

## REVIEW ARTICLE

**Genomic selection**M.E. Goddard<sup>1,2</sup> & B.J. Hayes<sup>2</sup>

1 University of Melbourne, Attwood, Australia

2 Department of Primary Industries, Attwood, Australia

**Keywords**

genomic selection; linkage disequilibrium; prior distribution.

**Correspondence**

Mike Goddard, Department of Primary Industries, 475 Mickleham Rd, Attwood Victoria 3049, Victoria. Tel: +61392174152; Fax: +61392174299; E-mail: mike.goddard@dpi.vic.gov.au

Received: 30 July 2007;

accepted: 15 October 2007

**Summary**

Genomic selection is a form of marker-assisted selection in which genetic markers covering the whole genome are used so that all quantitative trait loci (QTL) are in linkage disequilibrium with at least one marker. This approach has become feasible thanks to the large number of single nucleotide polymorphisms (SNP) discovered by genome sequencing and new methods to efficiently genotype large number of SNP. Simulation results and limited experimental results suggest that breeding values can be predicted with high accuracy using genetic markers alone but more validation is required especially in samples of the population different from that in which the effect of the markers was estimated. The ideal method to estimate the breeding value from genomic data is to calculate the conditional mean of the breeding value given the genotype of the animal at each QTL. This conditional mean can only be calculated by using a prior distribution of QTL effects so this should be part of the research carried out to implement genomic selection. In practice, this method of estimating breeding values is approximated by using the marker genotypes instead of the QTL genotypes but the ideal method is likely to be approached more closely as more sequence and SNP data is obtained. Implementation of genomic selection is likely to have major implications for genetic evaluation systems and for genetic improvement programmes generally and these are discussed.

**Introduction**

Traditional genetic improvement of livestock, using information on phenotypes and pedigrees to predict breeding values, has been very successful. However, breeding values should be able to predict more accurately by using information on variation in DNA sequence between animals. Research towards marker-assisted selection (MAS) has been extensive but implementation has been limited and increases in genetic gain have been small (Dekkers 2004). The factors governing the additional gains from MAS (Goddard and Hayes 2002) are:

- The accuracy of the existing estimated breeding values (EBV). If the accuracy is already high, there can be little gain. Consequently, gains are larger where traditional selection is most difficult, e.g. traits displayed only in females.
- The proportion of the genetic variance explained by the DNA markers.
- The accuracy with which the effect of the marked quantitative trait loci (QTL) alleles are estimated.
- The ability to reduce the generation interval by selecting MAS and breeding at an earlier age than previously.

As most economic traits are influenced by many genes, tracking a small number of these through DNA markers will only explain a small proportion of the genetic variance. In addition, individual genes are likely to have small effects and so a large amount of data is needed to accurately estimate their effects. This problem is exacerbated if a haplotype of markers is used to track QTL as many haplotype effects must be estimated.

The markers used for MAS can be linked to the QTL but in linkage equilibrium with it; in linkage disequilibrium (LD), the QTL or the marker can actually be the QTL (Dekkers 2004). If the marker is in linkage equilibrium with the QTL, all QTL alleles in founder animals are considered to be different and hence the number of QTL alleles whose effects must be estimated is further increased. Despite these difficulties, Boichard *et al.* (2006) show how gains can be made, although a very large amount of genotyping was necessary.

To overcome these difficulties, Meuwissen *et al.* (2001) proposed a variant of MAS that they called genomic selection. The key features of this method are that markers covering the whole genome are used so that potentially all the genetic variance is explained by the markers; and the markers are assumed to be in LD with the QTL so that the number of effects per QTL to be estimated is small. Using simulation, they showed that the breeding value could be predicted with an accuracy of 0.85 from marker data alone.

The major limitation to the implementation of genomic selection has been the large number of markers required and the cost of genotyping these markers. Recently both these limitations have been overcome in most livestock species following the sequencing of the livestock genomes, the subsequent availability of hundreds of thousands of single nucleotide polymorphisms (SNP), and dramatic developments in SNP genotyping technology which allow genotyping an SNP for as little as 1 US cent per animal. In cattle a commercial assay with 10 000 SNP has been available for 2 years and assays with 25–50 000 SNP are becoming available soon. In human genetics, assays with 500 000 SNP are routinely used (e.g. Parks *et al.* 2007).

As a result of these developments there are many livestock breeding companies planning to implement genomic selection in the near future. The purpose of this paper is to review the requirements for maximum benefits to be derived from genomic selection.

### Statistical analysis to calculate EBV from genome-wide DNA markers

It is convenient to think of the process in three steps:

1. Use the markers to deduce the genotype of each animal at each QTL.
2. Estimate the effects of each QTL genotype on the trait.
3. Sum all the QTL effects for selection candidates to obtain their genomic EBV (GEBV).

#### Use the markers to deduce the genotype of each animal at each QTL

The simplest method to deduce QTL genotypes is to treat the markers as if they were QTL and to estimate the effects of the marker alleles or genotypes. The key parameter here is the proportion of the QTL variance explained by the markers ( $r^2$ ). This is dependent on the LD between the QTL and one marker or a linear combination of markers. The extent of LD and hence  $r^2$  are highly variable. Average  $r^2$  declines as the distance between the two loci increases. In Holstein cattle the average  $r^2$  when loci are 50 kb apart is 0.35 (Goddard *et al.* 2006). To obtain an average spacing of 50 kb requires 60 000 evenly spaced markers. As the markers are unlikely to be evenly spaced, and due to the variable nature of the LD, we could still not expect that all QTL would have an SNP in complete LD with them. This suggests that we need denser markers than are currently available. The technology to achieve this is available (e.g. Parks *et al.* 2007).

An additional problem arises if we wish to estimate the effect of each marker across more than one breed. Then we require not only high  $r^2$  in each breed but the same linkage phase between the marker and the QTL in each breed. Goddard *et al.* (2006) found that this occurred between Angus and Holstein cattle only for markers <10 kb apart.

An alternative to using single marker genotypes is to construct haplotypes based on several markers. A QTL that is not in complete LD with any individual marker may be in complete LD with a multi-marker haplotype. Using 9323 SNP genotypes from Angus cattle, and considering a randomly chosen SNP as a surrogate for a QTL, Hayes *et al.* (2007) found that the proportion of variance explained by a haplotype of the surrounding markers increased from 0.2 for the nearest marker to 0.58 for a six-marker haplotype. The use of multiple marker

genotypes but without deducing haplotypes, e.g. with multiple marker regression, will be in between these two extremes (Goddard 1991). Typically there are many haplotypes present in a population and so the amount of data with which to estimate the effect of each one is reduced and this will reduce the accuracy with which each haplotype effect is estimated. However, Hayes et al. (2007) and Grapes et al. (2004) still found that the increase in QTL variance explained from using marker haplotypes more than compensated for the decrease in accuracy of estimating a greater number of haplotype effects, so that haplotypes predicted the effect of the QTL alleles more accurately than a single marker. Calus et al. (2007) compared the accuracy of GEBV when haplotypes or single markers were used to predict the QTL effects, at different levels of  $r^2$  between adjacent markers. They found that the advantage of haplotypes over single markers decreased as the  $r^2$  between adjacent markers increased. At  $r^2 = 0.215$  between adjacent markers, the haplotype approach and single marker approach gave very similar accuracies.

As the total number of animals with phenotypes and marker genotypes increases, the accuracy of estimating marker genotype effects will approach 1.0 and so will the accuracy of estimating the haplotype effects. However, the accuracy for the haplotype will approach 1.0 more slowly than the accuracy of estimating SNP effects because there are more than two haplotype effects per QTL to be estimated. Therefore, the advantage of the haplotypes over single markers is expected to increase as the amount of data for estimation increases, especially at lower marker densities (e.g. Zhao et al. 2007). In fact the accuracy of using single markers can be greater than using marker haplotypes if there is a limited number of phenotypic records to estimate the effects and the level of LD between the single markers and QTL is very high (Zhao et al. 2007).

An alternative to treating a haplotype of markers as if it were a QTL allele is to treat every gamete as carrying a different QTL allele but to estimate the correlation among the effects of these alleles based on the surrounding markers. A linkage analysis traces the QTL alleles through the known pedigree using the markers and calculates the probability that any two alleles are identical by descent (IBD) from a common ancestor within the pedigree. The probability that two QTL alleles are IBD because of a common ancestor outside the pedigree can be assessed from the similarity of the marker alleles surrounding the QTL by assuming an evolutionary model for the

LD between the markers and the QTL (Meuwissen and Goddard 2007). The linkage analysis and the LD analysis can be combined to estimate a matrix of IBD probabilities between all QTL alleles and this can be used to estimate the effects of all QTL alleles (Meuwissen et al. 2002). Errors in the positioning of the markers on the genome will reduce the accuracy of the inferring haplotypes, and therefore the accuracy of GEBV resulting from both the haplotype and the IBD approaches.

As one moves from using single marker genotypes to haplotypes or further to IBD probabilities the computational burden increases. At least at low marker densities (e.g.  $r^2$  between adjacent markers less than 0.2) this computational burden appears to be justified. Calus et al. (2007) demonstrated in this situation, that the IBD approach performed better than the haplotype approach, which in turn performed better than the single marker approach. At high marker densities on the other hand, the additional computational burden may not be justified as in this case the accuracies from all three methods are similar.

#### Estimating the effect of each QTL genotype on the trait

Genetic gain is greatest if the estimate of breeding value ( $\mathbf{g}$ ) has the property  $\text{GEBV} = E(\mathbf{g}'\text{data})$ . As the EBV will be calculated by summing the estimated effects of all QTLs ( $\mathbf{u}$ ), the desired property for the EBV can be achieved if we estimate each QTL effect by  $\hat{\mathbf{u}} = E(\mathbf{u}'\text{data})$ . The appropriate estimator is:

$$\hat{\mathbf{u}} = \frac{\int \mathbf{u} * p(\text{data}|\mathbf{u})p(\mathbf{u})d\mathbf{u}}{\int p(\text{data}|\mathbf{u})p(\mathbf{u})d\mathbf{u}}$$

where  $p(\text{data}|\mathbf{u})$  is a likelihood and  $p(\mathbf{u})$  is a prior distribution of QTL effects.

This shows that the ideal estimator of the QTL effects depends on the prior distribution of QTL effects. As we typically test for a QTL at many positions (e.g. 10 000 SNP), we expect that there is no QTL at most positions. Therefore, the prior distribution  $p(\mathbf{u})$  must have a high probability for  $p(0)$ . To know how high, we need to know how many QTL affect our trait. For milk production traits in dairy cattle we estimated that there are 150 QTL (Hayes et al. 2006), but no doubt there are even more because our power to detect them was not 100%.

For these 150 QTL we have estimated the distribution of their effects to be approximately exponential (Chamberlain et al. 2007) and others have reached a similar conclusion (e.g. Xu 2003; Weller et al. 2005).

Although an explicit model of the distribution of QTL effects seems logical, it is not the only option. Least squares is an alternative. However, if least squares is used to estimate the QTL effects, there are three disadvantages. First, it will be impossible to estimate the effects of all QTL simultaneously, as the number of effects to estimate (e.g. haplotypes or marker effects) will almost always be much larger than the number of records. Second, because the effects must therefore be estimated one at a time and used only if they exceed some imposed significance threshold, the significant effects are greatly overestimated, often by twofold (Beavis 1994). Third, the resulting correlation between EBV and true BV is low (Meuwissen et al. 2001). Least squares estimates correspond to assuming a prior distribution of QTL effects with an infinitely large variance which is obviously incompatible with the known total genetic variance. The other problem with least squares is that, because a significance threshold is imposed, only the QTL with large effect will be 'detected' and used, so consequently not all of the genetic variance will be captured by the markers. An alternative is to assume that the QTL effects are drawn from normal distribution with constant variance across chromosome segments, in which case the estimates are best linear unbiased prediction (BLUP) estimates, and all effects can be estimated simultaneously (Schaeffer 2006). This results in estimates that are better correlated with the true BV than from the least squares analysis but not as well as from a Bayesian analysis that uses a more appropriate prior distribution of QTL effects (Meuwissen et al. 2001). In the situation where most 'QTL' have zero effect, least squares and BLUP result in these zero effects being estimated to be small but non-zero and their cumulative effect adds noise to the estimates.

Better estimates are obtained where many possible QTL are estimated to have zero effect or, equivalently, excluded from the model. If all the QTL effects were from a reflected exponential distribution (i.e. without extra weight at zero), an estimator called the LASSO is the appropriate one (Tibshirani 1996). However, in the situation where many true effects are zero, LASSO still estimates too many non-zero effects. A pragmatic alternative is to exclude from the model all but the most highly significant effects. For instance, Wray et al. (2007) found that setting a significance threshold so that only one false positive per genome was expected to give EBV highly correlated with the BV. However, if the effects of these significant QTL are estimated by least

squares, the effects will still be overestimated. This overestimation can be corrected by using cross-validation (Whittaker et al. 1997). This involves estimating the effects in two independent parts of the data, and calculating the regression of one set of solutions on the other. The solutions are then regressed back by this regression coefficient to give unbiased estimates. Cross-validation can also be used to choose between competing models. Within a dataset adding extra QTL always increases the accuracy of prediction but the accuracy of GEBV in an independent dataset can be used to judge whether the accuracy has really increased.

Although these pragmatic methods may be satisfactory, it seems more logical to use an explicit prior. Meuwissen et al. (2001) assumed that QTL effects were drawn from a normal distribution but that the variance of that distribution varied between QTL, and that the distribution of variances followed an inverted chi-square distribution. The choice of these distributions was partly for computational efficiency, and did not exactly match the distribution used to simulate the data, but did give more accurate EBV than the BLUP estimates assuming constant variance. An advantage of using the correct prior is that the estimates of the biggest or most significant QTL are not overestimated. This means that the effects can be estimated from all available data regardless of whether the data was part of that used to discover the QTL or not. This should be an important advantage as genomic selection becomes implemented in the industry and it becomes impossible to clearly distinguish between discovery data (where least squares estimates are biased) and independent, validation data (where they are unbiased).

In all of the aforementioned methods to estimate BV, a polygenic term can be added to the model to account for the genetic variance not explained by the markers. When one marker at a time is tested for significance, we find that if the polygenic term is omitted from the model, twice as many false positives occur as indicated by the significance threshold (MacLeod et al. 2007). This is because, within a dataset, all markers and QTL are correlated through the pedigree relationship among the animals. Consequently, any marker can, by chance, be correlated with a QTL some distance away or even on another chromosome and so appear to have an effect which is actually an artefact of the pedigree structure. Even when all QTL are fitted simultaneously, it may still be desirable to fit a polygenic effect, as this will capture, to some extent, those QTL that are not associated with markers or haplotypes at high levels of  $r^2$ .

When others and we implemented genomic selection, using the markers as if they were QTL, we found that the accuracy of the prediction in a new dataset of unrelated animals is not as good as expected from randomly drawn subsets of the original data. This may be because some markers are correlated with a QTL within one set of families but not in other families. This problem warrants further study to assess its importance and find solutions.

Accurate estimates of QTL effects will require large numbers of animals with marker genotypes and phenotypes. Meuwissen et al. (2001) found that accuracy of GEBV was higher when 2000 records were used to estimate QTL effects than when 1000 records were used to estimate QTL effects, for a trait with a heritability of 0.5. Under less favourable assumptions about the LD structure (such as larger ancestral effective population size), even more records might be needed.

It should be pointed out that the simulation studies which have been conducted to date (e.g. Meuwissen et al. 2001) have used a genome size which is smaller than that of most livestock species (e.g. 10 Morgans were simulated while cattle have a genome of 30 Morgans for cattle). The number of segregating QTL simulated is also usually smaller than the number of QTL which have been estimated to have an effect on a typical quantitative trait (e.g. Hayes et al. 2006). One consequence will be that the effect sizes of the QTL used in simulation may be substantially larger than the effect sizes of real QTL, allowing the effects of simulated QTL to be predicted more accurately than may be possible in a real data set. However, even if the true model of quantitative trait variation approaches the infinitesimal model, with a very large number of QTL each of very small effect, genomic selection will still predict more accurate breeding values than is possible with pedigree and phenotypes alone. This is because genomic selection will exploit the Mendelian sampling that occurs during gamete formation, effectively capturing the realized relationship matrix rather than the average relationship matrix. Villanueva et al. (2005) demonstrated using the realized relationship matrix rather than the predicted relationship matrix in the calculation of breeding values could lead to higher accuracies of selection, using simulation. They proposed that marker information used in this way could offer benefits in selection programmes when no QTL has been mapped or when the underlying genetic model can be considered the infinitesimal model, where no QTL has a significant large effect.

## Implementation of genomic selection

The requirements to implement genomic selection in breeding programmes are relatively simple. Generally there will be a discovery dataset where a large number of SNP have been assayed on a moderate number of animals who have phenotypes for all the relevant traits. A prediction equation that uses markers as input and predicts BV is derived from this data. There should then be a validation sample (which can be smaller than the derivation sample) where a larger number of animals are recorded for the traits and genotyped at least for the markers that are proposed to be used commercially. The prediction equation is tested to assess its accuracy on this independent sample. Then selection candidates are genotyped for the markers and the prediction equation estimated in the discovery data used to calculate GEBV, but their accuracy is assumed to be that found in the validation sample. In practice, the process may be more complex but the distinction between discovery, validation and selection candidates is still useful. For instance, it makes clear that the estimation of QTL effects can be carried out on animals that are completely separate from the selection candidates. In fact the selection candidates do not need to have phenotypes recorded at all. As discussed next, this could lead to large changes in the structure of the livestock breeding programmes. We will refer to the combined discovery and validation datasets as the 'reference' population.

## Implementation of genomic selection in national genetic evaluation schemes

In the near future, there will be DNA marker data as well as phenotypes and pedigrees on potential selection candidates. It might be desirable to combine all this data to estimate improved EBV. This has the great advantage that the livestock breeder does not need to interpret masses of marker data but simply selects on EBV as at present. However, there are technical difficulties in implementing this approach. The major difficulty is that, for the near future at least, the amount of marker data available on each animal will be highly variable, with most animals having none. Three methods can be suggested to calculate EBV using such data. The first is to calculate traditional EBV from phenotypes and pedigrees and genomic EBV from markers separately and use a selection index to combine the two EBV on each animal into one final EBV for use. This is an approximate solution but is easily implemented. The second

method is to infer marker genotypes for all animals and use these to calculate genomic EBV. This may be the long-term solution when DNA marker data is widespread. The third method is that described by Meuwissen and Goddard (1999) which absorbs the equations for animals that do not have markers. This can be done provided these animals do not have genotyped descendants. The second and the third methods could be combined by inferring genotypes on the necessary ancestors and absorbing ungenotyped descendants.

If marker genotypes are inferred on all animals, the accuracy of the EBV is not as critically dependent on the accuracy of the inferred genotypes as might be thought. For instance, if no animals had actual genotypes and all were inferred, the inferred genotypes would simply replace the pedigree-derived relationship matrix in the calculation of EBV, provided enough QTL were contributing to the trait.

In time we believe DNA marker data will be available on most animals but the same markers will not be available on all animals. Therefore, the second method will be very attractive but there will be a need for a method of inferring missing genotypes that is very computationally efficient.

Cattle and sheep producers in most developed countries currently have EBV that compare all the available breeding stocks. In the case of dairy cattle, Interbull provides EBV that compare bulls from around the world. This situation is favourable to livestock producers but is threatened by genomic selection. Unless all genomic EBV are calculated using the same prediction equation(s), they will no longer be comparable. A country is most likely to preserve a single EBV system if the database used to estimate the prediction equation is held by the national genetic evaluation system. However, this may inhibit commercialization of genomic selection because it makes it difficult for a company to recover the investment needed to develop and market a set of markers for genomic selection. The alternative would be for each company to have its own markers, database and EBV. How the future will unfold is not clear.

### Implications of genomic selection for the design of breeding programmes

Genomic selection has the potential to radically alter the structure of livestock breeding programmes. For instance, in dairy cattle, it will be used initially to select young bull calves for progeny testing. However, if it is successful, sires of sons and sires of

replacement heifers will be selected based on genetic markers and formal progeny testing will disappear. This would potentially cut the cost of operating dairy breeding companies by 92% (Schaeffer 2006).

More generally, genomic selection will cause a tendency to shorten generation intervals because the markers can be genotyped at birth or even before. This will lead to the use of reproductive technology to decrease the age at first breeding and increase number of offspring (e.g. juvenile *in vitro* embryo transfer; Armstrong et al. 1997).

Genomic selection separates the animals in the reference population, which is used to estimate the prediction equation, from the selection candidates. The animals in the reference population must have phenotypes and genotypes but the selection candidates need to have only marker genotypes. Thus, there could be a huge saving in the cost of producing stud animals because there will be no need to record pedigrees or phenotypes. Except for the cost of genotyping, 'stud' animals could be produced as cheaply as commercial animals. Then traditional stud breeders who record their animals with a breed association would be unable to compete with breeders who relied entirely on marker data.

The separation of the reference population from the selection candidates has other effects. The reference population could be composed of commercial animals that are extensively recorded including traits not usually recorded on stud animals (alternatively, the stud breeders may enter the reference population and be paid for their recording). This might include carcass and meat quality traits, feed consumption, response to disease challenge and performance under commercial management conditions. For instance, in developing countries, the reference population could be managed under typical village conditions rather than in an atypical stud farm. If commercial animals are usually cross-breds, the reference population could be cross-bred and the merit of QTL alleles in a cross-bred animal estimated. This should increase the rate of genetic gain for commercially relevant traits.

The existence of even small genotype  $\times$  environment interactions ( $G \times E$ ) affects current breeding programmes. If the genetic correlation between production in two environments, such as two countries, is  $<0.8$ , different animals tend to be selected for the two environments. If animals can only be evaluated for their 'own' environment, then the two populations diverge. However, if all animals can be evaluated for production in both environments, then animals suited to either environment can be selected

from one population and the overall effective population size is smaller (Goddard 1992). Genomic selection means that all selection candidates can be evaluated for any environment for which the prediction equation is known. Consequently, many divergent populations, each specialized to a particular environment, are likely to be replaced by one general purpose population with a decrease in the total effective population size. For instance, some countries might abandon their own Holstein breeding programme and simply buy bulls from the rest of the world based on a prediction equation derived in their own conditions.

### Use of DNA markers to predict the phenotype

Markers can be used to predict genetic value including non-additive genetic effects such as dominance and epistasis, and hence phenotypic value. Xu and Jia (2007) described a genomic selection approach for the detection of epistatic QTL. Although this prediction will always be imperfect because of environmental effects it still might be useful. For instance,

- Animals might be bought and sold based on estimated phenotypic values (EPV) derived from the markers.
- Animal products, such as meat and milk, might be paid for based on genetic markers.
- Animals might be allocated to management options or environments based on genetic markers.
- Animals might be mated to achieve favourable non-additive gene combinations.
- DNA markers could be used to determine pedigree.
- DNA markers could be used to trace animals and their products.

Given these many uses, we believe DNA marker profiles will become widely used in livestock in the near future as the cost decreases and the benefits increase. In fact, a major research objective may be to make best use of this DNA data in commercial animal production.

### Conclusions

We have described genomic selection as a variant of MAS but the change from one to the other is likely to have major effects on the research agenda and commercialization of the technology. Research on MAS tended to focus on mapping a few QTL precisely in the hope that the gene that was the QTL could be identified. This often involved testing SNP in candidate genes based on their physiological

action. Genomic selection does not attempt to identify functional mutations but uses a random set of genome-wide markers to predict breeding value. The hope is that, because this method will track all the genetic variance, it will yield accurate EBV even without phenotypic measurements on the selection candidates. The results so far are encouraging but need more validation.

In the short-term, use of genomic selection will complicate calculation of EBV especially by national genetic evaluation systems. However, in the long-term we believe that models for estimation of breeding values and genetic values will be based entirely on DNA markers obtained from genotyping the animal or inferred from other animals as was envisaged by Goddard (1998). The ideal method to estimate the effect of individual markers or QTL and hence the breeding value of the animal is to compute the conditional mean of the posterior distribution of QTL effects. This requires an estimate of the prior distribution of QTL effects.

Widespread use of DNA markers will have a major impact on the structure of the breeding programmes and a significant impact on production systems more generally. Breeding animals will be reared cheaply with minimum recording of phenotypes and pedigree. Selection will be based on a prediction equation derived from a reference population that has extensive phenotypic recording and genotype data. To take maximum advantage of the genomic selection, generation intervals will be shortened as much as reproductive technology will allow.

### References

- Armstrong D.T., Kotaras P.J., Earl C.R. (1997) Advances in production of embryos in vitro from juvenile and prepubertal oocytes from the calf and lamb. *Reprod. Fertil. Dev.*, **9**, 333–339.
- Beavis W.D. (1994) QTL analysis: power, precision and accuracy. In: A.H. Paterson (ed.), *Molecular Dissection of Complex Traits*. CRC Press, Boca Raton, pp. 145–162.
- Boichard D., Fritz S., Rossignol M.N., Guillaume F., Colleau J.J., Druet T. (2006) Implementation of marker assisted selection: practical lessons from dairy cattle. In: 8th World Congress on Genetics Applied to Livestock Production, August 13–18. Belo Horizonte, MG, Brasil.
- Calus M.P.L., Meuwissen T.H.E., de Roos A.P.W., Veerkamp R.F. (2008) Accuracy of genomic selection using different methods to define haplotypes. *Genetics*, (inpress).
- Chamberlain A.J., Bowman P.J., Meuwissen T.H.E., McPartlan H.M., Goddard M.E. (2007) Estimating the

- distribution of QTL effects for milk production traits in dairy cattle. *Genetics*, e-pub ahead of print; doi: 10.1534/genetics.107.077784.
- Dekkers J.C. (2004) Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *J. Anim. Sci.*, **82** E-Suppl, E313–E328.
- Goddard M.E. (1991) Mapping genes for quantitative traits using linkage disequilibrium. *Genet. Sel. Evol.*, **23**(Suppl. 1), 1315–1345.
- Goddard M.E. (1992) Optimum effective population size for the global population of black and white dairy cattle. *J. Dairy Sci.*, **75**, 2902–2911.
- Goddard M.E. (1998) Gene based models for genetic evaluation – an alternative to BLUP? *Proc. 6th World Congr. Genet. Appl. Livest. Prod.*, **26**, 33–36.
- Goddard M.E., Hayes B.J. (2002) Optimisation of response using molecular data. *Proc. 7th World Congr. Appl. Livest. Prod.*, **33**, 3–10.
- Goddard M.E., Hayes B., McPartlan H., Chamberlain A.J. (2006) Can the same genetic markers be used in multiple breeds? 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Brazil, **8** pp. 22–14.
- Grapes L., Dekkers J.C., Rothschild M.F., Fernando R.L. (2004) Comparing linkage disequilibrium-based methods for fine mapping quantitative trait loci. *Genetics*, **166**, 1561.
- Hayes B.J., Chamberlain A., Goddard M.E. (2006) Use of linkage markers in linkage disequilibrium with QTL in breeding programs. *Proc. 8th World Congr. Genet. Appl. Livest. Prod.*, Belo Horizonte, Brazil, **8**, pp. 30–06.
- Hayes B.J., Chamberlain A.C., McPartlan H., McLeod I., Sethuraman L., Goddard M.E. (2007) Accuracy of marker assisted selection with single markers and marker haplotypes in cattle. *Genome Res.*, in press.
- MacLeod I.M., Hayes B.J., Savin K., Chamberlain A.J., McPartlan H., Goddard M.E. (2008) Power of dense bovine single nucleotide polymorphisms (SNPs) for genome scans to detect and position quantitative trait loci (QTL). *Genetics*, (inpress).
- Meuwissen T.H.E., Goddard M.E. (1999) Marker assisted estimation of breeding values when marker information is missing on many animals. *Genet. Sel. Evol.*, **31**, 375–394.
- Meuwissen T.H.E., Goddard M.E. (2007) Multipoint IBD prediction using dense markers to map QTL and estimate effective population size. *Genetics*, **176**, 2551–2560.
- Meuwissen T.H.E., Hayes B.J., Goddard M.E. (2001) Prediction of total genetic value using genome wide dense marker maps. *Genetics*, **157**, 1819–1829.
- Meuwissen T.H.E., Karlsen A., Lien S., Olsaker I., Goddard M.E. (2002) Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics*, **161**, 373–379.
- Parkes M., Barrett J.C., Prescott N.J., Tremelling M., Anderson C.A., Fisher S.A., Roberts R.G., Nimmo E.R., Cummings F.R., Soars D., Drummond H., Lees C.W., Khawaja S.A., Bagnall R., Burke D.A., Todhunter C.E., Ahmad T., Onnie C.M., McArdle W., Strachan D., Bethel G., Bryan C., Lewis C.M., Deloukas P., Forbes A., Sanderson J., Jewell D.P., Satsangi J., Mansfield J.C., Cardon L., Mathew C.G. (2007) Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat. Genet.*, **39**, 830–832. DOI: 10.1038/ng2061
- Schaeffer L.R. (2006) Strategy for applying genome-wide selection in dairy cattle *J. Anim. Breed. Genet.*, **123**, 218–223.
- Tibshirani R. (1996) Regression shrinkage and selection via the Lasso. *J. R. Statist. Soc., Ser. B*, **58**, 267–288.
- Villanueva B., Pong-Wong R., Fernández J., Toro M.A. (2005) Benefits from marker-assisted selection under an additive polygenic genetic model. *J. Anim. Sci.*, **83**, 1747–1752.
- Weller J.I., Shlezinger M., Ron M. (2005) Correcting for bias in estimation of quantitative trait loci effects. *Genet. Sel. Evol.*, **37**, 501–522.
- Whittaker J.C., Haley C., Thompson R. (1997) Optimal weighting of information in marker-assisted selection. *Genet. Res.*, **69**, 137–144.
- Wray N.R., Goddard M.E., Visscher P.M. (2007) Prediction of individual risk to disease from genome-wide association studies. *Genome Res.*, **17**, 1520–1528.
- Xu S. (2003) Estimating polygenic effects using markers of the entire genome. *Genetics*, **163**, 789–801.
- Xu S., Jia Z. (2007) Genome-wide analysis of epistatic effects for quantitative traits in barley. *Genetics*, **175**, 1955–1963.
- Zhao H.H., Fernando R.L., Dekkers J.C.M. (2007) Power and precision of alternate methods for linkage disequilibrium mapping of quantitative trait loci. *Genetics*, **175**, 1975–1986.