Australian Government

Department of Agriculture and Water Resources

AUSTRALIAN Pork™

Industry information paper on genomic selection

Final Report
APL Project 2015/064

February 2017

Animal Genetics and Breeding Unit

Rob Banks and Susanne Hermesch

University of New England

Armidale NSW 2351

Acknowledgements

This project is supported by funding from Australian Pork Limited and the Department of Agriculture and Water Resources.

Executive Summary

Genomic methods involve reading DNA to some level of precision and analysing data to estimate associations between genotype and phenotype. These methods are being increasingly widely used in most farmed species of livestock and plants, because they offer increased accuracy of estimation of genetic merit for individuals that have limited phenotypes available at the time of selection. Depending on the species, this provides scope for earlier and/or more accurate selection and hence faster genetic progress.

While there are a range of approaches to using DNA information in estimating genetic merit, the most widespread approach depends on having a reference population of animals with both performance records (phenotypes) and SNP data (genotypes) recorded. This reference population provides the basis for estimating genetic merit for the recorded traits based solely on the genomic similarity between candidate animals and the reference population. This approach is known as genomic selection.

Implementing genomic selection almost invariably increases costs of a breeding program, mainly due to the cost of genotyping. Where additional traits are recorded, this will also increase costs, but this is to increase the power of genomic selection, rather than to make it possible in the first instance. Accordingly, careful benefit-cost analysis should be conducted prior to introducing genomic selection.

The key facts about genomic selection can be listed as:

a) It provides scope to increase the accuracy of selection, and to enable more accurate selection in animals that have limited phenotypic information available at selection.
b) A reference population of animals with the relevant phenotypes and genotypes is essential and it must be maintained over time. Its size determines the benefit arising from genomic selection which becomes an essential cost to the breeding program.
c) Implementation of genomic selection requires considerable expertise in using appropriate software to analyse the extensive data and in designing cost-effective breeding programs.
d) The estimated additional gains from implementation of genomic selection can be considerable. Increases in accuracy of selection for profit between 10-50%, with corresponding increases in rate of progress, can be obtained.
e) Economic benefits for value chains or vertically integrated operations can be substantial, but will usually require increased investment in recording and genotyping. Provided these additional costs are converted to faster genetic progress, and this additional progress is harvested through a large enough commercial sector, the investment will be profitable.

There is scope for improvement in rates of genetic progress in Australian pig populations through use of genomic selection. Determining the extent of this scope will require some platform R&D around the economics of genomic selection at the individual enterprise or company level, and around the genetic relationships between herds in the Australian industry. Depending on the effective breeding size of individual herds, and on the strength of those relationships, a pooled or industry approach to developing a reference population may be needed in order to exploit the power of genomic selection for the Australian industry.

Table of Contents

List of Tables

List of Figures

1. Background Information about Genetic Evaluation Systems

This paper aims to provide a basis for pork industry consideration of the opportunities and challenges around genomic technologies, to stimulate discussion, and to outline some potential strategies for R&D and implementation at the industry and enterprise level.

The starting point for this discussion is to describe the current standard approach to genetic improvement in pigs. This is important in order to pinpoint what can be different when genomic technologies are available and what the consequences of those differences are for genetic improvement.

1.1 Principles of genetic evaluations

The core of genetic improvement is estimating the genetic merit of candidate boars and gilts, to assist in selecting those which will be used as parents of the next generation in the breeding nucleus or herd, and which in turn generate animals used to breed commercial pigs. Estimating genetic merit simply means using clues of various types, pedigree information, performance records and genomic information, to work out the value of each animal's genes for the traits we are interested in, and performance records for.

The standard method of estimating genetic merit in pig breeding in Australia, as with other countries and indeed most livestock industries in the developed world, is Best Linear Unbiased Prediction (BLUP) methods for analysis of pedigree and performance data. When applied diligently, these can enable genetic change at around 1% of trait means for single traits or for combinations of traits. These methods are used in Australia either through use of PIGBLUP or in-house software. Walters (2006) reported on the effectiveness of these tools used for the Australian pig industry and the benefit-cost ratios of APL-funded research in genetic improvement systems. The benefit-cost ratios of the APL investment in genetic improvement tools varied from 2.2:1 to 6.5:1 due to new users of BLUP technologies and varied from 4.5:1 to 9.9:1 due to higher genetic gains in existing users. Adoption of genomic tools will increase genetic gain which will have to be compared to the costs of implementing new technologies that use genomic information. Further, Walters (2006) predicted that individual breeding companies will become larger and the number of seed stock suppliers will decrease over time due to increasing competition to supply fewer, larger producers. Walters (2006) stated that 'those that remain will increasingly need specialist genetic and technical advice to 'stay in the game'. Although genomic selection was not specifically mentioned, this statement is certainly true for incorporating genomic information in pig breeding programs effectively and economically.

Genetic evaluations based on BLUP analyses generate Estimated Breeding Values (EBVs) for the recorded traits, usually expressed in units of the trait such as g/day for growth rate, mm for backfat and muscle depth traits or number of piglets born alive for reproductive traits of sows. The records used in the analyses may be taken on the selection candidates themselves prior to selection, as is the case for growth traits or backfat or muscle depth traits recorded on finisher pigs prior to selection. Alternatively, records may be collected on relatives of the selection candidate animals, as is the case for carcase or meat quality attributes recorded on

full- or half-sibs of selection candidates or reproductive records of sows recorded on dams and other female relatives of the selection candidate. Genomic selection tools will be most useful for traits that have limited information available for selection candidates prior to selection.

To apply BLUP methods, it is essential to have records of each animal's pedigree, and some performance data. Because both pedigree and performance recording can be expensive and logistically challenging to obtain, such recording is frequently restricted to animals in a breeding nucleus. In the nucleus, the animals with the highest genetic merit expressed in EBVs will be used as parents of the next nucleus generation, and those with lower merit as parents of commercial animals, or to breed boars for use as sires of the commercial animals. This structuring, based on recording and genetic evaluation concentrated in the nucleus and genetic improvement in each generation then disseminated into the commercial pig population to produce slaughter pigs is well established in pig (and poultry) breeding. This approach is economically extremely efficient, as long as the traits contributing to profit are recorded so that genetic improvement can be made in those traits, and the multiplication of genetic merit from nucleus to commercial population is high.

1.2 Why then are livestock industries taking up genomic selection?

Existing BLUP methods are very successful and reliable, and when implemented in appropriate breeding and production structures, can support very good economic returns. Why then the interest in genomic methods?

Genomic selection in general can be thought of as the ability to read the DNA makeup of an individual, and subject to the availability of data on some aspect of performance, to make an estimate of the genetic merit of the individual for that aspect of performance.

This underpins two main types of opportunity for the use of genomic tools:

a) Genetic or DNA tests for specific conditions
b) Using genotypes to estimate genetic relationships between and genetic merit of pigs

2. Objectives of the Research Project

The project will deliver a review and information paper on opportunities and challenges around genomics and specifically genomic selection for the Australian Pork Industry, including:

- A high-level outline of the nature of the technologies,
- Summary of reports on applications of genomics in pigs, and key findings from other species,
- Outline of potential benefits to the Australian industry, and to individual breeding and production enterprises, together with potential implementation challenges at the enterprise and industry levels,
- Outline of potential R&D that would enhance industry evaluation of the opportunities, and/or help assess potential benefits,
- Bibliography of key references and information sources,
- Recommendations for industry actions.

3. Introductory Technical Information about Genomic Tools

3.1. Genetic or DNA tests for specific conditions

If genes or regions of the genome can be identified that are responsible for specific conditions, a genetic test can be developed. Here, specific conditions means some specific attribute, such as some inherited deformity leading to death, some metabolic weakness, or in some cases something beneficial, which is known or discovered to be due to or caused by one gene. A well-known example in pig breeding is the 'Halothane gene' which leads to improved leanness as well as inferior pork quality and increased risk of mortality in stress situations. The mutation in porcine ryanodine receptor associated with malignant hyperthermia was discovered by Fujii et al., (1991) and a commercial test was available for this gene. The effect of the Halothane gene on performance was evaluated in Australian pigs (McPhee et al., 1992; McPhee et al., 1994; Luxford, 1995). An overview of commercial gene tests was provided to Australian pig breeders by Rothschild (2010) as shown in Table 1. This overview demonstrates that multiple gene tests are available affecting economically important traits for pork production.

Table 1. Molecular Genetic Tests Used by the Pig Industries (from Rothschild, 2010).

Gene or test	Industry Use
Parentage tests	exclusive use within some companies, commercially available
HAL	meat quality, commercially available
ESR, EPOR	litter size, commercially available
KIT	white colour - exclusive use
MC1R	red/black colour, use unknown
MC4R	growth and fatness, commercially available
FUT1	edema E. coli F18, exclusive use
RN	meat quality, commercially available
AFABP, HFABP	intramuscular fat, use unknown
PRKAG3	meat quality, commercially available
CAST	tenderness, commercially available
IGF2	carcass composition, commercially available
Trade secret tests	several traits – many companies

Genomic technologies have made it much easier to locate such regions, where a single region or gene is controlling the trait. So, developing genetic or DNA tests for recessive disorders is now easier and faster, as was recently observed in one cattle breed where it took only a month or two to develop a new gene test for a recessive disorder.

It is important to note that development of such tests requires good phenotypes (i.e. observations or records of the trait) on a number of affected individuals.

It is also important to note that a genetic test that is very accurate in one breed or population may not automatically be as useful in another – that will depend on whether the gene affecting the trait or condition works identically in each population. This is by no means guaranteed. Accordingly, some form of validation is needed – finding affected individuals in each population, and determining whether the same gene or DNA region is affecting the condition, and whether to the same extent.

The use of each gene test in breeding programs needs to be evaluated for each case in regard to:

- optimal sampling strategies for genotyping;
- optimal mating strategies to increase or reduce the frequency of the favourable or unfavourable allele in a population;
- optimal testing strategy for performance recording;
- optimal use of the gene test to maximise genetic gain in the overall breeding objective; and
- management of effective population size and inbreeding in the population to ensure long-term genetic gain is possible.

These commercial gene tests offer additional opportunities for Australian pig breeding companies that need to be evaluated carefully in regard to the costs and benefits resulting from the implementation of these gene tests in pig breeding programs.

3.2. Using genotypes to estimate relationships

3.2.1. Identify sire and dam, and use this information in the BLUP calculations

If we can "read" the DNA to some level of precision, it follows that we can determine how similar or different individuals are at the level of the DNA. Another way of saying this is that we can determine quite accurately how similar individuals are in the genetic makeup.

This is potentially important, as it provides an alternative way of obtaining the pedigree information needed for BLUP genetic analysis. In BLUP analysis, we use records on animals' relatives to provide additional clues on their genetic makeup, over what can be recorded on the individuals themselves. It is important to know the pedigree in order to know which animals provide clues for which relatives and how much to weight the information from the various relatives.

3.2.2. Track the shares of an individual's DNA that it has received from its grand-parents via its parents (and this extends out through cousins etc.), in calculating the EBVs

Usually, the pedigree is built up by identifying the sire and dam of each individual or litter, which means that each individual has a pedigree that in principle includes all relatives. This pedigree information is based on units of relationship that are multiples of one-half: progeny get half their genes from their parents, half-sibs share a quarter of their genes, grand-parents and grand-progeny share a quarter in common, and so on. However, in reality this is not quite true. While each progeny gets half its genetic material from each of its parents, the sperm and eggs they receive from those parents have been through meiosis, which involves

recombination, and so it is possible for a sperm to be carrying marginally more DNA from one of the sire's parents (i.e. the grand-parents of the progeny) than the other. In simple terms, the genetic relationship between progeny and their grand-parents is not exactly one quarter. If we know this – by reading the DNA – we can place more accurate weights on each individual in the relationship to each other, and so make more accurate use of the records from those relatives, meaning that EBVs will be more accurate. The extent of this increase in accuracy is around 2-3% improvement.

3.2.3. Take account of genetic similarity for identifiable regions between members of the same population, in calculating the EBVs

When we use pedigree information in estimating genetic merit (whether based on identifying sire and dam, or on examining the similarity across the DNA), we are relying on the assumption that individuals that are related share genes in common, and the more related they are, the more genes they share. In simple terms, we assume overall similarity at one half, or one quarter and so on. However, not all regions of the DNA contribute equally to a trait – there are genes "for" traits (although as will be discussed in the next section, there are usually very many such genes). If we can track regions or chunks of DNA down through the pedigree, in principle we can take more accurate account of the regions that are affecting a particular trait.

This provides two opportunities. If we can identify regions or chunks of the DNA that have a detectable effect on a trait, we can weight that information in the data analysis, and that will help us more precisely estimate individuals' genetic merit. More simply, we can identify where even distantly related individuals share regions in common, and pick up some extra clues about genetic merit. So, even if two individuals have no direct pedigree connection, if they come from the same population and have a specific region of DNA in common, and our data tells us that that region has some detectable impact on performance, we can use the knowledge that individuals are similar at that region, in calculating the EBVs. Essentially, those 2 individuals are "related" at that region.

3.3. Using genotypes to estimate genetic merit

3.3.1. Maintaining a suitable reference population is critical

The population of individuals with performance records and genotypes is usually referred to as the reference population. This is very important because the second and third approaches outlined above only provide potential benefit when we have recorded phenotypes (i.e. trait records) and genotypes on substantial numbers of individuals in the population. In simple terms this is because the effects we are trying to detect and use – the differences from the one-half, one-quarter etc. relationships, and the shared DNA regions between very distantly related individuals - are relatively small and assuming the standard pedigree relationships (one-half, one-quarter etc.) is a very good approximation of the truth – which is why BLUP works so well.

And for most traits of interest, there are so many genes and DNA regions contributing to performance, that tracking chunks of DNA that are common to related individuals, provides in most circumstances only marginal benefit.

Theory has been developed that allows us to estimate the size of that benefit.

When we have a reference population, we can take the genotype of an animal with no records, and determine its genetic (or genomic, in this situation the terms are inter-changeable) relationship or similarity to the reference population, and estimate its genetic merit, i.e. produce an EBV for it. The accuracy of that EBV will depend on the degree of relationship between the animal in question and the reference population.

The important difference between the pedigree-based approach and the genomic approach is that the latter takes into account all relationships or similarity between the candidate animal and the reference, and this usually provides some extra accuracy of the resulting EBV, compared with that calculated using normal analysis.

Goddard and Hayes (2009, Figure 1) provide a simple graphic illustrating the reference population concept and how it underpins genomic selection:

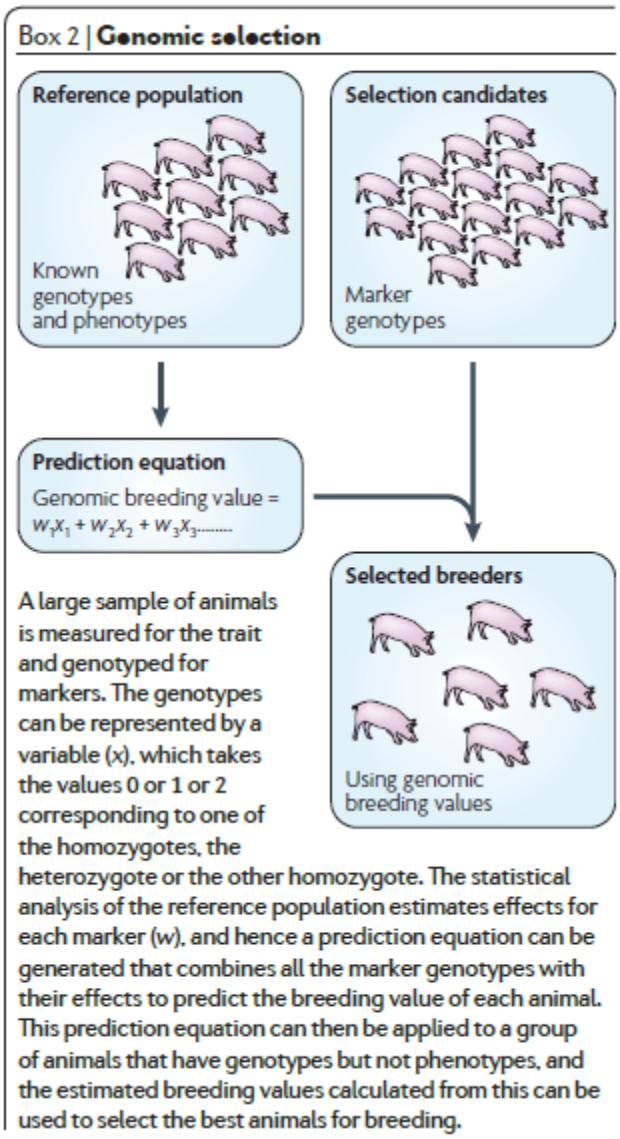

Figure 1. Genomic selection requires a reference population with known genotypes and phenotypes

Thus as long as there is a reference population, in which the appropriate traits are recorded, breeders and producers have the opportunity to estimate animals' genetic merit (i.e. get EBVs), without having recorded anything, but simply had a DNA sample processed. This can mean cost savings - depending on how the cost of the recording is recouped, and operational simplicity (it is easier to take a hair or tissue sample than go to the trouble to record traits). However, the overall costs will depend on the size of the reference population that is required to achieve accurate genomic selection.

Theory has been developed that allows us to estimate how accurate genomic selection will be depending on the size of the reference population, and the trait. The estimates are shown in the following chart (Goddard and Hayes, 2009, Figure 2):

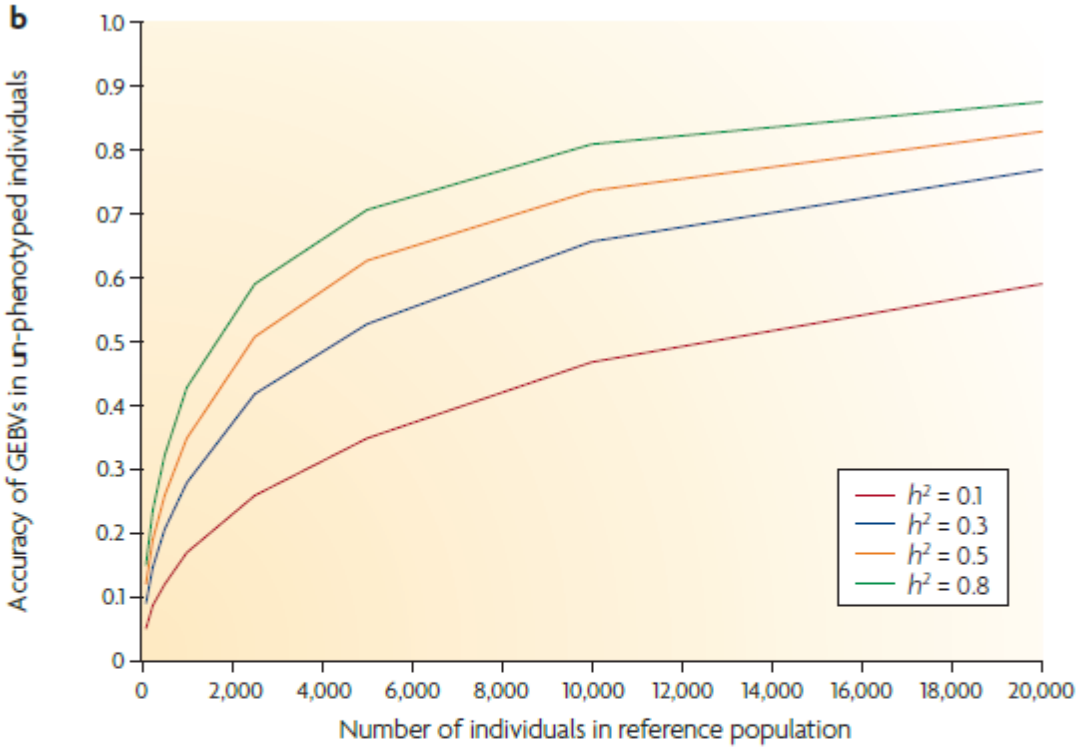

Figure 3 | **Calculation of number of animals in a reference population and accuracy of breeding values. a** | Number of animals needed in a reference population. To achieve an accuracy of 0.7 for estimated genomic breeding values (GEBVs) calculated from SNPs requires an increasing number of animals in the reference population as the heritability declines or the N_e of the population increases. **b** | Accuracy of GEBVs of un-phenotyped individuals with increasing number of phenotype records in the reference population used to estimated SNP effects, for different heritabilities (h^2). N_e was 100.

Figure 2. Number of animals required in the reference population to achieve a certain accuracy of estimated genomic breeding values (GEBVs) depends on the heritability of a trait.

This shows us that the reference population needs to be larger for lowly-heritable traits, such as health and fertility traits, in order to achieve a certain accuracy of genomic estimated breeding values (GEBVs) of individuals with no performance records. For example, in order to achieve accuracy of EBV of 50% for animals with only a genotype, for a trait with heritability of 30% we need a reference population of about 5,000 animals from the same population with phenotypes and genotypes recorded.

A fundamental requirement of the reference population is that it relates to the current population under selection. This reflects the fact that the reference population data helps us establish relationships or associations between genotype and phenotype, and these relationships change under selection, and can be different between different populations, even within the same breed.

The change in the relationships under selection basically means that the accuracy of selection using genomic methods will decline over time, unless the reference population is "refreshed" with new data at regular intervals. There is no precise or set method for doing this, but a rule

of thumb applied world-wide is to ensure that the reference population has been completely replaced over two generations. So for example, if the reference population is 4,000 to 6,000 individuals genotyped and phenotyped, and the generation length is 2 years, 1,000 to 1,500 new individuals should be genotyped and phenotyped each year.

These calculations have been extended in pig breeding for selection of purebred pigs to improve performance of crossbred populations (e.g. Cleveland et al., 2010). In this situation of developing a reference population to predict genetic merit for a second population (B) in order to select in the parent population (A), the reference population needs to either include individuals from both B and A, or solely B. Therefore, using genomic selection for genetic improvement of crossbred animals requires substantial organisational changes because it requires genotyping and phenotyping of crossbred pigs possibly from multiple farms, genotyping of purebred selection candidates and tracing the pedigree to connect crossbred pigs with purebred pigs.

Such a scheme was implemented by French breeding companies for the Pietrain breed. The genetic analyses of these data conducted at INRA (Tusell et al., 2016) found that the genomic model which accounted for genomic information was superior to the model that used pedigree-based relationships only. However, the more complex genomic model that used both purebred and crossbred pigs (reference population A and B) did not outperform a simpler genomic model that used either purebred (A) or crossbred (B) information to predict performance of purebred or crossbred pigs supporting findings by Cleveland et al., 2010). Tusell et al., (2016) concluded that the implementation of the genomic model 'is straight forward with available software but its use under field conditions needs to be further addressed in terms of predictive ability, genetic progress achieved, and costs.'

3.3.2. Improvements in accuracy due to genomic selection

The difference in accuracy of selection between the genomic approach and the pedigree-based approach is summarised in this chart:

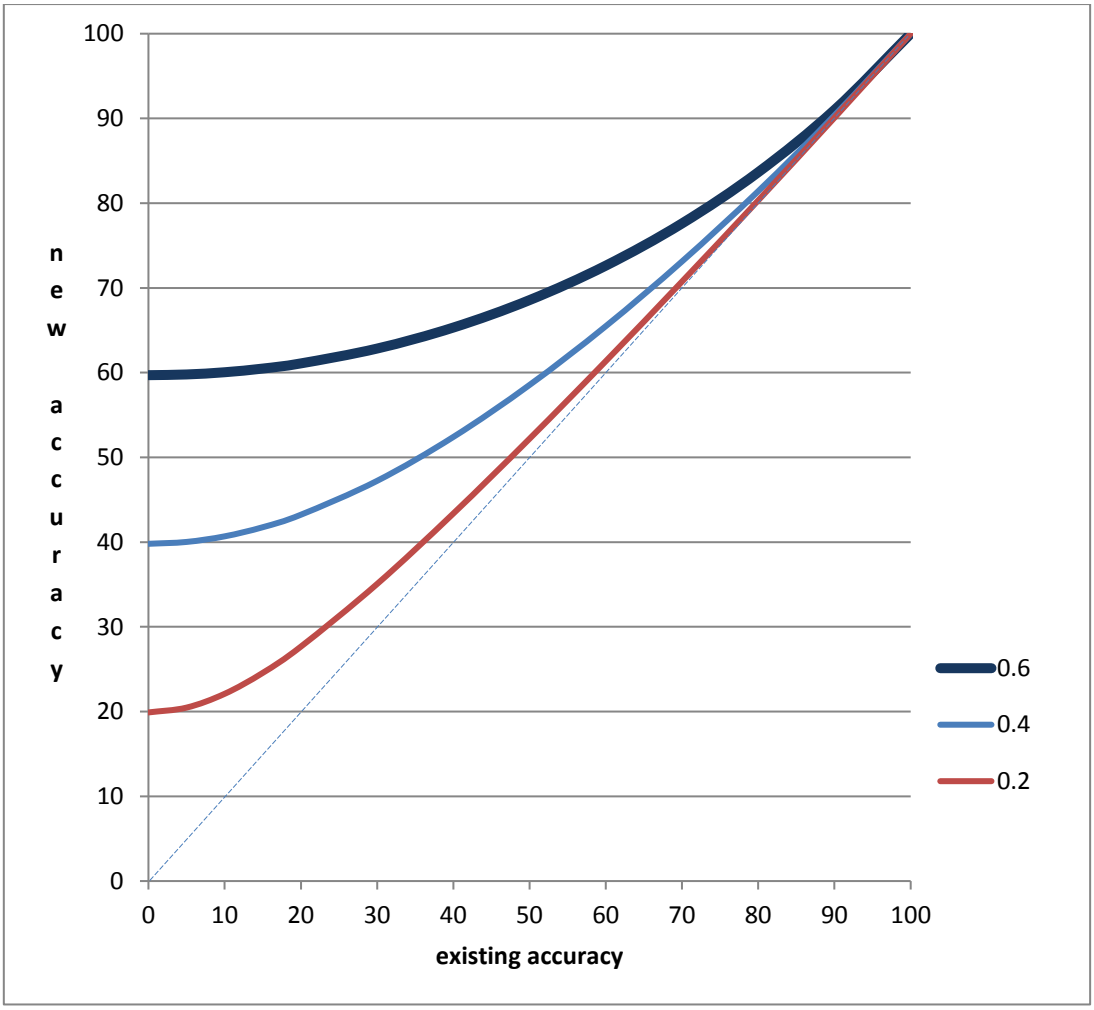

Figure 3: Improvement in accuracy of EBV from a genomic analysis (new accuracy) for accuracy of EBV from analysis using pedigree alone (existing accuracy) based on different genomic accuracies (0.2, 0.4 and 0.6), Source: D. Johnston, AGBU, personal communication.

The x-axis shows EBV accuracy from analysis using pedigree alone. The y-axis shows accuracy of EBV from a genomic analysis. The three lines are for genomic EBV accuracy for three traits with genomic accuracies of 20%, 40% and 60% which are determined by the size of the reference population. So, for example, for a trait with genomic accuracy of 40% (light-blue line), for an animal with existing (i.e. pedigree-based) EBV accuracy of 30%, the EBV accuracy it would get from a genotype alone would be about 45%. The gap between the solid lines and the dotted line is the advantage in accuracy from genotyping. This gap is what offers an advantage – in that animals' genetic merit can be more accurately estimated than when based on traditional BLUP methods alone.

3.3.3. Methods to use genomic information in genetic evaluations

Finally, for this background section, it is useful to outline how genomic information is used in the analysis to estimate breeding values. Collecting genotypes adds a potential source of information for our analysis to estimate genetic merit – meaning that we can have three types of information – pedigree (either traditionally recorded, or determined using DNA analysis), performance records, and genotype.

The prediction equations to predict breeding values for animals are derived from the reference population – animals with recorded phenotypes and genotypes. We can then use the prediction equation to calculate a breeding value (sometimes referred to as a Direct Genomic Value, or DGV) from animals' genotypes.

Accordingly, if we have a reference population, and genotypes of candidate animals, they can have three sources of information – pedigree, performance records, and DGVs. There are two approaches to combine these.

3.3.3.1. Blending genomic values into EBVs

Here there traditional BLUP EBV for each animal, and its DGV, are added together, weighted by the respective accuracies of each. The accuracy of the DGV is calculated from the reference population.

Here a prediction equation is generated from which the genetic merit (or breeding value) of individuals is calculated based solely on their genotype.

This approach has the advantage that it is relatively simple to calculate. At the same time it has the disadvantage that it assumes that all animals with DGVs have the same genetic relationship with the reference, and so their DGVs are weighted equally when being combined with the traditional EBVs.

3.3.3.2. Single Step Analysis

In this approach, the traditional BLUP analysis is modified by replacing the values in the pedigree with the actual genetic relationships amongst individuals derived from their genotypic similarities.

This properly accounts for all the relationships amongst genotyped individuals, and means that genotyped and non-genotyped individuals are analysed together. Accordingly, it is more accurate than the blending approach. At the same time, it is completely dependent on having a suitable reference population, and the advantage in accuracy over traditional BLUP depends on the size of the reference population.

This analysis is computationally more intense than traditional BLUP, but software is available to conduct such analysis here and overseas.

3.4. Definitions and terms (partly from Hu et al., 2011)

This section provides a brief glossary of key terms relevant to genomic technologies.

Additive genetic effects: the effect of an allele on animal performance, independent of the effect of the other allele at a locus. These effects of the two alleles at a locus add up (thus "additive"). Alleles at a locus may have other effects (*dominance*, *epistasis*), so that there are not genes that have just "additive" effects and other genes with only "dominance" effects. Additive genetic effects can be inherited; other genetic effects such as dominance and epistasis are the result of allele combinations that are lost between generations. The additive genetic effect that an animal has for a trait is equal to its breeding value.

Alleles: one of a pair, or series of alternative forms of a gene that can occur at a given locus on homologous chromosomes. Genes can exist in more than one form, and when this occurs, we refer to the different forms as alleles of the gene. For a very simple example, using the imaginary sequence of bases above, GCCTTTTACGGAAAA, if this was actually a defined gene (it is actually far too short for that), and we could also find individuals with GCCTTATACGGAAAA – and this made some difference to the phenotype, we would refer to the forms of the gene containing the T and A at this location, as alleles of that gene.

Animal model: a system for genetic evaluations that estimates breeding values of individual animals (males, females) at the same time. The system uses production data on all known relatives in calculating a genetic evaluation.

Bases, base pairs, nucleotide bases: a strand of DNA is made up of 2 strings of paired molecules, which are of 4 types – Guanine (G), Cytosine (C), Adenosine (A) and Thymine (T). The bases are always paired G-C (or C-G) and A-T (or T-A). Reading DNA results in strings like – GCCTTTTACGGAAAA.

Blended EBVs: a method of using genotype results to add information to traditional BLUP EBVs. In this method, Direct Genomic Values (DGVs) are added to the traditional Estimated Breeding Values (EBVs), weighted by the accuracy of the DGVs, which is assumed the same for all genotyped animals.

BLUP: the standard method of analysing livestock breeding data in order to estimate genetic merit. It uses pedigree relationships along with phenotype records in the analysis. It is particularly appropriate when performance data come from genetically diverse contemporary groups.

DNA: Deoxyribonucleic acid, the chemical material which carries information to code for a gene.

Direct Genomic Values (DGVs): estimated breeding values calculated using genotype data alone, summing the value of each SNP in the genotype based on analysis of the reference population data.

Dominant: applied to one member of an allelic pair of genes, which has the ability to express itself wholly or largely at the exclusion of the expression of the other allele.

Dominance genetic effects: the effect that an allele has on animal performance, which depends upon the genotype at the locus. For example, the "a" allele may have a different effect on animal performance in "aa" animals than in "Aa" animals. See *additive genetic effects*.

EBV (Estimated Breeding Value): these are estimates of the genetic merit or value of individuals, usually expressed in the units in which we measure the trait. Note that EBVs are not estimates of the genetic make-up – what genes the individual actually has; rather, EBVs are estimates of the value of those genes for the trait(s) of interest.

Functional traits: there is no strict definition of this term, or distinction between these traits and production traits, but functional traits is a term frequently used for traits that are more related to how the animal functions, than with aspects of the animal that directly lead to income or cost. The main point or purpose of differentiating functional and production is

to highlight that recording is frequently focussed on production traits, perhaps because they are usually somewhat simpler to record. The use of this term overlaps also with "hard-to-measure" – see next point.

Genes: definable or detectable lengths of DNA, that provide the information template for the construction of amino acids and proteins, and via this process, underpin all growth and development. This is a somewhat simplistic way of thinking about genes, since the pathways between gene and phenotype is usually extremely complex, involving many more than simply one gene, and almost invariably involving many interactions and feedback loops between genes and between genes and the external environment.

Genetic distances: a measure of gene differences between populations (hence genetic relationships among them) described by some numerical quantity; usually refer to the gene differences as measured by a function of gene frequencies.

Genetic drift: changes in gene frequency in small breeding populations due to chance fluctuations.

Genetic gain: the amount of increase in performance that is achieved through genetic selection after one generation of selection.

Genetic maps: See *linkage map.*

Genetic marker: a gene or DNA sequence having a known location on a chromosome and associated with a particular gene or trait; a gene phenotypically associated with a particular, easily identified trait and used to identify an individual or cell carrying that gene.

Genetic variance: variation in phenotype which results from variations in genetic composition among individuals.

Genotype: the genetic constitution of one or a few *gene*(s) or *locus* (*loci*), or total genetic make-up (genes) of an individual organism.

Genome: another term for the genetic makeup of the individual or species, essentially meaning the chromosomes of the individual or species, which are strings of DNA including defined sequences that are referred to as genes.

Genomics: the study of all the gene (DNA) in each cell or organism

Genotype: the genotype of an individual can have two meanings. In the general sense it is another way of saying "the genetic makeup of the individual" – it's the actual sequence of base pairs along its chromosomes. In the narrow sense, it is the result obtained when we read the DNA – again it is a sequence of identified base pairs, but may not be the whole sequence, depending on the SNP density (see below).

Hard-to-measure traits: traits that are not routinely or easily recorded in commercial populations, whether because of the expense of recording – individual feed intake, or aspects of meat quality, for example, and/or the time taken for data to accumulate – lifetime piglets weaned, for example.

Linkage: association of genes physically located on the same *chromosome*. A group of linked genes is called a *linkage group*.

Linkage map: a linear map of an experimental population that shows the position of its known genes and/or genetic markers relative to each other in terms of *recombination frequency*.

Locus, pl. loci: a fixed position on a *chromosome* occupied by a given *gene* or one of its *alleles*.

Major gene: a gene that has an easily recognizable and measurable effect on a characteristic.

Map: refers to knowing what genes are where along the DNA, or what SNPs are where.

Marker: specific and identifiable sequences of the DNA molecule. These markers may or may not be functional genes.

Marker—assisted selection (MAS): selection for specific alleles using genetic markers.

Pedigree: usually refers to *pedigree chart* or what a pedigree chart represents in genetics. It is a document to record the ancestry of an individual. A pedigree can also be used to illustrate the family structure or breeding scheme.

Phenotype: the phenotype of an individual is what we see or measure. For example, the weight of an animal at a particular age is a phenotype, the number of piglets born in a litter is a phenotypes, the intra-muscular fat % is a phenotype, and so on. The phenotype is determined by the genotype of the individual plus factors such as the amount and quality of the diet, as well as interactions between the genotype and external factors such as nutritional level.

Prediction Equation: an equation in which each of an identified series of SNP is given a weighting according to its effect on the trait, and the summed value is the genetic merit for the individual, based on its genotype alone. The prediction equation is used to calculate estimated breeding values based solely on genotype, known as Direct Genomic Values (DGVs).

Production traits: this is a simple catch-all term for traits such as growth rate, carcase leanness, litter size, and feed intake. Essentially it refers to traits that directly affect income or cost.

Qualitative trait: a trait that can generally be classified into a limited number of categories, and the animal can be said to "possess" the quality or not. Examples include hair colour, skin colour, and ear stature.

Quantitative trait: a trait that is represented by an almost continuous distribution of measurements. Examples include average daily gain, backfat thickness, and height.

Quantitative trait locus (QTL): a *locus* that affects a *quantitative trait*.

Reference Population: a sample or group of animals from a population that has been recorded for one or more traits, and also genotyped. The term reference is based on the fact that the patterns of relationship between the genotypes (the SNP sequences) and observed differences in phenotype act as a reference when we genotype a new individual – essentially

look for the useful or informative patterns which we know from our reference, in that new individual.

Single step analysis: a method of analysing pedigree, performance and genotype together, to generate an EBV for each animal. In this method, the relationship between each animal and the reference population determined from its genotype is taken into account.

SNPs (Single Nucleotide Polymorphisms): the explanation above for Alleles leads directly to SNPs, or Single Nucleotide Polymorphisms. In the example above, if we can detect or read the DNA at the point at which we can find T or A, that is a polymorphism (Greek for many forms) at a single nucleotide. Reading SNPs just means reading the DNA at particular defined points along the DNA.

SNP Chip: this is a glass slide engraved with large numbers of very small pits which can hold chemical reagents which bind to DNA at specific locations, and which can then be detected in genotype reading machines. In genotyping, a sample of the DNA of an individual essentially coats the area of the SNP chip with the pits, and this is then read after appropriate treatment to achieve the binding reactions. SNP chips come in various sizes, which basically refers to how many SNP they can test for – the number ranges from around 200 up to millions.

4. Using Genomic Selection in Breeding Programs

This section outlines three ways in which genomic tools are being used in genetic improvement of livestock, with brief discussion of what is actually generating the advantage.

4.1. Dairy cattle breeding – genomic selection to overcome the tyranny of time

The dairy industry has seen the most widespread implementation of genomic technologies to date, with essentially all programs in developed countries now including genomic selection to some, often large, extent.

Here the "classical" model of genomic selection is in operation:

- There is a reference population, in the form of a genotyped population of bulls with highly accurate EBVs for a range of (usually) production traits such as milk volume, fat and protein production. The bulls' EBVs are highly accurate because they are based on large numbers of daughters with milk records, reflecting the high level of AI use in most dairy industries, coupled with widespread milk recording (farmers collecting data on milk production) coordinated into central databases. These high accuracy bulls are genotyped using moderate to high density SNP chips (50k, 800k).
- Young bulls (around 1 year-old) are then genotyped and obtain genomic breeding values, using prediction equations developed from the reference population. These young bulls are then collected for AI use, and start generating daughters immediately.
- The accuracy of the genomic EBVs can be very high, depending on the number of bulls in the reference population and the accuracy of their own EBVs.

By contrast, the "traditional" model of dairy genetic improvement involves using a team of young bulls to generate a limited number of daughters in commercial herds via AI, and then waiting for those daughters' milking records to underpin breeding values for the bulls – breeding values that are based on progeny test. From those progeny test breeding values, the best of the bulls are then chosen to be the parents of a new crop of elite young bulls, via mating to cows with high EBVs.

This approach essentially involves two sorts of bulls:

- Young "unproven" bulls, used to breed progeny test daughters, and selected essentially on mid-parent EBVs (around 35-40%);
- Older "proven" bulls, used to breed the next crop of young bulls, selected on the basis of their daughters' performance, therefore with high accuracy (80% or higher).

The main feature of this traditional model is the long wait to obtain the breeding values based on daughters' production - meaning that the selection of the best bulls to become parents of the next crop of bulls takes approximately seven years.

In contrast, under the genomic selection approach, young bulls become parents essentially as early as is possible, and there is no distinction between young unproven bulls and older proven bulls – bulls are simply chosen on EBV, with the EBVs of young bulls being based on genomic information, with accuracies around 60-65% for young bulls with no daughter information.

What are the "pros" and "cons" of moving to genomic selection for dairy breeding?

The "pros" are:

- The reduction in the generation interval on the male side more than makes up for the accuracy of selection for young bulls – with only genomic information – being lower than that of the older "proven" bulls in the traditional approach. The result of this is potential for increases in rates of progress of up to 50% (Schaeffer, 2006) – dependent of course on the accuracy of the genomic EBVs.
- The predominant use of young bulls means that breeding companies do not have to hold onto large numbers of bulls waiting for their daughters' records to come through. This is a substantial cost under the traditional approach – as an example, one of the main breeding programs in Australia was holding approximately 150 bulls for 6-7 years, with all the associated management costs, and generating semen sales at the end of that period from a very small fraction of them.

The "cons" are:

- The accuracy of genomic selection depends on the amount of data collected on the cows in the population. Typically, the data collected is focussed on milk production and composition (kg of milk, kg of fat, kg of protein). Data collection for other traits is much less systematic and does not automatically allow accurate estimation of EBVs. For example, recording of data for traits related to fertility is poorer, and for traits such as feed intake is essentially absent.
- This raises the risk that the "balance" of selection, between production traits (milk, fat, protein) and functional traits such as health, fertility, feed intake, rather than being improved by using genomic selection, can in fact be worsened. If the reference population data is only for production traits, genomic selection will mean that progress for these traits is accelerated, but there is no improvement in the rate of progress for functional traits.

This in turn means that the design, funding and maintenance of the reference population become critical. Dairy breeding programs around the world are grappling with how to ensure that reference populations capture data on the "hard-to-measure" traits.

- More technically, under genomic selection it becomes possible for the breeding of bulls to be much faster than the accumulation of data on the performance of their close relatives – which is essential for accurate genomic selection.

This means that the reference population has to be very large for confidence in the genomic EBVs. Partly for this reason, breeding companies are tending to share datasets in order to pool as much reference data as possible.

This is also a driver for exploration of models for genomic selection informed by identifying the actual genes affecting traits – the reasoning being that using the animals' genotype for actual genes will be more stable in terms of estimating merit than using SNPs that happen to be somewhere near the actual genes. This approach is the subject of intense research in dairy

cattle and other species, and it is not yet clear how much improvement in terms of accuracy of genomic EBVs over time might be achieved.

Dairy breeding has switched almost completely over to genomic selection, reflecting the technical advantage of potential for faster genetic progress, coupled with the very significant cost reduction due to not having to hold large numbers of adult bulls for many years. Having genomic EBVs on young bulls has given industry confidence to use young bulls, where previously, a high proportion of dairy farmers avoided using young bulls as AI sires because of the perceived risk associated with their lower accuracy, mid-parent EBVs. Interestingly, it was suggested many years ago that the dairy industry would be better off using young bulls – it would achieve faster progress – by an expert looking at that industry from a pig breeding perspective (Bichard, 2002).

What messages are there for pig breeding in the dairy experience? The most significant message is around the design of the reference population: if that population captures records/data for hard-to-measure traits, then selection of young bulls and cows can be more accurate for these traits than is possible under the traditional progeny testing model, can be made earlier, and hence genetic progress can be faster and better balanced.

The other message extends the first, in that scale is important: the accuracy of genomic selection rises as the size of the reference population rises. Breeds with small populations struggle to obtain much benefit from genomic technologies. If such breeds are carefully recording production and functional traits (i.e. including any important but hard-to-measure traits), they can still obtain some benefit by using the genomic relationship rather than the traditional pedigree in estimating breeding values, but this benefit will be likely in the range 3-5% improvement in accuracy of selection.

4.2. Meat sheep in Australia – incorporating hard-to-measure traits into genetic improvement programs

Lamb production in Australia is based on a simple crossbreeding structure, in which terminal sire breed (such as Poll Dorset and White Suffolk) rams are mated to either Merino, or Merino-cross ewes. In broad terms it is similar to the breeding structure of pork production.

Genetic evaluation is well established, for terminal sire breeds focussed on improving growth rate, leanness and muscling and for maternal breeds those traits plus fertility/reproduction and to varying extents wool production.

Selection for the slaughter phase traits – growth rate, leanness and muscling – has been based on weights and ultrasonic fat and muscle depth. This is again similar to selection in terminal sire lines in pigs.

This approach has not included direct measurement of either carcase composition or meat quality to any significant extent. It has been known since the beginning of application of genetic evaluation in lamb in Australia (starting in the late 1980s), that there would be some risk of reducing meat quality if selection for leanness "went too far".

Industry began to tackle this risk directly around 2010, by establishing a reference population in which meat quality traits (as well as many others) would be measured directly. This reference population, known initially as the Information Nucleus, involved:

- mating sires from the 3 main breed groups (terminal, maternal and Merino), to either Merino or Merino-cross ewes;
- using approximately 5,000 ewes in flocks in 7 locations (providing the basis for exploration of the extent of genotype by environment interaction);
- mating approximately 100 sires per year to these ewes for 3 years;
- genotyping all progeny and sires; and
- taking meat samples for a proportion of lambs, for taste panel assessment under Meat Standards Australia protocols.

The overall cost of the Information Nucleus holding the reference population and genotyping in commercial research flocks was approximately $5 million over 6 years funded through the Sheep CRC. The genotyping costs are now being covered by a combination of breeders and MLA.

This reference population dataset provided the basis for development of new EBVs for sheep:

- lean meat yield, a direct measure of the amount of lean tissue in the carcase, as a more accurate measure that a prediction derived from weight and sub-cutaneous fat depth;
- shear force of the meat, measured by Warner-Bratzler (Warner, 1928); and
- intra-muscular fat % ("marbling") of the meat.

During the period 2014-2016, the lean meat yield (LMY) and intramuscular fat content (IMF%) EBVs were generated by single trait, single step calculation, and EBVs provided to industry for those sires with progeny measured in the reference population.

In 2016, the calculations were switched over to full multi-trait, single step analysis, and the LAMBPLAN analysis now generates EBVs for these traits for all animals.

How significant is this? We can answer this question in two ways. The first is to examine the genetic trends for the growth, carcase composition and meat quality traits under the selection using standard LAMBPLAN information (Figure 4):

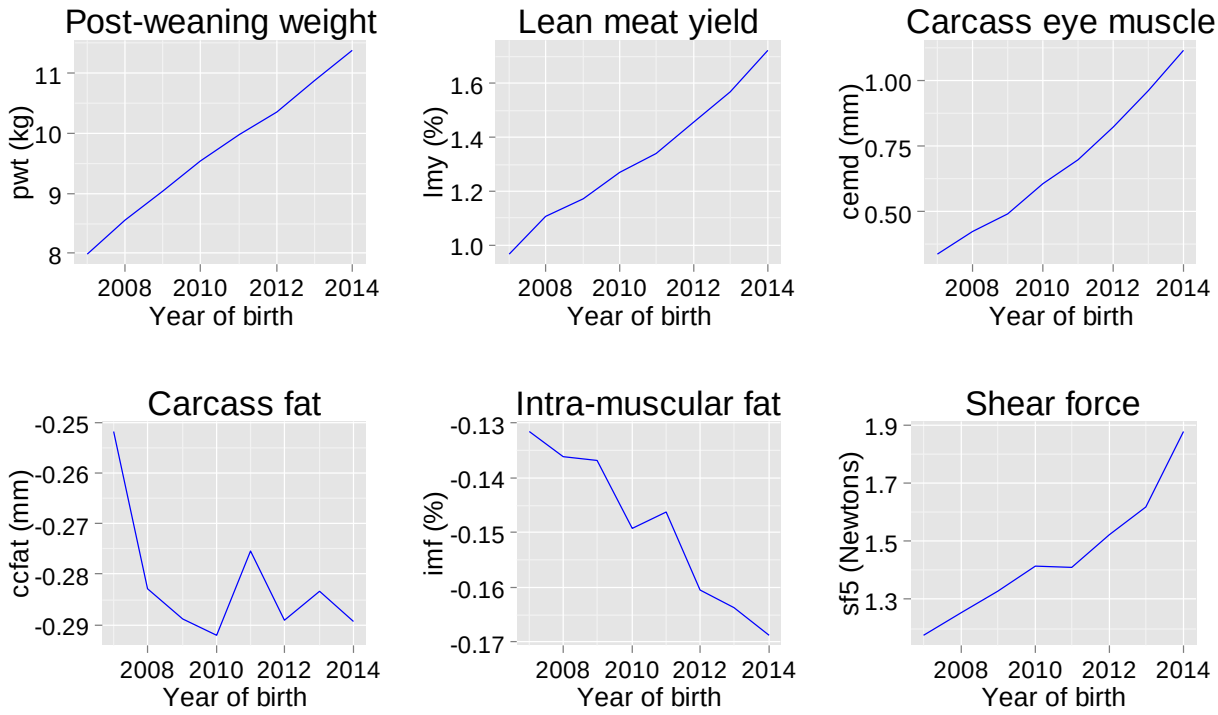

Figure 4. Genetic trends for the growth, carcase composition and meat quality traits under the selection using standard LAMBPLAN information.

These charts show that selection using the basic growth and sub-cutaneous fat and muscle measures was very effective – substantial genetic change was achieved in those 3 traits, and in lean meat yield.

At the same time, we now know that there was a slow but significant correlated response of declining IMF%, and a clear correlated response in increasing shear force, or toughness. These two trends can now be estimated because of the reference population data, and the charts for these two traits are those in the bottom right above – intramuscular fat has declined and shear force has increased.

From an industry perspective, neither of these trends had taken lamb outside a "consumer comfort zone", but if continued, would increase the prevalence of tougher and dryer meat.

The second way of exploring the significance of the developments is to examine the potential rate of direction of progress now that genomic EBVs are available for LMY, shear force and IMF% (Figure 5).

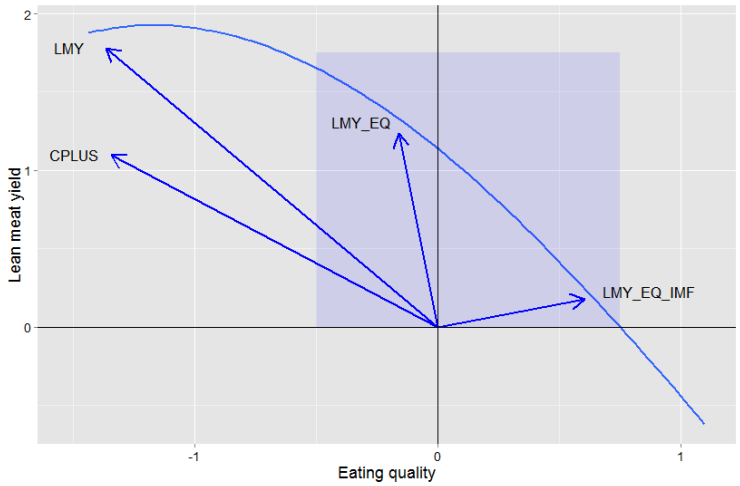

Figure 5. Potential rate of genetic gain based on genomic EBVs for lean meat yield (LMY) and eating quality (EQ) under 4 selection scenarios.

This chart shows the progress possible in two traits – Lean Meat Yield (LMY) and Eating Quality (EQ - which combines Shear Force and Intramuscular Fat% into an index of consumer EQ score), under 4 selection scenarios (CPLUP, LMY, LMY_EQ, LMY_EQ_IMF, Swan, 2016):

- CPLUS: selection using EBVs for growth, subcutaneous fat and eye muscle depth alone as selection criteria. This will result in genetic progress in the positive direction LMY, but backwards for EQ. This scenario mimics what has happened over the last 25 years.
- LMY: is selection as for CPLUS but with LMY added to the breeding goal, and with LMY EBVs available via genomic testing. This results in faster progress for LMY than under CPLUS, but no difference in outcome for EQ.
- LMY_EQ: this is as for LMY, but with EQ added to the breeding goal, and EQ EBVs available via genomic testing. This scenario results in less rapid progress in LMY, but minimal change in EQ.
- LMY_EQ_IMF: as for LMY_EQ, but with added emphasis on IMF in the breeding goal, and again with EBVs for LMY and the EQ traits available via genomic testing. This essentially maximises progress for EQ, but at the expense of taking LMY in the wrong direction.

The results for these scenarios have two key messages:

- By collecting data on both LMY and EQ traits, industry has more capacity to monitor genetic change, and the basis for changing the direction of selection away from LMY alone
- The data collected in the reference population provides the basis for generating genomic EBVs, which can be used in selection of rams and ewes for LMY and EQ direct, rather than indirectly via EBVs for growth, leanness and muscling. Clearly it is not possible to directly measure live selection candidates for either LMY or the EQ traits – so by collecting data on the reference population, and using genomic testing to estimate genetic merit for these traits, direct selection on them becomes possible.

Meat sheep breeders are now conducting genomic testing on young rams and ewes, and through this genomic selection, obtaining the EBVs for LMY and EQ traits, and industry has now developed new indexes which include these traits. Accordingly, the lamb industry is now better able to balance genetic progress for LMY and EQ.

4.3. Examples from the pork industry world-wide

A number of review or summary papers are available that discuss the application of genomic selection in pig breeding (Newman, 2013; Henryon et al., 2015; Ibanez-Escriche et al., 2014; Knol et al., 2016; Samore and Fontanesi, 2016). Genomic selection is expected to increase genetic progress by reducing the generation interval as is the case in dairy or by increasing accuracy of genetic predictions at the same age (Knol et al., 2016; Samore and Fontanesi, 2016). In pigs the largest impact of genomic selection can be derived from the increased accuracy of genetic predictions and for the possibility to predict maternal traits in boars and pork quality traits in selection candidates.

The size of the training population and its closeness to the prediction population is crucial for the precision of genomic selection. A drop in accuracies is expected to occur already after the first generation and regular re-training is required in closed populations such as chicken (Wolc et al., 2011) or pigs.

Different SNP panels exist with high or low density. Samore and Fontanesi (2016) discuss the development of different porcine SNP panels. The effectiveness of SNP panels for genomic predictions depends on the level of linkage disequilibrium between markers and QTL (Hayes et al., 2009). Pig populations in Australia have been maintained separately from pig populations overseas and more specific SNP panels may be required for Australian pig populations in order to improve accuracies of genomic predictions.

Knol et. al., (2016) provide a comprehensive overview of genomic technologies with particular focus on commercial pig breeding. The paper includes some comparison with the applications in dairy, poultry and potato breeding, in each case highlighting the advantages that genomics provides for that industry, and comment on whether that advantage applies in pig breeding.

They then provide 3 specific examples in pigs:

- Selection for teat number, where genomic testing provides scope for increased accuracy of selection;
- Post-weaning mortality, where genomic selection provides scope for increased accuracy of selection, and the authors comment on the challenges involved in continuing collection of reference data; and
- Using single step information for culling after the first litter, where genomic selection again provides scope for increased accuracy of selection, but care must be taken in a) collecting the appropriate data to underpin genomic accuracy, and b) in weighting the genomic information in the evaluation.

They also provide data from a single selection line where rate of progress in selection index was able to be increased by 50% with use of genomic information. In comparison, the increase in accuracies or reliabilities of genetic predictions varied from 0 to 91% for maternal traits

and from 8 to 39% for performance and other traits in the review of studies by Samore and Fontanesi (2016).

The concluding remarks by Knol et al., (2016) point to some of the research challenges still inherent in genomic methods, particularly around whether and how to weight known genetic markers, and on how best to use data from crossbred reference populations in genetic evaluation of parent and grand-parent lines. Cleveland et. al., (2010) investigated this question for a commercial breeding company situation, including estimating the improvement in accuracy of selection in the parent and grand-parent lines.

Newman (2013) provided for a beef industry audience an overview of the development of implementation of genomic methods in a large pig breeding organisation. Similarly to Knol et. al., (2016), he showed significant increases in selection accuracy across a wide range of traits. He also outlined some of the operational challenges involved – around the development of reference population recording, the sampling and storage of tissue and DNA samples, and the computing requirements.

The overall message around genomics for the pig industry is that increases in rates of progress are available, even when existing recording and selection procedures are good. The most significant challenge is in ensuring that appropriate reference data is being collected.

Genomic selection has been implemented successfully in dairy cattle where the costs of genotyping a few high-value bulls are low relative to the commercial value of AI bulls. This situation is very different in pigs where the production cost of the animal is relatively low and the costs of genotyping effectively across the pool of selection candidates is a more important issue (Knol et al., 2016, Samore and Fontanesi, 2016). The economic aspects of implementing genomic evaluations in a pig sire line breeding scheme were evaluated by Tribout et al., (2013) who concluded that "implementing genomic evaluations was the most efficient approach when major expenditure [annual extra cost of over $300,00] was possible, whereas increasing phenotypes was preferable when limited resources were available".

5. Genomic Pilot Studies conducted at AGBU in Pig Breeding

AGBU has conducted some initial scoping studies of the value of genomic methods for the Australian industry. An initial exploration of the potential value of using genomic methods for the Australian industry was conducted by Dominik et al., (2016) including the following features:

- A crossbreeding structure with 2 maternal and a sire line;
- Breeding objectives that mirror those in use in Australian breeding operations;
- Basic recording included average daily gain, carcase fat depth, post-weaning survival and daily feed intake measured on selection candidates;
- A reference population of 1,000 animals;
- BLUP selection, without and with recording of carcass traits on half-sibs of selection candidates. The carcass traits measured did not include EQ traits;
- Genomic selection, with and without recording of carcass traits on half-sibs of selection candidates. The carcass traits measured did not include EQ traits;
- Recording costs representative of industry practice, including a realistic cost for genotyping; and
- It was assumed that the breeding program represented the whole Australian pig industry as one commercial population to recover the costs of the breeding program.

The results can be briefly summarised:

- Recording the carcass traits on unselected half-sibs of candidate animals increased accuracy of selection for the breeding objective by 5% in the sire line under BLUP selection.
- Recording feed intake records increased accuracy for the breeding objective of selection by 13% in the sire line and by 3% in the dam line. The benefits of recording feed intake is lower in dam lines because it has a lower economic importance in dam lines. The larger increase in dam lines was expected because in dam lines sex-limited traits and traits that are expressed late in life (longevity, sow mature weight) are important.
- Use of genomic methods increased selection accuracy by approximately 35% for the maternal lines, and approximately 23% for the sire line.
- Costs of recording increased only slightly when carcase traits were recorded (8%). Recording feed intake increased costs of recording by 400% (4 times higher) in comparison to the simple scheme while genomic selection lead to an increase of costs by 1,500% (15 times higher).
- Economic returns over a 10-year period increased by approximately 50% through use of genomic selection based on the assumption that the commercial population is representative of the whole Australian pig industry.

Two features of the model underpin these results:

- The population was modelled as a single integrated breeding program, with a nucleus in which all recording and selection was focussed – this is technically and economically highly efficient. However, this scenario does not exist currently and would require close collaboration between Australian breeding companies to achieve one national breeding program.
- A reference population of 1,000 animals was established, resulting in accuracies of genomic EBVs of 0.7-0.8 for male line traits, and 0.3-0.5 for maternal traits in this simulation study.

These features and results highlight conditions under which genomic selection can be advantageous:

- The breeding objective includes hard-to-measure traits which can be recorded in a reference population which is large enough that the accuracy of selection can be increased compared to that achievable under more conventional recording approaches, including when half-sibs are slaughtered for carcass trait assessment.
- The additional costs imposed by genotyping can be spread over a sufficiently large commercial population.

It is worth considering whether these results might be relevant for pork quality traits, not modelled in this study. These traits are likely to be unfavourably correlated with growth and carcass composition traits, especially with leanness, and can only be measured on slaughtered animals. Accordingly, they could be included in conventional BLUP selection as long as meat samples were collected on half-sibs of selection candidates. However, if this recording was a part of a reference population, the additional accuracy benefit of genomic selection would be available. This would allow balanced selection for production efficiency and pork quality to be made at almost certainly faster rates than would be possible under conventional BLUP selection – with the advantage dependent on the size of the reference population and the resulting accuracy of genomic selection.

AGBU has also conducted a small study into inbreeding levels in a sample of Australian herds using pedigree and genomic analysis (D'Augustin, 2016), which has demonstrated that:

- Inbreeding is accumulating in herds using conventional recording and BLUP selection, at rates which warrant some careful consideration of options;
- Genotype analysis of a small sample of boars suggests that inbreeding levels estimated using genotypes can be lower or higher than those estimated from pedigree alone – reflecting for example that half-sibs can get more or less similar samples of their grand-parents' genomes than the one-quarter expected on average; and
- Single step analysis of Australian data is feasible – AGBU has conducted small test runs using the dataset collected for this small study.

Together, these examples from dairy cattle, meat sheep and pigs illustrate some important general principles for the potential of genomic technologies for the Australian industry:

- Increased rates of genetic progress can be achieved using genomic selection, arising from a combination of earlier selection being possible and/or increased accuracy of determination of animals' relationships; and
- Better balanced selection – by enabling faster selection for hard-to-measure traits.

At the same time, it is clear that appropriately designed and maintained reference populations are essential.

6. Opportunities for the Australian Pork Industry

The discussion in the Introduction and Overview, and Review of R&D and Implementation, outlines the two main ways in which genomics can assist in animal breeding and production:

- Through development of very precise tests for genetic conditions; and
- Through enabling faster and/or better targeted genetic improvement, either through enabling earlier selection (so reducing the generation interval) and/or enabling selection for traits previously very difficult to change, principally because of the difficulty of measuring the trait(s) involved (Hard-to-Measure traits).

The discussion here will focus on the second of these two ways.

To consider how genomics might help the Australian pork industry, we need to include the question "what is possible now". In simple terms this means considering what traits are selected for now, how much progress is being made for them, and whether there is scope to make faster progress. From that we can consider whether genomics offers advantages.

Genetic gains of 28 pig populations were summarised by Hermesch (2006). The average population achieved an annual genetic gain of $1.3 per pig while the top 25% of herds achieved an annual genetic gain of $2.6 per pig based on today's economic importance of traits. This summary was based on traits that were available in all herds (growth, backfat, feed conversion ratio, muscle depth and litter size). More genetic progress is possible when more traits are considered. In particular survival traits (pre- and post-weaning), maternal effects on progeny growth, sow mature weight and sow longevity are economically important traits (Amer et al., 2014; Hermesch et al., 2014). Most of these traits are candidate traits for genomic selection because they are sex-limited traits and require a long time period until records are available. Assuming that genomic selection may increase genetic gain by 10 to 50%, a higher annual genetic gain of 0.25 to 1.25 $ per pig may be expected for the superior breeding programs. These potential benefits have to be compared with the additional costs of implementing and maintaining genomic selection. The pilot study modelled the Australian industry as if it were one population. This is not the case, in that there are a number of independent breeding programs supplying boars for commercial use. Further R&D should therefore explore ways to develop either individual reference populations, or an industry approach.

To model the individual example, the pilot study could be expanded to cover a range of commercial population sizes and reference population sizes, and including benefit-cost analysis in each case.

At the same time, there is potential value in considering an industry approach. This would involve a number of breeding programs sharing phenotype and genotype data, presumably on a confidential basis, in order to collect together enough records to support accurate genomic selection. This would require an initial scoping step to investigate the degree of genomic relationship across breeding herds. This would involve sampling a number of influential (i.e. high relationship with their own herd) sires per line, genotyping them, and then constructing a relationship matrix across the herds. This would show the level of genomic relationship or

similarity within- and across-herds, and would also provide the basis for predicting the level of accuracy of genomic selection possible.

This proposal may seem idealistic, but has in fact been applied by a number of dairy breeding programs world-wide, from a number of countries, to develop a reference population dataset to underpin selection for feed efficiency in dairy cows (Berry, 2015). The countries/programs ranged widely in the number of animals sampled for genotypes and phenotypes, but all countries obtained a benefit through having access to a larger reference population – meaning that selection for feed efficiency became possible when it would not otherwise have been.

7. Implications & Recommendations

7.1. Use of commercial gene tests should be evaluated better

A number of commercial gene tests are available for multiple traits describing lean meat growth, carcase and pork quality, piglet survival and litter size. The use of these gene tests in breeding programs needs to be evaluated for each case in regard to:

- optimal sampling strategies for genotyping;
- establishing that the gene tests actually affects economically important trait in Australian pig populations;
- optimal mating strategies to increase or reduce the frequency of the favourable or unfavourable allele in a population;
- optimal testing strategy for performance recording;
- optimal use of the gene test to maximise genetic gain in the overall breeding objective; and
- management of effective population size and inbreeding in the population to ensure long-term genetic gain is possible.

The use of these gene tests requires technical expertise to evaluate the costs and benefits of implementing each gene test in a breeding program. Various commercial gene tests have been implemented in some pig breeding programs in Australia. Only the Halothane test has been investigated in scientific studies. The effectiveness of other gene tests in Australian pig populations is not publicly known and has not been evaluated in independent research projects.

7.2. Use of genomic information to better estimate genetic relationships between pigs

The costs of the 80K SNP chip continues to decrease. Evaluation of a first sample of boars indicated that genetic relationships between boars based on genomic information differs from estimates of genetic relationships based on pedigree information. It is not possible to import new genetics into Australia and managing effective population sizes and inbreeding levels in pig populations is important. Genotyping costs per pig are still high (currently $75.00 per sample) and breeders may target boars as a first step to evaluate genetic relationships between boars better. Breeders have been encouraged to collect tissue samples of boars and possibly dams in order to enable future genotyping of pigs of interest. This recommendation continues to be highly important.

7.3. Guidelines for setting up reference populations are required

The reference population is the group of pigs that has both performance data (phenotypes) and genomic information (genotypes) available. The size of the reference population depends on the heritability for each trait and needs to be larger for lowly heritable traits. It is important that the reference population is closely related to the selection candidates and continues collection of phenotypes and genotypes is required for pigs that are part of the reference population. Cost-effective strategies to maintaining reference populations in Australia need to be explored.

7.4. Steps towards implementation of single-step genetic evaluations that use genomic information

Single-step genetic evaluations are now the method of choice for genetic evaluations that use genomic information. Software is available to implement single-step genetic evaluations at AGBU. However, these genetic evaluations are technically more advanced and will require considerable technical expertise for implementation and running of procedures. Individual breeding companies in Australia will require considerable technical support from either AGBU or international collaborators in order to evaluate and potentially implement genomic selection in pig breeding programs. The costs of implementing and maintaining an effective breeding program that incorporates genomic selection will have to be compared with expected returns from the commercial sector.

8. Literature cited

Amer, PR, Ludemann, CI, Hermesch, S (2014) Economic weights for maternal traits of sows, including sow longevity. *Journal of Animal Science* 92, 5345-5357.

Berry, DP, Kearney, JF, McHugh, N, Wall, E, Purfield, D, Cromie, AR (2015) Integration of genomic information into national cattle and sheep evaluations – past, present and future. Proc. Assoc. Advmt. Anim. Breed. Genet. 21: 257-260.

Bichard, M. (2002) *Genetic Improvement in Dairy Cattle – an Outsider's Perspective.* Livestock Production Science 75, p. 1-10.

Cleveland, MA, Forni, S, Garrick, DJ, Deeb, N (2010) Prediction of genomic breeding values in a commercial pig population. In *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production, Leipzig, Germany: August 2010*. Paper 47.

D'Augustin, O (2016) Estimating genetic relationships of Australian boars using pedigree and genomic information. Master thesis, University of Rennes 1.

Dominik, S, Hermesch, S, Boerner, V, and Arnal, M (2016) *Modelling a crossbred selection scheme to study the efficiency of measurement strategies and genotyping in the Australian pig industry* Genetique Selection Evolution (*in preparation*)

Fujii, J, Otsu, K, Zozato, F, Leon, SD, Khanna, VK, Weiler, JE, O'Brien, PJ, MacLennan, DH (1991) Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science* 253, 448-451.

Goddard, ME, Hayes, BJ (2009) Mapping Genes for Complex Traits in Domestic Animals and their Use in Breeding Programs. *Nature Reviews Genetics* 10, 381-391.

Hayes, BJ, Bowman, PJ, Champerlain, AJ, Goddard, ME (2009). Invited Review: genomic selection in dairy cattle: progress and challenges. *Journal of Dairy Science*, 92: 433-443.

Henryon, M, Ostersen, T, Ask, B, Sorensen, AC, Berg, P (2015) Most of the long-term genetic gain from optimum-contribution selection can be realised with restrictions imposed during optimisation. *Genetics Selection Evolution* 47:21.

Hermesch, S (2006) From genetic to phenotypic trends. In '2006 AGBU Pig Genetics Workshop Notes. Armidale'. 59-65. (AGBU).

Hermesch, S, Ludemann, CI, Amer, PR (2014) Economic weights for performance and survival traits of growing pigs. *Journal of Animal Science* 92, 5358-5366.

Hu, ZL, Park, CA, Reecy, JA (2011) Standard Genetic Nomenclature of the Pig, with Glossaries. In 'The Genetics of the Pig, 2nd Edition.' (Eds MF Rothschild, A Ruvinsky.) pp. 473-494. (CAB International: Wallingford, Oxfordshire, UK).

Ibanez-Escriche, N, Forni, S, Noguera, JL, Varona, L (2014) Genomic information in pig breeding: Science meets industry needs. *Livestock Science* 166, 94-100.

Knol, EF, Nielsen, B, Knap, PW (2016) Genomic Selection in Commercial Pig Breeding. *Animal Frontiers* 6, 15-22.

Luxford, BG (1995) Evaluation of the halothane gene in several nucleus lines in a temperate environment, Proc. Austr. Assoc. Anim. Breed. Genet., Vol. 11, 639-642.

McPhee, CP, Daniels, LJ, Kramer, HL, Noble, JW, Trout, GR (1992) The effect of halothane gene on performance, carcass and meat quality in a fat and a lean line of pigs, Proc. Aust. Assoc. Anim. Breed. Genet. Vol. 10, 76-79.

McPhee, CP, Daniels, LS, Kramer, HL, Macbeth, GM, Noble, JW (1994) The effects of selection for lean growth and the halothane allele on growth performance and mortality of pigs in a tropical environment. *Livestock Production Science* 38, p. 117-123.

Newman, S (2013). Use of genomic information in pig genetic improvement programs. In: Integrating DNA Information into Beef Cattle Production Systems. National Beef Cattle Evaluation Consortium, Kansas City, MO. http://www.nbcec.org/producers/workshops.html

Rothschild, MF (2010) Genetic and genomic technologies from A-Z. In '2010 AGBU Pig Genetics Workshop Notes.' (Eds S Hermesch, K Dobos.) pp. 25-38. (AGBU: Armidale)

Samore, AB, Fontanesi, L (2016) Genomic selection in pigs: state of the art and perspectives. *Italian Journal of Animal Science*, 5, 211–232.

Schaeffer LR (2006) Strategy for applying genome-wide selection in dairy cattle. *Journal Animal Breeding Genetics* 123, 218-223.

Swan, A. A. (2016). http://www.sheepgenetics.org.au/Updates/Analysis-changes May 2016

Tribout, T, Larzul, C, Phocas, F (2013) Economic aspects of implementing genomic evaluations in a pig sire line breeding scheme. *Genetics Selection Evolution, 45:40*

Tusell, L, Gilbert, H, Riquet, J, Mercat, MJ, Legarra, A, Larzul, C (2016) Pedigree and genomic evaluation of pigs using a terminal-cross model. *Genetics Selection Evolution* 48:32.

Walters, R. (2006). *Review of APL-Funded Quantitative Pig Genetics in Australia*. Final Report for Australian Pork Limited, pp. 76.

Warner, K.F. (1928). Progress report of the mechanical test for tenderness of meat. *Proceedings American Society Animal Production,* 21, 114–116.

Wolc, A, Arango, J, Settar, P, Fulton, JE, O'Sullivan, NP, Preisinger, R, Habier, D, Fernando, R, Garrick, DJ, Dekkers, JCM (2011). Persistence of accuracy of genomic estimated breeding values over generations in layer chickens. *Genetics Selection Evolution*, 43:23.