

RESEARCH ARTICLE

Open Access



# A conditional multi-trait sequence GWAS discovers pleiotropic candidate genes and variants for sheep wool, skin wrinkle and breech cover traits

Sunduimijid Bolormaa<sup>1,2\*</sup>, Andrew A. Swan<sup>2,3</sup>, Paul Stothard<sup>4</sup>, Majid Khansefid<sup>1,2</sup>, Nasir Moghaddar<sup>2,5</sup>, Naomi Duijvesteijn<sup>2,6</sup>, Julius H. J. van der Werf<sup>2,5</sup>, Hans D. Daetwyler<sup>1,2,7</sup> and Iona M. MacLeod<sup>1,2</sup>

## Abstract

**Background:** Imputation to whole-genome sequence is now possible in large sheep populations. It is therefore of interest to use this data in genome-wide association studies (GWAS) to investigate putative causal variants and genes that underpin economically important traits. Merino wool is globally sought after for luxury fabrics, but some key wool quality attributes are unfavourably correlated with the characteristic skin wrinkle of Merinos. In turn, skin wrinkle is strongly linked to susceptibility to “fly strike” (*Cutaneous myiasis*), which is a major welfare issue. Here, we use whole-genome sequence data in a multi-trait GWAS to identify pleiotropic putative causal variants and genes associated with changes in key wool traits and skin wrinkle.

**Results:** A stepwise conditional multi-trait GWAS (CM-GWAS) identified putative causal variants and related genes from 178 independent quantitative trait loci (QTL) of 16 wool and skin wrinkle traits, measured on up to 7218 Merino sheep with 31 million imputed whole-genome sequence (WGS) genotypes. Novel candidate gene findings included the *MAT1A* gene that encodes an enzyme involved in the sulphur metabolism pathway critical to production of wool proteins, and the *ESRP1* gene. We also discovered a significant wrinkle variant upstream of the *HAS2* gene, which in dogs is associated with the exaggerated skin folds in the Shar-Pei breed.

**Conclusions:** The wool and skin wrinkle traits studied here appear to be highly polygenic with many putative candidate variants showing considerable pleiotropy. Our CM-GWAS identified many highly plausible candidate genes for wool traits as well as breech wrinkle and breech area wool cover.

## Background

More than half of the 70 million sheep in Australia are pure Merino that are traditionally bred for high value wool which is globally sought after for luxury fabric manufacture. The gross value of wool produced in Australia was 3 billion AU dollars in 2020 (<https://woolp>

[producers.com.au/about-us/wool-trade/](https://woolproducers.com.au/about-us/wool-trade/)). The value of a sheep’s wool fleece depends on several key attributes, including fleece weight, fibre diameter, staple strength and length, as well as crimp (or curvature) [1]. Each of these attributes are moderately to highly heritable and show some degree of genetic correlation between them. Some traits are correlated in a favourable direction (e.g. clean fleece weight and staple length [2]) while for other trait pairs there are significant unfavourable correlations that affect both wool quality (e.g. fleece weight and fibre diameter) and welfare traits (e.g. breech skin wrinkle

\*Correspondence: [bolormaa.sunduimijid@agriculture.vic.gov.au](mailto:bolormaa.sunduimijid@agriculture.vic.gov.au)  
<sup>1</sup> Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, VIC 3083, Australia  
Full list of author information is available at the end of the article



© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

and fibre diameter) [3]. While breech skin wrinkle and breech area wool cover are not of direct economic value, they are positively genetically correlated to susceptibility to “flystrike” (*Cutaneous myiasis*) [4]. When the skin around the sheep’s rear (breech area) is heavily wrinkled and woolly, this area is more likely to become damp and soiled by faeces (thus attracting flies to lay eggs) compared to less wrinkled and bare skin. When the eggs hatch the larvae can break the skin to feed on exudate (*Cutaneous myiasis*). Flystrike imposes a heavy economic and welfare burden in Australian sheep flocks and was recently estimated to cost the Australian industry over \$170 million annually [5].

To our knowledge, only two genome-wide association studies (GWAS) have previously mapped quantitative trait loci (QTL) for a range of wool traits or skin wrinkle in sheep, however, these studies used only 50k [6] or high-density (600k) SNP array genotypes [7]. The estimated effect of a SNP on phenotype depends on the linkage disequilibrium (LD) between the SNP and the causal variant. Standard SNP arrays are very unlikely to include causal variants and, furthermore, the SNPs on arrays are preselected to be (highly) polymorphic across breeds. This can result in rarer or breed-specific causal variants that are not in strong LD with the array SNPs, and thus their effects may not be captured in GWAS. In contrast to SNP arrays, whole-genome sequence (WGS) data should include all or at least many of the causal variants. Recently, we demonstrated that the use of imputed whole-genome sequence (WGS) variants increased the accuracy of genomic selection for wool traits in sheep [8]. Thus, it is of considerable interest to undertake genome-wide association studies (GWAS) with imputed sequence to fine-map putative causal variants and genes underpinning economically important wool traits as well as breech wrinkle. This would provide a better understanding of the underlying biology and pleiotropy across traits. Also, variants with large effects for single traits or with multiple pleiotropic effects are of particular interest to enrich custom SNP panels to improve accuracy of genomic prediction [9].

Use of multi-trait meta-GWAS analysis that combines results from individual trait GWAS has been demonstrated to be beneficial for identifying pleiotropic variant effects among traits for SNP array genotypes [7, 10–12], but has not yet been tested for sequence-level GWAS in sheep. Pausch et al. [13] demonstrated that, while it is possible to identify causal variants using GWAS with accurately imputed WGS, precise mapping can be difficult due to long-range LD resulting in many variants being associated with each real causal variant. An alternative approach to GWAS for fine-mapping QTL is to fit all the variants simultaneously in a model, such as BayesR

[14, 15] that accounts for linked markers, fits a sparse distribution of QTL effects and allows for diverse genetic architectures. With many thousands of individuals, BayesR with WGS becomes computationally challenging, however a subset of the sequence variants can be tested instead.

Therefore, the main objective of this study was to fine-map putative pleiotropic causal variants for 16 wool and breech wrinkle traits measured in 7218 Merino sheep. We did this by extending the approach of Bolormaa et al. [10] and undertook a stepwise conditional multi-trait GWAS (CM-GWAS) with 31 million WGS variants. We present a list of annotated putative causal variants and genes and indicate their pleiotropic effects across traits.

## Methods

### Phenotype data and traits

These Merino animals were sourced from the mixed breed Information Nucleus (IN) flock of the Cooperative Research Centre for Sheep Industry Innovation (Sheep CRC) [16, 17]. The project made available a “Q matrix” with breed proportion for each animal from among a range of breeds, as well as strains of Merino sheep, based on pedigree recording information [18]. For this study, we used only “pure” Merino (MER) animals, where “pure” was defined as the sum of their breed proportions being more than 90% Merino. Merinos were identified as belonging to one of three strains based on wool type: ultra-fine, fine to medium, and broad fibre diameter wool [18]. Up to 7218 animals were measured for 16 traits. Wool production traits were measured at two ages defined as “yearling” (150 days < Age < 550 days) and “adult” (Age ≥ 550 days) and these were considered as different traits. Subjective scores for breech wool cover and breech skin wrinkle were also analysed because they are important flystrike susceptibility indicator traits [4]. Not all sheep were measured for all traits. Trait definitions and numbers of records for each trait of the corrected phenotypes for the animals with phenotypes and genotypes are in Table 1. Phenotypes were obtained from the official Sheep Genetics industry genetic evaluation database [19], and were pre-adjusted for various fixed effects (contemporary group, flock, drop year, sex, birth type, and rearing type, age of measurement and age of dam) at AGBU, NSW, Australia. A complete description of the measurement and recording of wool production and quality assessments is in Hatcher et al. [20] and Bolormaa et al. [7].

### Whole-genome sequence genotype data

WGS genotypes of 7218 Merinos used in this study were imputed together with a larger group of sheep (47,000 animals), following Bolormaa et al. [21] and, which we

**Table 1** Trait names and acronyms (a = adult and y = yearling), number of records and units of measure

Trait acronym	Full description	Units of measure	Number of records
ygfw	Yearling greasy fleece weight	kg	6741
agfw	Adult greasy fleece weight	kg	4307
ycfw	Yearling clean fleece weight	kg	6105
acfw	Adult clean fleece weight	kg	3096
ysl	Yearling staple length	mm	4644
asl	Adult staple length	mm	3159
yfd	Yearling mean fibre diameter	µm	6359
afd	Adult mean fibre diameter	µm	3157
ydcv	Yearling fibre diameter coefficient of variation	%	6365
adc	Adult fibre diameter coefficient of variation	%	3158
ycuv	Yearling mean fibre curvature	°/mm	4203
acuv	Adult mean fibre curvature	°/mm	3187
yss	Yearling staple strength	N/k tex	4639
ass	a. staple strength	N/k tex	3155
ebwr	Breech wrinkle	1–5 score	7218
ebcov	Breech cover	1–5 score	5747

briefly described here for clarity. The 47k animals were a mix of purebreds, crosses and composites were imputed together as part of a research project that demonstrated that the use of sequence variants increased the accuracy of genomic prediction [8]. The Merinos used in our study were jointly phased and imputed to whole-genome sequence because empirical testing showed that this approach surpassed the accuracy of imputing breed groups separately [21]. The reference sequence data had an average read depth of ~10x, and details of the bioinformatic pipeline for quality control and variant calling are described in Bolormaa et al. [21].

The Merino animals were previously genotyped using low-density (“12k”), medium-density (“50k”), or high-density (“HD” with ~500k) SNP panels. Quality control of SNP array genotypes, imputation of sporadic missing genotypes within each SNP panel, and imputation of the genotypes from lower density to medium density, and then to HD SNP panels are described in [11, 22]. The imputation from HD SNPs to WGS variants was performed using the Minimac3 algorithm (version 2.0.1; [23]). Minimac3 requires pre-phased genotypes, in both reference (WGS) and target sets. The Eagle (version 2.3; [24]) software was used for pre-phasing in both reference WGS and HD target sets. The WGS data of 726 animals representing multiple European breeds and crosses [25] were used as a reference set as detailed in Bolormaa et al. [21]. Prior to imputation, we removed variants from the reference sequence data with less than 5 minor allele counts, variants with more than 2 alleles (2.97 million), and the variants on the X chromosome,

leaving 39,844,235 variants. After imputation to WGS, all imputed SNPs and indels with the Minimac3  $R^2$  statistic higher than 0.4 were retained (removing approximately 20% of the variants), resulting in a final set of 31,154,082 imputed variants for the 47,000 sheep animals. Bolormaa et al. [21] showed that the Minimac3  $R^2$  statistic is a good proxy for empirical imputation accuracy. According to their study, a Minimac3  $R^2$  higher than 0.4 corresponded to the empirical imputation accuracy of 0.87 (measured as the correlation between real and imputed genotypes). This threshold does remove a large proportion of very rare variants because these are the most difficult to impute accurately. However, very rare variants can also result in false positive results in GWAS, thus we also imposed a MAF threshold for the GWAS. All variants were fully annotated for a range of characteristics using the NGS-SNP pipeline [26].

**Sequence GWAS**

**Single-trait GWAS**

The mixed model used for the GWAS fitted each sequence variant as a covariate, one at a time, and tested for association with the trait:

$$y = \mathbf{1}_n\mu + \mathbf{s}_i\alpha_i + \mathbf{Q}\mathbf{q} + \mathbf{g} + \mathbf{e}, \tag{1}$$

where  $\mathbf{y}$  is the vector of phenotypic values pre-adjusted for fixed effects of the animals,  $\mathbf{1}_n$  is an  $n \times 1$  vector of 1s ( $n$ =number of animals with phenotypes),  $\mu$  is the overall mean,  $\mathbf{s}_i$  is a vector of genotypes (coded as 0, 1, and 2) for each animal at the  $i$ -th variant,  $\alpha_i$  is the corresponding variant effect,  $\mathbf{Q}$  is a matrix with Merino strain

proportions calculated from the pedigree ( $\mathbf{q} \sim N(\mathbf{0}, \mathbf{I}\sigma_q^2)$ ) [18],  $\mathbf{g}$  is a vector of genomic breeding values (GEBV)  $\sim N(\mathbf{0}, \mathbf{G}\sigma_g^2)$ , where  $\sigma_g^2$  is the genetic variance and  $\mathbf{G}$  is the genomic relationship matrix (GRM) calculated from HD genotypes, and  $\mathbf{q}$  and  $\mathbf{e}$  are the vectors of random effects of Merino strain proportions and residual error, respectively. Merino strains included ultra-fine, fine to medium, and broad fibre diameter wools. For a variant to be included in the GRM, its minor allele frequency (MAF) had to be higher than 0.005 in the entire dataset. The analysis was performed using the Wombat software [27]. Variance components for random effects were first computed without fitting SNP effects ( $s_i\alpha_i$ ) in the model (Eq. 1) using Wombat. Following Bolormaa et al. [28], the false discovery rate (FDR) for variants associated with each trait at  $P$ -values of  $10^{-5}$  and  $10^{-6}$  was calculated as:  $FDR = P(1 - A/T)/((A/T)(1 - P))$ , where  $P$  is the  $P$ -value tested,  $A$  is the number of SNPs that were significant at the  $P$ -value tested and  $T$  is the total number of SNPs tested.

**CM-GWAS**

Here, we extend the multi-trait meta-analysis method (M-GWAS) described in Bolormaa et al. [10] to a conditional multi-trait meta-analysis (CM-GWAS). This analysis used WGS variant effects estimated from the 16

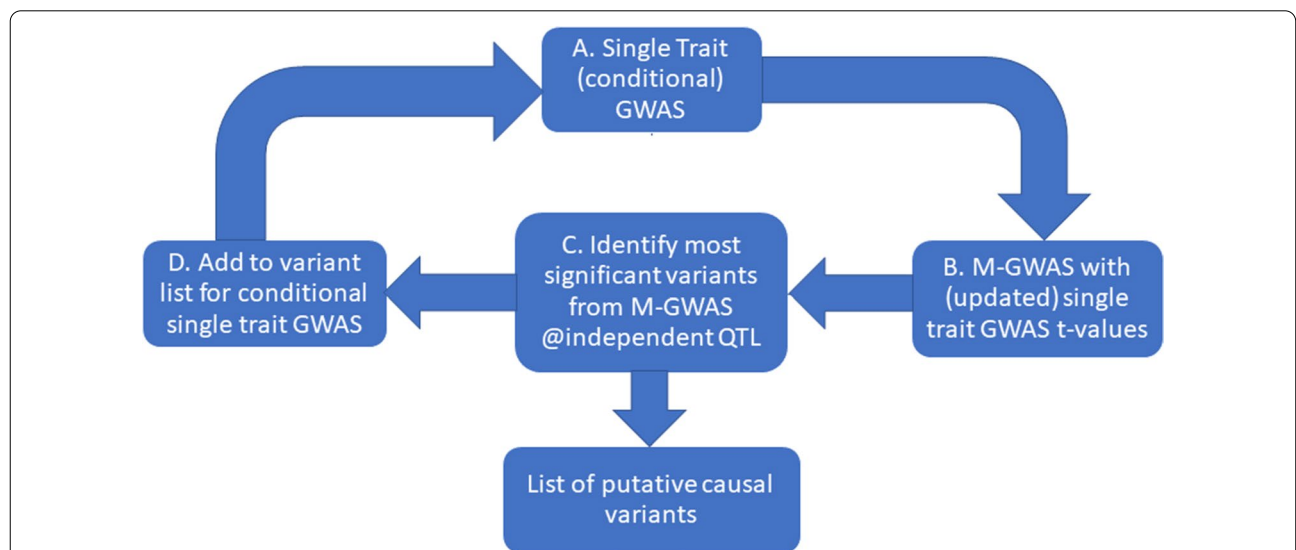
single-trait GWAS to identify pleiotropic variants that affected wool traits, breech wrinkle and breech cover. The multi-trait meta-analysis (M-GWAS)  $\chi^2$  statistic has degrees of freedom equal to the number of traits analysed and was calculated as described in [10]:

$$\text{multi-trait } \chi^2 = \mathbf{t}'_i \mathbf{V}^{-1} \mathbf{t}_i, \tag{2}$$

where  $\mathbf{t}_i$  is a vector of the signed t-values of the effects of the  $i$ -th SNP for the 16 traits and  $\mathbf{V}^{-1}$  is the inverse of the  $16 \times 16$  correlation matrix where the correlation was calculated over the all estimated SNP effects (signed t-values) between each pair of traits. Details of the multi-trait statistic property are in [10].

The CM-GWAS approach cycles back and forward between the single-trait GWAS and M-GWAS to re-test variants conditional on jointly fitting the most significant putative causal variants from independent QTL regions (where we defined significant as  $P < 10^{-5}$ ). The independence of QTL regions was defined by LD levels among variants. The steps for the CM-GWAS are outlined below and visualized in Fig. 1.

1. A single-trait GWAS was performed for all 16 traits.
2. The signed t-values ( $\mathbf{t}_i$ =each SNP effect from GWAS divided by its standard error) and correla-



**Fig. 1** Workflow for the conditional multi-trait meta-GWAS (CM-GWAS). The CM-GWAS cycles between single-trait GWAS for all traits and a multi-trait meta-GWAS (M-GWAS). After the first cycle, steps from A to D are repeated, jointly fitting the most significant independent M-GWAS variants in a conditional single-trait GWAS for all traits. The t-values from the conditional single trait GWAS (step A) are then updated to re-run the M-GWAS (step B). This continues for one or more cycles until no more significant variants from the M-GWAS are detected in step C. At step C, to determine independence, first the most significant M-GWAS variant from each chromosome is selected and added to the list of putative causal variants. If the pairwise linkage disequilibrium (LD) between this variant and any other significant variant on the same chromosome has  $r^2 > 0.1$ , these other variants are considered as potentially tagging the same causal variant and are not considered as independent QTL for this cycle. Then, from the remaining significant variants in  $LD\ r^2 \leq 0.1$ , the next most significant is selected on each chromosome, LD is tested between these and the remaining significant variants and so on, until no more significant variants can be identified in this cycle

tion between 16 traits (**V** matrix) were calculated to perform M-GWAS where the multi-trait  $\chi^2$  was calculated using Eq. (2). Then, the  $\chi^2$  values were converted to *P*-values with 16 degrees of freedom.

3. The most significant variant ( $P < 10^{-5}$ ) from M-GWAS (excluding any previously selected most significant variants) was identified on each chromosome. The LD between the most significant variant and every other significant variant on the same chromosome was then tested using the PLINK v1.9 software [29]. If the LD between the most significant variant and any other significant variant that exceeded an  $r^2$  of 0.1, these other variants were considered as potentially tagging the same causal variant and hence were not eligible for selection. The remaining next most significant variant per chromosome (in LD, i.e.  $r^2 \leq 0.1$  with the most significant variant) was then selected, and again LD between this and all remaining variants on the same chromosome was tested to determine if, in this cycle, there were any other remaining significant variants that were independent ( $r^2 \leq 0.1$ ). This selection of the most significant independent QTL variants continued for each chromosome until no further significant variants representing independent QTL were eligible for selection in this cycle.
4. Then, each of the 16 single-trait GWAS was re-run per chromosome while jointly fitting the selected sets of independent significant QTL variants and conditionally re-testing all remaining variants. If there were no significant variants selected on a particular chromosome, then no conditional single-trait GWAS was required for this chromosome for any trait.
5. The newly derived t-values from each conditional single trait GWAS except for the significant variants selected from previous cycle(s) were used to recalculate the M-GWAS chi-squared statistic (Eq. (2)). The original t-values were used for the variants selected from previous cycle(s) that were fitted conditionally in the single-trait GWAS.
6. For the next cycle (and any following), the above steps 3 to 5 were repeated until no more variants were significant in the CM-GWAS.

Having identified all independent significant variants from the CM-GWAS, we then created a secondary expanded list that included all the significant M-GWAS variants ( $P < 10^{-5}$ ) in strong LD ( $r^2 > 0.80$ ) with the selected independent variants. We did this because it is possible that the most significant imputed sequence variant is in strong LD with the actual causal variant but the latter may show a slightly lower significance [13]. The nearest genes to each of these variant positions on the

reference genome OAR3.1 were identified from UCSC Genome Bioinformatics (<http://genome.ucsc.edu/>) and Ensembl ([www.ensembl.org/biomart/](http://www.ensembl.org/biomart/)).

The power of QTL detection in the M-GWAS and the single-trait GWAS was investigated by comparing the FDR in each analysis [28]. Also, Q–Q plots were generated for all traits and the genomic inflation factor (lambda) for each trait was calculated to check if there was any inflation of observed chi-squared statistic due to population structure.

## Results

### Sequence GWAS

#### Single-trait GWAS

GWAS were performed for 16 wool quantity and quality traits (measured as yearling and adults) as well as breech wool cover and breech skin wrinkle scores for up to 7218 pure Merinos that had imputed genotypes for ~30 million WGS variants (Table 1). Estimates of the FDR for significant levels  $P < 10^{-5}$  and  $P < 10^{-6}$  varied between traits (Table 2). At  $P < 10^{-6}$ , ten traits had more than 300 significant variants with an FDR lower than 9.1%, while at  $P < 10^{-5}$  FDR was generally higher, ranging from 9 to 51%, excluding mean fibre diameter at adult age (afd) and breech cover (ebcov). An alternative presentation of the

**Table 2** Number of significant variants (N) ( $P < 10^{-5}$  and  $P < 10^{-6}$ ) and their false discovery rates (FDR, %) for each trait from the single-trait GWAS and multi-trait GWAS

Trait <sup>a</sup>	$P < 10^{-5}$		$P < 10^{-6}$	
	N	FDR	N	FDR <sup>b</sup>
ygfw	1891	15.1	716	4.0
agfw	2164	13.2	910	3.1
ycfw	1145	24.8	606	4.7
acfw	1153	24.8	274	10.4
ysl	2182	13.1	1166	2.4
asl	753	37.9	103	27.7
yfd	1340	21.2	314	9.1
afd	312	91.4	16	
ydcv	3162	9.0	1722	1.7
adcv	1301	21.9	493	5.8
ycuv	682	41.8	145	19.7
acuv	562	50.8	183	15.6
yss	2516	11.3	1547	1.8
ass	1611	17.7	911	3.1
ebwr	1174	24.3	518	5.5
ebcov	354	81.0	14	
M-GWAS	9879	2.9	7431	0.4

N number of significant variants, M-GWAS multi-trait GWAS

<sup>a</sup> Trait names are as defined in Table 1

<sup>b</sup> FDR in empty cells are either not available or higher than 100%

results to the calculation of FDR is a Q–Q plot shown for each trait (see Additional file 1: Fig. S1). In addition, there was no indication of inflation of the test statistic due to population structure because the genomic inflation factor lambda was very close to 1 for all traits (see Additional file 2: Table S1). This is expected because population structure has been captured in our analysis by fitting both a random Merino strain effect and a genomic relationship matrix (see the model used in single-trait GWAS in “Methods” section).

Figure 2 shows the number of significant variants ( $P < 10^{-5}$ ) per chromosome for each trait. Many significant variants (>100) with pleiotropic effects were identified on *Ovis aries* (OAR) chromosome OAR3, 6, 8, 11, 22, 23 and 25. In addition, other clusters of variants that appear to affect only a single trait [e.g. variants on OAR17 for mean fibre diameter at yearling age (yfd)] were found. In a few cases, we observed regions with clusters of highly significant variants that appeared for traits measured at yearling but not for traits measured at adult age and vice versa. For example, over 600 significant variants were found on OAR6 for clean fleece weight at adult age (acfw), but none for clean fleece weight at yearling age (ycfw), whereas significant variants were observed on OAR25 for greasy fleece weight at yearling age (ygf) and ycfw, but not for greasy fleece weight at adult age (agf) and acfw. To demonstrate that this was not due to different numbers of animals being recorded and analysed at yearling and adult ages for greasy fleece weight, we re-analysed only the 4222 overlapping phenotypes available at both ages. A similar pattern of significance regions was still observed (see Additional file 3: Fig. S2).

**Multi-trait GWAS (M-GWAS and CM-GWAS)**

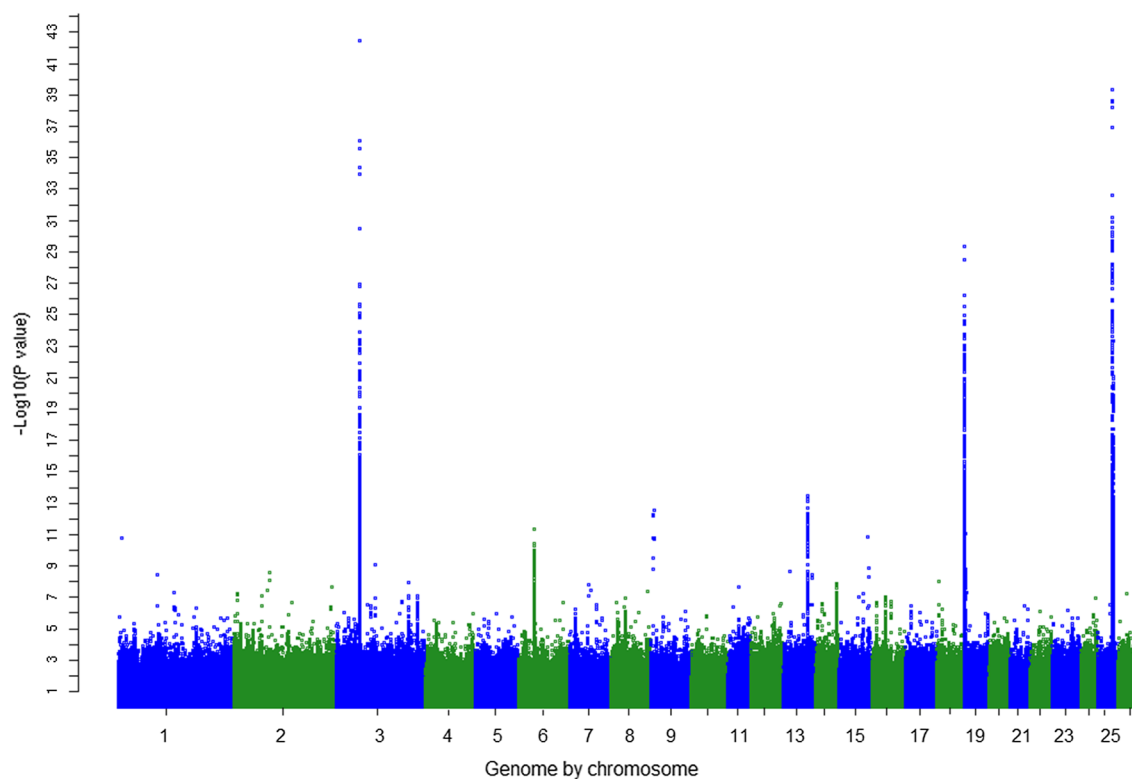
In the M-GWAS that combined the single-trait GWAS results, there were 9879 and 7431 significant variants

at thresholds of  $P < 10^{-5}$  and  $P < 10^{-6}$ , respectively. This corresponded to an FDR of 2.9 and 0.4%, respectively, which was lower than for any individual trait tested in the single-trait GWAS (Table 2). Many highly significant variants from the M-GWAS were found within narrow regions on OAR3, 6, 9, 13, 14, 16, 19 and 25 (Fig. 3). Within these clusters, many of the variants were in strong LD, therefore, to be able to identify independent putative causal variants we extended the published M-GWAS method to include a CM-GWAS. We selected only the most significant (‘top’) variants from each cycle of the CM-GWAS ( $P < 10^{-5}$ ; LD  $r^2 < 0.10$  with other ‘top’ variants) and identified 178 independent putative causal variants across the genome (see Fig. 2 for number per chromosome). These 178 variants were separated from each other by at least 127 kb: the LD regions, variant effects (signed *t*-values >2 or <-2), and the genes closest to these variants are presented in (see Additional file 2: Table S2). In total, 20 independent variants (out 178 variants in Additional file 2: Table S2) were located on OAR3, while only two remained on OAR25. All 178 significant CM-GWAS variants are putatively causal and most of them showed pleiotropic effects on the wool traits as would be expected from this analytical approach.

In addition, the list of 178 significant CM-GWAS variants was expanded to include other significant variants from the standard M-GWAS ( $P < 10^{-5}$ ), which showed strong LD ( $r^2 > 0.80$ ) with these 178 top variants, because the most significant imputed variant may not always be the causal variant. This resulted in 1510 variants and we annotated the nearest genes located within 100 kb on either side of each variant (see Additional file 2: Table S3). Thirty-one of the 178 independent variants were each in strong LD with 5 to 283 other significant variants. The length of these 31 LD regions ranged from 0.86 to 733 kb (see Additional file 2: Table S2). Many of

trait	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	All	
ygf	53	4	16	12	48	170	13	4	11	8	5	4	9	31	4	5	1,156	12	5	8	17	296					1,891	
agf	56	11	41	3	5	1,614	7	7	9	4	114	90	10	2	52	6	7	53	13	1	11	39				1	2,164	
ycf	20	38	6	14				4	29	15	43	46	1	21	1	3	3	12	7	2	3	6	2	868		1	1,145	
acf	181	24	30	22	648	14			3	75	51	4			55	32		12	1							1	1,153	
ysl	12	2	1,769	11	13	6	4	6	11		4		8		1	13		2		3	1	71	17	5	223		2,182	
asl	23	10	184	2			2	284	1			2	13	8	6	8			5			25	5	4	170	1	753	
yfd	34	46	403	1		6	17	152	1		9		97	62	59	169	100		37	3				19	121	4	1,340	
a fd	5	17	4	54		107	7	31		1	5	3	5	15	3	1	7	9	32		2						4	312
ydcv	12	5	207	1	2	3	16	1	5		103		408	243	3		1			20		126	232	1	1,770	3	3,162	
adc	35	4	193		1	16	11	5			31	57	122	29	1	3				4	2	3	1	1	782		1,301	
ycuv	11	10	2			62	23	21		19	87	1	7		19		5		5	6		135	2	159	108		682	
acuv	66	29	8	20		50		19		8	47	8	5	2	3	1	2	2	2			69	199	13	7	2	562	
yss	6	47	10	1	33	30	32	1		1		2	2	4	10		3		90	2			44	8	2,190		2,516	
ass	17	85	7	16			2		1			28		1			2		1	24			4	20	1,403		1,611	
ebwr	63	100	758	5	3	2	5	14	7	25	1	137	1	9	1	5	5			24						6	2	1,174
ebcov	11	37	17	9		13		38	34				7	2		50		108				21		1		6	354	
M-GWAS	40	78	1,907	28	36	602	89	202	275	2	19	35	495	190	25	373	31	16	2,927	9	10	5	13	29	2,440	3	9,879	
CM-GWAS	17	12	20	6	5	9	10	7	1	6	10	7	6	9	3	5	5	5	6	5	2	8	4	2	3	178		

**Fig. 2** Number of significant ( $P < 10^{-5}$ ) variants for individual traits at yearling (y) and adult (a) ages from single-trait GWAS and multi-trait-GWAS and a total number of QTL identified on each of the 26 ovine chromosomes (OAR)



**Fig. 3** Manhattan plot of multi-trait meta-analysis (M-GWAS) before analysis of the conditional multi-trait GWAS (CM-GWAS) model. The  $-\log_{10}(P\text{-value})$  of the multi-trait test on the y axis were calculated using SNP effects from the single-trait GWAS for 16 wool traits and genome positions on the 26 ovine autosomes are on the x axis

the 178 CM-GWAS significant variants differed in some of their pleiotropic trait associations (see Additional file 2: Table S2), which suggests that the putative causal variants influence a range of physiological pathways that modify the expression of these traits.

#### Putative causal variants and candidate genes

The largest number of highly significant variants in the M-GWAS (Fig. 2) was found within narrow regions on OAR3, 6, 13, 19 and 25 (Fig. 3) and (see Additional file 2: Table S2). In each of these regions, the CM-GWAS identified the most significant single independent putative causal variant as detailed below.

##### Variant OAR3:58,729,692

This putative causal variant (CM-GWAS  $P=3.2 \times 10^{-43}$ , MAF=0.06) was located within an intronic region of the *THNSL2* gene (*threonine synthase like 2*) and lies approximately 257 kb upstream of the *FOXI3* gene (58,986,757 to 58,990,671 bp). The alternative allele was associated with almost every trait: increased fleece weight, fibre diameter coefficient of variation, breech wrinkle and breech coverage, as well as decreased staple length,

mean fibre diameter, mean fibre curvature and staple strength. Although this variant was not in very strong LD ( $r^2 > 0.80$ ) with any other significant M-GWAS variants, 627 significant variants (M-GWAS:  $P < 10^{-5}$ ) were in moderate LD ( $0.10 < r^2 < 0.74$ ) with this variant within the 58.5–60.8 Mb region, encompassing the *FOXI3* gene.

##### Variant OAR25:35,305,108

This variant (CM-GWAS  $P=4.8 \times 10^{-40}$ , MAF=0.11) falls within an intron of the *MAT1A* gene (*methionine adenosyltransferase 1A*) and was in strong LD ( $r^2 > 0.80$ ) with 117 other significant variants. One of the variants in strong LD was a missense variant in *MAT1A* and was the third most significant variant ( $2.3 \times 10^{-39}$ ). The alternative allele of OAR25:35,305,108 was associated with increased fleece weight, staple length, mean fibre diameter and staple strength, as well as decreased fibre diameter coefficient of variation and mean fibre curvature. The *t*-values (effect/SE) were consistently larger for the yearling traits than for the same traits measured in adults. This is probably due to larger numbers of animals with yearling age records. There was no significant effect of this variant for breech wrinkle or breech cover. Of the

2440 significant variants on OAR25 in the M-GWAS, only one other variant (OAR25:30,581,246) remained significant ( $P < 10^{-5}$ ) after jointly fitting OAR25:35,305,108 in the CM-GWAS model.

#### **Variant OAR19:840,732**

This is a missense variant (CM-GWAS  $P = 4.4 \times 10^{-30}$ , MAF = 0.12) in the *EGFR* gene (*epithelial growth factor receptor*) and was in strong LD ( $r^2 > 0.8$ ) with 246 other significant variants (M-GWAS:  $P < 10^{-5}$ ). The alternative allele was associated with increased greasy fleece weight and staple length, as well as decreased fibre diameter coefficient of variation and mean fibre curvature. There were no significant effects on clean fleece weight, staple strength, breech wrinkle or breech cover.

#### **Variant OAR13:62,835,771**

This is an intergenic variant (CM-GWAS  $P = 3.3 \times 10^{-14}$ , MAF = 0.12) located 16 kb from the *RALY* (*RALY heterologous nuclear ribonucleoprotein*) gene and 71 kb from the *EIF2S2* (*eukaryotic translation initiation factor 2 subunit beta*) gene. It was in strong LD ( $r^2 > 0.8$ ) with 283 other significant variants (M-GWAS:  $P < 10^{-5}$ ) spanning a region encompassing both the *RALY* and *EIF2S2* genes. The alternative allele was associated with increased fleece weight, staple length, fibre diameter, and fibre diameter coefficient of variation, as well as decreased mean fibre curvature, staple strength, and breech wrinkle.

#### **Variant OAR6:37,676,407**

This is an intergenic variant (CM-GWAS  $P = 4.9 \times 10^{-12}$ , MAF = 0.39) and was in strong LD ( $r^2 > 0.8$ ) with 249 other significant variants (M-GWAS:  $P < 10^{-5}$ ) spanning 0.7 Mb, located between 36.97 and 37.71 Mb. Among the M-GWAS significant variants, 12 variants were within the *LCORL* gene (*ligand dependent nuclear receptor corepressor like*) and 24 in the *NCAPG* gene (*non-SMC condensin I complex subunit G*). Although the variants within these genes included three missense variants, they showed a somewhat lower significance level ( $7.9 \times 10^{-8}$ ) than the OAR6:37,676,407 variant. The latter variant was associated with increased fleece weight and fibre diameter coefficient of variation as well as decreased mean fibre diameter and staple length.

#### **Additional CM-GWAS candidate genes**

While it is not possible to discuss all of the 178 candidate genes and variants (see Additional file 2: Table S2), we highlight several additional variants that showed strong pleiotropy across both fibre and skin wrinkle traits and were identified near or within interesting candidate genes: *ALX4*, *EIF2AK2*, *ESRP1*, *HAS2*, *MC5R* and *MX2* (Table 3).

## **Discussion**

To our knowledge, this is the first study to undertake a large-scale GWAS using whole-genome sequence genotypes for wool traits and skin wrinkle in sheep. Our results fine-mapped 178 significant putative causal variants and candidate genes for wool traits and breech skin wrinkle in independent QTL regions from over 30 million whole-genome sequence variants. Independent QTL regions were identified by applying our CM-GWAS that cycled through multiple rounds of single- and multi-trait GWAS to conditionally identify the most significant variants in the multi-trait GWAS. The selection of the most significant variants was also conditional on a very stringent pairwise LD ( $r^2 < 0.1$ ) with other significant variants. The CM-GWAS could be applied to any species measured for a set of traits. The threshold  $P$  value for the multi-trait will depend on the power of the analysis, and we also recommend a stringent LD threshold to determine QTL independence. We used LD  $r^2$  of 0.1 so that in each cycle there was little risk of selecting more than a single variant from each QTL, particularly for variants that had relatively large effects. Then, if a variant in LD  $r^2 > 0.1$  remained significant in the following cycle after conditional analysis, then the variant could be selected in that cycle.

This relatively conservative approach enabled identification of large and/or highly pleiotropic variant effects for wool and breech wrinkle traits that may be the causal variants or be in very strong LD with causal variants. Imperfect sequence imputation may result in a causal variant having a lower  $P$ -value than one or more other variants in strong LD with it [13]: to mitigate this, we annotated other significant variants that were in strong LD ( $r^2 > 0.8$ ) with the most significant putative causal variant (see Additional file 2: Table S3). Indeed, among the five most significant putative causal variants, four were in strong LD with over 100 other significant variants (Chr6:37,676,407; Chr19:840,732; Chr25:35,305,108; and Chr13:62,835,771). This strong LD spanned over 150 kb suggesting that there has been strong selection across these regions.

The direction of the various trait effects for pleiotropic mutations are of great interest to the wool industry because they may decrease or increase overall wool value. In general, the most desirable wool attributes are increased clean fleece weight (cfw) and staple strength (ss) with decreased fibre diameter (fd), and coefficient of variation in fibre diameter (dcv). In addition, reduced breech cover (ebcov) and reduced breech skin wrinkle (ebwr) would help reduce the incidence of breech flystrike that has a large economic impact on the Merino industry. Among these traits, the strongest economic indicators are for cfw and fd, however there is



**Table 3** List of putative causal variants with the signed significant t-values ( $|t| \geq 2.0$ ) across wool traits at yearling and adult ages

OAR pos. (bp) for most significant SNP	Annotation class for the most significant variant	M-GWAS Pval	ygfw	agfw	ycfw	acfw	ysl	asl	yfd	afd	ydcv	adcw	ycuv	acuv	yss	ass	ebwr	ebcov	Flanking genes <sup>a</sup>
Chr3:58,729,692	Intronic	3.21E-43	4.6	2.9	6.7	3.3	4.6	-10.9	-5.5	-6.7	-3.9	3.3	5.1	-3.2	-2.3	-4.2	6.9	3.5	THNSL2 <sup>0</sup>
Chr25:35,305,108	Intronic	4.85E-40	5.1	4.9	4.1	3.3	4.6	4.6	4.4	2.8	-9.6	-6.8	-4.5	-3	10.4	8.8			MATTIA <sup>0</sup>
Chr19:840,732	Missense	4.44E-30	4.9	4.1	4.1	2	2	2	2	2	-2.5	-2.1	-2.6						EGFR <sup>0</sup>
Chr13:62,835,771	Intergenic	3.30E-14	4	4	2	3.8	4	3.1	4.7	3.4	6.2	4.7	-2.2	-2.2	-2.5	-2.5			RALY <sup>6</sup> EIF2S2 <sup>71</sup>
Chr6:37,676,407	Intergenic	4.87E-12	3.9	6.2	3.6	5.2	-2.6	2.9	6.6	3.8	3.5	2.1							LCORL <sup>224</sup>
Chr15:72,567,959	Intronic	5.92E-09	2	2.3	2.3	2	2.9	2.9	6.6	3.8	5.1	3.7							ALX4 <sup>0</sup>
Chr3:86,858,557	Intronic	1.29E-07	2	2.3	2.3	2	2.9	2.9	6.6	3.8	5.1	3.7							EIF2AK2 <sup>0</sup>
Chr9:82,254,047	Intronic	8.29E-07	2	2.3	2.3	2	2.9	2.9	6.6	3.8	5.1	3.7							ESRP1 <sup>0</sup>
Chr23:43,896,965	Upstream	3.07E-06	2	2.3	2.3	2	2.9	2.9	6.6	3.8	5.1	3.7							MC2R <sup>22</sup>
Chr9:29,973,331	Intergenic	4.59E-06	3.2	5.2	5.2	5.2	3.4	2.8	-3.3	-2.9	5.4	3.4							MCSR <sup>27</sup>
Chr3:58,952,741	Intergenic	3.41E-06	3.2	5.2	5.2	5.2	3.4	2.8	-3.3	-2.9	5.4	3.4							ZHX2 <sup>327</sup>
Chr1:259,689,936	Intronic	1.97E-06	2.2	2	2	2	2	2	3.1	3.1	3.1	3.1							HAS2 <sup>970</sup>

<sup>a</sup> Number in superscript shows the distance to each gene in kb  
<sup>b</sup> MX2 lies very close to LOC101102526, which is a keratin-associated protein 10-12-like gene

an unfavorable positive genetic correlation between *fd* and *cfw* in Merinos ( $\sim 0.3$ ; [30], which suggests that the more common pattern of pleiotropy would be that where effects follow the same direction. Indeed, among the 178 variants, 24 showed pleiotropy for *cfw* and *fd*, and of these, 15 were in the same direction, while only nine showed effects in opposing directions (see Additional file 2: Table S2). Of the 68 variants with an effect on adult *cfw*, for all but one, the alternative (i.e. non-reference) allele was associated with increased fleece weight. This is perhaps somewhat expected given that the OAR3.1 reference genome was obtained from a Texel sheep: a breed with much lower fleece weights than Merinos. There is also an undesirable positive genetic correlation between breech wrinkle and fleece weight, as well as an undesirable negative genetic correlation between wrinkle and fibre diameter [4]. Again, among the 178 most significant CM-GWAS variants, those that decreased wrinkle also decreased fleece weight (and vice versa) and/or increased fibre diameter (and vice versa).

The variant on OAR6 (37,676,407 bp) was in a strong LD region that is better known for associations with body size in a range of mammalian species including humans [11, 31–35] although evidence for the exact mechanism and causal variants remain elusive. This most significant variant in our study was intergenic ( $P = 5 \times 10^{-12}$ ) and was in strong LD with several missense variants (among other significant variants) in the *NCAPG* and *LCORL* genes ( $P = 8 \times 10^{-8}$ ) that have been reported to be located within a signature of selection observed between Merino and Churra sheep [36]. However, it is not clear what effect the genes in this region may have on wool production.

A second variant (OAR19:840,732) showing an extended region of strong LD was a missense mutation in the *EGFR* gene. The protein encoded by *EGFR* is a cell surface protein that binds to epidermal growth factor (EGF) and is well documented as playing a key role in skin and hair follicle development, as reviewed by [26]. Previously, a point mutation in the *EGFR* gene has been demonstrated to be responsible for the mouse “wavy hair” phenotype [37]. Furthermore, mice that harbor a targeted disruption of the *epidermal growth factor receptor* (*EGFR*) allele exhibit a severely disorganized hair follicle phenotype and fuzzy coat [38]. In our study, the missense putative causal variant was associated with fibre curvature, staple length and greasy fleece weight but, interestingly, had no effect on clean fleece weight. This suggests that the increase in greasy fleece weight is either due to some modification of the sebaceous gland output of suint, or that the modified fibre characteristics resulted in the fleece holding more dust and/or suint. Related to the effect of *EGFR* on wool traits was

a different significant variant on OAR11 (11,374,041 bp) located in an intergenic region showing moderate LD and 11 kb from the *PIK3R1* gene that has a GO term: “epidermal growth factor receptor signalling pathway”. This variant had the strongest association with staple length and fibre curvature.

A third putative causal variant (Chr25:35,305,108) in a region of strong LD was in an intron of the *MATIA* gene. This gene codes for a key enzyme (methionine adenosyltransferase 1A) in the pathway that converts methionine to cysteine. These two sulphur amino acids have been shown to be rate limiting for wool production in sheep [39], such that supplementing sheep (per abomasum to bypass the rumen) with these amino acids resulted in increased wool weight, staple strength and fibre diameter. Interestingly, these effects are in keeping with the pleiotropic effects observed for our putative causal variant. Furthermore, the strongest effects of this mutation were observed for the yearling rather than the adult traits, which suggests that this may be due to stronger competition for sulphur amino acids in young growing sheep. There was a missense variant in strong LD ( $r^2 > 0.8$ ) within the same gene, OAR25:35,301,334 (see Additional file 2: Table S3), which lies just 3774 bp away. It is plausible that this missense mutation maybe the true causal variant because it was the third most significant with a *P* value ( $2.3 \times 10^{-39}$ ) very close to that of the most significant variant ( $4.8 \times 10^{-40}$ ). To our knowledge there is only one other report of a QTL for wool traits close to the *MATIA* gene and, in that study, a subset of the same animals as ours was analysed but only with high-density marker genotypes [7].

The fourth highly significant pleiotropic variant in strong LD with many surrounding variants was an intergenic variant on OAR13 (62,835,771 bp) and the LD region spanned variants in and close to the *RALY* and *EIF2S2* genes. While the *RALY* gene has been reported to show a strong signature of selection between sheep breeds [40], it is not documented as having any reported effects on fibre production. However, the *EIF2S2* gene has been reported to play a role in a major mutation that discriminates between the ancestral sheep with a coarse longer hair-like coat from the modern domestic descendants with a finer and shorter wool coat [41]. These authors describe the mutation resulting in the “woolly” coat as an antisense *EIF2S2* retrogene (called *asEIF2S2*) that is inserted into the 3' UTR of the *IRF2BP2* gene. The hybrid mRNA that is transcribed from the disrupted *IRF2BP2* gene was demonstrated to affect the expression of both the *EIF2S2* and *IRF2BP2* mRNA. The pleiotropic effects of our putative causal variant in our study are in keeping with these authors' observations, i.e., increased fleece weight, staple length and fibre diameter

as well as reduced curvature. It is possible that there is a causal variant among the significant variants that we detected, which alters the expression or the transcript of the *EIFS2S* gene. The strong LD across this region is potentially accentuated because *EIFS2S* lies close to a large 149-kb duplication encompassing the *ASIP* gene that results in the dominant white colour of Merinos [42]. Interestingly we found an intronic variant in the related *EIF2AK2* gene which like *EIF2S2* also had the strongest effect on the coefficient of fibre diameter and breech wrinkle (Table 3). Similarly, a variant in the *MX2* gene (Table 3) was very strongly associated with fibre curvature but also breech wrinkle and a number of other traits. The *MX2* gene has a GO term: “regulation of cell cycle” and has been found to be significantly differentially expressed in the skin of super fine Merino sheep compared to a coarse wool sheep breed [43].

The most highly significant variant in our CM-GWAS on OAR3 (58,729,692 bp) showed only moderate to low LD with other significant variants ( $0.10 < r^2 < 0.74$ ) and had a low MAF (0.06), which indicates that it may be a relatively recent mutation. We observed that after fitting this variant in the conditional GWAS all 1644 significant variants in the 57.5–61.2 Mb region on OAR3 (except 11 variants) became no longer significant, and thus this variant seems unlikely to be spurious. It is of course also possible that the imputation in this region is less accurate than average for some reason, and that this has affected the LD pattern in the region. While this mutation was associated with a favourable effect for the sheep industry, i.e. increasing wool weight and decreasing fibre diameter, this is also accompanied by an unfavourable increase in breech wrinkle as well as breech wool cover. Breech wrinkle and breech cover are indicator traits associated with breech flystrike because deep wrinkles (characteristic of some Merinos) combined with wool cover in the breech area increases the risk of urine and faecal staining. This provides a moist environment for flystrike and therefore breech wrinkle and cover traits are used by some breeders to select for resistance to breech flystrike. About 40 variants (out of 178) had a significant effect on breech wrinkle and most of these had an antagonistic effect either with mean fibre diameter and/or fleece weight (e.g. decreased wrinkle and increased fibre diameter). The putative causal variant on OAR3 (58,729,692 bp) lies in an intronic region of the *THNSL2* gene, which to our knowledge has not been previously reported as being associated with wool/fibre characteristics or skin wrinkle. Furthermore, the GO terms associated with this gene and extensive literature searches suggest no obvious role for *THNSL2* in the traits analysed. However, the closest flanking protein coding genes are *FABP1* (*fatty acid binding protein 1*) and *FOXI3* (*forkhead box I3*) with the latter

lying ~250 kb downstream. Previously, a 7-bp duplication in an exon of *FOXI3* (causing a premature stop codon) has been reported to cause a hairless phenotype in some dog breeds [44]. Our putative causal mutation (OAR3: 58,729,692) was associated with much shorter and finer hair, but also with the largest effect observed for skin wrinkle. We also observed another apparently independent putative causal variant closest to the *FOXI3* gene that was also strongly associated with breech wrinkle (58,952,741 bp: Table 3). Although there were no previously documented reports of associations between *FOXI3* and skin wrinkle, the *FOXI3* gene is associated with the GO term “cell differentiation” (GO:0030154). In the Sheep Tissue Atlas [45] (see <https://www.ncbi.nlm.nih.gov/gene/?term=FOXI3++sheep>), the highest tissue expression of *FOXI3* was in skin. The family of *FOX* genes encode transcription factors that are involved in functions such as: cell differentiation and proliferation [46]. Interestingly, another forkhead box protein gene, *FOXN1*, also highly expressed in the skin was reported to cause hypotrichosis (abnormal hair growth) and highly wrinkled skin in Birman cats [47]. The putative causal variant in the *THNSL2* gene also lies within a long non-coding RNA (within this gene), so it could even be possible that this plays some regulatory role on the adjacent *FOXI3* gene. We found two additional significant pleiotropic intergenic variants close to the *FOXK1* and *FOXBI* genes.

The significant variant with an undesirable effect on fibre diameter was in the *ALX4* gene which is known to have a role in hair follicle growth and cycling [48, 49]. Kijas et al. [50] also found strong evidence for a sheep selection signature in the region around the *ALX4* gene. A breech wrinkle variant that was strongly associated with staple length but did not associate with either fleece weight or fibre diameter was located in the *ESRP1* gene (*epithelial splicing regulatory protein 1*). This gene has a GO term “fibroblast growth factor receptor signalling pathway”. This is very interesting because the *FGF5* (*fibroblast growth factor 5*) gene is known to act as an inhibitor of anagen phase of hair cycle that has been demonstrated to have a direct impact on fibre growth across numerous species [51, 52]. The *ESRP1* gene regulates alternative splicing events in epithelial cells of the epidermis, and double knockouts of the *ESRP1* and *ESRP2* in mice impacted hair follicle maturation and epidermal thickness [53]. *ESRP1* is also very highly expressed in sheep skin (Sheep Atlas displayed on NCBI Gene information: [45]). Incidentally, we located a significant CM-GWAS variant at Chr16:31,054,761 that lies close to the *FGF10* gene and another one (Chr10:35,310,820) in a gene desert, but near *FGF9*, and both were associated with staple length as well as some other traits. We

found two significant variants that reduced breech cover while not having very strong effects on other traits: one (Chr1:197,166,904) in the *LPP* gene (*LIM domain containing preferred translocation partner in lipoma*) and the second (Chr12:71,641,048) was just upstream of the *IRF6* (*interferon regulatory factor*) gene. The *IRF6* gene is very highly expressed in skin and is linked to the GO terms: keratinocyte proliferation and differentiation, cell cycle arrest. These findings may be useful for future research to help reduce breech wrinkle and wool cover and to lower the incidence of flystrike.

A very interesting significant variant that was associated with reduced breech wrinkle, increased fleece weight and staple length, but was not associated with fibre diameter (Table 3) and lies in a gene desert on Chr9 (29,973,331 bp) downstream from the *ZHX2* gene (*zinc fingers and homeoboxes 2*) and upstream of the *HAS2* gene (*hyaluronan synthase 2*; see Table 3). In the Shar-Pei dog breed a large ~16-kb duplication approximately 0.35 Mb upstream of the *HAS2* gene has been found to enhance the expression of the *HAS2* gene causing the characteristic thick skin folds particularly around the head of this breed [54].

There are few GWAS reports on sheep wool traits for us to compare our results to. In a single-trait GWAS with 50k SNP array genotypes in Chinese Merino sheep, Wang et al. [55] identified 28 SNPs located within 12 genes that affect fibre diameter, fibre diameter coefficient of variance, and crimp. However, we did not find any significant variants ( $P < 10^{-5}$ ) within these genes, which is perhaps not surprising given the low density of their genotypes and smaller number of animals studied. The only other report was our own study using a subset of the animals in this study but using only a high-density SNP array, therefore we have not included a comparison of those previous results here.

It is of interest to understand if most candidate genes associated with wool traits in sheep have already been identified in other mammalian studies. Thus, we extended our investigation further by compiling a set of genes (Published-Gene set) that were previously reported in the literature as associated with wool, fur, fibre or hair characteristics across humans, mice and other mammals (see Additional file 2: Table S4), and fine-mapped and estimated the variance explained by a subset of sequence variants in and around several hundred genes associated with hair, fur or wool fibre characteristics across different mammals (for details see Additional file 4). Interestingly, the Published-Gene set explained 6 to 17% more genomic variance than randomly selected sets (Random-Gene sets) for the traits related with fibre quality characteristics, but there was no clear difference between Random-Gene and Published-Gene sets for fleece

weight, breech wrinkle or breech cover traits (see Additional file 4). Although there was some overlap between the CM-GWAS candidate genes and the genes previously reported in the literature as associated with mammalian fibre traits, many were not included in this gene list, which indicates that our study brings new knowledge to this field.

## Conclusions

To our knowledge, this is the first multi-trait meta GWAS study of wool traits using imputed sequence data in sheep. Our findings confirm the high polygenic and pleiotropic nature of the variants that affect wool traits as well as breech wrinkle and cover. Our CM-GWAS meta-analysis enabled the detection of 178 independent QTL regions and identified putative candidate variants and genes that may affect these traits such as a missense variant in *EGFR* (a gene known to affect fur growth in mammals). Novel candidate genes identified here included the *MATIA* gene that encodes an enzyme in the sulphur metabolism pathway critical to production of wool proteins in high-producing Merino sheep, and the *ESRPI* gene. We also discovered a significant breech wrinkle variant upstream of the *HAS2* gene, which is known to cause the exaggerated skin folds of the Shar-Pei dog breed. These findings and additional functional annotation will further enhance our ability to exploit sequence mutations to improve wool and welfare traits in sheep.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12711-021-00651-0>.

**Additional file 1: Figure S1.** Quantile–quantile plot of P-values from single-SNP genome wide association study (GWAS) for each of the 16 traits studied (dark orange) and from multi-trait meta GWAS (dark magenta). Observed and expected P-values would fall on the light blue line if there was no association.

**Additional file 2: Table S1.** Inflation factor ( $\lambda$ ) for each trait. The inflation of observed chi-squared statistic due to population structure was checked. **Table S2.** Annotation of pleiotropic effects on individual traits for 178 putative causal variants identified from the multi-trait conditional GWAS. The annotation includes positions, functional classification, rs ID where available and signed significant t-values ( $|t| > 1.96$ ) were provided for 178 variants across the 16 traits at two ages. Gene names and distances from the nearest genes are also provided if variants are located in or within 100 kb of the gene. **Table S3.** Annotation of 1510 sequence variants in strong LD with the 178 most significant variants from multi-trait conditional GWAS. The annotation includes the positions, rs ID where available, functional classification and P values of multi-trait  $\chi^2$  statistic. The names of the nearest genes are also provided if variants are located in or within 100 kb of the gene. **Table S4.** List of 453 published genes previously associated with fibre characteristics in mammals (“Published-Gene set”) used to identify sequence variants to be included in BayesR and GBLUP analysis. Ensembl identification, name of genes and gene location (chromosome number, start and end position in bp).

**Additional file 3: Figure S2.** Manhattan plot of single-trait GWAS for greasy fleece weight at yearling (a) and adult (b) ages using all animals and at yearling (c) and adult (d) ages using the same animals. The red points represent significant variants at  $P < 10^{-6}$ .

**Additional file 4.** Analyses of variants located around previously reported wool and hair genes. A set of variants in and close to 453 previously reported genes associated with wool/fur/hair traits across mammals (Published-Gene set) were identified including 1619 coding variants and 11,712 variants up- and down-stream of the genes. A BayesR fine-mapping analysis identified which of these variants were most associated with wool traits and a GBLUP tested the proportion of genetic variance that the sequence variant set explained for each of the wool traits.

### Acknowledgements

The authors gratefully acknowledge funding or use of data from the Cooperative Research Centre for Sheep Industry Innovation, SheepGENOMICS project, Sheep Genetics, Australian Wool Innovation Ltd, and FarmIQ. The authors would like to extend their gratitude to Klint Gore (University of New England, Armidale, NSW, 2351, Australia) for preparing and cleaning genotype data and managing the CRC information nucleus database, and Amanda J. Chamberlain, Brett Mason, Coralie M. Reich, and Claire P. Prowse-Wilkins for generating and processing sequence and genotype data, and all staff involved at the Sheep CRC sites and SG sites across Australia. The authors thank James Kijas, Rudiger Brauning, Alan McCulloch, and Sean McWilliams for their contribution to SheepGenomesDB (<https://sheepgenomesdb.org/>) and all institutions that have made their sheep WGS data public. The authors also acknowledge constructive comments made by anonymous reviewers to improve the manuscript.

### Authors' contributions

SB, IMM, HDD, and JHJvdW conceived the study. SB performed all analyses. SB, AAS, MK, NM and ND were involved in preparing pedigree, phenotypes and genotypes. SB and IMM were involved in processing and imputing sequence data and wrote the manuscript. PS generated the annotations for the sequence variants. IMM, SB, HDD, JHJvdW and AAS contributed to the design of the study. All authors read and approved the final manuscript.

### Funding

This work was funded by the Australian Sheep Cooperative Research Centre (AU) (Grant No. Program 3).

### Availability of data and materials

All the SheepGenomesDB Run2 sequence data used for imputation to sequence is freely available. A list of the 935 SRA Sample ID here: Brauning R, McWilliam S, Daetwyler H, McCulloch A, Clarke S. SheepGenomesDB run2—list of animals. figshare. Dataset. 2020. <https://doi.org/10.6084/m9.figshare.11530665.v1>.

### Declarations

#### Ethics approval and consent to participate

No ethical approval was required for this study.

#### Consent for publication

All authors have given consent to publish.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, VIC 3083, Australia. <sup>2</sup>Cooperative Research Centre for Sheep Industry Innovation, Armidale, NSW 2351, Australia. <sup>3</sup>Animal Genetics and Breeding Unit, University of New England, Armidale, NSW 2351, Australia. <sup>4</sup>Faculty of Agricultural, Life & Environmental Sciences, University of Alberta, Edmonton, AB T6G 2R3, Canada. <sup>5</sup>School of Environmental and Rural Science, University of New England, Armidale, NSW 2351, Australia. <sup>6</sup>Hendrix Genetics, Boxmeer, The Netherlands.

<sup>7</sup>School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3086, Australia.

Received: 29 January 2021 Accepted: 29 June 2021

Published online: 08 July 2021

### References

- Nolan E. The economic value of wool attributes phase 2. Report prepared for Australian Wool Innovation Limited. Sydney: The University of Sydney School of Economics; 2014.
- Safari E, Fogarty NM, Gilmour AR, Atkins KD, Mortimer SI, Swan AA, et al. Genetic correlations among and between wool, growth and reproduction traits in Merino sheep. *J Anim Breed Genet*. 2007;124:65–72.
- Swan AA, Purvis IW, Piper LR. Genetic parameters for yearling wool production, wool quality and bodyweight traits in fine wool Merino sheep. *Aust J Exp Agric*. 2008;48:1168–76.
- Hatcher S, Preston JW. Genetic relationships of breech cover, wrinkle and wool coverage scores with key production traits in Australian Merino sheep. *Small Ruminant Res*. 2018;164:48–57.
- GHD Pty Ltd, Lane J, Jubb T, Shephard R, Webb-Ware J, Fordyce G. Priority list of endemic diseases for the red meat industries. North Sydney: Meat & Livestock Australia Limited; 2015.
- Wang ZH, Yang H, Wang S, Rong E, Pei W, Li H, et al. Genome-wide association study for wool production traits in a Chinese Merino sheep population. *PLoS One*. 2014;9:e107101.
- Bolormaa S, Swan AA, Brown DJ, Hatcher S, Moghaddar N, van der Werf JH, et al. Multiple-trait QTL mapping and genomic prediction for wool traits in sheep. *Genet Sel Evol*. 2017;49:62.
- Moghaddar N, Khansefid M, van der Werf JHJ, Bolormaa S, Duijvesteijn N, Clark SA, et al. Genomic prediction based on selected variants from imputed whole-genome sequence data in Australian sheep populations. *Genet Sel Evol*. 2019;51:72.
- Moghaddar N, MacLeod IM, Duijvesteijn N, Bolormaa S, Khansefid M, Al-Mamun H, et al. Genomic evaluation based on selected variants from imputed whole-genome sequence data in Australian sheep populations. In: Proceedings of the 11th World Congress on Genetics Applied to Livestock Production: 11–16 February 2018; Auckland; 2018.
- Bolormaa S, Pryce JE, Reverter A, Zhang Y, Barendse W, Kemper K, et al. A multi-trait, meta-analysis for detecting pleiotropic polymorphisms for stature, fatness and reproduction in beef cattle. *PLoS Genet*. 2014;10:e1004198.
- Bolormaa S, Hayes BJ, van der Werf JHJ, Pethick D, Goddard ME, Daetwyler HD. Detailed phenotyping identifies genes with pleiotropic effects on body composition. *BMC Genomics*. 2016;17:224.
- Xiang R, van den Berg I, MacLeod IM, Hayes BJ, Prowse-Wilkins CP, Wang M, et al. Quantifying the contribution of sequence variants with regulatory and evolutionary significance to 34 bovine complex traits. *Proc Natl Acad Sci USA*. 2019;116:19398–408.
- Pausch H, MacLeod IM, Fries R, Emmerling R, Bowman PJ, Daetwyler HD, et al. Evaluation of the accuracy of imputed sequence variant genotypes and their utility for causal variant detection in cattle. *Genet Sel Evol*. 2017;49:24.
- Kemper KE, Reich CM, Bowman PJ, vander Jagt CV, Chamberlain AJ, Mason BA, et al. Improved precision of QTL mapping using a nonlinear Bayesian method in a multi-breed population leads to greater accuracy of across-breed genomic predictions. *Genet Sel Evol*. 2015;47:29.
- Erbe M, Hayes BJ, Matukumalli LK, Goswami S, Bowman PJ, Reich M, et al. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *J Dairy Sci*. 2012;95:4114–29.
- van der Werf JHJ, Kinghorn BP, Banks RG. Design and role of an information nucleus in sheep breeding programs. *Anim Prod Sci*. 2010;50:998–1003.
- White JD, Allingham PG, Gorman CM, Emery DL, Hynd P, Owens J, et al. Design and phenotyping procedures for recording wool, skin, parasite resistance, growth, carcass yield and quality traits of the SheepGENOMICS mapping flock. *Anim Prod Sci*. 2012;52:157–71.

18. Swan AA, Brown DJ, van der Werf JHJ. Genetic variation within and between sub-populations of the Australian Merino breed. *Anim Prod Sci*. 2014;56:87–94.
19. Brown DJ, Swan AA, Boerner V, Li L, Gurman PM, McMillan AJ, et al. Single-step genetic evaluations in the Australian sheep industry. In: *Proceedings of the 11th World Congress on Genetics Applied to Livestock Production: 11–16 February 2018; Auckland*; 2018.
20. Hatcher S, Hynd PI, Thornberry KJ, Gabb S. Can we breed Merino sheep with softer, whiter, more photostable wool? *Anim Prod Sci*. 2010;50:1089–97.
21. Bolormaa S, Chamberlain AJ, Khansefid M, Stothard P, Swan AA, Mason B, et al. Accuracy of imputation to whole-genome sequence in sheep. *Genet Sel Evol*. 2019;51:1.
22. Daetwyler HD, Swan AA, van der Werf JHJ, Hayes BJ. Accuracy of pedigree and genomic predictions of carcass and novel meat quality traits in multi-breed sheep data assessed by cross-validation. *Genet Sel Evol*. 2012;44:33.
23. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet*. 2012;44:955–9.
24. Loh PR, Danecek P, Palamara PF, Fuchsberger C, Reshef YA, Finucane HK, et al. Reference-based phasing using the haplotype reference consortium panel. *Nat Genet*. 2016;48:1443–8.
25. Daetwyler HD, Brauning R, Chamberlain AJ, McWilliam S, McCulloch A, Vander Jagt CJ, et al. 1000 bull genomes and sheep genomes projects: enabling cost-effective sequence level analysis globally. In: *Proceedings of the 22nd Association for the Advancement of Animal Breeding and Genetics Conference: 2–5 July 2017; Townsville*; 2017.
26. Grant JR, Arantes AS, Liao X, Stothard P. In-depth annotation of SNPs arising from resequencing projects using NGS-SNP. *Bioinformatics*. 2011;27:2300–1.
27. Meyer K. “WOMBAT—a tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *J Zhejiang Univ Sci B*. 2007;8:815–21.
28. Bolormaa S, Hayes BJ, Savin K, Hawken R, Barendse W, Arthur PF, et al. Genome wide association studies for feedlot and growth traits in cattle. *J Anim Sci*. 2011;89:1684–97.
29. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–75.
30. Huisman AE, Brown DJ, Ball AJ, Graser HU. Genetic parameters for body-weight, wool, and disease resistance and reproduction traits in Merino sheep. 1. Description of traits, model comparison, variance components and their ratios. *Aust J Exp Agric*. 2008;48:1177–85.
31. Tetens J, Widmann P, Kühn C, Thaller G. A genome-wide association study indicates LCORL/NCAPG as a candidate locus for withers height in German Warmblood horses. *Anim Genet*. 2013;44:467–71.
32. Setoguchi K, Watanabe T, Weikard R, Albrecht E, Kühn C, Kinoshita A, et al. The SNP c.1326T>G in the non-SMC condensin I complex, subunit G (NCAPG) gene encoding a p.Ile442Met variant is associated with an increase in body frame size at puberty in cattle. *Anim Genet*. 2011;42:650–5.
33. Yurchenko AA, Denisovskaya TE, Yudin NS, Dotsev AV, Khamiruev TN, Selionova MI, Egorov SV, et al. High-density genotyping reveals signatures of selection related to acclimation and economically important traits in 15 local sheep breeds from Russia. *BMC Genomics*. 2019;20:294.
34. Bouwman AC, Daetwyler HD, Chamberlain AJ, Ponce CH, Sargolzaei M, Schenkel FS, et al. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nat Genet*. 2018;50:362–7.
35. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet*. 2014;46:1173–86.
36. Gutiérrez-Gil B, Esteban-Blanco C, Wiener P, Chitneedi PK, Suarez-Vega A, Arranz JJ. High-resolution analysis of selection sweeps identified between fine-wool Merino and coarse-wool Churra sheep breeds. *Genet Sel Evol*. 2017;49:81.
37. Luetke NC, Phillips HK, Qiu TH, Copeland NG, Earp HS, Jenkins NA, et al. The mouse waved-2 phenotype results from a point mutation in the EGF receptor tyrosine kinase. *Genes Dev*. 1994;8:399–413.
38. Hansen LA, Alexander N, Hogan ME, Sundberg JP, Dlugosz A, Threadgill D, et al. Genetically null mice reveal a central role for epidermal growth factor receptor in the differentiation of the hair follicle and normal hair development. *Am J Pathol*. 1997;50:1959–75.
39. Reis PJ, Schinkel PG. The growth and composition of wool II. The effect of casein, gelatin, and sulphur-containing amino acids given per abomasum. *Aust J Biol Sci*. 1964;17:532–47.
40. Li X, Yang J, Shen M, Xie XL, Liu GJ, Xu YX, et al. Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. *Nat Commun*. 2020;11:2815.
41. Demars J, Cano M, Drouilhet L, Plisson-Petit F, Bardou P, Fabre S, et al. Genome-wide identification of the mutation underlying fleece variation and discriminating ancestral hairy species from modern woolly sheep. *Mol Biol Evol*. 2017;34:1722–9.
42. Norris BJ, Whan VA. A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. *Genome Res*. 2008;18:1282–93.
43. Zhang L, Sun F, Jin H, Dalrymple BP, Cao Y, Wei T, et al. A comparison of transcriptomic patterns measured in the skin of Chinese fine and coarse wool sheep breeds. *Sci Rep*. 2017;7:14301.
44. Drögemüller C, Karlsson EK, Hytönen MK, Perloski M, Dolf G, Sainio K, et al. A mutation in hairless dog implicates *FOXI3* in ectodermal development. *Science*. 2008;321:1462.
45. Jiang Y, Xie M, Chen W, Talbot R, Laddox JF, Faraut T, et al. The sheep genome illuminates biology of the rumen and lipid metabolism. *Science*. 2014;344:1168–73.
46. Hannehalli S, Kaestner KH. The evolution of Fox genes and their role in development and disease. *Nat Rev Genet*. 2009;10:233–40.
47. Abitbol M, Bossé P, Thomas A, Turet L. A deletion in *FOXN1* is associated with a syndrome characterized by congenital hypotrichosis and short life expectancy in Birman cats. *PLoS One*. 2015;10:e0120668.
48. Duverger O, Morasso MI. To grow or not to grow: hair morphogenesis and human genetic hair disorders. *Semin Cell Dev Biol*. 2014;25–26:22–33.
49. Kayserili H, Uz E, Niessen C, Vargel I, Alanay Y, Tuncbilek G, et al. ALX4 dys-function disrupts craniofacial and epidermal development. *Hum Mol Genet*. 2009;18:4357–66.
50. Kijas JW, Lenstra J, Hayes BJ, Boitard S, Porto Neto L, San Cristobal M, et al. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol*. 2012;10:e1001258.
51. Higgins CA, Petukhova L, Harel S, Ho YY, Drill E, Shapiro L, et al. FGF5 is a crucial regulator of hair length in humans. *Proc Natl Acad Sci USA*. 2014;111:10648–53.
52. Li WR, Liu CX, Zhang XM, Chen L, Peng XR, He SG, et al. CRISPR/Cas9-mediated loss of *FGF5* function increases wool staple length in sheep. *FEBS J*. 2017;284:2764–73.
53. Bebee TB, Park JW, Sheridan KI, Warzecha CC, Cieply BW, Rohacek AM, et al. The splicing regulators *Esrp1* and *Esrp2* direct an epithelial splicing program essential for mammalian development. *Elife*. 2015;4:e08954.
54. Olsson M, Meadows JRS, Truvé K, Pielberg GR, Puppo F, Maucci E, et al. A novel unstable duplication upstream of *HAS2* predisposes to a breed-defining skin phenotype and a periodic fever syndrome in Chinese Shar-Pei dogs. *PLoS Genet*. 2011;7:e1001332.
55. Wang Z, Zhang H, Yang H, Wang S, Rong E, Pei W, et al. Genome-wide association study for wool production traits in a Chinese Merino sheep population. *PLoS One*. 2014;9:e107101.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.