19th Jan 09

Genomic Selection Procedures for Ireland (Version 1.1)

Donagh Berry[a] and Francis Kearney[b1],

[a]*Teagasc Moorepark,* [b]*Irish Catlle Breeding Federation.*

Introduction

Genomic selection, as we have come to know it was first described by Meuwissen et al. (2001) and has been described as "the most promising application of molecular genetics in livestock populations since work began almost 20 years ago" (Sellner et al., 2007). It is based on the simultaneous selection for many thousand of single nucleotide polymorphisms (SNPs) that densely cover the entire genome exploiting linkage disequilibrium between the SNPs and the quantitative trait nucleotides. The objective of this document is to clearly outline, step by step, the procedures implemented to generate genomic estimated breeding values (EBVs) in Ireland for Spring 2009. In this document, direct genomic values (DGVs) are breeding values estimated using only genomic information; genomic EBVs (GEBVs) are breeding values estimated after blending DGVs with traditional EBVs.

Collation of biological material and DNA extraction

Collection of semen from dairy and beef AI bulls began in March 2007 with the purchase of 2 liquid nitrogen semen tanks and collection of semen stocks from 34 dairy and beef bulls at Department of Agriculture premises, Abbotstown. This was funded by the Department of Agriculture, Conservation of Genetic Resources for Food, Agriculture & Forestry, Grant Aid, 2006. Semen of dairy and beef bulls was subsequently purchased from breeding organizations, funded by Department of Agriculture, Research Stimulus Fund; The National Cattle Breeding Center donated free of charge semen they had in stock. Requests for semen from Irish farmers were also undertaken with ICBF paying

[1] Many people from many national and international research, breeding and genetic evaluation organisations have contributed to this project

€30/straw; approximately 200 Holstein-Friesian bulls were obtained from farmers. Semen straws from farmers were collated centrally at the ICBF and mailed to the DNA extraction lab at Teagasc, Athenry.

DNA extraction began on the 17th January 2009. All DNA extraction was undertaken in Teagasc Athenry. The protocol for DNA extraction as of January 2009 is given in Appendix 1. Where DNA concentration from a sample was less than 50 µg/ml, an additional sample was requested; where possible 2-3 semen straws were always requested. As of January 2009, DNA from 1,998 semen straws (1,290 unique bulls) was extracted. There was no obvious effect on DNA quality from mailing semen straws to the laboratory although a small proportion of straws were broken. DNA was also extracted by Weatherby's Ireland from hair samples and genotypes compared with genotypes from semen straws. Results from this study are summarized in Appendix 2. DNA was also obtained from collaborators at the Scottish Agricultural College and Poland. DNA from Poland had undergone whole genome amplification.

Genotyping

Genotyping using the Illumina Bovine50 Beadchip was carried out by AROS Applied Biotechnology, Denmark (http://www.arosab.com/). Other companies which offer a genotyping service using the Illumina Bovine50 Beadchip are given in Appendix 2. The Beadchip requires a minimum of 200 ng DNA at a concentration of 50 µg/ml. Samples were sent, in dry ice, in batches of 192 to 384 samples with 400 ng to 1,000 ng DNA per sample. A total of 4 batches of samples were sent (192 samples in January, 288 in July, and two batches of 192 and 386 in October). Throughput was 192 samples per week usually starting the Monday immediately after receiving the samples. The cost of the chip is $145 and the entire cost of genotyping (including cost of chip and hybridizing but not DNA extraction and postage) was $372 (~€272) for the first 192 samples and €210 for the remaining samples. The lower cost after the first 192 samples was because 1,056 chips were ordered after the first 192 samples and the price per chip drops by $37 if purchasing >1,009 chips.

Excluding DNA samples extracted from hair or supplied by Polish collaborators following whole-genome amplification the average call rate of the remaining 1,032 samples was 99.2%. The mean number of dissimilar calls for the 24 duplicate samples (excluding the duplicates involving the hair samples) from DNA extracted from semen was 21 varying from 6 to 45. Call rates from the hair samples are given in Appendix 2. Call rates from the 12 samples received from Poland following whole genome amplification was 95.2%. Although AROS offered a facility to call genotypes, genotype calling was undertaken at Moorepark using the customized Illumina software, Beadstudio. Basic operations of Beadstudio are given in Appendix 4.

Editing of genotype data

Genotype data was exported from beadstudio in a space delimited file with 6 columns, sample ID (alphanumeric), SNP_index (unique SNP identifier ranging from 1 to 54,001), Allele1_AB (either "A" or "B"), Allele2_AB (either "A" or "B"), GC_score (numerical score varying from 0.15 to 0.99 on the "accuracy" of the SNP call for that sample), GT_Score (numerical score varying from 0.00 to 0.99 on the "accuracy" of the SNP call across all samples). Two additional files were also generated linking the sample_IDs to ICBF generated animal techids and linking the SNP_index to the respective chromosome and position on the chromosome. The latter was required for haplotyping. Genotypes from over 2,000 Holstein-Friesian sires were obtained from Livestock Improvement, New Zealand; 158 of these sires that had daughters in Ireland were retained.

All SNP editing was carried out in SAS. Genotypes were recoded into 0 (AA), 1 (AB) and 2 (BB). For imputing missing SNPs, haplotypes 7 SNPs long were generated for each individual at each location based on chromosome and position; no haplotypes were generated for SNPs that had no known position. Sire haplotypes, where genotyped, were merged with the progeny haplotypes by SNP. Missing progeny SNPs were imputed from the sire's genotype where the 3 SNPs either side of the missing SNP were identical in sire-son pairs. A similar approach was used to impute the SNP of an individual based

on its progeny. Haplotype frequency per genome location were calculated across all samples. In order of descending population frequency, where the 3 SNPs either side of a missing SNP were identical to a haplotype in the population the missing SNP was imputed from that haplotype. Where the same individual was genotyped more than once, the non-missing genotype (following imputing the missing SNPs) was retained.

Autosomal genotypes were also compared between sire and sons. If the progeny was homozygous AA then the sire should not be homozygous BB and vice versa. These discrepancies were identified and SNPs with >5% incorrect calls were discarded (n=218). Furthermore, sire-son pairs with >5% improper calls were identified. Where an individual with a large proportion of improper calls was identified its progeny's genotypes were extracted and these were used to determine if the inconsistencies between the sire-son was most likely attributable to the sire or the son. When it was not possible to determine which sample was likely incorrect both samples were discarded. A total of 15 individuals were discarded.

Several editing criteria were imposed based on individual SNP statistics. Following the removal of SNPs where there were 5% inconsistencies between sire-son genotypes, 53,189 SNPs remained. SNPs were subsequently discarded, in the following order, if they were on the X-chromosome (n=747), there was no genotype called for that SNP (n=15), were monomorphic (n=3,090), were not segregating (n=59), had a minor allele frequency <2% (4,463), had more than 5% missing calls (n=1,004), the clustering for the SNP was not of high quality (GenTrainScore <0.55; n=306), if there was a proportion of heterozygotes but neither of the two homozygotes were present (n=585), deviated significantly ($P<0.1*10^{-9}$) from Hardy-Weinberg equilibrium (n=0) and had a ratio of heterozgotes to homozygotes within SNP of >90% (n=0). A total of 42,920 SNP remained. Any missing genotypes which still existed were imputed as follows. If the frequency of the genotype of the missing SNP in the population was greater than 90% then the missing SNP was given that value. If the frequency of the genotype of the missing SNP in the population was <90% and the sire's genotype at that SNP was known then the progeny's genotype assumed the sire's genotype, otherwise the most prevalent

genotype was used. Finally SNPs in complete linkage disequilibrium across the sample population were discarded (n=322). A total of 42,598 SNPs on 1,209 animals remained.

Estimation of direct genomic values

Direct genomic values are estimated using mixed models equations by replacing the traditional numerator relationship matrix with a genomic relationship matrix as outlined by VanRaden (2008).

The dependent variable included in the genomic evaluation is the deregressed traditional EBV of the animal ($\tilde{y}$) as outlined by (Harris and Johnson, 2008):

$$\tilde{\mathbf{y}} = \mathbf{R}(\mathbf{R}^{-1} + \mathbf{A}^{-1})\hat{\mathbf{a}},$$

where $\hat{\mathbf{a}}$ is a vector of EBVs from traditional BLUP evaluations, **R** is a diagonal matrix containing one divided by the animal's reliability from his daughters less one, and **A** is the numerator relationship matrix. Domestic EBVs were used in the deregression when the associated domestic reliability was ≥90%; otherwise INTERBULL MACE EBVs were used.

The genomic relationship matrix is derived using the approaches outlined by VanRaden (2008). Prediction of DGVs are by mixed models (VanRaden, 2008) as:

$$\mathbf{D}\hat{\mathbf{G}}\mathbf{V} = \mathbf{G}(\mathbf{R} + \mathbf{G})^{-1}\tilde{\mathbf{y}}$$

where **G** is the genomic relationship matrix calculated from the markers; **R** is a diagonal matrix containing one divided by the animal's reliability from his daughters less one and $\tilde{\mathbf{y}}$ is the deregressed EBV for the trait under investigation.

DGVs for animals with no phenotypes are predicted by substituting the leftmost **G** matrix in the immediately previous equation with the genomic relationships between the animals with genotypes plus phenotypes and the animals with genotypes but no phenotypes.

Expected reliabilities of DGVs are calculated by direct inversion of the mixed model equations as follows for animals included with phenotypes and genotypes, and animal with just genotypes as follows, respectively (VanRaden, 2008):

$$\mathrm{Re}\, l_{WITH\ PHENOTYPES} = G \cdot [G + R]^{-1} \cdot G,$$

$$\mathrm{Re}\, l_{WITHOUT\ PHENOTYPES} = C \cdot [G + R]^{-1} \cdot C',$$

where G is the genomic relationship matrix between animals with genotypes and phenotypes, C is the genomic relationship matrix between animals without phenotypes and animals with phenotypes, R is a diagonal matrix containing one divided by the animal's reliability from his daughters less one.

Estimation of genomic EBVs

Genomic EBVs are the combination of the DGVs and the traditional EBVs. This blending procedure is done because not all animals in the pedigree are genotyped and therefore not all information is included in the genomic evaluation. Care must be taken in the blending to ensure no "double-counting" of pedigree effects.
The blended EBV (i.e., GEBV) and blended reliability made publicly available in Ireland are calculated as follows:

$$\boldsymbol{EBV}_{BLEND} = \frac{(-1+\boldsymbol{R}_{GS}) \cdot (\boldsymbol{EBV}_{NAT} + \boldsymbol{EBV}_{GA}(-1+\boldsymbol{R}_{NAT}) - \boldsymbol{EBV}_{NAT} \cdot \boldsymbol{R}_{GA}) + \boldsymbol{EBV}_{GS}(-1+\boldsymbol{R}_{NAT} + \boldsymbol{R}_{GA} - \boldsymbol{R}_{NAT} \cdot \boldsymbol{R}_{GA})}{-1+\boldsymbol{R}_{GS} \cdot (\boldsymbol{R}_{NAT} - \boldsymbol{R}_{GA}) - (-2+\boldsymbol{R}_{NAT}) \cdot \boldsymbol{R}_{GA}}$$

$$\boldsymbol{R}_{BLEND} = \frac{\boldsymbol{R}_{NAT} + \boldsymbol{R}_{GS} \cdot (1+\boldsymbol{R}_{NAT}(-2-\boldsymbol{R}_{GA})) - \boldsymbol{R}_{GA}}{1+(-2+\boldsymbol{R}_{NAT}) \cdot \boldsymbol{R}_{GA} + \boldsymbol{R}_{GS}(-\boldsymbol{R}_{NAT} + \boldsymbol{R}_{GA})}$$

where R_* is the reliability and EBV_* is the estimated breeding value for the different components of the selection index with the subscripts GS, NAT and GA representing the values obtained from the genomic evaluation, national evaluation and a traditional genetic evaluation including only relationships among genotyped animals. This is

identical to the approach of VanRaden (2008) but avoids the requirement to invert the 3x3 V-matrix.

The relative weighting on genomic information over and above that already contributed through the national genetic evaluations using traditional methods is calculated as:

$$\boldsymbol{WEIGHT_{GENOMICS}} = \frac{(-1+\boldsymbol{R_{GS}})\cdot(-1+\boldsymbol{R_{NAT}})+(-1+\boldsymbol{R_{NAT}}+\boldsymbol{R_{GA}}-\boldsymbol{R_{NAT}}\cdot\boldsymbol{R_{GA}})}{-1+\boldsymbol{R_{GS}}\cdot(\boldsymbol{R_{NAT}}-\boldsymbol{R_{GA}})-(-2+\boldsymbol{R_{NAT}})\cdot\boldsymbol{R_{GA}}}$$

Testing the accuracy of genomic selection in Ireland

Predicted transmitting ability and associated reliabilities used to derive the input to the genomic selection were obtained from ICBF on Holstein-Friesian sires with daughters in Ireland. Predicted transmitting ability and associated reliabilities calculated from INTERBULL were also obtained from the ICBF on all sires. Each trait was treated separately when applying edits with the exception of milk, fat and protein yield which have identical heritability in the genetic evaluation, the same number of progeny, and therefore the same reliability within animal.

To test the accuracy of genomic selection using Irish data only genotyped sires with at least 40 milking daughters in Ireland were retained (n=803). This dataset was divided into sires born prior to 1997 (n=596; training dataset) and sires born after 1996 (n=207; validation dataset). DGVs and GEBVs were predicted for the validation dataset. The accuracy of genomic selection was quantified by the mean bias and RMSE as well as the correlation and regression of actual EBVs (as estimated using the traditional methods) on genomic EBVs. Results are summarized in Table 1 for the traits included in the EBI. The results in Table 1 may be artificially superior since daughters of the sires in the validation dataset were included in the genetic evaluation of sires in the training dataset. Regression coefficients varied from 0.61 to 0.99.

For the actual genomic evaluation of young test bulls, all sires with daughters in Ireland were included in the training population (n=945). Genomic and blended EBVs as well as reliabilities were calculated for a total of 246 young bulls with no daughters in

Ireland. Summary statistics are detailed in Table 2 for the traits included in the EBI. The average increase in reliability for the blended EBVs over and above those obtained from parental average using traditional methods varied from 0.01 (locomotion) to 0.18 (fertility sub-index); the weighting on genomic information per individual varied from 0 to 48%. The main reason for the poorer response to the addition of genomic information in Ireland compared to others such as the US (VanRaden et al., 2009) and LIC (Harris et al., 2008) is most likely due to the smaller training population size in Ireland. The genomic reliability of individual bulls for EBI increases as their relatedness to the training populations of bulls increased (Figure 1).

Table 1. *Mean bias and root mean square error (RMSE) of the predicted trait and index values from blended genomic and traditional proofs in the group of validation bulls (n=207).*

Index / Trait	Bias	RMSE	r	b (se)
Economic Breeding index	1.3	39.2	0.68	0.67 (0.05)
Production sub-index	8.6	19.4	0.80	0.84 (0.04)
Fertility sub-index	-9.4	32.4	0.79	0.62 (0.03)
Calving sub-index	3.5	8.3	0.73	0.93 (0.06)
Beef sub-index	-3.2	6.4	0.59	0.71 (0.07)
Health sub-index	1.7	5.4	0.70	0.89 (0.06)
Milk yield	58.8	125.8	0.83	0.76 (0.04)
Fat yield	1.4	4.2	0.76	0.78 (0.05)
Protein yield	1.8	3.5	0.81	0.80 (0.04)
Calving interval	0.4	2.4	0.80	0.64 (0.03)
Survival	0.0	0.0	0.66	0.61 (0.05)
Direct calving difficulty	-0.8	1.2	0.65	0.77 (0.06)
Maternal calving difficulty	1.0	1.3	0.76	0.81 (0.05)
Direct gestation length	-0.3	0.8	0.72	0.90 (0.06)
Direct calf mortality	0.0	0.4	0.73	0.99 (0.06)
Progeny carcass weight	-1.5	4.6	0.68	0.74 (0.06)
Progeny carcass conformation	-0.2	0.2	0.80	0.81 (0.04)
Progeny carcass fat	0.0	0.1	0.78	0.82 (0.05)
Cull cow weight	-0.4	5.2	0.81	0.76 (0.04)
Somatic cell score (*1000)	-0.1	0.9	0.68	0.85 (0.06)
Locomotion	0.1	0.8	0.50	0.67 (0.08)

Table 2. *Mean, standard deviation and reliabilities for the different indexes and traits from parental averages and blended evaluations as well as the weight on genomics in the blended proofs and the correlation between the blended proof and parental average in the young bulls.*

Index / Trait	Mean		SD		Reliability		Weight	r
	PA	Blend	PA	Blend	PA	Blend		
Economic Breeding index	117	122	39	41	0.30	0.45	0.19	0.80
Production sub-index	61	69	24	27	0.37	0.50	0.20	0.87
Fertility sub-index	41	38	29	38	0.21	0.39	0.18	0.83
Calving sub-index	21	25	8	9	0.34	0.46	0.18	0.85
Beef sub-index	-6	-10	6	7	0.27	0.42	0.19	0.80
Health sub-index	-1	-1	3	2	0.29	0.43	0.18	0.83
Milk yield	168	188	140	175	0.37	0.50	0.20	0.90
Fat yield	10.7	12.1	4.6	5.3	0.37	0.50	0.20	0.85
Protein yield	9.1	10.3	4.1	5.0	0.37	0.50	0.20	0.90
Calving interval	-2.22	-2.70	1.95	2.72	0.22	0.40	0.20	0.85
Survival	1.31	0.51	0.80	0.79	0.19	0.36	0.19	0.74
Direct calving difficulty	-3.33	-4.08	0.79	0.94	0.35	0.47	0.17	0.63
Maternal calving difficulty	2.90	3.68	1.09	1.38	0.34	0.47	0.19	0.81
Direct Gestation length	-1.57	-1.88	0.74	0.80	0.34	0.46	0.19	0.87
Direct calf mortality	-0.77	-0.95	0.41	0.45	0.27	0.36	0.18	0.83
Prog. carcass weight	-1.71	-3.42	4.57	5.62	0.27	0.42	0.12	0.86
Prog. carcass conformation	-0.46	-0.64	0.22	0.26	0.26	0.42	0.19	0.75
Prog. carcass fat	-0.05	-0.06	0.12	0.17	0.27	0.41	0.19	0.84
Cull cow weight	-0.79	-2.00	5.43	7.21	0.25	0.40	0.18	0.89
SCS (*1000)	12.4	8.9	53.9	37.9	0.33	0.47	0.18	0.82
Locomotion	-0.19	-0.29	0.58	0.73	0.30	0.31	0.14	0.77

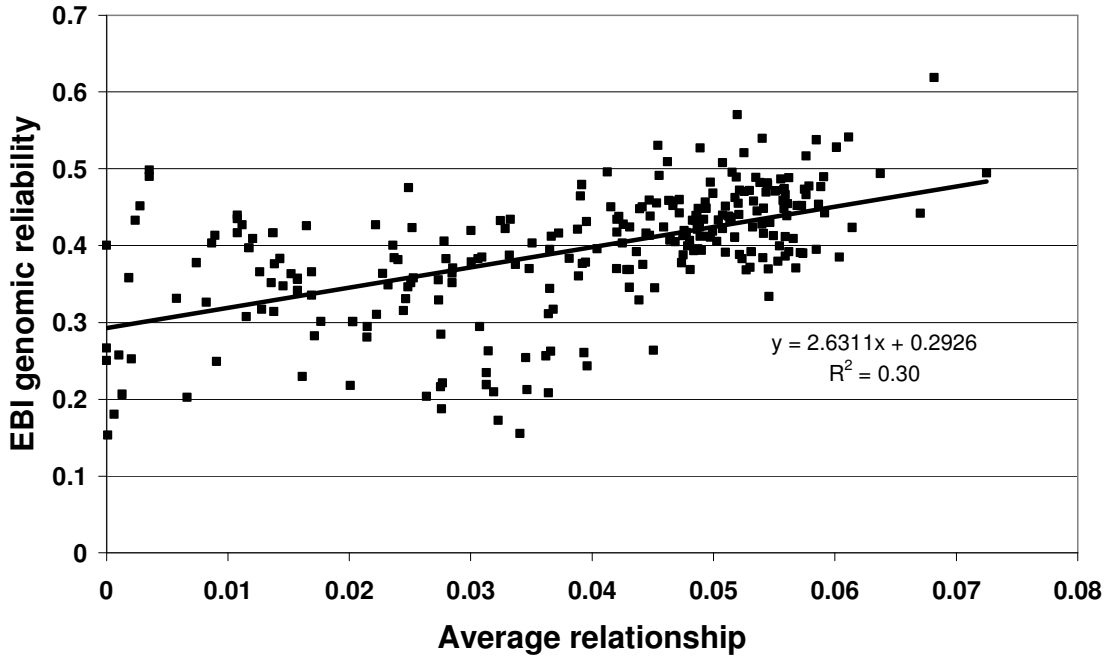

Figure 1. *Association between genomic reliability of young bulls and their average expected relationship (calculated only from pedigree) with animals in the training population.*

Implementation

Evaluations for bulls in Ireland based on daughter performance have been routinely published four times a year. Three of these runs are concurrent with INTERBULL and the fourth is a domestic run. Proofs are published in January, April, September, and November. Bulls tested in Ireland or who have a domestic reliability of >90% for an Interbull trait will have their Irish EBV published; otherwise a bull's Interbull proof is published. A parent average is assigned to bulls that do not have information for a particular trait. All proofs are available either through the individual bull search or the bull files on the website. In addition, the top 75 bulls across all breeds based on EBI with semen available (the Active Bull List) is published twice per year, once in the spring, and once in the autumn. The minimum reliability for a bull to be on the list was at 58% (in 2008). Conversion equations are used to convert genetic proofs for

Norwegian Red and Jersey that are not yet part of Interbull. Test-bulls have a maximum of 1000 doses available for use, but no limits were placed on widespread bulls.

It is anticipated that there will be no change to the current evaluation schedule with the introduction of genomic data. All bulls will receive updated proofs four times per year. A stand-alone process may be put in place to aid the turnaround time for calves that AI companies and breeders are specifically interested it. In terms of publication, currently information based on progeny performance is used where available. If a bull has no progeny performance, but has a genotype, he gets a blended genomic value and reliability for that trait. In future this may change to the use of all blended information and reliabilities. The introduction of genomic selection bulls has seen changes made to the Active Bull List. We are now able to calculate a GEBI for these bulls based on a genotype provided by an AI company. It is essential for a bull to have a calving proof based on progeny performance to be included on the list. In the case of Jersey, we are using conversion equations and the country of origin (genomic) figures, to produce a GEBI but these bulls are not eligible for the Active Bull List. The recommended limits for genomically selected bulls and their eligibility to be included on the Active Bull List are:

1. For Holstein Friesian bulls;

a. Those with an EBI reliability of 50% or greater (based on the Irish genomic key) and with a calving difficulty proof in the country of 1st proof with a reliability of 50% or greater, then there be maximum of 10,000 doses used in Ireland. **(*eligible for Active Bull List*)**

b. Those with an EBI reliability of 35% to 50% (based on the Irish genomic key) and with a calving difficulty proof in the country of 1st proof with a reliability of 50% or greater, then there be maximum of 5,000 doses used in Ireland. **(*eligible for Active Bull List*)**

c. Those with an EBI reliability of less than 35% (based on the Irish genomic key) and with a calving difficulty proof in the country of 1st proof with a

reliability of 50% or greater, then there be maximum of 3,000 doses used in Ireland. **(NOT *eligible for Active Bull List as Reliability < 35%)***

2. For bulls of other breeds; (not eligible for Active Bull List)

a. For bulls with an evaluation based on the genomic key for another country and a calving difficulty proof in the country of 1st proof with a reliability of 50% or greater, then there be maximum of 3,000 doses used in Ireland.

b. As there is no Irish genomic key for such bulls & breeds, then these bulls (as in 2 a) will be required to be incorporated into an expanded G€N€ IR€LAND progeny test program where 350 straws per bull is used alongside progeny test bulls to ensure more accurate proofs for these bulls & breeds in the future.

ICBF will be monitoring the usage of all bulls used in AI and plan to review the situation on completion of the spring mating season. In addition, a second list will be produced in spring as some of the bulls will have reached their limits prior to the end of the breeding season. We also expect to go to a 'live' active bull list, such that anyone can get an up-to-date list of the top 75 available bulls by requesting it from the website. It is anticipated that the AI companies will have an active roll in making sure their semen inventories are kept current.

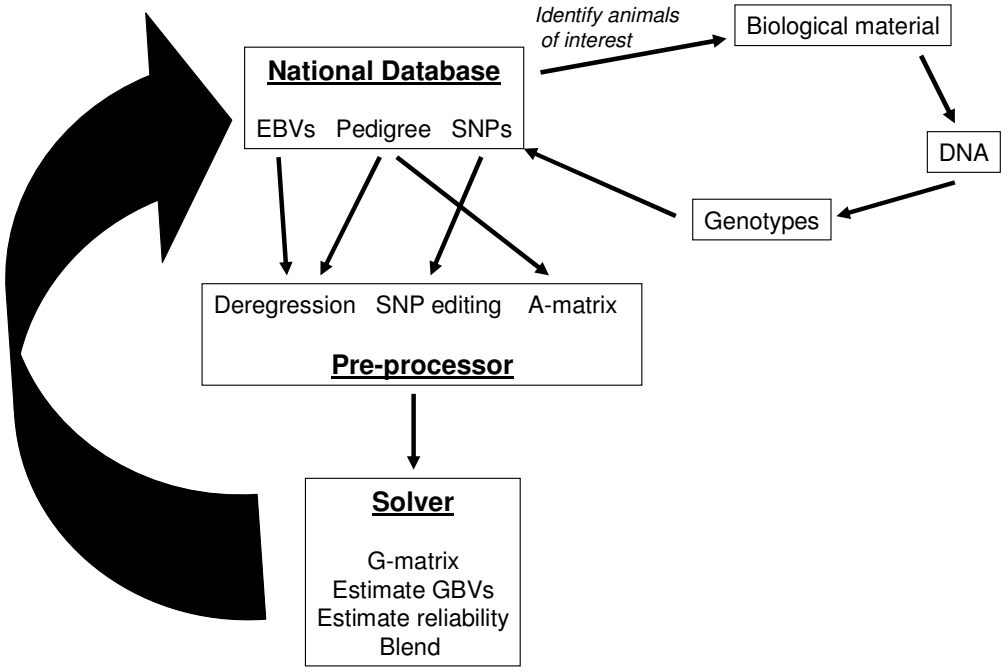

Figure 2. *Summary of the procedures involved in genomic selection in Ireland*

Acknowledgements

This project has been funded by Teagasc, the Department of Agricultural, Conservation of genetic resources for food and agriculture, Stimulus Research Fund, the Irish Cattle Breeding Federation, and the National Cattle Breeding Centre. Gratitude is also extended to the breeding organizations and farmers who provided biological material as well as the intellectual input from Irish and international scientists.

References

Harris, B.L., and D.L. Johnson (2008). Outline of the methods for integration of genomic information in the national genetic evaluation system. Technical report to the New Zealand Animal Evaluation Unit.

Harris, B.L., D.L. Johnson, and R.J. Spelman. (2008). Genomic selection in New Zealand and the implications for national Genetic Evaluation. Proceedings of the 36th ICAR Conference. Niagara Falls, NY. 16th – 20th June. Pp 35.

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. (2001). Prediction of total genetic value using genome-wide dense marker maps. Genetics. 157:1819-1829.

Sellner, E.M., Kim, J.W., McClure, M.C., Taylor, K.H., Schnabel, R.D., and Taylor, J.F. (2007). Application of genomic information in livestock. J. Anim. Sci. 85: 3148-3158.

VanRaden, P.M. (2008). Efficient methods to compute genomic predictions. J. Dairy Sci. 91: 4414-4423.

VanRaden, P.M., C.P. Van Tassell, G.R. Wiggans, T.S. Sonstegard, R.D. Schnabel, J.F. Taylor, and F.S. Schenkel. (2009). Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92: 16-24.

19th Jan 09

APPENDIX 1.

SOP for DNA extraction from semen straws

Prepared/Revised By: Dawn Howard **Department:** Molecular Lab, Athenry

Date: 16th September 08

1.0 Purpose:

1.1 To describe the procedure to be followed when extracting DNA from semen straws.

2.0 Scope:

2.1 All DNA extracted from semen straws.

3.0 Responsibility:

3.1 It is the responsibility of the molecular Research Officer to update this SOP as required.

4.0 Equipment:

4.1 Maxwell 16 instrument (cat. # AS1000)

4.2 Maxwell 16 Tissue DNA purification kit (cat. # AS1030)

4.3 Microcentrifuge

4.4 Waterbath or oven at 60°C

4.5 1.5 µl microcentrifuge tubes

4.6 1X Phosphate buffered saline (PBS) (pH 7.4)

4.7 Extraction buffer (10mM Tris pH 8, 10mM EDTA pH8.0, 1% SDS, 100mM NaCl) warmed to 60°C.

4.8 2-Mercaptoethanol

4.9 Proteinase K solution (20mg/ml)

4.10 Sterile scalpel blades.

5.0 Procedure:

5.1 Overnight Lysis:

5.1.1 Empty semen straw into a 1.5 ml microcentrifuge tube.

5.1.2 Add 1ml of 1X PBS .

5.1.3 Centrifuge at 5000g for 5 minutes to pellet the sperm.

5.1.4 Aspirate the supernatant.

5.1.5 Resuspend the pellet in 1ml 1X PBS.

5.1.6 Repeat step 5.1.3 and 5.1.4,

5.1.7 Resuspend the pellet in 450µl of warmed extraction buffer.

5.1.8 Add 15µl of 2-Mercaptoethanol and incubate at 55°C for 15min.

5.1.9 Add 10µl of the Proteinase K solution (20mg/ml).

5.1.10 Incubate overnight at 60°C. Samples can also be stored at -20°C at this stage for future processing.

5.2 Maxwell purification:

5.2.1 Following incubation, arrange the Maxwell 16 cartridges (1 per sample) into the holding rack with the ridged side of the cartridge facing the numbered side of the rack.

5.2.2 Remove the seal from each cartridge.

5.2.3 Place a plunger into well #7 of each cartridge (well #7 is the well closest to the ridged side of the cartridge).

5.2.4 Transfer the entire contents of the 1.5ml microfuge tube into well #1.

5.2.5 Place the blue elution tubes into the magnetic holding rack in the same arrangement as the cartridges. Add 300ul of elution buffer into each tube.

5.2.6 Transfer the reagent cartridges and the elution tubes into the Maxwell 16 platform. If processing less than 16 samples centre the

cartridges on the platform and space them evenly outwards from the centre.

5.2.7 Run the Maxwell16 using the DNA/Tissue program. This program takes approx 40 minutes.

5.2.8 At the end of the run replace the elution tubes in the holding rack.

5.2.9 Transfer the eluted samples into storage tubes using a pipette. To avoid particle transfer do not remove the elution tubes from the rack and pipette the samples from the side of the elution tube next to the numbered edge of the rack.

5.2.10 Dispose of the used elution tubes into a Biohazard waste bag.

6.0 DNA quantity analysis:

6.1 The concentration, 260/280 and 230/280 values for each DNA sample is determined by optical density analysis on the NanoDrop® spectrophotometer.

6.2 Refer to SOP 2003 for detailed instructions on operation of the NanoDrop® instrument.

7.0 QC of semen extracted DNA:

7.1 DNA extracted from semen must undergo QC testing to ensure the DNA extracted is of the expected quality and free of any inhibitors which might interfere with PCR. The quantity of the DNA will already have been tested in step 6.0 above.

7.2 Every extraction run for the Maxwell® instrument will be tested to ensure correct operation, therefore at least one sample from each run of 16 must be tested.

7.3 Refer to SOP 3004 for QC procedure and requirements.

19th Jan 09

APPENDIX 2.

Report on the potential to extract DNA from bovine hair samples to be used with the Illumina BovineSNP50 Beadchip

Donagh Berry[1], John Flynn[2], Dawn Howard[3], Brian Wickham[4]

[1]*Teagasc, Moorepark Dairy Production Research Centre, Fermoy, Co. Cork.*

[2]*Weatherbys Ireland Ltd., Naas, Co. Kildare.*

[3]*Teagasc, Animal Bioscience Center, Grange, Dunsany, Co. Meath.*

[4]*Irish Cattle Breeding Federation, Highfield House, Bandon, Co. Cork.*

Objective

The objective of this study was to determine the feasibility of extracting sufficient quality and quantity of DNA from bovine hair follicles for genotyping using the Illumina BovineSNP50 Beadchip

Study design

DNA from hair samples of 11 bulls was extracted in Weatherby's Laboratory, Naas, Ireland. In brief, the DNA was extracted using a lysis buffer containing 0.3g Tris 0.93g KCL,1.25ml Tween made up to 250mls with H2O and Proteinase K 5mg/ml. For 50 samples, 5ml lysis solution was added to 50µl Proteinase K. 100µl of this solution was added to 4-6 hair follicles per tube which was incubated for 45 minutes at 56ºC followed by 95ºC for 15 minutes. Extracts were spun at 11,000 RPM for 30seconds immediately prior to PCR. DNA was also extracted from frozen semen straws on 30 bulls, 6 of which were common with the hair samples extracted by Weatherbys. The remaining 24 animals were genotyped twice using DNA from semen. DNA extraction procedures are outlined in Appendix 1. DNA quality of all samples was quantified using the same Nanodrop® Spectrophotometer (Thermo Scientific, US) at Teagasc, Athenry. All genotyping was carried out by AROS, Aarhus, Denmark using the Illumina BovineSNP50 Beadchip. Call rates and level of concordance between genotype calls on the same animal was quantified for each sample.

Results

The quality of DNA extracted from the hair and corresponding semen samples are detailed in Table 1. The quality of DNA from the hair samples was in most cases considerably poorer than the quality of DNA from the semen straws. Furthermore, all 260:280 ratio values were less than 1.8 for the hair samples while only one DNA sample extracted from semen was less than 1.8; a pure sample of DNA has a 260/280 ratio of 1.8 and is relatively free from protein contamination. A DNA preparation that is contaminated with protein will have a 260/280 ratio lower than 1.8. Genotype call rates for the hair samples varied from 85.3% to 98.5% while genotype call rates for the semen samples were all greater than 98.8%. Of the 1,032 samples previously genotyped with the Illumina Bovine50 Beadchip the average call rate was 99.2% while excluding the 10 samples with the lowest call rate, the next lowest call rate was 95%. These results suggest that the DNA extracted from hair, using the methods of extraction in the present study, provided on average poor quality DNA which in most cases did not genotype very well. However, good genotype calls were obtained from some hair samples.

Degree of concordance between the genotypes called for the 6 animals genotyped using DNA extracted from hair or from semen are summarised in Table 2. Only SNPs where a call existed for both samples were retained. The mean number of dissimilar calls for the 24 duplicate samples all from DNA extracted from semen was 21 varying from 6 to 45. A considerably poor level of concordance was observed between the genotypes from DNA extracted from hair and from semen. This suggests that the genotype results obtained from DNA extracted from hair are not always consistent with the DNA extracted from semen on the same animal.

Table 1. *Comparison of DNA concentrate (ng/ul)[1] and 260/280 ratio for 11 hair samples and corresponding 6 semen samples. Also included is the call rate for each sample[2].*

AICODE	Hair sample					Semen sample		
	ng/ul	260/280	ng/ul	260/280	Call rate	ng/ul	260/280	Call rate
ARX	15.2	1.17	15.2	1.12	0.973	94.3	1.83	0.994
KSY	28.0	1.19	25.0	1.18	0.872	28.0	1.19	0.990
BSG	18.5	1.08	17.4	1.11	0.853	112.9	1.84	0.989
KQU	13.5	1.24	13.5	1.22	0.980	116.7	1.85	0.989
BCY	75.5	1.26	77.4	1.30	0.976	69.1	1.85	0.994
PVH	44.7	1.21	43.8	1.21	0.978	90.8	1.81	0.989
KDI	40.4	1.25	41.3	1.22	0.985			
KIJ	178.4	1.38	173.1	1.37	0.934			
KTX	186.0	1.39	182.3	1.38	0.944			
RHU	37.7	1.26	38.0	1.27	0.979			
HTO	41.4	1.24	39.7	1.26	0.971			

[1] DNA concentration and 260:280 ratio was quantified twice for the hair samples but only once for the semen samples

[2] Call rate is the proportion of the 54,001 SNPs that were called with sufficient accuracy (i.e., GC score>0.15)

Table 2. *Degree of concordance between called genotypes for the 6 samples genotyped both from hair and semen samples*

AICODE	No SNPs correct	No SNP errors	Total No SNPS	Reproducibility
ARX	52067	228	52295	0.998
BCY	52338	193	52531	0.998
BSG	41622	1096	42718	0.987
KQU	52385	141	52526	0.999
KSY	43679	819	44498	0.991
PVH	52208	138	52346	0.999

Conclusions and recommendations

- It appears that DNA extracted from hair samples, using the DNA extraction protocols used in the present study, do not consistently provide DNA of high quality which gives high genotyping call rates and concurs well with genotypes from DNA extracted from semen straws

- One recommendation would be to repeat the study on the same 6 animals using alternative protocols for DNA extraction from hair

19th Jan 09

APPENDIX 3.

Companies that offer a genotyping service for the Illumina Bovine50SNP Beadchip

Richard Wood
Regional Account Manager
Illumina UK Ltd.
Mobile: 07799 036188
Email: rwood@illumina.com
Web: www.illumina.com

Mark S. Thornber,
Director of Commercial Operations,
AROS Applied Biotechnology AS
Web: www.arosab.com
Email: mark.thornber@arosab.com
Mob: +44 7876 337432
Tel: +45 7027 1170

Hein van der Steen
StoneBridge breeding ltd
The Gate House, Abbotswood,
Evesham, Worcs, WR11 4NS, UK
hein.vandersteen@stonebridgeg.com
phone: +44 1386 41189

Daniel Pomp
GeneSeek Inc.
4665 Innovation Dr Ste 120
Lincoln, NE 68521 USA
Mobile: 402-770-0506
dpomp@geneseek.com

19th Jan 09

APPENDIX 4

Steps when using the Illumnia BeadStudio software

Donagh Berry

04/11/08

Importing data

To import additional samples click on "file" → "Load additional samples". In the dialog box click on "load sample intensities by selecting directories with intensity files". Then just point at the folder that holds all the folders for each plate. No need to click on each file, just the folders.

Once the data is in

Priority should be given to getting as many samples called for each SNP and to remove SNPs with poor overall clustering.

When all data is in, I suggest re-clustering under the heading "Analysis" → "Cluster all SNPs". Then if you click on the tab "SNP table" of the table with all the SNPs you can sort by "# no calls" and find the SNPs where poor numbers were called. Also by looking at the GenTrain score you can move around the clusters on the SNP graph (by holding shift and moving with the mouse) to get a poor GenTrain score. In the editing software SNPs with a GenTrain score <0.55 are removed. Also, you can select a number of SNPs and right click and click "zero SNP". This will set the GenTrain score to zero for these SNPs

You can import a text file linking the sample_IDs to AIcodes. On the Samples table click on "Import columns into the table". Make sure that the heading of one of the columns is "Sample ID" (with the space). You can identify duplicates by clicking on "Analysis" →

"Edit Replicates" and than match up the duplicates. You can generate a report on how many SNPs differed by duplicate by clicking on "Analysis" → "Reports" → "Create reproducibility and heritability report"

You can also load or make a SNP cluster file which stores the clustering used after modifying. You can save it through "File" → "Import/export cluster position"

Exporting the file

Beadstudio will export files in two main ways. As a matrix with one row per SNP and multiple columns per animal or as a file with one row per SNP-by-animal. The latter is the preferred option so the number of rows is 54001 times the number of animals.

Click on "analysis" → "report wizard". Click on all samples and then you get the dialog box below. Drag across the fields: Sample_ID, SNP_index, Allele1_AB, Allele2_AB, GC_score, GT_score. I save as a tab delimited file.