SELECTION FOR NOVEL TRAITS: AN INTERNATIONAL GENOMICS PERSPECTIVE

Donagh P. Berry[1]
[1]*Animal & Grassland Research and Innovation Centre Teagasc, Moorepark, Ireland*

Introduction

Genomic selection is being heralded as the "..most promising application of molecular genetics in livestock production since work began almost 20 years ago" (Sellner et al., 2007). The objective of genomic selection is to increase the accuracy of identifying genetically elite (and inferior) animals at a younger age but also at a lower cost per animal. Genetic gain may be defined as (Rendel and Robertson, 1950):

$$\Delta G = \frac{i \cdot r \cdot \sigma}{L}$$

where ΔG is annual genetic gain; i is the intensity of selection; r is the accuracy with which you know the genetic merit of each animal, σ is the genetic standard deviation (i.e., the square root of the genetic variance or simply just a measure of the genetic differences among animals), and L is the generation interval. Genomic selection attempts to alter i, r and L. It may also influence the detected genetic variation. Genomic selection, however, does not necessarily improve all three components simultaneously as it may reduce the accuracy of selection (i.e., r) compared to traditional methods but reduce the generation interval (i.e., L) proportionally more thereby increasing annual genetic gain. Because the cost of "testing" a young bull with genomic selection is approximately 0.3% (i.e., 0.003) the cost of progeny testing the selection intensity can be increased considerably thus advancing genetic gain.

Genomic selection (and genomics in general) is particularly advantageous for traits that are:

- Sex linked (e.g. milk yield and female fertility)
- Take a long time to measure (e.g., cow longevity)
- Exhibit low heritability (e.g., female fertility)
- Difficult and/or expensive to measure (i.e., novel traits like feed intake complex, meat quality)

Genomic selection (GS) has been successfully implemented into national dairy cattle genetic evaluations in many countries since 2009 (Spelman et al., 2013). Retrospective analysis (McParland et al., 2014) signifies that GS is up to 29% more accurate at predicting an animal's true genetic merit (based on progeny performance) compared to just parental average. However, the breeding structures of dairy and beef are quite different and this has implications for the successful implementation of genomic selection in beef but also the justification for international cooperation, especially for novel traits.

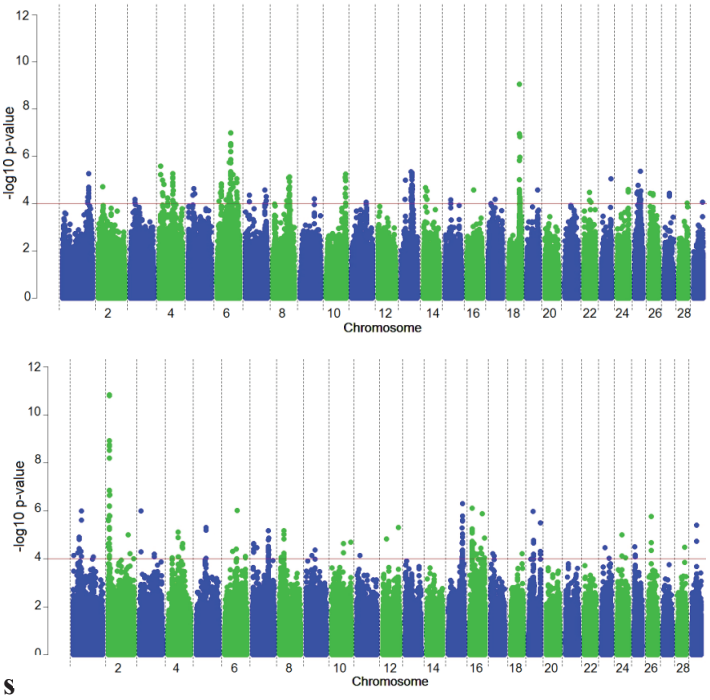
The objective of this article is to discuss the potential for international collaboration in genomics in beef cattle; although examples will be given for novel traits the relevance of the discussion is applicable to all traits although the marginal benefit is greatest for novel traits where the population of phenotyped and genotyped animals may be smaller (discussed later).

Differences between Dairy and Beef Breeding Structures and Implications for Genomic Selection

Many differences exist between dairy and beef breeding structures so therefore the approaches applied to-date in dairy cattle may not be directly applied in beef, although there are obvious similarities.

Breed. One breed (i.e., Holstein-Friesian) predominates the dairy cattle populations in most developed countries making it relatively easy and inexpensive to develop large informative reference populations for the generation of accurate genomic predictions. It is now well known that the stronger the genomic relationship between the reference population of genotyped and phenotyped animals with the candidate animals, the greater will be, on average, the accuracy of genomic predictions (Habier et al., 2007; Pszczola et al., 2012). Accurate across-breed genomic predictions have to-date been elusive (Karoui et al, 2012; Berry 2012) in cattle. Figure 1 shows a genome wide association study for direct calving difficulty in Irish Holstein and Charolais animals. The scoring system for calving difficulty is the same across both breeds and the genetic evaluations are across breeds. A genomic region with a large association (2.49% of genetic variation) with calving difficulty was detected on chromosome 18 in the Holstein-Friesian population and, although these SNPs were also segre-

gating in the Charolais population, no association was detected in this region of the genome. Similarly a genomic region associated with calving difficulty in Charolais (3.13% of the genetic variation) was detected on Chromsome 2 but not in Holstein-Friesians despite the SNPs segregating in both populations. Moreover, the sign of the allelic effects for 50% of SNP differed when estimated in either the Holstein-Friesian population or the Charolais population. This is likely due to differences in linkage phase between breeds and background polygenic effects and is undoubtedly a contributor to the sometimes observed negative correlations between genomically predicted EPDs and progeny-based EPDs when the population being tested is not adequately represented in the genomic reference population. This difference between dairy and beef and the current inability for genomic algorithms and genomic information to be useful for acrossbreed genomic evaluation implies that each breed has to generate (and therefore incur the cost) of generating its own reference population. The same is true for novel traits implying a large cost for each country to implement and genomic selection program.



s

Figure 1. Manhattan plots of the single nucleotide polymorphisms associated with direct calving difficulty in 770 Holstein-Friesian (Top figure) and 927 Charolais (bottom figure) (Purfield et al., 2014).

Effective Population Size. The effective population size globally of Holstein-Friesians is likely to be somewhere between 40 and 100 (McParland et al., 2007; Saatchi et al., 2011). The global effective population size of beef breeds is likely to be larger (McParland et al., 2007; Saatchi et al., 2011) given the vast differences in breeding policies implemented in the different populations. The accuracy of genomic predictions is a function of the size of the reference population, the heritability of the trait under investigation, and the effective population size of the population (Daetwyler et al., 2008). Larger effective population sizes require larger reference populations to achieve the equivalent accuracy of genomic predictions compared to populations with smaller effective population sizes.

The number of independent genomic segments is likely to vary with effective population size. The number of independent loci (M_e) in a 30 Morgan genome can be derived deterministically for a range of different effective population sizes as (Goddard, 2009):

$$M_e = \frac{2N_e L}{Log_{10}(4N_e L)}$$

where N_e is the effective population size and L is the length of the genome in Morgans. The number of animals (N) required to achieve a given accuracy (i.e., square root of the reliability) can then be derived as (Calus et al., 2012):

$$N = \frac{r^2 M_e}{h^2 (q^2 - r^2)}$$

Where q^2 is the proportion of genetic variance captured by the SNPs (here assumed to be 0.8) and h^2 is the heritability of the traits (here assumed to be 0.20). Figure 2 illustrates the number of animals that need to be phenotyped and genotyped to achieve a given accuracy for different effective population sizes. The larger the effective population size the larger the dataset of phenotyped and genotyped animals that is required to achieve an equivalent accuracy of genomic predictions compared with populations with smaller effective population sizes.

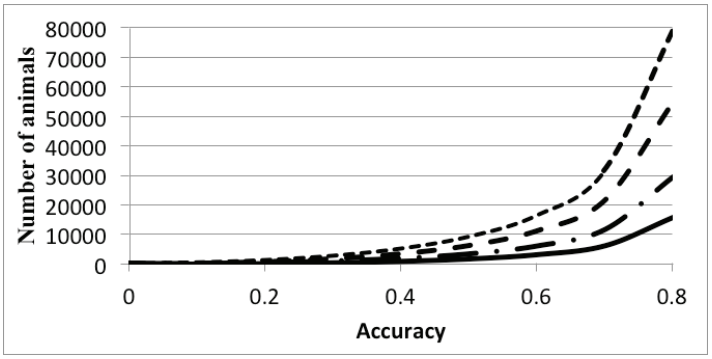


Figure 2. Number of animals that need to be both genotyped and phenotyped to achieve different levels of accuracy (i.e., square root of the reliability) when the effective population size is 50 (solid line), 100 (long-dashed line), 200 (shorter dashed line) and 300 (smallest dashed line)

Greater Usage of AI in Dairy. In general, there is a greater usage of AI in dairy cattle than in beef. The accuracy of an animal's EPD from traditional genetic evaluations increases with increasing quantity of progeny records; therefore the accuracy of progeny tested bull can be very high. Because the heritability statistic measures the strength of the resemblance between the phenotypic value of an animal and its true genetic merit, the effective heritability of high accuracy EPDs is close to unity.

Figure 3 illustrates the number of genotyped and phenotyped animals required to achieve different accuracy levels of genomic predictions. Clearly to achieve the same accuracy of genomic predictions, less genotyped and phenotyped animals are required for higher heritability traits (or animals with higher accuracy EPDs). Dairy cattle genomic breeding programs firstly focused on the genotyping of thousands of AI progeny tested bulls because of their greater accuracy and thus greater effective heritability. Using this approach for a trait with a heritability of 0.20, 7903 genotyped animals with own performance records would be equivalent to 1756 bulls (i.e., less than one quarter) with an EPD accuracy (i.e., square root of reliability) of 0.95. Hence, all else being equal, the implementation of genomic selection in beef where less high reliability sires exist will be considerably more expensive than in dairy. Collaboration can help reduce this cost. Because novel traits do not generally tend to be measured on large populations of animals, the generation of high accuracy EPDs for a large population of sires is generally not achievable. Therefore, a large population of phenotyped and genotyped animals will be required to achieve an acceptable accuracy of genomic predictions.

Less Phenotypes and Parentage Recording in Beef. Accurate recording of detailed phenotypes on large populations of commercial animals is generally the norm in most dairy cow populations. Furthermore, parentage of most dairy females is known facilitating accurate EPDs of their pedigree. Although phenotypic recording exists in many beef populations it is, however, lacking (for some traits at least) in some populations. As alluded to previously, genotypes from animals with high accuracy EPDs can be more informative than genotypes of animals with lower accuracy EPDs. One-step genomic procedures will not alleviate this issue as animals with non-recorded pedigree will still have to be genotyped to allocate the animal to its pedigree. Ultra-low cost genomic tools for parentage assignment may aid in allocating animals to parents and thus increase the accuracy of traditional genetic evaluations for some animals. Lack of pedigree information and phenotypes is generally not of concern for animals with novel phenotypes since if the resources are being expended in generating the phenotypes then the pedigree is usually also recorded.

Lack of Participation in International Genetic Evaluations. Many genomic evaluations in dairy cattle, including Ireland, operate a two-step procedure where

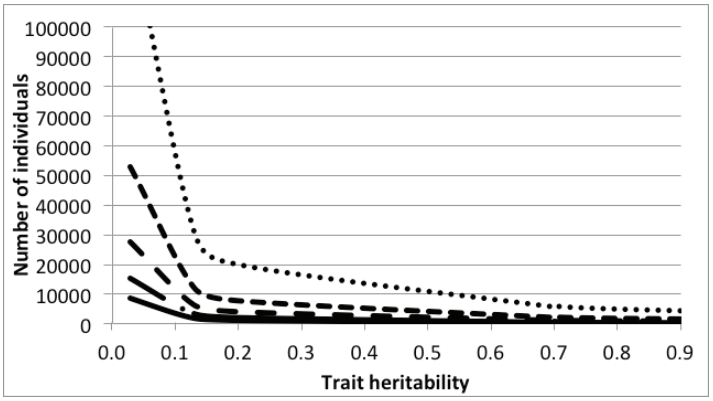


Figure 3. Number of phenotyped and genotyped animals that are required to achieve an accuracy of genomic prediction of 0.4 to 0.8 (length of dashes decrease and the accuracy increases); calculations are based on the assumption of 1000 independent genomic regions and the genomic markers explaining 80% of the total genetic variance.

derivatives of EPDs are used as input phenotypes for the development of the genomic prediction equations. Many countries, especially those with a small breeding program, exploit MACE evaluations generated by INTERBULL. Therefore, phenotypic information is available on bulls even if they do not have any daughter performance records in that country. The level of participation of beef breeds in international genetic evaluations is less although initiatives such as BreedPlan and INTERBEEF as well as pan-American are underway. If participating in international genetic evaluations the extent of genotype-by-environment interactions should be quantified and the appropriate approach taken thereafter in the genetic evaluation. To estimate precise genetic correlation between populations, good genetic connectedness is needed (Berry et al., 2014b). This is particularly true for novel traits which tend to be mueasured in research herds.

Type of International Genomic Cooperation Initiatives in Genomics

Several alternative strategies of international cooperation in genomics exist and some of these are briefly discussed. It should be noted that there is widespread international collaboration among dairy populations both in the sharing of genotypes and phenotypes. This is despite the points previously raised to that genomic selection in dairy is arguably considerably easier (i.e., less expensive) than in beef.

Sharing of Information on What Animals Have Been Genotyped. One of the easiest and least controversial approaches to achieving useful international collaboration in the selection of novel traits is the sharing of information on what animals have been genotyped and on what genotyping panel. An example of the current international list compiled (currently only participated in by Ireland, the UK, France and Australia) is in Table 1; Australian information was deleted to fit in the page as were several other columns with bull aliases. Moreover, whether DNA is available or if the bull is of particular importance in a country for genotyping can be noted. The complete list can be obtained from the author (donagh.berry@tagasc.ie). Furthermore, requests to join the list can be directed to the author. What is immediately obvious from Table 1 is that already some bulls have been genotyped more than once representing a squandering of funds. Figure 4 outlines the number of dairy bulls that were genotyped more than once up to the year 2010 across 10 different countries. Almost 700 genotypes were genotyped more than once; each genotype at the time cost approximately €160 implying a squandering of over €110,000.

Advantages: Ability to identify animals that have already been genotyped and thus engage with sharing of genotypes to avoid duplication of genotyping; no competitive advantage is gained by genotyping the same animal twice

Disadvantages: I cannot think of any disadvantage other than for some unknown reason not wanting others to know what animals have been genotyped in a given population. Information on what dairy animals are genotyped is generally freely available. Which dairy animals are genotyped and on what genotype platform in the US is freely available at https://www.cdcb.us/eval.htm

Table 1. A small section of the international list of beef bulls genotyped

ANIMAL_NAME	ID	DOB	Brd	IRL	UK	FRA
BLUEBELL AIGLON	CHLFRAM007185101623	24/01/1985	CH	Want		
ABOUKIR	CHLFRAM007185119662	07/01/1985	CH	Illum_HD		
COMMANDEUR	CHLFRAM007187126401	01/01/1987	CH	Have DNA		
BANDIT	LMSFRAM008786003322	14/02/1986	LM	Want		Illum_54K
IMPERIAL	LMSFRAM008793000421	10/01/1993	LM	Want		Illum_54K
TANHILL RUMPUS	LMSIRLM000000FBR092	24/04/1980	LM		Illum_HD	
ESPOIR	LMSFRAM008789003720	02/03/1989	LM			
HIGHLANDER	LMSFRAM001692111209	01/01/1992	LM	Illum_HD		Illum_54K
OMAR	LMSFRAM001930098242	24/12/1998	LM	Illum_HD		Illum_54K
KILKELLY DUKE	AANIRLM272061330257	01/03/2007	AA	Illum_HD		
DELFUR T-BONE	SIMGBRM523461601799	09/04/2006	SI	Illum_HD		

types benefit the exporting country more. The greater the genomic relationships between the reference population and the candidate population the greater, on average, will be the accuracy of the genomic predictions (Habier et al., 2007; Pszczola et al., 2012). Therefore, if the back-pedigree of animals being exported into a country exists within that country's reference population (assuming they also have phenotypic measures such as EPDs from pedigree or other descendants) then the accuracy of genomic predictions for those candidate animals will be greater. Hence the sharing of genotypes benefits both countries; the exporting country receives more accurate genomic proofs on their candidate bulls while the importing country gains access to potentially genetically elite individuals.

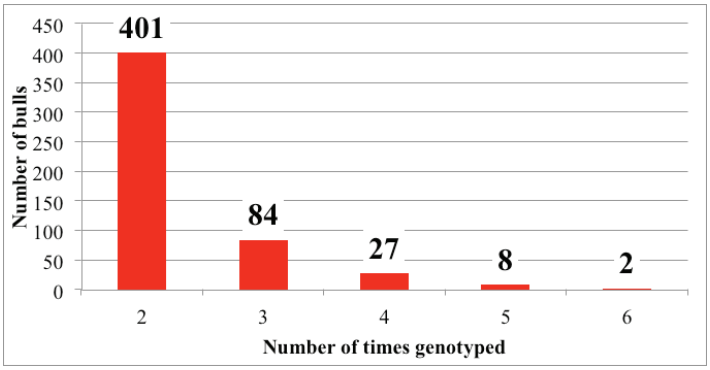


Figure 4. Presentation by Donagh Berry at the INTERBULL business meeting in 2010 on the number of animals that had been genotyped more than once across ten participating countries. The cost per animal genotype at that time was €160

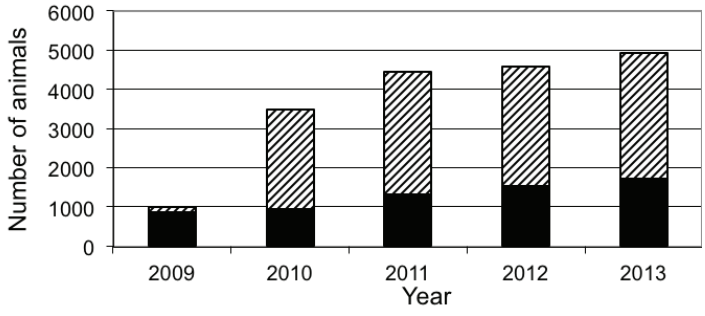


Figure 5. Number of bull genotypes included in the Irish genomic selection reference population for each year which were genotyped by Irish funding (black bars) or obtained through bilateral sharing (diagonal bars)

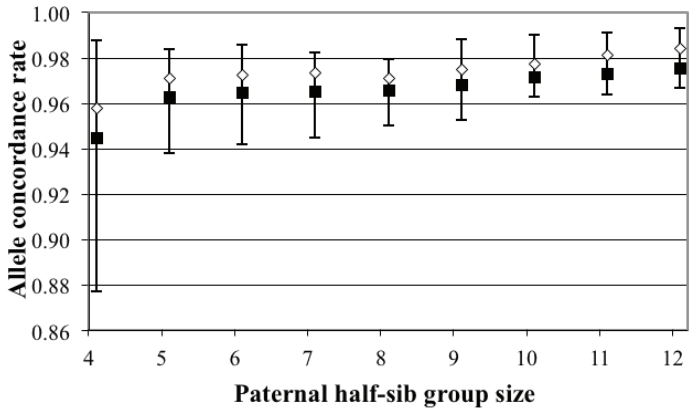


Figure 6. Mean animal allele concordance rate for Illumina Bovine50 Beadchip genotypes across different paternal half-sib progeny group sizes. Errors bars represent the, within animal, lowest and greatest mean concordance rate. Also represented (diamonds) is the mean animal allele concordance rate for a subset of the data across different parental half-sib progeny groups sizes when the paternal grand sire's genotype is also available (Berry et al., 2014a).

Sharing of genotypes is also advantageous even if the animal has no phenotype in the country. This is because the genotype of a non-genotyped animal (but potentially with a phenotype) can be imputed from its progeny genotypes. Figure 6 shows that parental alleles can be imputed with, on average, ≥96% accuracy if genotypes on ≥5 progeny are available (Berry et al., 2014a). Sharing of different genotype platforms (i.e., different densities or different commercial providers) can also relatively easily be facilitated through imputation across genotyping panels (Druet et al., 2010; Berry et al., 2014c). Furthermore, access to genotypes of animals even with no phenotypes can be used to improve the accuracy of imputation of their descendants or pedigree genotyped on lower density panels.

The precedence already exists in the sharing of genomic information. Sharing of genotypes in dairy is occurring among many populations. The 1000 bulls' genome project has collated to-date sequence data from over 1000 dairy and beef bulls. Furthermore, several thousand SNP and microsatellite genotypes were collated for the development of algorithms to convert SNP data to microsatellite data for parentage testing (McClure et al., 2013). This approach has benefited the en-

tire global cattle industry by eliminating the necessity to SNP genotype animals already genotyped back-pedigree for microsatellite. The introduction of this tool in Ireland in March 2013 has already saved the beef industry €200,000 by not having to re-genotype back-pedigree. Ireland is moving to parentage testing using just SNP data in both dairy and beef cattle. The advantage of this approach is that the SNPs, if undertaken as part of a larger panel, can also be used for genomic selection. Also, in Ireland the custom genotyping panel includes almost 2,000 research SNPs which can be used to validate in an independent commercial population. This can be particularly useful to elucidate, at no cost, if any SNPs strongly associated with novel traits are also antagonistically correlated with other performance traits in commercial populations.

Sharing of genotypes of young bulls can be of particular benefit if the genotypes are run through each country's genomic prediction equations and the genetically elite animals identified and subsequently imported. Such an approach benefits the exporter (sells the germplasm) and importer (access to genetically elite germplasm). However as previously alluded to, this approach is best achieved if the genotypes and phenotypes of the back-pedigree of these candidate animals are already in the importing country's genomic reference population.

At the very least the genotypes of each animal for the SNP parentage panel should be available without restriction. This panel cannot be used in genomic selection but is extremely useful in parentage testing.

An example of a document that could be used in the bilateral sharing of genotypes is given in Appendix I.

Advantages: The reference population size can be increased dramatically; in the case of the Irish dairy genomic selection breeding program, the size of the reference population was increased 300% (Figure 5). This will increase the accuracy of genomic evaluations (Figure 3). The marginal benefit of additional genotypes is greater when the reference population is smaller (Figure 3) as is usually the case for novel traits. For the (larger) exporting country the accuracy of genomic predictions on their candidate animals in the importing country is, on average, expected to be greater. The approach of sharing of genotypes should not be construed as an approach to facilitate the generation of genomic breeding values for bodies that decided not to invest in genomic selection; it involves bilateral sharing so there must therefore be (equal) investment. Genotypes can also be used to achieve more accurate imputation from lower density panels.

Disadvantages: Populations with a very large reference population may have little to gain from sharing of genotypes if they already have most of the other available genotypes in their reference population, Furthermore, the marginal benefit of additional genotypes in a reference population diminishes as the size of the actual reference population increases (Figure 3). There is still however a marginal benefit of additional genotypes on phenotyped animals even with many thousands of animals in the reference population. There is sometimes a perception that genotype sharing should not be undertaken because it was expensive to generate the population; however it is usually exactly the same expense for the other country assuming equal numbers of genotypes are shared. Sharing is a less expensive way to achieving a large reference population.

Sharing of Phenotypes. Many dairy cow population share phenotypic information through the international genetic evaluation at INTERBULL. Some beef populations also share phenotypic data via INTERBEEF, Breedplan and Pan-American initiatives.

Advantages: access to "phenotypes" on a large population of animals which increases the accuracy of genomic prediction with the marginal benefit being greatest when the reference population is relatively small as is usually the case for novel traits. The sharing of phenotypes and genotypes can also be used as an independent population to evaluate the precision and robustness of developed genomic selection algorithms. This is particularly important for novel traits where the population size is small.

Disadvantage There is background intellectual property associated with the generation of phenotypes, especially for novel traits and acquiring such phenotypes are usually costly. There is therefore reluctance among some to provide these data free of charge to others. To overcome this however the approach described previously on equal exchange of genotypes could be imposed for phenotypes. This however is only sensible if undertaking univariate (within country) genetic/genomic evaluations and excluding phenotypes from a multi-population evaluation would not be recommended. Again to overcome this, a price per unit phenotype could be generated; this could be relatively easily achieved using selection index theory. Then a value on each population's

correlations with other populations could be generated. The consortium may purchase these phenotypes or may pay an annual licensing fee to have access to these data for use in the multi-population evaluation. The price paid per population will differ based on the information content of the phenotypes (i.e., coheritability between populations) but also on how much that population is also contributing to the database of phenotypes.

Sharing of Genomic Keys. Collection of novel traits is generally a costly exercise and therefore the number of traits collected is usually limited. For example a population may deeply phenotype for one health trait but not for others. Sharing of genomic keys among populations that have phenotyped for a different suite of novel traits could provide potentially useful information on the likely correlated responses in other (not measured) novel traits. The validity of genomic keys from other populations could be relatively easily tested by phenotyping a smaller number of animals and relating their phenotypic values to those predicted from the shared genomic keys. Furthermore, visibility on the genomic keys from other population could help inform genomic prediction algorithms in that population for the same phenotype and therefore place greater emphasis on genomic regions detected as significant from more than one populations. Combined genomic keys can also provide a greater in depth knowledge of the underlying biological pathways governing differences in performance facilitating more powerful biological pathway analysis.

Advantages: Could remove the necessity to phenotype for all novel traits possible but could also be useful in elucidating the genetic merit of a population for example for diseases that currently do not exist in that population.

Disadvantages: There may be discontent in the sharing of genomic keys that required considerable resources to generate. Agreements can be put in place *a priori* on either selling the genomic keys or direct sharing (with some financial remuneration if differences exist between populations in the size of the reference population of the cost of generating the phenotypes).

Pan International Bull List to Increase Connectedness. Genetic connectedness is fundamental for the estimation of precise genetic correlations among populations (Berry et al., 2014b). There may be an advantage of generating a small list of bulls (varying every year with some crossreference bulls across years) that should be used in different populations to improve connectedness. Connectedness algorithms could be used to identify populations that could benefit most from such an initiative; such algorithms are commonly used to improve connectedness between flocks in sheep (Fouilloux et al., 2008). Such an initiative is particularly important for novel traits which tend to be recorded mainly in research herds; thus a pan-global list of bulls for recommended use in research herds could be generated. Only a few progeny need to be generated thereby having a likely minimal impact on the objectives of the research projects.

Advantages: More precise estimates of genetic correlations among populations necessary for inclusion in multi-trait genomic evaluations across populations to increase the accuracy of genomic predictions

Disadvantages: Could be difficult to reach a consensus on such a list of bulls given the likely different breeding objectives in different populations and may reduce the statistical power of the experiment.

Validation or Fine-mapping of Putative Causal Variants. Genomic selection requires the continual regeneration of genomic predictions including more recent generations of phenotyped animals in the prediction process. Moreover, for novel traits generally measured in research herds, the genomic relationships between these animals and the candidate population may be low; the weaker the genomic relationship between the animals in the reference population and the candidate population the lower will be the accuracy of genomic predictions (Habier et al., 2007; Pszczola et al., 2012). The reason the accuracy of genomic predictions is expected to decline over generations is due to recombination during meiosis because the SNPs exploited in genomic predictions are very unlikely to be the causal mutation and thus the linkage disequilibrium between the genotyped SNP and the true causal mutation can break down during meiosis. Hence, many research projects are engaged with attempted to locate the actual causal mutation thereby avoiding the necessity to continually re-estimate the genomic prediction equations. To facilitate the discovery of causal mutations, a very large population of animals is required to ensure adequate statistical power. Therefore, few, if any, animals exist to validate the discoveries or fine-map the genomic regions further. Different populations tend to have different linkage phases so therefore using alternative

populations could be extremely beneficial in fine-mapping further and eventually identifying the causal mutation if segregating in the validation population.

Advantage: Detection of the causal mutation or mutations in very strong linkage disequilibrium with the causal mutation should increase the accuracy of genomic predictions (especially across breeds) and the predication accuracy will be less subjected to erosion over generations.

Disadvantage: Many think they will make millions out of patenting of causal mutations. A good example of how this does not always materalise is the K232A polymorphism in DGAT1 which has a very large effect

Table 2. Number of lactations (N) as well as the mean, genetic standard deviation, heritability (h^2) and repeatability (t) of dry matter intake in all countries (i.e., all countries) or each individual country.

Country	N	Mean (kg DM/day)	σ_g (kg DM/day)	h^2	t
Cows					
All	10,641	19.7	1.13	0.27 (0.02)	0.66 (0.01)
Canada	411	22.2	1.01	0.11 (0.11)	0.46 (0.06)
Denmark	668	22.1	1.48	0.46 (0.12)	0.62 (0.04)
Germany	1141	20.2	0.64	0.16 (0.06)	0.84 (0.05)
Iowa	398	23.5	1.48	0.58 (0.12)	
Ireland	1677	16.7	0.88	0.29 (0.07)	0.64 (0.02)
Netherlands	2956	21.4	1.15	0.38 (0.04)	0.54 (0.03)
UK	2840	17.4	1.07	0.30 (0.06)	0.72 (0.02)
Wisconsin	447	24.9	0.90	0.19 (0.13)	
Australia	103	15.6			
Heifers					
Australia		8.3	0.77	0.39 (0.08	
New Zealand		7.6	0.66	0.25 (0.07)	

Table 3. Genetic correlations (below diagonal; standard errors in parenthesis) between dry matter intake measured in groups of countries[1] as well as the number of sires common (above diagonal; sires plus maternal grandsires in common in parenthesis) between the groups of countries.

Region	North-America	EU high-input	EU low-input	Grazing
North-America		39 (72)	4 (10)	6 (8)
EU high-input	0.76 (0.21)		125 (144)	23 (28)
EU low-input	0.79 (0.38)	0.84 (0.14)		4 (4)
Grazing	0.14 (0.43)	0.33 (0.20)	0.57 (0.43)	

[1] North-America = Iowa + Wisconsin + Canada; EU-high input = Netherlands+Germany+Denmark+high input feeding treatment in the UK; EU-low input= low input feeding treatment in the UK; Grazing = Ireland + Australia;

on milk production traits in dairy cattle (Berry et al., 2010). Royalties must be paid if this extremely large genomic effect is to be used in a breeding program but to my knowledge few, if any bodies actually exploit this mutation in their breeding program. There is a growing consensus that discovered causal mutations should be published in the scientific literature. An alternative it to retain the mutation as a trade secret within the company that made the discovery. The downside of this approach is that others may detect the mutation in the near future and publish it.

Case Study of International Genetic and Genomic Evaluations in Dairy Cows for a Novel Trait– Dry Matter Intake

Despite the large contribution (~60%) of feed to the variable costs of production in dairy cattle systems (Ho et al., 2005; Shalloo et al., 2004), feed intake is currently not explicitly included in the breeding goal of any dairy cattle population. This omission is principally due to an absence of sufficient feed intake information to estimate breeding values of individual animals. Collation of international data on feed intake and associated information from research herds and nucleus breeding herds is one approach to increase the quantity of feed intake data available for estimation of breeding values. This was the motivation of the global Dry Matter Intake initiative (gDMI) participated in by 9 countries. A total of 224,174 feed intake test-day records from 10,068 parity one to five records from 6,957 Holstein-Friesian cows, as well as records from 1,784 growing Holstein-Friesian heifers were collated from 9 countries in the US, Europe and Austral-Asia. Animal and back-pedigree genotypes were also pooled (Pryce et al., 2014) with the aim of undertaking an international genomic evaluation for feed intake which is still being researched (de Haas et al., 2014). Genetic parameter estimates for the different populations are in Tables 3 and 4, respectively (Berry et al., 2014b). Of less specific interest here are the actual results, but the point being made is that nine countries understood that the only way to achieve accurate genomic predictions for this novel trait was to pool their respective datasets. The same approach can be easily applied to other novel phenotypes or to different breeds. For example three countries pooled information on milk quality to derive more accurate rapid predictors of these novel traits from infrared spectroscopy in milk (Soyeurt et al., 2011); could a similar approach be adopted in the prediction of meat quality using rapid measures? The number of countries participating in prediction of milk quality initiative has since more than doubled. This is because the benefit of collaborating far outweighs the benefits of not.

Four Simple Steps Required for International Collaboration in Genotype Sharing to Happen!

1. Decide whether or not you want accurate genomic predictions for your breed or population – if so then international collaboration is the best approach to achieve this
2. Email Donagh.berry@teagasc.ie if you are willing to participate in a publically available excel spreadsheet on what animals are genotyped in your population and on what genotyping platform
3. You can also let it be known whether or not you are willing to share genotypes, either bilaterally or multilaterally.
4. Exchange genotypes

Literature Cited

Berry, DP. 2012. Across-breed genomic evaluations in cattle. *European Association of Animal Production.* Bratislava, Slovakia, *27-31 August 2012. P 127*.

Berry, D.P., McParland, S., Kearney, J.F., Sargolzaei, M., and Mullen, M.P. 2014a. Imputation of un-genotyped parental genotypes in dairy and beef cattle from progeny genotypes. *Animal* (in press).

Berry, D.P., Coffey, M.P., Pryce, J.E., de Haas, Y., Løvendahl, P., Krattenmacher, N., Crowley, J.J., Wang, Z., Spurlock, D., Weigel, K., Macdonald, K, and Veerkamp, R.F. 2014b. International genetic evaluations for feed intake in dairy cattle through the collation of data from multiple sources. *J. Dairy Sci.* (in press).

Berry, D.P., Howard, D., O'Boyle, P., Waters, S., Kearney, J.F. and McCabe, M. 2010. Associations between the K232A polymorphism in the diacylglycerol-O-transferace 1 (DGAT1) gene and performance in Irish Holstein-Friesian dairy cattle. *Irish J. of Agric. & Food Res.* 49:1-9.

Berry, D.P., McClure M.C., and Mullen, M.P. 2014c. Within and across-breed imputation of high density genotypes in dairy and beef cattle from medium and low density genotypes. *J. Anim. Breed. Gen.* (in press).

Calus, M.P.L., Berry, D.P., Banos, G., de Haas, Y. and Veerkamp, R.F. 2013. Genomic selection: the option for new robustness traits?. *In: Advances in Animal Biosciences, Cambridge University,* 4 : 618-625.

Daetwyler H.D., Villanueva B. and Woolliams J.A. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* 3, e3395.

de Haas Y., Pryce, J.E., Berry, D.P., and Veerkamp, R.F. 2014. Genetic and genomic solutions to improve feed efficiency and reduce environmental impact of dairy cattle. *Proc. World Cong. on Gen. Appl. to Livest. Prod.* Vancouver. *August 2014.*

Druet, T., Schrooten, C., and de Roos A.P. 2010. Imputation of genotypes from different single nucleotide polymorphism panels in dairy cattle. *J. Dairy Sci.* 93:5443-5454.

Fouilloux, M.N., Clément, V., and Laloë, D., 2008. Measuring connectedness among herds in mixed linear models: From theory to practice in large-sized genetic evaluations. *Genet. Sel. Evol.* 40:145-159.

Goddard, M 2009. Genomic selection: prediction of accuracy and maximization of long term response. *Genetica* 136: 245–257.

Habier D., Fernando R. and Dekkers J. 2007. The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177: 2389–2397.

Ho, C., Nesseler R, Doyle P, Malcolm B. 2005. Future dairy farming systems in irrigation regions. *Austral. Farm Bus. and Manage. J.* 2: 59–68.

Karoui S., Carabaño M.J., Díaz C. and Legarra A. 2012. Joint genomic evaluation of French dairy cattle breeds using multiple-trait models. *Genet. Sel. Evol.* 44:39.

Saatchi M., McClure M.C., McKay S.D., Rolf M.M., Kim J., Decker J.E., Taxis T.S., Chapple R.H., Ramey H.R., Northcutt S.L., Bauck S., Woodward B., Dekkers J.C.M., Fernando R.L., Schnabel R.D., Garrick D.J. and Taylor J.F. 2011. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. *Genet. Sel. Evol.* 43:40.

McClure, M.C., Sonstegard, T.S., Wiggans, G.R., Van Eenennaam, A.L., Weber, K.L., Penedo, C.,T., Berry, D.P., Flynn, J., Garcia, J., Carmo, A.S., Regitano, L.C.A., Albuquerque, M., Silva, M.V.G.B., Machado, M.A., Coffey, M., Moore, K., Boscher, M.Y., Gene 2013. Imputation of microsatellite alleles from dense SNP genotypes for parentage verification across multiple Bos taurus and Bos indicus breeds. *Frontiers in Genet.* 4 : 1-11.

McParland, S., Berry, D.P. and Kearney, J.F. 2014. Retrospective analysis of the accuracy of genomic selection in Irish dairy cattle. *Proc. Irish Agric. Res. Forum.* March 10-11, 2014. Tullamore, Ireland.

McParland, S., Kearney, J.F., Rath, M. and Berry, D.P. 2007. Inbreeding trends and pedigree analysis of Irish dairy and beef cattle populations. *J. Anim. Sci.* 85:322-331.

Prendiville R and McHugh, N. 2014. Comparative live weight, body condition score at breeds, onset of puberty and age at first calving for heifers of high and low maternal Index. *Proceedings of the Irish Agricultural Research Forum.* March 10-11, 2014. Tullamore, Ireland. Pp117.

Pryce, J.E., Johnston J, Hayes B.J., Sahana G, Weigel K.A., McParland S., Spurlock D., Krattenmacher N., Spelman R.J., Wall E., Calus M.P.L. 2014. Imputation of genotypes from low density (50,000 markers) to high density (700,000 markers) of cows from research herds in Europe, North America, and Australasia using 2 reference populations. *J. Dairy Sci.* 97: 1799-1811

Pszczola M., Strabel T., Mulder H.A. and Calus M.P.L. 2012. Reliability of direct genomic values for animals with different relationships within and to the reference population. *J. Dairy Sci.* 95: 389–400.

Purfield D.C., Bradley, D.G, Evans R.D., Kearney, J.F. and Berry, DP. 2014. Genome-wide association study for calving performance using high density genotypes in dairy and beef cattle. *BMC genomics*. (submitted).

Rendel J., and Robertson A. 1950. Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. *J. Genetics*. 50:1-8.

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

Shalloo L., Dillon P., Rath M. and Wallace, M. 2004. Description and validation of the Moorepark Dairy Systems Model (MDSM). *J. Dairy Sci*. 87:1945-1959.

Soyeurt, H., Dehareng, F., Gengler, N., McParland, S., Wall, E., Berry, D.P. and Coffey, M. 2011. Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries. *J. Dairy Sci.* 94: 1657-1667.

Spelman, R., Hayes, B.J. and Berry, D.P. 2013. Use of molecular technologies for the advancement of animal breeding: Genomic selection in dairy cattle populations in Australia, Ireland and New Zealand. *Anim. Prod. Sci.* 53: 869-875.

APPENDIX I.

Agreement

between

BODY1 NAME
BODY1 ADDRESS
BODY1 ADDRESS
BODY1 ADDRESS

hereinafter referred to as "**XXXXXX**"

and

BODY2 NAME
BODY2 ADDRESS
BODY2 ADDRESS
BODY2 ADDRESS

hereinafter referred to as "**YYYYYY**"

Agreement dated this [insert date] and continuing until this [insert date + three years] or until terminated under guidelines in Article 5.

1 *Purpose of the agreement*

XXXXXX and **YYYYYY** agree to collaborate in the area of genomic evaluation and selection namely through:

- exchanging information about methods used for genomic evaluation and selection in cattle,
- exchange of animals genotypes to avoid multiple genotyping of the same animals, and
- exchange of genotypes of animals in general.

2 *Exchange of information about methods used for genomic evaluation and selection*

(a) **XXXXXX** and **YYYYYY** agree to exchange the following information solely for the purposes referred to herein (the **Purpose**):

i. their respective methods used for genomic evaluation and selection in cattle;

ii. animal genotypes and the identity of animals proposed to be genotyped;

iii. the estimation of effects at the single nucleotide polymorphisms (SNP) and the conclusions derived for the corresponding breeding program; and

iv. such other information as is referred to herein (the **Confidential Information**)

(b) The Confidential Information exchange will take place on a regular basis at the conferences and other forums as agreed between the parties.

(c) For the avoidance of doubt any exchange of Confidential Information specifically excludes a license to a software package that one party may use to implement genomic evaluation or selection.

(d) Nothing in this agreement shall be construed as assigning or otherwise transferring any proprietary rights including Intellectual property rights in a party's Confidential Information to the other party.

3 Confidentiality

(a) All Confidential Information given by a party to the other party under the terms of this agreement is valuable information of the disclosing party and the receiving party undertakes to keep the Confidential Information secret and to protect and preserve the confidential nature and secrecy of the Confidential Information.

(b) A receiving party:

i. must not disclose Confidential Information of the disclosing parties to any person except as permitted under this Agreement;

ii. must not permit unauthorised persons to have access to the disclosing party's Confidential Information;

iii. must not make or assist or permit any person including its officers and employees, agents or advisors (Representatives) to make any unauthorised use, disclosure or reproduction of the disclosing Party's Confidential Information

iv. must take reasonable steps to enforce the confidentiality obligations imposed or required to be imposed by this Agreement and must co-operate with the disclosing party in any action that it may take to protect the confidentiality of the Confidential Information disclosed under this Agreement.

(c) A receiving party must only use the disclosing party's Confidential Information for the Purpose and must only disclose Confidential Information to its Representatives for the conduct of the Purpose and then only on a need to know basis.

(d) Each party must ensure that its Representatives do not do or omit to do anything which if done or omitted to be done by the receiving party would be a breach of the receiving party's obligations under this agreement.

4 Avoid multiple genotyping of the same animals

XXXXXX and **YYYYYY** will use the same genotyping technology and platform, namely the same SNP-chip. Currently, the Illumina Bovine SNP 50™ BeadChip and the Illumina Bovine SNP HD™ BeadChip will be used. **XXXXXX** and **YYYYYY** will exchange the identity of the animals they have or plan to genotype.

5 Exchange of animals genotypes

(a) **XXXXXX** is granted the right to obtain genotypic information of genotyped sires from **YYYYYY** and **YYYYYY** are granted the right to obtain genotypic information of genotyped sires from **XXXXXX**. Each party shall exchange approximately equal numbers of genotyped sires to the other. Each party, **XXXXXX** and **YYYYYY**, will retain ownership of the genotyping information they provided.

(b) The genotyping information **XXXXXX** obtains from **YYYYYY** may be used by **XXXXXX** for genetic evaluation in the **COUNTRY** base and scale and selection purposes only. All results and products originating from genotyping information obtained from **YYYYYY** belong to **XXXXXX**.

(c) The genotyping information **YYYYYY** obtains from **XXXXXX** may be used by **YYYYYY** for genetic evaluation in the **COUNTRY** base and scale and selection purposes only. All results and products originating from genotyping information obtained from **XXXXXX** belong to **YYYYYY**.

(d) **XXXXXX** and **YYYYYY** may extract the genotype for parentage testing SNPs and provide these to third parties for the purpose of validating parentage of animals in their respective cattle populations.

6. Termination

(a) **XXXXXX** and **YYYYYY** have the right to terminate this agreement by giving 3 months written notice to the other party.

(b) Termination of this agreement will be without prejudice to any other rights and remedies of the parties arising out of any default which occurs before the termination and will be without prejudice to any claim for money payable at the time of termination in respect of work done, genotyping information exchanges or liabilities incurred before the termination.

(c) Upon termination or expiration of this agreement, or the request of the disclosing party, the receiving party will deliver to the disclosing party (or with the disclosing party's prior consent, destroy or erase);

i. all material forms of the other party's Confidential Information (including biological or other samples) in its possession or the possession of any of its Representatives; and

ii. a statutory declaration made by an authorised officer of the party declaring that the provisions of this article have been complied with.

(d) Return of material forms of Confidential Information does not release a party or its Representatives from the confidentiality obligations set out in this agreement

(e) The obligations of confidentiality contained herein survive termination or expiration this agreement.

7 Dispute Resolution

In the event of any dispute between the parties in relation to the terms and conditions of this agreement, the parties will first seek to resolve such dispute by promptly giving notice of the dispute to the other party and in good faith endeavour to resolve such dispute. If the dispute remains unresolved for 20 days, the parties will first seek a resolution through the use of mediation. If the dispute still remains unresolved, a resolution through the use of arbitration shall be tried and only as a last resort, resolution is pursued through courts. Nothing in this Agreement will be interpreted as preventing a party from seeking urgent interlocutory relief through the courts to protect its interest in the Confidential Information disclosed to the other party.

8 Governing Law

This Agreement shall be governed by the laws of **COUNTRY**.

Signed for XXXXXXX:			**Signed for YYYYYYYY:**	
Signature	Date		Signature	Date
Signature	Date		Signature	Date