

# Determination of the genetic cause of an internationally unique, naturally occurring muscular dystrophy in Western Australian Merino sheep

---

This thesis is presented for the Honours degree in Biomedical Science at Murdoch University

**Jez Supreme BSc. Molecular Biology BSc. Biotechnology**

**11/1/2013**

Supervisors:

A/Prof Kim Carter

*Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia, Perth, Australia*

A/Prof Kristen Nowak

*Western Australian Institute for Medical Research, Centre for Medical Research, University of Western Australia, Perth, Australia*

A/Prof Wayne Greene

*Murdoch University, Perth, Australia*

Prof. Nigel Laing

*Western Australian Institute for Medical Research, Centre for Medical Research, University of Western Australia, Perth, Australia*

## Declaration Page

I declare this thesis is my own account of my research and contains as its main content, work which has not previously been submitted for a degree at any tertiary educational institution.

.....

**Jez Supreme**

## Abstract

Muscular dystrophies (MDs) are neuromuscular disorders characterised by chronic, usually progressive, skeletal muscle weakness. Individuals often lose walking ability and can suffer terminal cardiorespiratory complications. Determining the genetics of a disease helps provide diagnosis, prognosis, genetic counselling, and the basis for rational therapeutic design.

A naturally occurring sheep model of autosomal recessive congenital MD was identified in WA in the 1950's and preserved as a research colony. The pathological features and distribution of this MD is novel. A sheep model of MD is incredibly valuable; sheep have similar skeletal muscle mass to humans, representing a significant improvement over smaller mammals in which to trial therapies. Successfully characterising the causative gene(s) would enable a possible target for new therapies and may open new lines of investigation into better understanding and treatment of MD in humans.

This project utilised a two-pronged approach to investigate the genetics of this ovine MD. First, bioinformatics analysis of SNP genotyping for multiple individuals in the flock by a 50,000 SNP array in combination with the latest sheep genome reference build released by the International Sheep Genome Consortium, enabling homozygosity mapping, genetic linkage and association mapping. Second, molecular biological approaches further explored the identified prime candidate gene by cDNA sequencing.

This research project identified *ROCK2* as the prime candidate gene most likely harbouring a mutation causing the muscular dystrophy in this internationally unique ovine model. It also demonstrated for the first time in sheep the existence of *ROCK2m*, an isoform of *ROCK2* preferentially expressed in skeletal muscle. This work has set the stage for further investigations into *ROCK2m* and the ovine MD which will hopefully pinpoint the causative disease mutation.

## Table of Contents

|   |          |
|---|----------|
| Declaration Page.....   | I        |
| Abstract.....   | II       |
| Acknowledgements .....  | VII      |
| Table of Figures.....   | 1        |
| List of Tables .....  | 3        |
| Definitions & Abbreviations .....                                   | 4        |
| <b>1. Introduction and Literature Review .....</b>                  | <b>5</b> |
| 1.1. Overview.....  | 5        |
| 1.2. Genetics & Disease .....                                       | 6        |
| 1.2.1. Importance of Genetics.....                                  | 6        |
| 1.2.2. DNA to Protein.....  | 6        |
| 1.2.3. Genetic Disease .....  | 9        |
| 1.2.4. Heritable Disease.....                                       | 10       |
| 1.2.5. Disease & Treatment Investigation .....                      | 10       |
| 1.2.6. Animal models of Human Disease .....                         | 11       |
| 1.3. Genomics & Bioinformatics Challenges.....                      | 11       |
| 1.3.1. Genomics.....  | 11       |
| 1.3.2. Traditional and Second-Generation DNA Sequencing .....       | 14       |
| 1.3.3. Bioinformatics Challenges for NGS.....                       | 14       |
| 1.3.4. Single Nucleotide Polymorphisms .....                        | 17       |
| 1.3.5. Linkage .....  | 19       |
| 1.3.6. Linkage Disequilibrium and Genetic Association Analysis..... | 20       |
| 1.4. Neuromuscular Disorders.....                                   | 20       |
| 1.4.1. Muscular Dystrophy .....                                     | 21       |
| 1.4.2. Congenital Muscular Dystrophy.....                           | 21       |
| 1.4.3. Duchenne Muscular Dystrophy .....                            | 21       |

|           |  |           |
|-----------|--|-----------|
| 1.4.4.    | Myotonic Dystrophy .....                                       | 22        |
| 1.4.5.    | Nemaline Myopathy .....  | 23        |
| 1.4.6.    | Limb Girdle Muscular Dystrophy Type 2 .....                    | 24        |
| 1.5.      | A Sheep Model of Muscular Dystrophy .....                      | 25        |
| 1.5.1.    | The OCPMD Flock.....   | 25        |
| 1.5.2.    | Inheritance .....  | 25        |
| 1.5.3.    | Disease Morphology.....  | 25        |
| 1.5.4.    | Diagnosis .....  | 28        |
| 1.5.5.    | Gene Investigations to Date.....                               | 28        |
| 1.6.      | Comparisons of the Ovine Model with Human Disease .....        | 29        |
| 1.6.1.    | Commonalities with Human Disease .....                         | 29        |
| 1.6.2.    | Differences to Human Disease .....                             | 29        |
| 1.7.      | Bioinformatics .....   | 30        |
| 1.7.1.    | Origins of Bioinformatics .....                                | 30        |
| 1.7.2.    | Importance of Bioinformatics to this project.....              | 31        |
| 1.8.      | Research hypothesis and aims .....                             | 32        |
| 1.9.      | Significance of the Project .....                              | 33        |
| <b>2.</b> | <b>Materials and Methods .....</b>                             | <b>34</b> |
| 2.1.      | The Ovine Congenital Progressive Muscular Dystrophy Flock..... | 34        |
| 2.1.1.    | History .....  | 34        |
| 2.1.2.    | Description and definition of phenotype .....                  | 34        |
| 2.1.3.    | Overview of Pedigree .....                                     | 36        |
| 2.1.4.    | Collection of biological samples .....                         | 38        |
| 2.1.5.    | Histology and biomarkers.....                                  | 38        |
| 2.2.      | Genetics and Genomics .....                                    | 39        |
| 2.2.1.    | Sheep Genome Project.....                                      | 39        |
| 2.2.2.    | SNP Array.....   | 39        |

|   |           |
|---|-----------|
| 2.2.3. Whole-Genome Sequencing .....                            | 40        |
| 2.3. Bioinformatics .....                                       | 40        |
| 2.3.1. Analysis .....   | 40        |
| 2.3.1.1. Homozygosity Mapping.....                              | 41        |
| 2.3.1.2. Association using PLINK .....                          | 43        |
| 2.3.1.3. Linkage analysis.....                                  | 46        |
| 2.3.1.4. Identification of the Candidate Geneset.....           | 47        |
| 2.3.1.5. Biological Plausibility Investigation Candidates ..... | 47        |
| 2.3.2. Data Manipulation .....                                  | 48        |
| 2.3.2.1. SNP Array Data Manipulation.....                       | 48        |
| 2.3.2.2. SNP validation .....                                   | 48        |
| 2.3.2.3. PLINK.....   | 48        |
| 2.3.2.4. Merlin.....  | 49        |
| 2.3.2.5. Genome Sequencing .....                                | 49        |
| 2.4. Molecular Biology.....                                     | 50        |
| 2.4.1. Primer Design .....                                      | 50        |
| 2.4.2. RNA Extraction .....                                     | 52        |
| 2.4.3. cDNA synthesis .....                                     | 52        |
| 2.4.4. PCR Amplification .....                                  | 52        |
| 2.4.5. Electrophoresis of PCR products.....                     | 53        |
| 2.4.6. Sequencing of Candidate Gene.....                        | 53        |
| 2.4.7. Sequence Analysis .....                                  | 54        |
| <b>3. Results.....</b>  | <b>56</b> |
| 3.1. Summary .....  | 56        |
| 3.2. Results from Bioinformatics analyses.....                  | 56        |
| 3.2.1. Homozygosity Mapping .....                               | 57        |
| 3.2.2. Association Analysis .....                               | 59        |

|   |           |
|---|-----------|
| 3.2.3. Linkage Analysis .....   | 62        |
| 3.2.4. Whole-genome sequencing.....                                   | 63        |
| 3.2.5. Candidate Geneset .....  | 63        |
| 3.2.6. Biological plausibility investigation.....                     | 63        |
| 3.3. Molecular Biological Investigations of Prime Candidate Gene..... | 64        |
| 3.3.1. RNA Extraction .....   | 64        |
| 3.3.2. cDNA Synthesis, PCR Amplification & Electrophoresis .....      | 65        |
| 3.3.3. Sequencing.....  | 66        |
| <b>4. Discussion and Future Directions.....</b>                       | <b>68</b> |
| <b>5. Bibliography.....</b>   | <b>78</b> |
| <b>6. Appendices.....</b>   | <b>92</b> |

## Acknowledgements

This thesis would not have been possible without the support, care and guidance of myriad people. First, I must gratefully acknowledge the expert knowledge, insight, encouragement, and hours of patient teaching provided by my supervisors A/Prof Kim Carter, A/Prof Kristen Nowak, and Prof. Nigel Laing (and thanks for handling the Murdoch paperwork, A/Prof. Wayne Greene!).

With no experience in bioinformatics beyond being a videogame geek and thinking it was a cool area, Kim took me under his wing and spent FAR more time than he had to spare guiding me through the fundamentals of UNIX, bioinformatics methodologies and scripting in Perl. It might be awhile before he takes on another biology major for a bioinformatics project, but I am extremely grateful he wasted his time with me. Non-model organisms, eh?!

To Kristen, who seemed to be available and enthusiastic so much more than a human could be capable of: THANK YOU. Without your consistent enthusiasm and expert advice on everything molecular biology-related this project could never have begun (and I certainly couldn't have completed it). Without you and Kim I'm not sure my scientific career would be happening either. You guys rock.

Nigel, you are a neuromuscular disease Jedi; thanks for stooping to deal with such an ignorant Padawan.

To everyone in S203: Dr. Richard Francis (you're a doctor now, right?), Denise Anderson, and Matt Cooper, thank you so much for being so welcoming and helping me when I needed it. Thanks especially to Richard for being my backup bioinformatician whenever Kim was unavailable. You guys made TICHR way more fun than it otherwise would have been.

To Royston Ong & Elyshia McNamara: thank you both SO MUCH for your help. You guys both spent a LOT of time and effort helping me get the Mol. Bio. side of things running and disabusing my ignorance. Thanks also to Kyle Yau for proving once and for all that the linkage problems weren't my fault. Thanks too to Dr. Rachael Duff for your sequence analysis expertise, and to everyone else at WAIMR who helped me out.

Finally, thank you Nicole. Thank you for Egg.

## Table of Figures

|   |    |
|---|----|
| <b>Figure 1.1.</b> The central dogma of molecular biology. DNA is transcribed into RNA and then translated into protein. Figure from (Ianello Giasseti et al., 2013).....   | 7  |
| <b>Figure 1.2.</b> Exon splicing. Introns are removed from the pre-mRNA and exons joined. (Figure from Sparknotes, 2013).....   | 8  |
| <b>Figure 1.3.</b> Meiosis in diploid organisms. (Figure adapted from Marston and Amon, 2004).....  | 8  |
| <b>Figure 1.4.</b> Change in DNA sequencing costs over time. (Figures from Wetterstrand, 2013).....   | 13 |
| <b>Figure 1.5.</b> RNA-sequence analysis pipeline. (Figure from Mutz et al., 2013).....   | 17 |
| <b>Figure 1.6.</b> Haplotype as unit of statistical testing. (Figure from Chao, 2012).....  | 18 |
| <b>Figure 1.7.</b> Histology of healthy and dystrophic muscle in Duchenne muscular dystrophy. (Figure adapted from Davies and Nowak, 2006).....   | 22 |
| <b>Figure 1.8.</b> Genetics and pathology of myotonic dystrophy types I & II. (Figure from Turner and Hilton-Jones, 2010).....  | 22 |
| <b>Figure 1.9.</b> Child affected with nemaline myopathy.....   | 23 |
| <b>Figure 1.10.</b> Gomori trichrome staining in nemaline myopathy. (Adapted from Ilkovski et al., 2001).....   | 23 |
| <b>Figure 1.11.</b> Proteins involved in pathogenesis of limb girdle muscular dystrophy. (Figure from Zatz et al., 2003).....   | 24 |
| <b>Figure 1.12.</b> Adipose replacement in OCPMD-affected skeletal muscle tissue of the anconaeus of a 2 year old ewe from the ovine congenital muscular dystrophy flock. (Figure adapted from Richards et al., 1988a)..... | 26 |
| <b>Figure 1.13.</b> Most severely affected muscles in OCPMD-affected sheep.....   | 26 |
| <b>Figure 1.14.</b> Transverse section of vastus intermedius from a 14-week old dystrophic ram. (Richards et al., 1988a).....   | 26 |
| <b>Figure 1.15.</b> Disorganised fibrillar material containing small nemaline bodies beneath the cell membrane of dystrophic fiber (Richards and Passmore, 1989).....   | 27 |
| <b>Figure 1.16.</b> Longitudinal section of the vastus intermedius in an OCPMD affected sheep. (Richards et al., 1988a).....  | 27 |
| <b>Figure 2.1.</b> Full OCPMD pedigree inclusive of all information.....  | 36 |
| <b>Figure 2.2.</b> Pedigree of initial analysis.....  | 37 |
| <b>Figure 2.3.</b> Pedigree of final analysis.....  | 37 |

|  |    |
|--|----|
| <b>Figure 2.4.</b> Flowchart for bioinformatics analyses leading to candidates for further investigation.....                      | 41 |
| <b>Figure 2.5.</b> Flowchart for molecular biological investigation of prime candidate gene for the OCPMD pathology.....           | 50 |
| <b>Figure 3.1.</b> Electrophoresis of PCR products from <i>ROCK2</i> primer pairs.....   | 65 |
| <b>Figure 3.2.</b> Snapshot of the sequencing chromatogram highlighting the beginning of the variable region of <i>ROCK2</i> ..... | 67 |

## List of Tables

|  |    |
|--|----|
| <b>Table 1.1.</b> Subtypes of limb girdle muscular dystrophy type 2, with associated genomic positions, affected genes and associated proteins (Adapted from Laval and Bushby, 2004) ..... | 24 |
| <b>Table 2.1.</b> Source of cDNA for <i>ROCK2</i> human, mouse and cow species .....   | 51 |
| <b>Table 2.2.</b> Primer sets designed for targeting of <i>ROCK2</i> cDNA .....  | 52 |
| <b>Table 2.3.</b> PCR components for targeted amplicons within <i>ROCK2</i> .....  | 53 |
| <b>Table 2.4.</b> Sequencing reaction components for <i>ROCK2</i> PCR products .....   | 54 |
| <b>Table 2.5.</b> Thermocycler protocol for sequencing reaction.....   | 54 |
| <b>Table 3.1.</b> Longest homozygous regions observed in the SNP genotypes of affected individuals in the initial dataset.....   | 57 |
| <b>Table 3.2.</b> Longest homozygous regions observed in the SNP genotypes of affected individuals in the final dataset.....   | 58 |
| <b>Table 3.3.</b> SNP genotypes of the affected individuals within the <i>ROCK2</i> -containing run of homozygosity.....   | 58 |
| <b>Table 3.4.</b> SNP genotypes of the carrier individuals within the <i>ROCK2</i> -containing run of homozygosity.....  | 59 |
| <b>Table 3.5.</b> Top SNPs by significance for initial association analysis.....   | 60 |
| <b>Table 3.6.</b> Association analysis results of SNPs within the <i>ROCK2</i> gene in the initial dataset.....  | 60 |
| <b>Table 3.7.</b> Top SNPs by significance for final association analysis.....   | 61 |
| <b>Table 3.8.</b> Association analysis result of SNPs within the <i>ROCK2</i> gene in the final dataset.....   | 62 |
| <b>Table 3.9.</b> Linkage analysis for SNPs within the <i>ROCK2</i> -containing homozygous region.....   | 62 |
| <b>Table 3.10.</b> RNA extraction concentrations from skeletal muscle tissue.....  | 64 |
| <b>Table 3.11.</b> Primer sets amplified to the predicted size based on electrophoresis of PCR products.....   | 65 |
| <b>Table 3.12.</b> PCR products that were successfully sequenced and aligned to the bovine <i>ROCK2</i> cDNA reference.....  | 66 |

## Definitions & Abbreviations

**Allele** Alternative DNA sequence at a locus

**Amino acid** Individual components of a protein. Dictated by a 3 base codon of DNA.

**DNA** Deoxyribonucleic acid

**Exon** Segment of a gene which is translated to an end product. A given gene may have multiple exons

**Exon splicing** Process by which the exons of a gene are joined together as a step towards the production of a mature RNA product. May have alternative forms in which exons are absent, introns preserved, or the inclusion of an alternative acceptor site changes the upstream or downstream exon

**Gene** A section of DNA that gets transcribed into RNA and then into protein

**Genetic locus** A particular point in the genome

**Genetic Polymorphism** A DNA variant occurring at an incidence of less than 99%

**GWAS** Genome-Wide Association Study

**Intron** Segment of a gene which is not translated into a product. A given gene may have multiple introns.

**Map Distance/Unit** usually expressed as either map unit (MU) or centimorgan (cM) is a measure of recombination frequency between two genetic loci. One cM represents the genetic distance between genes at which one product of meiosis in 100 will lead to a recombinant product (recombination frequency of 1%).

**Moore's Law** Observation that over the history of computing hardware the speed of processors approximately doubles every 2 years; with a concomitant drop in cost

**MyD** Myotonic Dystrophy

**MD** Muscular Dystrophy

**NM** Nemaline myopathy

**NMD** NeuroMuscular Disorders

**Nucleotide** A single base of DNA. DNA is made up of four nucleotide bases; A, T, G, C