Use of Genome Editing in Animal Breeding Programs

J. Jenko[1], G. Gorjanc[1], G. Mészáros[1], C.B.A. Whitelaw[1], J.A. Woolliams[1], M.A. Cleveland[2] and J.M. Hickey[1].
[1]The Roslin Institute and R(D)SVS, The University of Edinburgh, Easter Bush, United Kingdom, [2]Genus plc, Hendersonville, Tennessee, USA

ABSTRACT: The purpose of this work was to evaluate the potential of using genome editing (**GE**) in conjunction with genomic selection (**GS**) for increasing the rate of genetic progress in livestock breeding programs. Results showed between 1.05 and 3.20 times higher response to selection when combining GE with GS as compared to using GS alone. In the short term these differences could be even higher - between 1.00 and 5.40 for generation 5. Traits determined by a smaller number of causative variants, had larger rates of genetic improvement under the GE approach compared to the traits controlled by more causal variants. Traits with lower heritability showed more benefit from GE than traits with high heritability. Improvement was larger with more edits per sire and more sires edited. GE has great potential for use alongside GS in livestock breeding programs.
Keywords: genome editing; genomic selection

Introduction

Genomic selection (**GS**) is routinely used in most advanced breeding programs to drive genetic progress. The commercial success of GS will result in huge data sets of densely phenotyped and genotype individuals being generated in several species. In the next decade it is not inconceivable that these huge data sets could comprise many hundreds of thousands or millions of individuals. Analysis of such huge data sets will enable large proportions of the causal variants for important traits to be identified. Using conventional selection methods to bring the favourable alleles at these causal variants to fixation will be slow because there are likely many variants and the low levels of recombination that occurs in livestock prevents all of the favourable alleles from arising in the selected individuals. Genome editing (**GE**) is a potential new way to overcome this problem.

Genome editing enables modification of genetic material in targeted ways (Niu et al., 2014; Cong et al., 2013). In the context of animal breeding one use of GE could be to fix a small number of undesirable alleles in individuals that have otherwise high breeding values. Such an approach could make GE very complimentary to GS. Individuals could be first selected on the basis of GS and then have some of their unfavorable alleles "fixed".

The objective of this study was to use simulation to quantify the potential of combining GE with GS in livestock breeding programs.

Materials and Methods

Simulation. Ten replicates of 78 different scenarios were simulated using AlphaDrop (Hickey and Gorjanc, 2012). A number of features were common across all scenarios: (i) the genome comprised 10 chromosomes each 1 Morgan in length; (ii) 50 generations were simulated and in each generation there was 1000 individuals (500 male, 500 female); (iii) in each generation 25 males were selected using genomic estimated breeding values (**GEBV**) that were predicted using genomic best linear unbiased prediction (**GBLUP**) to become the sires of the next generation, all 500 females were selected; (iv) the GBLUP prediction equation was trained using phenotype and genotype data of the 1000 individuals in the previous generation using 20,000 markers; (v) the causal variants effects were assumed known for GE; and (vi) GE involved making the sire homozygous for favourable allele of the largest causal variants that it was not already homozygous for. The different scenarios involved different number of causal variants (1,000 and 10,000), trait heritability ($\mathbf{h^2}$) (0.05, 0.15, and 0.30), and the number of genome edits per selected sire (0, 1, 5, 10, and 20). Additionally, three different strategies for the use of GE were tested: editing all selected sires, only the best 10 selected sires or only the worst 10 selected sires.

Validation. Response to selection was computed as a difference between the mean true breeding value of the current and the first generation and divided by the standard deviation of the true breeding values in the first generation. Different scenarios were compared using a relative change in response to selection in comparison to the scenarios with GS only (i.e. zero genome edits to selected sires). Additionally, prediction accuracies and average changes in allele frequency were calculated. Prediction accuracy was measured as the correlation between the true and the estimated breeding values, while average change in allele frequency was calculated as the average change of allele frequency for all the alleles between the two subsequent generations.

Results

Using GE in conjunction with GS resulted in bigger rates of genetic improvement in comparison to using GS alone and using many edits (i.e. 20) per sire was much more beneficial than using a few (Figure 1).

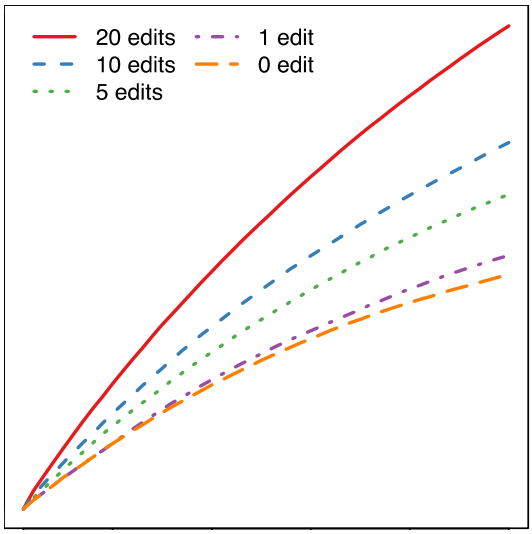

Figure 1. Response to selection over 50 generations of genomic selection for a trait with a heritability 0.3 under the effect of 10,000 QTL and different genome editing number on all of the 25 selected sires.

To enable comparison of the many scenarios in detail results are presented for scenarios with GE relative to scenarios with no GE after 5 generations (Table 1). The impact of GE was greatest when h^2 was lowest, when the number of causal variants was lowest, and when the number of edits per sire was highest. For the trait with a h^2 of 0.05 and 1,000 causal variants 20 edits resulted in a standardised genetic merit that was 5.4 times higher than the no GE scenario. In comparison 5 edits only resulted in a standardised genetic merit that was 2.4 times higher. Using 20 edits per sire for the trait controlled by 1,000 causal variants resulted in a standardised genetic merit that was 5.4 times higher than the no GE scenario when the h^2 was 0.05 but only 3.37 times higher when the h^2 was 0.30. Comparing the scenario with 1,000 causal variants versus 10,000 causal variants, when the h^2 was 0.05 using 20 edits per sire resulted in a standardised genetic merit that was 5.4 times higher than the no GE scenario for the 1,000 causal variants scenario and only 2.94 times higher for the 10,000 causal variants scenario. Despite GE being of least benefit for the trait with the high h^2 (0.30) and largest number of causal variants (10,000) the benefit of GE for this trait was still large. In fact the standardised genetic merit was 25% higher when using 5 edits and almost two times higher when using 20 edits compared to using GS only.

In comparison to editing only the top or bottom 10 selected sires, editing all of the selected sires resulted in the higher genetic improvement (results not shown). There was inconsistency with regard to whether editing the top 10 sires or the bottom 10 sires was better.

Table 1: Relative change in response to selection in scenarios with genome editing[§] in comparison to scenarios with no genome editing.

	h^2					
	0.05	0.15	0.30	0.05	0.15	0.30
Edit[¥]	1000 QTL			10,000 QTL		
0	1.00	1.00	1.00	1.00	1.00	1.00
1	1.40	1.29	1.19	1.12	1.09	1.00
5	2.40	1.95	1.74	1.50	1.45	1.25
10	3.40	2.67	2.33	1.94	1.73	1.50
20	5.40	4.19	3.37	2.94	2.32	1.96

[§]20 genome edits per each of 25 selected sires.
[¥]Number of genome edits per sire per generation.

Where GE had a strong impact on the response to selection (e.g. low h^2, small number of causal variants) the accuracy of the GEBV from GBLUP reduced with increasing generation number (Figure 2). This reduction in accuracy was not observed for scenarios where GE did not have as big an impact on response to selection.

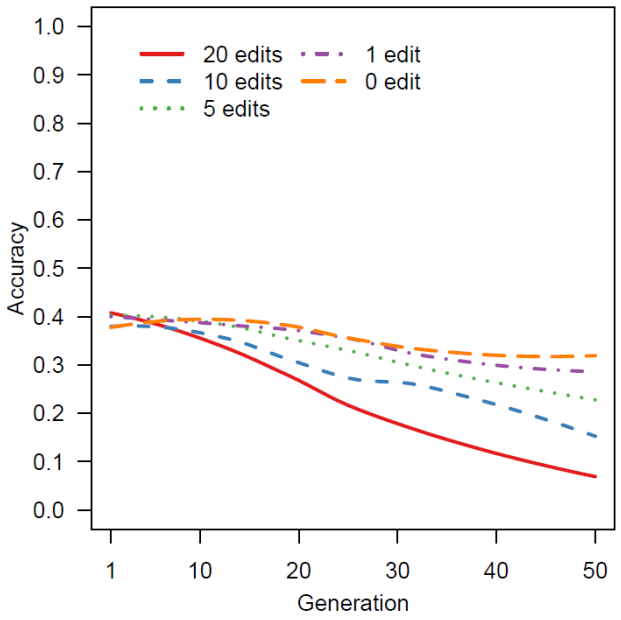

Figure 2. Accuracy of predicted breeding values over 50 generations of genomic selection for a trait with a heritability 0.05 under the effect of 1000 QTL and different genome editing number on all 25 selected sires.

Discussion

Genome editing is a powerful complimentary technology to genomic selection for increasing the rate of genetic improvement for polygenic traits in livestock. Its impact is greatest for traits with a lower h^2 and a lower number of causal variants but it is also powerful for traits with higher h^2 and many causal variants if these variants are accurately identified.

In this study the effects of the causal variants were assumed to be known with complete accuracy for GE, but not for GS. This high degree of accuracy undoubtedly contributed to the relative power of GE, but a perhaps more important reason for its power was the ability of GE to

increase the effective recombination rate such that favourable permutations of causal variants could be generated more often. One of the major barriers to genetic improvement for polygenic traits is that meiotic recombination occurs relatively rarely and thus permutations containing most or all of the favourable alleles take a long time to arise. By using GE to "fix" selected sires such permutations can be generated in much shorter periods of time.

When GE was powerful it resulted in the accuracy of GS becoming lower. There were two reasons for this. Firstly, the larger causal variants became fixed and thus no longer contributed to both the variance in GEBV and true breeding values. Larger effects are easier to estimate than smaller effects and consequently GEBV containing larger effects will be more accurate. Secondly, in the absence of sequence information the accuracy of GEBV depends on the correlation (linkage and linkage disequilibrium) between the markers in the genome and the causal variants. Where GE is powerful the correlation structure between the markers in the genome and the causal variants breaks down, which in turn reduces the accuracy of GEBV. This is one reason as to why sequence information in large quantities will probably be needed to enable GE have maximal efficacy as a compliment to GS in animal breeding programs.

The economic value of GS will enable huge data sets (e.g. many hundreds of thousands of individuals with phenotypes and sequence data) be assembled in many animal breeding programs. With such huge data sets it will be possible to finely map the causal variants for large proportions of the genetic variation for polygenic traits, or at least it will be possible to accurately identify the causal variants that are comparatively of larger effect. Having these causal variants accurately identified will be essential for GE. In this study we used small data sets (i.e. 1000 phenotyped and genotyped individuals), which would have resulted in rather inaccurate estimates for the causal variants and thus a poor performance for GE. For this reason we feel justified in our assumption that we knew the effects of the causal with complete certainty.

This is a first attempt at how one might use GE to improve polygenic traits in livestock breeding programs. Many other approaches are possible (e.g. using GE to create new variation). If the cost and efficacy of the molecular biology techniques for performing GE continues to improve it may soon become an important and widely used tool in animal breeding programs.

Literature Cited

Cong, L., F.A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P.D. Hsu, X. Wu, W. Jiang, L.A. Marraffini, et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science. 339:819–823.

Hickey, J.M. and G. Gorjanc. (2012). Simulated data for genomic selection and genome-wide association studies using a combination of coalescent and gene drop methods. G3. 2:425–427.

Niu, Y., B. Shen, Y. Cui, Y. Chen, J. Wang, L. Wang, Y. Kang, X. Zhao , W. Si, W. Li, et al. (2014). Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. Cell. 156:836–843.