

# Using a 1.5K ovine SNP array to expand the sheep linkage map

Jillian F. Maddox<sup>1</sup>, Allan M. Crawford<sup>2</sup>, Roxann Ingersoll<sup>3</sup>, James W. Kijas<sup>4</sup>, Brian P. Dalrymple<sup>4</sup>, Ian R.W. Evans<sup>5</sup>, Frank W. Nicholas<sup>6</sup>, David J. Townley<sup>4</sup>, John McEwan<sup>2</sup>, Tracey van Stijn<sup>2</sup>, Herman W. Raadsma<sup>6</sup>, and the International Sheep Genomics Consortium<sup>7</sup>

<sup>1</sup> Department of Veterinary Science, University of Melbourne, Australia; <sup>2</sup> AgResearch, New Zealand; <sup>3</sup> Johns Hopkins University, Institute of Genetic Medicine, USA; <sup>4</sup> CSIRO Livestock Industries, Australia; <sup>5</sup> irwevans@bigpond.net.au; <sup>6</sup> Reprogen, University of Sydney, Australia; <sup>7</sup> [http://www.sheephapmap.org/isgc\\_participants.htm](http://www.sheephapmap.org/isgc_participants.htm)

## Background

Sheep are important both for their role in agriculture and as a model organism for humans for studying the normal development of various physiological systems, as well as abnormal states caused by genetic and acquired diseases. Many of the traits of importance in sheep have been shown by breeding studies to be heritable, and research groups worldwide are conducting mapping experiments to identify the loci that influence them. The sheep linkage reference map has been developed as part of an international collaboration by genotyping the International Mapping Flock (IMF), a flock that was produced by AgResearch in the early 1990s (Crawford *et al.*, 1995). The IMF is a three-generation, full-sibling pedigree that consists of 127 sheep in nine families and contains 222 informative meioses (Figure 1). The first version of the sheep linkage map was released in 1995 (Crawford *et al.*, 1995) and the growth of the linkage map is shown in Figure 2. The most recent version, 4.7, released in December 2006 spans approximately 3,600 cM and comprises 1,425 markers representing 1,381 loci. The majority of the mapped markers on version 4.7 were microsatellites (87%), and 51% were derived from markers initially developed for cattle. Further details on the sheep linkage map can be found at [rubens.its.unimelb.edu.au/~jillm/jill.htm](http://rubens.its.unimelb.edu.au/~jillm/jill.htm).

## Genome Wide SNP Identification

A 12x bacterial artificial chromosome (BAC) sheep library (CHORI-243, [bacpac.chori.org/library.php?id=162](http://bacpac.chori.org/library.php?id=162)) was constructed under the auspices of the International Sheep Genomics Consortium (ISGC) ([www.sheephapmap.org](http://www.sheephapmap.org)) in 2002, and in 2005 the BAC ends of the library were sequenced with the sequence information being used to construct a Virtual Sheep Genome (VSG, Dalrymple *et al.*, 2007, [www.livestockgenomics.csiro.au/vsheep/](http://www.livestockgenomics.csiro.au/vsheep/)). The 370k sheep BAC end sequences (BES) data set, together with a 140k sheep expressed sequence tag (EST) data set provided by Ovita, were used to identify targets for resequencing for SNP identification. Target selection was mainly based on VSG positional information. The process aimed to identify targets that were either spaced as evenly across the genome as possible, or targets that anchored or oriented VSG contigs with uncertain positioning or orientation, or ESTs of positional interest. Primers for PCR amplification were designed using an automated pipeline at the Australian Genome Research Facility. The average target size was 508 nucleotides. In total, 2,644 genomic loci were re-sequenced using a panel of 9 sheep from 9 different breeds (Awassi, Gulf Coast Native, Katahdin, Lacaine, Merino, Poll Dorset, Red Maasai, Romney and Texel) with 2,562 of targets (97%) yielding fragments that were suitable for sequence analysis. Sequences were deposited in GenBank (GenBank GSS ET114816-ET163596). 6,021 ovine SNPs were identified corresponding to an average of 4.9 SNP per kb, and 1,536 SNPs (comprising 1,486 BES, 49 ESTs, 1 Y chromosome SNP), were selected (Table 1) for construction of an Illumina BeadArray. SNPs were selected to represent every ovine chromosome with an average of 56 SNPs per chromosome ranging from one SNP on OARY to 151 SNPs on OAR1, and 30 SNPs were included (29 ESTs, 1 BES) that lacked positional information ([www.sheephapmap.org/1536\\_shp\\_snp.xls](http://www.sheephapmap.org/1536_shp_snp.xls)).

## CRI-MAP and MultiMap modifications

The CRI-MAP (Green *et al.*, 1990) and MultiMap (Matise *et al.*, 1994) programs were modified so that they could handle a larger volume of genotyping data and run on a 64-bit platform. The CRI-MAP code was ANSI-fied and modified so that it would compile with a C++ compiler. A C++ module was developed that implemented a memory constrained product cache algorithm that minimised the amount of product recalculation required for a constrained amount of memory. The NMAX value used for shuffling was increased from 9 to a value between 20 and 50 that would allow sufficient shuffling to occur to avoid memory allocation problems. Integer overflow problems were removed by converting the scoring system from integers to long doubles. Product calculations were changed to use long doubles to avoid a loss of precision for large data sets. A single source code was used for both CRI-MAP and the modified version used by MultiMap (lisperi) so that both versions could be built simultaneously. The output of MultiMap was modified so that it would automatically produce an output map file containing all loci, with non-framework loci being positioned in the first interval of a framework map where the position had the greatest level of support.

## IMF Genotyping

The 1.5K SNP array was used to genotype 117 members of the IMF (Figure 1) at Johns Hopkins University. Initial quality control checking was performed at JHU and 1,437 (93.6%) SNPs, comprising 1,390 BES SNPs, 46 EST SNPs and the oY1 Y chromosome SNP, passed this phase. Three of the sheep (IMF17, IMF18 and IMF19) had a higher failure rate and lower GC scores than the remainder. IMF17 had the highest failure rate and was scored for only 1,027 SNPs (410 failures). However, as the three sheep were grandparents and thus contribute phase information the genotyping data from these sheep was included. Further quality control testing using CRI-MAP with the IMF pedigree identified 1,820 non-inheritance issues for 115 SNPs and 99 individuals. The majority of the non-inheritance issues found by CRI-MAP were caused by SNPs where males appeared homozygous and only females were heterozygous suggesting that these SNPs mapped to the X chromosome. The second allele was adjusted for male sheep for the 92 SNPs of this type. This reduced the number of non-inheritance issues to 313 (31 SNPs, 99 individuals). Many of these SNPs showed poor genotype clustering and these SNPs were removed from the dataset. The IMF contained more than one genotype for 1,282 of the SNPs, 1,257 SNPs had "CRI-MAP" informative meioses, and all three genotypes were present for 1,017 SNPs. The number of informative meioses ranged from 3 to 201, and the number of phase known informative meioses ranged from 0 (107 SNPs) to 171. 1,078 SNPs had more than 30 informative meioses, and 828 SNPs had more than 60 informative meioses.

## Linkage Map Construction

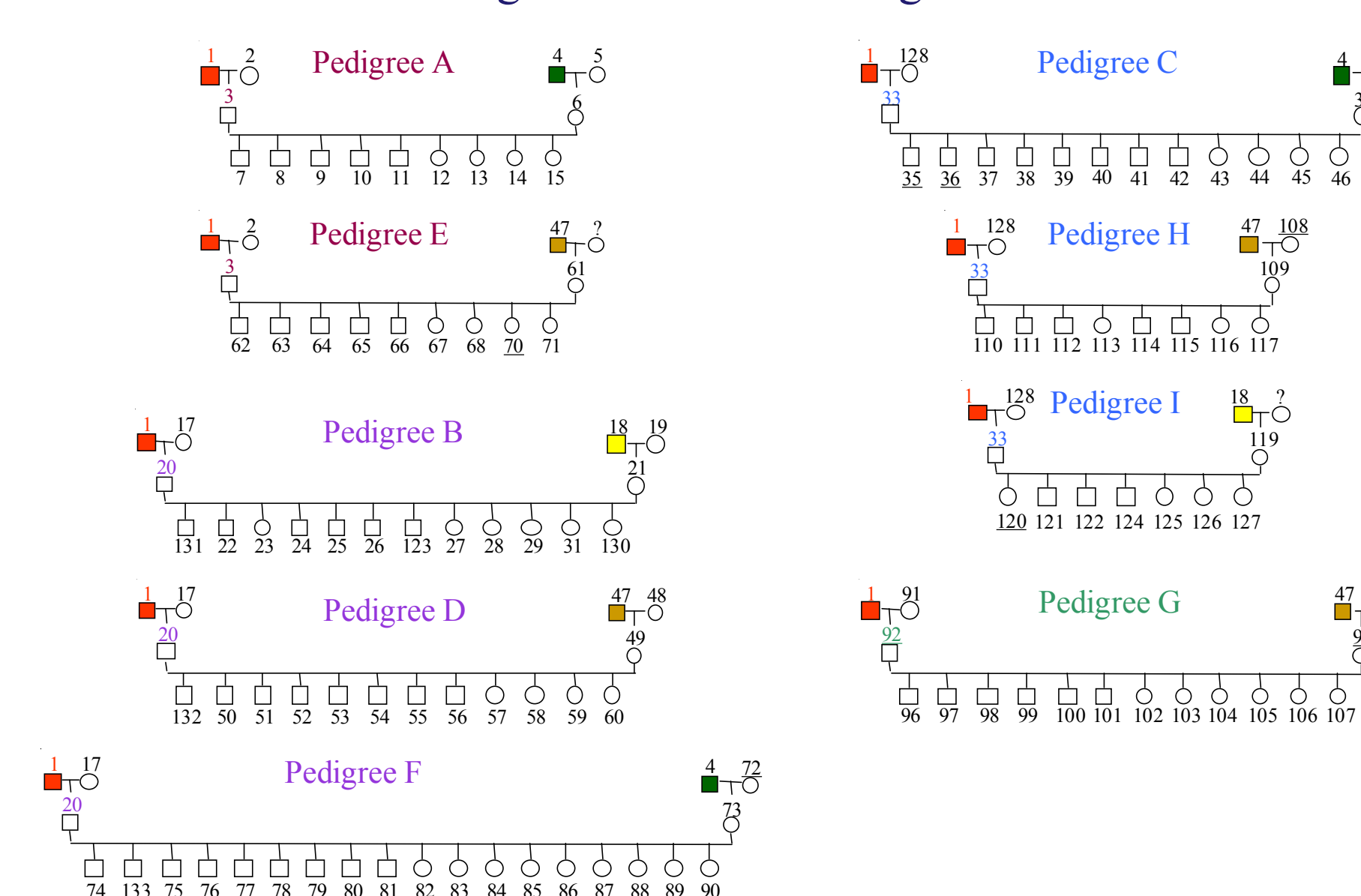
The majority of map assembly was performed on a Linux dual quadcore machine with 32 Gb physical RAM. Initially, the find-all-linkage-groups option of MultiMap (theta 0.2, lod 4) was used to group the SNPs into linkage groups. This generated 64 linkage groups of 2 or more SNPs, and 35 unlinked SNPs. The 64 groups and unlinked SNPs were assigned to chromosomes by CRI-MAP twopoint against a skeleton microsatellite map. The process used for map constructions was to first build a lod3 framework map with MultiMap, and then a lod2 framework map, and then a lod1 framework map. Remaining markers were then added into the first interval with the highest likelihood, and the orders were refined by running CRI-MAP flips<sub>n</sub> (where *n* was between 3 and 7) until no orders with better likelihoods could be found. As multiple orders exist with the same likelihood, the linkage map order was then compared to the VSG order to determine whether any rearrangements to better suit the VSG order were possible with a similar likelihood.

The new version of the linkage map, version 5, is still being constructed and the new map will contain approximately 2,528 markers including 1,111 SNPs from the array (Table 2). The addition of the SNPs to the sheep linkage map has increased the overall length of the autosomes by 221.5 cM (note, chromosome 2 has not yet been completed), of which 167.8 cM corresponds to new regions added to the ends of chromosomes and 53.7 cM is due to internal expansion (Table 2). Some of the internal map expansion is a consequence of double recombinants, and these events will be investigated. It is noteworthy that some of the previous suspect double recombination events disappeared with the addition of more markers to the map.

## Comparisons between version 5 of the Sheep Linkage Map and the VSG 1.2 and conclusions

Whilst, in general there is good correlation between the positions in the linkage map and the VSG v1.2, 33 markers appear on different chromosomes in the VSG, and there are a number of intrachromosomal order differences. Many of these are minor and are probably due to the lack of resolution of the linkage map. However some of these, such as CP88 and HRH1 shown in Table 3 for OAR19, represent likely errors in the VSG. All chromosomes show some order differences between the linkage map and the VSG, with the minimum number of discrepancies being one difference for OAR21. In conclusion, as the expanded linkage map was built completely independently of the VSG positional information, the expanded linkage map will be an important resource for checking the ordering and orientation of contigs for a sheep genome sequence map.

Figure 1 The IMF Pedigree



Grandparents and sires are colour coded. Individuals not genotyped on the SNP chip are underlined.

Figure 2 Growth of the sheep linkage map

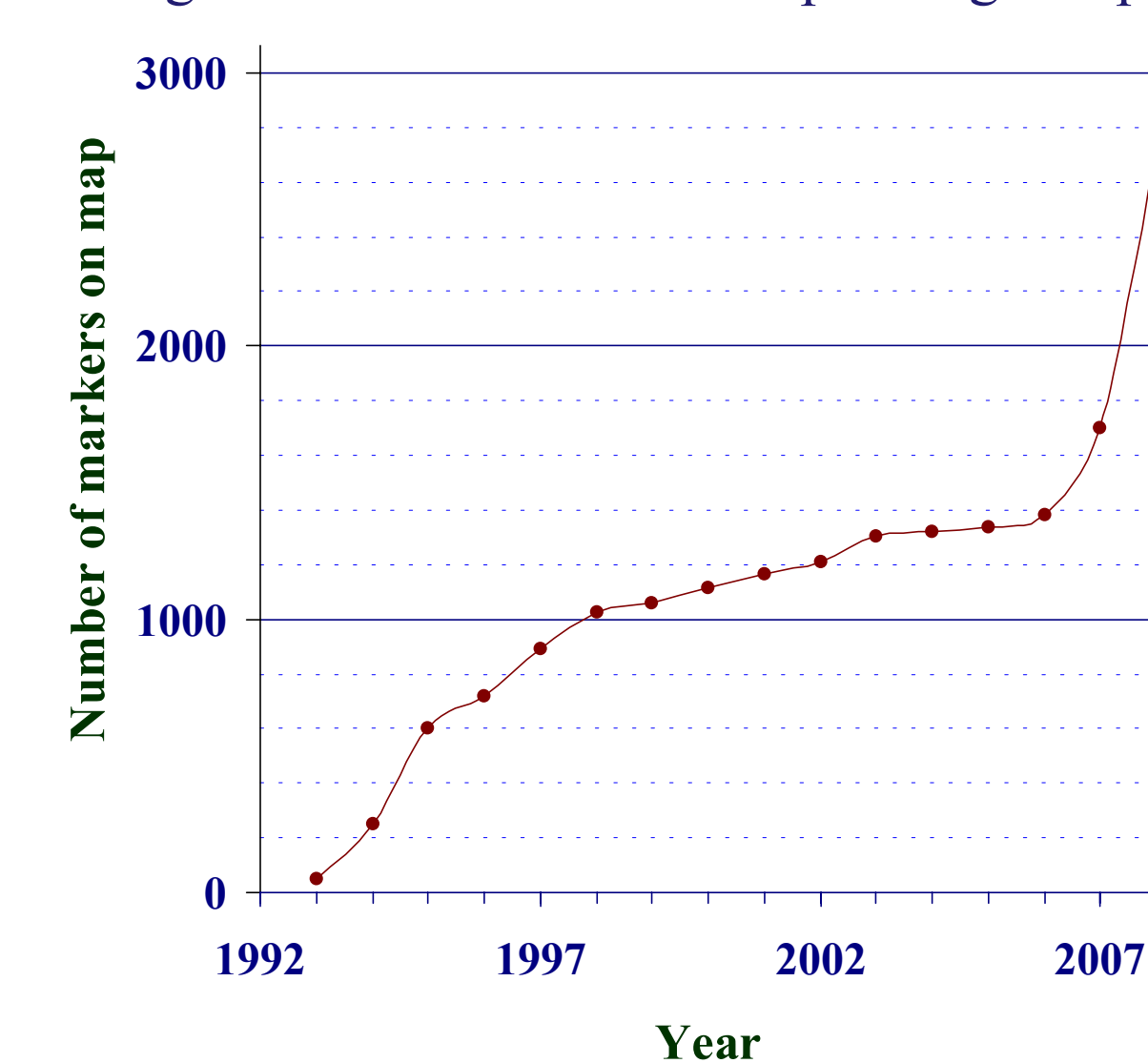


Figure 3. Distribution of markers on the sheep linkage map

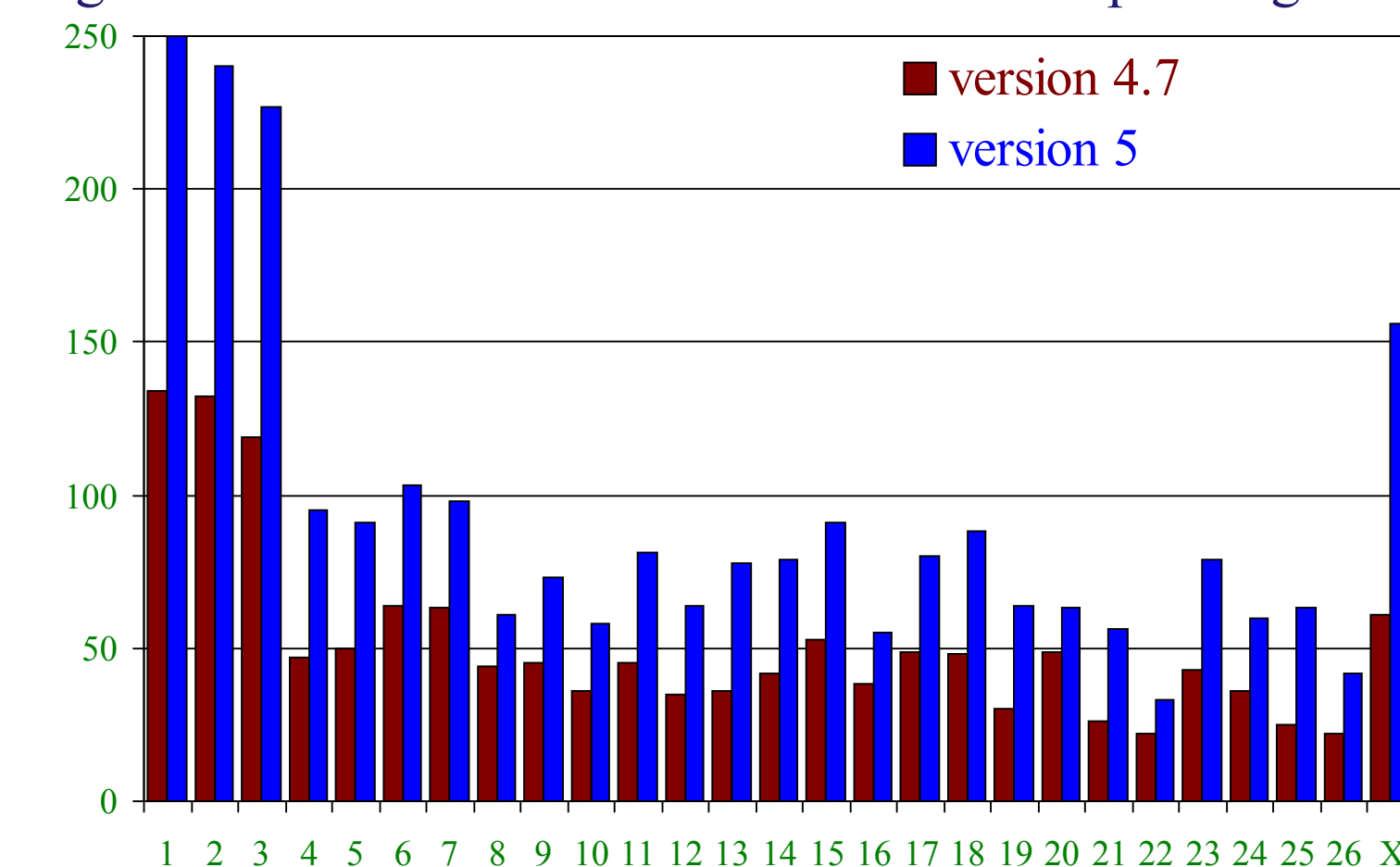


Table 1. SNPs on the 1.5k array

Criteria for Target selection	Number of SNPs
Polymorphism (SNP) identified by re-sequencing project(MAF >=0.2, Illumina score >=0.6)	1085
Filling gaps on virtual sheep genome to ensure good coverage.	197
Targets chosen to bridge gaps in virtual genome build	114
Additional SNP on large comparative genome contigs (see virtual genome)	90
Additional EST-based SNPs	22
Targets also present on RH panels	15
Targets to address over-clustering issues in virtual genome build	12
Y-chromosome SNP	1

Table 2. Chromosome details for linkage maps

Chromosome	Version 4.7 (cM)	Version 5 (cM)	Map expansion at ends top + bottom (cM)	Version 5 number of markers
1	325.2	340.0	11.3 + 0	250
2	315.2			240
3	302.3	314.4	0 + 0	227
4	147.8	166.1	8.3 + 3.9	95
5	158.9	168.4	4.8 + 0	91
6	155.7	154.6	0 + 0	103
7	148.4	147.9	0.7 + 1.7	98
8	127.8	131.5	0 + 4.9	61
9	126.9	138.0	0 + 9.8	73
10	100.2	118.3	16.5 + 0	58
11	109.6	117.7	0 + 3.7	81
12	106.4	120.4	0 + 8.9	64
13	128.3	135.0	0 + 4.4	78
14	118.9	125.3	0 + 4.3	79
15	123.8	130.8	0 + 1.4	91
16	84.7	100.2	0 + 15.9	55
17	130.0	130.4	0 + 0	80
18	127.7	130.8	0 + 0	88
19	71.8	84.7	8.6 + 3.2	64
20	99.9	100.9	0.8 + 0	63
21	75.5	78.4	0 + 0	56
22	82.9	104.6	20.2 + 0	33
23	83.9	93.3	0 + 13.4	79
24	81.6	81.6	0 + 0	60
25	68.3	84.0	3.8 + 8.7	63
26	71.1	81.8	6.7 + 1.9	42
Total autosomal	3472.8	3379.1	167.8	2372
X				156

Table 3. Comparison between SM5 and VSG1.2 positions for OAR 19

Marker	cM	VSG Mb	Informative Meioses	Phase Known Inf Meioses	VSG CGC <sup>1</sup>	VSG Confidence <sup>2</sup>
DU264531 <sup>1</sup>	0.0		109	35	885	7
DU317310 <sup>1</sup>	0.9		85	82	10017	
DU461343 <sup>1</sup>	0.9		153	22	10018	
DU446631	1.8	1.05	114	114	635	5
DU459528	2.8	0.50	78	0	635	5
DU330416	7.6	2.83	117	41	636	4
INRA26	8.6		151	130		
L51143	8.6	5.19	66	43		
PZ963	19.6	7.28	193	180		
DU272265	22.1	9.50	138	47	637	4
DU382242	22.1	10.25	131	39	638	4
BMS517	23.9		120	112		
DU473765	23.9	10.57	17	17	638	4
BM1558	23.9	11.29	203	190		
CSSM06	26.4	11.99	170	159		
CZ921601	31.9	14.79	71	34	639	1
DU519086	33.2	20.13	56	9	621	1
OXTR	33.2	20.60	106	34		
2HF3B	34.6	20.89	203	160		
BM1303	34.6	21.26	160	142		
DU488497	34.6	21.26	93	17	621	1
BM3406	36.2	21.95	138	124		
AE119	36.2	22.12	151	133		
DU486233	36.6	30.09	23	23	647	3
UCD015	37.1		98	98		
DU278349	40.9	25.01	138	0	621	1
MILVET8	40.9	25.77	184	140		
DU447987	40.9	26.34	39	6	621	1
DU476540	45.3	28.92	82	49	620	3
DU428471	46.4	29.27	129	28	619	3
DU198540	48.8	30.44	65	37	647	3
DU205548	49.8	31.37	83	31	646	1
CSSM58	51.8	33.00	124	113		
BM3628	52.6	35.26	124	79		
BMS390	54.3	36.22	46	0		
DU518795	54.3	35.62	58	43	646	1
MCM61A	54.3	37.45	144	122		
DU222179	56.8	40.26	109	39	645	4
BM2613	57.7	41.45	182	152		
DU49256	57.7	41.45	87	0	645	4
DU174410	57.7	44.12	47	47	644	5
BMS693	66.5	49.36	16	0		
BMS875	68.3	49.18	101	79		
DU447672	71.0	46.71	53	9	643	1
UCD034	71.1		42	38		
DU485754	71.1	45.23	62	29	644	5
BMS340	71.1	50.58	176	102		
DU298844	71.9	51.02	113	10	643	1
DU359756	73.7	53.39	93	14	643	1
FCB304	73.7	54.42	131	60		
LSCV14	75.3		104	79		
DU415553	75.3		112	101		
DU388595	75.3	18.67	20	12	621	1
DU502734	75.3	58.11	103	27	642	4
MCM111	75.3	58.17	159	146		
DU177621	78.4	59.67	58	27	641	6
CP88	79.1	19.05	123	72		
HRH1	81.5	18.11	103	97		
RHO	81.5	61.57	104	85		
DU411432	82.4	60.31	62	62	622	4
DU423183	82.4	61.76	129	12	669	4
DU258053	83.3	60.47	140	27	622	4
DU507401 <sup>1</sup>	84.7		87	55	623	7
DU186267 <sup>1</sup>	84.7		110	38	623	7

<sup>1</sup> VSG 1.2 comparative genome contig (CGC) positions are given for SNPs on the 1.5K array.

<sup>2</sup> VSG confidence for position of CGC: 1 (highest) to 7 (lowest).

<sup>3</sup> Chromosome discrepancies between VSG and linkage map: DU264531 on VSG1.2 chromosome 4; DU317310 and DU461343 on VSG1.2 chromosome 26; DU507401 and DU186267 on VSG1.2 chromosome 1. Other discrepancies in order are indicated in red.

## References

- Crawford A.M. *et al.* (1995) An autosomal genetic linkage map of the sheep genome. *Genetics* **140**, 703-724.  
Dalrymple B.P. *et al.* (2007) Using comparative genomics to reorder the human genome sequence into a virtual sheep genome. *Genome Biology* **8**:R152.  
Green P. *et al.* (1990) Documentation for CRI-MAP, version 2.4.  
Matise T.C. *et al.* (1994) Automated construction of genetic linkage maps using an expert system (MultiMap): a human genome linkage map. *Nature Genetics* **6**: 384-390.

## Acknowledgements

The work was funded by an Australian Government International Sciences Linkage grant and sheepGENOMICS a joint initiative of Meat and Livestock Australia and Australian Wool Innovation. The authors thank other members of the ISGC who contributed to the project; Annette McGrath and Peter Wilson of the Australian Genome Research Facility who performed the SNP resequencing work; AgResearch; and Ovita.

