



EDITORIAL

How far and how fast can genomics take us towards precision breeding?

All sectors of animal breeding need to deliver more complex objectives. In livestock sectors, the challenges are for more outputs for less inputs, reduced environmental impact per unit of product, improved animal health and welfare. Where breeding companies have global aspirations, these have to be delivered in widely different environmental conditions. With companion animals, especially dogs, there is a public demand for objectives of health and welfare to be prioritized more firmly ahead of form. Addressing complex objectives requires recording that is sufficient for their scope, an understanding of the interrelationships between the traits that are either being recorded or in the objectives, and the sharing out of the selection intensity across the criteria, most commonly using an index. In short, complex objectives demand precision breeding. This is particularly true where the breeding pyramids that evolve multiply the deviations from what is desired.

So what do we need to set as targets to advance the practice of precision breeding? Flint and Woolliams 2008 (*Phil Trans Royal Soc Series B Biol Sci* **363**: 573) provide three goals. First, the 'quantomics' aim, of increasing the scope and precision of predictors of the outcomes of breeding, which clearly encompasses within it the prediction of breeding values. Second, a capacity to avoid the introduction and advance of characteristics deleterious to animal well-being or, more generally, the well-being of the species. Third, managing the genetic variance between and within breeds in accord with the Convention of Biological Diversity and, more recently, the Interlaken Declaration and Global Plan of Action. At first sight, it is paradoxical to set a target of maintaining variation for the purpose of precision breeding, but given the dynamics of breeding objectives, the management of variation underpins the continuity of precision. This special issue contains papers (references without journal name) that advance towards these goals, primarily the first and third of these goals.

For the most part, the papers in this special issue are concerned with using genomic information. There are three notable papers concerned with geno-

mic evaluation, from Aguilar *et al.*, Goddard *et al.* and Meuwissen *et al.*, which are overlapping in their advances, and tackle the first of the precision breeding goals. It will be a common problem, for example where schemes are an open or extended nucleus, for important data to be available on animals that are not genotyped as well as on animals that are. Achieving the highest possible accuracy demands using both sources of data and a natural first step to achieving this is through the use of what is termed 'single-step' GBLUP. Aguilar *et al.* and Meuwissen *et al.* both highlight the computational challenges that are faced as we move away from the rule-based inversion of **A** (pedigree-based relationship matrix). However, the primary deliverable from the paper of Meuwissen *et al.* is a novel framework for unbiased genomic EBV (gEBV) in a 'single-step' approach. Unbiasedness of EBV has come to be expected by the breeders who use the (g)EBV, and the presence of bias in gEBV has the potential to slow the uptake of genomic evaluations, especially where candidates have varying amounts of information – historically, the removal of bias by the adoption of BLUP was a major advance over what had gone before. One of the important concepts used by Meuwissen *et al.* concerns the correction for errors in relationships that are estimated from genomic data and the regression back to the pedigree (expected) relationship which was originally developed by Goddard *et al.* in their very interesting paper. This paper gives a theoretical underpinning of the properties of the constructed genomic relationship matrix, **G**, demonstrates that the way **G** is constructed does matter and provides deterministic tools for calculating the true accuracy of prediction. On a different note, Cole *et al.* increase the scope of the predictions made from genomic evaluations by decomposing the gEBV to predict limits to selection in dairy cattle. The authors sum the haplotype effects estimated from the genomic evaluations. To what extent the additivity of the effects will be sustained as selection progresses is open to question, but this assumption is also testable over time, and the emerging answer will be very intriguing.

A further set of papers (de Gara *et al.* and Engelsma *et al.*, Cervantes *et al.*) address the third goal of precision breeding, in managing genetic variation in a population. The first two of these explore the value of genomic data for maintaining genetic variance in a conserved closed population. Both papers show that there are benefits compared to pedigree, but that these can be small. De Gara *et al.* demonstrate that the step change in the density of SNPs across the genome has been important in giving this benefit, as low-density SNP performed worse than pedigree. Engelsma *et al.* demonstrate that where pedigree is unreliable or absent then SNP data are potentially very important. Cervantes *et al.* explore further criteria for managing the genetic variance observed in a population, primarily by changing the degree of assortative mating.

The continuing development of genomics is fundamental to advancing precision breeding, as pedigree and phenotypes can only take us so far. In particular, genomic evaluations should allow for faster progress towards breeding objectives, in directions that are less restricted by varying heritabilities among the traits and the differential availability of records. It will allow epidemiological traits of susceptibility and infectivity to be better addressed in the absence of continuing epidemics, providing sufficient recording is made when epidemics are in progress. Beyond this, genomics will be essential for expanding the scope of our prediction equations as urged in the first goal for precision breeding. In particular, there is the hope that we may be able to proactively predict genetic correlations to our selection decisions by a better understanding of how genotypes are built upon in the development of phenotypes. Underlying this challenge is the regulation of gene expression, including its genetic component. The idea of an eQTL is already well established, where QTL influence mRNA abundance associated with a target locus, but more recent evidence suggests widespread mQTL may exist, where the QTL affect the methylation status of a target locus (e.g. Gibbs *et al.*, 2010 *PLoS Genetics* **6**: e1000952). These fields of genetical genomics and genetical epigenetics are made more

feasible by 'next generation' techniques, which are capable of producing vast quantities of data even on a small number of phenotypes.

However, the volume of data will in turn place a challenge for the quantitative scientists in delivering the next generation of genomic tools. One aspect of this challenge will be the dimensionality alluded to above, with $p \gg n$, where p and n are the number of parameters and phenotypes, respectively. This problem has been tackled successfully in some approaches to genomic evaluations, but building a comprehensive model of expression will require dealing with a possible parameter space that is increased by further orders of magnitude. How can this be performed? It seems to me that success will rely on scientists from other fields providing informative prior information upon how to structure models. For example, this is already performed in genomic analysis, where the predictions of SNP effects frequently assume additive effects; this is justified by utility, the observations of additive genetic variance, and an underpinning theory that suggests that other components of the broader genotypic variance are likely to be small in comparison (e.g. as explored recently by Hill, Goddard & Visscher 2008 *PLoS Genetics* **4**: e1000008). In the context of being proactive over the genetic consequences of breeding, the priors will build upon (i) richer information to be extracted from sequence data, including the use of comparative genomics (ii) *in vitro* genetics, providing more readily attainable phenotypes for hypothesis testing and validation, and (iii) cell and systems biology improving our physiological understanding of pathways. This will not be achieved soon, but if we aspire to precision breeding then it needs to be tackled, and we need quantitative biologists to ensure the predictions will have validity beyond the test data.

John A. Woolliams
The Roslin Institute & R(D)SVS,
University of Edinburgh
E-mail: john.woolliams@roslin.ed.ac.uk