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

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 moleculr 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

List of Tables

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