

SUPPLEMENTARY MATERIALS

Comparing linkage and association analyses in sheep points to a better way of doing GWAS

Kathryn E. Kemper^{1*}, Hans D. Daetwyler^{2,3}, Peter M. Visscher^{4,5}, Michael E. Goddard^{1,2}

¹Department of Agriculture and Food, University of Melbourne, Parkville, Victoria 3010, Australia

²Victorian Department of Primary Industries, AgriBiosciences Centre, LaTrobe Research and Development Park, Bundoora, Victoria 3083, Australia

³Cooperative Research Centre for Sheep Industry Innovation, Armidale, NSW, 2351, Australia

⁴University of Queensland Diamantina Institute, University of Queensland, Princess Alexandra Hospital, Brisbane, Queensland 4102, Australia

⁵The Queensland Brain Institute, The University of Queensland, Brisbane, Queensland 4072, Australia

Part A. Algorithm for tracking the paternal alleles

First, uninformative (homozygous) stretches of more than 10 SNP were identified in the sire genotype. Then starting at the first SNP position on each chromosome stretches of >50 SNP alleles were identified as matching to either the 1st or 2nd paternal chromosome. One mismatch per section was permitted to account for genotyping and map errors. When 2 mismatches occurred and the section was ≤ 50 SNP, the algorithm restarted at the second SNP position of the mismatched section. If 2 mismatches occurred and matched section was >50 SNP, alleles from the first to the penultimate position in the section were assigned to the matched chromosome and the algorithm restarted at the final mismatched allele. If >50% of the alleles were uninformative in a matched section, the algorithm did not assign inheritance and reverted to the second SNP position of the section. The final linkage data consisted of 48,640 SNP positions where each animal had the paternal SNP allele assigned as either inherited from the 1st or 2nd paternal chromosome, uninformative or unassigned. Unassigned and uninformative alleles were treated as missing data in all subsequent linkage analyses.

Part B. Cross-validation method to estimate overestimation bias in linkage effects

The data was split into two halves and the regression slope of the sire allele effect estimated from each half of the data was obtained at 5061 random SNP positions (Hayes *et al.* 2006; Whittaker *et al.* 1997). Thus if α_1 and α_2 are the estimates of the sire allele effect from each half of the data, then the estimate with all the data ($\bar{\alpha}$) is $0.5(\alpha_1 + \alpha_2)$ and $\alpha_1 = g + e_1$, where g is the true effect, e_1 is the error associated with the first estimate. We want to assess the bias by estimating $\beta^* = \frac{\text{var}(g)}{\text{var}(\bar{\alpha})}$, such that $\beta^* < 1$ indicates overestimation of $\text{var}(g)$.

Let the regression of α_1 on α_2 be $\beta = \frac{\text{cov}(\alpha_1, \alpha_2)}{\text{var}(\alpha_1)}$ and so,

$$\beta^* = \frac{\text{cov}(\alpha_1, \alpha_2)}{0.5[\text{var}(\alpha_1) + \text{cov}(\alpha_1, \alpha_2)]}$$

$$\beta^* = \frac{2\beta}{1 + \beta}$$

For comparison to the random model ([3]), an estimate of the phenotypic variance explained by sire allele ($\sigma_{\text{sire.snep}}^2 / \sigma_p^2$) can be obtained by

$\left[\frac{1}{M} \sum_{j=1}^M \frac{W'W(1-\beta^*)}{\beta^*} \right]^{-1}$, assuming that $\sigma_{\text{sire.snep}}^2$ is small relative to the phenotypic variance, W is as defined in [3] and M is the number of SNP positions used in the regression.

Figure A1 shows the regression of α_1 on α_2 and estimates β^* as 0.173. Thus the phenotypic variance explained by the paternally inherited allele using the

cross-validation method is 0.0025. This is slightly lower than the random effect model (where $\sigma_{sire.snp}^2 / \sigma_P^2 = 0.0037$), probably because this method is less precise to account for differences in the size of the sire families, particularly the variation in number of tracked alleles for offspring at each position. The cross-validation method suggests that the fixed effect model overestimates the allele effects by a factor of 0.173^{-1} or about 6 times the true effect.

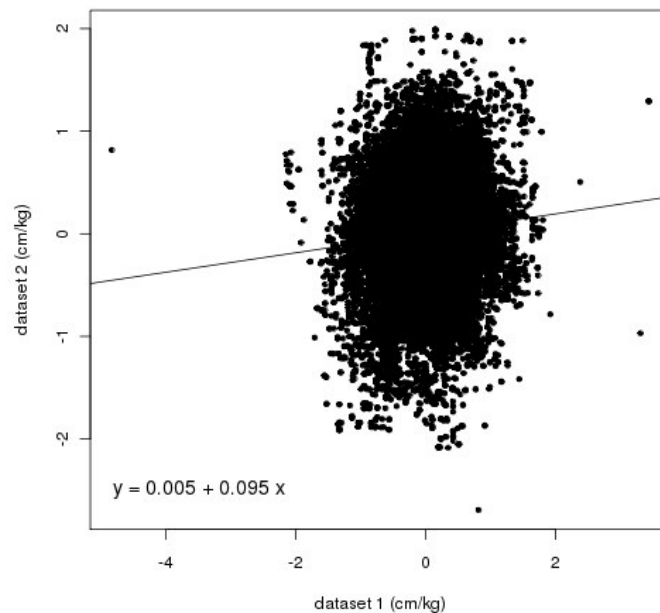


Figure A1. Regression of the estimated linkage effect using the fixed effect model ([3]) for two random splits of the data at 5061 randomly selected SNP positions.

Part C. Predicting the genetic merit of sires

The mean SNP effects from model [5] were used to predict each sires genetic merit for eye muscle depth. This marker based genetic merit was compared to the genetic merit estimated from the pedigree. Two different predictions of genetic merit with markers were calculated, first when the marker effects were estimated when all sires were included in the analysis and second when the sire to be predicted was excluded from the analysis to estimate the marker effects.

To calculate each sires genetic merit from the markers, the sum of the sires allele effects was calculated. That is, the genetic merit was $\mathbf{t}\boldsymbol{\gamma}$, where \mathbf{t} is a vector of each sires genotype and $\boldsymbol{\gamma}$ is a vector of marker effects from [5]. Marker effects were estimated either with all sires or excluding the sire to be predicted.

The model used to calculate the genetic merit of each sire from linkage, that is using the pedigree, was

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{v}' + \mathbf{e} \quad [7]$$

where \mathbf{y} is a vector of progeny phenotypes, \mathbf{X} is a design matrix assigning progeny to fixed effects, \mathbf{b} is a vector of fixed effect solutions, \mathbf{Z} is a design matrix allocating phenotypes to sires, \mathbf{v}' is a vector of random sire effects [distributed $N(0, \mathbf{I}\sigma_{sire}^2)$] and \mathbf{e} is a vector of residuals [distributed $N(0, \mathbf{I}\sigma_e^2)$, where \mathbf{I} is an identify matrix). Fixed effects in \mathbf{b} were year of birth (2 levels), age in days (covariate, mean age 304 days), birth and rearing type (3 levels) and sex nested within year (4 levels). Principal components were not fitted

when predicting sire genetic merit as they explain much of the between sire variation. The model was fitted using ASReml (Gilmour 2006) and the vector of sire solution (\mathbf{v}') used as the estimates of genetic merit.

The correlation between the pedigree and marker predictions was highest when all data was used to predict genetic merit (0.76, Table A1). This is in accordance with expectations because the same data was used for both analyses. The proportion of variance in the pedigree genetic merited explained by the SNP predictions (58%) is comparable to the variance explained in the linkage results by the sum of association effects in Figure 6a (52%).

Table A1. Correlation between the genetic merit predicted from pedigree and marker effects estimated with either all data or excluding the sire to be predicted (independent data). R^2 gives the proportion of variance in the pedigree based estimate of genetic merit explained by the merit estimated with markers.

	Data set used to estimate marker effects	
	all data	independent data
correlation	0.76	0.20
regression slope	1.06 (\pm 0.28)	0.45 (\pm 0.70)
regression intercept	0.33 (\pm 0.16)	0.13 (\pm 0.28)
R^2	0.58	0.04

When independent data was used, i.e. when sire to be predicted was excluded from the analysis estimating the marker effects, the correlation between the estimates and the proportion of genetic merit explained was

much lower (0.20 and 4% respectively, Table A1). This is not unexpected as the dataset is small and probably lacks predictive power. For example, Goddard and Hayes (2009) estimate a reference set of 100,000 animals is required for a correlation of about 0.8 between true and marker predicted genetic merit for a population with similar properties to Australian merino sheep (i.e. an effective population size, N_e , of about 800; Kijas *et al.* 2012). The regression slope for the independent dataset is not significantly different from zero, although the correlation suggests an improved ability to predict between sires differences with independent data compared to the ability to predict linkage results with the sum of association effects (Figure 6b).

Supplementary Tables

Table S1. The 136 SNP identified as significant in the linkage analysis and association analysis. Given is the SNP number, name and position (base pairs, bp); F-ratios and P-values from the linkage and association analyses and the SNP allele effect (mm) with standard error (s.e.)

SNP no.	SNP Name	chr.	position	linkage analysis		association analysis			
				F ratio	P value	F ratio	P value	effect	s.e.
4915	s69813.1	1	267298347	2.83	0.002	7.48	0.006	-0.189	0.069
4920	s68196.1	1	267617643	2.84	0.002	7.55	0.006	0.214	0.078
4955	OAR1_269640647.1	1	269640647	2.61	0.005	8.08	0.005	0.191	0.067
4982	s43468.1	1	272038126	2.55	0.007	8.65	0.003	-0.192	0.065
5002	s26271.1	1	273092616	2.51	0.004	11.28	0.001	-0.252	0.075
5015	s54564.1	1	274075318	2.77	0.001	12.36	0.000	-0.232	0.066
5017	s63115.1	1	274129594	2.78	0.001	9.84	0.002	0.259	0.083
11154	OAR3_28864450.1	3	28864450	2.32	0.008	8.85	0.003	-0.204	0.068
11190	OAR3_30728905.1	3	30728905	2.57	0.002	7.8	0.005	-0.227	0.081
11199	s24401.1	3	31149791	2.65	0.002	10.66	0.001	-0.309	0.095
11200	OAR3_31172991.1	3	31172991	2.65	0.002	7.58	0.006	-0.192	0.070
11428	s08514.1	3	42590500	2.51	0.003	14.09	0.000	0.325	0.087
11440	OAR3_43171845.1	3	43171845	2.62	0.003	8.82	0.003	-0.198	0.067
11441	OAR3_43209409.1	3	43209409	2.62	0.003	14.56	0.000	-0.262	0.069
11443	s51799.1	3	43281085	2.69	0.002	7.04	0.008	-0.183	0.069
11445	s34225.1	3	43355052	2.7	0.002	7.81	0.005	-0.249	0.089
11449	OAR3_43504420.1	3	43504420	2.85	0.001	7.31	0.007	-0.183	0.068
11461	OAR3_43953410.1	3	43953410	2.86	0.001	7.96	0.005	0.230	0.082
11468	OAR3_44397353.1	3	44397353	2.93	0.001	8.25	0.004	-0.241	0.084
11475	s72391.1	3	44712439	2.82	0.001	7.21	0.007	-0.181	0.067
11479	OAR3_45012699.1	3	45012699	2.96	0.001	6.94	0.008	0.169	0.064
11480	OAR3_45074818_X.1	3	45074818	2.95	0.001	10.91	0.001	0.249	0.075
11498	s26968.1	3	45865724	2.79	0.001	8.41	0.004	-0.200	0.069
11500	OAR3_45946367.1	3	45946367	2.79	0.001	6.86	0.009	0.176	0.067
11505	OAR3_46132071.1	3	46132071	2.72	0.002	11.38	0.001	-0.267	0.079
11506	OAR3_46242384.1	3	46242384	2.68	0.002	9.31	0.002	-0.222	0.073
11514	s42822.1	3	46712983	2.56	0.003	9.2	0.002	0.208	0.069
11515	OAR3_46785511.1	3	46785511	2.54	0.003	11.73	0.001	0.235	0.068
11517	OAR3_46866444.1	3	46866444	2.58	0.003	13.61	0.000	-0.276	0.075
11538	OAR3_47817469.1	3	47817469	2.32	0.006	7.18	0.007	0.205	0.077
11542	s16062.1	3	48510938	2.29	0.007	7.31	0.007	-0.190	0.070
11546	s31911.1	3	48655576	2.33	0.006	10.49	0.001	0.217	0.067
11550	OAR3_48886032.1	3	48886032	2.34	0.006	13.42	0.000	0.253	0.069
11553	OAR3_49000877.1	3	49000877	2.37	0.005	10.94	0.001	-0.253	0.077
11556	OAR3_49189484.1	3	49189484	2.37	0.005	12.67	0.000	-0.238	0.067
11562	OAR3_49379614.1	3	49379614	2.58	0.003	8.2	0.004	-0.193	0.068

Table S1. Cont.

SNP no.	SNP Name	chr.	position	linkage analysis		association analysis			
				F ratio	P value	F ratio	P value	effect	s.e.
11563	OAR3_49394486.1	3	49394486	2.6	0.003	21.19	0.000	0.336	0.073
11564	OAR3_49467043.1	3	49467043	2.65	0.002	6.82	0.009	0.327	0.125
11565	OAR3_49563542.1	3	49563542	2.72	0.002	9.05	0.003	0.393	0.131
11574	OAR3_50288433.1	3	50288433	2.56	0.006	9.04	0.003	0.281	0.093
11578	OAR3_50474257.1	3	50474257	2.57	0.006	10.88	0.001	-0.267	0.081
11582	OAR3_50659624.1	3	50659624	2.61	0.005	13.08	0.000	-0.244	0.067
11584	OAR3_50739593_X.1	3	50739593	2.62	0.005	10.37	0.001	-0.281	0.087
11588	OAR3_51035614.1	3	51035614	2.71	0.004	7.3	0.007	-0.196	0.073
11590	OAR3_51138929.1	3	51138929	2.71	0.004	7.39	0.007	-0.212	0.078
11600	OAR3_51547857.1	3	51547857	2.43	0.010	7.57	0.006	-0.305	0.111
11627	OAR3_53056523.1	3	53056523	2.38	0.006	11.93	0.001	0.219	0.064
11628	OAR3_53082077.1	3	53082077	2.36	0.007	10.56	0.001	0.231	0.071
11638	OAR3_53557748.1	3	53557748	2.38	0.006	9.07	0.003	-0.217	0.072
11790	s50340.1	3	62213478	2.27	0.010	10.59	0.001	-0.412	0.127
15358	DU398618_415.1	4	6125705	2.27	0.010	7.5	0.006	-0.213	0.078
15360	OAR4_6204679.1	4	6204679	2.35	0.007	6.65	0.010	-0.195	0.076
15434	OAR4_10346722.1	4	10346722	2.26	0.010	7.77	0.005	-0.301	0.108
19965	OAR6_269130.1	6	269130	2.44	0.009	6.74	0.009	0.402	0.155
19966	OAR6_308395.1	6	308395	2.44	0.009	14.36	0.000	0.341	0.090
21518	OAR6_83251333.1	6	83251333	2.54	0.003	9.66	0.002	0.242	0.078
21521	s18172.1	6	83504534	2.67	0.002	14.58	0.000	0.404	0.106
21523	OAR6_83607627.1	6	83607627	2.67	0.002	7.2	0.007	0.190	0.071
21539	OAR6_84351693.1	6	84351693	2.52	0.004	7.56	0.006	0.195	0.071
21546	OAR6_84820452.1	6	84820452	2.51	0.004	8.36	0.004	0.202	0.070
21601	OAR6_87886231.1	6	87886231	2.79	0.002	6.83	0.009	-0.277	0.106
21654	OAR6_91313780.1	6	91313780	2.54	0.003	7.99	0.005	-0.569	0.201
21730	OAR6_95830790.1	6	95830790	2.97	0.001	9.08	0.003	-0.226	0.075
21777	OAR6_98017695.1	6	98017695	2.9	0.001	11.2	0.001	0.252	0.075
21826	s01688.1	6	100588031	2.66	0.002	7.24	0.007	-0.347	0.129
21847	OAR6_101422717.1	6	101422717	2.69	0.002	13.81	0.000	-0.285	0.077
21879	OAR6_102875580.1	6	102875580	2.61	0.003	7.12	0.008	0.184	0.069
21994	s38976.1	6	108822140	2.46	0.005	6.76	0.009	0.260	0.100
22032	OAR6_110771028.1	6	110771028	2.45	0.005	8.46	0.004	-0.192	0.066
22075	OAR6_112991064.1	6	112991064	2.49	0.004	6.8	0.009	-0.235	0.090
22252	OAR6_123166087.1	6	123166087	2.39	0.005	8.39	0.004	0.282	0.097
22255	s37950.1	6	123243237	2.53	0.003	10.71	0.001	0.252	0.077
22297	OAR6_125026886.1	6	125026886	2.73	0.002	17.34	0.000	-0.368	0.088
22298	OAR6_125088022.1	6	125088022	2.71	0.002	6.89	0.009	-0.198	0.076
22318	OAR6_126151313.1	6	126151313	2.42	0.004	8.54	0.004	0.206	0.071
22348	s47717.1	6	127882842	2.29	0.009	7	0.008	-0.187	0.071
22412	OAR7_1946256.1	7	1946256	2.28	0.007	7.21	0.007	-0.239	0.089
22442	OAR7_3297890.1	7	3297890	2.63	0.002	7.49	0.006	0.185	0.068
22519	OAR7_6864118.1	7	6864118	3.04	0.000	7.51	0.006	-0.211	0.077
22539	s32513.1	7	7726714	2.97	0.000	9.18	0.002	-0.248	0.082
22554	s44621.1	7	8351647	3	0.000	6.7	0.010	0.223	0.086
22591	OAR7_10235711.1	7	10235711	2.87	0.001	7.85	0.005	-0.477	0.170
22605	OAR7_10929173.1	7	10929173	2.82	0.001	9.2	0.002	0.210	0.069
22620	OAR7_11778784.1	7	11778784	2.65	0.002	6.94	0.008	0.219	0.083
22631	OAR7_12244180.1	7	12244180	2.53	0.003	7.42	0.007	-0.191	0.070
22638	s58081.1	7	12530462	2.47	0.003	19.83	0.000	0.311	0.070

Table S1. Cont.

SNP no.	SNP Name	chr.	position	linkage analysis		association analysis			
				F ratio	P value	F ratio	P value	effect	s.e.
22643	s24310.1	7	12858005	2.79	0.001	7.42	0.007	-0.203	0.075
22677	OAR7_14444904.1	7	14444904	3.06	0.000	6.82	0.009	0.488	0.187
22678	s27071.1	7	14505016	3.06	0.000	9.97	0.002	-0.287	0.091
24354	s39758.1	7	103295984	2.25	0.008	6.91	0.009	0.245	0.093
24835	OAR8_18770427.1	8	18770427	2.41	0.008	11.22	0.001	0.271	0.081
24840	OAR8_18997620.1	8	18997620	2.43	0.007	8.18	0.004	-0.237	0.083
24852	OAR8_19922267.1	8	19922267	2.42	0.007	14.92	0.000	0.275	0.071
24863	s41707.1	8	20652593	2.5	0.004	9.2	0.002	-0.273	0.090
24883	OAR8_21613515.1	8	21613515	2.71	0.002	8.34	0.004	0.191	0.066
27606	OAR9_62768786_X.1	9	62768787	2.59	0.003	8.4	0.004	-0.206	0.071
27658	OAR9_65469087.1	9	65469087	2.25	0.008	8.66	0.003	0.221	0.075
27688	s27135.1	9	67366454	2.41	0.004	9.14	0.003	-0.287	0.095
27949	OAR9_80183649.1	9	80183649	2.44	0.004	8.28	0.004	0.332	0.116
27978	OAR9_81415466.1	9	81415466	2.73	0.001	7.92	0.005	-0.180	0.064
28025	OAR9_83941099.1	9	83941099	2.7	0.001	7.11	0.008	-0.191	0.072
28136	s58148.1	9	88749619	2.47	0.005	7.1	0.008	-0.183	0.068
28137	OAR9_88784528.1	9	88784528	2.48	0.004	7.81	0.005	-0.209	0.075
28140	OAR9_88886896.1	9	88886896	2.43	0.005	6.71	0.010	-0.184	0.071
29552	OAR10_61107395.1	10	61107395	2.21	0.009	7.77	0.005	0.244	0.088
29573	OAR10_62158653.1	10	62158653	2.24	0.008	7.76	0.005	0.215	0.077
29574	OAR10_62219787.1	10	62219787	2.24	0.008	8	0.005	0.205	0.073
29578	OAR10_62334685.1	10	62334685	2.24	0.008	7.68	0.006	0.212	0.077
29579	OAR10_62361045.1	10	62361045	2.24	0.008	8.26	0.004	-0.249	0.087
29592	OAR10_62812314.1	10	62812314	2.37	0.005	11.45	0.001	-0.268	0.079
29593	OAR10_62872792.1	10	62872792	2.37	0.005	17.61	0.000	0.302	0.072
29595	OAR10_62928758.1	10	62928758	2.37	0.005	9.67	0.002	0.227	0.073
29596	OAR10_62938042.1	10	62938042	2.37	0.005	7.65	0.006	-0.235	0.085
29597	OAR10_62988399.1	10	62988399	2.37	0.005	11.69	0.001	-0.287	0.084
29628	OAR10_65943285.1	10	65943285	2.46	0.003	7.07	0.008	-0.269	0.101
29630	OAR10_65976077.1	10	65976077	2.43	0.004	10.19	0.001	0.253	0.079
29640	OAR10_66804684.1	10	66804684	2.39	0.005	6.76	0.009	-0.351	0.135
32053	OAR12_48310169.1	12	48310169	2.58	0.006	6.98	0.008	0.212	0.080
32062	s40079.1	12	48735033	2.77	0.003	8.08	0.005	0.217	0.076
33042	OAR13_16167298.1	13	16167298	2.41	0.006	9.1	0.003	-0.234	0.078
38251	OAR16_68076864.1	16	68076864	2.56	0.002	7.64	0.006	0.229	0.083
38261	s68943.1	16	68555554	2.67	0.001	7.06	0.008	-0.172	0.065
38282	OAR16_69575140.1	16	69575140	2.5	0.003	9.24	0.002	0.215	0.071
38309	OAR16_71075325_X.1	16	71075326	2.52	0.003	6.68	0.010	-0.170	0.066
38312	s53066.1	16	71283840	2.24	0.008	6.81	0.009	0.225	0.086
38332	OAR16_72087208.1	16	72087208	2.35	0.005	8.8	0.003	-0.216	0.073
44983	OAR22_49529205.1	22	49529205	2.44	0.009	7.78	0.005	-0.281	0.101
45046	s75918.1	22	52559544	2.89	0.001	7.15	0.008	-0.177	0.066
45065	s64601.1	22	53588682	2.9	0.001	9.88	0.002	-0.408	0.130
45067	s73684.1	22	53640797	2.88	0.001	10.83	0.001	-0.223	0.068
45069	s20215.1	22	53652273	2.83	0.002	7.77	0.005	-0.197	0.071
45077	s40969.1	22	53963300	3.06	0.001	9.56	0.002	0.230	0.074

Supplementary Figures

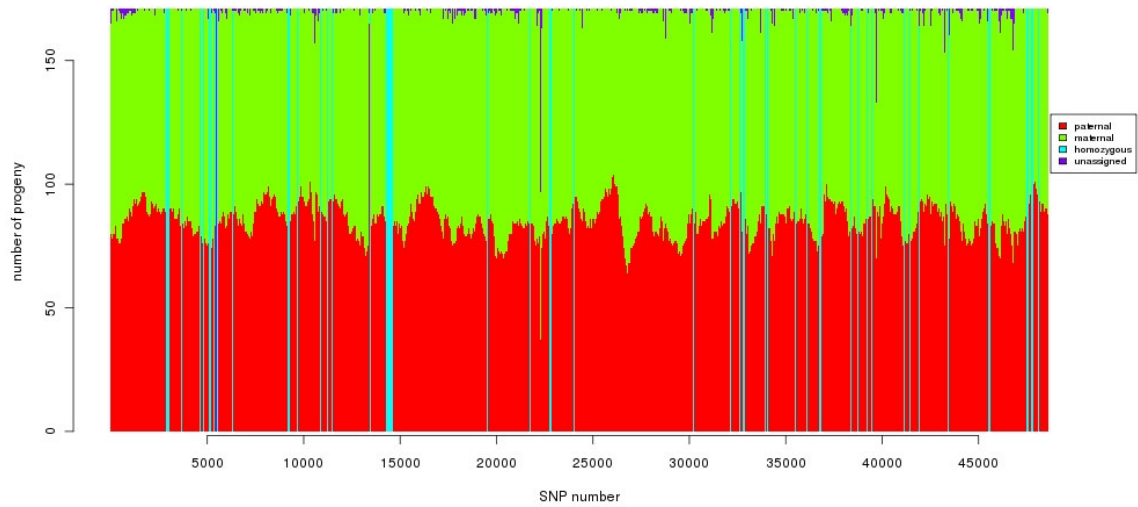


Figure S1. Number of progeny determined to inherit alleles from the 1st (red) or 2nd (green) chromosome and unassigned (purple) for a single sire family (“W4”). Uninformative positions are also shown (light blue).

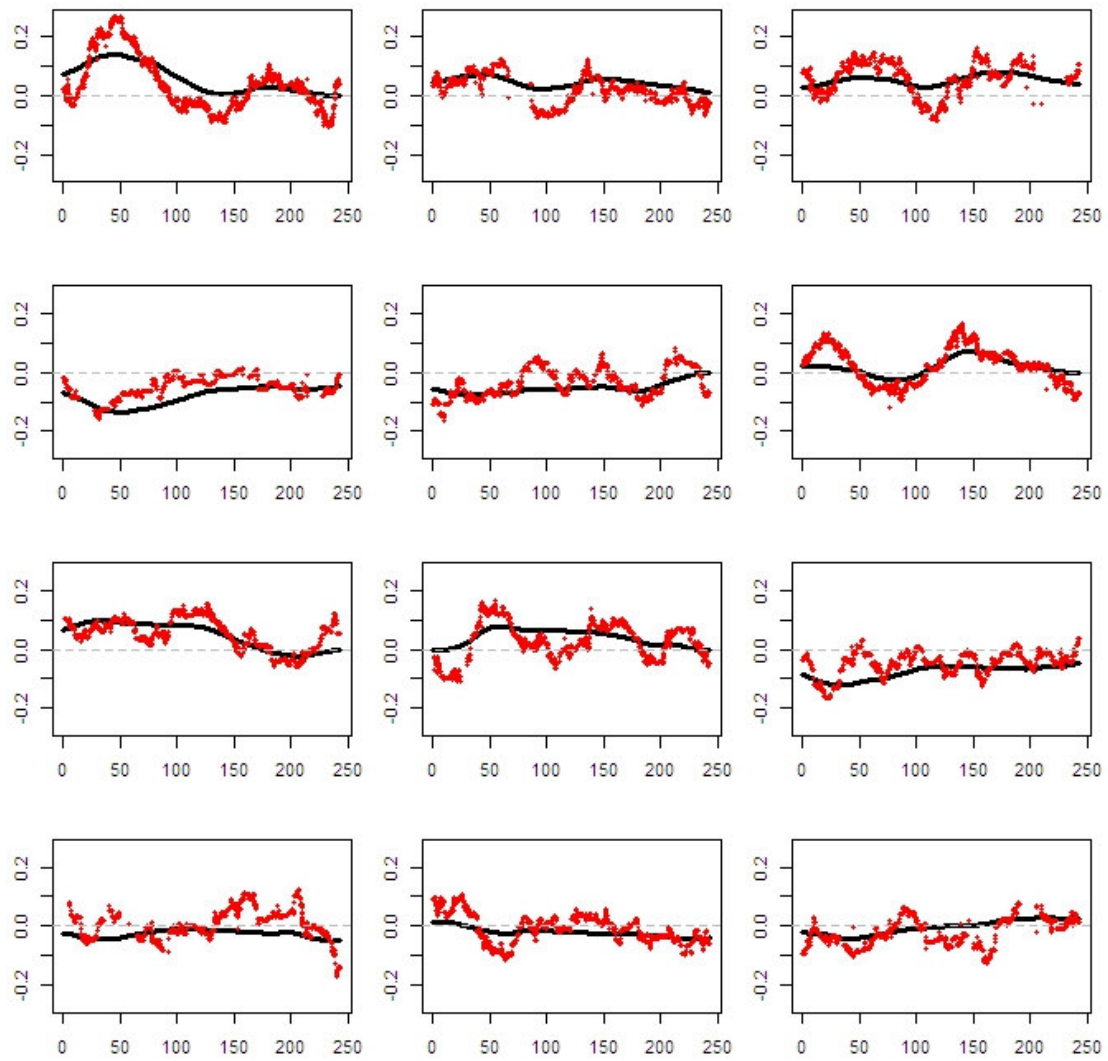


Figure S2. The size of QTL effects (mm) for chromosome 3 (Mbp) for the 12 sires using the sum of association SNP effects accounting for recombination (black) or within family linkage effects (red). The sum of association effects shows a fair correlation with the linkage results, particularly when the linkage effect is large (i.e. at around 50 Mbp for the sire at the top left).

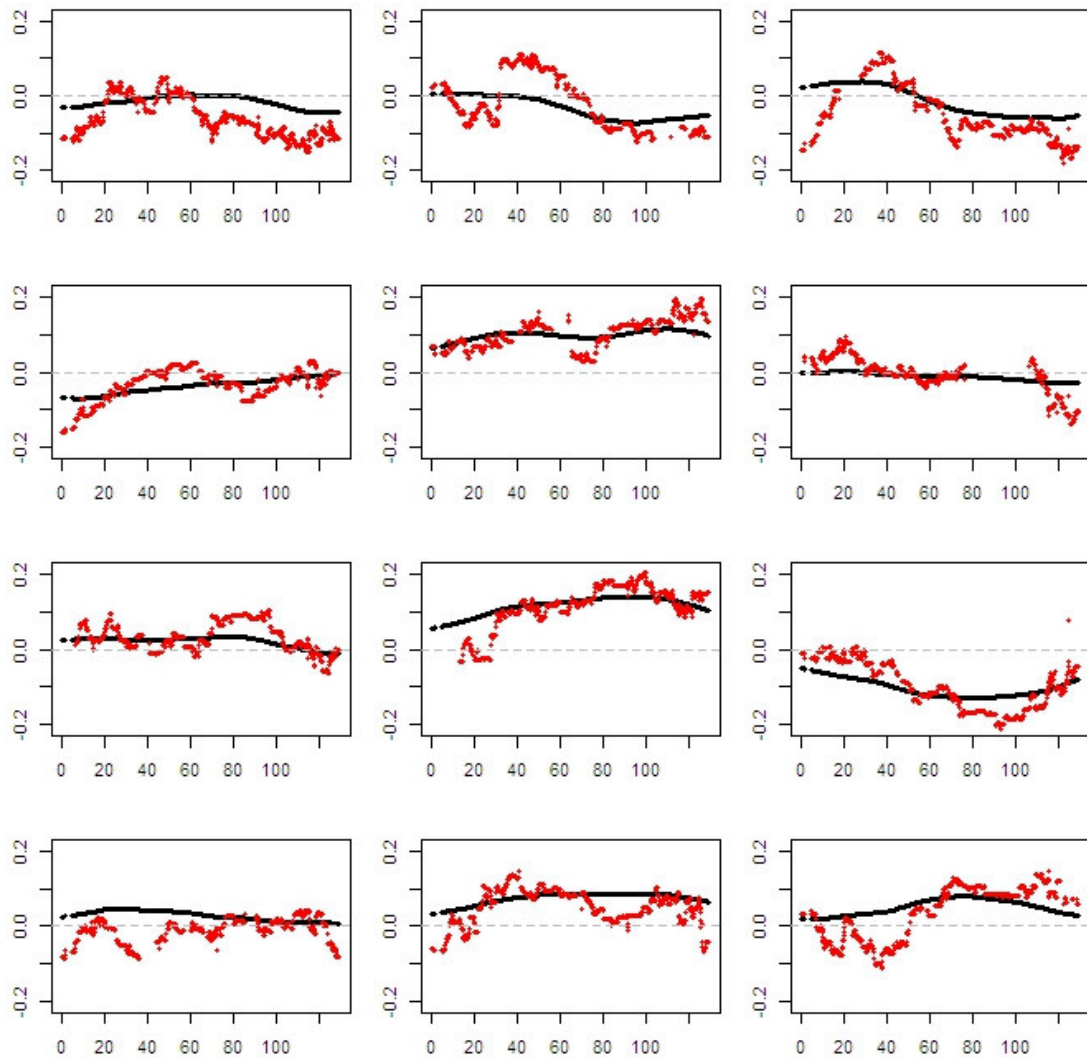


Figure S2 (cont.). The size of QTL effects (mm) for chromosome 6 (Mbp) for the 12 sires using the sum of association SNP effects accounting for recombination (black) or within family linkage effects (red). Similar to above, the sum of association effects shows a fair correlation with the linkage results for large linkage effects (i.e. at around 100 Mbp for 2 sires to the right in the 3rd row).

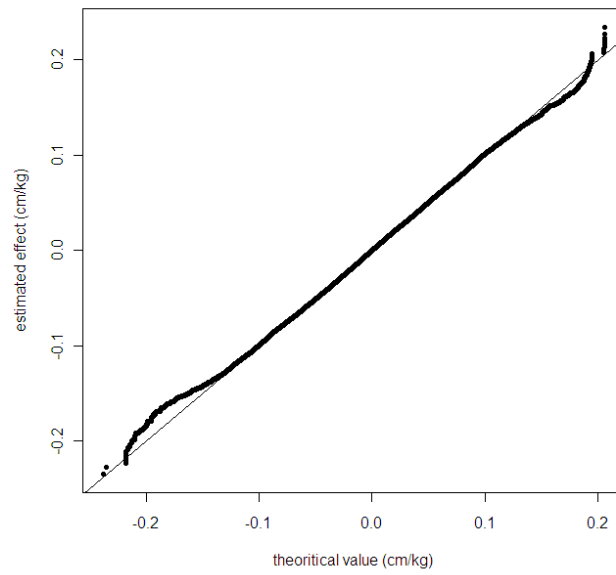


Figure S3. The qq-plot of residuals from the regression of linkage effects on the sum of the association effects. Residuals show a normal distribution with $\sigma^2_e=0.0026$.

References

- Gilmour, A. R., Gogel, B. J., Cullis, B. R. & Thompson, R. (2006). ASReml user guide 2.0. VSN International Ltd., Hemel Hempsted, UK.
- Goddard, M.E. and Hayes, B.J. (2009) Mapping genes for complex traits in domestic animals and their use in breeding programs. *Nature Reviews Genetics*. **10**:381-391.
- Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Porto Neto, L. R., San Cristobal, M., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raadsma, H., McEwan, J., Dalrymple, B. & other members of the International Sheep Genomics, C. (2012). Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biology* 10(2): e1001258.
- Hayes, B. J., Chamberlain, A. J. & Goddard, M. E. (2006). *8th World Congress on Genetics Applied to Livestock Production*, Belo Horizonte, Brasil, p. 30-06.
- Whittaker, J. C., Haley, C. & Thompson, R. (1997). Optimal weighting of information in marker-assisted selection. *Genetics Research* 69, 137-144.