

RESEARCH ARTICLE

Open Access

# Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland

Brian K Meredith<sup>1,5</sup>, Francis J Kearney<sup>3</sup>, Emma K Finlay<sup>4</sup>, Daniel G Bradley<sup>4</sup>, Alan G Fahey<sup>5</sup>, Donagh P Berry<sup>2</sup> and David J Lynn<sup>1\*</sup>

## Abstract

**Background:** Contemporary dairy breeding goals have broadened to include, along with milk production traits, a number of non-production-related traits in an effort to improve the overall functionality of the dairy cow. Increased indirect selection for resistance to mastitis, one of the most important production-related diseases in the dairy sector, via selection for reduced somatic cell count has been part of these broadened goals. A number of genome-wide association studies have identified genetic variants associated with milk production traits and mastitis resistance, however the majority of these studies have been based on animals which were predominantly kept in confinement and fed a concentrate-based diet (i.e. high-input production systems). This genome-wide association study aims to detect associations using genotypic and phenotypic data from Irish Holstein-Friesian cattle fed predominantly grazed grass in a pasture-based production system (low-input).

**Results:** Significant associations were detected for milk yield, fat yield, protein yield, fat percentage, protein percentage and somatic cell score using separate single-locus, frequentist and multi-locus, Bayesian approaches. These associations were detected using two separate populations of Holstein-Friesian sires and cows. In total, 1,529 and 37 associations were detected in the sires using a single SNP regression and a Bayesian method, respectively. There were 103 associations in common between the sires and cows across all the traits. As well as detecting associations within known QTL regions, a number of novel associations were detected; the most notable of these was a region of chromosome 13 associated with milk yield in the population of Holstein-Friesian sires.

**Conclusions:** A total of 276 of novel SNPs were detected in the sires using a single SNP regression approach. Although obvious candidate genes may not be initially forthcoming, this study provides a preliminary framework upon which to identify the causal mechanisms underlying the various milk production traits and somatic cell score. Consequently this will deepen our understanding of how these traits are expressed.

## Background

Dairy production is an economically important sector of global agriculture with the top 10 leading dairy companies turning over in excess of \$114 billion in 2009 [1]. Dairy cows account for 84% of global dairy output [1] so consequently there is great interest placed upon the production potential and health of these animals. Until recently, the majority of international dairy breeding programmes selected solely for increased milk production, however,

breeding goals have diversified to include health and functional traits in an effort to minimise and reverse the decline in these traits [2]. Prominent among these health-related traits is mastitis (commonly measured using somatic cell score (SCS) as an indicator trait), which is one of the most important and costly production diseases in the dairy industry. Selection for improved milk production traits and reduced SCS (indicating increased mastitis resistance) can potentially be improved through the identification of quantitative trait loci (QTL) associated with these traits of interest by allowing geneticists to infer and comprehend the genetic and molecular mechanisms underlying the traits.

\* Correspondence: [david.lynn@teagasc.ie](mailto:david.lynn@teagasc.ie)

<sup>1</sup>Animal and Bioscience Research Department, Teagasc, Animal & Grassland Research and Innovation Centre, Grange, Dunsany, Co. Meath, Ireland  
Full list of author information is available at the end of the article

QTL associated with milk production and SCS have been extensively reviewed by, amongst others, Khatkar et al. [3] and Smaragdov et al. [4]. Milk production QTL have been reported most often on chromosomes 1, 3, 6, 10, 14 and 20 while SCS QTL were most frequently observed on chromosomes 5, 8, 11, 18 and 23. Many of the reviewed studies utilised family-based studies and microsatellite markers to identify areas of the genome associated with a particular trait, however, these studies were often limited by the relatively low number of genetic markers sparsely distributed across the genome.

The sequencing of the bovine genome and the subsequent HapMap project made large amounts of genetic markers available in the form of single nucleotide polymorphisms (SNPs). This massive increase in marker numbers allied with the emergence of high-throughput genotyping technologies allowed routine genome-wide association studies (GWAS) to be performed in cattle populations. GWAS allow screening of the genome utilising a large number of genetic markers spread across the entire genome to detect genetic variants associated with a particular disease or trait. The majority of recent GWAS employ SNPs as genetic markers. SNPs may not themselves be responsible for the variation observed in a trait, however, due to their close proximity to un-genotyped causal variants they have been co-inherited and so can act as proxies for the unknown causal variants [5]. In this way, SNPs significantly associated with a disease or trait may indicate a region of the genome which harbours genetic variants influencing the expression of that disease or trait. In general, GWAS studies act as an initial screening tool, from which significantly associated regions can be further refined using a higher marker density (potentially by re-sequencing), with the ultimate goal of identifying candidate genes believed to underlie the trait(s) of interest. Candidate genes can then be characterised further in an attempt to identify the functional mechanisms underlying a trait. This will lead to a greater understanding of molecular basis and regulation of the trait in question. A number of recent studies in dairy cattle have detected associations with production and functional traits using a GWAS approach [6-8].

In general, GWAS studies to date have identified important genomic regions using single locus association methods whereby each individual SNP is tested for association with the trait of interest. This method has proved useful in its straight-forward implementation and interpretation, however, it has been hampered by the large number of false positive results produced due to the large number of markers and hence individual statistical tests. This problem has been somewhat addressed through multiple correction techniques such as the false discovery rate (FDR) [9], yet markers with small effects are liable to be lost in this manner. Multi-locus and

validation-based approaches stand as possible alternatives to tackle the amount of false positives produced from a GWAS. An example of a validation-based approach is a two-stage GWAS where an initial genome-wide scan is performed in a large group of animals to identify a subset of significant SNP. This is then replicated in an independent population of animals to validate the significant associations. Such an approach has been used in a GWAS investigating bull fertility [10].

Irrespective of the statistical approach or study design used, most QTL studies to-date in cattle have largely been undertaken using phenotypic data originating from high-input, concentrate-based dairy production systems. Genotype by environment (GxE) interactions have, however, been reported between high- (i.e. animals predominantly kept in confinement and fed a concentrate-based diet) and low-input production (i.e. animals fed a predominantly forage-based diet in a pasture-based production system) systems [11-14]. In general, interaction effects tend to be scaling effects where animals retain the same ranking across different environments, however, re-ranking of bulls between pasture and total mixed ration feeding systems has been reported [15]. Indeed, Interbull [16] issues separate sets of results of its Multiple Across Country Evaluation to each participating country due to potential re-ranking of sires via GxE interactions. Consequently, we hypothesised that some QTL for milk production traits and/or SCS may vary across may vary across different production systems.

The objective of this study was to identify regions of the genome associated with milk production traits and SCS in cattle fed a basal diet of grazed grass (low-input system) using single SNP regression in a dataset of 914 Holstein-Friesian sires. A multi-locus, Bayesian approach in the same population and a single SNP regression analysis in a separate population of Holstein-Friesian cows were conducted to provide further support to the associations.

## Results

### Significant associations

Two populations of 914 Holstein-Friesian AI sires and 493 Holstein-Friesian cows were used to quantify associations between genotypic and phenotypic data. A summary of the various phenotypes and the correlations between them in the sire dataset are detailed in Table 1 and Table 2 respectively. The corresponding information for the cows is detailed in Additional file 1 and Additional file 2 respectively. A set of haplotype blocks was defined using the 40,668 SNPs from the sires. Using these SNPs, 10,958 haplotype blocks were defined which accounted for ~1 Gb of the genome. On average, the blocks were 97,006 Kb in length and contained 2.84 SNPs; the majority of SNPs ( $n = 31,157$  SNPs) were located within haplotype blocks. The remaining 9,511

**Table 1 Summary statistics for the phenotypic data in the sires**

	Trait	N	Mean	$\sigma$
	Milk Yield (kg)	914	158.3	231.2
	Fat Yield (kg)	914	6.1	7.6
Sires	Protein Yield (kg)	914	5.8	6.6
	Fat Percentage (kg) ( $\times 1000$ )	914	4.9	142.3
	Protein Percentage (kg) ( $\times 1000$ )	914	-0.96	82.7
	Somatic Cell Score ( $\log_{10}$ SCC) ( $\times 1000$ )	773	34.2	104.9

Summary statistics include the total number of phenotypic records (N), mean and standard deviation ( $\sigma$ ) for each trait in the sires. Phenotypes in the sires are expressed as daughter yield deviations on a PTA scale

SNPs were not located within any haplotype block. Two statistical approaches, a frequentist single SNP regression approach and a multi-locus Bayesian method, were used to detect significant associations with milk, fat and protein yield, milk fat and protein percentage and somatic cell score (SCS). In total, 1,529 and 37 associations were detected in the sires using a single SNP regression and a Bayesian method respectively. Of these, 276 associations were found to be novel (i.e. had a q-value  $\leq 0.05$  and were located outside any known QTL regions for the particular trait) in the sires using the single SNP regression approach. There were 103 associations in common between the sires and cows across all the traits. The results of the Bayesian analysis were tested for robustness to the prior probability of a SNP being associated with the phenotype. Only SNPs with a posterior probability of association  $\geq 0.8$  for two out of the three priors were considered significant. The results using different priors were generally consistent within a trait with the same SNP or a SNP in the same region of the genome indicated as significant across all priors. This Bayesian approach allowed a multi-locus statistical approach while also providing the opportunity to validate any significant associations found using the single SNP regression approach in the same population of sires (figure 1). Indeed numerous SNPs were detected as significant using the Bayesian approach which were also significant using the single SNP regression.

The use of a single SNP regression approach in a population of 493 Holstein-Friesian cows endeavoured

to replicate any previous associations in a semi-independent population. The average additive relationship among the sires was 0.046, was 0.040 among the cows, and was 0.038 between the sires and cows. Significant associations detected for the two statistical approaches in the sires and the single SNP regression in the cows are detailed for each trait in Additional file 3.

A number of SNPs were significantly associated with more than one trait. We combined all significant associations (i.e. from both sire statistical models and the cow validation dataset) detected for a trait and identified 232, 83, 52 and 3 SNPs which were significantly associated with 2, 3, 4 and 5 traits, respectively (Additional file 3).

### Milk yield

Using the single SNP regression approach, 370 SNPs were identified as being significantly associated ( $q \leq 0.05$ ) with milk yield at the 5% FDR in the sire population. The majority of these SNPs (311/370) were located in known QTL regions for milk yield [17-20]. Several genomic regions contain clusters of SNPs associated with the trait, including several on chromosome 14 in close proximity to the DGAT1 gene, which is known to affect milk yield [21-23]. Indeed, the top-ranked SNP for milk yield ( $q = 8.33 \times 10^{-12}$ ) was located 146 base pairs from the DGAT1. A cluster of SNPs on chromosome 20, in a region also known to be associated with milk yield [18,19,24], was also identified. Notably, 59 SNPs, which do not appear to be located within known QTL regions for milk yield, were detected on nine different chromosomes (Additional file 3). However, none of these novel SNPs were significant using the Bayesian model or in the cows. Of potential interest was a cluster of significant SNPs on chromosome 13 from ~45-49 Mb which contained numerous SNPs of moderate association (lowest q-value =  $5.14 \times 10^{-6}$ ). The genes in this region are listed in Additional file 3.

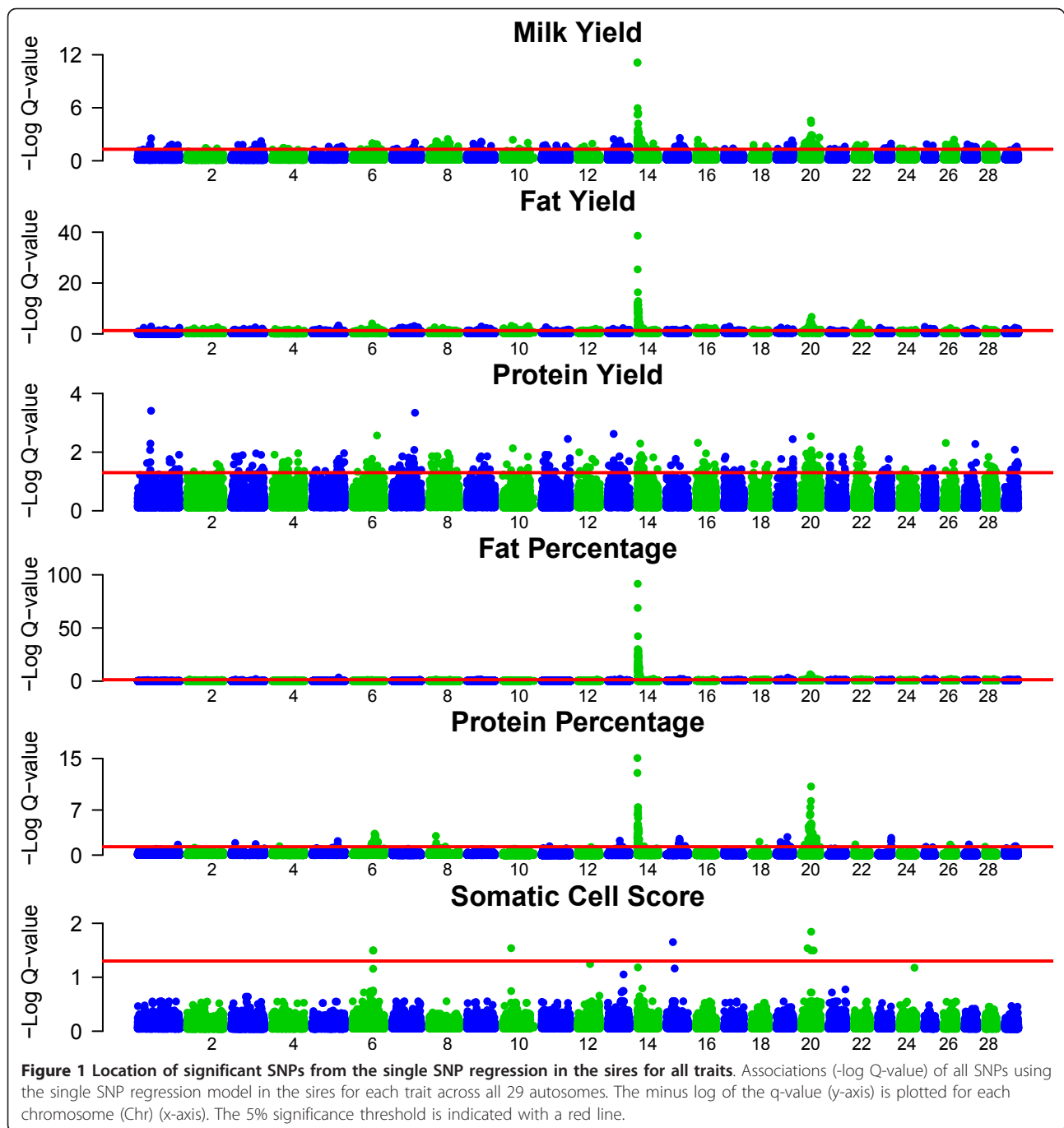
There were two SNPs significantly associated with milk yield using the Bayesian approach in the sires (Additional file 3). These two SNPs, rs42211964 located ~27 Mb on chromosome 8 and rs109421300 located ~0.4 Mb on

**Table 2 Pearson correlations between phenotypes for the sires**

Trait	Milk Yield	Fat Yield	Protein Yield	Fat %	Protein %	SCS
Milk Yield		0.57	0.87	-0.60	-0.70	0.12
Fat Yield	0.57		0.75	0.34	-0.02	0.16
Protein Yield	0.87	0.75		-0.25	-0.23	0.17
Fat %	-0.60	0.34	-0.25		0.76	0.01
Protein %	-0.70	-0.02	-0.23	0.76		0.01
SCS	0.12	0.16	0.17	0.01	0.01	

Fat % = fat percentage; Protein % = protein percentage; SCS = somatic cell score

Correlations between phenotypes in the sires expressed as daughter yield deviations on a PTA scale



chromosome 14 were also significantly associated with milk yield using the single SNP regression in the sires. Both of these occurred in known QTL regions [17,19] with the latter in close proximity to the DGAT1 gene known to affect milk yield in dairy cattle [21].

#### Fat yield

A total of 370 SNPs were significantly associated ( $q \leq 0.05$ ) with fat yield using the single SNP regression in the sires

(Additional file 3). As with milk yield, the majority (281/370) of these were located in known QTL regions for fat yield [22,25-27]. Additionally, the most significantly associated SNPs were also located close to the DGAT1 gene [21] with one SNP in this region, rs109421300 ( $q = 5.15 \times 10^{-39}$ ), being the most significant association for fat yield. However, 89 significant SNPs were detected outside known QTL regions from Cattle QTLdb [28] and were considered novel (i.e. a q-value  $\leq 0.05$  and located outside

any known QTL regions for fat yield). The most interesting and strongly-associated of these novel SNPs were two clusters of SNPs at 34 Mb and 39 Mb which are in close proximity to the GHR and PRLR genes, respectively. Of the 89 novel SNPs, one SNP in a similar region of chromosome 20 was also significant in the Bayesian method. We were unable to replicate these associations in the cow dataset, possibly due to the reduced power in that data (see discussion).

Using the Bayesian approach, 12 SNPs across eight chromosomes were significantly associated with fat yield in the population of 914 sires. Several of these SNPs (4/12), located on chromosomes 5, 7, and 14, were assigned posterior probabilities of association  $\geq 0.8$  using all three Bayesian priors and were located within known QTL regions [22,29,30]. One of the eight SNPs located in known QTL regions, rs29016908 on chromosome 5, was positioned 28 kb from the EPS8 gene which binds with the EGFR gene to alter responsiveness to EGF [31]. EGF is believed to affect various milk production traits [32]. An additional four significant SNPs, located on chromosomes 11, 20 and 25, did not overlap with known QTL regions and appear to be novel. Over half (7/12) of the SNPs significantly associated with fat yield in the Bayesian approach were also significant in the single SNP regression in the sires, providing further evidence of their association.

#### Protein yield

Using the single SNP regression model, 385 SNPs were significantly associated ( $q \leq 0.05$ ) with protein yield in the sires. Most of these SNPs (305/385) were found to lie within known QTL regions for protein yield [26,33,34]. Unlike milk and fat yield, the largest associations were not focused on chromosomes 14 and 20 but were instead distributed across numerous other chromosomes. The most significantly associated SNP across all chromosomes was rs42327956 ( $q = 3.86 \times 10^{-4}$ ) located at ~50,6 Mb on chromosome 1. Several significantly associated SNPs were clustered on chromosome 1 in a region which overlaps with known QTL regions for protein yield [33-35]. The genes in this region are supplied in Additional file 3. There were 80 significant SNPs located outside known QTL regions with a cluster of SNPs located on chromosome 8. However none of these 80 novel SNPs were significant in the Bayesian approach or in the cows.

There were two SNPs significantly associated with protein yield in the sire population using the Bayesian approach. These SNPs, located on chromosomes 11 and 27 were both located in known QTL regions for protein yield [34,36]. One SNP, rs41257411 located on chromosome 27, was also significantly associated with protein percentage using the single SNP regression in the population of 914 sires.

#### Fat percentage

Using the single SNP regression approach in the sires, a total of 216 SNPs were significantly associated ( $q \leq 0.05$ ) with fat percentage with 199 of these located within known QTL regions for fat percentage [22,24,35,37,38]. The location of significant associations was similar to that of fat yield with strongest associations harboured in a region encompassing the first 6 Mb of chromosome 14. Similar to milk yield, the strongest association ( $q = 3.91 \times 10^{-92}$ ) was detected for a SNP close to the DGAT1 gene, a gene known to heavily affect milk fat percentage [21,23]. Several significantly associated SNPs were also detected in a segment of chromosome 20 from ~34-37 Mb which are located close to the GHR gene, which is also known to affect milk production traits [38]. There were 17 potentially novel significant SNPs ( $q \leq 0.05$ ) detected outside known QTL regions for fat percentage. Of particular interest was a number of novel SNPs on chromosome 13 located at 46 Mb. A potentially novel association for milk yield was also detected in this region. However, none of these novel SNPs were detected as significant using the Bayesian approach or in the cows.

A total of 12 SNPs were significantly associated with fat percentage using the Bayesian approach in the sire population. Nine of these SNPs were located in known QTL regions [24,35,39]. Two SNPs, rs109421300 on chromosome 14 and rs110482506 on chromosome 20 located in close proximity to the DGAT1 and GHR genes, respectively. Surprisingly, only two SNPs significantly associated with fat percentage were also significantly associated with fat yield using the Bayesian approach, however, several significant SNPs were located in similar genomic regions to those associated with fat yield. Furthermore, three SNPs located on chromosomes 9, 21 and 27 did not occur in any known QTL regions for fat percentage. Of the 12 significant SNPs detected in the Bayesian approach, four of these were also significant for fat percentage in the sires using the single SNP regression approach.

#### Protein percentage

Using the single SNP regression in the sires, there were 229 SNPs significantly associated ( $q \leq 0.05$ ) with protein percentage of which 204 SNPs were within known QTL regions for protein percentage [19,24,26,30]. Like the majority of the milk production traits, clusters of associations were located on chromosomes 14 and 20. A cluster of significant SNPs in a region 0-6 Mb on chromosome 14 contained the strongest association ( $q = 7.44 \times 10^{-16}$ ) for a SNP close to the DGAT1 gene which has been shown to affect protein percentage [22,23]. In addition, a segment of chromosome 20 from 29 to 40 Mb, which contains the GHR gene, harboured a large number of significantly associated SNPs. Several strong associations were also detected

on chromosome 6 between 80 and 90 Mb, a region that contains the Casein gene cluster which has been associated with changes in the protein composition of bovine milk [40]. However, all the aforementioned associations were located within known QTL regions for protein percentage [19,24,26,41,42]. An additional 25 significant SNPs were found outside known QTL regions and these occurred across 10 different chromosomes. Of these rs41573791 ( $q = 9.94 \times 10^{-4}$ ) on chromosome 8 was the most strongly associated with protein percentage. In addition a number of novel SNPs located at 50 Mb on chromosome 15 were also moderately ( $q$ -value = 0.002) associated. One of these 25 novel SNPs, located on chromosome 5, was also detected as significant when using the Bayesian approach, however, none were significant in the cows.

There were eight SNPs significantly associated with protein percentage using the Bayesian approach in the sires with five of these SNPs overlapping known QTL regions [19,26,43]. The remaining three SNPs, located outside known QTL regions for protein percentage, were located on chromosomes 5, 12 and 18. Furthermore, five out of the eight significantly associated SNPs were also significantly associated with protein percentage using the single SNP regression in the sire population.

#### Somatic cell score

Only nine SNPs were significantly associated ( $q \leq 0.05$ ) with SCS using the single SNP regression in the sires, three of these were located within known QTL regions for SCS. These three SNPs were located on chromosomes 6 and 10 [34,44,45]. The remaining six SNPs, located outside known QTL regions for SCS, were spread across chromosomes 6, 15 and 20 with the most significant ( $q = 0.014$ ) of these located on chromosome 20. None of these six novel SNPs were significant in the Bayesian approach or in the cows. All SNPs significantly associated with SCS are listed in Additional file 3 along with genes close to or overlapping them.

Only a single SNP, rs41590209 located at ~97 Mb on chromosome 4, was significantly associated with SCS using the Bayesian approach and this fell into a known QTL region for SCS [46]. However, neither this SNP nor any SNP on chromosome 4 was significantly associated with SCS using the single SNP regression in the population of 773 sires.

#### Discussion

The objective of this study was to identify QTL associated with milk production traits and SCS in Holstein-Friesian cattle from a low-input production system. Both a frequentist and a Bayesian statistical approach were employed to test for association between genotypes and phenotypes. The QTL identified using both the sire and

cow populations were spread across all 29 autosomes; the location and frequency of these QTL were in general agreement with those previously reported [3,4].

A large number (1,529) of significant associations were detected across all traits. The majority of these significant associations were located within known QTL for the trait of interest. This shows that our methodology is effective in detecting associated regions of the genome. Also, our findings will help to further refine QTL regions previously detected with microsatellites [47]. The detection of a large number of known QTL regions in our study would suggest that a large number of QTL that are important in high-input, confinement, concentrate-based systems are also important in low-input, pasture-based systems such as ours. In spite of this, 276 novel SNPs were detected in the sires using the single SNP regression approach. Of these novel SNPs, a number of promising clusters of SNPs were identified for each trait which may indicate potential new QTL regions. These regions include an area of chromosome 13 significantly associated with milk yield and fat percentage. Also, significant novel associations were detected on chromosome 20 for fat yield and somatic cell score close to the GHR and PRLR genes reported to be associated with milk production traits and SCS [38,48]. In addition, particular areas of interest were separately detected for protein yield and percentage on chromosomes 8 and 15, respectively. These genomic regions may consist of QTLs that are unique to or advantageous in a low-input system such as ours.

Several significant associations, both within and outside known QTL regions, were detected for SCS. However, associations were considerably less numerous and weaker compared with those for the milk production traits. This may have been due to several inherent problems with the SCS phenotype resulting in reduced power to detect associations. Firstly, the reliability of the SCS proofs, which is an indicator of the amount of information available for an animal, was lower than that of the milk production traits in both the sires and cows. Decreased reliability of SCS means greater uncertainty as to the true breeding value of the animal for that trait. Furthermore, in the sire population, there were 138 fewer animals used to test for associations with SCS which would also decrease the power to detect significant associations. In addition, the lower heritability of SCS when compared to that of the milk production traits may also contribute to the weaker associations identified for SCS (i.e. a greater number of animals may be required for SCS).

The Bayesian analysis used provided a number of advantages/alternatives to the standard single SNP regression approach. This Bayesian approach fits all markers in the analysis simultaneously and it was noticeable that this approach detected only 1-2 significantly-associated SNPs

where the single SNP regression detected a cluster of numerous significantly-associated SNPs (i.e. on chromosome 14 for fat percentage). Additionally, the ability to allow *a priori* information to be factored into the statistical model appears to have merit where different traits may be controlled by varying numbers of genetic variants [49]. The prior appears to be robust, with similar genomic regions detected as significant even when using different priors.

Genome-wide association studies are susceptible to detection of false positives due to the large number of statistical approaches being performed. One method to confirm or validate a SNP association/QTL is via replication of the association in a separate population as the probability of detecting the same associated variant in two separate populations is small [50]. Our use of a separate population of Holstein-Friesian cows allowed validation of a number of associations from the sires, however, the size of this population was probably insufficient to validate SNPs of smaller effect (i.e. power was lower). Also the reliabilities of all traits in the cows were much lower than those of the sires resulting in potentially less accurate phenotypes to quantify the associations.

A number of SNPs in this analysis were significantly associated with more than one trait suggesting that genes with pleiotropic action may have been detected. Typically, in this study, a SNP affected multiple production traits with no association with SCS. Examples of this are three SNPs which were significantly associated with all five production traits (Additional file 3). This indicates that certain regions of the genome may affect various different production-related traits and this should be taken into consideration when selecting animals for a particular breeding goal. In addition, four SNPs were significantly associated with a production trait and SCS, in particular three SNPs on chromosome 20 were associated with a concurrent decrease in milk yield and SCS. This observation agrees with the well-known positive correlation that exists between milk yield and SCS [51]. Of these three SNPs, two lie in close proximity to the PRLR gene which has been reported to be associated with milk production [48] and changes in SCC [52]. QTL regions such as this may help elucidate how to select for increased milk yield without the associated detrimental effect on resistance to mastitis.

## Conclusion

A large number of significant associations were detected in this analysis which either overlapped known QTL regions or were novel associations found outside these QTL regions. Those associations found within known QTL regions can help to further refine large QTL regions potentially leading to the discovery of the underlying causal mechanisms. A number of strongly-associated regions

were detected among the 276 novel SNPs found outside known QTL regions. These genomic regions may indeed be unique to a low-input, pasture-based system. These QTL regions can form the basis towards identifying potential candidate genes and genetic variants underlying the traits of interest.

## Methods

### Sire DNA extraction

Thawed frozen semen was washed twice in phosphate-buffered saline (pH 7.4), and cell pellets were harvested via centrifugation and re-suspended in 450  $\mu$ L of pre-warmed extraction buffer (10 mM Tris, pH 8.0; 10 mM EDTA, pH 8.0; 1% SDS; 100 mM NaCl); 15  $\mu$ L of 2-mercaptoethanol was added. Samples were incubated at 55°C for 15 minutes followed by the addition of 10  $\mu$ L of proteinase K (20 mg/mL). Lysis occurred following overnight incubation at 60°C. DNA was then extracted using the Maxwell instrument (Promega Corp., Madison, WI) according to the manufacturer's instructions. Details of the DNA extraction method used in the cow dataset are available in Additional file 4.

### Sire genotypic and phenotypic data

In total 54,001 biallelic SNPs for 1,957 Holstein-Friesian AI sires with progeny in Ireland were genotyped in this study; the sires were representative of the germplasm used in Irish dairy herds in past years. All animal procedures were carried out according to the provisions of the Irish Cruelty to Animals Act (licenses issued by the Department of Health and Children). Sires were genotyped using the Illumina BovineSNP50 Genotyping Beadchip (Illumina Inc., San Diego, CA). SNP positions were based on BTAU 4.0. The 2,419 SNPs that were on the X chromosome or whose positions on the genome were unknown were eliminated from the dataset. SNPs on the X chromosome were removed due to large-scale discordance between SNP genotypes and animal sex (i.e. numerous SNP on the X chromosome, outside the pseudoautosomal regions, were called as heterozygous in male animals) indicating poor SNP calling. A further 230 SNPs were discarded that did not conform to Mendelian inheritance patterns between sire and son based on analysis of a larger dataset of animals [53]. The remaining 51,352 SNPs were subjected to additional SNP editing in the following order; SNPs were removed if they were monomorphic ( $n = 4,772$ ), had a minor allele frequency  $\leq 5\%$  ( $n = 5,087$ ), if greater than 5% of SNP calls were missing ( $n = 770$ ), if there was poor SNP clustering ( $n = 14$ ) or the proportion of heterozygotes for a SNP was  $> 90\%$  ( $n = 41$ ). Following all edits 40,668 SNPs remained.

Daughter yield deviations (DYD) for milk yield, fat yield, protein yield, fat percentage, protein percentage and SCS along with their respective reliabilities were

available from national genetic evaluations undertaken in January 2010 by the Irish Cattle Breeding Federation (ICBF). Summary statistics for the phenotypic data is presented in Table 1. DYD for the aforementioned milk production traits and average SCS (i.e.,  $\log_e$  SCC) are estimated in Ireland using a repeatability animal model across the first five lactations [54]. The parental contribution to the reliability of each trait was removed using the method described by Harris and Johnson [55]. Only sires with an adjusted reliability of  $\geq 80\%$  for milk production or  $\geq 70\%$  for SCS were retained. In total 914 sires met these criteria for inclusion in the analysis of milk production and 773 sires were included for SCS. Details regarding the genotypic and phenotypic data in the cows can be seen in Additional file 4.

### Statistical analyses

A single SNP regression model and a Bayesian method were both used to detect regions of the genome associated with milk production traits and SCS. In both approaches the dependent variable was either the DYD (sires) or YD (cows). SNPs were included in all models as continuous variables. A pedigree file, containing at least the previous four generations of the genotyped animals ancestors, was generated for each of the sire and cow datasets, respectively. The pedigree consisted of 5,479 animals for the sires and 4,476 individuals in the cows. The relatedness within and between the sires and cows was determined by calculating the average additive relationship between animals using the available pedigree information.

### Single SNP regression model

The dependent variable (i.e. the respective trait (daughter yield deviation) was regressed on each SNP individually in a mixed animal model accounting for relationships among animals using the additive genetic relationship matrix in ASReml [56]. The individual animal was included as a random effect. Estimates of SNP effect and the associated standard error were recorded for each SNP. A false discovery rate (FDR) approach as described by Storey and Tibshirani [9], was used to correct for multiple testing. This was done in R version 2.12.0 [57] using the q-value package [58] to calculate q values (q). Q-values  $\leq 0.05$  were defined as significant.

### Bayesian model

The Bayesian method used was BayesB as described by Meuwissen et al. [59] modified to allow weightings based on the reliability of the dependent variable. An inverse chi-square distribution ( $\nu = 4.234$   $S = 0.0429$ ) was used as a prior to reflect the assumption that, within a given population, only a subset of all loci truly affect the phenotype while the remainder have no effect at all on the phenotype. A prior value was assigned to  $\pi$  which quantifies the

probability a SNP is associated with a phenotype. In this case the same value of  $\pi$  was used for all SNPs within a trait and thus  $\pi$  reflects the overall proportion of SNPs assumed to be associated with a particular phenotype. Here,  $\pi$  was calculated for each trait separately by dividing the total number of significant SNPs from the single SNP regression model for that trait in the sires by the total number of SNPs in the analysis (i.e. 40,668 SNPs). Additional BayesB analyses were run with different values for  $\pi$  to test the sensitivity of the associations to this prior; these priors included twice or half the original prior and are provided in Table 3. Further values of  $\pi$  (e.g.  $\pi = 0.0001$ ,  $0.00001$ ) were tested in the Bayesian approach to gauge the influence of prior values on the Bayesian approach. The Markov Chain Monte Carlo chains were used to sample from the posterior distribution and these were run for 300,000 cycles with the first 150,000 cycles discarded as burn-in. Convergence was determined visually by plotting the model log-likelihood for all iterations. The number of times a SNP had a non-zero effect after the burn-in was recorded for each SNP; this number was then divided by the total number of iterations after burn-in to give the posterior probability of association (PPA). The PPA, ranging from zero (no association) to one (highly associated), indicated the strength of evidence from the posterior distribution that the particular SNP was associated with the phenotype in a manner roughly equivalent to a frequentist p-value.

### Assignment of SNP to genes

Genotypic data from the sires consisting of 40,668 SNPs typed across 914 individuals were used to estimate haplotype blocks. The 'Solid Spine of LD' method (based on

**Table 3 Values of  $\pi$  (prior) used in the Bayesian analysis in the sires**

	S.S. in Sires	Total	SSR	SSR/2	SSRx2
Milk Yield	370	40,668	0.0091	0.0045	0.0182
Fat Yield	370	40,668	0.0091	0.0045	0.0182
Protein Yield	385	40,668	0.0095	0.0047	0.0189
Fat %	216	40,668	0.0053	0.0027	0.0106
Protein %	229	40,668	0.0056	0.0028	0.0113
SCS	9	40,668	0.0002	0.0001	0.0004

Fat % = fat percentage; Protein % = protein percentage; SCS = somatic cell score; S.S. in Sires = number of significant SNPs from the single SNP regression model in the sires for a particular trait; Total = total number of SNPs in the analysis; SSR = value of  $\pi$  based on the number of significant SNPs from the single SNP regression model in the sires (i.e. for milk yield,  $370/40668 = 0.0091$ ); SSR/2 = half the value of  $\pi$  than that of the SSR prior; SSRx2 = twice the value of  $\pi$  than that of the SSR prior

For the Bayesian analysis three different priors were used for each trait; these priors represent the proportion of SNPs believed to affect the particular trait. The SSR prior for a trait is calculated by dividing the total number of significant SNPs from the single SNP regression analysis in the sires by the total number of SNPs in the analysis (i.e. 40,668). The other two priors, SSR/2 and SSRx2 are simply the half and double the SSR prior respectively



the D prime statistic) in Haploview [60], with default settings, was used to estimate the blocks. Bovine gene start and end positions were sourced from Ensembl [61] based on build BTAU 4.0. SNPs were assigned to genes using an in-house Perl [62] script which worked in the following manner. If a SNP was located within a particular haplotype block then all genes that overlapped with that haplotype block were assigned to the SNP. Any SNP located outside a haplotype block was assigned to the closest gene. The assignment of SNPs to genes allowed a list of associated genes to be generated for each trait from the significant SNPs detected across all statistical approaches in the sires and cows.

### Checking for overlap with known QTL regions

All significant associations were examined to determine if they overlapped with known QTL regions using an in-house Perl script; if any significant SNP did not overlap with a known QTL region it was considered novel. The location of known QTL regions for the milk production traits and SCS were sourced from Cattle QTLdb [28]. We used the 'QTL Span' and/or 'QTL Peak Location' values in CattleQTLdb to define the boundaries of these known QTL. The keywords used for searches in the database were 'milk yield', 'milk fat yield', 'milk protein yield', 'milk fat percentage', 'milk protein percentage' and 'somatic cell score' for the traits milk yield, fat yield, protein yield, fat percentage, protein percentage and somatic cell score, respectively.

### Additional material

**Additional file 1: Summary statistics for the cow dataset.** This file contains summary statistics for the phenotypic data in the cow validation dataset mentioned in the manuscript. Summary statistics include the total number of phenotypic records (N), mean and standard deviation ( $\sigma$ ) for each trait in the cows. Phenotypes in the cows are expressed as yield deviations on a PTA scale.

**Additional file 2: Correlations between phenotypes for the cow dataset.** This file contains Pearson correlations between phenotypes in the cow dataset expressed as yield deviations on a PTA scale.

**Additional file 3: All Significant Associations.** This file contains all significant SNPs detected across all the analyses in the sires and cows. Significant associations are presented separately for each analysis with separate worksheets for each trait.

**Additional file 4: Cow Dataset Materials and Methods.** This file contains information on the materials and methods used for the cow validation dataset mentioned in the manuscript.

### Acknowledgements

The authors wish to acknowledge the SFI/HEA Irish Centre for High-End Computing (ICHEC) for the provision of computational facilities and support. We also acknowledge funding provided for this project through the Teagasc Walsh Fellowship scheme and Teagasc RMIS 6018. We would also like to thank Dr. Chris Creevey for helpful discussions. DGB is an SFI Principal Investigator (09/IN.1/B2642).

### Author details

<sup>1</sup>Animal and Bioscience Research Department, Teagasc, Animal & Grassland Research and Innovation Centre, Grange, Dunsany, Co. Meath, Ireland. <sup>2</sup>Animal and Bioscience Research Department, Teagasc, Animal & Grassland Research and Innovation Centre, Moorepark, Co. Cork, Fermoy, Ireland. <sup>3</sup>Irish Cattle Breeding Federation, Bandon, Co. Cork, Ireland. <sup>4</sup>Smurfit Institute of Genetics, Trinity College Dublin, Dublin 2, Ireland. <sup>5</sup>School of Agriculture and Food Science, University College Dublin, Dublin 4, Belfield, Ireland.

### Authors' contributions

DPB, DJL and AGF designed research; BKM performed the research, analyzed data and wrote the paper with DJL and DPB; JFK, DGB provided genotypic and phenotypic data; EKF carried out the DNA extraction. All authors read and approved the final manuscript.

### Competing interests

The authors declare that they have no competing interests.

Received: 7 October 2011 Accepted: 26 March 2012

Published: 26 March 2012

### References

1. Whitley L: The world dairy situation. *Int J Dairy Technol* 2010, **63**(3):471-472.
2. Miglior F, Muir BL, Van Doormaal BJ: Selection Indices in Holstein cattle of various countries. *J Dairy Sci* 2005, **88**(3):1255-1263.
3. Khatkar M, Thomson P, Tammen I, Raadsma H: Quantitative trait loci mapping in dairy cattle: review and meta-analysis. *Genet Sel Evol* 2004, **36**(2):163-190.
4. Smaragdov M: Genetic mapping of loci responsible for milk production traits in dairy cattle. *Russ J Genet* 2006, **42**(1):1-15.
5. Balding DJ: A tutorial on statistical methods for population association studies. *Nat Rev Genet* 2006, **7**(10):781-791.
6. Pryce JE, Bolormaa S, Chamberlain AJ, Bowman PJ, Savin K, Goddard ME, Hayes BJ: A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. *J Dairy Sci* 2010, **93**(7):3331-3345.
7. Cole JB, VanRaden PM, O'Connell JR, Tassell CPV, Sonstegard TS, Schnabel RD, Taylor JF, Wiggans GR: Distribution and location of genetic effects for dairy traits. *J Dairy Sci* 2009, **92**(6):2931-2946.
8. Jiang L, Liu J, Sun D, Ma P, Ding X, Yu Y, Zhang Q: Genome wide association studies for milk production traits in Chinese Holstein population. *PLoS One* 2010, **5**(10):e13661.
9. Storey JD, Tibshirani R: Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 2003, **100**(16):9440-9445.
10. Feungang J, Kaya A, Page G, Chen L, Mehta T, Hirani K, Nazareth L, Topper E, Gibbs R, Memili E: Two-stage genome-wide association study identifies integrin beta 5 as having potential role in bull fertility. *BMC Genomics* 2009, **10**(1):176.
11. McCarthy S, Horan B, Dillon P, O'Connor P, Rath M, Shalloo L: Economic comparison of divergent strains of Holstein-Friesian cows in various pasture-based production systems. *J Dairy Sci* 2007, **90**(3):1493-1505.
12. Horan B, Dillon P, Faverdin P, Delaby L, Buckley F, Rath M: The interaction of strain of Holstein-Friesian cows and pasture-based feed systems on milk yield, body weight, and body condition score. *J Dairy Sci* 2005, **88**(3):1231-1243.
13. Berry DP, Buckley F, Dillon P, Evans RD, Rath M, Veerkamp RF: Estimation of genotype  $\times$  environment interactions, in a grass-based system, for milk yield, body condition score, and body weight using random regression models. *Livest Prod Sci* 2003, **83**(2-3):191-203.
14. Fahey AG, Schutz MM, Lofgren DL, Schinckel AP, Stewart TS: Genotype by environment interaction for production traits while accounting for heteroscedasticity. *J Dairy Sci* 2007, **90**(8):3889-3899.
15. Kolver ES, Roche JR, De Veth MJ, Thorne PL, Napper AR: Total mixed rations versus pasture diets. evidence for a genotype  $\times$  diet interaction in dairy cow performance. *Proc New Zeal Soc An* 2002, **62**:246-251.
16. Interbull. [http://www.interbull.org/].
17. Grisart B, Coppieters W, Farnir Fdr, Karim L, Ford C, Berzi P, Cambisano N, Mni M, Reid S, Simon P, et al: Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1

- gene with major effect on milk yield and composition. *Genome Res* 2002, **12**(2):222-231.
18. Plante Y, Gibson JP, Nadesalingam J, Mehrabani-Yeganeh H, Lefebvre S, Vandervoort G, Jansen GB: **Detection of quantitative trait loci affecting milk production traits on 10 chromosomes in Holstein cattle.** *J Dairy Sci* 2001, **84**(6):1516-1524.
  19. Bagnato A, Schiavini F, Rossoni A, Maltecca C, Dolezal M, Medugorac I, Sölkner J, Russo V, Fontanesi L, Friedmann A, et al: **Quantitative trait loci affecting milk yield and protein percentage in a three-country brown swiss population.** *J Dairy Sci* 2008, **91**(2):767-783.
  20. Bennewitz J, Reinsch N, Grohs C, Levéziel H, Malafosse A, Thomsen H, Xu N, Looft C, Kühn C, Brockmann GA, et al: **Combined analysis of data from two granddaughter designs: a simple strategy for QTL confirmation and increasing experimental power in dairy cattle.** *Genet Sel Evol* 2003, **35**(3):319-338.
  21. Grisart B, Farnir Fdr, Karim L, Cambisano N, Kim J-J, Kvasz A, Mni M, Simon P, Frère J-M, Coppieters W, et al: **Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition.** *Proc Natl Acad Sci USA* 2004, **101**(8):2398-2403.
  22. Thaller G, Kramer W, Winter A, Kaupe B, Erhardt G, Fries R: **Effects of DGAT1 variants on milk production traits in German cattle breeds.** *J Anim Sci* 2003, **81**(8):1911-1918.
  23. Berry DP, Howard D, O'Boyle P, Waters S, Kearney JF, McCabe M: **Associations between the K232A polymorphism in the diacylglycerol-O-transferase 1 (DGAT1) gene and performance in Irish Holstein-Friesian dairy cattle.** *Irish J Agr Food Res* 2010, **49**(1):1-9.
  24. Arranz JJ, Coppieters W, Berzi P, Cambisano N, Grisart B, Karim L, Marcq F, Moreau L, Mezer C, Riquet J, et al: **A QTL affecting milk yield and composition maps to bovine chromosome 20: a confirmation.** *Anim Genet* 1998, **29**(2):107-115.
  25. Rodriguez-Zas SL, Southey BR, Heyen DW, Lewin HA: **Detection of quantitative trait loci influencing dairy traits using a model for longitudinal data.** *J Dairy Sci* 2002, **85**(10):2681-2691.
  26. Heck JML, Schennink A, van Valenberg HJF, Bovenhuis H, Visker MHPW, van Arendonk JAM, van Hooijdonk ACM: **Effects of milk protein variants on the protein composition of bovine milk.** *J Dairy Sci* 2009, **92**(3):1192-1202.
  27. Szyda J, Liu Z, Reinhardt F, Reents R: **Estimation of quantitative trait loci parameters for milk production traits in German Holstein Dairy cattle population.** *J Dairy Sci* 2005, **88**(1):356-367.
  28. **CattleQTLdb.** [<http://www.animalgenome.org/QTLdb/cattle.html>].
  29. Olsen HG, Gomez-Raya L, Vage DI, Olsaker I, Klungland H, Svendsen M, Adnøy T, Sabry A, Klemetsdal G, Schulman N, et al: **A genome scan for quantitative trait loci affecting milk production in Norwegian dairy cattle.** *J Dairy Sci* 2002, **85**(11):3124-3130.
  30. Ashwell MS, Da Y, Van Tassell CP, Vanraden PM, Miller RH, Rexroad CE Jr: **Detection of putative loci affecting milk production and composition, health, and type traits in a united states Holstein population.** *J Dairy Sci* 1998, **81**(12):3309-3314.
  31. Fazioli F, Minichiello L, Matoska V, Castagnino P, Miki T, Wong WT, Di Fiore PP: **Eps8, a substrate for the epidermal growth factor receptor kinase, enhances EGF-dependent mitogenic signals.** *EMBO J* 1993, **12**(10):3799-3808.
  32. Ogorevc J, Kunej T, Razpet A, Dovc P: **Database of cattle candidate genes and genetic markers for milk production and mastitis.** *Anim Genet* 2009, **40**(6):832-851.
  33. Nadesalingam J, Plante Y, Gibson JP: **Detection of QTL for milk production on Chromosomes 1 and 6 of Holstein cattle.** *Mamm Genome* 2001, **12**(1):27-31.
  34. Daetwyler HD, Schenkel FS, Sargolzaei M, Robinson JAB: **A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map.** *J Dairy Sci* 2008, **91**(8):3225-3236.
  35. Zhang Q, Boichard D, Hoeschele I, Ernst C, Eggen A, Murkve B, Pfister-Genskow M, Witte LA, Grignola FE, Uimari P, et al: **Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree.** *Genetics* 1998, **149**(4):1959-1973.
  36. Viitala SM, Schulman NF, de Koning DJ, Elo K, Kinosh R, Virta A, Virta J, Mäki-Tanila A, Vilkki JH: **Quantitative trait loci affecting milk production traits in finnish ayrshire dairy cattle.** *J Dairy Sci* 2003, **86**(5):1828-1836.
  37. Kaupe B, Brandt H, Prinzenberg EM, Erhardt G: **Joint analysis of the influence of CYP11B1 and DGAT1 genetic variation on milk production, somatic cell score, conformation, reproduction, and productive lifespan in German Holstein cattle.** *J Anim Sci* 2007, **85**(1):11-21.
  38. Blott S, Kim J-J, Moisis S, Schmidt-Kuntzel A, Cornet A, Berzi P, Cambisano N, Ford C, Grisart B, Johnson D, et al: **Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the Bovine growth hormone receptor is associated with a major effect on milk yield and composition.** *Genetics* 2003, **163**(1):253-266.
  39. Thaller G, Kühn C, Winter A, Ewald G, Bellmann O, Wegner J, Zühlke H, Fries R: **DGAT1, a new positional and functional candidate gene for intramuscular fat deposition in cattle.** *Anim Genet* 2003, **34**(5):354-357.
  40. Schopen GCB, Koks PD, Van Arendonk JAM, Bovenhuis H, Visker MHPW: **Whole genome scan to detect quantitative trait loci for bovine milk protein composition.** *Anim Genet* 2009, **40**(4):524-537.
  41. Näslund J, Fikse WF, Pielberg GR, Lundén A: **Frequency and effect of the Bovine Acyl-CoA:diacylglycerol Acyltransferase 1 (DGAT1) K232A polymorphism in Swedish dairy cattle.** *J Dairy Sci* 2008, **91**(5):2127-2134.
  42. Boichard D, Grohs C, Bourgeois F, Cerqueira F, Faugeras R, Neau A, Rupp R, Amigues Y, Boscher M, Leveziel H: **Detection of genes influencing economic traits in three French dairy cattle breeds.** *Genet Sel Evol* 2003, **35**(1):77-101.
  43. Ashwell MS, Heyen DW, Sonstegard TS, Van Tassell CP, Da Y, VanRaden PM, Ron M, Weller JL, Lewin HA: **Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle.** *J Dairy Sci* 2004, **87**(2):468-475.
  44. Bennewitz J, Reinsch N, Guiard V, Fritz S, Thomsen H, Looft C, Kuhn C, Schwerin M, Weimann C, Erhardt G, et al: **Multiple quantitative trait loci mapping with cofactors and application of alternative variants of the false discovery rate in an enlarged granddaughter design.** *Genetics* 2004, **168**(2):1019-1027.
  45. Rupp R, Boichard D: **Genetics of resistance to mastitis in dairy cattle.** *Vet Res* 2003, **34**(5):671-688.
  46. Tal-Stein R, Fontanesi L, Dolezal M, Scotti E, Bagnato A, Russo V, Canavesi F, Friedmann A, Soller M, Lipkin E: **A genome scan for quantitative trait loci affecting milk somatic cell score in Israeli and Italian Holstein cows by means of selective DNA pooling with single- and multiple-marker mapping.** *J Dairy Sci* 2010, **93**(10):4913-4927.
  47. Ashwell MS, Rexroad CEJ, Miller RH, Van Raden PM, Da Y: **Detection of loci affecting milk production and health traits in an elite US Holstein population using microsatellite markers.** *Anim Genet* 1997, **28**(3):216-222.
  48. Viitala S, Szyda J, Blott S, Schulman N, Lidauer M, Mäki-Tanila A, Georges M, Vilkki J: **The role of the Bovine growth hormone receptor and prolactin receptor genes in milk, fat and protein production in Finnish Ayrshire dairy cattle.** *Genetics* 2006, **173**(4):2151-2164.
  49. Hayes BJ, Pryce J, Chamberlain AJ, Bowman PJ, Goddard ME: **Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits.** *PLoS Genet* 2010, **6**(9):e1001139.
  50. Visscher PM: **Sizing up human height variation.** *Nat Genet* 2008, **40**(5):489-490.
  51. Carlén E, Strandberg E, Roth A: **Genetic parameters for clinical mastitis, somatic cell score, and production in the first three lactations of Swedish Holstein cows.** *J Dairy Sci* 2004, **87**(9):3062-3070.
  52. Boutet P, Sulon J, Closset R, Detilleux J, Beckers JF, Bureau F, Lekeux P: **Prolactin-induced activation of nuclear factor κB in Bovine mammary epithelial cells: role in chronic mastitis.** *J Dairy Sci* 2007, **90**(1):155-164.
  53. Kearney JF, Cromie AR, Berry DP: **Implementation and uptake of genomic selection in Irish Holstein-Friesian Dairy Cattle.** *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production: August 1-6 2010 Leipzig, Germany*; 2010.
  54. Berry DP, Shalloo L, Cromie AR, Veerkamp RF, Dillon P, Amer PR, Kearney JF, Evans RD, Wickham B: *The economic breeding index: a generation on Technical Report to the Irish Cattle Breeding Federation* 2007, 29-34.
  55. Harris B, Johnson D: **Approximate reliability of genetic evaluations under an animal model.** *J Dairy Sci* 1998, **81**(10):2723-2728.
  56. Gilmour AR, Cullis BR, Gogel BJ, Welham SJ, Thompson R: *ASReml User Guide Release 3.0* UK: VSN International Ltd, Hemel Hempstead, HP1 1ES; 2009.
  57. **The R Project for Statistical Computing.** [<http://www.r-project.org/>].

58. **Q-value.** [<http://genomics.princeton.edu/storeylab/qvalue/>].
59. Meuwissen TH, Hayes BJ, Goddard ME: **Prediction of total genetic value using genome-wide dense marker maps.** *Genetics* 2001, **157**(4):1819-1829.
60. **Haploview.** [<http://www.broad.mit.edu/mpg/haploview/>].
61. **Ensembl.** [<http://www.ensembl.org/>].
62. **Perl.** [<http://www.r-project.org/>].

doi:10.1186/1471-2156-13-21

**Cite this article as:** Meredith *et al.*: Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. *BMC Genetics* 2012 **13**:21.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

