final report

Project code: B.SGP.0500

Prepared by: Hutton Oddy and Frank Nicholas

University of New England and
University of Sydney

Date published: September 2015

PUBLISHED BY
Meat & Livestock Australia Limited
PO BOX 1961
NORTH SYDNEY NSW 2059

Sheep Genomics Program (SheepGenomics) Core Technologies Sub-program

Meat & Livestock Australia acknowledges the matching funds provided by the Australian Government to support the research and development detailed in this publication

Abstract

The Core Technologies Sub-program (CTS) of the Sheep Genomics Program (SGP) had three objectives:

1. Discovery and delivery of DNA markers from the Falkiner Memorial Field Station (FMFS) project and other mapping resources
2. Provide an integrated bioinformatics/biostatistics resource for use by the entire SGP
3. Develop a suite of leading-edge genomics research tools, funded wherever possible by external funds leveraged via active membership of the International Sheep Genome Consortium (ISGC)

All three of these objectives have been well and truly met.

The immediate practical outcomes of the CTS are DNA markers for breeding programs:

a) a single DNA marker for horns/polled. The SNP is located on chromosome OAR10 at position 30059832. If an individual is homozygous for allele 1 of this SNP there is a 99.8% chance the individual is homozygous for the polled mutation.
b) sets of DNA markers that enable the calculation of Australian Sheep Breeding Values (ASBVs) for animals (including new-born lambs) that have never been phenotyped, to make selection without phenotyping worthwhile right now for worm egg count, and in the near future for a range of other traits

Even though the SGP has now formally ended, the phenotypic/genotypic resources generated from the FMFS project will remain as an invaluable research resource for many years to come.

Of even more importance to the industry is the future legacy of a suite of genomics research tools. These include the infrastructure for ongoing provision of bioinformatics services to sheep and practical research tools which include the 50k ovine SNP chip, a draft sheep genome sequence, physical and comparative maps of the sheep genome and other species and beginnings of an understanding of the evolutionary biology of the sheep genome. This information makes it possible for the Sheep CRC to continue the application of high throughput genotyping for estimation of breeding values and will in the longer term make it possible to conduct gene discovery in sheep at a previously unimagined level.

Executive Summary

The mission of Core Technologies Sub-program (CTS) was to facilitate progress across the Sheep Genomics Program (SGP) by providing tools, resources and expertise for a comprehensive genomics research program.

All of CTS's activities were aimed at the three following objectives:
1. Discovery and delivery of DNA markers from the FMFS project and other mapping resources
2. An integrated bioinformatics/biostatistics resource for use by the entire SGP
3. A suite of high quality, leading-edge genomics research tools, funded wherever possible by external funds leveraged via active membership of the International Sheep Genome Consortium (ISGC)

All three of these objectives have been well and truly met.

The first of the three objectives (discovery and delivery of DNA markers) has been to an extent far beyond anything envisaged during the planning of the SGP. It is difficult for anyone not directly involved in the SGP to appreciate the extent of effort ─ way beyond the call of duty ─ that was involved in bringing about this achievement (especially because much of it took place in conditions of rather severe drought). The Herculean efforts of the large team of people involved in the creation, management, phenotyping and genotyping of the FMFS flock cannot be over-emphasised. We are not aware of any similar project conducted anywhere in the world on such a scale at the time the project was conceptualised and initiated. The model was taken up and further matured in the SheepCRC through the creation of the Information Nucleus Flocks.

Importantly, even though the SGP has now formally ended, the phenotypic/genotypic resources generated from the FMFS project will remain as an invaluable research resource for many years to come. Time and again in the future, sheep genetics researchers will return to this resource to identify further DNA markers for difficult-to-measure traits.

The success in achieving the first objective was heavily dependent upon CTS achieving its other two objectives, namely the provision of a bioinformatics/biostatistics team for the entire SGP, and the creation and provision of a suite of genomic tools, few of which existed at the commencement of the program. In other words, the discovery and delivery of a set of DNA markers for use by industry was achieved only because of very substantial behind-the-scenes efforts in creating and providing the essential genome research tools (culminating in the development of the 50K SNP chip) and in creating a team of talented researchers who created these tools and analysed the vast quantities of data that were generated. Much of the cost of the development of these tools was obtained by CTS personnel from outside funds, leveraged substantially by engagement with (and leadership of) the International Sheep Genomics Consortium (ISGC).

Objectives 2 and 3 above extended further than their role in achieving objective 1: they also played a key role in enabling the other SGP sub-programs to achieve their results (as described in the Final Reports of those other sub-programs). Thus, the bioinformatics team also provided an integrated bioinformatics web-base portal that provided a one-stop integrated "shop" for all the relevant bioinformatics tools and resources (some of which were created in-house), and support in the use of those resources; and the biostatisticians analysed large sets of gene-expression data collected by the other sub-programs, pointing to biochemical and physiological pathways that are relevant to key

biological processes involved in the production of wool and muscle (meat), and in resistance to internal parasites (all described in the Final Reports of the other sib-programs).

The immediate practical outcomes of the CTS are DNA markers for breeding programs:

c) a single DNA marker for horns/polled. The SNP is located on chromosome OAR10 at position 30059832. If an individual is homozygous for allele 1 of this SNP there is a 99.8% chance the individual is homozygous for the polled mutation.
d) sets of DNA markers that enable the calculation of Australian Sheep Breeding Values (ASBVs) for animals (including new-born lambs) that have never been phenotyped, to make selection without phenotyping worthwhile right now for worm egg count, and in the near future for a range of other traits

These results have also delivered a major result for the second SGP stream, namely gene function. By identifying quantitative trait loci (QTL) for each of the traits measured in the FMFS project, the genome scan with the 50K SNP chip has identified the regions of the sheep genome that contribute to genetic variation in each of the traits recorded in the FMFS project. For example, small genome regions on chromosomes 1, 3, 4 and 5 are associated with worm egg count in the FMFS flock and in the industry sires. Inheriting a favourable SNP allele in the region on chromosome 1 results in an improvement in PFEC of 23 ASBV units in the industry sires. Similarly, a cluster of SNPs on chromosome 4 is highly significant for bare breech width and depth. Also, a number of SNPs across 17 chromosomes are associated with total lean meat yield in the FMFS flock and with scanned eye muscle depth in the industry sires. When combined with the virtual sheep genome created within CTS, these results have provided, in effect, a list of genes worthy of research in relation to every trait that was recorded in the FMFS flock. Some of these genes (the most obvious ones) have already been the subject of intense research in the other SGP sub-programs. The remainder, which comprise an exceedingly rich resource, await research in future genomics projects. The accuracy of prediction of breeding values using genomic predictors generally remains rather low due to as yet inadequate sample numbers for most traits. The combined power of the FMFS and CRC data will enable the delivery of some of these new ASBVg; including for some novel traits to breeders. The pre-commercial evaluation of the technology is now underway.

The Sheep Genomics Program through the CTS contributed to an explosion of knowledge of the sheep genome at an international level. Through the activities of researchers in CTS and other SGP sub-programs the sheep has moved from a species where the community had scant knowledge of genome structure at the commencement of the program to where today a high quality reference genome sequence is about to be published. This would not have been possible if it were not for the leadership shown by the CTS in providing bioinformatic skills, vision as to what was possible with the rapidly changing technical landscape and negotiating skills to obtain funds form diverse sources within Australia and throughout the world. This was an extremely cost effectively way to provide new tools for the international community to conduct sheep research and to position Australia both at the forefront of that endeavour, and as immediate recipients of the best sheep genome tools in the world. This work is continuing through the International Sheep Genomics Consortium.

Contents

1 Context and structure of the CTS

1.1 Overview

1.1.1 Context in which the Sheep Genomics Program was developed

The following is an extract from the Executive Summary of the Final Report by Campbell and McGuirk (2002) that was used to scope the development of the Sheep Genomics Program (SGP). It provides the context in which the program was developed and goes on to describe the body of work that became the Core Technologies Sub-program (CTS).

"The proposed Sheep Genomics R&D initiative offers a unique chance to AWI and MLA to contribute significantly to improving industry profitability, by enabling them to address problems and opportunities with genomic tools that have not been available previously.

This new situation reflects the enormous developments in biotechnology generally and specifically in genomics, the study of the basic units of inheritance. These developments have occurred in many species, especially in humans, and parallel studies are now occurring in important species of domestic livestock such as cattle and pigs, and in a range of organisms that cause sheep diseases, including internal parasites. This foundation work has provided us with much-improved knowledge of inheritance, high throughput capabilities in genomics, especially in sequencing the DNA of an organism, and in related technologies such as proteomics…" Campbell and McGuirk (2002)

The authors of this report then went on to say:

"We suggest that there should be five "capability" projects, to improve underpinning technologies, namely:

- Ovine genome information collation and dissemination centre
- SNP development
- BAC library development
- Population design and data analysis
- Genetic evaluation methodology."

To a large extent, these suggestions became the body of work that was the CTS. The "Ovine genome information collation and dissemination centre" became the bioinformatics component of the sub-program. "SNP development" and "BAC library development", together with end–sequencing, were conducted as a joint venture between the CTS and the International Sheep Genomics Consortium (ISGC), using a small amount of funds from the SGP and significant contributions from a Commonwealth Government International Science Linkages (ISL) grant and from international collaborators. "Population design and data analysis" became a major component of CTS, comprising the detailed designing of a resource flock for identification of DNA markers for use in sheep selection programs, the creation and running of that resource flock in the form of the Falkiner Memorial Field Station (FMFS) flock, the recording of hundreds of phenotypes on each sheep in that flock, the genotyping of each sheep for approximately 50,000 DNA markers, the collation of the phenotypic and genotypic data into a database, and the analysis of all the data to identify DNA markers for use in sheep selection programs. Finally, the development of "Genetic evaluation methodology" became a key aspect of the analysis of the FMFS data.

The CTS was the epicentre of the SGP-wide gene mapping activity. This came about as a consequence of the design process for the SGP, details of which are briefly described here.

During 2002 a major workshop of all potential researchers in the field was held to put flesh on the bones of the Campbell and McGuirk report, i.e. to discuss advances in their fields, and what would be required to carry out "industrial strength" research in sheep genomics. This highlighted the following complementary strategies: 1) an extension of past work in quantitative genetics and QTL mapping, leading to gene discovery and industry utility via marker assisted selection, and 2) working out how identified genes affect phenotype to discover the mechanism of use and potentially develop alternatives to genetic improvement.

To inform subsequent discussion and decision-making, a number of short papers that described the state of the art in the field were commissioned. These included:

- Summaries of sheep genomic resources (past and present), including a catalogue of genetic resources (animals, DNA, phenotypes) available for research into meat, parasite resistance, wool and reproduction traits
- A summary of all QTL studies conducted to that time
- A statement of resources that would be required to carry out fine mapping
- A plan for bioinformatics support required to conduct gene mapping studies
- A description of the requirements and potential outcomes from a phenomics study (including mutagenesis). This subsequently became the modus operandi for the Australian Phenomics Centre located at ANU
- A list of genotyping tools available (up to that time)
- A description of the then available sheep EST libraries and how they might be used
- The physical and knowledge requirements of a sheep transcriptome project

This workshop, and subsequent scoping documents, set the shape of the SGP. It provided the rationale for breaking the objectives of the program into

a) discovery of useful DNA markers and
b) investigating gene function

as the two major streams of operation. From the outset, it was recognised that a CTS would be required to underpin these two major streams.

1.1.2 Overview of the Core Technologies Sub-Program (CTS)

At the outset of the SGP, there was a lack of high quality genomic resources publicly available for sheep research. Although a number of studies to find QTL affecting production and disease traits in sheep had been conducted in Australia and elsewhere (Crawford, 2002), they had largely been unsuccessful at finding useful QTL. The consensus view during development of the SGP was that past experiments were underpowered (in terms of animal numbers) to detect QTL of moderate effect (see later in this report). Moreover, it was obvious that unlike other livestock species (cattle and pigs), the genomic resources available to the sheep community were inadequate to enable fine mapping of regions of interest, when they were discovered.

The CTS was designed to support the product-focused sub-programs (meat, wool, parasite resistance), by developing and providing access to generic resources (animals, phenotypes and research tools including genomic information and the databases to catalogue and make them accessible to researchers). The CTS provided the vehicle in which the program-wide generation of

phenotyped animal resources was carried out – the FMFS project was a major activity undertaken under scientific guidance of the CTS.

The objective of providing the basic infrastructure for the SGP covered the following areas:

1. Design and implement necessary research activities to provide whole-of-program access to phenotype and genotypic data
   a) Design of a whole-of-program sheep QTL detection population
   b) Provide the resources to execute the FMFS phenotyping and genotyping project
   c) Overview and coordination of trait selection and data collection across all sheep resources funded through SGP
   d) Design, creation and implementation of a phenotype database
   e) Design, creation and implementation of a laboratory sample database
   f) Provision of a genotype database
   g) Sample management and storage, DNA extraction and co-ordination of genotyping

2. Provide oversight into the development of genomic resources to be used to create DNA markers (single nucleotide polymorphisms; SNPs) for genotyping the FMFS flock
   a) Assist development and end-sequencing of a sheep Bacterial Artificial Chromosome (BAC) library
   b) Provide intellectual input into development of a virtual sheep genome
   c) Co-fund together with others (ISL grant, USDA, BBSRC, Scottish Executive and Ovita–AgResearch and Meat and Wool NZ) the discovery of sufficient SNPs to make it practical to build a SNP genotyping platform
   d) Contribute to development of a draft sheep genome assembly

3. Provision of bioinformatic services for use in:
   a) Development of (in whole or in part) physical sheep genome resources (BAC Library, BAC-end sequence, native ovine genome sequence, SNPs and SNP chips – 1536 and 50k)
   b) Generation of (in whole or in part) knowledge resources (integrated sheep map, virtual sheep genome, sheep genome sequence assembly v.1, SNP chip design, sheep linkage maps, comparative maps between sheep and other species)
   c) Improvement of genomics tools (improved CRI-MAP software, sequence and SNP browsers, integrated gene expression and experiment databases, updated ovine and bovine sequence matched to protein libraries)
   d) Evaluation of gene expression technologies (MPSS, EST libraries, gene expression arrays)
   e) Co-ordination of the development and provision of bioinformatic services across SGP participants, principally through the SGP web portal

4. Biostatistical services and analysis
   a) Design of, and analysis of results from, the FMFS phenotyping and genotyping project, including inference of haplotypes to assist with allowing for otherwise unknown breed effects and to assign missing parents. This included:
      1. QTL mapping for traits measured at FMFS and elsewhere
      2. Estimation of marker effects for traits measured at FMFS
   b) Coordination of genotyping, and joint/complementary analyses of data, with Sheep CRC and Sheep Genetics. This included:

1. The use of marker effects estimated from FMFS data to calculate genomic EBVs for industry sires

c) Analysis of microarray data
d) Generation and maintenance of web-based analysis of phenotype/genotype data from other sheep resources funded through SGP

5. Creation and maintenance of web-based SGP resources and services, including:
   a) Building and maintaining the SGP website www.sheepgenomics.com
   b) Developing and maintaining a program-wide bioinformatics portal www.sgpbioinformatics.agresearch.co.nz
   c) Building and maintaining the SGP phenotype database (FMFS Phenotype Database http://pheno.sg.angis.org.au/)
   d) Populating a genotype database with data generated from the 50k ovine chip
   e) Assisting with development and maintenance of the website of the ISGC www.sheephapmap.org
   f) Updating the major sheep linkage map and making it freely available at http://rubens.its.unimelb.edu.au/~jillm/jill.htm
   g) Upgrading the sheep section of Online Mendelian Inheritance in Animals (OMIA) http://omia.angis.org.au
   h) Including all available sheep mapping information in Oxford grids and other comparative mapping frameworks http://oxgrid.angis.org.au
   i) Expanding and enhancing the Australian Sheep Gene Mapping Website http://rubens.its.unimelb.edu.au/~jillm/jill.htm
   j) Developing the sheep QTL map website (SheepQTLdb) http://sphinx.vet.unimelb.edu.au/QTLdb/

A simplified diagram (Figure 1) showing how these functions fit together is presented below.

Figure 1. A simplified diagram of the activities of the Core Technologies Sub-program (CTS)

This infrastructure necessary to conduct genomics R & D for sheep was developed within a space of two years, and was utilized by SGP researchers for the remaining life of SGP, at a total cost to MLA/AWI of $10,155,000 over the 5 years of SGP.

In the case of the FMFS project, additional costs were borne by each subprogram to complete the phenotyping. The total cost to MLA and AWI of the FMFS project was >$9m (see Table 1, below, for a simplified breakdown of costs/activities – Note this does not include any costs incurred by research organisations in generation and analysis of the FMFS data).

Core Technologies Sub-Program

Data Collection and Access	Bioinformatics – and genomic resource development	BioStatistics
FMFS funded and managed within core-tech	Bio-informatics database	Analysis of gene expression microarray data
FMFS activities – general flock	Bioinformatics support for development of sheep genome resources	Analysis of MPSS data
Procurement, management and live trait measurement	Virtual sheep genome	Analysis of genomic data in
FMFS Technical Officer	Selection of re-sequencing targets	FMFS progeny
FMFS traits collected by specific Sub-Programs	Assembly of 454 and Illumina RR sequence for SNP discovery	Method development and testing
CSIRO RoxMo flock, the missing FMFS Family	Selection of optimal SNP for chip development	Haplotypes, sire identification and dam lines
Phenotype database	Assembly and reporting of 1st sheep genome sequence	Generation of QTLS/trait
Genotype database	BAC library	The horn/Poll locus
Sample management, DNA extraction and arraying	End sequencing BAC library	GEBVs
Genotyping of FMFS animals and industry sires	Development and refinement of sheep linkage map	FMFS Prototypes and industry validation
The sheepGENOMICS website	Positioning misaligned SNPs using the FMFS flock	Preparation for joint analysis with Sheep CRC
Support for IGSC and ASGMWS websites	Developed improved mapping software – upgraded CRI-MAP	
	Comparative maps and sheep consensus map	
	Support for proteomics analysis Access to, and evaluation of, gene expression tools	

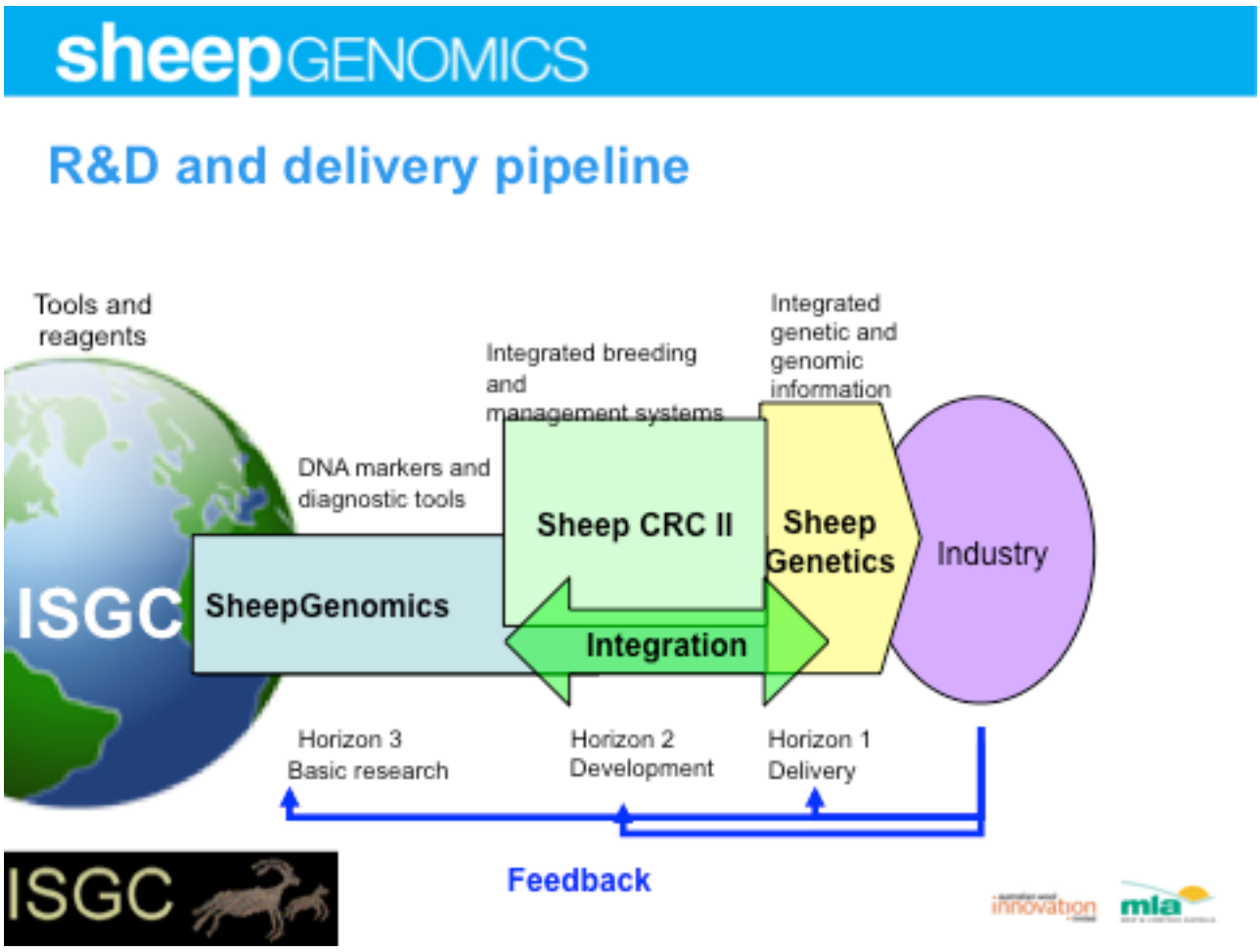
Table1. Summary of costs incurred to generate FMFS animals, phenotypes and subsequent genotypes and analysis thereof ($1000's)

Activity	Core Tech	Muscle	Host Resistance to Parasites	Wool	Reproduction
Generation of Animals and standard phenotyping	$1,450				
Sub-Program Specific phenotyping		$1,504	$2,000 (estimate)	$1,000 (estimate)	$568
Data collation and database management	$250 (estimate; does not include Mohammad's time)				
DNA Extraction and Genotyping	$1,200				
Data Analysis	$1,090				

In terms of generic sheep genomic resources (generation of sheep sequence and discovery of new markers, mainly SNPs), the investments of SGP were highly leveraged through the ISGC. The ISGC co-ordinated international investment of approx AUD$6.3million, of which less than AUD$1m came from MLA/AWI sources, and another AUD$857,250 from the Australian Government via an ISL grant, to generate the sequence necessary to build a commercial 50k ovine SNP chip that has subsequently been used to genotype ~ 4000 FMFS animals, > 600 sires from the Australian sheep industry, >5000 Sheep CRC animals, >3000 sheep in the International Haplotype Mapping Project and many thousands of other sheep around the world.

The level of international co-investment made possible through the influence of SGP has hastened the development of these resources, and made them available at substantially less cost to the Australian sheep industry than would otherwise have been the case.

Figure 2 illustrates the relationship between the SGP (and Sheep Genetics and the Sheep CRC) and the ISGC. It highlights the role of the SGP (principally the CTS) in contributing to the development of tools and reagents for the conduct of sheep genomics research (see Oddy et al., 2007).



2 Objectives of the CTS

2.1 Building the capacity to conduct "Industrial Strength" genomics research on sheep

The mission of CTS was to facilitate progress across the SGP by providing tools, resources and expertise for a comprehensive genomics research program.

All of CTS's activities were aimed at the three following objectives:
1. Discovery and delivery of DNA markers from the FMFS project and other mapping resources
2. Provide an integrated bioinformatics/biostatistics resource for use by the entire SGP
3. Provide a suite of validated, leading-edge genomics research tools, funded wherever possible by external funds leveraged via active membership of the International Sheep Genome Consortium (ISGC)

Given the state-of-the-art of genomics applied to sheep at the commencement of the CTS, it was clear that a major effort was required to build the resources necessary to facilitate an integrated

sheep genomics R&D program. The efforts of CTS to fulfil its mission can best be described under the following headings.

2.1.1 People

At the outset of the SGP, the number of people with skills in the field of genomics applied to sheep was extremely low. It was reasoned that by making a program such as SGP sufficiently attractive, we could both attract skilled people from other disciplines and species to work in sheep, and that these people would attract and train new scientists to the area.

2.1.2 Animal Resources

Up to the development of the SGP, analysis of experiments designed to find QTL for production traits in sheep indicated little success (Crawford, 2002; Robinson and Goddard, 2002; Goddard et al., 2002). Following a review by Goddard et al. (2002), it was decided to build a specific QTL-mapping resource that had sufficient power to detect QTL that accounted for >0.3 phenotypic standard deviations in a population designed to allow both linkage analysis (LA) and linkage disequilibrium (LD) (association) analysis for as many traits as possible. This became the FMFS project, which, together with the CSIRO flock generated by crossing an intestinal-parasite-resistant Romney ram over susceptible Merino ewes, became the major resource for gene discovery in the SGP.

2.1.3 Support for program-wide phenotype recording and sample handling

The scope of the FMFS project required a specific database to store and recall phenotypic information and a sample database to track samples for subsequent DNA extraction and genotyping. A specific DNA genotype database was also required. After extensive evaluation of existing databases, it was decided to create a phenotype database from scratch, and to adapt existing databases for samples (AgResearch) and DNA genotype (DPIVic).

2.1.4 Genomics Tools and Reagents

In 2002 state-of-the-art tools for genotyping and QTL discovery were microsatellite markers placed on linkage maps, with discovery of causative polymorphisms facilitated by sequencing the appropriate clones recovered from BAC libraries. By 2005 the ground had shifted to SNP-based genotyping and physical location of SNPs and other markers on physical maps informed by genome sequence. This rapid change in the technological (sequencing and mapping) and information (physical cf linkage maps) base was the driving force behind CTS's support for development of new genomic reagents. Much of this work was undertaken in collaboration with scientists in the USA, NZ and elsewhere. In addition to genotyping tools, tools for measurement of gene expression evolved rapidly during this time. The CTS provided access to these tools (ranging from small, generally unannotated, long cDNA-based chips, generated from knowledge of Expressed Sequence Tags, to Affymetrix arrays of small molecules designed with mRNA sequence from cattle). These gene expression array tools required testing for intended application and subsequent analysis.

2.1.5 Bioinformatics – bringing genomics tools together

At the commencement of the SGP there was very little in the way of sheep-based genomic information readily accessible to scientists in the program. The task in this element of CTS was to gather sheep-specific genome information and place it in readily accessible databases, with database tools that made searching for genome sequence, gene ontology, protein structure and

function easily able to be conducted by SGP scientists. In addition to facilitating access to extant information, the bioinformatics component of CTS played a vital role in generating the tools to build a virtual sheep genome, evaluated methods for SNP discovery, provided the design work for both the prototype 1536 SNP chip and the final ovine SNP50 beadchip, and worked with the sheep research community to refine the sheep genome sequence.

2.1.6 BioStatistics – analysis of gene expression and genotyping data

It was clear during the development of the SGP that the skills requirement to analyse gene expression data were common across sub-programs. The other major task involved analysis of genotype/phenotype data from the FMFS project and other Australian sheep genotyped with the 50K SNP chip. CTS appointed geneticists/statisticians process the enormous quantity of data generated in both these areas, and to conduct the analyses.

3 Results and Discussion

3.1 Building the capacity to conduct "Industrial Strength" genomics research on sheep

3.1.1 People: The Human Face of the Sheep Genomics Program Core Technologies Sub-Program

The following is a list of people who contributed to the CTS. Many were on the CTS payroll for at least part of the life of the SGP.

Site	Supervisors	Staff	Skills area
University of Sydney	Frank Nicholas Herman Raadsma Nick Sangster	Jonathan Usmar, Oscar Luo, Mohammad Shariflou	Database development and web delivery (JU), the integrated sheep map (OL), sample management and storage, data quality control and arrangement/coordination of SNP genotyping (MS), comparative mapping (OL, JU)
University of New England	Julius van der Werf	Sang Hong Lee, Cedric Gondro, John Hickey, Karen Marshall	QTL analysis, analysis of gene expression data, development of analytical methods, including detection of haplotypes
	Hutton Oddy	Jason Siddell	Field data collection; initial database entry and quality control
University of Adelaide	Cindy Bottema	Greg Nattrass	Gene expression analysis
University of Melbourne	Jillian Maddox	Gavin Baker	Linkage mapping, development of the Falkiner Phenotype Database, development of new tools
	Jason White		SheepGENOMICS web portal: www.sheepGENOMICS.com
CSIRO Livestock Industries	Brian Dalrymple	Sean McWilliam, David Townley, Abhirami Radnakumar, Wes Barris,	

	John Henshall Nick Andronicus Ian Purvis James Kijas	Antonio Reverter, Margaret O'Grady, Haja Kadarmideen, Peter Hunt, Caroline Lee Sonja Dominik, Amy Bell	
AgResearch	Alan McCulloch	Nauman Maqbool, Jason Mitchell	
DPIVic	Mike Goddard Ben Hayes	Jenny Pryce, Iona McLeod	
AWI	Troy Fischer Graham Cam	Nigel Strutt, John Murray, George Nichols Alistair Donaldson John Owens	
MLA	Terry Longhurst Alex Ball		
UWA	Jason White	Melissa Berg Guy Ben-Ary	sheepGENOMICS.com

The skills in bioinformatics and analysis of genomics data applied to sheep are unique and invaluable assets for the sheep R&D community (and through this, the practical sheep industry). The CTS significantly increased the number of people able to work in this field and the overall knowledge base for the species. At least 20 new people were exposed to, and contributed their expertise to, expand sheep-specific information. Some of these will contribute to the sheep industry in the future.

3.1.2 Animal Resources and Supporting Databases

At the 2002 workshop that gave rise to the SGP, the consensus was that then extant animal/DNA and phenotype resources were largely underpowered to discover QTL and find genes that were causative for particular traits. Moreover, it was considered that the trait-specific nature of existing sheep resources (i.e. find QTL and then genes within trait groups) was a weakness if the data were to be used on an industry-wide basis. Accordingly, one recommendation from the workshop was to explore the design of a comprehensive QTL-mapping flock comprising families of sufficiently large size to have the power to detect QTL with effect greater than approximately 0.3 phenotypic standard deviations, and that measured the full range of industry relevant phenotypes in all progeny.

The ideas above were further developed by Mike Goddard, John Henshall and Julius van der Werf in a review commissioned to assess the strength of evidence in then extant QTL studies and explore the design criteria for a study with sufficient power to overcome past weaknesses.

The key elements of their recommendations were to:-

1. Establish a resource of animals with large half-sib families on which measurements of meat, wool, parasite resistance and reproductive traits are made on each (relevant) progeny

2. Use this resource to find genes/polymorphisms causing variation in the target traits or markers very closely linked to these genes
3. Given the marker-genotyping limitations at that time, concentrate initially on a limited number of QTL already identified in other studies, leaving open the possibility of conducting a genome scan if the necessary resources were to become available

The steps to implement this research plan were:

a) Select six chromosomes that are most likely to have relevant QTL (based on past projects)
b) Verify these QTL and re-estimate their effects
c) Determine if these QTL segregate in commercial sheep
d) Map the QTL to <5cM
e) Identify the QTL and mutation causing the phenotypic effect
f) Confirm the QTL and estimate effects on other traits
g) Use four sires heterozygous for these QTL to generate 400 progeny/sire
h) Use 15 industry sires to generate 200 progeny/sire
i) Use ewes from diverse backgrounds
j) Record a standard set of phenotypes on all progeny, and genotype them all for 30 markers on six chromosomes
k) Map the QTL to within 10 cM by fine-mapping heterozygous offspring
l) Reduce confidence interval for a QTL to <5 cM by LA-LD mapping
m) Select candidate genes in the QTL regions and search for polymorphisms associated with the trait in animals with known QTL genotype
n) Use identified SNPs across a wide range of offspring to confirm they are in LD with the QTL
o) Refine the confidence interval and find additional useful SNPs
p) Genotype additional animals for the putative causative QTL and estimate effect on all traits

This design was implemented as the SGP QTL-mapping flock at the Falkiner Memorial Field Station (FMFS), near Deniliquin NSW, together with one of the QTL flocks at CSIRO Chiswick (near Armidale NSW).

From mid-2004 until late-2009, FMFS was home to the largest sheep QTL-mapping flock ever assembled.

The FMFS was chosen as the site for this project primarily because, as a result of a recent substantial investment in upgrading its flood-irrigation system by its owner AWI, the station was "guaranteed" to be drought-proof ─ an important consideration when one is contemplating the collection of phenotypes on a scale never before contemplated anywhere. Unfortunately, no sooner was the decision made to conduct the project on the newly-upgraded FMFS, than the water crisis in the Murray-Darling system resulted in a zero allocation of irrigation water to the research station! With 2004 turning out to be a drought year, the project was faced with enormous challenges that necessitated the purchase of expensive fodder. Despite these challenges, the project was implemented: in 2004 and 2005, 9522 ewes were artificially inseminated with semen from 20 sires to generate more than 4700 lambs of which at least 4632 were weaned. A full description of the data collected in this study is attached in appendix 1 (attach draft FMFS phenotype paper).

Briefly, a wide range of phenotypes (including the basic traits captured in Sheep Genetics) were measured on all progeny. In addition to the basic Sheep Genetics traits, new traits included:

i) Body composition and retail yield (estimated by Dual emission X-Ray Absorption – DXA)
j) Meat quality – pH, tenderness, intramuscular fat, meat colour
k) Muscle fibre number size and oxidative state

l) Temperament
m) Worm egg counts after challenge infection with *H. contortus* and *T. colubriformus*
n) Blood cell count and PCV
o) Wool and skin traits
p) Breech score, polledness, body conformation

In the >2000 ewe progeny, many reproductive traits were measured following syndicate joining, including:

q) number of lambs in utero, number lambed, number weaned, ewe behaviour around the birth site, and derivative traits.

In total, the FMFS resource flock was phenotyped for 306 traits (obviously, not all traits were able to be recorded on all sheep, e.g. carcase traits were recorded only on males, and reproductive traits only on females). This enormous effort required approximately 540,000 animal handlings over 2.5 years, and generated over 927,000 phenotypic records from 4,655 progeny, all of which were deposited in the tailor-made phenotype database. Around 350 kg of faeces were collected for worm egg counts, and 400 litres of blood were collected for DNA extraction. By any standards, this was a huge logistical operation!

While most of the phenotyping was completed by 2007, the recording of reproductive traits on the ewe progeny necessitated retaining the ewes for longer. In late 2009, with AWI having decided to decommission FMFS and lease the property, the ewe-progeny flock was relocated to the Sheep CRC for further study of ewe traits, including reproduction, methane production and further study of the biology of a myostatin mutation that has been the subject of much research within the Muscle sub-program.

By the time the blood samples had been collected for extraction of DNA, the restrictions that formed the basis of the Goddard-Henshall-van der Werf recommendations had been overtaken by the development of the 50K SNP chip (developed, as explained elsewhere in this report, with enormous input from CTS). Consequently, it was decided to proceed straight to conducting a genome scan of as many FMFS progeny as could be afforded within the CTS genotyping budget. In the end, after another enormous effort by CTS personnel, and after the application of various quality-control procedures, genotypes for each of 48,640 SNPs were available for each of 3,860 FMFS progeny on which phenotypic data had been collected; a total of 187,750,400 genotypes were determined and deposited in the genotype database, ready for analysis! This unbelievably large dataset was then combined with the more than 927,000 records in the FMFS phenotype database in a monumental analysis conducted by a joint CTS/Sheep-CRC biostatistics team. The results of this enormous analysis have been reported in detail in the Final Report of the SGP biostatistics support team (projects SGP 558, 559, 560), by Ben Hayes, John Henshall, Sonja Dominik, Julius van der Werf, and are summarised in section 3.1.4 below.

3.1.3 Genomic Tools and Reagents

The tools and reagents created specifically for the SGP (and which could never have been created without the CTS) and which are now available for sheep R&D generally are listed below.

- A high-coverage sheep BAC Library was constructed in the laboratory of Pieter de Jong at the Childrens Hospital Research Institute (CHORI) in Oakland California, USA. This library consists of >210,000 clones of average length ~120 kilobases (kb) which provides a coverage of more than 11x the sheep genome. The CHORI-243 ovine BAC Library was

constructed from DNA from an inbred Texel ram (obtained from USDA Meat Animal Research Laboratory). See http://bacpac.chori.org/library.php?id=162

- The CHORI-243 BAC library was end-sequenced (using Sanger sequencing, with an average read length of 550 bases) at The Institute for Genome Research (TIGR) in Maryland, USA. The end-sequence has been deposited in Genbank, at http://www.livestockgenomics.csiro.au/perl/gbrowse.cgi/sheepbacend/

- The end-sequence data from the CHORI-243 library was assembled into a virtual sheep genome using human, cattle, horse and dog genomes as reference (Dalrymple *et al.*, 2007). The latest (second) version of this resource is viewable at: http://www.livestockgenomics.csiro.au/perl/gbrowse.cgi/vsheep2/

- The sheep linkage map was revised no fewer than four times from 2001-2009 to incorporate the positions of new markers as they became available: http://rubens.its.unimelb.edu.au/~jillm/jill.htm

- An integrated sheep map that combined all available mapping information from linkage and physical maps was constructed and was then used to create a virtual sheep map, which informed marker positions on the virtual sheep genome. The integrated sheep map is viewable at http://compldb.angis.org.au

- Comparative maps of sheep and other species were constructed and are available as Oxford grids at http://oxgrid.angis.org.au/

- An enhanced sheep section of Online Mendelian Inheritance in Animals (OMIA), providing up-to-date references for all sheep loci with a known phenotypic effect, plus additional hyperlinks to other relevant databases in the USA and UK, and to the integrated map described above, as well as to the virtual sheep genome. This is viewable at http://omia,angis.org.au

- A SNP (single nucleotide polymorphism) discovery project was carried out that had three major phases:

    - In the first phase, resequencing of DNA from a diverse selection of sheep breeds (as individual sheep and as pools) was carried out using either BAC-end-sequence or expressed-sequence-tag (EST) sequence as the starting point. From this, more than 6000 SNP were discovered, of which 1536 were developed into a SNP chip and were used to conduct a pilot haplotype mapping project and were positioned on the International Mapping Flock (linkage map) and on the USDA Radiation Hybrid Panel. Information from this phase of the SNP discovery project is available at http://www.sheephapmap.org/pilot.php. See also Kijas *et al.*, 2009 for a description of the key results of the pilot haplotype mapping project.
    - In the second phase, six sheep of different breeds (Poll Dorset, Merino, Awassi, Scottish Blackface, Romney and Texel) were each sequenced to a depth of 0.5x (approx 1.5Gb of sequence/sheep). This information was assembled against bovine sequence for SNP discovery and positions were assigned using the virtual sheep genome. Overall more than 500,000 new SNPs were discovered in this process. Sequencing was conducted at the University of Otago (New Zealand), using

AgResearch/Meat and Wool New Zealand funds, and at the Human Genome Sequencing Center, Baylor College of Medicine, Texas, USA, using funds from the Australian ISL grant and SGP. More information about this phase of the project is available at http://www.sheephapmap.org/genseq.php

  - In the third phase of the SNP discovery project, reduced-representation sequencing of pooled DNA from 60 sheep was carried out using Illumina/Solexa sequencing technology. This resulted in discovery of a further 76,000 high-quality SNPs (see link above). Over all phases of the SNP discovery project, more than 600,000 SNPs were discovered. These became the starting material for development of an ovine SNP chip comprising more than 50,000 of the best SNPs.

- The ovine SNP50 beadchip was jointly designed by scientists in the SGP, the ISGC and by Illumina Inc.
  (See http://www.illumina.com/products/ovine_snp50_whole_genome_genotyping_kits.ilmn) Since its release in January 2009, more than 20,000 chips have been sold. Included among the sheep genotyped with the chip are >4000 FMFS progeny, >500 industry sires and >5000 sheep-CRC Information Nucleus animals. In addition, the chip has been the basis of the public domain sheep haplotype mapping project in which 2890 sheep from 64 breeds were genotyped; see http://www.sheephapmap.org/hapmap.php

- To position the SNPs accurately, a draft sheep genome sequence has been constructed http://www.livestockgenomics.csiro.au/sheep/oar1.0.php. This resource and the virtual sheep genome are key inputs into current activities of ISGC participants to generate a sheep reference genome.

These outputs from the SGP and from the ISGC will remain a lasting legacy of the CTS. They provide the resources for linkage and association (LD) analysis, for development of additional genotyping capacity, for analysis of re-sequencing data and gene discovery in sheep generally. The processes used for development of the virtual sheep genome were instrumental in discovering errors in the assembly of the bovine genome and contributed greatly to improvements in the bovine genome sequence. The contribution of the team of scientists assembled by the SGP is continuing through the Sheep CRC and the CRC for Beef Genetic Technologies in Australia and through a range of international projects conducted within the International Sheep Genomics Consortium.

Issues that did NOT eventuate included undesirable consequences of leakage of intellectual property in this space. In fact, by making the basic resources largely available in the public domain, others freely contributed and built the resource base to a level far beyond that initially envisaged. Details of many of these outcomes which have been deposited in the public domain can be obtained on the ISGC website http://www.sheephapmap.org and on the website of the National Center for Biotechnology Information (NCBI) http://www.ncbi.nlm.nih.gov/projects/genome/guide/sheep/

3.1.4 Bioinformatics tools and resources

To develop the genomics resources (above) is a pre-requisite to success, but did not ensure their availability to scientists throughout the SGP and their subsequent widespread application. Making genomics resources available to the research community in a readily accessible and appropriate format is essential for successful implementation. The bioinformatics team in the CTS had this responsibility.

Originally conceived as a service component of the CTS, bioinformatics became both a service provider and a major contributor to the discovery process. The discovery component of genomics tool development was an integral part of the outputs described in 3.1.3 above.

The delivery component was more specific to the SGP. It included:

- Responsibility for the SGP website http://www.sheepgenomics.org which became the repository of all the operational aspects of the SGP and the only place in which all reports from the projects in the program were held in an easily accessible form. The SGP website was constructed by the University of Western Australia and administered under the supervision of Dr Jason White (University of Melbourne).
- Development of the bioinformatics portal at https://sgpbioinformatics.agresearch.co.nz/, which provided program-specific bioinformatic resources to the SGP (developed by AgResearch). These included a sheep genomics database for sharing all types of data files generated within SGP, stored in such a way as to be searchable in an integrated manner with respect to, e.g., gene, animal, sample, sequence, ontology, protocols; together with integrated access to:
    - virtual sheep genome browser
    - sheep BAC-end browser
    - bovine genome browser
    - human genome browser
    - OASIS (the CSIRO database of in-silico sheep SNPs)
    - international sheep linkage map
    - Online Mendelian Inheritance in Animals (OMIA)
    - Oxford Grids
    - integrated sheep map
    - sheep gene index
    - database of candidate genes
    - software for comparison and analysis of DNA sequence
    - other external genomics resources
- Access to gene-expression arrays, including testing of the arrays to ensure suitability for the proposed purpose. Although the SGP sub-programs used a number of gene expression arrays and even Massively Parallel Signature Sequencing (MPSS), the array used by most programs was the Affymetrix bovine array, which worked quite satisfactorily in sheep. The wool sub-program used a cDNA-based array available through AgResearch.
- Assistance with annotation of gene-expression arrays and with analysis of gene-expression data. Scientists were appointed at CSIRO Livestock Industries, University of New England and University of Adelaide to help those scientists working on gene-expression analyses to interrogate and analyse their data.
- Development of a pipeline to automate routine analysis tasks from raw microarray data to lists of differentially expressed genes
- Assistance with annotation of protein data obtained using proteomics techniques. At the outset of these studies, there were little sheep-specific protein sequence data available and the bioinformatics team assisted the scientists involved in proteomics with both protein and gene annotation of output from their studies
- Quality control for the SNP genotyping data, conducted by Jill Maddox, Mohammad Shariflou and Ben Hayes. An automated process was created by Jill Maddox and Cedric Gondro to speed up the data cleaning/selection. This is a very tedious (but absolutely essential) task that must be conducted prior to performing any analyses on SNP genotype data

- Adapting the CRI-MAP and MultiMap linkage mapping programs so that they can handle larger data sets with more loci and more animals (Jill Maddox)
- Continuing development of the Australian Sheep Gene Mapping web site (ASGMWS) by Jill Maddox. The ASGMWS is an Australian resource that is primarily devoted to supporting the interests of the sheep gene mapping community. It includes a sheep marker database, comparative map viewing tools (CMap), a sheep QTL database (SheepQTLdb), and information about other genomics resources that are relevant to sheep trait mappers: http://rubens.its.unimelb.edu.au/~jillm/jill.htm
- Maintaining the ISGC web site from August 2006 to February 2009 (Jill Maddox)
- Clustering and annotation of sheep ESTs (CSIRO, in collaboration with AgResearch)
- Annotation of sheep genes (CSIRO)
- Predicting SNPs from sheep ESTs - in collaboration with AgResearch (CSIRO)
- Analysis of sheep BAC-end sequences (CSIRO)
- Rationalisation of sheep markers (CSIRO and Jill Maddox)
- Design and data analysis of sheep BAC-end and EST resequencing project (CSIRO)
- Creation of refined bovine genome assembly Btau3.6x as scaffold for building virtual sheep genome v 2 (CSIRO)
- Creation of version 1 of the sheep genome assembly on the scaffold of the virtual sheep genome 2 (CSIRO, in collaboration with AgResearch)
- Including a range of quality control analyses of the sheep genome assembly (CSIRO)
- Development of associated browsing tools for providing access to SNP and genome assembly and annotation information (CSIRO)
- SNP calling from Illumina sequencing of sheep-reduced representational datasets (CSIRO)
- Design of the Illumina Ovine SNP50k BeadChip (CSIRO, in collaboration with AgResearch)

3.1.5 BioStatistics – analysis of genotyping data

The results from the enormous phenotyping/genotyping/analysis effort are by far the most important immediate-impact output from the entire SGP (and from CTS). The results from extensive analyses of the FMFS phenotype/genotype data, together with similar data from industry sires, are detailed in the MLA final report for projects SGP 560, 559, 558 (entitled “Genomic Breeding Values and Genome Wide Association in the Falkiner Flock”, by Ben Hayes, John Henshall, Sonja Dominik, and Julius Van Der Werf). To provide an indication of the essence of the results, the abstract from that final report is reproduced below:

DNA or genomic information has potential to increase the accuracy of selection and therefore the rate of genetic gain in the sheep meat and wool industries. In this report we demonstrate how genomic information can be translated into predicted breeding values for key traits in sheep meat and wool production. Some traits are determined by a mutation at a single gene - the most striking example of this is horns/polled, and here we report a single DNA marker which predicts genotype for horns or polled with 99.8 accuracy (though this marker should be validated in another set of animals). However for most traits we investigated, the effect of individual markers were small - which we demonstrated by using the Falkiner Memorial Field Station flock (FMFS) to predict the effect of each of 48640 markers on a range of traits. Therefore we tried an approach called genomic selection. This approach used all 48640 markers simultaneously to predict breeding values. The accuracy of the genomic breeding values (GEBVs) was assessed in Merino, maternal and terminal industry sires. The accuracies of GEBV for Merinos were moderate for most traits and high for a few of them. The accuracy of GEBV for worm egg count traits was particularly encouraging given the low heritability of this trait. However for terminals and maternals, accuracies of breeding value were low,

reflecting the fact that FMFS is largely based on Merino genetics. Genotyping a large group of industry sires is recommended to improve the accuracy of GEBV, and therefore potential rates of genetic gain, in these breed types.

The following paragraph from the Executive Summary of that report is also of interest:

The impact of this technology will depend on the price with which it can be delivered on farm. The current price of the Ovine50SNP chip is sufficiently high that its use in the industry to predict breeding values is likely to be limited to a few elite flocks. However the cost of the technology is dropping. If sufficiently low prices are achieved, use of the technology could become widespread across the stud sector with the gains predicted by Van der Werf (2009) realised.

As also explained in that report, the predictions from van der Werf (2009 are that:

if the accuracy of genomic breeding values for breeding objective traits was as high as the square root of the heritability of the trait, genomic selection could increase the overall response for a terminal sire index by about 30%, and a fine wool merino index by about 40%. For some traits we are close to achieving this. For instance for FEC the approximate accuracy was 0.43. The heritability of FEC is approximately 0.2, so it's square root is 0.44. For other traits, improvements in the accuracy of GEBV are necessary. For example for EMD [eye muscle depth], the accuracy of GEBV, even adjusted for the accuracy of ASBV, is still below the square root of the heritability of EMD. However once the FMFS and Sheep CRC reference populations are merged [which has now happened], we will be closer to achieving these accuracies.

In essence, therefore, the first of the two streams of the SGP has delivered (via CTS):

e) a single DNA marker for horns/polled. The SNP is located OAR10 at position 30059832. If an individual is homozygous for allele 1 of this SNP there is a 99.8% chance the individual is homozygous for the polled mutation.
f) sets of DNA markers that enable the calculation of ASBVs for animals (including new-born lambs) that have never been phenotyped, to make selection without phenotyping worthwhile right now for worm egg count, and in the near future for a range of other traits

These results have also delivered a major result for the second SGP stream, namely gene function. By identifying QTL for each of the traits measured in the FMFS flock, the genome scan with the 50K SNP chip has identified the regions of the sheep genome that contribute to genetic variation in each of the traits recorded in the FMFS flock. For example, small genome regions on chromosomes 1, 3, 4 and 5 are associated with worm egg count in FMFS and in the industry sires. Inheriting a favourable SNP allele in the region on chromosome 1 results in an improvement in PFEC of 23 ASBV units in the industry sires. Similarly, a cluster of SNPs on chromosome 4 is highly significant for bare breech width and depth. Also, a number of SNPs across 17 chromosomes are associated with total lean meat yield in FMFS and with scanned eye muscle depth in the industry sires. When combined with the virtual sheep genome created within CTS, these results have provided, in effect, a list of genes worthy of research in relation to every trait that was recorded in the FMFS flock. Some of these genes (the most obvious ones) have already been the subject of intense research in the other SGP sub-programs. The remainder, which comprise an exceedingly rich resource, await research in future genomics projects.

3.1.6 Biostatistics – analysis of gene expression data

The second leg of the Biostatistics effort was the analysis of gene expression data generated by the muscle, wool and host resistance to parasites subprograms

Within each subprogram, extensive studies of gene expression were conducted which required expert analysis and interpretation. The CTS enabled and/or assisted the following scientists to assist the subprograms with statistical and bioinformatics analysis of gene expression and other experimental data:
Toni Reverter (CSIRO Livestock Industries, Brisbane)
Cedric Gondro (UNE – Armidale)
Greg Nattrass (SARDI – University of Adelaide)
Among other things, this team developed a pipeline to automate routine analysis tasks from raw microarray data to lists of differentially expressed genes. The pipeline analysis generates:
• quality control checks
• expression summaries
• differentially expressed genes using an empirical Bayes method
• conventional log-fold change in association with a moderated t-test

Among datasets that have been analysed are:
• Fleece-rot microarray experiments
• Gastro-intestinal nematodes candidate gene expression
• Gastro-intestinal nematodes microarray experiments
• Experiments comparing sheep with different ASBVs for muscle traits

This led to very fruitful collaboration and a number of publications on topics ranging from associations between gene expression and phenotype in response to challenge by intestinal parasites, to gene network analysis of effects of genetic control of muscling on gene expression in muscle (see papers published). More importantly, the work resulted in development of tools that enabled gene expression microarray experiments to be appropriately analysed, e.g. strategies were developed for pathway analyses to identify over-represented pathways among differentially expressed genes; this being the first step along a path towards systems analysis.

3.2 The future for sheep genomics resources

The work described here underpins the future of genetics and genomics research in sheep and the delivery of SNP-based tools to the sheep breeding and production industries. At the moment the link between the development of the resources and users of the resources is being stretched very thinly by the gap in ongoing funding to support the continued development, or even just the maintenance, of the resources. Much of the ongoing work is funded by international collaborators in the ISGC including USDA, Chinese Academy of Sciences, BGI (Beijing Genomics Institute, Shenzhen), BBSRC and the European Union Framework 7 funding arrangements, and residual funds from the Australian Government International Science Linkage (ISL) Grant. If no further Australian industry development of the resources occurs, the benefits of close links for the application of what has been delivered so far will be lost.

Of course, the current generation of tools is not the end of the road for genomics-based genetic improvement of sheep. The balance of the ISL grant is sufficient to fund a substantial additional sequencing project on the six animals already sequenced to 0.5 X each by 454 technology. This will provide more SNPs and a better sheep genome assembly less reliant on the bovine genome

assembly. The availability of more SNPs will enable researchers to find large numbers of SNPs in any area of interest from the public resources, reducing the costs to them of more sequencing of target regions of the sheep genome. At this stage, a major new SNP-chip design initiative is not planned until a sheep reference genome is complete. Generaton of a reference genome will by itself lead to the discovery and placement of more than 2 million new SNPs, and will come without a specific investment in SNP discovery, which logically should follow the assembly of a sheep reference genome.

In addition, Noelle Cockett is coordinating a USDA grant to complete the sheep genome sequence to reference genome standard. This will underpin the long-term application of comparative genomics of sheep and cattle to truly understand the genetic basis of sheep production traits.

Most recently (from March 2010), the ISGC has welcomed scientists from the Kunming Institute of Zoology (KIZ) and the Beijing Genomics Institute, Shenzhen (BGI) into the consortium. BGI and KIZ had until then independently sequenced a sheep genome (female Texel) and were seeking assistance with access to ISGC resources to complete their assembly. Now, the Chinese researchers are collaborating with scientists from the Roslin Institute (Edinburgh, Scotland), CSIRO Livestock Industries and AgResearch to complete a single high quality sheep genome sequence assembly that will be published in the second half of 2010. This will greatly hasten the development of a reference quality sheep genome sequence.

4 Success in Achieving Objectives

4.1 Success in Achieving Objectives

As summarised in section 3.1.4 above, the first of the three objectives (discovery and delivery of DNA markers) has been well and truly met. Indeed, the DNA-marker results summarised above (and described in detail in the Final Report of the biostatistics team) go far beyond anything envisaged during the planning of the SGP. It is difficult for anyone not directly involved in the SGP to appreciate the extent of effort – way beyond the call of duty – that was involved in bringing about this achievement (especially because much of it took place in conditions of rather severe drought). The Herculean efforts of the large team of people involved in the creation, management, phenotyping and genotyping of the FMFS flock cannot be over-emphasised. We are not aware of any similar project conducted anywhere in the world on such a scale.

Importantly, even though the SGP has now formally ended, the phenotypic/genotypic resources generated from the FMFS project will remain as an invaluable research resource for many years to come. Time and again in the future, sheep genetics researchers will return to this resource to identify further DNA markers for difficult-to-measure traits.

The success in achieving the first objective was heavily dependent upon CTS achieving its other two objectives, namely the provision of a bioinformatics/biostatistics team for the entire SGP, and the creation and provision of a suite of genomic tools, few of which existed at the commencement of the program. In other words, the discovery and delivery of a set of DNA markers for use by industry was achieved only because of a huge behind-the-scenes effort in creating and providing the essential genome research tools (culminating in the development of the 50K SNP chip) and in creating a team of talented researchers who created these tools and analysed the vast quantities of data that were generated. This is a major legacy of the SGP CTS team, which will enable future work in the field to more rapidly deliver new knowledge of what genes do what in sheep and, but underpin delivery of new ASBVs and gene markers to industry.

In summary, it is evident from the paragraphs above that all three objectives of the CTS have been achieved. That is not to say, however, that the entire CTS operation was not free of problems. The following section attempts to summarise important issues that arose during the life of the CTS, and lessons that can (hopefully) be applied by anyone contemplating a similar venture in the future.

4.2 Problems encountered and lessons learned

As explained in detail in the final report for project SG543, the development of the phenotype database was not helped by the resignation of its designer, Gavin Baker, very soon after the bare bones of the database had been created. Financial constraints at the time of Gavin's resignation resulted in responsibility for the database being handed to Jonathan Usmar, a bioinformatician who had been employed for other tasks and who, as it turned out, was less on the wavelength of the end users than was desirable, and who was severely hampered by illness. Accordingly, the phenotype database required by the researchers conducting analyses on the FMFS data never eventuated in a useful way, and today bears the scars of this failure to interact with the users. This was not the case with the genotype database, which was built by DPIVic explicitly using experience obtained from developing a similar database for dairy cattle.

More generally, the unwieldy management structures imposed by (or within) the CTS component of SGP actually impeded the production of the tools required by the end users.

If such a program were to be repeated, it is strongly recommend that the researchers be provided with more autonomy to generate the tools and resources they require. To work through a bureaucratic process such as imposed by the entire SGP management structure impeded rapid and responsive decision-making and implementation of appropriate cost-effective solutions.

The following are paraphrased comments from Brian Dalrymple (CSIRO), and leaders of Muscle, Wool and Host Resistance to Intestinal Parasites sub-programs who are well placed to comment on the interactions between the CTS and the target-specific sub-programs.

1. The Core Tech program started after the functional programs. This meant that the opportunity to build an interactive resource–function relationship was unable to be fully developed. Most functional programs were well on the way to their goals with the expectation that Core Tech could provide the resources. This expectation could not be met, because the work was not jointly developed. In many ways Core Tech was playing catch up, and the functional programs were often unaware of the opportunities. This was mainly because of delays in getting the Core Tech program in place.
2. In some cases, personalities determined the researchers served by the Core Tech program. For example, at CSIRO Livestock Industries, Ross Tellam and Tony Vuoccolo were designated to work with Cedric Gondro at UNE, instead of Tony Reverter at CSIRO LI. This was because of past experience and preconceived notions of difficulty in relationships. It ignored that fact that at UNE, Cedric was a first year Post-Doc, learning the field, and that travel between Armidale and Brisbane for face to face meetings was difficult. It also did not fully capitalise on Brian Dalrymple's group's (of which Tony Reverter was a part) interest in functional genomics of muscle.
3. When pressed to engage with the complexity of the data generated by (for example) the Muscle Sub-Program (data on gene expression, organelle specific proteomics, cell biology and sub cellular phenotypes – localization of genes and gene products where affected by major genes) the Bioinformatics program saw this integration work as outside their brief. Accordingly the anticipated help from the Bioinformatics team was not always forthcoming. To be fair they did provide help with annotation of protein sequence with gene sequence for protein identification/gene annotation where

available, but this was a low level job compared with the integrative task that was required to resolve the depth of data collected within the Muscle sub-program.
4. The Wool sub-program did not get the service that it required. It got AgResearch (NZ) to do some work in NZ and CSIRO in Australia. In short, scientists in the Wool sub-program worked with those they already had relationships with, rather than teach and learn new things and build new relationships with the Core Tech sub-program.
5. Host resistance to intestinal parasites sub-program had less problems accessing resources for, say, gene expression analysis, because they were working in the CSIRO LI labs and used the local (i.e. Brisbane) connections. In short, the relationships did not need to be rewritten between individuals within the program (they were simply continuations).

5 Impact on Meat and Livestock Industry

5.1 Impact Now

The immediate outputs from CTS were listed in section 3.1.4 and are repeated here for convenience:
a) a single DNA marker for horns/polled. The SNP is located OAR10 at position 30059832. If an individual is homozygous for allele 1 of this SNP there is a 99.8% chance the individual is homozygous for the polled mutation.
b) sets of DNA markers that enable the calculation of ASBVs for animals (including new-born lambs) that have never been phenotyped, to make selection without phenotyping worthwhile right now for worm egg count, and in the near future for a range of other traits
c) a 50k ovine SNP chip with positions located on a draft sheep genome will continue to facilitate mapping of traits to positions on the genome, and to underpin continued genotyping in current studies such as those undertaken within the Sheep CRC
d) functional expertise in both genome structure (to build for example genotyping resources) and in statistical analysis to conduct gene discovery and prediction of genomic breeding values.
e) a database of >4000 animals with deep phenotypes and genotypes which has in the short term been combined with information from the Sheep CRC Information Nucleus to provide an immediately useful resource for both further research and, as noted, above estimation of ASBVs.

5.2 Impact in five years time

This topic was also canvassed in section 3.1.4. A slightly expanded version is presented here:

Van der werf (2009) suggested that if the accuracy of genomic breeding values for breeding objective traits was as high as the square root of the heritability, genomic selection could increase the overall response for a terminal sire index by about 30%, and a fine wool merino index by about 40%. Banks et al. (2009) suggested the economic impact of achieving this could be substantial. For some traits we are close to achieving this. For instance for FEC the r(GEBV,ASBV) was 0.28. If we divide this by the accuracy of FEC ASBV in Merinos, which was 0.65, we get an estimate of r(GEBV,TBV), where TBV is true breeding value, of 0.43. The heritability of PFEC is approximately 0.2, so it's square root is 0.44. For other traits, improvements in the accuracy of GEBV are necessary. For example for EMD, the accuracy of GEBV, even adjusted for the accuracy of ASBV, is still below the square root of the heritability of EMD. However once the FMFS and Sheep CRC reference populations are merged, we will be closer to achieving these accuracies.

It is also important to note that there is extra information that can be used to improve the accuracy of genomic breeding values, in the form of parent average or pedigree information. In the current data set we could not assess how much this would add to the accuracy of breeding value, as this information was not available for almost all of the Merino industry sires.

The SNP we have discovered which predicted poll genotype offspring from Merino sires could be used to select for polledness following a relatively small validation step.

The long term impact of this technology will depend on the price with which it can be delivered on farm. The current price of the Ovine50SNP chip is sufficiently high that its use in the industry to predict breeding values is likely to be limited to a few elite flocks. However the cost of the technology is dropping rapidly - for example in cattle the bovine50K SNP Chip will drop in price from AU $250 to AU $100 next year when the bovine 500K SNP chip arrives. If such a price point could be reached for the Ovine50SNP chip, its use could become widespread across the stud sector with the gains predicted by Van der Werf (2009) realised.

The contribution of the CTS to the development of the tools that paved the way for a reference sheep genome cannot be underestimated. This has already facilitated more rapid assembly of the sequence of a Texel female by the team at Kunming Institute of Zoology and Beijing Genomics Institute and the ISGC work at the Roslin research Institute and elsewhere.

5.3 Impact in the longer term

As explained in section 1.1.1 of this report, the SGP always had two overall aims, namely:

a) discovery of useful DNA markers and
b) investigating gene function

It was always recognised that the first of these aims would be the first to deliver practical benefits to the sheep and wool industries. And, as described above, this has already been realised, with much more to come in the next five years.

Just as it was realised that DNA markers would be the first fruits of the SGP, so too, was it realised that the benefits of the second aim (investigating gene function) would take longer to bear fruit. The results achieved from the genomics research conducted within the other SGP sub-programs are described in the Final Reports of those sub-programs.

It is useful to compare the SGP with the Human Genome Project, in which the vast bulk of activity fits into the second aim above. The driving force behind the Human Genome Project was the realisation that genomics research enabled the asking of key questions about life that could never before be asked. For the first time, for example, we could ask: which genes are switched on when a mammal (human or sheep) is infected with internal parasites? Even more importantly, we could, for the first time, ask: what is the difference in gene expression between mammals (humans or sheep) that are resistant and susceptible to an internal parasite? As a result of the Human Genome Project and the SGP, we now have initial answers to questions such as these for humans and for sheep. In other words, we now have a far, far better idea of what biochemical and physiological processes are involved when internal parasites infect humans or sheep, and we also have some idea of the differences in those biochemical and physiological processes between resistant and susceptible individuals. However, translating that knowledge into practical products of direct use to sheep producers is not a simple business. History teaches us that the fruits of such basic biological research will be plentiful in the fullness of time, in ways that we cannot today envisage.

The strength of the SGP was that it was structured so as to deliver something of immediate benefit (DNA markers) while laying the absolutely essential foundations and conducting initial genomics research that will provide key outputs for sheep production in the longer term.

6 Conclusions and Recommendations

The first three sub-sections below are paraphrased from the Final Report of the biostatistics team.

6.1 Genomic breeding values and Genomic Selection

We have demonstrated how genomic information can be translated into genomic breeding values for the meat sheep and wool industries. The accuracy of these genomic breeding values varies widely according to breed type and trait. In the FMFS project, there are insufficient terminal or maternal chromosome segments to allow useful predictions for terminal or maternal breed types in industry. However, as the FMFS project is largely based on Merino genetics, the accuracies of GEBV for Merinos were moderate for most traits and high for a few of them. The accuracy of GEBV for worm egg count traits was particularly encouraging, given the low heritability of this trait.

6.2 Genome-wide association studies

For most traits of interest in meat and wool sheep, the effect of individual SNP are small, so the preferred approach to using the SNP information is to use large panels of markers to predict breeding values. This ensures as much genetic variation as possible is captured. However the obvious exception is polledness – the SNP we found associated with this trait explains 99.8% of the trait variation. Another important exception is a mutation in regulatory region of the myostatin (GDF8) gene that reduces active myostatin protein in tissues resulting in increased muscling and reduced fatness.

While genome-wide association studies generally do not discover SNPs associated with sufficient genetic variance to make a large contribution to improving accuracy of genomic breeding values, they do uncover some of the key genes and pathways involved in the relevant trait. This information may be very useful, for example for anthelminthic development in the case of intestinal worms.

6.3 Recommendations relating to DNA markers

We recognise that some of the following are already happening but are included for completeness.

1. The SNP associated with polledness should be validated and potentially commercialised if the validation is successful

2. The bare area SNPs be validated in another population

3. The SNPs associated with total lean meat yield be validated in conjunction with the CRC

4. The Terminal and maternal reference population needs to be expanded to improve the accuracy of GEBV in these breedtypes. The most cost effective way to do this may be to genotype a large cohort of industry sires and use these as the reference population.

5. The FMFS and Sheep CRC reference populations should be merged to improve the accuracy of GEBV

6. A denser SNP chip would improve predictions in the Merino breed and allow information to be used in across breed predictions

7. Sheep Genetics Australia could consider GBLUP as the method to calculate genomic breeding values, given its simplicity. This method also has the advantage that it can predict the relationship derived component of breeding value if pedigree is unavailable, provided relationship is not corrected for in the calculation of GEBV.

6.4 Recommendations relating to bioinformatic issues

1. A key recommendation is that there continue to be funding for bioinformatics resources for sheep genomics so as to expedite development and utilisation of possible genomics tools and resources by the Australian sheep industry.

2. That support be provided for an Australian sheep genomics web site and that useful material from the current ASGMWS site be housed on such a site. This includes the microsatellite marker database, sheep CMap application, and either SheepQTLdb or an alternative sheep trait and QTL database.

3. That consideration be given to the concept of developing a linkage map based on the FMFS 50k SNP genotyping data. Such a map would provide vital sheep-specific information on the ordering of loci. It is further recommended that if the decision is made to produce a high density linkage map, then the possibility of including genotyping data produced by other organisations should be investigated, and that additional investment be made in linkage mapping software improvements and hardware so as to increase the speed, capacity and reliability of map construction.

4. That further investigations and simulations be made as to the numbers of SNPs needed for parentage panels for different population types, including FMFS, the AWI merino parentage panel and other flocks, before decisions are made as to the size of the SNP parentage panel.

5. That MLA and AWI, on behalf of the sheep industries, develop their own set of minimum requirements for quality of genotyping data from service providers.

6.5 Broader issues

There is no doubt that the SGP was instrumental in catalysing the development of the wide range of genomics tools for sheep that have become available over the last five years. By facilitating the ISGC and, in particular, linkages with AgResearch, the ISL grant and especially with Noelle Cockett at Utah State University and with Genesis Faraday, the SGP has changed the direction of sheep genetics in Australia. However, looking to the future, it is hard to see how this role will be continued. Will MLA, AWI, and/or the Sheep CRC fill the gap? Will one of the universities, or the DPIs or CSIRO fill the gap? Perhaps it is the ISGC that will now continue to lead the sheep genomics revolution.

Whoever does take the lead, it is clear that MLA and AWI need to be active participants. A world class and world leading team (for sheep) has been put together, but genomics is such a broadly

applicable and exciting field that any substantial pause could see this team dismantled as quickly as it was put together.

6.5.1 Significant learnings

Apart from the delivery of DNA markers to industry, the most successful activities were those with clear and challenging goals, e.g. those led by the bioinformaticians, namely the genome assemblies and development of the SNP chip. The many small interactive tasks aligned with the biology were piecemeal and not as successful in most cases. Also the underpinning sciences structure created a bit of a barrier. In part this was due to the bioinformatics not being an integral part of the original planning process; it was too much seen as a support service, rather than a driver of what could be achieved with smart integration of the biological questions and the bioinformatics. The onerous requirements for establishing contractual arrangements significantly impeded the desired cross discipline activity.

6.5.2 General recommendations

That MLA and AWI remain engaged in the generation and use of genomics tools for sheep breeding for the long haul. The time will come when they prove their worth, but there will be a lot of work to be done in the meantime, including moving well beyond the current approaches. Disengaging now runs the risks that Australia's leading role in the development will decline, and researchers will move on, possibly never to return to sheep research and hence never having a long term impact on the competitive advantage of the industry.

The MLA and AWI should pick winners and support them through the ups and downs.

6.5.3 What we could have done differently

The value of the integrated database was never realised, as it was not funded sufficiently to really deliver, and the SGP was not sufficiently big and/or focussed to justify the investment that would have been required. In hindsight, probably its development should have been less ambitious: it would have been better to have aimed only at a data repository and to not have tried to become a web-enabled interactive database integrating datasets together.

Overall, the bioinformatics area could have benefitted from much tighter management / coordination of its activities. We were successful because we self-managed. Hopefully we picked the right targets.

6.5.4 Future pieces of the puzzle

Much exciting and cutting-edge work was undertaken in the SGP and much will continue outside of the sphere of influence of MLA and AWI. This has laid the foundation for the ultimate successful application of genomics to sheep breeding, perhaps long after the linkage with the SGP is forgotten. A particular focus will be integration of genetic and non-genetic data to predict phenotype, utilising knowledge-intensive approaches in combination with numerical analyses.

7 Bibliography

Publications arising

Full (refereed) papers and book chapters:

Bidwell C.A., Cockett N.E., Maddox J.F. & Beever J.E. (2008) Genome Mapping in Sheep. In: Genome Mapping and Genomics in Domestic Species. Editors: N. E. Cockett and C. Kole. Springer Publishers, Germany pp. 33-46.

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

Dodds K.G., McEwan J.C. & Davis G.H. (2007) Integration of molecular and quantitative information in sheep and goat industry breeding programmes. *Small Ruminant Research* **70**: 32-41.

Goldammer T., Brunner R.M., Rebl A., Wu C.H., Nomura K., Hadfield T., Maddox J.F. & Cockett N.E. (2009) Cytogenetic anchoring of radiation hybrid and virtual maps of sheep chromosome X and comparison of X chromosomes in sheep, cattle, and human. Chromosome Research 17, 497-506.

Goldammer T., Di Meo G.P., Lühken G., Drögemüller C., Wu C.H., Kijas J., Dalrymple B.P., Nicholas F.W., Maddox J.F., Iannuzzi L. & Cockett N.E. (2009) Molecular Cytogenetics and Gene Mapping in Sheep (Ovis aries, 2n = 54). Cytogenetic and Genome Research 126, 63-76.

Kijas J.W., Townley D., Dalrymple B.P., Heaton M.P., Maddox J.F., McGrath A., Wilson P., Ingersoll R.G., McCulloch R., McWilliam S., Tang D., McEwan J., Cockett N., Oddy V.H., Nicholas F.W., Raadsma H.; for the International Sheep Genomics Consortium. (2009) A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. PLoS ONE 4(3): e4668.

Maddox J.F. & Cockett N.E. (2007) An update on sheep and goat linkage maps and other genomic resources. *Small Ruminant Research* **70**: 4-20.

Maddox J.F., Tammen I. & Dominik S. (2007) Using PCR and linkage mapping to identify single genes and QTL for livestock traits. in PCR: Methods Express, edited by S. Hughes & A. Moody, pp 81-106, Scion Publishing Ltd, Bloxham, UK.

Marshall K., van der Werf J.H.J., Maddox J.F., Graser H-U., Zhang, Y., Walkden-Brown S.W. & Kahn L. (2005) A genome scan for quantitative trait loci for resistance to the gastrointestinal parasite Haemonchus contortus in sheep. Proc. Assoc. Advmt. Anim. Breed. Genet. 16 115-118.

Marshall K., Maddox J.F., Lee S.H., Zhang Y., Kahn L., Graser H.U., Gondro C., Walkden-Brown S.W. & van der Werf J.H. (2009) Genetic mapping of quantitative trait loci for resistance to Haemonchus contortus in sheep. Animal Genetics 40, 262-272.

McEwan J.C. (2007) Developing genomic resources for whole genome selection. *New Zealand Society of Animal Production* **67**: 148-152.

Meadows J.R.S., Hanotte O., Drögemüller C., Calvo J., Godfrey R., Coltman D., Maddox J.F., Marzanov N., Kantanen J. & Kijas J.W. (2006) Globally dispersed Y chromosomal haplotypes in wild and domestic sheep. Animal Genetics 37, 444-453.

Oddy V.H., Longhurst T.J., Nicholas F.W., Maddox J.F. & McDonagh M.B. (2005) Sheep genomics – core technologies and resources. Proc. Assoc. Advmt. Anim. Breed. Genet. 16 209-212.

Oddy H., Dalrymple, B., McEwan, J., Kijas, J, van der Werf, J., Emery, D, Hynd, P., Longhurst, T. Fischer, T., Ferguson, D., Forage, R., Cockett, N. and Nicholas, F.. (2007) SheepGenomics and the International Sheep Genomics Consortium. Proc. Assoc. Advmt. Anim. Breed. Genet. 17, 411-417.

Rowe, A., C. Gondro, D. Emery and N. Sangster (2008). "Genomic analyses of Haemonchus contortus infection in sheep: abomasal fistulation and two Haemonchus strains do not substantially confound host gene expression in microarrays." Veterinary Parasitology 154 (1-2): 71-81.

Rowe, T., C. Gondro, D. Emery and N. Sangster (2009). "Sequential microarray to identify timing of molecular responses to Haemonchus contortus infection in sheep". Veterinary Parasitology 161(1-2): 76-87.

Tetens J., Goldammer T., Maddox J.F., Cockett N.E., Leeb T. & Drögemüller C. (2007) A radiation hybrid map of sheep chromosome 23 based on ovine BAC end sequences. Animal Genetics 38, 132-140.

Wu C.H., Nomura K., Goldammer T., Hadfield T., Dalrymple B.P., McWilliam S., Maddox J.F., Womack J.E. & Cockett N. E. (2008) A high-resolution comparative radiation hybrid map of ovine chromosomal regions that are homologous to human chromosome 6 (HSA6). Animal Genetics 39, 459-467.

Wu C.H., Jin W., Nomura K., Goldammer T., Hadfield T., Dalrymple B.P., McWilliam S., Maddox J.F. & Cockett N. E. (2009) A radiation hybrid comparative map of ovine chromosome 1 aligned to the virtual sheep genome. Animal Genetics 40, 435-455.

Working Papers:

Sequencing the Sheep Genome Working paper prepared by: Hutton Oddy, John McEwan, Brian Dalrymple, Jill Maddox, Frank Nicholas, Chris Warkup and Alex Clop. 23 November 2006

Papers close to submission at time of report (if not submitted already):

Beh, K.J., N.M. Kingsford, C. Gondro, L.J. Chappel and H.W. Raadsma (2010). "Extremes of genetic resistance to gastro intestinal nematodes show limited effects of acquired immunity to adult worms of Haemonchus contortus during primary infestations". Veterinary Parasitology (in preparation).

Beh, K.J., L.J. Chappel, N.M. Kingsford, C. Gondro and H.W. Raadsma (2010). "Extremes in genetic resistance to Haemonchus contortus fail to show differential larval establishment on primary infestation in naïve lambs". Veterinary Parasitology (in preparation).

Byrne, K, T. Vuocolo, C. Gondro, J. White, N.E. Cockett, T. Hadfield, C.A. Bidwell, J.N Waddell and R. Tellam (2010). "A gene network switch enhances the oxidative capacity of skeletal muscle during late fetal development". BMC Genomics (accepted).

Conference papers:

Dalrymple B.P. et al. (2008) PAG-XVI (W95) Toward A Sheep Genome Assembly.

Dalrymple B.P. et al. (2009) PAG-XVII (W083) Identification Of Sheep SNPs Using Illumina Sequencing And Design Of The Sheep 60K Illumina Iselect SNP BeadChip. - PowerPoint Talk

Gondro, C. (2009). "Summarization methods and quality problems in Affymetrix microarrays". Proceedings of the 18th Conference of the Australian Association for the Advancement of Animal Breeding and Genetics. 18: 304-307.

Gondro, C. and J. H. J. van der Werf (2008). "On the importance of QC measures in microarray analyses". Proceedings of the Canadian Society of Animal Science 2008.

Kadarmideen, H.N., N. Watson-Haigh, J.W. Kijas, T. Vuocolo, K. Byrne, C. Gondro, H. Oddy, G.E. Gardner and R.L. Tellam (2010). "Genetics of global gene expression patterns and gene networks affecting muscling in sheep". Proceedings of the 9th World Congress on Genetics Applied to Livestock Production.

Kijas J., TownleyD. , Dalrymple B., McWilliam S., McEwan J.C., Wilson P., McGrath A., Nicholas F. & the ISGC 2007 The International Sheep Genomics Consortium and SNP Discovery in Sheep (poster) SABRE / EADGENE joint consortium meeting in Utrecht 4-8 June 2007.

Kijas J. et al. (2008) PAG-XVI (W92) Preliminary Survey Of SNP Variation Within The Sheep Genome.

Kijas J.W. et al. (2009) PAG-XVII (W082) Determining The Genetic Structure Of Sheep And The ISGC Hapmap Experiment.

Maddox J.F. (2005) A presentation of the differences between the sheep and goat genetic maps. Genet. Sel. Evol. 37 (Suppl. 1), S1-S10.

Maddox J.F., Dalrymple B.P., MCWilliam S.M., MCCulloch A., Kirkness E.F., Townley D.J., Barris W.C., Kijas J.K., Ratnakumar A., MCEwan J.C. & Nicholas F. (2006) Enhanced mapping tools for the sheep genome - BACs, microsatellites, SNPs and the virtual map. Poster E507, International Society of Animal Genetics (ISAG), 20-25 August 2006, Porto Seguro, Brazil

Maddox J.F. et al. (2008) ISAG - Sheep Genome Project - PowerPoint Talk.

Maddox, Jillian F., Allan M. Crawford, Roxann Ingersoll, James W.Kijas, Brian P. Dalrymple, Ian R.W. Evans, Frank W. Nicholas, David. J. Townley, John McEwan, Tracey van Stijn, Herman W. Raadsma, and the International Sheep Genomics Consortium (2008) Using a 1.5K ovine SNP array to expand the sheep linkage map ISAG Poster

McEwan J.C. et al. (2008) PAG-XVI (P543) Using Roche 454 Flx Sequencing For SNP Discovery In Sheep.

McEwan J.C. et al. (2008) PAG-XVI (W93) 454 Sequencing And SNP Discovery In Sheep: Project Design And Interim Results.

McEwan J.C. et al. (2009) PAG-XVII (W081) Skim Sequencing The Sheep Genome, And SNP Discovery, Using Roche 454 FLX Sequencing.- PowerPoint Talk

Oddy H. et al. (2008) PAG-XVI (W96) A Reference Genome For Sheep.

Payne G.M. et al. (2008) PAG-XVI (P544) A Comparison Of 454 Vs Traditional Sequencing Approaches.

Rowe, T., C. Gondro, D. Emery and N. Sangster (2007). “Imune inhibition of establishment and development of Haemonchus contortus in sheep.” Proceedings of the Australasian Society for Immunology 2007.

Papers and reports Cited:

8 Appendices

8.1 Appendix 1

Draft paper of the methods used to develop the FMFS research flock and the phenotyping process used to generate the data used for initial gene discovery and estimation of molecular estimates of Australian Standard Breeding values.

“Design and phenotyping procedures for determination of wool, skin, parasite resistance, growth, carcass yield and quality traits of the SheepGENOMICS mapping flock”

Jason D White, Peter Allingham, Chris M Gorman, David Emery, Philip Hynd, John Owens, Amy Bell, Jason Siddell, Ben Hayes, Jonathan Usmar, Mike Goddard, John Henshall, Julius van der Werf, Frank Nicholas, Robyn Warner, Chris Hofmyer, Terry Longhurst, Paul Swan, Rob Forage, V. Hutton Oddy

To be submitted to Livestock Production.