3.4.3 Milestone 4.2 Control of Johne's Disease

Description	The objective is to conduct a genetic expression study with selection lines of red deer that have been bred for putative resistance (R) or susceptibility (S) to Tb and Johne's disease and which have been experimentally challenged with M. paratuberculosis. RNA expression levels in the mesenteric lymph nodes will be measured using 454 Corporation technology. Copy DNA will be produced from RNA recovered from the tissue samples and then this technique allows high-throughput DNA sequencing of this c-DNA using a novel massively parallel sequencing-by-synthesis approach. The DNA sequences will be matched with the bovine genome to identify genes associated with protection and disease. The ultimate aim is to identify genes associated with resistance to Johne's Disease in deer cattle and sheep. Candidate genes and SNP's will be investigated. The results of this study will add to the understanding of the Pathobiology of Johne's and may lead to better diagnostics and will underpin marker assisted selection.
Alignment with JDRC Strategy	Tool: Gene markers for identification of Johne's disease resistant animals
Status	Ongoing program. On target, achieved contracted milestones 2008-2011

2011 Science Report

JDRC Full Scientific Report (due 22 March 2011)

Milestone Title: Control of Johne's

Colin Mackintosh

Milestone short title:
Identification of genes with expression patterns associated with resistance and susceptibility to Johne's disease

Achievement Measures:
1. *Identification of genes associated with resistance to Johne's disease.*
2. *A publication in an international journal describing the results of the expression study.*
3. *Development of improved diagnostic tests and validation of tests developed for marker assisted selection of Johne's resistant animals.*

Year 3 Project description:
Complete a gene expression study on tissues taken from selectively bred deer showing resistance (R) or susceptibility (S) phenotypes to Johne's disease after experimental challenge with MAP. In years 1 and 2 of this study the field aspects of the two challenge studies of selectively bred deer were completed. Selected samples from the most resistant (R) and most susceptible (S) animals from the challenge studies are undergoing massive parallel sequencing using SOLiD technology. In year 3 of this study sequencing data will be analyzed to identify key genes associated with response over time to infection with MAP and the disease outcome. These key genes will be functionally annotated using comparative genomic methods, and incorporating knowledge of biological pathways and networks. Ultimately this will lead to hypotheses as to the roles specific genes play in an immune response to MAP.

Executive Summary

This project is in the third year of a four year study. Preliminary results from the 2008 study (3R and 3S animals at Weeks 4, 13 and 50) and the 2009 study (3R and 3S animals at Week 4) are exciting and differential up/down regulation of a number of key genes have been identified. We expect that the results from the full set of 2009 animals will be even more exciting.
This is leading edge research, which is dependent on a unique ruminant MAP challenge model and the latest in "next generation" sequencing to show differential gene expression between known R and S deer at key times during the disease process.
The AgResearch Invermay deer MAP challenge model is the only internationally recognized model that can reliably deliver natural Johne's disease outcomes in deer of known RS phenotype/genotype within a 12 month period.
In both studies, lymph nodes samples were collected at three time points (Week 4, Week 12/13 and at euthanasia of clinically affected deer or at Week 49/50), and frozen in liquid nitrogen to preserve messenger-RNA (mRNA). SOLiD SAGE sequencing was carried out and preliminary analysis point to a number of candidate genes differentially up- or down- regulated between R and S animals and with time. The analysis of the first set of results is scheduled to be completed by June 2011. As soon as approval is given by the JDRC, the full set of samples from 2009 is due to undergo "next generation" sequencing using Illumina RNA-Seq, which is now the preferred option.
Understanding these mechanisms is fundamental to improving the efficiency of selection, the development of better diagnostic tests, the design and development of better vaccines and the targeting of therapeutics that could improve host responses against Mycobacteria in deer, sheep and cattle.

Introduction

Worldwide, there is a universal need to understand the mechanisms of resistance to diseases, especially in the case of serious, untreatable, chronic diseases caused by intracellular pathogens such as MAP and tuberculosis (Tb). Our current methods of diagnosis, vaccination and therapeutics of these diseases are crude and largely inefficient or ineffective. Currently there are no means of preventing MAP infections in livestock and diagnostic tests are not accurate enough to eradicate disease using "test-and-slaughter". Nor are there any effective treatments and the vaccines only ameliorate disease but do not prevent it. Only by gaining a better understanding of the immunological mechanisms of protection can we hope to make significant progress in any of these areas of control against Johne's disease and other similar diseases. Knowledge of these complex mechanisms and targeting key pathways will enhance our ability to improve the efficiency of selection, develop better diagnostic tests, develop better vaccines and discover therapeutics that increase an animal's ability to cope with infection, such as somatic gene therapy.
Johne's disease affects all ruminants, and in sheep and cattle a small proportion of infected animals develop slowly progressive disease that becomes clinical in 3-4 years. Red deer develop very similar granulomatous enteritis lesions to sheep and cattle, but the disease progresses more quickly and clinical signs often occur in rising yearling animals. This has enabled us to develop an experimental challenge model that reproduces a typical range of infection/disease outcomes, but within a one year time frame. Another key advantage with the deer model is that we have identified strains of deer that are either very resistant or very susceptible to MAP infection/disease, which has not been achieved in other ruminants. We believe that the results that we obtain using the deer model will be applicable to cattle and sheep on the basis that the immune pathways related to resistance will be very similar across ruminants.

The objective of this four year study is to identify genes with expression patterns associated with resistance (R) and susceptibility (S) to Johne's disease. It is believed that the immune events quite early in the disease determine the pathway that relatively susceptible or resistant animals take in terms of progressing to disease or recovery. We have also previously observed that infection quickly spreads from the initial sites of invasion and establishment in the intestine, to the draining jejunal and ileocaecal lymph nodes, where the immune response to infection is centred. We therefore chose to surgically biopsy the jejunal lymph nodes at two time points early in the

infection viz 4 and 12/13 weeks post challenge (pc), and then at euthanasia in clinically affected animals or at the end of the study in subclinically affected animals viz 49/50 weeks pc.

This study, which was originally funded directly by FRST, started in March 2007 with the breeding of hinds by AI using two stags that had been shown to be resistant (TBR) or susceptible (TBS) to infection with *Mycobacterium bovis* (TB). Their offspring were born in November 2007 and they were challenged with MAP in April 2008. On 30 September, 2008 the JDRC took over funding this project.
The field aspects of the 2008 field study were completed in March 2009 with the slaughter of the animals. The diagnostic laboratory aspects of the study were completed in the following 4-5 months.
The 2009 field study started in April 2009 with the challenge of the offspring of two stags R or S to Johne's disease. The field aspects of the 2009 field study were completed in March 2010 with the slaughter of the animals. The diagnostic laboratory aspects of the study were completed in the following 4-5 months.
Subsequently lymph node samples underwent new generation sequencing and the challenging task of biometrical analysis and rationalizing the very large amounts of data is being undertaken in the current year.

Methodology
The materials and methods used in the 2008 and 2009 field challenges are described in the attached papers (Appendices I and II).

Samples of jejunal lymph node were collected from the 2008 study animals at 4, 12 and 50 weeks post challenge (pc), and from 2009 animals at 4, 13 and euthanasia or 49 weeks pc. Duplicate samples were frozen in liquid N immediately after collection and then stored at -80C in two different -80C freezers.

When the study was first conceived in 2006, the plan was to use cattle micro-arrays to measure gene expression in the lymph node samples collected during both studies, but by 2007 new generation sequencing became available at reasonable cost. Initially, 454 pyrosequencing was believed to offer the best prospects for measuring gene expression in deer. However, by 2009, the deer genome sequence had been built at Invermay and the bioinformatics group at Invermay, headed by Dr Rudiger Brauning, had developed the expertise necessary for dealing with shorter gene transcripts. We therefore decided to use SOLiD sequencing and SAGE (serial analysis of gene expression) libraries, which produces shorter sequences than 454 but considerably more copies and better measurement of gene expression, which was the main objective.

In the 2010-11 year selected frozen samples from the 3 most resistant and the 3 most susceptible offspring at the 3 time points (4, 12 and 50 weeks pc) from the 2008 and the Week 4 samples from the 3 most resistant and the 3 most susceptible offspring from the 2009 study were processed and submitted for new generation gene expression sequencing at the University of Otago using the Life Technologies SOLiD platform. Three SOLiD plates were divided into 8 segments, which is why we used 24 samples to fill all the plates efficiently (this explains why we added the Week 4 samples to make up the number). The budget was insufficient to include all samples from both studies and also it was considered more sensible to take a staged approach to sequencing and data analysis, so that lessons learnt from the first set could be applied to the second set.

Subsequently, advancements in the Illumina sequencing platform and in bioinformatic data handling has made this the most attractive method of sequencing to complete the sequencing of the 2009 samples. For comparison (see Appendix 1 for Glossary of terms):
SOLiD SAGE

- gives 1 tag (27 base pair fragment) per transcript. This facilitates statistical evaluation greatly

- is solely aimed at measuring expression levels
- works best with a known transcriptome
- did not show the expected uptake by the research community
- can be complemented by RNA-Seq

Illumina RNA-Seq (non-normalized library)

- gives all full-length transcripts (including alternative transcripts) that can be assembled
- measures expression levels
- mapping is relatively straight forward compared with SOLiD-SAGE
- is the favourite protocol of the research community
- more tools are available to handle sequence space (i.e. Illumina RNA-Seq) compared to colour space sequences

At the start of this expression study the recommendation was to use SOLiD SAGE as it was competitively priced and because statistical issues with RNA-Seq had not been worked out. Subsequently these bioinformatics issues have been overcome and now the recommendation is to use Illumina RNA-Seq over SOLiD SAGE in all cases where there is only an incomplete reference transcriptome available.
The Week 4 samples have already been sequenced using SOLiD SAGE, but the plan is to sequence the full 2009 series (ie including the Week 4 samples) using Illumina RNA-seq. This will allow valuable linkage of results using the two different systems.

Results

This study started in March 2007 with the breeding of hinds by AI using two stags that had been shown to be resistant (TBR) or susceptible (TBS) to infection with *Mycobacterium bovis* (TB). Their offspring were born in November 2007 and they were challenged with MAP in April 2008. The deer underwent surgical biopsy of their jejunal lymph nodes 4 and 12 weeks pc and were killed 50 weeks pc and necropsied. During the study they were weighed and blood sampled at regular intervals. The biopsy and necropsy samples were split 3 ways for histopathology, culture and frozen in liquid nitrogen for future gene expression studies. There were no clinical cases of Johne's disease, but the scoring of histopathology and culture of biopsy and necropsy samples showed a range of disease severity from very mild to very severe. Although there was no correlation with stag genotype, there was clear separation of the 3 least affected and the 3 most severely affected, and samples of these 6 animals at the 3 time points were processed and sequenced using SOLiD SAGE technology (see Appendix 2 for draft paper on the field aspects of the 2008 study).

In March 2008 a second group of unselected red hinds were inseminated with semen from two stags that had been shown to be R or S to Johne's disease. This was based on a genetic study carried out by Professor Frank Griffin using stud records and repeated blood testing of animals on prominent NZ deer stud, and analyzed by Dr Ken Dodds at AgResearch Invermay. The 18 offspring (9R and 9S) were born in November 2008 and they were challenged with MAP in April 2008, in a repeat of the previous study. The deer underwent surgical biopsy of their jejunal lymph nodes 4 and 13 weeks post challenge (pc). Three offspring (2S and R) developed clinical Johne's disease and were euthanised and necropsied. The remaining deer were euthanised 49 weeks pc and necropsied. During the study they were weighed and blood sampled at regular intervals. As before, the biopsy and necropsy samples were split 3 ways for histopathology, culture and frozen in liquid nitrogen for gene expression studies. The offspring of the R and S sires had significantly different disease outcomes ($P<0.05$).

Fig 1: Paralisa for 3 clinically affected deer

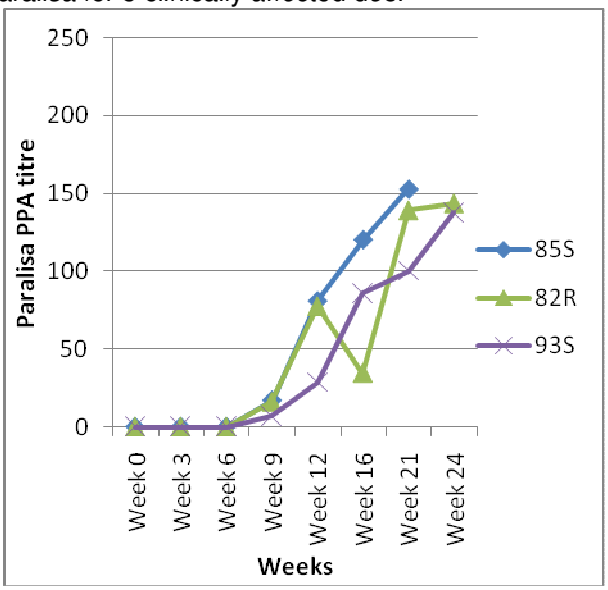

Fig 2: Paralisa for 4 minimally affected deer

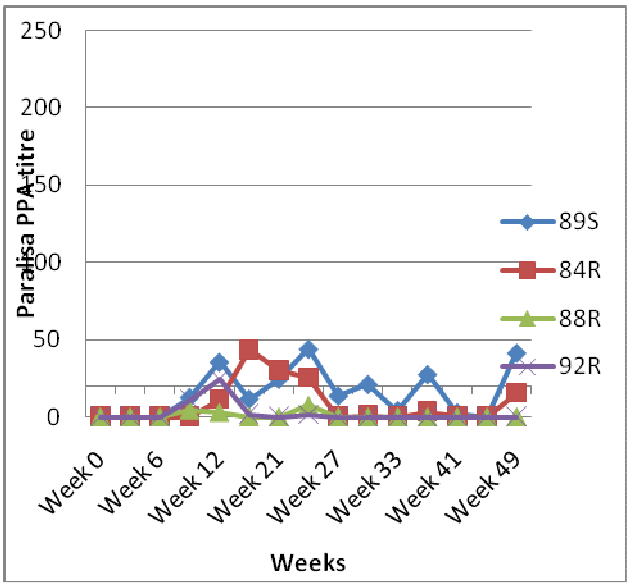

Fig 3: Paralisa for remaining deer (R offspring grey dots, S offspring black lines)

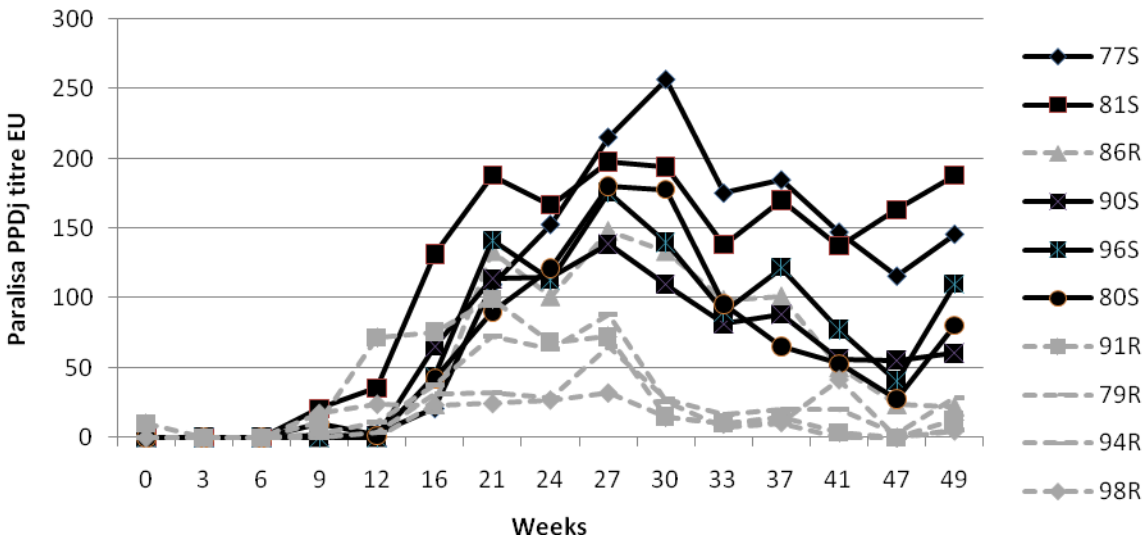

Table 1.
Histopathological lesion severity scores (LSS) for biopsy samples at Weeks 4 and 13 post challenge (pc) and gross lesion score, LSS, lesion description as multibacillary (MB) or paucibacillary (PB) and Johne's disease (JD) outcome for the offspring of the resistant sire (R), the susceptible sire (S) at euthanasia or slaughter 49 weeks pc.

			Week 4	Week 13			Week 49/euthanasia/death		
Sire	Tag	Sex	LSS	Gross	LSS	AFO	Gross	LSS	AFO
R	92	F	0	0	6	PB	0	0	None
R	91	M	0	0	6	PB	0	2	None
R	88	M	0	1	6	PB	0	5	PB
R	84	M	0	4	11	PB	0	5	PB
R	79	M	0	2	13	PB	0	5	PB
R	94	F	0	0	13	PB	0	5	PB
R	98	M	0	0	11	PB	0	5	PB
R	86	F	0	0	11	PB	0	13	MB
R	82	F	0	1	11	PB	3	13	MB
	Ave		0	0.9	9.8		0.3	5.9	
S	89	M	0	0	11	PB	0	3	None
S	77	M	5	0	13	PB	5	13	MB
S	80	M	2	2	13	MB	0	11	PB
S	81	M	0	1	13	MB	3	13	PB
S	90	M	0	0	6	PB	0	13	MB
S	96	M	0	2	13	PB	5	13	MB
S	87*	F	0	1	13	MB	0	13	MB
S	93	M	0	0	11	MB	5	13	MB
S	85	F	0	0	13	MB	5	13	MB
	Ave		0.8	0.7	11.8		2.9	11.7	

* Animal 87S died 24 hours after the Week 13 biopsy surgery, but full necropsy confirmed LSS 13 MB very severe JD

The Paralisa antibody patterns were very different for the R and S animals (Figs 1, 2 and 3). All 18 animals were shown to be infected at Week 13 and had clear signs of mild to severe lesions in the biopsy samples (Table 1). Subsequently, the S sire had 2 animals clinically affected and overall 7/9 had severe lesions. The R sire had one clinical case and overall only 2/9 had severe lesions. One of the R offspring appeared to completely "cure" itself between Weeks 13 and 49, and the remaining 6 R offspring only had very mild disease. This clearly demonstrates that resistance to paratuberculosis is strongly heritable. A paper on this study has recently been submitted to Veterinary Immunology and Immunopathology (see Appendix 3). Preliminary results have been presented at a number of forums and received very positive feedback.

The six Week 4 lymph nodes samples from 3 least affected and the 3 most severely affected were included with the full set of 18 samples (6 animals by 3 time points) from the 2008 study for the SOLiD SAGE sequencing in order to fill all the available sections on 3 plates (each plates was divided into 8 sections).

Expression study results to date

The SOLiD technology sequencing of 18 samples (3R and 3S at 3 time points) from 2008/9 samples and 6 samples (3R and 3S at 4 weeks pc) was successfully completed at Otago University under the supervision of Dr Jo Stanton. The initial sequence data was analyzed at AgResearch Ruakura by Dr Paul Maclean, under the supervision of Dr Nauman Maqbool, and the analysis of gene expression has been undertaken by Dr Rudiger Brauning, head bioinformatician at AgResearch Invermay.

i. Sequencing
   - A total of 566 million reads (24 samples).
   - 15-30 million reads were generated per sample.

ii. Quality control
   - 6.7 million reads (1.2%) were discarded because of ambiguities.
   - No bias for a specific sequence composition was detected.
   - The remaining 560 million reads represent between 9.9 and 19.4 million different sequences per sample point with copy numbers ranging from 1 to 167,389.
   - Interestingly between 93.9% and 96.2% of all different sequences per sample point are present only in a single copy.

iii. Converting raw reads into tags
   - 373 million tags with lengths between 26 and 28 bases were generated.

iv. Converting tags into expression counts per gene
   - Genes were mapped against three datasets; the deer and elk transcriptome (in-house), cattle transcriptome (Ensembl cattle genes) and the bovine genome (Btau 4). Of the 373 million tags, 31,500 genes were uniquely mapped with 18,000 genes also been annotated.

v. Analysis of gene networks
   - Of the above 18,000 genes, 17,500 were recognized by the Ingenuity Pathway Analysis (IPA) programme.
   - The expression level for the 3 samples at each time point were averaged and an expression value was calculated for each gene and this was expressed as a Log ratio, Fold change and p-value (significance of Fold change).
   - Using IPA genes were filtered by p-value ($p < 0.05$), organ, tissue type, immune response or inflammation, leaving 238 genes across all the 4 sample groups

- The expression of each gene has been compared between various treatments and time points. These results were put onto known mammalian pathways to elucidate the responses in resistant / susceptible deer.

vi. Preliminary results to date:

Relative Upregulation R between and S

The following Table is a summary of genes that were filtered out using IPA for each group in a comparison between R and S animals at each time point, including only datapoints where $p<0.05$. The first 3 groups are 2008 animals at Weeks 4, 12 and 50, and the fourth group (RS 0409) are the 2009 animals at Week 4. It is interesting to see that in all cases the S group had more genes relatively up-regulated in each of the ranges and in the top range the S group almost always had the higher degree of up-regulation (except RS 0309).

Group	S n	S Upregulation Fold Range	R n	R Upregulation Fold Range
RS 0408	17	8-33.5	10	8-22.8
	58	4-7.9	20	4-7.9
	60	3-3.9	5	3-3.9
	34	2.5-2.9	5	2.5-2.9
RS 0608	110	8-142	26	8-16.9
	142	4-7.9	75	4-7.9
	86	3-3.9	54	3-3.9
	81	2.5-2.9	41	2.5-2.9
RS 0309	18	8-25.9	14	8-32.9
	55	4-7.9	22	4-7.9
	52	3-3.9	27	3-3.9
	63	2.5-2.9	21	2.5-2.9
RS 0409	89	8-178.3	43	8-34.3
	67	4-7.9	73	4-7.9
	3	3-3.9	61	3-3.9
	0	2.5-2.9	53	2.5-2.9

Interesting candidates relatively upregulated in R or S animals at most time points

Symbol	Synonym(s)	Entrez Gene Name	Location	Type(s)
CCL5	D17S136E, MGC17164, MuRantes, RANTES, RNTES, SCYA5, Similar to merantes, SISd, TCP228	chemokine (C-C motif) ligand 5	Extracellular Space	cytokine
CCRL1	A630091E18Rik, CC-CKR-11, CCBP2, CCR-11, CCR10, CCX-CKR, CCX-CKR1, CKR-11, Cmkbrl1, LOC684092, PPR1, VSHK1	chemokine (C-C motif) receptor-like 1	Plasma Membrane	G-protein coupled receptor
FABP4	422/AP2, A-fab, A-FABP, Adipocyte fatty acid binding, ALBP, ALBP/AP2, AP2, Ap2 lipid-binding protein, Ap4, Lbpl	fatty acid binding protein 4, adipocyte	Cytoplasm	transporter
GGH	gamma-GH, GH, LOC667301, MGC105496	gamma-glutamyl hydrolase (conjugase, folylpolygammaglutamyl hydrolase)	Cytoplasm	peptidase
IFNG	IFG, IFI, IFN GAMMA, IFN TYPE II, IFNG2, INTERFERON GAMMA, TYPE II INTERFERON	interferon, gamma	Extracellular Space	cytokine
IL1A	IL-1, IL1-ALPHA, IL1F1, Interleukin-A	interleukin 1, alpha	Extracellular Space	cytokine
MYD88	MYD88D	myeloid differentiation primary response gene (88)	Plasma Membrane	other
NOS2	CALCIUM-INDEPENDENT NOS, HEP-NOS, Hepatocyte NOS, Inducible NOS, INOS, NOS, NOS-II, NOS2A	nitric oxide synthase 2, inducible	Cytoplasm	enzyme
PTGS2	COX-2, CYCLO-OXYGENASE 2, GRIPGHS, hCox-2, INDUCIBLE CYCLOOXYGENASE, MTCO2, PGG/HS, PGH synthase 2, PGHS-2, Pgi2 synthase, Pgs2, Pgsi, PHS-2, Prostaglandin endoperoxide synthase 2, TIS10	prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)	Cytoplasm	enzyme
TOLLIP	4930403G24Rik, 4931428G15Rik, FLJ33531, IL-1RAcPIP	toll interacting protein	Cytoplasm	other

Subsets of these genes (CCL5, FABP4, IFNG, IL1A, MYD88, NOS2, PTGS2) in various combinations, are involved in infiltration of granulocytes, cell movement of neutrophils, inflammatory response, activation of macrophages, cell movement of antigen presenting cells, immune response, development of helper T lymphocytes, TH1 immune response, apoptosis, priming and activation of T lymphocytes.

Interesting candidates in RS 0409 animals

One of the most interesting sets is the comparison of R and S 2009 animals at Week 4, which had a more extreme polarization than the 2008 animals. Although the 2008 animals showed some polarization along the RS spectrum, none of them died, whereas the three S animals in 2009 all went on to die of clinical Johne's 18-23 weeks pc, while the R animals were all very mildly affected or resolved completely. In this set the following disease/immune/inflammation

related genes were prominent (note in the RS comparison the positive value is up-regulation of the gene in R animals (shades of green), while a negative value is for up-regulation of the gene in S animals (shades of red):

Symbol	Log Ratio	Fold Change	p-value	Entrez Gene Name	Location
CAMP	4.242	18.916	4.01E-03	cathelicidin antimicrobial peptide	Cytoplasm
MYD88	1.84	3.581	1.07E-02	myeloid differentiation primary response gene (88)	Plasma Membrane
TNFSF15	2.04	4.113	3.45E-02	tumor necrosis factor superfamily, member 15	Extracellular Space
DAP	-2.416	-5.337	1.91E-02	death-associated protein	Cytoplasm
TOLLIP	-5.494	-45.07	3.16E-04	toll interacting protein	Cytoplasm

Paired RS upregulated genes

In the RS 0409 animals, pairs of genes from the same family showed relative upregulation with R and S animals. The functions of many of these genes or how it might relate to RS is not known as yet.

Symbol	Log Ratio	Fold Change	p-value	Entrez Gene Name	Location	Type(s)
CYP2B6	-3.63	-12.383	1.53E-03	cytochrome P450, family 2, subfamily B, polypeptide 6	Cytoplasm	enzyme
CYP4X1	2.133	4.386	3.45E-03	cytochrome P450, family 4, subfamily X, polypeptide 1	Cytoplasm	enzyme
EIF2C1	-3.763	-13.581	3.05E-03	eukaryotic translation initiation factor 2C, 1	Cytoplasm	translation regulator
EIF3B	2.067	4.191	3.54E-02	eukaryotic translation initiation factor 3, subunit B	Cytoplasm	translation regulator
FAM100B	2.49	5.616	1.53E-03	family with sequence similarity 100, member B	unknown	other
FAM184B	-3.456	-10.975	2.10E-03	family with sequence similarity 184, member B	unknown	other
MAP3K1	-2.785	-6.893	2.31E-02	mitogen-activated protein kinase kinase kinase 1	Cytoplasm	kinase
MAP3K11	5.05	33.138	1.87E-06	mitogen-activated protein kinase kinase kinase 11	Cytoplasm	kinase
SERPINA3	-3.587	-12.019	8.94E-03	serpin peptidase inhibitor, clade A, member 3	Extracellular Space	other
SERPINA5	2.603	6.075	4.83E-04	serpin peptidase inhibitor, clade A, member 5	Extracellular Space	other
SERPINB8	2.737	6.666	2.06E-03	serpin peptidase inhibitor, clade B, member 8	Cytoplasm	other
TCEA2	2.285	4.874	1.09E-02	transcription elongation factor A (SII), 2	Nucleus	transcription regulator
TCEB2	-3.552	-11.727	8.75E-04	transcription elongation factor B (SIII), 2	Nucleus	transcription regulator

Symbol	Log Ratio	Fold Change	p-value	Entrez Gene Name	Location	Type(s)
TMEM101	1.561	2.95	3.41E-02	transmembrane protein 101	Extracellular Space	other
TMEM141	3.394	10.509	3.74E-02	transmembrane protein 141	unknown	other
TMEM149	-5.755	-54.009	2.36E-05	transmembrane protein 149	unknown	other
TMEM164	1.626	3.088	2.52E-02	transmembrane protein 164	unknown	other
TMEM179B	1.512	2.852	3.53E-02	transmembrane protein 179B	unknown	other
TMEM182	3.048	8.271	1.10E-02	transmembrane protein 182	unknown	other
ZBTB38	-4.278	-19.398	6.74E-03	zinc finger and BTB domain containing 38	Nucleus	transcription regulator
ZCCHC24	-3.46	-11.002	4.30E-02	zinc finger, CCHC domain containing 24	unknown	other
ZNF653	3.03	8.167	3.74E-02	zinc finger protein 653	unknown	other
ZNF830	-3.138	-8.801	9.03E-03	zinc finger protein 830	Nucleus	other
ZSWIM3	2.057	4.162	4.56E-02	zinc finger, SWIM-type containing 3	unknown	other

Discussion

Studies on gene expression rarely give black and white results, but rather shades of grey. When analyzing gene expression levels for deer we are faced with combining these shades of grey at various levels.

1) Generation of sequence tags: The restriction enzymes used are not 100% accurate and the restriction is also not 100% efficient. This results in tags that differ in length and it also results in more than one tag per transcript.
2) Sequencing of tags: Although highly accurate when looked at as a whole, individual sequences can contain ambiguous bases as well as sequencing errors.
3) Mapping of tags: We want to analyze expression of genes. To do that tags have to be mapped uniquely to a gene. Even if we assume all tags are perfect (which they are not; see points 1. & 2.) there are more challenges. Genes can have alternative transcripts with different tags. Here we have to combine all the transcripts and thereby all the tags that belong to a specific gene, carefully trying to avoid combining transcripts of similar genes. Ideally we would have the complete transcriptome of deer available as a reference. In reality we have an incomplete transcriptome available that consists of deer and elk transcripts. To compensate for the incomplete deer & elk transcriptome dataset we use a reference set of cattle genes. This resource can be considered as nearly complete but it is cattle and not deer. To compensate for the incomplete set of cattle genes we use the cattle genome.
4) Annotating genes: Having mapped tags uniquely to a gene, we are interested in the functional annotation of that gene. Annotation is mostly derived from sequence comparisons with annotated genomes of popular species like human. The bovine genome consortium invested considerable effort in annotating the cattle genome. Despite this there remain genes that either have not enough similarity with other known annotated genes or they do have enough similarity but the function of the known gene has not been determined.
5) Examining networks: To look at the interaction of genes we use Ingenuity Pathway Analysis (IPA). IPA collects and curates network information from various databases and from the primary literature. It is human, mouse and rat centric. To analyze our deer genes with IPA

they get translated into their human/mouse/rat equivalent. This translation does not work for all genes. The disease examples are also human-centric.

Thus, although the SOLiD SAGE sequencing produced 566 million reads, this came down to 373 million tags, 31,500 uniquely mapped genes, of which 18,000 were annotated and 17,500 were recognized by IPA. Careful filtering brought this down to 200-300 genes up or down regulated across all the time points.

We have only had the complete set of data for a few weeks prior to the review submission deadline and therefore the results are very preliminary, given the complexity of the dataset and biological networks that we are dealing with.

The list of interesting candidates is exciting and will be added to and rationalized over the next two months, before the final Milestone reporting time at the end of June.
The strength of the approach in this study is that it can identify important genes without having any prior knowledge of their function. This is in marked contrast to the candidate gene approach. The most important genes in R / S may be ones for which their function is as yet unknown.

Conclusions
The gene expression of deer that are R or S to Johne's disease is entering a very informative phase. The results to date appear to be meaningful, but are incomplete. The results that are expected to be produced from the full 2009 set are likely to be even more informative in the final year of this programme. These future results should provide a list of genes that can be added to a qPCR panel using microtitre plates (currently under development by Otago University DRL) for *in vitro* testing of high value deer on deer studs and the benefits "trickle down" to commercial herds using selected stags. An additional year would allow testing of these deer candidate genes in sheep and cattle.

Recommendations for future direction (i.e. outline for further research)

To complete this project:

(a) Analyze data from the second sequencing.
(b) Validate results by choosing gene targets for further investigation. In collaboration with University of Otago DRL, design primers based on identified gene sequence and perform qPCR on lymph node and gut tissues collected in the 2008 and 2009 studies from known R and S phenotypes, as well as animals showing intermediate levels of susceptibility. Samples will also be collected from animals of known genotype on the original deer stud and on offspring of 4 of the stags (for which "Johne's Breeding Values" have been calculated) that are being used in an AI programme on Invermay this autumn. R and S phenotype animals from commercial farms may also be tested.
(c) Nominate candidate genes that can be included in a panel of qPCR tests for *in vitro* assessment of high value stud animals for R and S to Johne's disease.
(d) Complete publication of the 2008 and 2009 field studies.
(e) Publish findings from the first and second gene expression studies.
(f) Publish lay summaries of information on the genetics of resistance to Johne's disease for wider audiences.

Future studies:

a) In association with Otago University DRL, develop and test a panel of qPCR tests for in vitro assessment of high value stud deer for R and S to Johne's disease.
b) Apply a range of candidate genes to samples from sheep and cattle of known R and S phenotype to Johne's disease.
c) Develop and test a panel of qPCR tests for in vitro assessment of high value stud beef and dairy cattle and sheep for R and S to Johne's disease.

Appendix 1. Glossary of molecular genetic terms.
Genome: The entire set of genetic instructions found in a cell.

Gene: The gene is the basic physical unit of inheritance. There are around 35,000 genes in the cattle genome.

Transcript: To go from genes to proteins requires an intermediate molecule, the messenger RNA (mRNA). The mRNA produced for a particular gene is called a transcript, because it is transcribed gene information. Each gene consists of various transcribed bits (exons) separated by non-transcribed bits (introns). Exons can be reshuffled during transcription, resulting in alternative transcripts.

Transcriptome: The collection of all transcripts.
Tag: A short (27 bp) piece of a transcript, cut out from the end (3' end) of that transcript.
Read: The sequence generated by a sequencing machine is referred to as read. A read consists of a tag plus some generic adapter sequences which have been added to allow the generation and sequencing of tags.

Appendix 2.
See http://www.jdrc.co.nz/members_area/sciencereview.html for a copy of the draft paper: "Longitudinal pathogenesis study of young red deer after experimental challenge with *Mycobacterium avium* subsp. *paratuberculosis*" C.G. Mackintosh, R.G. Clark, B. Tolentino, S Liggett, G.W. de Lisle, J.F.T. Griffin

Appendix 3.
See http://www.jdrc.co.nz/members_area/sciencereview.html for a copy of the draft paper submitted to Veterinary Immunology and Immunopathology: "Immunological and pathological responses of red deer resistant or susceptible genotypes, to experimental challenge with *Mycobacterium avium* subsp. *paratuberculosis*" C.G. Mackintosh, R.G. Clark, B. Tolentino, G.W. de Lisle, S. Liggett, J.F.T. Griffin