Challenges from Day 1: Taking Genomic Evaluation to the Next Level

R. Mark Thallman
USDA, ARS, U.S. Meat Animal Research Center

United States Department of Agriculture
Agricultural Research Service

Terminology

- SNP BLUP = Marker Effects Model
- GBLUP is genomic relationship matrix applied to only genotyped animals
- ssGBLUP = GBLUP + ungenotyped animals
- Hybrid model = marker effects & imputation residuals of ungenotyped
- SuperHybrid model = marker effects & polygenic effects of ungenotyped
- Super-duper? Hybrid model = SuperHybrid + Extra-Polygenic effect

Comparison of Approaches

- The two approaches are more similar than they appear on the surface and would converge to the same predictions as the amount of information available to estimate marker effects approaches infinity
- The differences are in the priors
- The differences occur when there is insufficient information to estimate marker effects well

Comparison of Approaches

- Computational efficiency of superHybrid model is directly impacted by number of markers included
- Computational efficiency of ssGBLUP is **not** directly impacted by number of markers included
- **But**, it seems plausible that computational efficiency of ssGBLUP could be improved without changing predictions by dropping markers with estimated effects very close to zero, through reduction in number of core animals required.

Integrating Genotyped and Non-genotyped Animals into Same Analysis

- Far less trivial than it appears from a distance
- Potential for bias and scaling problems between genotyped and non-genotyped that are not directly solved by putting them together in the same analysis.
- The people integrally involved in the evaluations seem to be highly aware of these issues and working hard to manage them.

Benefits and Pitfalls of Marker Selection and Weighting

- With many genotyped animals with phenotypes (or phenotyped progeny) for the traits of interest, benefits and pitfalls are both small
- Otherwise, some weighted methods will outperform unweighted while many will perform worse.
  - We know quite a bit about how to choose methods that will perform better
  - The best choice depends on genetic architecture
  - We are probably a long way from a method that is optimal in any sense
  - Optimal solution requires balancing the benefits of a better-fitting model with the pitfalls of overfitting

Benefits of Adding Causal Variants

- No consensus
- Different problem than weighting/selecting markers on an existing chip
- But, essentially same considerations as on previous slide

Discovery Pipeline for New Variants and Systematic Evaluation of Whether They Are Better

- We are missing an opportunity

Multiple Trait Analyses

- Report accuracy of genetic RFI instead of dropping correlated traits from analysis