section of 96 nucleotides that is hypothesised to be exon 1 is included upstream of exons 2 and 3. Non-coding regions are cross hatched. The completed *RP1* sheep ortholog proposed in this thesis is detailed in green and includes all exons that were identified as having 67% similarity to the human annotation for *RP1*. **Figure generated using BioRender.**

3.1.1 Identifying the intron/exon boundaries in the ovine *RP1* gene

To further support the assertion that the gene we compiled is the sheep ortholog of *RP1*, we compared the position of dinucleotide motifs in the genomic DNA of the human gene to the sheep genome using NCBI's Genome Viewer genome (NC_056062.1 Chromosome 9 Reference ARS-UI_Ramb_v3.0 Primary Assembly). These dinucleotide motifs indicate the 5' and 3' ends of introns and are motifs that allow the pre-mRNA of a transcribed gene to undergo intron splicing before being translated into protein. These motifs were identified visually, and the genomic location of the resulting intron/exon boundaries is shown in Table 3.2:

Table 3.2: predicted intron/exon boundaries of the proposed complete annotation of *RP1* based on genomic dinucleotide positions in *O. aries*

Nucleotide sequence type	*Genomic location*
Exon 1	NC_056062.1 (chr9: 35307169-35307264)
Intron 1	NC_056062.1 (chr9: 35307265-35311678)
Exon 2	NC_056062.1 (chr9: 35311679-35312302)
Intron 2	NC_056062.1 (chr9: 35312303-35313536)
Exon 3	NC_056062.1 (chr9: 35313537-35313708)
Intron 3	NC_056062.1 (chr9: 35313710- 35315326)
Exon 4	NC_056062.1 (chr9: 35315327:35320860)

To illustrate these findings, Figure 3.3 depicts the location of these dinucleotide motifs within the wider structure of the proposed sheep gene.

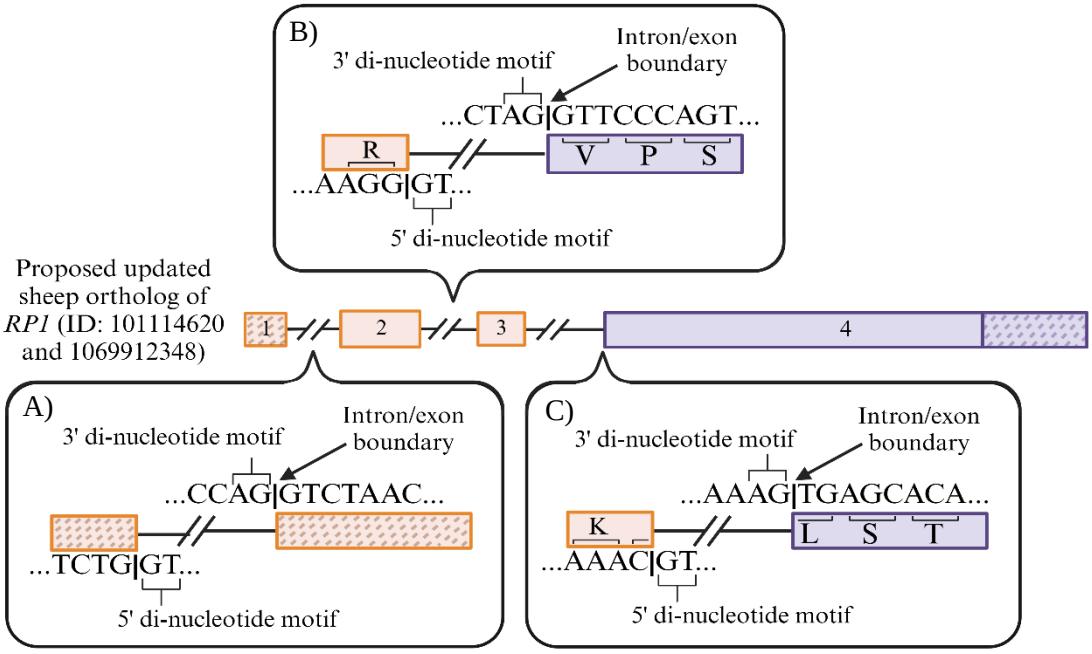


Figure 3.3: Identification of dinucleotide splicing motifs and proposed gene structure for sheep ortholog of RP1. A) shows the splice motifs identified between exons 1 and 2, **B)** shows the splice motifs between exons 2 and 3, **C)** shows the splice motifs between Exon 3 and 4, which are annotated in genes 101114620 and 1069912348 respectively. **Figure generated using BioRender.**

3.1.2 Summary of the sheep ortholog RP1 gene

The gene model of the sheep ortholog of human *RP1* that we compiled contains the four exons characteristic of the human *RP1* gene and has 67% similarity to the human *RP1* gene. This model sequence also contains the expected dinucleotide motifs that are required for mRNA splicing. The sheep gene model intron/exon

structure is similar to RP1 orthologs in the multi-species alignment in Figure 1.1. We used this gene model to determine whether we could generate a cell model of the R677X truncating mutation that causes Retinitis Pigmentosa in humans.

3.2 Sequence characterisation of the gene editing target region of exon 4 of *RP1* in *Ovis aries*

A multi-animal analysis was undertaken to determine whether there are any Single Nucleotide Polymorphisms (SNPs) that might interfere with the recognition of the guide sequence in both the immortalised foetal fibroblast cell line and other South Australian Merino sheep for any future work in generating a sheep model. The position of the R677X mutation in the sheep ortholog was identified in exon 4 at position 2023 (Appendix 4) and primers were designed for this region as described in section 2.2. These primers were used to amplify a 480 base amplicon from this region in DNA samples from 10 individual sheep, and the fibroblast line using PCR. The amplicons were resolved by agarose gel electrophoresis (Figure 3.4).

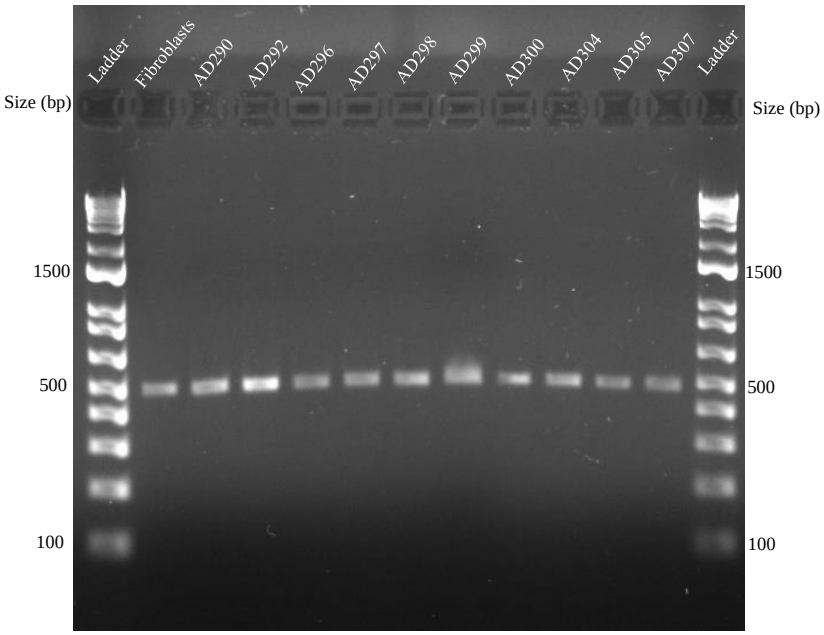


Figure 3.4: Electrophoretically resolved DNA amplicons from 10 sheep and the immortalised fibroblast cell line incorporating the target ovine RP1 exon 4. The image above shows the DNA for SNP analysis that was amplified and subsequently purified for Sanger sequencing. Individual sheep are labelled 'AD' with an animal number following. All samples amplified well and were the expected length (480 bp) produced by the primers designed in section 2.2.

The residual PCR product was sequenced using Sanger sequencing. These Sanger sequences were compared to the sequence of the sheep ortholog compiled in section 3.1 using Geneious Prime. This analysis revealed two SNPs in this amplicon upstream of the expected mutation site at nucleotide 2023 in sheep (Appendix 4) that encode a different amino acid than the reference genome sequence. A Cytosine to Adenine (C>A) change at nucleotide position 2000 (22 bases upstream of the mutation site) encodes a substitution of Aspartic Acid at amino 667 for Alanine. Of the 11 DNA samples sequenced, three sheep and the fibroblast line were heterozygous for this SNP, while the remaining seven sheep were homozygous for Aspartic Acid. The second polymorphism was a Cytosine to Thymine (C>T) change at nucleotide position 1967 (55 bases upstream of the mutation site) encoding a substituion of Proline in the reference sequence to Leucine. Four sheep were heterozygous for this SNP, while the remaining six sheep were homozygous, three of which carried the SNP. The immortalised fibroblast line did not contain this SNP.

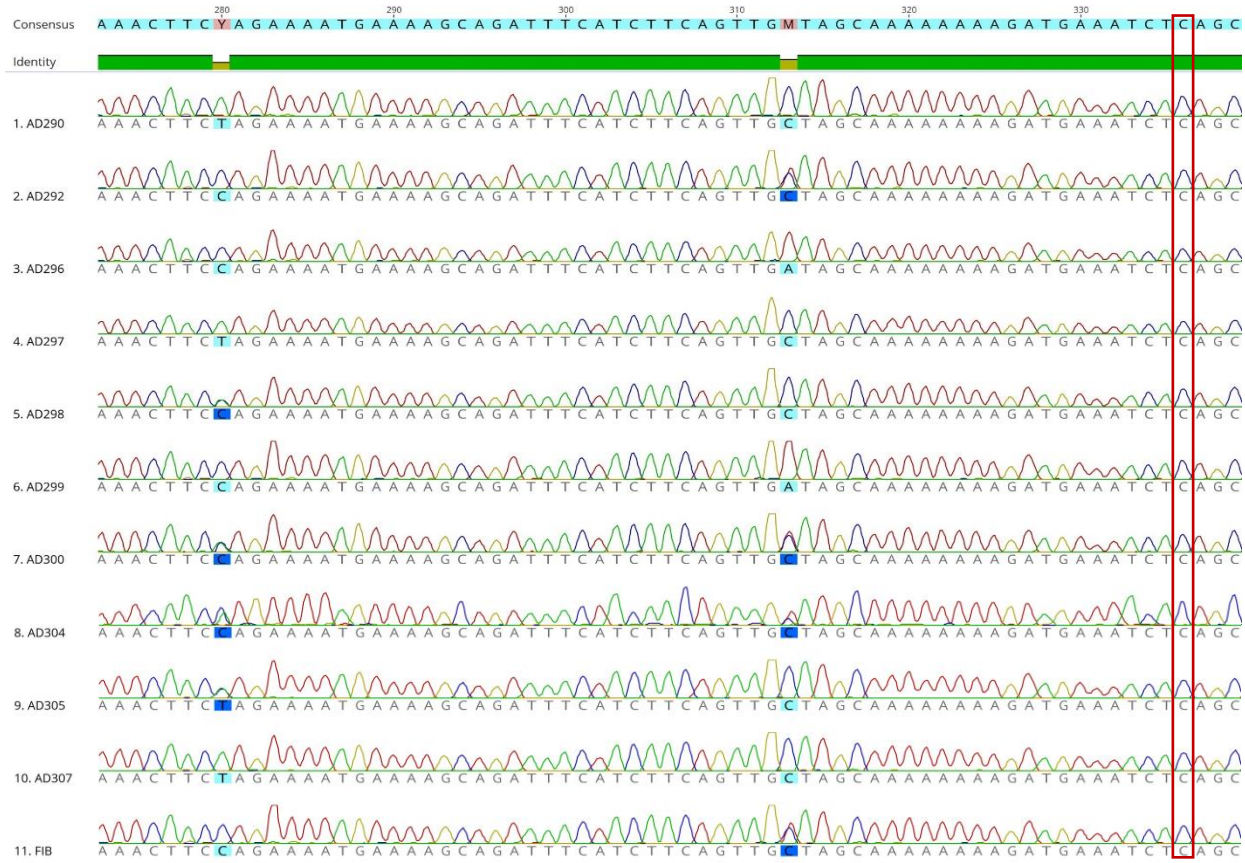


Figure 3.5: Sanger sequences of 10 South Australian Merino sheep. These images show the amplicons of 10 different sheep using the primers described in section 2.2. In each sample, each sheep is labelled 'AD' with an animal number following. The fibroblast line that was used was also sequenced to identify any SNPs in the region. These traces show two common SNPs identified 22 and 55 bases upstream of the corresponding position for the R677X mutation in humans is outlined in red. Heterozygotes are in dark blue and homozygotes are in light blue. **Generated by Geneious version 2022.1 created by Biomatters.**

The corresponding positions of both SNPs in the human *RP1* protein were searched in gnomAD (V4.0.0) which identified that the SNP at sheep position 2000 is polymorphic in humans (gnomAD V4.0.0 p.Ala669Asp, variant ID: 8-54625888-C-A) so is potentially a neutral substitution in sheep. No polymorphisms were identified for the syntenic position of the SNP at nucleotide position 1967. The SNPs that we identified should not interfere with the proposed guide binding sites as they are not contained within either the SpGCas9 H678 guide or the Cas9 I694 guide.

3.3 Using the alternate Cas enzyme SpGCas9 with a less specific PAM site to target the desired edit site

Previous experimental work in the Applied Translational Genetics (ATG) group to induce the R677X coding mutation using Cas9 in sheep fibroblasts was unsuccessful. The lack of editing that was seen is

theorised to be due to the location of the PAM sites of the two guides that were trialled, H685 and Q686 (Figure 2.3). To reduce the distance between the cut site of the guide and the expected mutation site, we used a guide for the alternate species of Cas, SpGCas9, which recognises an NGN PAM site (Figure 3.6). Unlike the delivery of Cas9 and its guide as an RNP Complex, SpGCas9 is not available as a purified protein, therefore SpGCas9 was expressed in the target cells through the delivery by transfection of a plasmid. The SpGCas9 gene was encoded on a lentivirus backbone SpG-HF1 Cas9 pLentiCas9-Blast plasmid (Figure 2.3) and was transfected, rather than transduced. The guide for SpGCas9 was delivered through an additional transfection 24-hours later to allow time for the cells to express the enzyme prior to the introduction of the guide.

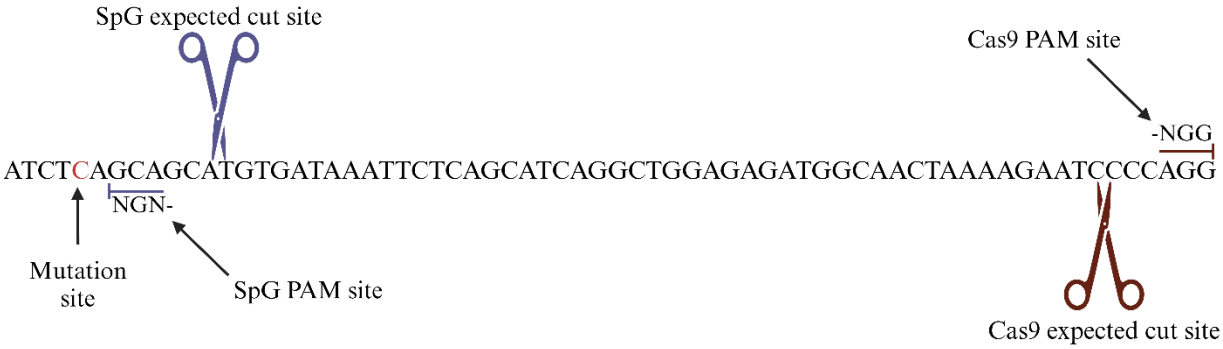


Figure 3.6: Schematic of expected cut sites for the SpG enzyme and Cas9 enzyme. This figure shows the location of the mutation site in respect to the expected cut sites of the two enzymes that were used. The PAM site for SpGCas9 is directly beside the mutation site, with the cut site occurring 7 bases downstream. This is significantly closer than the Cas9 PAM site that used the most efficient guide (I694) sequence with fewest off-target effects for this region of the gene. This is shown in red and is 55 bases downstream of the mutation site. **Figure generated using BioRender.**

3.3.1 Amplification purification and characterisation of the SpG-HF1 Cas9 pLentiCas9-Blast plasmid

Before transfecting fibroblast cells with the SpG-coding plasmid, we needed to amplify and isolate the DNA for the SpG-HF1 Cas9 pLentiCas9-Blast plasmid (labelled "query" in Figure 3.3). Purified plasmid DNA from separate colonies were digested using the *Hin*dIII enzyme to confirm the correct banding pattern and hence the integrity of the plasmid. Previous research in this lab has used this plasmid as a lentivirus to generate stably integrated cells. A portion of the plasmid DNA from that research was provided by Dr Alex Trevarton to use as a positive control. Five DNA fragments are expected following digestion of SpGCas9 with *Hin*dIII; 556 bp, 584 bp, 1020 bp, 5030 bp and 5541 bp. In this digest, we observed bands at the expected lengths and despite two pairs of bands (556 and 584, and 5030 and 5541) being difficult to distinguish due to the similarities in length, the positive control demonstrated the expected sizes.

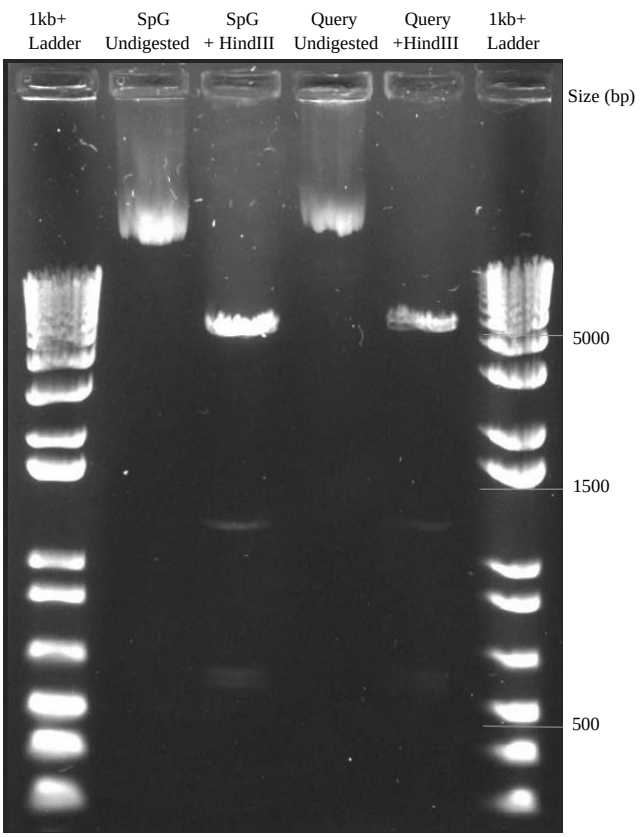


Figure 3.7: Restriction digest of SpG and query plasmid using the HindIII enzyme. This gel compares 500 ng of a known isolation of the SpG plasmid in lanes 1 and 2 and 500 ng of the plasmid that was isolated from *E. coli* for this thesis, labelled "query" in lanes 3 and 4. The desired fragments lengths are 556bp, 584bp, 1020bp, 5030bp and 5541bp, which can be faintly seen in a similar pattern to the control digest.

3.3.2 Plasmid transfection optimisation

Along with the SpG-HF1 Cas9 pLentiCas9-Blast plasmid, a positive control pmaxGFP™ (Green Fluorescent Protein) was also transfected to indicate that transfection was successful prior to administering the SpGCas9 guide. The pmaxGFP™ plasmid was used as the positive control and an experiment was run according to the Lipofectamine™ 3000 protocol to determine the optimal concentration of DNA and of Lipofectamine prior to transfecting cells with the SpG-HF1 Cas9 pLentiCas9-Blast plasmid. Because SpGCas9 requires at least 24 hours for the enzyme to be transcribed and translated prior to the introduction

of the guide, the cells in this optimisation experiment were transfected at the same confluency (~40%) as they would when the SpGCas9 plasmid would be introduced. Although the size of the GFP plasmid is much smaller than SpGCas9, at 3,486 bp compared to 12,913 bp. this control was useful for determining that the Lipofectamine™ 3000 kit would transfect any measurable plasmid as expected. Once complete, the cells were imaged using a fluorescence microscope and the efficiencies of the transfection quantified by counting using a manual cell counter (Figure 3.8). Approximately 40,000 cells per well were transfected in a 24-well plate then the transfection gradient was run according to the manufacturer's recommendations.

The lipofectamine was tested at two volumes: 0.75 µL and 1.5 µL, labelled L1 and L2 respectively. The DNA amount was 250 ng, 500 ng and 1000 ng, D1, D2 and D3. The combination of these concentrations generated six experimental conditions as seen in Figure 3.8.

Table 3.3: Transfection efficiency of pmaxGFP™ (Green Fluorescent Protein) in immortalised sheep fibroblasts quantified from images in from Figure 3.8

Fluorescent level	**L1D1**	**L1D2**	**L1D3**	**L2D1**	**L2D2**	**L2D3**
Bright	0.51%	1.75%	0.92%	1.59%	4.06%	1.9%
Medium	0.85%	4.97%	1.39%	3.06%	6.23%	4.3%
Faint	4%	11.1%	4.78%	11.45%	25.47%	10.2%
Transfection efficiency	*4.5%*	*17.8%*	*7.1%*	*16%*	*36.2%*	*16.4%*

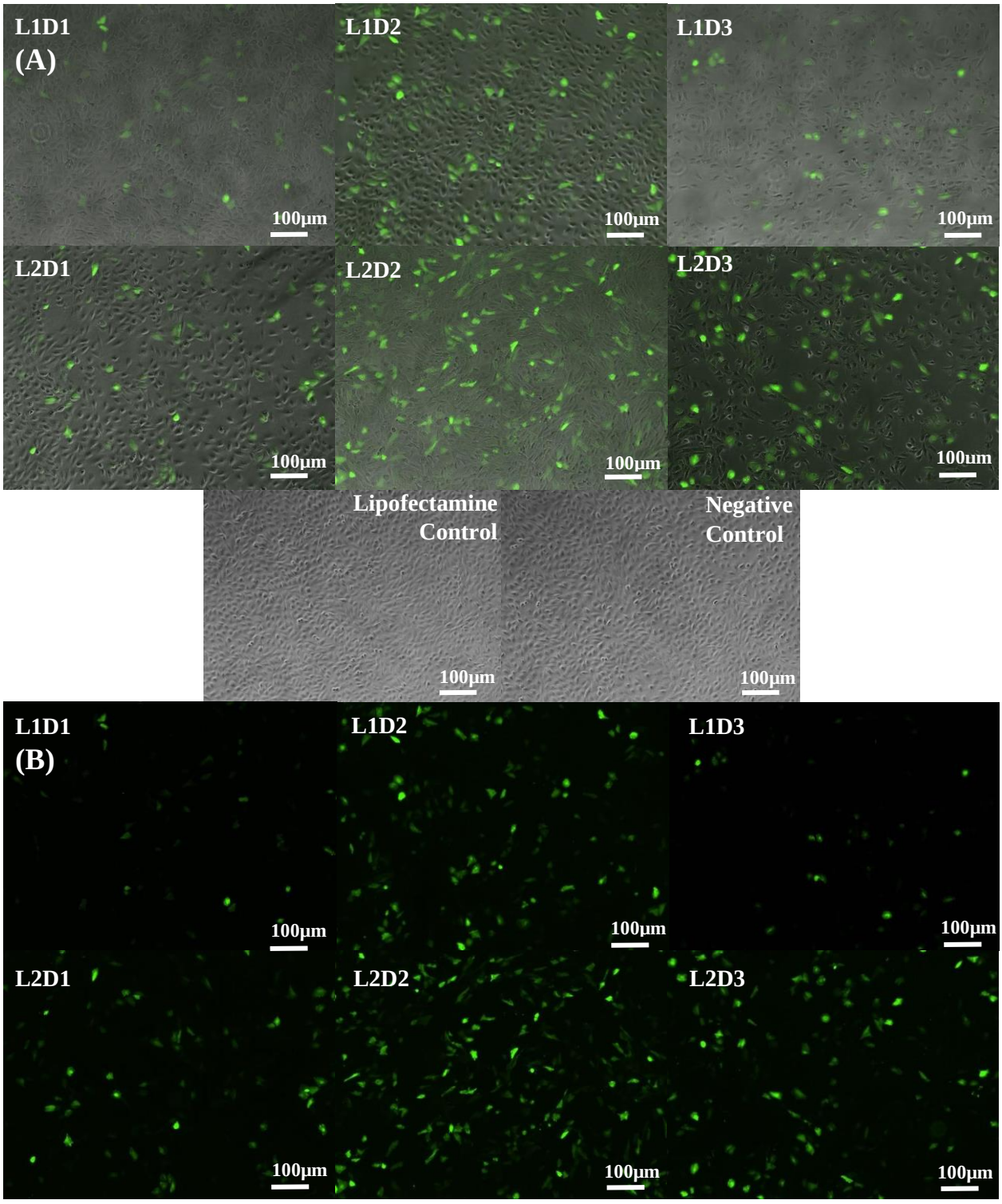


Figure 3.8: GFP gradient test for optimising plasmid and lipofectamine concentrations. These images show fibroblasts that were transfected with varying concentrations of both Green Fluorescent Protein (GFP) and Lipofectamine viewed under fluorescent and brightfield microscope. Images labelled "L1" contain 0.75 µL of lipofectamine and "L2" contain 1.5 µL of lipofectamine. "D1" contains 250 ng of GFP DNA, "D2" contains 500 ng and "D3" 1000 ng. This GFP plasmid was used as a control for transfection of the SpGCas9 plasmid to check the efficiency of the Lipofectamine 3000 protocol. Controls consisted of a Lipofectamine-only control with no GFP added, and a negative control.

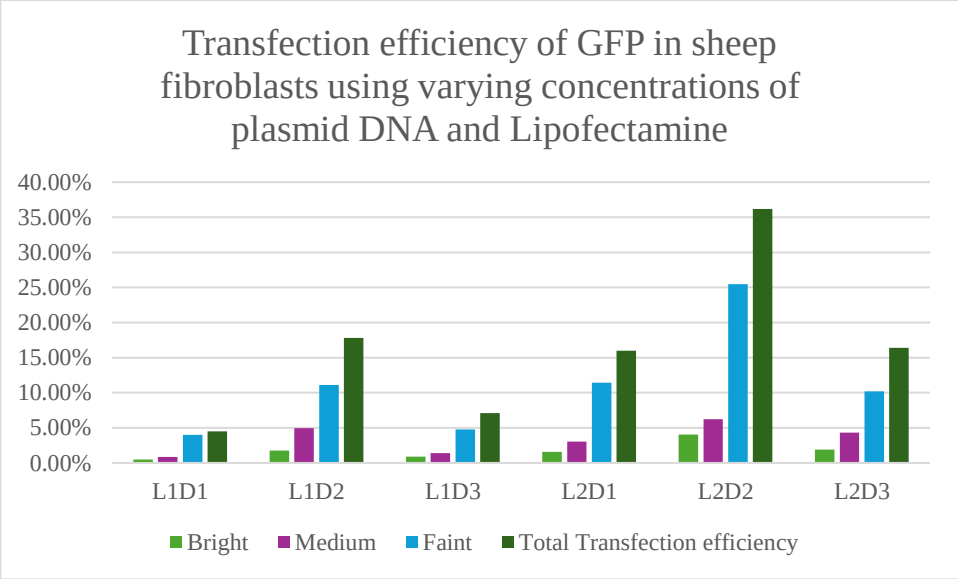


Figure 3.9: Graph of transfection efficiency of South Australian Merino Immortalised Foetal Fibroblasts with GFP using Lipofectamine™ 3000. This graph shows the transfection efficiency of GFP in cells using increasing concentrations of plasmid DNA SpG-HF1 Cas9 pLentiCas9-Blast plasmid (D1:250, D2:500 and D3:1000 ng), and of lipofectamine (L1: 0.75 µL and L2: 1.5 µL). The combination that worked most efficiently was L2D2, with a total transfection efficiency of 22.5%. The cells were transfected in a 24-well plate and at the same confluence that we would transfect the SpGCas plasmid.

The transfection gradient ranging from 250 ng to 1000 ng of plasmid DNA (Figure 3.8) indicated that the most efficient transfection of in sheep fibroblasts was 1.5 µL of lipofectamine and 500 ng of plasmid DNA, labelled L2D2 in Figures 3.8 and 3.9. This was therefore the amount that was used during transfections of the SpG plasmid.

3.3.3 MinION sequence analysis of SpGCas9 edited cells showed no indication of editing

As detailed in section 2.12.2, cells were transfected using the optimal lipofectamine volume of 500ng of the SpGCas9 plasmid and 1.5 µL of Lipofectamine™ 3000, then were transfected again 48 hours later with the guide using CRISPRMAX™.

The genomic DNA from transfected cells was then extracted and the targeted region amplified using the primers designed in section 2.2. The PCR products were purified (section 2.5) and sequenced using the MinION protocol (section 2.7.2). Visualising the MinION reads using Geneious Prime revealed less than 5% deletions and nucleotide at the expected H678 cut site (Table 3.4). Similar changes were observed in the negative control, so it was concluded that no editing had occurred, or that the editing frequency was not

higher than the MinION sequencing error of 6% (Delahaye & Nicolas, 2021). MinION sequencing error will be further discussed in the next section.

Table 3.4: Deletions and base changes at the expected cut site in cells transfected with SpGCas9

SpGCas9 guide concentration	*T>A*	*T>C*	*T>G*	*Deletions*	*Total base changes*
Control	0.08%	0.47%	0.08%	0.7%	**1.3%**
3.9pmol RNP Complex	0.09%	0.9%	0.35%	0.84%	**2.2%**
7.8pmol RNP Complex	0.21%	0.52%	0.14%	0.99%	**1.8%**
15.65pmol RNP Complex	0.20%	0.63%	0.1%	0.92%	**1.8%**
31.3pmol RNP Complex	0.14%	0.7%	0.20%	1.09%	**2.1%**

These results show that, when the H678 guide was transfected via CRISPRMAX using concentrations ranging from 3.9 pmol to 31.2 pmol, the SpGCas9 enzyme and *RP1* guide did not induce base changes or deletions at the intended site. Therefore, editing using SpGCas9 was not continued.

3.4 Cas9 RNP complex liposome-mediated editing

This section describes the testing of liposome-mediated RNP complex delivery to immortalised foetal fibroblasts. The first step was to determine the cutting efficiency of the RNP complex (using guide I694) *in vitro* by testing its ability to cut the target DNA within an amplicon generated via PCR. The second step was to determine whether the RNP complex would efficiently edit the DNA of cultured fibroblast cells *in vitro*. Despite the I694 guide cut site being located further away from the mutation than the H685 and Q686 guides, the decision was made to use this guide because of a higher predicted editing efficiency with fewer off-targets. The I694 guide cut site is located 55 bases downstream of the targeted mutation site, as shown in Figure 3.6.

3.4.1 *In Vitro* RNP Cleavage Assay

The result of an *in vitro* RNP cleavage assay using the I694 guide RNP complex on the 480 bp amplicon of sheep DNA (Figure 3.10) showed that the RNP complex was able to cleave the target sequence efficiently, generating two bands of 394 bp and 86 bp. The assay revealed increasing cleavage efficiency at higher concentrations, with the largest fragment of uncut PCR amplicon of 480 bases gradually reducing in brightness as the RNP complex was increased, with a corresponding increase in the expected 394 bp and 86 bp fragments. The control well only contained uncut PCR product to facilitate measuring the placement of the bands. Despite the bands on this gel travelling at a higher molecular weight than they should when compared to the ladder, multiple *in vitro* cleavage assays produced similar results –the bright diffuse signal below the 250 nM, 500 nM and 1000 nM lanes is the guide RNA which is theorised to cause the samples in the gel to run more slowly than anticipated. Due to time constraints, optimisation of this *in vitro* cleavage assay was limited, therefore future work should re-evaluate the ratio of Cas9 to PCR product, as well as an RNAse step during heat inactivation to digest any residual guide. In summary, the *in vitro* cleavage assay demonstrated successful guide-mediated DNA cleavage.

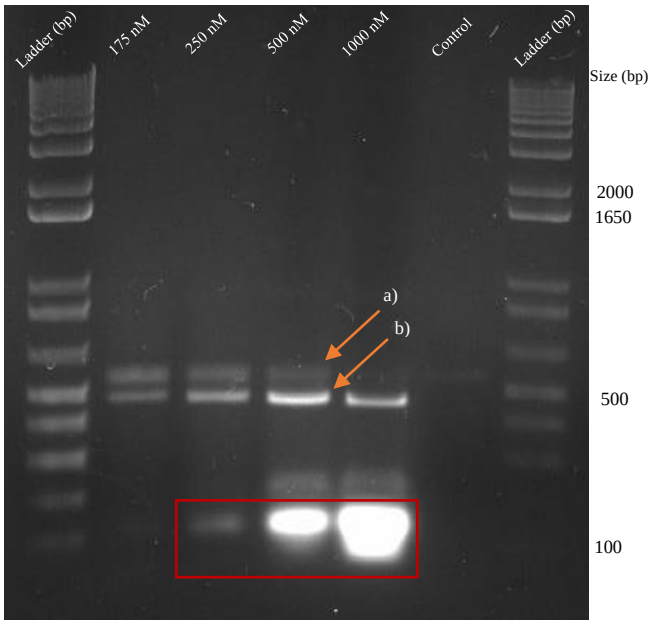


Figure 3.10: Image of the in vitro cleavage assay using the standard Cas9 and corresponding guide run on a 1.5% agarose gel. The volumes of RNP complex range from 175nM to 1000nM. Arrows a) and b) show the uncut and the cut PCR product respectively. The rightmost lane contains the DNA-only control and the residual RNA is outlined in red.

3.4.2 Indication of base changes and deletions through MinION sequencing

The next step was to deliver the I694 guide RNP complex using CRISPRMAX to edit cells. This experiment was conducted use two Protocols (A and B), as described in section 2.12.1. For the first treatment of immortalised fibroblast cells using RNP complex, the cells were cultured in Full Media until they reached approximately 80% confluence, then were transfected in a 24-well plate using Protocol A, detailed in

Section 2.12.1.1. This protocol had been successfully applied to generate edited cell lines in previous experiments conducted in this lab. Following transfection, cells were grown to confluence and DNA was extracted, amplified, and sequenced in the same manner as SpGCas9-treated cells (detailed in section 3.3.3). These samples were sequenced with the Oxford Nanopore MinION device.

Initial analysis of the MinION sequence reads from the amplicons that included the targeted site (1.9 pmol to 31.3 pmol of RNP complex) indicated what we thought was a series of deletions at the expected cut site in DNA tested across all experimental wells. The percentage of deletions and total base changes observed in this experiment are detailed in table 3.5 below:

Table 3.5: Base changes and deletions observed at the expected cut site for the I694 guide in cells transfected with Cas9

RNP complex concentration	*C>A*	*C>G*	*C>T*	*Deletion*	*Total base change %*
No Cas9 Control	0.38%	0.47%	1.19%	6.17%	**8.2%**
1.96 pmol	1.00%	1.27%	1.71%	13%	**16.98%**
3.9 pmol	0.91%	0.42%	1.04%	11.21%	**13.58%**
7.8 pmol	0.8%	0.65%	1.37%	10.04%	**12.8%**
15.65 pmol	0.81%	0.78%	1.66%	10.43%	**13.67%**
31.3 pmol	0.84%	0.98%	2.1%	11.83%	**15.75%**

As seen in Figure 3.11, sequencing indicated nucleotide substitutions at the cut site of the I694 guide, however the percentage of deletions was low, and it was difficult to ascertain whether these deletions were due to Cas9-mediated editing or due to technical sequencing error.

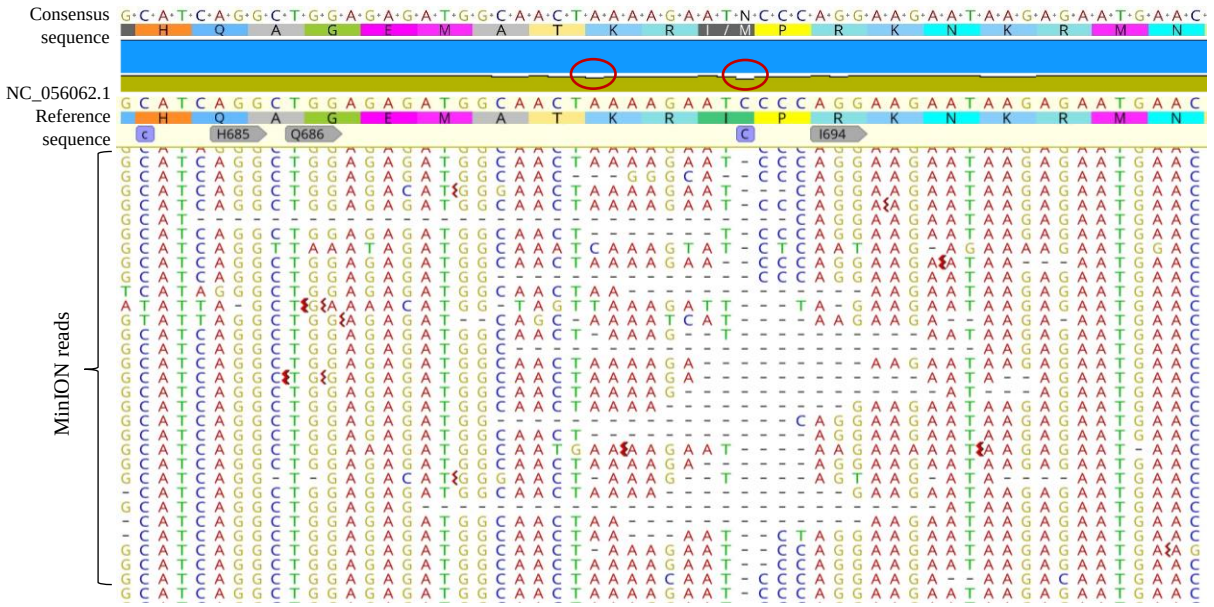


Figure 3.11: Evidence of low levels of editing using the I694 Cas9 guide transfected with the Lipofectamine™ CRISPRMAX™ Cas9 reagent. This figure shows the MinION reads from the Cas9 transfection using the I694 guide when transfected with 31.3 pmol of RNP complex. This image in representative of some of the large deletions that were observed at the cut site for this guide (indicated by the blue C). The red circles indicate the reduction in the level of coverage (the yellow horizontal band) that is observed across all reads in this dataset, with the dip for the expected cut site being similar to that seen at the end of the homopolymer run upstream. **Generated by Geneious version 2022.1 created by Biomatters.**

3.4.3 Transfection using higher RNP complex concentrations

The second treatment of immortalised fibroblast cells using RNP complex followed Protocol B (section 2.12.1.2), which uses significantly higher concentrations of the complex than Protocol A and a reverse transfection method in a 96-well plate. Cells were grown in a T-25 flask until there was a population of approximately 240,000 cells. The RNP complexes were assembled and the Lipofectamine™ CRISPRMAX™ Cas9 reagents were added to each experimental well and were gently mixed with ~60,000 cells suspended in 50 µL of media. Contrary to what was expected, the resulting MinION reads from this transfection experiment showed similar levels and types of deletions in the 15.65 pmol samples and experimental samples, rather than higher levels of editing (Table 3.6).

Table 3.6: Base changes observed using higher concentrations of the RNP Complex

RNP complex concentration	*C>A*	*C>G*	*C>T*	*Deletions*	*Total base change %*
15.65pmol	1.2%	2.33%	6.49%	14.82%	**24.84%**
100 nM	1.05%	1.35%	6.17%	14.28%	**22.85%**
500 nM	1.5%	2.01%	5.88%	14.55%	**23.94%**

3.4.4 Comparison to the SpGCas9 results

Because the experimental conditions showed very little difference to the control, there was some suspicion that the majority of the editing results so far may have been artifactual and solely represent MinION sequencing errors. To evaluate this possiblity, the sequence variants identified through the SpGCas9 and the RNP complex editing were compared. A table incorporating the site specific edits for the SpGCas9 editing experiment is below (Table 3.7):

Table 3.7: Nucleotide substitutions observed at the I694 cut site in cells treated with SpGCas9

SpGCas9 guide concentration	*C>A*	*C>G*	*C>T*	*Deletions*	*Base change %*
Control	0.48%	0.95%	1.78%	6.39%	**9.6%**
3.9pmol	0.53%	0.62%	1.29%	7.13%	**9.57%**
7.8pmol	0.60%	1.02%	1.45%	7.30%	**10.37%**
15.65pmol	0.66%	0.79%	1.25%	5.54%	**8.24%**
31.3pmol	0.28%	0.87%	1.23%	6.02%	**8.4%**

The results from Table 3.6 show that we are likely observing MinION sequencing error as high as 10.37%, as the negative control in the SpGCas9 samples had the same levels of deletions at the I694 cut site (nucleotide position 2079), and there were no editing reagents targeted to that site that could induce deletions. Furthermore, the amplicons from the SpGCas9 samples had been sequenced at the same time as those from the cells that were edited with low concentrations of the RNP complex (Table 3.4) and had similar levels of deletions at that position. To confirm whether there was truly an edit at the cut site (nucleotide position 2079), we Sanger-sequenced amplicons from the wild type, the experimental control and the 500 pmol editing experiment.

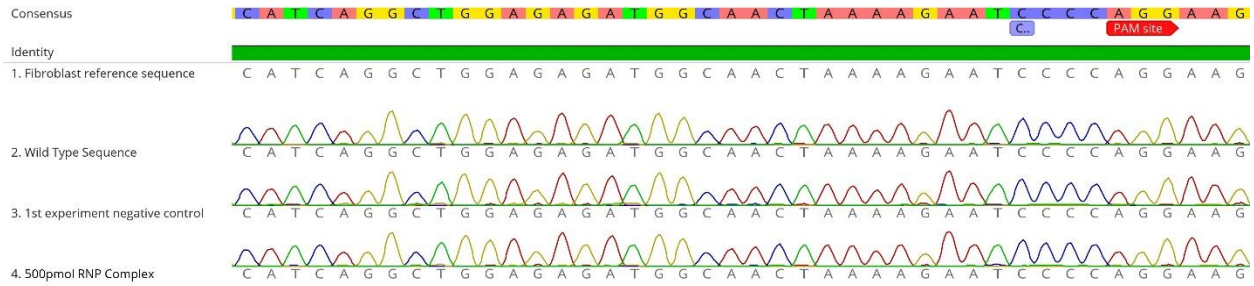


Figure 3.12: Sanger sequences from there different sequencing runs shows no difference from the unedited fibroblast control. Sanger sequences above show the original fibroblast sequence, the 500pmol of RNP complex and two wild type sequences from separate editing experiments. The deletion that was observed in the MinION sequences at the expected Cas9 cut site, labelled 'C' in blue, is not observed in the Sanger traces. **Generated by Geneious version 2022.1 created by Biomatters.**

The Sanger sequences in Figure 3.12 demonstrate that editing is unlikely to have been the cause for such high levels of apparent deletions at the I694 cut site (nucleotide position 2079). The apparent sequence variation seen in the experiment using 500pmol of RNP complex with a purported deletion rate of 24% would have been visible in the sequences (Table 3.5). Taken together with the results of the SpGCas9 samples, the observations suggests that editing reagents are failing to edit the cells when transfected using CRISPRMAX™.

3.5 Hypotheses generated from the errors and problems identified

Considering the lack of editing seen thus far, two potential explanations were developed: firstly, that the RNP complex was incapable of cleaving the target site in the cell line we were using; secondly, that CRISPRMAX™ was not effectively delivering the editing reagents to the cells. A final experiment was conducted in an attempt to resolve this editing issue.

3.5.1 Delivering the I694 RNP editing complex into an alternate cell line using nucleofection

Using dual-guide RNAs that were used to successfully generate a cell model for the *FRM1* gene, members of the Snell lab set up a comparative experiment. When editing using the dual guide system for *FMR1* is successful, a portion of the *FMR1* gene is deleted. The effectiveness of this *FMR1* editing can be revealed by gel electrophoresis of amplicons that include the edited region. It was therefore used as a positive control.

Both *FMR1* and I694 guides were tested in the primary cell line 2D636 from adult South Australian Merino sheep. This cell line was developed by the South Australian Research and Development Institute (SARDI).

The RNP complexes for both *FMR1* and I694 (at 100 nM, 500 nM and 1000 nM) were formed as described in Protocol B and delivered using CRISPRMAX™. These *FMR1* and I694 RNP complexes were also delivered to another 96-well plate of cells using nucleofection. The resulting gel for the *FMR1* dual guide system demonstrated this shorter amplicon in 2D636 cells when RNP complexes were delivered via nucleofection, but not in those delivered via CRISPRMAX™.

DNA from cells nucleofected with the I694 guide RNP complex at 100, 500 and 1000 nM was isolated, PCR amplicons were generated as before, then sequenced using MinION. The reads demonstrated deletions at frequencies much higher than those observed in the negative control as shown in Table 3.8:

Table 3.8: Nucleotide substitutions and deletions observed at the I694 cut site for RNP Complexes delivered to 2D636 primary sheep fibroblasts via Nucleofection

RNP Complex concentration	*C>A*	*C>G*	*C>T*	*Deletions*	*Total base change %*
Negative Control	0.9%	1.79%	2.33%	10.13%	**15.1%**
100 nM	1.32%	1.92%	3.31%	13.22%	**19.8%**
500 nM	1.68%	0.93%	2.93%	56.65%	**62.2%**
1000 nM	1.73%	1.23%	2.86%	54.72%	**60.5%**

The MinION sequences visualised in Geneious Prime indicated that editing had occurred in the 500 nM and 1000 nM samples. This was because the percentage of MinION reads that were carrying large deletions around the CRISPR-Cas9 mediated cut site was significantly higher than that of the negative control.

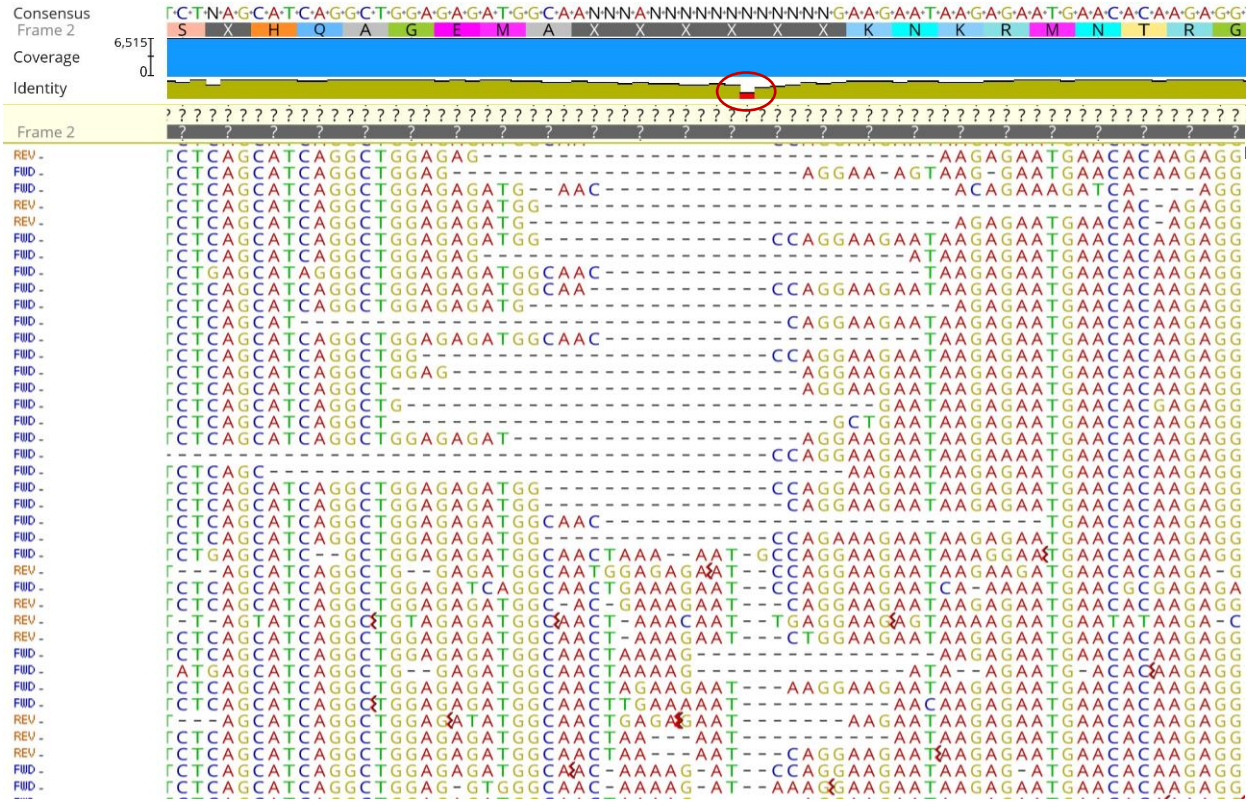


Figure 3.13: MinION reads of DNA extracted from 2D636 primary fibroblast cells treated with 500nM RNP complex for the I694 guide. This figure shows the amplicons from ovine primary fibroblasts that had the highest percentage of deletions when compared to the negative control. The evidence for editing is fairly conclusive, with large sections of deletions encompassing the cut site in 48% of the MinION reads. The drop in coverage at the expected cut site (circled in red) demonstrates that this base saw the highest number of deletions. The lack of amino acid determination in the consensus sequence also supports this conclusion. **Generated by Geneious version 2022.1 created by Biomatters.**

When accounting for MinION error seen in the negative control of 15.15%, the 100, 500 and 1000 nM samples have adjusted editing rates of 4.6%, 47% and 45.4% respectively. This increased editing efficiency relative to the liposome-mediated editing experiment is encouraging. The expectation for future experiments would be that in the incorporation of the HDR template (section 2.9.1.2), we would observe cells carrying the R677X mutation.

3.6 The I694 Cas9 guide has the capability to generate an isogenic cell line without the need for HDR

Following close examination of the MinION amplicon sequences from the nucleofected I694 CRISPR/Cas9-edited 2D636 cells, it was apparent that deletions generating *RP1* frameshifts would generate premature stop codons in the region of the gene that causes a dominant form of the condition. Approximately 33% of the reads in the 500 nM RNP complex sample show a single base “C” deletion at

the exact guide-mediated cleavage site. Unfortunately, this is also the location that is mis-sequenced using the MinION methodology, with anomalous deletions of this “C” nucleotide in 10.1% of reads in unedited cells. An optimistic view of these results would suggest that having deducted the 10.1% sequencing error rate, as high as 23% of all transcripts truly have an editing-induced single-base deletion at this position. If this is correct, the frameshift that is induced would result in a premature stop codon 6 amino acids downstream from the cut site (Figure 3.14).

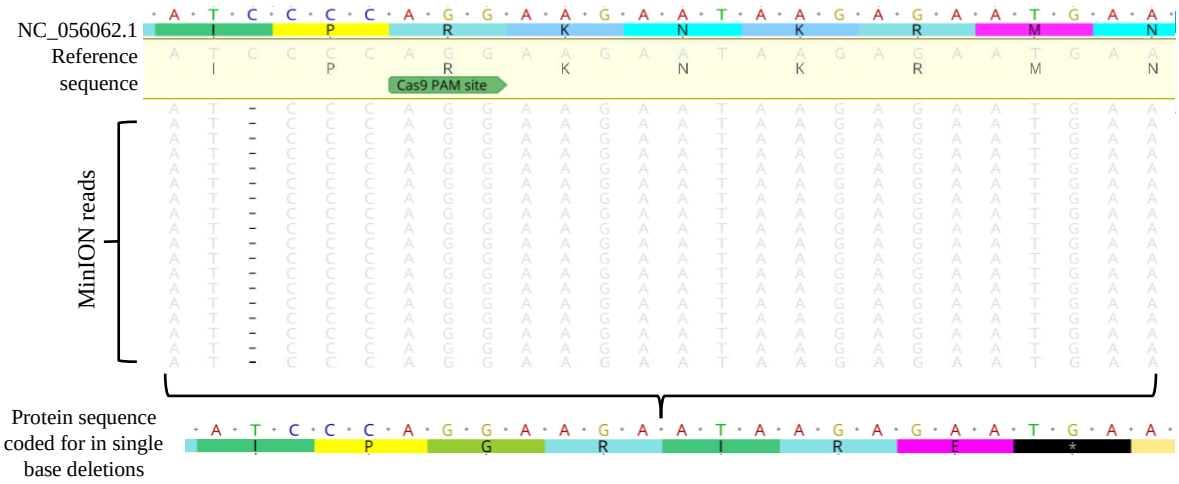


Figure 3.14: Deletions present in MinION reads for cells treated with 500 nM of RNP complex via nucleofection, with the predicted protein sequence after the deletion causes a frameshift. The reads from the sample of DNA nucleofected 500 nM of RNP complex showed series of deletions at the expected cut site 4 bases upstream of the Cas9 PAM site (labelled in green). These single base deletions generate a frameshift mutation that causes a TGA stop codon to be generated 6 amino acids downstream of the I694 cut site. **Generated by Geneious version 2022.1 created by Biomatters.**

This means that in cells with a single base deletion at the I694 cut site, the *RP1* gene is truncated at amino acid 701 (amino acid 702 in the human sequence). This has significantly more widespread implications than initially anticipated, as the stop codon at position 702 is also the location for several clinical variants that cause frameshift mutations that have been confirmed as pathogenic or likely pathogenic in ClinVar. These variants are detailed in Table 3.9:

Table 3.9 Clinical Variants that cause a frameshift causing a stop codon at amino acid 702

ClinVar reference	**Classification/Review status**	**Position of stop codon amino acid number**
NM_006269.2(RP1):c.2053del (p.Tyr**685**fs)	Pathogenic	702
NM_006269.2(RP1):c.2072del (p.Ala**691**fs)	Pathogenic	702
NM_006269.2(RP1):c.2079del (p.Gly**694**fs)	Pathogenic	702
NM_006269.2(RP1):c.2083del (p.Ile**695**fs)	Likely pathogenic	702

These data demonstrate the potential for generating a cell model of a truncating mutation in the *RP1* gene with the same phenotype as known pathogenic variants without needing to replicate each variant's specific mutation.

Regardless, the high-level of multi-nucleotide deletions incorporating the cut site that we observed in the 500 nM and 1000 nM samples indicate that editing was relatively efficient (Table 3.7). Therefore, the editing induced by delivering the I694 RNP complex using nucleofection demonstrated more conclusive, consistent editing in the 2D636 primary fibroblasts than Lipofection using CRISPRMAX™.

4 Discussion

4.1 Retinitis Pigmentosa

Retinitis Pigmentosa (RP) is one of the leading causes of inherited blindness, affecting an estimated 1.5 million people, or 1 in every 4000, globally (Wu et al., 2023). As a progressive ocular disease, the age of onset ranges from a person's twenties to early forties. A person diagnosed with RP has a normal lifespan, but will be completely blind by their fifties, a distressing and debilitating reality for most. Retinitis Pigmentosa currently has no effective treatment to slow or arrest disease progression, with the mechanism of cell death and melanocyte migration still the subject of some debate (Verbakel et al., 2018; O'Neal & Luther, 2023). As an inherited disease, many families have multiple members afflicted which can arise through dominant or recessive acting mutations. Retinitis Pigmentosa poses a significant unmet health need, creating a financial burden on families and health systems. It is therefore an important disease requiring a treatment.

People with the recessive form of RP due to loss of function mutations of the *RPE65* gene have been subject to gene replacement therapy trials. This has proven effective with the replacement vectors delivered by intraocular injection transducing the retinal pigment epithelial (RPE) cells (Ducloyer et al., 2020). The results of these clinical trials resulted in statistically significant improvements in full-field light sensitivity. This has led to this approach being the first FDA-approved treatment for a genetic form of RP (Ducloyer et al., 2020). Closer to home, this treatment has also been approved for use by the New Zealand Medicines and Medical Device Safety Authority (MedSafe) as of January of 2023. However, Luxturna is only effective for those who carry recessive-acting mutations in the *RPE65* gene. Treatments for forms of RP caused by autosomal dominant-acting mutations have not yet been approved.

The aim of this thesis was to develop and test the gene editing tools that could be applied to make a sheep model of RP that could be used for preclinical therapy testing. The target gene was *RP1,* and the mutation of particular interest is the relatively common p.R677X. Truncating mutations in this region of the gene result in the production of a dominantly acting gain of function protein. Simply put: the mutant protein interferes with the functions of the normal copy of the protein. Practically, the development of the editing

tools involved characterising the ovine *RP1* gene, then designing and testing CRISPR/Cas9 editing reagents in sheep fibroblasts. It was envisaged that in the future this cell model could also be used to test reagents that could completely knock out the mutant allele. The assumption is that the removal of the mutant copy of the gene through allele-specific editing of somatic cells would leave a patient hemizygous for the unaffected copy of the *RP1* gene. The ultimate aim following on from this thesis is that this could be tested in the yet-to-be-made sheep model to determine whether it is an effective treatment for autosomal dominant RP.

4.2 Annotation of the *RP1* gene in sheep

At the commencement of this project, it was apparent that the annotated ovine version of the *RP1* gene was inaccurate, including missing exons. Therefore, in order to create an ovine *RP1* model, a correctly annotated transcript needed to be established. This was required to ensure that the location of CRISPR/Cas9 guides and the HDR template design was accurate. Initial searches for the sheep ortholog in the ovine genome indicated that the *RP1* gene was not structured in the way that we expected, and we concluded that the annotation was likely incomplete. This incomplete annotation of the genome is common in species that are not the primary focus of genetic research. Automatic annotation of genes with predictive tools are known to sometimes annotate those genes incorrectly.

The human *RP1* gene is comprised of four exons, the first of which is non-coding (Figure 3.2). The initial alignment of the automatically annotated sheep *RP1* gene to the human gene had a similarity score of 61%. However, this similarity was only through the alignment of human exon 4, the largest of *RP1*'s exons to the automated annotation of ovine *RP1*. The first three exons of the ovine gene were identified by using NCBI blast tools to align the human *RP1* DNA, RNA and translated protein sequence to the sheep genome (section 3.1). Translation of the then-predicted ovine *RP1* mRNA and alignment with human *RP1* gene revealed a similarity score of 67% (Appendix 6 and 7). The intron/exon boundaries were identified in the same relative positions as they occur in humans' gene due to the presence of canonical dinucleotide splicing motifs (Figure 3.3). In summary, this analysis confirmed that the exons we identified were very likely to be exons 1, 2 and 3 that were missing from the predicted annotation of the ovine *RP1* gene in Genbank (NCBI).

With the sheep *RP1* gene fully annotated, this was used as the starting point for guide design to induce a truncation close to the desired R677X protein-coding mutation site. In the future, this annotation will enable the generation of further cell models of RP targeting different mutation sites.

4.3 Trialling SpGCas9 enzyme for *RP1* editing

The SpGCas9 enzyme (PAM site -NGN) is an exciting addition in the CRISPR toolkit as the PAM site required less specificity. In effect the use of the Cas9 variant increases the number of available cleavage sites for targeting areas of a genome that do not have a standard -NGG PAM site in close proximity to the desired edit site. It is known that the efficiency of cleavage of the SpGCas9 enzyme varies depending on the nucleotides flanking the invariant "G" position. It has been reported that for some sites the SpGCas9 can achieve a cleavage rate equivalent to the original Cas9 enzyme and others, lower. Research conducted by Walton et al. in their 2020 paper demonstrated high efficiency (on-site editing efficiency of 51.2% for -NGG sites) and minimal off targeted editing in HEK293T cells. The researchers also demonstrated, using a method they developed called high-throughput PAM determination assay (HT-PAMDA), that the SpGCas9 enzyme had varying levels of editing efficiency dependent on the nucleotides that followed the "G" position of the NGN PAM site. They found that the SpGCas9 enzyme had close to the same level of efficiency as Cas9 when using the PAM site -NGAG (Walton et al., 2022). The target site in the sheep fibroblasts used in this thesis utilised a PAM site with NGCG, which may have significantly reduced the editing efficiency of the enzyme.

Considering the prior difficulties that were faced by previous students in the Applied Translational Genetics lab to induce editing in *RP1* in the immortalised fibroblast cell line using Cas9 PAM sites, a closer PAM site to the R677X locus for a different species of Cas was investigated. The application of SpGCas9 in this thesis was a reasonable method for increasing the likelihood of replicating the mutation; the closest Cas9 PAM site to the mutation of interest, R677X, was 23 bases downstream and the next PAM site nearby was 54 bases upstream of this position in the sheep ortholog of the gene (Figure 3.6). The use of the closest SpGCas9 guide site to the mutation meant that the distance between the expected cut site for the SpGCas9 enzyme and the mutation site that we hoped to replicate was just 7 bases from the mutation site, as opposed

to 23 with Cas9. Furthermore, the manner in which we intended to replicate the R677X mutation was through the incorporation of the HDR template (section 2.9.1.2) and there is research that supports the assertion that HDR efficiency is much higher when the cleavage site is close to the nucleotides that are intended to be replaced by the template (Schubert et al., 2021).

The SpGCas gene coding for the enzyme was contained in the SpG-HF1 Cas9 LentiCas9-Blast plasmid. This plasmid had previously been used by my research group to produce a lentivirus for stable cell line production. It had also been tested by simple liposome mediated plasmid transfection and shown to successfully facilitate genome cleavage. In the experiments presented here that SpG-HF1 Cas9 LentiCas9-Blast plasmid was transfected into cells using Lipofectamine 3000™. In order to give time for the Cas9 protein to be produced in the target cells the guide RNA was subsequently transfected 24 hours later. The volume of plasmid to be delivered to the cells was determined by a transfection gradient using GFP as the positive control, which showed that the highest transfection efficiency that could be obtained using Lipofectamine 3000™ was 36.2%. After the immortalised fibroblast cells were transfected using the concentrations in Protocol A, we did not see any editing occurring, despite transfection controls indicating the methodology worked.

We identified several complications with using the alternative Cas enzyme SpGCas9 to edit cells. Despite SpGCas9's purported efficiency, it is not as high as Cas9 for our editing site. The delivery mechanism that has worked most effectively in our lab for editing cells is to form the RNP complexes and administer the enzyme and guide together in one transfection using CRISPRMAX™ Lipofectamine. This is not possible for SpGCas9, as the enzyme is delivered on a plasmid and needs to be expressed in the cells before the guide can be administered. This means that the guide needs to be administered at the point at which the highest level of SpGCas9 has been translated. Furthermore, the success of the plasmid transfection control could be due to the use of Lipofectamine 3000™ for transfecting the SpGCas9 enzyme and its control, GFP, while CRISPRMAX™ was used to transfect the guide. If there is an issue with the CRISPRMAX™ lipofection kit, as it has now been determined, then it is more likely that the lack of editing is due to the guide not being delivered, rather than a problem with the SpGCas9 enzyme *per se*.

Despite the excitement and promise that engineered species of Cas9 bring, it is important to note that experimentation into improving the efficiency of alternative Cas enzymes, including SpGCas9, is still ongoing. Recent research published after the experimental findings of this thesis criticise the effectiveness of the enzyme and state that the editing efficiency and off-target effects generated by engineered species of Cas need to be improved (Villiger et al., 2024; Zheng et al., 2024). It is also important to note that the delivery of the SpGCas9 enzyme via plasmid rather than through a pre-formed RNP complex, as is used in Cas9 experiments, can significantly affect the ability of the enzyme to edit as desired in cells (Wei et al., 2020). This will improve when the purified protein from these engineered Cas9 species is available commercially.

4.4 The editing of fibroblasts using the I694 guide RNP complex was successful when delivered by nucleofection, but not lipofection

4.4.1 Lipofectamine was less effective in delivering the RNP complex to cells

As detailed in section 3.4.2, we used the CRISPRMAX™ Lipofectamine kit to deliver the RNP complex containing purified Cas9 protein and the I694 guide to immortalised fibroblast cells using Protocols A and B (section 2.12.1). In these experiments we tested a wide range of RNP complex concentrations (1.9 pmol to 1000 nM). The MinION sequencing results from PCR amplicons of the genomic DNA from edited cells revealed that the deletions and nucleotide substitutions were at comparable levels to unedited cells. Previous experiments in our laboratory using both Protocols A and B that were used in this thesis (section 2.12.1) had successfully produced very significant levels of editing (up to 80%). In order to see if these unexpected results were cell line or transfection methodology-dependent, an experiment was conducted according to Protocol B which compared Lipofection to Nucleofection using 2D636 primary sheep fibroblasts. A previously used dual guide RNA combination targeting the *FMR1* gene was included in this experiment as a positive control (section 3.5.1). This experiment demonstrated again the failure of the liposome transfection method, while the nucleofection approach was very successful (as shown in Table 3.7 and figure 3.13).

4.4.2 Expired or faulty reagents may have been the cause of failure in lipofection-based transfections

Considering the successful editing that was induced in primary fibroblasts using nucleofection (section 3.5.1), it is possible that the RNP complexes had not been delivered to the cells due to issues with the batch of CRISPRMAX™ lipofectamine used. It is not completely clear as to why lipofectamine, a previously successful method of transfecting the immortalised fibroblast cell line in the lab, was inefficient during the course of this thesis. It is possible that the reagents had expired or were accidentally frozen, which is known to alter the structural integrity and inactivate the lipid particles (ThermoFisher, 2024).

4.4.3 Nucleofection succeeded in a primary fibroblast line

The decision to test whether we could generate editing in a different cell line was to clarify whether the lack of editing we initially saw was due to issues with the cell line or the method of transfection. When both immortalised fibroblast cells and 2D636 cells were transfected with lipofectamine, there was no editing observed in the positive control, the guide for the *FMR1* gene that had been used successfully to edit these cells. There was also no editing that was easily distinguished from the reads for the negative control in the MinION sequences. Changing the transfection method to nucleofection in the 2D636 primary fibroblasts produced the editing we were hoping for (detailed in section 3.5.1) and confirmed that the lack of editing was likely due to issues with reagents, rather than the guide or cells. Due to time constraints, the experiment was not repeated in the immortalised fibroblast cell line used for the original experiments presented in this thesis.

When accounting for the read errors caused by Oxford Nanopore MinION sequencing, the highest editing rate observed, 47%, was in the experimental well that used 500nM of RNP delivered by nucleofection complex using a reverse transfection method (a component of Protocol B in section 2.12.1.2). This is comparable to other experiments conducted in the Snell Lab that have successfully edited sheep fibroblasts to create isogenic cell lines for modelling disease. Notably, the I694 guide induces a 1bp deletion at its predicted cut site (the 3rd nucleotide of amino acid 694), which causes a frameshift mutation to occur (notated as p.I694fs). This mutation causes a premature stop codon to be induced at amino acid 701 in sheep, which would correspond to amino acid 702 in humans.

4.4.4 Clinical variants with a truncating frameshift mutation

With this in mind, the Clinical Variant database for variations that relate to human disease (ClinVar) was used to determine whether there were any documented frameshift mutations in humans also due to a stop codon at this position. One likely pathogenic and three confirmed pathogenic clinical variants for Retinitis Pigmentosa were identified with this specific mutated codon (table 3.9). As a result, if this experiment were repeated and optimised for our immortalised fibroblast line, then the cells could be single-cell sorted and tested to identify lines with a single base deletion at I694. The resulting isogenic cell line could subsequently be a relevant model of the three pathogenic variants for Retinitis Pigmentosa; p.Tyr685fs, p.Ala691fs and p.Gly694fs. As mentioned, these pathogenic variants cause the RP1 protein to truncate at aa702, thereby missing 1454 amino acids, or 67.4% of the RP1 protein. This is similar to the truncation seen in those with R677X, where 1479 amino acids, 68.5% of the protein is not translated.

4.4.5 An HDR template introducing the R677X mutation may not be required to generate the same phenotype in an isogenic cell line

This thesis has been focused on generating an isogenic cell line carrying the R677X mutation by introducing an HDR template that replicates it, however research into the effect of the truncated RP1 protein suggests that we may be able to produce the same gain of function phenotype without needing to replicate the mutation exactly (Chen et al., 2010; Nanda et al., 2019; Wang et al., 2021). As mentioned in section 1.3.1, a mutational hotspot was proposed by Chen et al. in 2010 and has since been the subject of much research, with a review of 147 pathogenic truncating mutations identifying three distinct regions, one of which contained most autosomal dominant variants (Figure 4.1) (Wang et al., 2021).

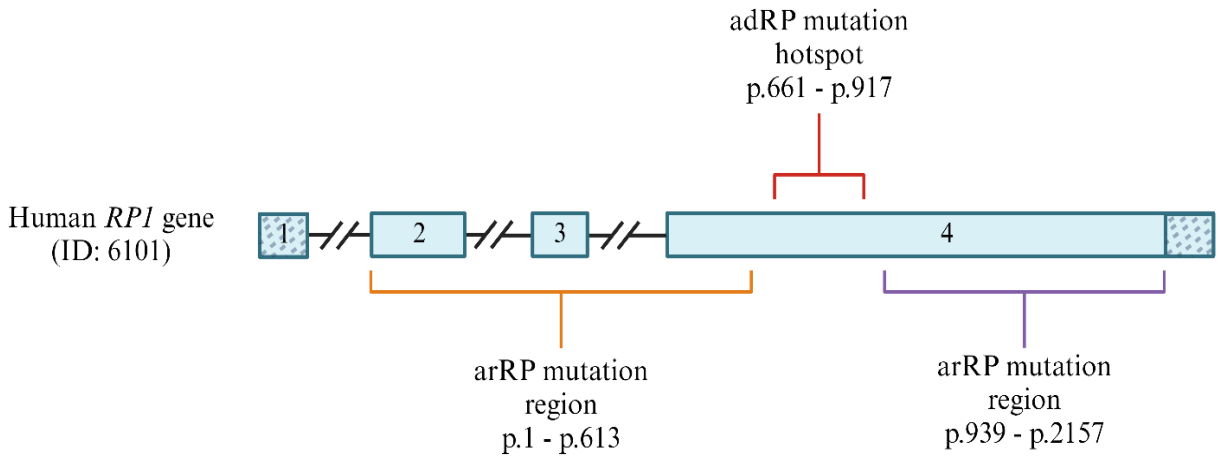


Figure 4.1: Schematic of human RP1 gene showing the autosomal dominant hotspot proposed by Wang et al. (2021) after categorising 147 pathogenic truncating RP mutations. This schematic shows the 4 exons of the human *RP1* gene (ID: 6101) with non-coding regions shaded, and the three distinct regions that were identified where autosomal recessive and autosomal dominant mutations for *RP1* are known to occur. The vast majority of autosomal dominant variants fell within p.661 and p.917, which is in keeping with mutational hotspots proposed by Chen et al. (2010) and Nanda et al. (2019). **Figure generated using BioRender.**

It is important to note that even in the event that the HDR template is not incorporated and subsequently induces our desired mutation, lines carrying single base deletions near the I694 cut site will still be incredibly useful as an isogenic line for directly modelling the human Y685fs, A691fs and G694fs mutations in *RP1*. The R677X mutation generates a protein that is 677 amino acids long, while Y685fs, A691fs and G694fs all generate a 702 amino acid protein – with only 25 amino acids difference in length, deletion induced by I694 will likely still be relevant for research purposes especially as the ultimate aim is to use editing reagents to knock out the mutant allele. Considering that most of the clinical variants for this disease in ClinVar are pathogenic when premature stop codons are induced by mutations in this region, we may not need to generate the exact R677X mutation to adequately model its phenotype.

4.4.6 Truncating mutations in the *RP1* gene

The current working theory is that the severe truncation mutations within the autosomal dominant hotspot have a dominant negative effect, with the truncated protein preferentially binding to the doublecortin binding domain in rod cells (Liu et al., 2012; Mizobuchi et al., 2021). This theory is supported by research undertaken by Lafont et al. (2011) where they examined the rate of degradation caused by autosomal dominant and autosomal recessive forms of the disease. Their findings showed that despite visual acuity

decreasing with age in both adRP and arRP, those with recessive mutations had more severe symptoms, recognised reduced visual acuity much earlier, and had a higher rate of decline in loss of their visual field than dominant cases. This is in keeping with what is expected of biallelic disease variants; however, the exact molecular mechanism has not yet been sufficiently elucidated. More research identifying the mechanisms of the protein is required to fully understand how mutations in the autosomal dominant hotspot of exon 4 have an affect on the function of the 3' end of the protein.

4.4.7 Advantages and disadvantages of Nucleofection

As previously discussed, the I694 guide that successfully edited primary sheep fibroblasts was delivered via Nucleofection. This method of transfection uses electroporation to disrupt the electrical charge of the cell membrane and is known to have a high transfection efficiency in cells coupled with the certainty of delivering RNP complexes directly to the cell nucleus (Leclere et al., 2005; Ren et al., 2022). Nucleofection circumvents several theorised complications related to the accessibility of the DNA in the immortalised fibroblast line. Because *RP1* is not expressed in fibroblasts, the window in which we could successfully edit the DNA is restricted to when the cells are actively replicating, as genes are packed away when they are not being expressed and are therefore less accessible due to the closed chromatin structure. The process of nucleofection delivers an electrical current in a manner that transfects the RNP complexes directly into the nucleus of the cell, rather than transfecting into the cytoplasm like lipofection does and therefore does not rely on the chromatin of cells being open as it is during replication.

Nucleofection has its drawbacks: it requires optimising of the reagents and selected program on the Nucleofector that administers the electrical current to the cells. Dependent on the cell type and line being used, the viability of the cell reduces as the transfection efficiency increases, so the immortalised fibroblast cell line would need to be tested for its optimal transfection efficiency and required reagents to maximise the editing rate and cell survival. There is also some evidence that suggests nucleofection can alter the metabolic activity of transfected cells, as well as negatively impact the membrane polarity of certain cell types due to the nature of electroporation (Mello de Queiroz et al., 2012; Mo et al., 2010). However, for the purposes of generating a cell model, these potential complications are outweighed by the benefits of

delivering RNP complexes directly to the cell nucleus, as we struggled to generate editing that was high enough to distinguish from sequencing error when Lipofectamine was used.

As mentioned in section 4.3, we failed to generate any editing in the cells that were treated using the SpGCas9 enzyme. If we wanted to continue using SpGCas9 to attempt to replicate the R677X mutation, it is unlikely that nucleofection could be used. This is because of the requirement of double transfection, with the enzyme and guide being transfected 24 hours apart and self-assembling into RNP Complexes that could cleave the DNA. To nucleofect the cells twice would likely have a significant impact on cell viability and greatly increase cell death. This challenge of dual transfection would need to be addressed before we could use nucleofection to introduce SpGCas9 and its guide into cells efficiently enough to induce editing.

4.5 Issues encountered with editing

As previously discussed in section 4.4, there were multiple issues with getting editing to occur in the immortalised fibroblast line at any of the available PAM sites in the region, all of which should be highly efficient Cas9 guides (as determined by CRISPOR). Initial examination of the reads produced by MinION sequencing for the cells treated with the I694 RNP complex delivered via lipofectamine indicated that there had been sporadic deletions outside the expected cut site for the I694 guide. These deletions were spread across the entire length of the 480 bp PCR product that we amplified and were usually at the end of homopolymer runs of longer than 4 bases. One theory as to how this may have occurred was that the treatment of a mycoplasma infection with the antibiotic Ciprofloxacin during the early passages of the cell line had damaged the DNA, creating low level mosaicism in the immortalised fibroblast cells prior to both SpGCas9 and Cas9 transfections.

4.5.1 Ciprofloxacin and double-stranded DNA breaks

Ciprofloxacin is a fluoroquinolone that is used to treat bacterial infections in human patients (Thai, Salisbury & Zito, 2023). Its use in cell culture to treat Mycoplasma infections has been associated with double and single-stranded DNA breaks through the production of reactive oxidative species (ROS) which causes oxidative stress in cells (Hincal, Gürbay & Favier, 2003). Studies have also found that fluoroquinolones for bacterial infections in humans produce ROS and free radicals that cause cell damage

in vivo (Talla and Veerareddy, 2011). It is well documented in cancer research that free radicals will induce double-stranded DNA breaks, much of which is repaired by Non-Homologous End Joining (NHEJ). This would explain why all samples (edited and non-edited) had similar levels of deletions (between 6 and 10%) in a single site in the genome which is 34 bp 5' from the expected cut site. Having said this, it is well documented that MinION sequencing struggles with reading homopolymer runs, and it is possible that at least a portion of the deletions that were observed in successfully edited cells using the I694 guide were caused by sequencing error due to the cut site being at the 5' end of a 4-base homopolymer run.

4.5.2 Sequencing through homopolymer runs with Oxford Nanopore Sequencing Technology

In this thesis it was difficult to extrapolate conclusive data from the MinION samples for both lipofected and nucleofected cells. This was due to the basecalling by the MinION sequencer which suggested that there were 1bp deletions 4 bases and 66 bases 5' of the expected cut site for I694. These regions were part of two homopolymer runs comprised of 4 nucleotides and 8 nucleotides respectively. For our purposes, a homopolymer run is defined as a section of DNA that has the same nucleotide 4 or more times in a row. In simplistic terms, the MinION sequencer struggles with identifying the exact length of these runs because as DNA strands are pulled through the pore of the sequencer, there is an electrical current that is disrupted in a manner that is unique to each of the four nucleotides A, T, C and G (Wick et al., 2019). When there are long sections of DNA with the same nucleotide in each position, the signal does not change as it normally would, which can cause the DNA strand to feed through too quickly and one or two few nucleotides to be called for that read. Once the electrical signal is translated to its corresponding base and the read is aligned to the desired reference sequence, a deletion is identified at that site, thereby suggesting that the DNA in the cell the reads were translated from differed from the reference sequence at that position, influencing the overall results of the reads from that sample. This is a common error seen in MinION sequencing; however, as researchers in the Applied Translational Genetics group identified, MinION also has issues with the identification of indels adjacent to a homopolymer run (Whitford et al., 2022). With this in mind, close analysis of sequences and in some cases, validation via Sanger, is required in future modelling of this disease locus, especially for identifying and accounting for MinION error when determining editing

efficiency. For this thesis, we determined the editing efficiency by deducting the percentage of base changes and deletions that were seen in the negative controls from the experimental samples. This approach was applied to all experiments regardless of the species of enzyme or transfection method.

4.5.3 Chromatin structure may have reduced the ability to edit the region of interest using Lipofectamine

It is clear from these results and previous experimental work in the Snell Lab that there have been difficulties in targeting the R677X mutation in *RP1* for editing. The use of fibroblasts could theoretically have reduced the editing capabilities of Cas9 due to the lack of expression of *RP1* in these cells and how accessible that section of the DNA was at the point of transfection. This is because of the way in which the DNA is packed in the cell nucleus and the restriction that this reduced level of expression poses. When the cell is in its stationary phase, the chromatin is closed and genes that are not required for metabolism are less accessible. As previously mentioned, the RP1 protein is a microtubule-associated protein that stabilizes the structure of rod cells, and according to The Human Protein Atlas, is not expressed in fibroblasts, which reduces the window of opportunity for editing the gene. This means that the only period in which the DNA is accessible is during cell replication and division, which requires close attention to be paid to the stages of the cell cycle. Attempting to transfect cells with RNP complexes when they are not actively dividing will significantly reduce the likelihood of editing.

This was a proof-of-principle experiment which intended to show whether the DNA could be edited, and the determinants of the experiment were focused on base changes at the DNA level rather than protein expression. One could argue that a different cell line with higher expression levels of *RP1*, such as a retinoblastoma line, could have produced better results. However, targeting the cells during their proliferative phase, during which the chromatin is open for cell replication, would mitigate this issue and the successful nucleofection of fibroblasts with the I694 guide negates these concerns. Coupling this with the complications that arise from applying research conducted in cancer-derived cell lines to disease modelling, the use of fibroblasts was sufficient in demonstrating the ability to edit the desired site in the sheep genome. Although the editing that we obtained was in primary fibroblast cells rather than immortalized fibroblasts we were wanting to generate an isogenic line from, we have concluded that the

lipofectamine was causing this reduced editing efficiency and we now have a method of transfecting the immortalised fibroblast cell line via nucleofection that we know is effective.

4.6 Future directions

4.6.1 Optimisation of editing in immortalised fibroblast cells using the I694 guide

The results of this thesis (section 3.6) determined that a frameshift mutation leading to a premature stop codon downstream could be induced in sheep fibroblast cells by generating a single base deletion using the I694 CRISPR/Cas9 guide. To produce this result, a primary cell line, 2D636, was used to allow us to determine whether the lack of gene editing was caused by an issue with the immortalised fibroblast line or the method of transfection (lipofection). Given primary cell lines have a limited life, it is more challenging to create an isogenic cell line compared with an immortalised cells line. It is therefore desirable to induce the editing that was seen in the primary fibroblasts in the immortalised fibroblast cell line for future investigations into treatment using CRISPR/Cas9. We determined that the likely cause of the failure to edit cells in the first instance was the lipofectamine, rather than the cell line, so we could test this by repeating the experiment using new reagents. Additionally, considering that nucleofection is known to induce higher levels of editing than lipofection, optimising the nucleofection protocol for the immortalised fibroblast line is also a possibility for generating an isogenic line for *RP1* using immortalised fibroblasts (Ren et al., 2022). Once transfected, cells can then be cell sorted using a BD FACS Aria II cell sorter (BD Biosciences) and grown up in the same manner described in section 2.12.1 until populations are large enough to freeze down for storage. The DNA from a small, randomly chosen subset of cell populations and sequenced to determine the type of editing that was induced and whether this is suitable for the development of an isogenic cell line.

4.6.2 Trialling nucleofection with two PAM sites closer to desired mutation site

It is clear that the most likely cause for a lack of editing observed in the past in the two closer Cas9 guides, H685 and Q686, is due to the mode of transfection. As a result, these guides may still prove efficient at cutting at the site closer to our desired mutation site, particularly as they are predicted to be high efficiency guides with few off-target effects. To test this hypothesis, I would set up an experiment using the 2D636 primary fibroblast line to determine whether we could induce any indels with these guides using

nucleofection and MinION sequencing. This would facilitate the generation of an isogenic line through single-cell sorting. It is worth noting that any single base deletions that might be induced using the H685 and Q686 guides would also cause a frameshift that terminates at amino acid 701, as we observed with the I694 guide. If they were shown to be effective in causing Cas9 to cut at their expected cut sites, these two guides would be better used for providing an opportunity for the HDR template to be incorporated.

4.6.3 Incorporation of HDR template to determine whether we can generate an isogenic cell line with the R677X mutation

If we wanted to continue trying to model the R677X mutation specifically, we would need to determine whether we could incorporate the HDR template designed in section 2.9.1.2 using the I694 guide (or guide H685 or Q686 if editing was successful). This template includes silent SNPs spread throughout the length of the template as well as two premature stop codons: one at the site where the R677X mutation would occur and another that removes the Q686 PAM site that had previously been trialled by other members of the Snell lab. To incorporate the HDR template, cells need to be in the S or G2 phase of replication, so timing the transfection is fundamental to the success of HDR incorporation (Lin et al., 2014). For this reason, it would be interesting to investigate the cell synchronisation method proposed by Lin et al. whereby cells are arrested at the S phase using specific drugs that impede part of the cell cycle (2014). They are then able to be nucleofected with the desired RNP complex and the likelihood of HDR incorporation increases dramatically. The incorporation of the HDR template will still occur without cells being synchronised as described above, but if the incorporation is not high enough to identify an isogenic line, the synchronisation method could further improve the transfection and subsequent editing efficiency. Successful HDR incorporation can be assessed by looking for the introduced SNPs and stop codons close to the cut site.

4.6.4 Generation of a large animal model

Once an isogenic cell line has been generated, we can go through the process of generating a sheep model for a truncation mutation within the *RP1* autosomal dominant hotspot. This would involve using CRISPR to edit single cell embryos from South Australian Merino sheep by administering the I694 RNP complex to the embryo's pronuclei via microinjection. Previous research in the Snell laboratory has led to the development of a transgenic sheep model of Huntington's Disease using this method of microinjection

(Jacobsen et al., 2010). This HD sheep model has facilitated the tracking of biochemical and cellular functions that could not be achieved in smaller animal models of the disease or in humans. The generation of a sheep model in *RP1* would facilitate the testing of potential treatments and provide crucial information regarding disease characteristics, onset and other related functionalities that cannot easily be obtained from people affected by those conditions. Considering the still-debated aspects of the mechanism by which the truncated RP1 protein causes the disease phenotype, a sheep model would shed light on the mechanisms of the disease that have not been definitively characterised, thereby facilitating the development of a treatment.

4.6.5 Proposed treatment for autosomal dominant *RP1* truncation mutations

We theorise that we could use CRISPR/Cas9 to knock out the disease allele in a sheep model. This might be achieved by categorising SNPs at the beginning of the coding region of the disease allele and developing a guide system that would generate a frameshift or nonsense mutation that silences the dominant allele. For example, there are 3 clinical variants with synonymous mutations in the 5' region of Exon 2 in *RP1* that are classified as likely benign that generate novel PAM sites: pPro5= (Accession number: VCV002095546.2), p.Thr7= (Accession number: VCV002005268.2) and p.Pro23= (Accession number: VCV002791725.1). These SNPs are outside of the doublecortin binding domains that coded for in amino acids 36-229 and generate a novel PAM site on the antisense strand. Using the PAM sites that are generated by these SNPs could facilitate an allele-specific CRISPR knock-down of the disease that silences the mutant copy of the allele if the novel PAM sites are present on that allele.

It may also be worth evaluating whether it is possible to induce an autosomal dominant mutant allele to become recessive by replicating a mutation in the arRP region upstream of the dominant hotspot (Figure 4.1). Liu et al. (2012) evaluated the family of an adult patient with RP who carried a truncation mutation at p.Pro229fs (Accession number: VCV002735168.1) – this is in the region of the proposed autosomal recessive region of *RP1*. Both parents and the affected person's sibling were heterozygous for this truncating mutation but developed no disease phenotype (Liu et al., 2012). It is possible that this autosomal recessive mutation, or one of the many others that demonstrates no disease phenotype in heterozygous individuals, could be replicated by targeting a nearby SNP that creates a PAM site. This could cause the

resulting protein to express the same truncation as that seen in people who are heterozygous for the recessive form of RP, thereby silencing the toxic gain-of-function mutant allele. This approach would require a more comprehensive understanding of the protein structure and function of the truncated toxic gain-of-function protein before testing this treatment in cellular or animal models.

4.7 Limitations

4.7.1 Applicability of disease research from a sheep model to humans

Despite sheep being a comparatively easy animal model to care for and examine, there are some fundamental morphological differences between the retina of a sheep and that of a human. Sheep have a rectangular retina, which contains a small band of ganglion cells, and rod and cone cells (Shinozaki et al., 2010). As a result, sheep have significantly better peripheral vision than humans, which may alter the ability to better understand the pathophysiology of Retinitis Pigmentosa. It may also impede the progression of the disease, as the tunnelling effect that is observed in humans where cells on the outermost portion of the retina die off first, might not occur in a similar manner.

4.8 Implications

The editing induced by the I694 Cas9 guide has exciting, widespread implications for the generation of a large animal model for Retinitis Pigmentosa. The stop-codon introduced at amino acid 701 in the reads from the primary sheep fibroblasts shows that the I694 guide for Cas9 has the potential to replicate the disease phenotype of multiple autosomal dominant pathogenic clinical variants who carry a truncating mutation in the *RP1* gene. The successful editing of these cells means that we can now further optimise existing protocols to generate isogenic cell lines in which we can test a gene knock-down treatment. This is also promising for the development of a sheep model of RP, which will improve the understanding of the molecular mechanisms of the disease as well as be a suitable model for testing treatments. This experimental work has the possibility to be foundational for the development of treatments for an otherwise untreatable condition.

4.9 Concluding remarks

In early March of 2023, the Third International Summit on Human Genome Editing was held in London, hosting keynote speakers from all areas of science and attracting thousands of scientists from across the globe. The opening statement from the Summit's organising committee reads as follows:

> *"Remarkable progress has been made in somatic human genome editing, demonstrating it can cure once incurable diseases. To realise its full therapeutic potential, research is needed to expand the range of diseases it can treat and to better understand its risks and unintended effects."*

Retinitis Pigmentosa presents a serious challenge for treatment development, as the protein that is produced by any of mutations in the *RP1* gene is not something that could be supplemented in a person's diet, for example. Longterm, the potential for reversing any rod cell death that has already occurred is low, as the retina as part of the central nervous system is non-regenerative, however using the results of this thesis we may be able to cease its progression and alleviate symptoms of the disease. Determining whether we can use CRISPR/Cas9 gene editing to generate a model of the disease, allowing researchers to test treatments, is a key part of addressing the unmet health need that exists for people with Retinitis Pigmentosa. It is also fundamental to recognising the distress that many people with incurable conditions like RP experience. Existing research in this disease is encouraging, with an approved gene augmentation treatment for an autosomal recessive form of RP; however, this still leaves a significant number of people who have a different mode of inheritance with limited treatment options. For a disease with such a high global prevalence, autosomal dominant forms of Retinitis Pigmentosa are an unmet need in clinical research. The results of the editing induced in cells during the course of this thesis are foundational for generating a model of this disease. Working under the established framework of large animal model generation, there is every hope that in the future this previously incurable disease could soon find a treatment in the CRISPR/Cas9 gene editing system.

Appendices

Appendix 1 Table from Daiger et al.'s 2013 paper on Genes and mutations causing retinopathies

Location	Protein	Type of retinitis pigmentosa	Other diseases	Associated Conditions	Number
ABCA4	1p22.1	ATP-binding cassette transporter—retinal	Autosomal recessive	Recessive macular dystrophy; recessive fundus flavimaculatus; recessive cone-rod dystrophy	680
BEST1	11q12.3	Bestrophin 1	Autosomal dominant; autosomal recessive	Dominant vitreoretinochoroidopathy; recessive bestrophinopathy; dominant Best type macular dystrophy	232
CRB1	1q31.3	Crumbs homolog 1	Autosomal recessive	Recessive Leber congenital amaurosis; dominant pigmented paravenous chorioretinal atrophy	183
CRX	19q13.32	Cone-rod otx-like photoreceptor homeobox transcription factor	Autosomal dominant	Recessive, dominant and de novo Leber congenital amaurosis; dominant cone-rod dystrophy	51
EYS	6q12	Eyes shut/spacemaker (*Drosophila*) homolog	Autosomal recessive		118
NR2E3	15q23	Nuclear receptor subfamily 2 group E3	Autosomal dominant; autosomal recessive	Recessive Stargardt disease; Goldmann-Favre syndrome; recessive enhanced S-cone syndrome	45
OFD1	Xp22.2	Oral-facial-digital syndrome 1 protein	X-linked	Orofaciodigital syndrome 1, Simpson-Golabi-Behmel syndrome 2	127
PDE6B	4p16.3	Rod cGMP phosphodiesterase beta subunit	Autosomal recessive	Dominant congenital stationary night blindness	39
PRPF31	19q13.42	Human homolog of yeast pre-mRNA splicing factor 31	Autosomal dominant		65
PRPH2	6p21.1	Peripherin 2	Autosomal dominant; digenic with ROM1	Dominant macular dystrophy; dominant vitelliform MD; dominant cone-rod dystrophy; dominant central areolar choroidal dystrophy	123
RDH12	14q24.1	Retinol dehydrogenase 12	Autosomal dominant; autosomal recessive	Recessive Leber congenital amaurosis	66
RHO	3q22.1	Rhodopsin	Autosomal dominant; autosomal recessive	Dominant congenital stationary night blindness	161
RP1	8q12.1	RP1 protein	Autosomal dominant; autosomal recessive	Autosomal dominant and recessive	67
RP2	Xp11.23	Retinitis pigmentosa 2 (X-linked)	X-linked		76
RPE65	1p31.2	Retinal pigment epithelium-specific 65 kDa protein	Autosomal dominant; autosomal recessive	Recessive Leber congenital amaurosis	134
RPGR	Xp11.4	Retinitis pigmentosa GTPase regulator	X-linked	X-linked cone dystrophy 1; X-linked atrophic macular dystrophy	151
SAG	2q37.1	Arrestin (s-antigen)	Autosomal recessive	Recessive Oguchi disease	11

TULP1	6p21.31	Tubby-like protein 1	Autosomal recessive	Recessive Leber congenital amaurosis	31
USH2A	1q41	Usherin	Autosomal recessive	Recessive Usher syndrome	392
Others less than 1%	C2ORF71, C8ORF37, CA4, CERKL, CLRN1, CNGA1, CNGB1, DHDDS, FAM161A, FSCN2, GUCA1B, IDH3B, IMPDH1, IMPG2, KLHL7, LRAT, MAK, MERTK, NRL, PDE6A, PDE6G, PRCD, PROM1, PRPF3, PRPF6, PRPF8, RBP3, RGR, RLBP1, ROM1, RP9, SEMA4A, SNRNP200, SPATA7, TOPORS, TTC8, ZNF513				312
Total					**3064**

Appendix 2 Comparative table for canine orthologs of the most common mutations causing Retinitis Pigmentosa and other retinopathies using information from Bunel et al.'s 2019 paper.

Gene	Location	Protein	Type of retinitis pigmentosa	Associated Conditions	Corresponding Canine chromosome
C2ORF71	2p23.2	Chromosome 2 open reading frame 71	Autosomal recessive		(CFA17)
CNGA1	4p12	Rod cGMP-gated channel alpha subunit	Autosomal recessive		(CFA13)
CNGB1	16q13	Rod cGMP-gated channel beta subunit	Autosomal recessive		(CFA2)
FAM161A	2p15	Family with sequence similarity 161 member A	Autosomal recessive		(CFA10)
MERTK	2q13	c-mer protooncogene receptor tyrosine kinase	Autosomal recessive		(CFA17)
PDE6B	4p16.3	Rod cGMP phosphodiesterase beta subunit	Autosomal recessive	Dominant congenital stationary night blindness	(CFA20)
PRCD	17q25.1	Progressive rod-cone degeneration protein	Autosomal recessive		(CFA9)
RHO	3q22.1	Rhodopsin	Autosomal dominant; autosomal recessive	Dominant congenital stationary night blindness	(CFA20)
SAG	2q37.1	Arrestin (s-antigen)	Autosomal recessive	Recessive Oguchi disease	(CFA25)
TTC8	14q32.11	Tetratricopeptide repeat domain 8	Autosomal recessive	Recessive Bardet-Biedl syndrome	(CFA8)
With the exception of RHO, which makes up approximately 5.25% of cases, all others genes in this list make up less than 1% of cases identified in Daiger et al.'s 2013 paper.					

Appendix 3 Genes Associated with Non-syndromic Autosomal Dominant Retinitis Pigmentosa (adRP) from Fahim et al.'s summary paper.

Gene	Frequency	OMIM Phenotype Description
NRL	Rare	613750
PRPF3	Rare	601414
PRPF6	Rare	613983
PRPF8	1% of persons with adRP	600059
PRPF31 [4]	Unlikely cause of disease [3]	600138
PRPH2	5% in Japan; rare in UK	608133
RDH12	Rare	608830
RHO	2%-3%	613731
ROM1	1%-2%	608133
RP1	1%-2%	180100
RP9 (formerly *PAP1*)	Rare	180104
RPE65	1%	613794
SEMA4A	Rare	610282
SNRNP200	2%-3%	610359
SPP2	Rare	602637
TOPORS	1%	609923

Appendix 4 Proposed mRNA for the *Ovis aries* ortholog of the *RP1* gene

Note: The mutation site corresponding to the position of the R677X mutation in humans is highlighted in red. The I695 guide is highlighted in green and its cut site is underlined.

ATGAGTGAAACTCCTTCCACCAGTTTCTCCATGGTTCGCCGGATCTCCTCTGAAGGTCAGCTT
CCTTCTCTTCGCCAGTCGGGCATCACCCAGCCTGTTGTGGCCAAAAGGATCAGTTTCTACAA
GAGCGGAGACCCTCAGTTCGGCGGGGTCCGGGTGGTGCTGAACCCTCGTTCCTTCAAGACAT
TCGATGCTCTGCTGGACAACCTGTCGGGCAAGGTGCCCCTGCCCTTCGGGGTGCGGAACATC
AGCACCCCCCGCGGGAGGCACAGCATCACGCGCCTGGAAGAGCTGGAGGACGGTCAGTCAT
ACCTGTGCTCCCACGGCAGGAAGGTGCAGCCGGTGGACCTGGACAAGGCCCGGCGGCGCCC
GCGGCCCTGGCTCAGCAGCCGAGCTCTCAGCACGCATGTGCAGCGAGGCCCCGCCCCTGCTG
CTCCCGGTATGCTGCGCGCGCCACGACGGCTCGTGGTCTTCAGAAATGGCGACCCAAAGACG
AGGCGTGCAATCGTGCTCAACAGGAGGATCACGCAGAGCTTCGAGGTCTTCCTTCAATACCT
GACACAGGTCATGCAGCGCCCGGTGACCAAGCTGTATGCCACAGACGGAAGGGTTCCCAGT
CTGCAGGCTGTGATCCTGAGCTCCGGAGCTGTGGTGGCAGCAGGAAGGGAACCGTTTAAAC
CAGGAAATTATGACATCCAAAAGTACTTGCTTCCTGCTAGATTACCAGGCATCTCTCGTCGT
GTGTACCCCAAGGGAAATGCTAGGTCAGAAAGCAGAAAACTGAGCACACATGTACCTTCAA
GCCCAACGTCTCAGATTTATTCTCTTTCTTCTGAGAAAATGCAGAGTAATGATTGCTACTCAG
ATCATTCTTTTGCTTCTGAAAATTACTTGGCATTAGAAAAAAATGATTCTCAGAATTTATTGA
TATATCCTTCTGAAGATGATGTTGAGAAATCAATTATTTTTAATCAAGATGGCACTATGACA
GTTGAGATGAAAATTCGATTCAAGATAAAAGAGGAAGAAACCATAAAATGGACAACCACTC
TCTGTAGAGCTGATCTGTCCAATAATGGTGAAAAAAGTGAAATAAGCAGTCTCCCAGGGAG
AACGGATGATCGATCATCTGGTGTAAAGATTACTGCATGTTCATTGTCCACAGACATCTCAC
CTCTGGAGAAAGGTGGTAGTCTGGTGGACAGTCTAGCGGAGGAGGTGAACACTCAAGTGAA
AGATCAAGATGTTGAAACTCGCAGTTCTACCAGCTTGGAGAACCCTGCTATGGACACAGATG
CCACCCAGGGAACTCAGGATCGAGTGAAACATCGTTTCTACAGGCCCCCTACACCTGGACCA
AGGAGAATGAGGCAGAAGAAGTCTGTGATAGGGAGTGTGACCTTAGTATCTGAAACTGAGG

TTCAAGAGAAAATGATTGGGCAGTTTTCCTACAATGAAGAAAGGAGAGATTGGGAAAACAA

GTCCGAGTATCACATGGTCACACATTCTTGCAGTAAAATGTCATCTGTGTCCAACAGACCCA

TAGTTGTTGAAGTTGATAATGATGAGCAGGTAGCATCATCTTTAGAAAGAAAAAAGGAAAG

CAGGTTGCTCAAATCAAATGCAGTAAGTGCTGGTGTTGTAGAAATTACAAGTCAGAAGATGT

TAGAGATGTCCCATAATGGTGGCTTGCCACAGACTACATCAGAAAACTCCATTGTGGAGGAA

GGAATAGTTGATAATGTCACAGCAGACAACAAAGCTAGGGTCAGGAATTTAAGAACTTATG

GTAACACCGATGATAGATCCAGCCCTTTCTTAGCAGATGCAGCTCATTCTTCAAGTAACAAC

TCTGGAACTGACAAAACTATTTCCAAGACCCCAGCTTCAGTAGGATCCTCTACTGTCACTAC

AAGAATCGACCAACTGATTCATGAATTTTCTCAGTGTGGTTTAACAAAACTTCCAGAAAATG

AAAAGCAGATTTCATCTTCAGTTGATAGCAAAAAAAAGATGAAATCTCAGCAGCATGTGAT

AAATTCTCAGCATCAGGCTGGAGAGATGGCAACTAAAAGAATCCCCAGGAAGAATAAGAGA

ATGAACACAAGAGGTAGAATTGCACAGGAAACCATATTGCGAGATTCACGTAGTCCCCTCA

AAGGGACCATACTTTGTGAGAAAGACCTCCATGCAAGTGATACAGTAATTGAATCAAATTAT

TTTTCTTCAAAAGGTAATAATCCTGTGAATTCCAGAAATTTCCATAGAAATAAATTAAATAC

TATTCATAAACCTAAGGTTCAAGGACTTTTAGCCAGAAGAAAATCCAGACCACTAAACAAA

GTAAACTTGGGGGGACCTACAAAAAGAGAAATTGATCAAGGAGAGAAAGTGTTTTCCCATA

ATGAGGTTGGATATTGCAAAAATACCTTTGAAAATCAAAATTTGTTTCATTTGTTTAACTTCC

TTGAGCAAAAACCCAGTGCTTTTTGTGGGCCAGAGTCTCAGGCAGAAACAGCATCTTGGTAT

TTGAGAGGAACGTCAAGGAGGAGTTTAGTTTCAAAAGTTAATAACTCACACATAACTTTAAG

GAGCCAGAAAAAACAAAAAAGGGATAAGTTGAAATCAGATACTACTGTAAGTAAGCAGCAT

GTCACAACTAGGGCAAATTCCTTGGCTTCTTTGGAAAAAGCTGTTTTTCCTGAGAATGTTACC

CATCATTCAGTTCAAAGTTATGTACAAAGATGGTTGCAGAACTTAAGTCCACAAGCAGCTTT

GCAGCTTGGCAAGTCAGCTCCAGTATACAAAAAGGAAAGGAGTGTGGCGAGTTACAACAAC

GGTTTTCTTCCAGGAAACAGTTCCTGCACTAGTTCTGGAAAAAGAAATGATTCTGTTATGCA

AAGTAATAGACACACAACTAAAAGTGACAGTTTGACAGGAGACAATCTAGATAAGAAAGTA

GGTATGTCTTTTGACAAAGATAGCAGTGAAGAACTCATCCAGGATCACTGTGAGAGCCAGA

CTGACTCTCTGAACGATACTTACTTGCTTTCTGTTCATGAATTCTGTACTTTGACACAGTCAG

CTATGGATGATCCTAATGCTAAAAGTCAAGTATCTGCTGCAAAGTCAGGGCAAGAGATGAG

CCTTGTTTACAAAGACATAAACCTTGCTGCAAAAGGGCCAAGCGTAGAGACTGCCGTACAG

GTAGATCTGGAAGGGGACGCCCCACAGCACTTGTCACCAGTCCAGCTGCTTCGCCAGCTGCA

AGCTTTGGCTCCTAGCAGTCCCAAGGCTCAAAATGGAGTTGTTCAGATGCCAGGTTCACTTT

CAGAAGTTCCTTTCCCTTCTTTGATATGTAATTCCTCCACTAATGTACTCCTAGCTTGGCTCCT

GGTGCTAACCCTAAAGGGAGGTGTGAATAGCTTCTGTCCAGGTGACGCTCTCAAGGCGACCA

GTGGAAGTCCAGAAACACTTGCACTGTTGGAGGTGCTGAAGCACATTGCTGTCATAGAGGA

AGCTGATGACTTGAAGGCCGCCGTGGCCAGCTTAGTGGAATCAACCACGAATCACTTTGGAC

TCACTGAGAAAGAACAGGATGTGGTTCCAATAGGTGTTTCTGCAAATTGCTCTACACCCAGC

ATTCAGATAATTCCTCAGTGTGCTGAAAATGAGAAAACACAGAAAATCTCTTTAGATGGAAG

CCATACTGCCAGTGAGGAAGTCTCTGAAGTCTGTGTTACAGCAGTGACTTGCTCTCCATGTA

AAATGGACACTGTAGTTAAGACTTACCCTCCAAAAGAGACTTGTCACCTCATTGAAGATTCT

TTCCCCAGTAATGACTGTACCACGGATCAGACTTCCATGAACAAGGCTTGTTTCTTAGGAGA

CATCTCTTCACTTACTGATGCTGTGTCTTCTCATGAGGGTTGTGCTTATGAACAAAACCATAG

CTATGAGAGAGCTGATAATTTGGAATTGACCGAAGAGTTAGAAAGAGTTGATGAAGTTCAG

AAGGACAGAAATATTTTGGCAGACCCTGAGTGTAAACACGGCTCTAATATGTTGGTGTCACA

CCAAAATATCAGTAGTTTAAGCCACTGTGGCTCTTTCCAAAATACAACTGAATCAGAGCTTG

ATGGAGAACATAGTTTTTTAGATAAATCTGGAAGTTGTTCATTAAAGAAATTTCAGGATAAA

AATGTATATACATCTTTTGATAAGGAGGATTCAAAGACTTCTGAAGAACCAGGCTCAACAAG

CAACAGCATGACATCAAGTGAAAGAAACGTCTCAGAAATGGAATCTTTTGAGGAATTAGAA

AACCAGAACACTGATATCTTTAATATAAAGGTAAATTCAGGGGAGCAACTGACTGAAGAAT

TGATCCAAAAGGAGTTAGAGGCTAGTCAAAGTTTGGAATTGATCAACGTGTCCAGCAGAAA

TGATGCTGAAGAAGGAAAGGATGGTATAATTTGTGAGACAATCAGTAGGAAACTGGTGACA

CCACCGTCATTAGTATTTTGCTATGATTCTAAGCAAAATACAGAAAAGGAGCCCAGTGAAGG

AGAAACTAAAACAAAAGTCAGAAAGATGGTGGAAAGCTTGGAAGCTGGAAGTTCTGCAGA

GTCTCCTCTTAATTTTAAAAATGGCCTAAGAAGGTCAGGAACTTCTGATTGGTCAGATTATA

GACAAAACAGTGAGAATGAACAGTCATACAAAACATCCAGCGATGGCCCCAGTGACAGTGA

TGAGGAGATGATCCCTGAGAAAGAATGCAACAAAGGATTTGTTAAAAGGACAATAGAGAAA
CTTTATGGTAAAGCAGAGATGATGAGACCATCTTTTTTTGCTGGATCTACACACACATCTCA
GGTTTATCCTTGTGATTCTGTGGAATTTCAGGGCACTGGGAAAGTAGGTCTTTATGATCCTGA
AGGTCAGTCACTTGCCTCTTTGGAACGGGTGTCTAGTAATTCAGCTGTGTTGCAGAAATTTCC
GGAGCAAAAACGAGATAAATGTGATGTTAATAACGTGAGGGACAGTTCTCCCAGGGAAGAC
ATTGCAGAACATGGTACAAAACAGAATGATCATAAAAGAATCCTCAGGGACAGGGAAGAG
GGAGTACTGATTGACAAAGGCAAGTGGCTCCTGAAAGAGAATCATTTGCTAAGAGTATCAT
CGCTGGAATGTTCTGGCCCGTGTGGCCATGCAGACACCACATCAGTGGATACTCTACTGGAT
AATAGCAGCAACGAGGTTCCGTATTCACATTTTGGAAACTTGGCTCCAGGCCCAAACATGGC
TGAACTATCCTCCTCAGAGCTAGAGGAACTGACTCAGCCTCCTGAGCTGAGATGCAATTATT
TTAATGTGCCTCATTGTAGTGACTCGGAGCCCTTCCATGACGATGAGCTGGATAGTCAAGAT
GAAGCTTGTGCTCAGGAGAGAAAACCCAATCACTCAGCAGAGGAGAAGGGTAACCTTAGAT
CAGAGAGAGTGTGTACGTCTGCCACTCATGTCTTCGCGTCTGCTGGTAACAAAGTCCATCCT
GTCTCTGATGGTGCTGTTAGGAACCAACCGTTGGCTGGTAGTAATGTAATTCATGGTGCCCTT
CAGGAAGGCGACTCTTTGGATAAACTCTATAATATCTGTGGTCAGCATTGCCCGATACTAAC
TGTGATTAACCAGCCTGTAAATGAGGAACACCGAGGATTTGCATATTGCAAAGATTCTGATG
TTGAAAATTCTTTGGGTCTCCAGTTATGGATGAAAATACACCCATGTTTACGACAGTCAAGC
AAAACCATGTTCAGAGACAAGAACAATAAAACAAGAAGTAGAAGAGCACTTACTGATAATG
CCGTTGGAAACACACATGATTGGGCTCATTTTAATAACACACTTGACTTGATGGACAGAAGG
AGAAAATTAAAACAAAGTAACTGCTTGGGCTTAGAGGAAGAAAATAATTTCAATAAATTTC
AGTCATATTTAAAGAGTTTCTTGCACACGTTGTTGTCAGTTGTGGGTCAGGTGAATTCAAATA
CACAAGACCCCAGCAGTCAGACAAAAGAAATCTTTGAAGTAGTTGATGAGAACAACAACTT
ATTAAACAGCAGATTCCAGAACTCAGGAACGAATCTCAACCAAGTAGTCAGAGAACACAGC
TATCATTTGTCCTTTGAAATGCTTGGCCAAGCCCGCCTGTTTTGCCAAGTTGAGACATTCTTA
GGTATTAGCAACAGAAGTATCTTAGAAATATTTTATATTTTTGAAGATGAAAATCTTTTCATT
TGGGAAGAGGAAAACTAA

Appendix 5 Proposed protein sequence for *RP1* in *Ovis aries* with the amino acid location of the R677X mutation highlighted in yellow (total length of 2105 amino acids)

MSETPSTSFSMVRRISSEGQLPSLRQSGITQPVVAKRISFYKSGDPQFGGVRVVLNPRSFKTFDALL
DNLSGKVPLPFGVRNISTPRGRHSITRLEELEDGQSYLCSHGRKVQPVDLDKARRRPRPWLSSRA
LSTHVQRGPAPAAPGMLRAPRRLVVFRNGDPKTRRAIVLNRRITQSFEVFLQYLTQVMQRPVTK
LYATDGRVPSLQAVILSSGAVVAAGREPFKPGNYDIQKYLLPARLPGISRRVYPKGNARSESRKL
STHVPSSPTSQIYSLSSEKMQSNDCYSDHSFASENYLALEKNDSQNLLIYPSEDDVEKSIIFNQDGT
MTVEMKIRFKIKEEETIKWTTTLCRADLSNNGEKSEISSLPGRTDDRSSGVKITACSLSTDISPLEK
GGSLVDSLAEEVNTQVKDQDVETRSSTSLENPAMDTDATQGTQDRVKHRFYRPPTPGPRRMRQ
KKSVIGSVTLVSETEVQEKMIGQFSYNEERRDWENKSEYHMVTHSCSKMSSVSNRPIVVEVDND
EQVASSLERKKESRLLKSNAVSAGVVEITSQKMLEMSHNGGLPQTTSENSIVEEGIVDNVTADNK
ARVRNLRTYGNTDDRSSPFLADAAHSSSNNSGTDKTISKTPASVGSSTVTTRIDQLIHEFSQCGLT
KLPENEKQISSSVDSKKKMKSQQHVINSQHQAGEMATKRIPRKNKRMNTRGRIAQETILRDSRSP
LKGTILCEKDLHASDTVIESNYFSSKGNNPVNSRNFHRNKLNTIHKPKVQGLLARRKSRPLNKVN
LGGPTKREIDQGEKVFSHNEVGYCKNTFENQNLFHLFNFLEQKPSAFCGPESQAETASWYLRGTS
RRSLVSKVNNSHITLRSQKKQKRDKLKSDTTVSKQHVTTRANSLASLEKAVFPENVTHHSVQSY
VQRWLQNLSPQAALQLGKSAPVYKKERSVASYNNGFLPGNSSCTSSGKRNDSVMQSNRHTTKS
DSLTGDNLDKKVGMSFDKDSSEELIQDHCESQTDSLNDTYLLSVHEFCTLTQSAMDDPNAKSQV
SAAKSGQEMSLVYKDINLAAKGPSVETAVQVDLEGDAPQHLSPVQLLRQLQALAPSSPKAQNG
VVQMPGSLSEVPFPSLICNSSTNVLLAWLLVLTLKGGVNSFCPGDALKATSGSPETLALLEVLKHI
AVIEEADDLKAAVASLVESTTNHFGLTEKEQDVVPIGVSANCSTPSIQIIPQCAENEKTQKISLDGS
HTASEEVSEVCVTAVTCSPCKMDTVVKTYPPKETCHLIEDSFPSNDCTTDQTSMNKACFLGDISS
LTDAVSSHEGCAYEQNHSYERADNLELTEELERVDEVQKDRNILADPECKHGSNMLVSHQNISS
LSHCGSFQNTTESELDGEHSFLDKSGSCSLKKFQDKNVYTSFDKEDSKTSEEPGSTSNSMTSSERN
VSEMESFEELENQNTDIFNIKVNSGEQLTEELIQKELEASQSLELINVSSRNDAEEGKDGIICETISR
KLVTPPSLVFCYDSKQNTEKEPSEGETKTKVRKMVESLEAGSSAESPLNFKNGLRRSGTSDWSD
YRQNSENEQSYKTSSDGPSDSDEEMIPEKECNKGFVKRTIEKLYGKAEMMRPSFFAGSTHTSQVY

PCDSVEFQGTGKVGLYDPEGQSLASLERVSSNSAVLQKFPEQKRDKCDVNNVRDSSPREDIAEH

GTKQNDHKRILRDREEGVLIDKGKWLLKENHLLRVSSLECSGPCGHADTTSVDTLLDNSSNEVP

YSHFGNLAPGPNMAELSSSELEELTQPPELRCNYFNVPHCSDSEPFHDDELDSQDEACAQERKPN

HSAEEKGNLRSERVCTSATHVFASAGNKVHPVSDGAVRNQPLAGSNVIHGALQEGDSLDKLYNI

CGQHCPILTVINQPVNEEHRGFAYCKDSDVENSLGLQLWMKIHPCLRQSSKTMFRDKNNKTRSR

RALTDNAVGNTHDWAHFNNTLDLMDRRRKLKQSNCLGLEEENNFNKFQSYLKSFLHTLLSVVG

QVNSNTQDPSSQTKEIFEVVDENNNLLNSRFQNSGTNLNQVVREHSYHLSFEMLGQARLFCQVE

TFLGISNRSILEIFYIFEDENLFIWEEEN

Appendix 6 Alignment of the human *RP1* gene mRNA using NM_006269.2 and the proposed ovine ortholog assembled from LOC101114620 and LOC106991348

```
Consensus  ATGAGTGAWACYCCTTCYACYRGTTTYTCCATSRTTCNNNNNATCNNNNNTCYTCTGAAG    60
H. sapiens ATGAGTGATACCCCTTCTACTGGTTTTTCCATCATTC-----ATCCTACGTCTTCTGAAG    55
O. aries   ATGAGTGAAACTCCTTCCACCAGTTTCTCCATGGTTCGCCGGATC-----TCCTCTGAAG    55

Consensus  GTCARSTTCCWYCYCYTCGCCAKTYGRGCMTCACYCAKCCTGTTGTGGCCAARMGRATCA   120
H. sapiens GTCAAGTTCCACCCCCTCGCCATTTGAGCCTCACTCATCCTGTTGTGGCCAAGCGAATCA   115
O. aries   GTCAGCTTCCTTCTCTTCGCCAGTCGGGCATCACCCAGCCTGTTGTGGCCAAAAGGATCA   115

Consensus  GTTTCTACAAGAGCGGAGACCCYCARTTCGGCGGGGTCMGGGTGGTGSTSAACCCTCGYT   180
H. sapiens GTTTCTACAAGAGCGGAGACCCCCAATTCGGCGGGGTCAGGGTGGTGGTCAACCCTCGCT   175
O. aries   GTTTCTACAAGAGCGGAGACCCTCAGTTCGGCGGGGTCCGGGTGGTGCTGAACCCTCGTT   175

Consensus  CCTTYAAGWCMTTYGATGCTCTGCTGGAYAACYTGTCNNGGNNAAGGTGCCCCTSCCYTT   240
H. sapiens CCTTTAAGTCCTTTGATGCTCTGCTGGATAACTTGTCCAGG--AAGGTGCCCCTCCCTTT   233
O. aries   CCTTCAAGACATTCGATGCTCTGCTGGACAACCTGTC--GGGCAAGGTGCCCCTGCCCTT   233

Consensus  YGGRGTGMGGAACATCAGCACCCCYCGSGGSAGGCACAGCATCACGCGCCTGGARGAGCT   300
H. sapiens TGGAGTGAGGAACATCAGCACCCCTCGGGGCAGGCACAGCATCACGCGCCTGGAGGAGCT   293
O. aries   CGGGGTGCGGAACATCAGCACCCCCCGCGGGAGGCACAGCATCACGCGCCTGGAAGAGCT   293

Consensus  GGAGGACGGYSAGTCMTACCTRTGYTCCCACGGCAGGAAGGTGCAGCCKGTRGACCTGGA   360
H. sapiens GGAGGACGGCGAGTCCTACCTATGTTCCCACGGCAGGAAGGTGCAGCCTGTAGACCTGGA   353
O. aries   GGAGGACGGTCAGTCATACCTGTGCTCCCACGGCAGGAAGGTGCAGCCGGTGGACCTGGA   353

Consensus  CAARGCCCGKCGGCGCCCGCGGCCCTGGCTCAGCAGCCGRGCYMTYAGCRCGCANTNNNC   420
H. sapiens CAAAGCCCGTCGGCGCCCGCGGCCCTGGCTCAGCAGCCGGGCCATTAGCGCGCACT---C   410
O. aries   CAAGGCCCGGCGGCGCCCGCGGCCCTGGCTCAGCAGCCGAGCTCTCAGCACGCA-TGTGC   412

Consensus  ASCGNNNCCCCRCCCCNNNNNNNTNGCTGCTCCCGGYATGCYSCGCSCSCCACGRMGSCT   480
H. sapiens ACCG---CCCCACCCCGTAGCCGTCGCTGCTCCCGGCATGCCCCGCCCCCCACGGAGCCT   467
O. aries   AGCGAGGCCCCGCCCC-------T-GCTGCTCCCGGTATGCTGCGCGCGCCACGACGGCT   464

Consensus  MGTGGTCTTCAGRAATGGCGACCCRAAGACGAGGCGTGCRRTYSTKCTSARCAGGAGGRT   540
H. sapiens AGTGGTCTTCAGGAATGGCGACCCGAAGACGAGGCGTGCGGTTCTTCTGAGCAGGAGGGT   527
O. aries   CGTGGTCTTCAGAAATGGCGACCCAAAGACGAGGCGTGCAATCGTGCTCAACAGGAGGAT   524

Consensus  CACSCAGAGCTTCGAGGYMTTYCTWCARYACCTGACASAGGTCATGCAGCGCCCKGTGRY   600
H. sapiens CACCCAGAGCTTCGAGGCATTTCTACAGCACCTGACAGAGGTCATGCAGCGCCCTGTGGT   587
O. aries   CACGCAGAGCTTCGAGGTCTTCCTTCAATACCTGACACAGGTCATGCAGCGCCCGGTGAC   584

Consensus  CAAGCTGTAYGCYACRGACGGANNNAGGGTTCCCAGYCTSCAGGCWGTGATCCTGAGCTC   660
H. sapiens CAAGCTGTACGCTACGGACGGAAGGAGGGTTCCCAGCCTCCAGGCAGTGATCCTGAGCTC   647
O. aries   CAAGCTGTATGCCACAGACGGA---AGGGTTCCCAGTCTGCAGGCTGTGATCCTGAGCTC   641

Consensus  YGGAGCTGTGGTGGCRGCAGGAAGGGARCCRTTTAAACCAGGAAATTATGACATCCAAAA   720
H. sapiens TGGAGCTGTGGTGGCGGCAGGAAGGGAGCCATTTAAACCAGGAAATTATGACATCCAAAA   707
O. aries   CGGAGCTGTGGTGGCAGCAGGAAGGGAACCGTTTAAACCAGGAAATTATGACATCCAAAA   701

Consensus  RTACTTGCTTCCTGCTAGATTACCAGGSATCTCTCRKCGTGTGTACCCCAAGGGAAATGC   780
H. sapiens ATACTTGCTTCCTGCTAGATTACCAGGGATCTCTCAGCGTGTGTACCCCAAGGGAAATGC   767
O. aries   GTACTTGCTTCCTGCTAGATTACCAGGCATCTCTCGTCGTGTGTACCCCAAGGGAAATGC   761

Consensus  WARGTCAGAAAGCAGAAANNNGAGCACACNNATGTNNCTTCAAGCYCAASGTCYCAGATT   840
H. sapiens AAAGTCAGAAAGCAGAAA---GAGCACACATATGT--CTTCAAGCTCAAGGTCCCAGATT   822
O. aries   TAGGTCAGAAAGCAGAAAACTGAGCACAC--ATGTACCTTCAAGCCCAACGTCTCAGATT   819
```

```
Consensus  TATTCTSTTTCTTCTGAGAAAAYRCAKARTAATGATTGCTACTYAGAYYATTCTTTTGYT   900
H. sapiens TATTCTGTTTCTTCTGAGAAAACACATAATAATGATTGCTACTTAGACTATTCTTTTGTT   882
O. aries   TATTCTCTTTCTTCTGAGAAAATGCAGAGTAATGATTGCTACTCAGATCATTCTTTTGCT   879

Consensus  YCTGAAAAKTACTTGGCMTTAGAAAARAATGATTCTCAGAATTTAYYRATATATCCTTCT   960
H. sapiens CCTGAAAAGTACTTGGCCTTAGAAAAGAATGATTCTCAGAATTTACCAATATATCCTTCT   942
O. aries   TCTGAAAATTACTTGGCATTAGAAAAAAATGATTCTCAGAATTTATTGATATATCCTTCT   939

Consensus  GAAGATGATRTTGAGAAATCAATTATTTTTAATCAAGAYGGCACTATGACAGTTGAGATG  1020
H. sapiens GAAGATGATATTGAGAAATCAATTATTTTTAATCAAGACGGCACTATGACAGTTGAGATG  1002
O. aries   GAAGATGATGTTGAGAAATCAATTATTTTTAATCAAGATGGCACTATGACAGTTGAGATG   999

Consensus  AAARTTCGATTCARRATAAAAGAGGAAGAAACCATAAAATGGACAACYACTSTCWGTARA  1080
H. sapiens AAAGTTCGATTCAGAATAAAAGAGGAAGAAACCATAAAATGGACAACTACTGTCAGTAAA  1062
O. aries   AAAATTCGATTCAAGATAAAAGAGGAAGAAACCATAAAATGGACAACCACTCTCTGTAGA  1059

Consensus  RCTGRTCYKTCYAATAATGRTGAAAARAGTGARATNNGNAGTYTYCCAGGRAGAACRGAN  1140
H. sapiens ACTGGTCCTTCTAATAATGATGAAAAGAGTGAGAT--G-AGTTTTCCAGGAAGAACAGA-  1118
O. aries   GCTGATCTGTCCAATAATGGTGAAAAAAGTGAAATAAGCAGTCTCCCAGGGAGAACGGAT  1119

Consensus  NANNTCGATCATCTGGTKTAAAGMTTRCWGCATGTTCATTSTCYRCAGAYRTSTCACCTM  1200
H. sapiens -AAGTCGATCATCTGGTTTAAAGCTTGCAGCATGTTCATTCTCTGCAGATGTGTCACCTA  1177
O. aries   GA--TCGATCATCTGGTGTAAAGATTACTGCATGTTCATTGTCCACAGACATCTCACCTC  1177

Consensus  TGGANNGAANNGYRGTARTCWRGWGGRCAGTYTRGCRGAGGAGRTRAACAYTCAARTGAM  1260
H. sapiens TGGAGCGAA--GCAGTAATCAAGAGGGCAGTTTGGCAGAGGAGATAAACATTCAAATGAC  1235
O. aries   TGGA--GAAAGGTGGTAGTCTGGTGGACAGTCTAGCGGAGGAGGTGAACACTCAAGTGAA  1235

Consensus  AGATCAAGWKGYTGAAACTYGCAGTTCTRCYAGYTKGGAGAANNNTGCTANNNTGGACAC  1320
H. sapiens AGATCAAGTGGCTGAAACTTGCAGTTCTGCTAGTTGGGAGAA---TGCTACTGTGGACAC  1292
O. aries   AGATCAAGATGTTGAAACTCGCAGTTCTACCAGCTTGGAGAACCCTGCTA---TGGACAC  1292

Consensus  AGATRYCAYCCAGGGAACTCARGAYCRAGYRAARCATCGTTTYTAYAGGCCCCCTACACC  1380
H. sapiens AGATATCATCCAGGGAACTCAAGACCAAGCAAAGCATCGTTTTTATAGGCCCCCTACACC  1352
O. aries   AGATGCCACCCAGGGAACTCAGGATCGAGTGAAACATCGTTTCTACAGGCCCCCTACACC  1352

Consensus  TGGACYAAGRAGARTGAGRCARAAGAARTCTGTGATWGGSAGTGTGACCTTAGTATCTGA  1440
H. sapiens TGGACTAAGAAGAGTGAGACAAAAGAAATCTGTGATTGGCAGTGTGACCTTAGTATCTGA  1412
O. aries   TGGACCAAGGAGAATGAGGCAGAAGAAGTCTGTGATAGGGAGTGTGACCTTAGTATCTGA  1412

Consensus  AACTGAGGTTCAAGAGAAAATGATTGGRCAGTTTTCMTAYARTGAAGAAAGGNGARANTK  1500
H. sapiens AACTGAGGTTCAAGAGAAAATGATTGGACAGTTTTCATATAGTGAAGAAAGG-GAAAGTG  1471
O. aries   AACTGAGGTTCAAGAGAAAATGATTGGGCAGTTTTCCTACAATGAAGAAAGGAGAGA-TT  1471

Consensus  GGGAAAACAAGTCYGAGTATCACATGKTYACACATTCTTGCAGTAAAATGTCATCWGTRT  1560
H. sapiens GGGAAAACAAGTCTGAGTATCACATGTTTACACATTCTTGCAGTAAAATGTCATCAGTAT  1531
O. aries   GGGAAAACAAGTCCGAGTATCACATGGTCACACATTCTTGCAGTAAAATGTCATCTGTGT  1531

Consensus  CYAACARACCMRTASTTGTTNNGANNTYRATAAYRATGAKCANNNGGNAGSANTCATCWT  1620
H. sapiens CTAACAAACCAGTACTTGTTCAGA--TCAATAACAATGATCAAATGG-AGGAGTCATCAT  1588
O. aries   CCAACAGACCCATAGTTGTT--GAAGTTGATAATGATGAGCA---GGTAGCA-TCATCTT  1585

Consensus  TAGAAAGAAAAAAGGAAARCAGKYTGCTYAARTCAARTGCARTAAGTGCTGGTGTTRTAG  1680
H. sapiens TAGAAAGAAAAAAGGAAAACAGTCTGCTTAAGTCAAGTGCAATAAGTGCTGGTGTTATAG  1648
O. aries   TAGAAAGAAAAAAGGAAAGCAGGTTGCTCAAATCAAATGCAGTAAGTGCTGGTGTTGTAG  1645

Consensus  AAATTACAAGTCAGAAGATGTTAGAGATGTCMCATAATRRTGGYTTGCCANCANACTAYA  1740
H. sapiens AAATTACAAGTCAGAAGATGTTAGAGATGTCACATAATAATGGTTTGCCATCA-ACTATA  1707
```

```
O. aries   AAATTACAAGTCAGAAGATGTTAGAGATGTCCCATAATGGTGGCTTGCCA-CAGACTACA  1704

Consensus  TCARAWAACTCMATTGTGGAGGAAGNNANNTAGTTGATWRTGTNNNRGYANNNGACAACA  1800
H. sapiens TCAAATAACTCAATTGTGGAGGAAG--ATGTAGTTGATTGTGT---GGTATTGGACAACA  1762
O. aries   TCAGAAAACTCCATTGTGGAGGAAGGAA--TAGTTGATAATGTCACAGCA---GACAACA  1759

Consensus  AARCTNNGGTNNCARGAAYTTMARAACTTATGGTAACACCRATGATAGRTYCAGYCCNNT  1860
H. sapiens AAACT--GGTATCAAGAACTTCAAAACTTATGGTAACACCAATGATAGGTTCAGTCCTAT  1820
O. aries   AAGCTAGGGT--CAGGAATTTAAGAACTTATGGTAACACCGATGATAGATCCAGCCC--T  1815

Consensus  TTCNNAGCAGATGCARCYCATTYTTCAAGTAAYAACTCTGGAACTGACAAAAMTATTTCY  1920
H. sapiens TTC--AGCAGATGCAACCCATTTTTCAAGTAATAACTCTGGAACTGACAAAAATATTTCT  1878
O. aries   TTCTTAGCAGATGCAGCTCATTCTTCAAGTAACAACTCTGGAACTGACAAAACTATTTCC  1875

Consensus  RAGRCYCCAGCTTCAGWAGSATCCTCTACTGTCACTRCAAGAATYGACMRACTRATTMAT  1980
H. sapiens GAGGCTCCAGCTTCAGAAGCATCCTCTACTGTCACTGCAAGAATTGACAGACTAATTAAT  1938
O. aries   AAGACCCCAGCTTCAGTAGGATCCTCTACTGTCACTACAAGAATCGACCAACTGATTCAT  1935

Consensus  GAATTTKCTCAGTGTGGTTTAACAAAACTTCCARAAAATGAAAAGMAGATTTNNNCATCT  2040
H. sapiens GAATTTGCTCAGTGTGGTTTAACAAAACTTCCAAAAAATGAAAAGAAGATTTTGTCATCT  1998
O. aries   GAATTTTCTCAGTGTGGTTTAACAAAACTTCCAGAAAATGAAAAGCAGATTT---CATCT  1992

Consensus  NNNGTTGMYAGCAAAAARAAGAWRAAATCTCRRCAGCAWGYRATAAATTCNCAGSNATCA  2100
H. sapiens ---GTTGCCAGCAAAAAGAAGAAAAAATCTCGACAGCAAGCAATAAATTC-CAGGTATCA  2054
O. aries   TCAGTTGATAGCAAAAAAAAGATGAAATCTCAGCAGCATGTGATAAATTCTCAGC-ATCA  2051

Consensus  RGMTGGASAGMTKGCAACYAAARGAATYCYYARKAAGAATRAGAGAATRAACACAARAGG  2160
H. sapiens AGATGGACAGCTTGCAACCAAAGGAATTCTTAATAAGAATGAGAGAATAAACACAAAAGG  2114
O. aries   GGCTGGAGAGATGGCAACTAAAAGAATCCCCAGGAAGAATAAGAGAATGAACACAAGAGG  2111

Consensus  TAGAATTRCAMAGGAAAYSATAKTGCRAGATTCASRTAGTCCCCTYAAANNNGGGANNNT  2220
H. sapiens TAGAATTACAAAGGAAATGATAGTGCAAGATTCAGATAGTCCCCTTAAAGGAGGGA---T  2171
O. aries   TAGAATTGCACAGGAAACCATATTGCGAGATTCACGTAGTCCCCTCAAA---GGGACCAT  2168

Consensus  ACTTTGTGAGRAAGACCTCCAKRMAAGTGATACWGTAATTGAATCAAATWMTTTTTSTTC  2280
H. sapiens ACTTTGTGAGGAAGACCTCCAGAAAAGTGATACTGTAATTGAATCAAATACTTTTTGTTC  2231
O. aries   ACTTTGTGAGAAAGACCTCCATGCAAGTGATACAGTAATTGAATCAAATTATTTTTCTTC  2228

Consensus  MAAARGTAATNNNAATYCYRYGAWTTCCARRAATTTCCATAGAAATAAATTAAATACTAY  2340
H. sapiens CAAAAGTAATCTCAATTCCACGATTTCCAAGAATTTCCATAGAAATAAATTAAATACTAC  2291
O. aries   AAAAGGTAAT---AATCCTGTGAATTCCAGAAATTTCCATAGAAATAAATTAAATACTAT  2285

Consensus  TCANAAANNCCNAAGGTTCAAGGACTTTTARCCARAAGAAAATCYAGAYCACTAAAYAAA  2400
H. sapiens TCA-AAATTCC-AAGGTTCAAGGACTTTTAACCAAAAGAAAATCTAGATCACTAAATAAA  2349
O. aries   TCATAAA--CCTAAGGTTCAAGGACTTTTAGCCAGAAGAAAATCCAGACCACTAAACAAA  2343

Consensus  RTAARCTTRGGRGSACCTAMAAAAAGAGAAATYGRTCAARGAGAKAAAGTGTTTYCYCAY  2460
H. sapiens ATAAGCTTAGGAGCACCTAAAAAAAGAGAAATCGGTCAAAGAGATAAAGTGTTTCCTCAC  2409
O. aries   GTAAACTTGGGGGGACCTACAAAAAGAGAAATTGATCAAGGAGAGAAAGTGTTTTCCCAT  2403

Consensus  AATGARKYTRRATATTGCAAAARTACYTTTGAAAANCAAAANTTTRTTTCATKTRTTTAA  2520
H. sapiens AATGAATCTAAATATTGCAAAAGTACTTTTGAAAA-CAAAAGTTTATTTCATGTATTTAA  2468
O. aries   AATGAGGTTGGATATTGCAAAAATACCTTTGAAAATCAAAA-TTTGTTTCATTTGTTTAA  2462

Consensus  CWTCCTTGAGCAAAAACCCARWGMTTTTTRTGSRCCRSARTCTCARGCAGAARYRGCATC  2580
H. sapiens CATCCTTGAGCAAAAACCCAAAGATTTTTATGCACCGCAATCTCAAGCAGAAGTGGCATC  2528
O. aries   CTTCCTTGAGCAAAAACCCAGTGCTTTTTGTGGGCCAGAGTCTCAGGCAGAAACAGCATC  2522

Consensus  TKGGTATTTGAGAGGAAYGKCAARGARGAGTTTAGTTTCAAAAGTTAMTRAYTCACACAT  2640
```

```
H. sapiens TGGGTATTTGAGAGGAATGGCAAAGAAGAGTTTAGTTTCAAAAGTTACTGATTCACACAT  2588
O. aries   TTGGTATTTGAGAGGAACGTCAAGGAGGAGTTTAGTTTCAAAAGTTAATAACTCACACAT  2582

Consensus  AACTTTAARRAGCCAGAAAAAACRWAAARGGGATAARKTGAAAKCARRTRCTAYTKTAAG  2700
H. sapiens AACTTTAAAAAGCCAGAAAAAACGTAAAGGGGATAAAGTGAAAGCAAGTGCTATTTTAAG  2648
O. aries   AACTTTAAGGAGCCAGAAAAAACAAAAAAGGGATAAGTTGAAATCAGATACTACTGTAAG  2642

Consensus  TAARCARCATGYYACAACYAGGGCAAATTCYTTRGCTTCTTTGRAAAAASCTGWTTTTCC  2760
H. sapiens TAAACAACATGCTACAACCAGGGCAAATTCTTTAGCTTCTTTGAAAAAACCTGATTTTCC  2708
O. aries   TAAGCAGCATGTCACAACTAGGGCAAATTCCTTGGCTTCTTTGGAAAAAGCTGTTTTTCC  2702

Consensus  TGAGRMTRTTRCYCATCATTCARTTCAAARTTATRTACARAGWTGGTTGCAGAACWTAAR  2820
H. sapiens TGAGGCTATTGCTCATCATTCAATTCAAAATTATATACAGAGTTGGTTGCAGAACATAAA  2768
O. aries   TGAGAATGTTACCCATCATTCAGTTCAAAGTTATGTACAAAGATGGTTGCAGAACTTAAG  2762

Consensus  TCCAYAWSCARCTTTRMAGCYTRKMAARTCAGCTCCAGTATRYARAAAKGAAASGAGTGT  2880
H. sapiens TCCATATCCAACTTTAAAGCCTATAAAATCAGCTCCAGTATGTAGAAATGAAACGAGTGT  2828
O. aries   TCCACAAGCAGCTTTGCAGCTTGGCAAGTCAGCTCCAGTATACAAAAAGGAAAGGAGTGT  2822

Consensus  GGYRARTTNNANCAAYAAYRGTTTTNTNNCAGGRAAYNGWTCCNNCNNNACWARTTCTGG  2940
H. sapiens GGTAAATTGTAGCAATAATAGTTTT-T--CAGGGAAT-GATCC--CCATACAAATTCTGG  2882
O. aries   GGCGAGTT--A-CAACAACGGTTTTCTTCCAGGAAACAGTTCCTGC---ACTAGTTCTGG  2876

Consensus  AAAAAKAARTRATTYTGTTATGSAAAGTAATARRCACAYAACTAAAAKTGMCRGTTTGAC  3000
H. sapiens AAAAATAAGTAATTTTGTTATGGAAAGTAATAAGCACATAACTAAAATTGCCGGTTTGAC  2942
O. aries   AAAAAGAAATGATTCTGTTATGCAAAGTAATAGACACACAACTAAAAGTGACAGTTTGAC  2936

Consensus  AGGAGAYAATCTAKRTAARRARGKAGRTAWGTCTTNNNTTGMCAAWGANNNCANNNGTGA  3060
H. sapiens AGGAGATAATCTATGTAAAGAGGGAGATAAGTCTTTTATTGCCAATGA---CACTGGTGA  2999
O. aries   AGGAGACAATCTAGATAAGAAAGTAGGTATGTCTT---TTGACAAAGATAGCA---GTGA  2990

Consensus  AGAANNNNNNNNNGATCWCYRTGAGASMCAGRYTGRMTCTCTGAAYGATRCTTAYTTGST  3120
H. sapiens AGAA---------GATCTCCATGAGACACAGGTTGGATCTCTGAATGATGCTTATTTGGT  3050
O. aries   AGAACTCATCCAGGATCACTGTGAGAGCCAGACTGACTCTCTGAACGATACTTACTTGCT  3050

Consensus  TYCNNTGNNCATGAAYWCTGTACTTTGWCACAGTCAGCTATKRATGATCMTAATRCTAAA  3180
H. sapiens TCCCCTG--CATGAACACTGTACTTTGTCACAGTCAGCTATTAATGATCATAATACTAAA  3108
O. aries   TTC--TGTTCATGAATTCTGTACTTTGACACAGTCAGCTATGGATGATCCTAATGCTAAA  3108

Consensus  AGTCANNTATNNCTGCTGMAAARTCAGGRCMAGAGAWRARMCTTGTTTACMARGAMATAA  3240
H. sapiens AGTCA--TATAGCTGCTGAAAAATCAGGACCAGAGAAAAAACTTGTTTACCAGGAAATAA  3166
O. aries   AGTCAAGTAT--CTGCTGCAAAGTCAGGGCAAGAGATGAGCCTTGTTTACAAAGACATAA  3166

Consensus  ACCTWGCTRSAAAARGGCMAAGYGTAGAGRCTGCCRTWCARGTAGATCNNNTRGAAGRGG  3300
H. sapiens ACCTAGCTAGAAAAAGGCAAAGTGTAGAGGCTGCCATTCAAGTAGATCCTATAGAAGAGG  3226
O. aries   ACCTTGCTGCAAAAGGGCCAAGCGTAGAGACTGCCGTACAGGTAGATC---TGGAAGGGG  3223

Consensus  AMRCYCCAMARSACYTSTYACCAGTCCWGMTGCTTCRCCARYTGCAAGCTTYRGYTCCTR  3360
H. sapiens AAACTCCAAAAGACCTCTTACCAGTCCTGATGCTTCACCAATTGCAAGCTTCAGTTCCTG  3286
O. aries   ACGCCCCACAGCACTTGTCACCAGTCCAGCTGCTTCGCCAGCTGCAAGCTTTGGCTCCTA  3283

Consensus  GYAKTCMCAAGRCTCARAATGGAGTTGTTCARATGCCAGGTTCACTTKCAGRWGTTCCYT  3420
H. sapiens GTATTCACAAGACTCAGAATGGAGTTGTTCAAATGCCAGGTTCACTTGCAGGTGTTCCCT  3346
O. aries   GCAGTCCCAAGGCTCAAAATGGAGTTGTTCAGATGCCAGGTTCACTTTCAGAAGTTCCTT  3343

Consensus  TYCMTTCNNTGNNATATGTAATTCMTCCACTAATSTMCTYCTAGCTTGGCTCYTGGTGCT  3480
H. sapiens TTCATTC--TGCAATATGTAATTCATCCACTAATCTCCTTCTAGCTTGGCTCTTGGTGCT  3404
O. aries   TCCCTTCTTTG--ATATGTAATTCCTCCACTAATGTACTCCTAGCTTGGCTCCTGGTGCT  3401
```

```
Consensus  AAMCCTAAAGGGARGTRTGAATAGCTTCTGTCMAGKTGAYGCTCWCAAGGCKACCANNRS  3540
H. sapiens AAACCTAAAGGGAAGTATGAATAGCTTCTGTCAAGTTGATGCTCACAAGGCTACCA--AC  3462
O. aries   AACCCTAAAGGGAGGTGTGAATAGCTTCTGTCCAGGTGACGCTCTCAAGGCGACCAGTGG  3461

Consensus  AARTCNNCAGAAACACTTGCAYTGTTGGAGRTKCTRAAGCACATWGCTRTCAYAGAGGAA  3600
H. sapiens AAATCTTCAGAAACACTTGCATTGTTGGAGATTCTAAAGCACATAGCTATCACAGAGGAA  3522
O. aries   AAGTC--CAGAAACACTTGCACTGTTGGAGGTGCTGAAGCACATTGCTGTCATAGAGGAA  3519

Consensus  GCTGATGACTTGAARGCYGCYGTKGCCARYTTAGTGGARTCAACYACRARYCACTTTGGA  3660
H. sapiens GCTGATGACTTGAAAGCTGCTGTTGCCAATTTAGTGGAGTCAACTACAAGCCACTTTGGA  3582
O. aries   GCTGATGACTTGAAGGCCGCCGTGGCCAGCTTAGTGGAATCAACCACGAATCACTTTGGA  3579

Consensus  CTCASTGAGAAAGAACARGAYRTGGTTCCAATAGRTSTTTCTGCAAATTGYTCNNCACSN  3720
H. sapiens CTCAGTGAGAAAGAACAAGACATGGTTCCAATAGATCTTTCTGCAAATTGTTC--CACGG  3640
O. aries   CTCACTGAGAAAGAACAGGATGTGGTTCCAATAGGTGTTTCTGCAAATTGCTCTACACC-  3638

Consensus  NCARCATTCAGAKWRTTCCTMANNGTGCNNTGAAAATGARARAACACANNGAAWMTCNTC  3780
H. sapiens TCAACATTCAGAGTGTTCCTAA--GTGCAGTGAAAATGAAAGAACACAAGGAATCTCCTC  3698
O. aries   -CAGCATTCAGATAATTCCTCAGTGTGC--TGAAAATGAGAAAACACA--GAAAATC-TC  3692

Consensus  TTTRGATGGARGNNNCNNTNCTGCCAGTGAGGMANGTNNCYCTGAAGTCTGTGTTWYRGM  3840
H. sapiens TTTGGATGGAGGTTGC--T-CTGCCAGTGAGGCATGTGCCCCTGAAGTCTGTGTTTTGGA  3755
O. aries   TTTAGATGGAAG---CCATACTGCCAGTGAGGAA-GT--CTCTGAAGTCTGTGTTACAGC  3746

Consensus  AGTGACTTGCTCTCCATGTRARATGKRCACTGTARWTAAGRCTTAYYCTCCAAAAGAGAC  3900
H. sapiens AGTGACTTGCTCTCCATGTGAGATGTGCACTGTAAATAAGGCTTATTCTCCAAAAGAGAC  3815
O. aries   AGTGACTTGCTCTCCATGTAAAATGGACACTGTAGTTAAGACTTACCCTCCAAAAGAGAC  3806

Consensus  WTGTMACCYCAKTGAMRMTTYTTTYCCYAGTRATGNNNAYNNTGTNNNNNGGATCAGACT  3960
H. sapiens ATGTAACCCCAGTGACACTTTTTTTCCTAGTGATGGTTATGGTGT-----GGATCAGACT  3870
O. aries   TTGTCACCTCATTGAAGATTCTTTCCCCAGTAATG---AC--TGTACCACGGATCAGACT  3861

Consensus  TCYATGAAYAAGGCTTGTTTCYTAGGAGASRTCTSTTCACTTACTGATRCTGTGTYTTCT  4020
H. sapiens TCTATGAATAAGGCTTGTTTCCTAGGAGAGGTCTGTTCACTTACTGATACTGTGTTTTCT  3930
O. aries   TCCATGAACAAGGCTTGTTTCTTAGGAGACATCTCTTCACTTACTGATGCTGTGTCTTCT  3921

Consensus  SATRAGGSTTGTGCTYANNAAMNNARAACCATASCTATGAGRGAGCNNNNNNNNNTGATN  4080
H. sapiens GATAAGGCTTGTGCTCA--AAAGGAGAACCATACCTATGAGGGAGCTTGCCCAATTGATG  3988
O. aries   CATGAGGGTTGTGCTTATGAAC--AAAACCATAGCTATGAGAGAGC---------TGAT-  3969

Consensus  NNNNNNNNNNNNNNNNNNNNNNNNNNAATNNNNTTGNNNNNNNNNNNNNNNNNGAANNNN  4140
H. sapiens AGACCTACGTTCCTGTCAATGTCTGCAATACCATTGACTTTTTAAACTCCAAAGAAAACA  4048
O. aries   --------------------------AAT----TTG-----------------GAA----  3978

Consensus  NNNNNNNNNNNNNNTTGNNNNNNACYGAAGAGTTAGAAAGAGKTGATGAMRTTCAGAARG  4200
H. sapiens CATATACTGATAACTTGGATTCAACTGAAGAGTTAGAAAGAGGTGATGACATTCAGAAAG  4108
O. aries   --------------TTG------ACCGAAGAGTTAGAAAGAGTTGATGAAGTTCAGAAGG  4018

Consensus  AYMKAAATATTTTGRCAGACCCTGARTRTAAAMAYGGMTYTAATAYRTTGGTGTCACAYC  4260
H. sapiens ATCTAAATATTTTGACAGACCCTGAATATAAAAATGGATTTAATACATTGGTGTCACATC  4168
O. aries   ACAGAAATATTTTGGCAGACCCTGAGTGTAAACACGGCTCTAATATGTTGGTGTCACACC  4078

Consensus  AAAATRTCAGTARTTTAAGCYMCTGTGGCNCTTTNCCWAARTNNAMAANNGAAKCAGARC  4320
H. sapiens AAAATGTCAGTAATTTAAGCTCCTGTGGC-CTTTGCCTAAGTGAAAAA--GAAGCAGAAC  4225
O. aries   AAAATATCAGTAGTTTAAGCCACTGTGGCTCTTT-CCAAAAT--ACAACTGAATCAGAGC  4135

Consensus  TTGATRNAGAANCATAGTTYTYTAGATRAWTYTGRAARTTGTTCAYTAARGAARTTTCAG  4380
H. sapiens TTGATA-AGAAACATAGTTCTCTAGATGATTTTGAAAATTGTTCACTAAGGAAGTTTCAG  4284
O. aries   TTGATGGAGAA-CATAGTTTTTTAGATAAATCTGGAAGTTGTTCATTAAAGAAATTTCAG  4194
```

```
Consensus  GATRAAAATGYATATACWTCYTTTGATAWGGARGAWYCAMRGACTTCTGAAGAACCAGGC  4440
H. sapiens GATGAAAATGCATATACTTCCTTTGATATGGAAGAACCACGGACTTCTGAAGAACCAGGC  4344
O. aries   GATAAAAATGTATATACATCTTTTGATAAGGAGGATTCAAAGACTTCTGAAGAACCAGGC  4254

Consensus  TCAAYAASCAACAGCATGACATCAAGTGAAAGAAACRTYTCAGAAWTGGAATCTTTTGAR  4500
H. sapiens TCAATAACCAACAGCATGACATCAAGTGAAAGAAACATTTCAGAATTGGAATCTTTTGAA  4404
O. aries   TCAACAAGCAACAGCATGACATCAAGTGAAAGAAACGTCTCAGAAATGGAATCTTTTGAG  4314

Consensus  GAATTAGAAAACCAKRACACTGATATCTTTAATAYARWGGTAAATKSAGGRGAGCAANNG  4560
H. sapiens GAATTAGAAAACCATGACACTGATATCTTTAATACAGTGGTAAATGGAGGAGAGCAA--G  4462
O. aries   GAATTAGAAAACCAGAACACTGATATCTTTAATATAAAGGTAAATTCAGGGGAGCAACTG  4374

Consensus  NNACTGAAGAATTRATCCAARARGAGKTAGAGGCTAGTMAAASTTTRGAATTGATMRACR  4620
H. sapiens CCACTGAAGAATTAATCCAAGAAGAGGTAGAGGCTAGTAAAACTTTAGAATTGATAGACA  4522
O. aries   --ACTGAAGAATTGATCCAAAAGGAGTTAGAGGCTAGTCAAAGTTTGGAATTGATCAACG  4432

Consensus  TSTCYAGNMAGAANATKATGNNGAAGAARNAAAGRATGNNNNGTATAATTTRTGARAYAA  4680
H. sapiens TCTCTAGTAAGAATATTATG--GAAGAAA-AAAGAATGAACGGTATAATTTATGAAATAA  4579
O. aries   TGTCCAG-CAGAA-ATGATGCTGAAGAAGGAAAGGATG----GTATAATTTGTGAGACAA  4486

Consensus  TCAGTARGARRCTGGYRACACCACCRTCWTTAGWWTTTTGCTATGATTCTAAGCAAAATA  4740
H. sapiens TCAGTAAGAGGCTGGCAACACCACCATCTTTAGATTTTTGCTATGATTCTAAGCAAAATA  4639
O. aries   TCAGTAGGAAACTGGTGACACCACCGTCATTAGTATTTTGCTATGATTCTAAGCAAAATA  4546

Consensus  SWGAAAAGGAGMCCARTGAAGGAGAAACTAARANNNYAAAANNNNNNAWGATGGTNGAAA  4800
H. sapiens GTGAAAAGGAGACCAATGAAGGAGAAACTAAGATGGTAAAA------ATGATGGT-GAAA  4692
O. aries   CAGAAAAGGAGCCCAGTGAAGGAGAAACTAAAA---CAAAAGTCAGAAAGATGGTGGAAA  4603

Consensus  RCTNTGGAARCTGGAAGTTMTKCAGAGNNTCCTCTNNTRATTTWAAAAAWKGCMTMARAA  4860
H. sapiens ACTATGGAAACTGGAAGTTATTCAGAG--TCCTCTCCTGATTTAAAAAAATGCATCAAAA  4750
O. aries   GCT-TGGAAGCTGGAAGTTCTGCAGAGTCTCCTCT--TAATTTTAAAAATGGCCTAAGAA  4660

Consensus  GKYCAGKRACTTCTGATTGGTCAGAYTATMGRCMWRACAGTGASARTGARCAGYCATAYA  4920
H. sapiens GTCCAGTGACTTCTGATTGGTCAGACTATCGGCCTGACAGTGACAGTGAGCAGCCATATA  4810
O. aries   GGTCAGGAACTTCTGATTGGTCAGATTATAGACAAAACAGTGAGAATGAACAGTCATACA  4720

Consensus  AAACATCCAGYGATGRYCCCARTGACAGTGNNNNNGMGANNTKAYCCMWGAGAAAGAATR  4980
H. sapiens AAACATCCAGTGATGATCCCAATGACAGTG-----GCGAACTTACCCAAGAGAAAGAATA  4865
O. aries   AAACATCCAGCGATGGCCCCAGTGACAGTGATGAGGAGA--TGATCCCTGAGAAAGAATG  4778

Consensus  YAACAWAGGATTTGTTAAAAGGRCAATAGARAAACTKTAYGGTAAAGCAGAKATKATSAR  5040
H. sapiens TAACATAGGATTTGTTAAAAGGGCAATAGAAAAACTGTACGGTAAAGCAGATATTATCAA  4925
O. aries   CAACAAAGGATTTGTTAAAAGGACAATAGAGAAACTTTATGGTAAAGCAGAGATGATGAG  4838

Consensus  ACCATCTTTTTTTSCTGGRTCTACMCRCAMATCTCAGGTTTRTCCTTRTRATTCTGTGGA  5100
H. sapiens ACCATCTTTTTTTCCTGGGTCTACCCGCAAATCTCAGGTTTGTCCTTATAATTCTGTGGA  4985
O. aries   ACCATCTTTTTTTGCTGGATCTACACACACATCTCAGGTTTATCCTTGTGATTCTGTGGA  4898

Consensus  ATTTCAGKGNNNCANNNGGAAAGYARGTCTTTATGATYCTGAAGGKCAGTCAYTTGSCTC  5160
H. sapiens ATTTCAGTGTTCCA---GGAAAGCAAGTCTTTATGATTCTGAAGGGCAGTCATTTGGCTC  5042
O. aries   ATTTCAGGG---CACTGGGAAAGTAGGTCTTTATGATCCTGAAGGTCAGTCACTTGCCTC  4955

Consensus  TTYKGAACRGGTRTCTAGTARTTCAKCTRTGTTGCAGRAATTYCMGGAGSAAARACRAGA  5220
H. sapiens TTCTGAACAGGTATCTAGTAGTTCATCTATGTTGCAGGAATTCCAGGAGGAAAGACAAGA  5102
O. aries   TTTGGAACGGGTGTCTAGTAATTCAGCTGTGTTGCAGAAATTTCCGGAGCAAAAACGAGA  5015

Consensus  TAARTGTGATGTTARTRMYGTGAGGGACARTTMTYSYAGGGRWGACATTGYAGAACMTGG  5280
H. sapiens TAAGTGTGATGTTAGTGCTGTGAGGGACAATTATTGTAGGGGTGACATTGTAGAACCTGG  5162
```

```
O. aries   TAAATGTGATGTTAATAACGTGAGGGACAGTTCTCCCAGGGAAGACATTGCAGAACATGG  5075

Consensus  TACAAAACARAATGATSATARMAGAATCCTCASRGACAKRGARGARGGAGTACTGATTGA  5340
H. sapiens TACAAAACAAAATGATGATAGCAGAATCCTCACAGACATAGAGGAAGGAGTACTGATTGA  5222
O. aries   TACAAAACAGAATGATCATAAAAGAATCCTCAGGGACAGGGAAGAGGGAGTACTGATTGA  5135

Consensus  CAAAGGCAARTGGCTYCTGAAAGARAATCATTTGCTAAGRRTRTCATNNCTGNAAWRTYC  5400
H. sapiens CAAAGGCAAATGGCTTCTGAAAGAAAATCATTTGCTAAGGATGTCAT--CTG-AAAATCC  5279
O. aries   CAAAGGCAAGTGGCTCCTGAAAGAGAATCATTTGCTAAGAGTATCATCGCTGGAATGTTC  5195

Consensus  TGGCMYGTGTGGCMATGCAGACACCACATCAGTGGAYACYCTACTKGATAATARCAGCAR  5460
H. sapiens TGGCATGTGTGGCAATGCAGACACCACATCAGTGGACACCCTACTTGATAATAACAGCAG  5339
O. aries   TGGCCCGTGTGGCCATGCAGACACCACATCAGTGGATACTCTACTGGATAATAGCAGCAA  5255

Consensus  YGAGGTWCCRTATTCACATTTTGGWAAYTTGGCYCCAGGCCCAAMSATGGMTGAACTMTC  5520
H. sapiens TGAGGTACCATATTCACATTTTGGTAATTTGGCCCCAGGCCCAACGATGGATGAACTCTC  5399
O. aries   CGAGGTTCCGTATTCACATTTTGGAAACTTGGCTCCAGGCCCAAACATGGCTGAACTATC  5315

Consensus  CTCYTCAGARCTMGAGGAACTGACTCARCCYCYTGARCTRARATGCAATTAYTTTAAYRT  5580
H. sapiens CTCTTCAGAACTCGAGGAACTGACTCAACCCCTTGAACTAAAATGCAATTACTTTAACAT  5459
O. aries   CTCCTCAGAGCTAGAGGAACTGACTCAGCCTCCTGAGCTGAGATGCAATTATTTTAATGT  5375

Consensus  GCCTCATKGTAGTGACTCRGARCCYTTYCATGASGANNTGNNCTGGATNGTNNNCANNAT  5640
H. sapiens GCCTCATGGTAGTGACTCAGAACCTTTTCATGAGGACTTG--CTGGAT-GTTCGCA--AT  5514
O. aries   GCCTCATTGTAGTGACTCGGAGCCCTTCCATGACGA--TGAGCTGGATAGT---CAAGAT  5430

Consensus  GAARCYTGTGCYMAGGARAGAAWASCMAATCANTCANRCAGAGGAGAAGGGTARNCMTYA  5700
H. sapiens GAAACCTGTGCCAAGGAAAGAATAGCAAATCA-TCATACAGAGGAGAAGGGTAGTCATCA  5573
O. aries   GAAGCTTGTGCTCAGGAGAGAAAACCCAATCACTCA-GCAGAGGAGAAGGGTAA-CCTTA  5488

Consensus  GNTCAGARAGAGTRTGYACRTCTGYCACTCATNTCNNNNNTTNNNNTCTGCTGGTAACAA  5760
H. sapiens G-TCAGAAAGAGTATGCACATCTGTCACTCAT-TCCTTTATT----TCTGCTGGTAACAA  5627
O. aries   GATCAGAGAGAGTGTGTACGTCTGCCACTCATGTC-----TTCGCGTCTGCTGGTAACAA  5543

Consensus  AGTCYAYCCTGTCTCTGATGRTGCTRTTARRAACCAACCRTTGSCTGGYAGTAATRTRAT  5820
H. sapiens AGTCTACCCTGTCTCTGATGATGCTATTAAAAACCAACCATTGCCTGGCAGTAATATGAT  5687
O. aries   AGTCCATCCTGTCTCTGATGGTGCTGTTAGGAACCAACCGTTGGCTGGTAGTAATGTAAT  5603

Consensus  TCATGGTRCMCTTCAGGAAGSYGACTCTTTGGATAAACTSTATRMTMTYTGTGGTCARCA  5880
H. sapiens TCATGGTACACTTCAGGAAGCTGACTCTTTGGATAAACTGTATGCTCTTTGTGGTCAACA  5747
O. aries   TCATGGTGCCCTTCAGGAAGGCGACTCTTTGGATAAACTCTATAATATCTGTGGTCAGCA  5663

Consensus  TTGCCCRATACTAACTGTKATTAWCCARCCNNTGNNAATGAGGAASACCGAGGATTTGCA  5940
H. sapiens TTGCCCAATACTAACTGTTATTATCCAACCCATG--AATGAGGAAGACCGAGGATTTGCA  5805
O. aries   TTGCCCGATACTAACTGTGATTAACCAGCC--TGTAAATGAGGAACACCGAGGATTTGCA  5721

Consensus  TATYGCAAAGAWTCTGATRTTGAAAATTYYTTGGGTYTYYAKTTATGGATGAAAATACAC  6000
H. sapiens TATCGCAAAGAATCTGATATTGAAAATTTCTTGGGTTTTTATTTATGGATGAAAATACAC  5865
O. aries   TATTGCAAAGATTCTGATGTTGAAAATTCTTTGGGTCTCCAGTTATGGATGAAAATACAC  5781

Consensus  CCATRTTTACKWCAGWCARRCAAAANNANNTGTTCAGRGAMRAGAACAATAAARCANNNA  6060
H. sapiens CCATATTTACTTCAGACAGACAAAA--ATGTGTTCAGGGAAGAGAACAATAAAGCA---A  5920
O. aries   CCATGTTTACGACAGTCAAGCAAAACCA--TGTTCAGAGACAAGAACAATAAAACAAGAA  5839

Consensus  GTAKRAGANCANNNNCTTAYTGATAATGCCRTTGGWRAYAYAYWTGATYRGKYTYATTTY  6120
H. sapiens GTATGAGA-CAAAATCTTATTGATAATGCCATTGGTGATATATTTGATCAGTTTTATTTC  5979
O. aries   GTAGAAGAGCA----CTTACTGATAATGCCGTTGGAAACACACATGATTGGGCTCATTTT  5895

Consensus  ARTAACACAYTTGACTTGATGGRYARAAGRAGAAAAYWAAAAMRAAKTAACTKCTTGGGS  6180
```

```
H. sapiens AGTAACACATTTGACTTGATGGGTAAAAGAAGAAAACAAAAAAGAATTAACTTCTTGGGG  6039
O. aries   AATAACACACTTGACTTGATGGACAGAAGGAGAAAATTAAAACAAAGTAACTGCTTGGGC  5955

Consensus  TTAGAGGAAGAARRTAATTTMAAKAAATTTCARYCAKATTTNNNNNAAAGNNNNNNNNNN  6240
H. sapiens TTAGAGGAAGAAGGTAATTTAAAGAAATTTCAACCAGATTTGAAGGAAAGGTTTTGTATG  6099
O. aries   TTAGAGGAAGAAAATAATTTCAATAAATTTCAGTCATATTT-----AAAG----------  6000

Consensus  ARTTTCTTGCACACRTYRTTGTYAGTTGTGGGTMAKGTGRATTCAAATACACAAGACCYC  6300
H. sapiens AATTTCTTGCACACATCATTGTTAGTTGTGGGTAATGTGGATTCAAATACACAAGACCTC  6159
O. aries   AGTTTCTTGCACACGTTGTTGTCAGTTGTGGGTCAGGTGAATTCAAATACACAAGACCCC  6060

Consensus  AGCRGTCAGACAAAWGAAATCTTTRAAGYAGTYGATGAGAAYAACAACTTATTAAAYARC  6360
H. sapiens AGCGGTCAGACAAATGAAATCTTTAAAGCAGTCGATGAGAATAACAACTTATTAAATAAC  6219
O. aries   AGCAGTCAGACAAAAGAAATCTTTGAAGTAGTTGATGAGAACAACAACTTATTAAACAGC  6120

Consensus  AGATTCCAGRRCTCARGAACRAATCTCAACCAAGTAGTMAGAGAAMANNNCARCTRTCAT  6420
H. sapiens AGATTCCAGGGCTCAAGAACAAATCTCAACCAAGTAGTAAGAGAAAATATCAACTGTCAT  6279
O. aries   AGATTCCAGAACTCAGGAACGAATCTCAACCAAGTAGTCAGAGAACA---CAGCTATCAT  6177

Consensus  TWSTYCTTTGAAATGCTTGGYCAAGCYYGCCTSTNNNNNNTTTGCCAAGTTGAGACMTYC  6480
H. sapiens TACTTCTTTGAAATGCTTGGTCAAGCTTGCCTCTTAGATATTTGCCAAGTTGAGACCTCC  6339
O. aries   TTGTCCTTTGAAATGCTTGGCCAAGCCCGCCTGT------TTTGCCAAGTTGAGACATTC  6231

Consensus  TTARRTATTAGCAACAGAARTATYTTAGANNNAYTTTRTATKTTTGARGRTGAAAATCTT  6540
H. sapiens TTAAATATTAGCAACAGAAATATTTTAGA---ACTTTGTATGTTTGAGGGTGAAAATCTT  6396
O. aries   TTAGGTATTAGCAACAGAAGTATCTTAGAAATATTTTATATTTTTGAAGATGAAAATCTT  6291

Consensus  TTCATTTGGGAAGAGGAARACNNNNTAAATTTAACTGATCTTGAAAGCAGTAGAGAACAA  6600
H. sapiens TTCATTTGGGAAGAGGAAGACATATTAAATTTAACTGATCTTGAAAGCAGTAGAGAACAA  6456
O. aries   TTCATTTGGGAAGAGGAAAAC----TAA-------------------------------  6347

Consensus  GAAGATTTATAA                                                6612
H. sapiens GAAGATTTATAA                                                6468
O. aries   ------------                                                6315
```

Appendix 7 Protein alignment of the human *RP1* gene using NP_006260.1 and the proposed ovine ortholog assembled from LOC101114620 and LOC106991348

```
Consensus  MSXTPSTXFSXXXXXSSEGQXPXXRXXXJTXPVVAKRISFYKSGDPQFGGVRVVXNPRSF    60
H. sapiens MSDTPSTGFSIIHPTSSEGQVPPPRHLSLTHPVVAKRISFYKSGDPQFGGVRVVVNPRSF    60
O. aries   MSETPSTSFSMVRRISSEGQLPSLRQSGITQPVVAKRISFYKSGDPQFGGVRVVLNPRSF    60

Consensus  KXFDALLDNLSXKVPLPFGVRNISTPRGRHSITRLEELEDGZSYLCSHGRKVQPVDLDKA   120
H. sapiens KSFDALLDNLSRKVPLPFGVRNISTPRGRHSITRLEELEDGESYLCSHGRKVQPVDLDKA   120
O. aries   KTFDALLDNLSGKVPLPFGVRNISTPRGRHSITRLEELEDGQSYLCSHGRKVQPVDLDKA   120

Consensus  RRRPRPWLSSRAJSXHXXXXPXAXAAPGMXRXPRXLVVFRNGDPKTRRAXXLXRRXTQSF   180
H. sapiens RRRPRPWLSSRAISAHSPPHPVAVAAPGMPRPPRSLVVFRNGDPKTRRAVLLSRRVTQSF   180
O. aries   RRRPRPWLSSRALSTHVQRGP-APAAPGMLRAPRRLVVFRNGDPKTRRAIVLNRRITQSF   179

Consensus  EXFLQXLTZVMQRPVXKLYATDGRXVPSLQAVILSSGAVVAAGREPFKPGNYDIQKYLLP   240
H. sapiens EAFLQHLTEVMQRPVVKLYATDGRRVPSLQAVILSSGAVVAAGREPFKPGNYDIQKYLLP   240
O. aries   EVFLQYLTQVMQRPVTKLYATDGRKVPSLQAVILSSGAVVAAGREPFKPGNYDIQKYLLP   239

Consensus  ARLPGISXRVYPKGNAXSESRKJSTHXXSSXXSQIYSXSSEKXXXNDCYXDXSFXXEXYL   300
```

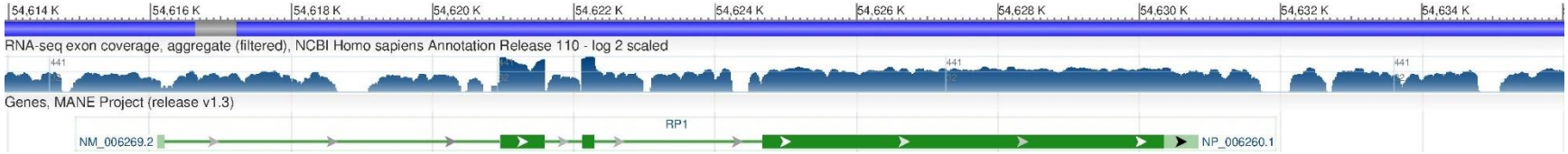
```
H. sapiens ARLPGISQRVYPKGNAKSESRKISTHMSSSSRSQIYSVSSEKTHNNDCYLDYSFVPEKYL   300
O. aries   ARLPGISRRVYPKGNARSESRKLSTHVPSSPTSQIYSLSSEKMQSNDCYSDHSFASENYL   299

Consensus  ALEKNDSQNLXIYPSEDDXEKSIIFNQDGTMTVEMKXRFXIKEEETIKWTTTXXXXXXSN   360
H. sapiens ALEKNDSQNLPIYPSEDDIEKSIIFNQDGTMTVEMKVRFRIKEEETIKWTTTVSKTGPSN   360
O. aries   ALEKNDSQNLLIYPSEDDVEKSIIFNQDGTMTVEMKIRFKIKEEETIKWTTTLCRADLSN   359

Consensus  NXEKSEXSXXPGRTXXRSSGXKJXACSXSXDXSPXEXXXXXXXSLAEEXNXQXXDQXXET   420
H. sapiens NDEKSEMS-FPGRTESRSSGLKLAACSFSADVSPMERSSNQEGSLAEEINIQMTDQVAET   419
O. aries   NGEKSEISSLPGRTDDRSSGVKITACSLSTDISPLEKGGSLVDSLAEEVNTQVKDQDVET   419

Consensus  XSSXSXENXXXDTDXXQGTQDXXKHRFYRPPTPGXRRXRQKKSVIGSVTLVSETEVQEKM   480
H. sapiens CSSASWENATVDTDIIQGTQDQAKHRFYRPPTPGLRRVRQKKSVIGSVTLVSETEVQEKM   479
O. aries   RSSTSLENPAMDTDATQGTQDRVKHRFYRPPTPGPRRMRQKKSVIGSVTLVSETEVQEKM   479

Consensus  IGQFSYXEERXXXENKSEYHMXTHSCSKMSSVSNXPXXVZXBNBXQXXXSSLERKKEXXL   540
H. sapiens IGQFSYSEERESGENKSEYHMFTHSCSKMSSVSNKPVLVQINNNDQMEESSLERKKENSL   539
O. aries   IGQFSYNEERRDWENKSEYHMVTHSCSKMSSVSNRPIVVEVDNDEQVA-SSLERKKESRL   538

Consensus  LKSXAXSAGVXEITSQKMLEMSHNXGLPXTXSXNSIVEEXXVDXVXXDNKXXXXNXXTYG   600
H. sapiens LKSSAISAGVIEITSQKMLEMSHNNGLPSTISNNSIVEEDVVDCVVLDNKTGIKNFKTYG   599
O. aries   LKSNAVSAGVVEITSQKMLEMSHNGGLPQTTSENSIVEEGIVDNVTADNKARVRNLRTYG   598

Consensus  NTBDRXSPXXADAXHXSSNNSGTDKXISXXPASXXSSTVTXRIDXLIXEFXQCGLTKLPX   660
H. sapiens NTNDRFSPISADATHFSSNNSGTDKNISEAPASEASSTVTARIDRLINEFAQCGLTKLPK   659
O. aries   NTDDRSSPFLADAAHSSSNNSGTDKTISKTPASVGSSTVTTRIDQLIHEFSQCGLTKLPE   658

Consensus  NEKXIXSSVXSKKKXKSXQXXINSXXQXGZXATKXIXXKNXRXNTXGRIXXEXIXXDSXS   720
H. sapiens NEKKILSSVASKKKKKSRQQAINSRYQDGQLATKGILNKNERINTKGRITKEMIVQDSDS   719
O. aries   NEKQISSSVDSKKKMKSQQHVINSQHQAGEMATKRIPRKNKRMNTRGRIAQETILRDSRS   718

Consensus  PLKGXILCEXDLXXSDTVIESNXFXSKXNXNXXXSXNFHRNKLNTXXXXKVQGLLXXRKS   780
H. sapiens PLKGGILCEEDLQKSDTVIESNTFCSKSNLNSTISKNFHRNKLNTTQNSKVQGLLTKRKS   779
O. aries   PLKGTILCEKDLHASDTVIESNYFSSKGN-NPVNSRNFHRNKLNTIHKPKVQGLLARRKS   777

Consensus  RXLNKXXLGXPXKREIXQXXKVFXHNEXXYCKXTFENXXLFHXFNXLEQKPXXFXXPZSQ   840
H. sapiens RSLNKISLGAPKKREIGQRDKVFPHNESKYCKSTFENKSLFHVFNILEQKPKDFYAPQSQ   839
O. aries   RPLNKVNLGGPTKREIDQGEKVFSHNEVGYCKNTFENQNLFHLFNFLEQKPSAFCGPESQ   837

Consensus  AEXASXYLRGXXXXSLVSKVXBSHITLXSQKKXKXDKXKXXXXXSKQHXTTRANSLASLX   900
H. sapiens AEVASGYLRGMAKKSLVSKVTDSHITLKSQKKRKGDKVKASAILSKQHATTRANSLASLK   899
O. aries   AETASWYLRGTSRRSLVSKVNNSHITLRSQKKQKRDKLKSDTTVSKQHVTTRANSLASLE   897

Consensus  KXXFPEXXXHHSXQXYXQXWLQNJXPXXXLXXXKSAPVXXXEXSVXXXXNXXXXGNXXXT   960
H. sapiens KPDFPEAIAHHSIQNYIQSWLQNINPYPTLKPIKSAPVCRNETSVVNCSNNSFSGNDPHT   959
O. aries   KAVFPENVTHHSVQSYVQRWLQNLSPQAALQLGKSAPVYKKERSVASYNNGFLPGNSSCT   957

Consensus  XSGKXXBXVMZSNXHXTKXXXLTGDNLXKXXXXSFXXBXXXEXXZDXXEXQXXSLNDXYL  1020
H. sapiens NSGKISNFVMESNKHITKIAGLTGDNLCKEGDKSFIANDTGE--EDLHETQVGSLNDAYL  1017
O. aries   SSGKRNDSVMQSNRHTTKSDSLTGDNLDKKVGMSFDKDSSEELIQDHCESQTDSLNDTYL  1017

Consensus  XXXHEXCTLXQSAXBDXNXKSXXXAXKSGXEXXLVYXXINLAXKXXSVEXAXQVDXJEXX  1080
H. sapiens VPLHEHCTLSQSAINDHNTKSHIAAEKSGPEKKLVYQEINLARKRQSVEAAIQVDPIEEE  1077
O. aries   LSVHEFCTLTQSAMDDPNAKSQVSAAKSGQEMSLVYKDINLAAKGPSVETAVQVD-LEGD  1076

Consensus  XPXXLXPVXXLXQLQAXXPXXXKXQNGVVQMPGSLXXVPFXSXICNSSTNXLLAWLLVLX  1140
H. sapiens TPKDLLPVLMLHQLQASVPGIHKTQNGVVQMPGSLAGVPFHSAICNSSTNLLLAWLLVLN  1137
O. aries   APQHLSPVQLLRQLQALAPSSPKAQNGVVQMPGSLSEVPFPSLICNSSTNVLLAWLLVLT  1136
```

```
Consensus  LKGXXNSFCXXDAXKATXXSXETLALLEXLKHIAXXEEADDLKAAVAXLVESTTXHFGLX  1200
H. sapiens LKGSMNSFCQVDAHKATNKSSETLALLEILKHIAITEEADDLKAAVANLVESTTSHFGLS  1197
O. aries   LKGGVNSFCPGDALKATSGSPETLALLEVLKHIAVIEEADDLKAAVASLVESTTNHFGLT  1196

Consensus  EKEQDXVPIXXSANCSTXXIQXXPXCXENEXTQXISXLDGXXXASEXXXXEVCVXXVTCS  1260
H. sapiens EKEQDMVPIDLSANCSTVNIQSVPKCSENERTQGISSLDGGCSASEACAPEVCVLEVTCS  1257
O. aries   EKEQDVVPIGVSANCSTPSIQIIPQCAENEKTQKIS-LDGSHTASEEVS-EVCVTAVTCS  1254

Consensus  PCXMXTVXKXYXPKETCXXXXXXFPSBXXXXDQTSMNKACFLGXXXSLTDXVXSXXXCAX  1320
H. sapiens PCEMCTVNKAYSPKETCNPSDTFFPSDGYGVDQTSMNKACFLGEVCSLTDTVFSDKACAQ  1317
O. aries   PCKMDTVVKTYPPKETCHLIEDSFPSNDCTTDQTSMNKACFLGDISSLTDAVSSHEGCAY  1314

Consensus  XZNHXYEXAXXXXXXXXXXXXXXXXXXXXXXXXXXXDNLXXTEELERXDXXQKDXNILXD  1380
H. sapiens KENHTYEGACPIDETYVPVNVCNTIDFLNSKENTYTDNLDSTEELERGDDIQKDLNILTD  1377
O. aries   EQNHSYERA--------------------------DNLELTEELERVDEVQKDRNILAD  1347

Consensus  PEXKXGXNXLVSHQNXSXLSXCGXXXXXXEXELDXXHSXLDXXXXCSLXKFQDXNXYTSF  1440
H. sapiens PEYKNGFNTLVSHQNVSNLSSCGLCLSEKEAELDKKHSSLDDFENCSLRKFQDENAYTSF  1437
O. aries   PECKHGSNMLVSHQNISSLSHCGSFQNTTESELDGEHSFLDKSGSCSLKKFQDKNVYTSF  1407

Consensus  DXEXXXTSEEPGSXXNSMTSSERNXSEXESFEELENXBTDIFNXXVNXGEQXTEELIQXE  1500
H. sapiens DMEEPRTSEEPGSITNSMTSSERNISELESFEELENHDTDIFNTVVNGGEQATEELIQEE  1497
O. aries   DKEDSKTSEEPGSTSNSMTSSERNVSEMESFEELENQNTDIFNIKVNSGEQLTEELIQKE  1467

Consensus  XEASXXLELIBXSSXNXXEEXXXBGIIXEXISXXLXTPPSLXFCYDSKQNXEKEXXEGET  1560
H. sapiens VEASKTLELIDISSKNIMEEKRMNGIIYEIISKRLATPPSLDFCYDSKQNSEKETNEGET  1557
O. aries   LEASQSLELINVSSRNDAEEGK-DGIICETISRKLVTPPSLVFCYDSKQNTEKEPSEGET  1526

Consensus  KXXVXXMVXXXEXGSXXESXXBXKXXJXXXXTSDWSDYRXBSXXEQXYKTSSDXPXDSXE  1620
H. sapiens KM-VKMMVKTMETGSYSESSPDLKKCIKSPVTSDWSDYRPDSDSEQPYKTSSDDPNDSGE  1616
O. aries   KTKVRKMVESLEAGSSAESPLNFKNGLRRSGTSDWSDYRQNSENEQSYKTSSDGPSDSDE  1586

Consensus  XXXXEKEXNXGFVKRXIEKLYGKAXXXXPSFFXGSTXXSQVXPXBSVEFQXXXKXXLYDX  1680
H. sapiens -LTQEKEYNIGFVKRAIEKLYGKADIIKPSFFPGSTRKSQVCPYNSVEFQCSRKASLYDS  1675
O. aries   EMIPEKECNKGFVKRTIEKLYGKAEMMRPSFFAGSTHTSQVYPCDSVEFQGTGKVGLYDP  1646

Consensus  EGQSXXSXEXVSSXSXXLQXFXEZXXDKCDVXXVRDXXXRXDIXEXGTKQNDXXRILXDX  1740
H. sapiens EGQSFGSSEQVSSSSSMLQEFQEERQDKCDVSAVRDNYCRGDIVEPGTKQNDDSRILTDI  1735
O. aries   EGQSLASLERVSSNSAVLQKFPEQKRDKCDVNNVRDSSPREDIAEHGTKQNDHKRILRDR  1706

Consensus  EEGVLIDKGKWLLKENHLLRXSSXEXXGXCGXADTTSVDTLLDNXSXEVPYSHFGNLAPG  1800
H. sapiens EEGVLIDKGKWLLKENHLLRMSS-ENPGMCGNADTTSVDTLLDNNSSEVPYSHFGNLAPG  1794
O. aries   EEGVLIDKGKWLLKENHLLRVSSLECSGPCGHADTTSVDTLLDNSSNEVPYSHFGNLAPG  1766

Consensus  PXMXELSSSELEELTQPXELXCNYFNXPHXSDSEPFHXDXLDXXBEXCAXERXXNHXXEE  1860
H. sapiens PTMDELSSSELEELTQPLELKCNYFNMPHGSDSEPFHEDLLDVRNETCAKERIANHHTEE  1854
O. aries   PNMAELSSSELEELTQPPELRCNYFNVPHCSDSEPFHDDELDSQDEACAQERKPNHSAEE  1826

Consensus  KGXXXSERVCTSXTHXFXSAGNKVXPVSDXAXXNQPLXGSNXIHGXLQEXDSLDKLYXJC  1920
H. sapiens KGSHQSERVCTSVTHSFISAGNKVYPVSDDAIKNQPLPGSNMIHGTLQEADSLDKLYALC  1914
O. aries   KGNLRSERVCTSATHVFASAGNKVHPVSDGAVRNQPLAGSNVIHGALQEGDSLDKLYNIC  1886

Consensus  GQHCPILTVIXQPXNEEXRGFAYXKXSDXENXLGXXLWMKIHPXLXQXXKXXFRXXNNKX  1980
H. sapiens GQHCPILTVIIQPMNEEDRGFAYRKESDIENFLGFYLWMKIHPYLLQTDKNVFREENNKA  1974
O. aries   GQHCPILTVINQPVNEEHRGFAYCKDSDVENSLGLQLWMKIHPCLRQSSKTMFRDKNNKT  1946

Consensus  XXRXXLXDNAXGBXXDXXXFXNTXDLMXXRRKXKXXNXLGLEEEXNXXKFQXXLKXXXXX  2040
H. sapiens SMRQNLIDNAIGDIFDQFYFSNTFDLMGKRRKQKRINFLGLEEEGNLKKFQPDLKERFCM  2034
O. aries   RSRRALTDNAVGNTHDWAHFNNTLDLMDRRRKLKQSNCLGLEEENNFNKFQSYLK-----  2001
```

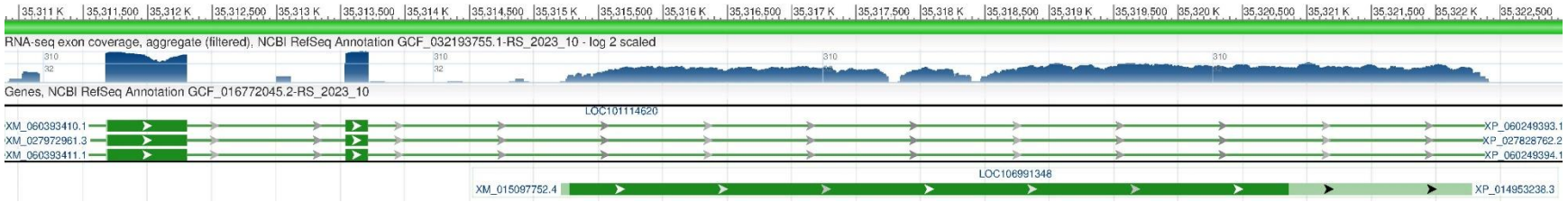
```
Consensus  XFLHTXLXVVGXVBSNTQDXSXQTXEIFXXVDENNNLLNXRFQXSXTNLNQVVREXXXXH  2100
H. sapiens NFLHTSLLVVGNVDSNTQDLSGQTNEIFKAVDENNNLLNNRFQGSRTNLNQVVRENINCH  2094
O. aries   SFLHTLLSVVGQVNSNTQDPSSQTKEIFEVVDENNNLLNSRFQNSGTNLNQVVREH-SYH  2060

Consensus  XXFEMLGQAXLXXXCQVETXLXISNRXILEJXXXFEXENLFIWEEEBXXXXXXXXXXXXX  2160
H. sapiens YFFEMLGQACLLDICQVETSLNISNRNILEL-CMFEGENLFIWEEEDILNLTDLESSREQ  2153
O. aries   LSFEMLGQARLF--CQVETFLGISNRSILEIFYIFEDENLFIWEEEN-------------  2118

Consensus  XXX                                                          2163
H. sapiens EDL                                                          2156
O. aries   ---                                                          2105
```

Appendix 8 Images of Gene annotation in NCBI's Genome Data Viewer for the *RP1* gene in *Homo sapiens* (GrCh38 p.14 Primary Assembly) and the two genes with similarity to *RP1* identified in the *Ovis* aries genome (ARS-UI_Ramb_v3.0 Primary Assembly).

RP1 in *Homo sapiens*



RP1 in *Ovis aries*



References

Ayuso, C., & Millan, J. M. (2010). Retinitis pigmentosa and allied conditions today: a paradigm of translational research. *Genome medicine, 2*(5), 34. https://doi.org/10.1186/gm155

Berson, E. L., Weigel-DiFranco, C., Rosner, B., Gaudio, A. R., & Sandberg, M. A. (2018). Association of Vitamin A Supplementation With Disease Course in Children With Retinitis Pigmentosa. *JAMA ophthalmology, 136*(5), 490–495. https://doi.org/10.1001/jamaophthalmol.2018.0590

Bhardwaj, A., Yadav, A., Yadav, M., & Tanwar, M. (2022). Genetic dissection of non-syndromic retinitis pigmentosa. *Indian journal of ophthalmology, 70*(7), 2355–2385. https://doi.org/10.4103/ijo.IJO_46_22

Bunel, M., Chaudieu, G., Hamel, C., Lagoutte, L., Manes, G., Botherel, N., Brabet, P., Pilorge, P., André, C., & Quignon, P. (2019). Natural models for retinitis pigmentosa: progressive retinal atrophy in dog breeds. *Human genetics, 138*(5), 441–453. https://doi.org/10.1007/s00439-019-01999-6

Chawla, H., & Vohra, V. (2023). Retinal Dystrophies. In *StatPearls*. StatPearls Publishing. https://pubmed.ncbi.nlm.nih.gov/33232049/

Chivers, M., Li, N., Pan, F., Wieffer, H., Slowik, R., & Leartsakulpanitch, J. (2021). The Burden of X-Linked Retinitis Pigmentosa on Patients and Society: A Narrative Literature Review. *ClinicoEconomics and outcomes research: CEOR, 13*, 565–572. https://doi.org/10.2147/CEOR.S297287

Concordet, J. P., & Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic acids research, 46*(W1), W242–W245. https://doi.org/10.1093/nar/gky354

Doudna, J. A., & Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science (New York, N.Y.), 346*(6213), 1258096. https://doi.org/10.1126/science.1258096

Daiger, S. P., Bowne, S. J., & Sullivan, L. S. (2015). Genes and Mutations Causing Autosomal Dominant Retinitis Pigmentosa. *Cold Spring Harbor perspectives in medicine, 5*(10), a017129. https://doi.org/10.1101/cshperspect.a017129

Delahaye, C., & Nicolas, J. (2021). Sequencing DNA with nanopores: Troubles and biases. *PloS one, 16*(10), e0257521. https://doi.org/10.1371/journal.pone.0257521

Diakatou, M., Manes, G., Bocquet, B., Meunier, I., & Kalatzis, V. (2019). Genome Editing as a Treatment for the Most Prevalent Causative Genes of Autosomal Dominant Retinitis Pigmentosa. *International journal of molecular sciences, 20*(10), 2542. https://doi.org/10.3390/ijms20102542

Ducloyer, J. B., Le Meur, G., Cronin, T., Adjali, O., & Weber, M. (2020). La thérapie génique des rétinites pigmentaires héréditaires [Gene therapy for retinitis pigmentosa]. *Medecine sciences:M/S, 36*(6-7), 607–615. https://doi.org/10.1051/medsci/2020095

Donders, F. C. (1857). Beiträge zur pathologischen Anatomie des Auges. *Archiv für Opthalmologie 3.* 139-165. https://doi.org/10.1007/BF02720685

Fahim, A. T., Daiger, S. P., & Weleber, R. G. (2023). Nonsyndromic Retinitis Pigmentosa Overview. In M. P. Adam (Eds.) et. al., *GeneReviews*®. University of Washington, Seattle. https://pubmed.ncbi.nlm.nih.gov/20301590/

Farvardin, M., Afarid, M., Attarzadeh, A., Johari, M. K., Mehryar, M., Nowroozzadeh, M. H., Rahat, F., Peyvandi, H., Farvardin, R., & Nami, M. (2018). The Argus-II Retinal Prosthesis Implantation; From the Global to Local Successful Experience. *Frontiers in neuroscience, 12*, 584. https://doi.org/10.3389/fnins.2018.00584

Ghani, N., Bansal, J., Naidu, A., & Chaudhary, K. M. (2023). Long term positional stability of the Argus II retinal prosthesis epiretinal implant. *BMC ophthalmology, 23*(1), 70. https://doi.org/10.1186/s12886-022-02736-w

Hincal, F., Gürbay, A., & Favier, A. (2003). Biphasic response of ciprofloxacin in human fibroblast cell cultures. *Nonlinearity in biology, toxicology, medicine, 1(*4), 481–492. https://doi.org/10.1080/15401420390271083

Hu, M. L., Edwards, T. L., O'Hare, F., Hickey, D. G., Wang, J. H., Liu, Z., & Ayton, L. N. (2021). Gene therapy for inherited retinal diseases: progress and possibilities. *Clinical & experimental optometry, 104*(4), 444–454. https://doi.org/10.1080/08164622.2021.1880863

Humayun, M. S., & Lee, S. Y. (2022). Advanced Retina Implants. *Ophthalmology. Retina, 6*(10), 899–905. https://doi.org/10.1016/j.oret.2022.04.009

Istrate, M., Vlaicu, B., Poenaru, M., Hasbei-Popa, M., Salavat, M. C., & Iliescu, D. A. (2020). Photoprotection role of melanin in the human retinal pigment epithelium. Imaging techniques for retinal melanin. *Romanian journal of ophthalmology, 64*(2), 100–104. https://pubmed.ncbi.nlm.nih.gov/32685774/

Jacobsen, J. C., Bawden, C. S., Rudiger, S. R., McLaughlan, C. J., Reid, S. J., Waldvogel, H. J., MacDonald, M. E., Gusella, J. F., Walker, S. K., Kelly, J. M., Webb, G. C., Faull, R. L., Rees, M. I., & Snell, R. G. (2010). An ovine transgenic Huntington's disease model. *Human molecular genetics, 19*(10), 1873–1882. https://doi.org/10.1093/hmg/ddq063

Kaur, G., & Singh, N. K. (2021). The Role of Inflammation in Retinal Neurodegeneration and Degenerative Diseases. *International journal of molecular sciences, 23(*1), 386. https://doi.org/10.3390/ijms23010386

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., Gonzales, A. P., Li, Z., Peterson, R. T., Yeh, J. R., Aryee, M. J., & Joung, J. K. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature, 523*(7561), 481–485. https://doi.org/10.1038/nature14592

Leclere, P. G., Panjwani, A., Docherty, R., Berry, M., Pizzey, J., & Tonge, D. A. (2005). Effective gene delivery to adult neurons by a modified form of electroporation. *Journal of neuroscience methods, 142*(1), 137–143. https://doi.org/10.1016/j.jneumeth.2004.08.012

Lin, S., Staahl, B. T., Alla, R. K., & Doudna, J. A. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife, 3*, e04766. https://doi.org/10.7554/eLife.04766

Liu, Q., Collin, R. W. J., Cremers, F. P. M., den Hollander, A. I., van den Born, L. I., & Pierce, E. A. (2012). Expression of Wild-Type Rp1 Protein in Rp1 Knock-in Mice Rescues the Retinal Degeneration Phenotype. *PLoS ONE, 7*(8), e43251. https://doi.org/10.1371/journal.pone.0043251

Liu, W., Liu, S., Li, P. & Yao, K. (2022). Retinitis Pigmentosa: Progress in Molecular Pathology and Biotherapeutical Strategies. *International Journal of Molecular Sciences 23*(9), 4883. https://doi.org/10.3390/ijms23094883

Mckean, N. E. (2022). *Creation of a gene edited PSEN1 E280A sheep model of Alzheimer's Disease*. [Doctoral dissertation, University of Auckland]. Research Space. https://researchspace.auckland.ac.nz/handle/2292/64141

Mello de Queiroz, F., Sánchez, A., Agarwal, J. R., Stühmer, W., & Pardo, L. A. (2012). Nucleofection induces non-specific changes in the metabolic activity of transfected cells. *Molecular biology reports, 39*(3), 2187–2194. https://doi.org/10.1007/s11033-011-0967-z

Miller A. (2013). Retinal implant system delivers limited sight to some blind people. *CMAJ: Canadian Medical Association journal = journal de l'Association medicale canadienne, 185*(14), E659–E660. https://doi.org/10.1503/cmaj.109-4499

Mizobuchi, K., Hayashi, T., Oishi, N., Kubota, D., Kameya, S., Higasa, K., Futami, T., Kondo, H., Hosono, K., Kurata, K., Hotta, Y., Yoshitake, K., Iwata, T., Matsuura, T., & Nakano, T. (2021). Genotype-Phenotype Correlations in RP1-Associated Retinal Dystrophies: A Multi-Center Cohort Study in JAPAN. *Journal of clinical medicine, 10*(11), 2265. https://doi.org/10.3390/jcm10112265

Mo, D., Potter, B. A., Bertrand, C. A., Hildebrand, J. D., Bruns, J. R., & Weisz, O. A. (2010). Nucleofection disrupts tight junction fence function to alter membrane polarity of renal epithelial cells. American journal of physiology. *Renal physiology, 299*(5), F1178–F1184. https://doi.org/10.1152/ajprenal.00152.2010

Mojica, F. J., Ferrer, C., Juez, G., & Rodríguez-Valera, F. (1995). Long stretches of short tandem repeats are present in the largest replicons of the Archaea Haloferax mediterranei and Haloferax volcanii and could be involved in replicon partitioning. *Molecular microbiology, 17*(1), 85–93. https://doi.org/10.1111/j.1365-2958.1995.mmi_17010085.x

Nanda, A., McClements, M. E., Clouston, P., Shanks, M. E., & MacLaren, R. E. (2019). The Location of Exon 4 Mutations in RP1 Raises Challenges for Genetic Counseling and Gene Therapy. *American Journal of Ophthalmology, 202*, 23–29. https://doi.org/10.1016/j.ajo.2019.01.027

Nangia, V., Jonas, J. B., Khare, A., & Sinha, A. (2012). Prevalence of retinitis pigmentosa in India: the Central India Eye and Medical Study. *Acta ophthalmologica, 90*(8), e649–e650. https://doi.org/10.1111/j.1755-3768.2012.02396.x

Nguyen, X. T., Moekotte, L., Plomp, A. S., Bergen, A. A., van Genderen, M. M., & Boon, C. J. F. (2023). Retinitis Pigmentosa: Current Clinical Management and Emerging Therapies. *International journal of molecular sciences, 24*(8), 7481. https://doi.org/10.3390/ijms24087481

O'Neal, T. B., & Luther, E. E. (2023). Retinitis Pigmentosa. In *StatPearls*. StatPearls Publishing. https://pubmed.ncbi.nlm.nih.gov/30137803/

Pan, C., Banerjee, K., Lehmann, G. L., Almeida, D., Hajjar, K. A., Benedicto, I., Jiang, Z., Radu, R. A., Thompson, D. H., Rodriguez-Boulan, E., & Nociari, M. M. (2021). Lipofuscin causes atypical

necroptosis through lysosomal membrane permeabilization. *Proceedings of the National Academy of Sciences of the United States of America, 118*(47), e2100122118. https://doi.org/10.1073/pnas.2100122118

Parmeggiani, F., Sato, G., De Nadai, K., Romano, M. R., Binotto, A., & Costagliola, C. (2011). Clinical and Rehabilitative Management of Retinitis Pigmentosa: Up-to-Date. *Current genomics*, *12*(4), 250–259. https://doi.org/10.2174/138920211795860125

Pickar-Oliver, A., & Gersbach, C. A. (2019). The next generation of CRISPR-Cas technologies and applications. *Nature reviews. Molecular cell biology, 20*(8), 490–507. https://doi.org/10.1038/s41580-019-0131-5

Pinnapureddy, A. R., Stayner, C., McEwan, J., Baddeley, O., Forman, J., & Eccles, M. R. (2015). Large animal models of rare genetic disorders: sheep as phenotypically relevant models of human genetic disease. *Orphanet journal of rare diseases, 10*, 107. https://doi.org/10.1186/s13023-015-0327-5

Piri, N., Grodsky, J. D., & Kaplan, H. J. (2021). Gene therapy for retinitis pigmentosa. *Taiwan journal of ophthalmology, 11*(4), 348–351. https://doi.org/10.4103/tjo.tjo_47_21

Ren, R., Guo, J., Liu, G., Kang, H., Machens, H. G., Schilling, A. F., Slobodianski, A., & Zhang, Z. (2022). Nucleic acid direct delivery to fibroblasts: a review of nucleofection and applications. *Journal of biological engineering, 16*(1), 30. https://doi.org/10.1186/s13036-022-00309-5

Richards, J. E. & Hawley, R. S. (2011). Chapter 5 – We Are All Mutants: How Mutation Alters Function. In Richards, J. E. & Hawley, R. S. (Eds.) *The Human Genome* (Third Edition, pp.143-195). Academic Press. https://doi.org/10.1016/B978-0-08-091865-5.00005-9

Sakai, D., Takagi, S., Totani, K., Yamamoto, M., Matsuzaki, M., Yamanari, M., Sugiyama, S., Yokota, S., Maeda, A., Hirami, Y., Mandai, M., Takahashi, M., Nakamura, M., & Kurimoto, Y. (2022). Retinal pigment epithelium melanin imaging using polarization-sensitive optical coherence tomography for patients with retinitis pigmentosa. *Scientific reports, 12*(1), 7115. https://doi.org/10.1038/s41598-022-11192-x

Sarna, M., Olchawa, M., Zadlo, A., Wnuk, D., & Sarna, T. (2017). The nanomechanical role of melanin granules in the retinal pigment epithelium. *Nanomedicine: nanotechnology, biology, and medicine, 13*(3), 801–807. https://doi.org/10.1016/j.nano.2016.11.020

Schubert, M. S., Thommandru, B., Woodley, J., Turk, R., Yan, S., Kurgan, G., McNeill, M. S., & Rettig, G. R. (2021). Optimized design parameters for CRISPR Cas9 and Cas12a homology-directed repair. *Scientific reports, 11*(1), 19482. https://doi.org/10.1038/s41598-021-98965-y

Shinozaki, A., Hosaka, Y., Imagawa, T., & Uehara, M. (2010). Topography of ganglion cells and photoreceptors in the sheep retina. *The Journal of comparative neurology, 518*(12), 2305–2315. https://doi.org/10.1002/cne.22333

Schwartz, S. G., Wang, X., Chavis, P., Kuriyan, A. E., & Abariga, S. A. (2020). Vitamin A and fish oils for preventing the progression of retinitis pigmentosa. *The Cochrane database of systematic reviews*, *6*(6), CD008428. https://doi.org/10.1002/14651858.CD008428.pub3

Silva, R. S., Salles, M. V., Motta, F. L., & Sallum, J. M. F. (2020). Retinitis Pigmentosa Due to Rp1 Biallelic Variants. *Scientific reports, 10*(1), 1603. https://doi.org/10.1038/s41598-020-58243-9

Solano F. (2020). Photoprotection and Skin Pigmentation: Melanin-Related Molecules and Some Other New Agents Obtained from Natural Sources. *Molecules (Basel, Switzerland), 25*(7), 1537. https://doi.org/10.3390/molecules25071537

Taylor, K. R., Holzer, A. K., Bazan, J. F., Walsh, C. A. & Gleeson, J. G. (2000). Patient Mutations in Doublecortin Define a Repeaed Tubulin-binding Domain*. *Molecular Basis of Cell and Developmental Biology 275(44),* 34442-34450. https://doi.org/10.1074/jbc.M007078200

Thai, T., Salisbury, B. H. & Zito, P. M. (2023). Ciprofloxacin. In *StatsPearls.* StatsPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK535454/

ThermoFisher Scientific (2012). T123 NanoDrop Lite Interpretation of Nucleic Acid 260/280 Ratios. https://assets.thermofisher.com/TFS-Assets/CAD/Product-Bulletins/T123-NanoDrop-Lite-Interpretation-of-Nucleic-Acid-260-280-Ratios.pdf

ThermoFisher Scientific (2024). *Lipid-Based Transfection – Troubleshooting*. ThermoFisher Scientific. https://www.thermofisher.com/nz/en/home/technical-resources/technical-reference-library/transfection-support-center/lipid-transfection-support/lipid-transfection-support-troubleshooting.html

Tycko, J., Myer, V. E., & Hsu, P. D. (2016). Methods for Optimizing CRISPR-Cas9 Genome Editing Specificity. *Molecular cell, 63*(3), 355–370. https://doi.org/10.1016/j.molcel.2016.07.004

Verbakel, S. K., van Huet, R. A. C., Boon, C. J. F., den Hollander, A. I., Collin, R. W. J., Klaver, C. C. W., Hoyng, C. B., Roepman, R., & Klevering, B. J. (2018). Non-syndromic retinitis pigmentosa. *Progress in retinal and eye research, 66*, 157–186. https://doi.org/10.1016/j.preteyeres.2018.03.005

Villiger, L., Joung, J., Koblan, L., Weissman, J., Abudayyeh, O. O., & Gootenberg, J. S. (2024). CRISPR technologies for genome, epigenome and transcriptome editing. *Nature reviews. Molecular cell biology*, 10.1038/s41580-023-00697-6. Advance online publication. https://doi.org/10.1038/s41580-023-00697-6

Voisin, A., Pénaguin, A., Gaillard, A., & Leveziel, N. (2023). Stem cell therapy in retinal diseases. *Neural regeneration research, 18*(7), 1478–1485. https://doi.org/10.4103/1673-5374.361537

Walton, R. T., Christie, K. A., Whittaker, M. N., & Kleinstiver, B. P. (2020). Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science (New York, N.Y.), 368(6488)*, 290–296. https://doi.org/10.1126/science.aba8853

Wang, J., Xiao, X., Li, S., Wang, P., Sun, W., & Zhang, Q. (2021). Dominant RP in the Middle While Recessive in Both the N- and C-Terminals Due to RP1 Truncations: Confirmation, Refinement, and Questions. *Frontiers in cell and developmental biology, 9*, 634478. https://doi.org/10.3389/fcell.2021.634478

Wei, T., Cheng, Q., Min, Y. L., Olson, E. N., & Siegwart, D. J. (2020). Systemic nanoparticle delivery of CRISPR-Cas9 ribonucleoproteins for effective tissue specific genome editing. *Nature communications, 11*(1), 3232. https://doi.org/10.1038/s41467-020-17029-3

Wen, F., Wang, Y., He, D., Liao, C., Ouyang, W., Liu, Z., Li, W., & Liao, Y. (2022). Primary Culture of Porcine Retinal Pigment Epithelial Cells. *Journal of visualized experiments: JoVE*, (187), 10.3791/64244. https://doi.org/10.3791/64244

Wick, R. R., Judd, L. M., & Holt, K. E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome biology*, *20*(1), 129. https://doi.org/10.1186/s13059-019-1727-y

Wu, K. Y., Kulbay, M., Toameh, D., Xu, A. Q., Kalevar, A., & Tran, S. D. (2023). Retinitis Pigmentosa: Novel Therapeutic Targets and Drug Development. *Pharmaceutics*, *15*(2), 685. https://doi.org/10.3390/pharmaceutics15020685

Yamashita, T., Liu, J., Gao, J., LeNoue, S., Wang, C., Kaminoh, J., Bowne, S. J., Sullivan, L. S., Daiger, S. P., Zhang, K., Fitzgerald, M. E., Kefalov, V. J., & Zuo, J. (2009). Essential and synergistic roles of RP1 and RP1L1 in rod photoreceptor axoneme and retinitis pigmentosa. *The Journal of Neuroscience: the official journal of the Society for Neuroscience*, *29*(31), 9748–9760. https://doi.org/10.1523/JNEUROSCI.5854-08.2009

Yoon, C. K., Bae, K., & Yu, H. G. (2022). Longitudinal Microstructure Changes of the Retina and Choroid in Retinitis Pigmentosa. *American journal of ophthalmology*, *241*, 149–159. https://doi.org/10.1016/j.ajo.2022.05.002

Zhang, J. H., Adikaram, P., Pandey, M., Genis, A., & Simonds, W. F. (2016). Optimization of genome editing through CRISPR-Cas9 engineering. B*ioengineered*, *7*(3), 166–174. https://doi.org/10.1080/21655979.2016.1189039

Zheng, Y., Li, Y., Zhou, K., Li, T., VanDusen, N. J., & Hua, Y. (2024). Precise genome-editing in human diseases: mechanisms, strategies and applications. *Signal transduction and targeted therapy*, *9*(1), 47. https://doi.org/10.1038/s41392-024-01750-2