Use of DNA pooling for validation of genomic predictions

L.A. Kuehn, J.W. Keele, T.G. McDaneld and R.M. Thallman

USDA, ARS, U.S. Meat Animal Research Center

USDA United States Department of Agriculture Agricultural Research Service

The USDA is an equal opportunity employer

Genomics in NCE (National Cattle Eval)

- Maturing (no longer just 'new')
- Methodology has been refined from multi-step to single- (one-) step approaches
  - Were using two-trait MBV models and blending
  - Now primarily unweighted genomic prediction to form a G-matrix that replaces/compliments pedigree

Genomically-Enhanced EPDs

American Gelbvieh Association — AGA GE-EPDs

Gelbvieh and Balancer® Breed Average EPDs (For active sires, as of the Spring 2016 International Cattle Evaluation)

	Growth and Maternal EPDs									
	CED	BW	WW	YW	MK	TM	CEM	HP	PG30	ST
Gelbvieh	9	1.3	67	94	29	63	7	4	1	7
Balancer®	11	0.1	65	100	26	59	7	4	1	4

Ranch Tested. Rancher Trusted. Red Angus

CHAROLAIS — ACA Charolais Genomics

Update on Genomic For Quantitative

SimGenetics PROFIT THROUGH SCIENCE

American HEREFORD ASSOCIATION

BEEFMASTER GENOMICS +

American SHORTHORN ASSOCIATION Performance with Purpose

Production									
CED Acc % Prog	BW Acc % Prog	WW Acc % Prog	YW Acc % Prog	RADG Acc % Prog	DMI Acc % Prog	YH Acc % Prog	SC Acc % Prog	Doc Acc % Prog
+7 .95 40% 4747	+1.5 .98 55% 14201	+36 .97 90% 14469	+71 .96 85% 9287	+.25 .80 20% 20	-.31 .80 10% 20	+.9 .97 15% 4223	+.93 .96 40% 3555	+11 .90 55% 109

ANGUS THE BUSINESS BREED News & Information

LIMOUSIN NORTH AMERICAN LIMOUSIN FOUNDATION

AGI GE-EPD

BRANGUS GE-EPDs

EPDs are enhanced by genomic results generated by: GENESEEK zoetis

Genomics in NCE

- Some question on need for validation
  - How are genomic predictions working?
  - Before going further:
    - I am confident that current approaches to genomically enhanced EPDs (GE-EPD) are improving predictions
    - EPD accuracy is improving and with further use and implementation, genetic trends will improve
    - Validation more important in models where marker effect are predicted/fitted in a weighted fashion
- With that background, what can be done?

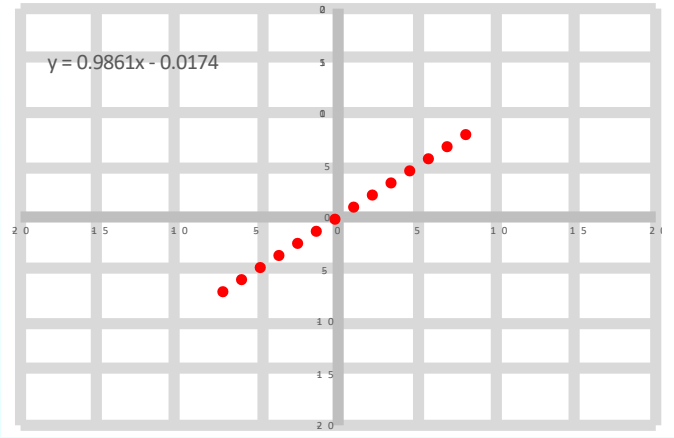
Diagnostics for GE-EPD

- From 2012 Genetic Prediction Workshop
  - Dorian Garrick
  - "Confirm that genetic predictions are behaving as expected"
  - Regression of GE-EPD on non-genomic EPD
    - Expectation of 1 (Method R, Reverter *et al.*, 1994)
    - Deviation from 1 indicates over- or under-shrinkage
  - Correlation of (GE-EPD – EPD) with EPD
    - Expectation of 0 (just as likely to increase as decrease)

Diagnostics for GE-EPD

- These diagnostics were used heavily in blending
  - Post-adjustment of EPD for genomic predictions (MBV/DGV)
  - Other diagnostic was regression of performance on EBV/DGV/MBV
    - Similar to across breed EPD system (Notter, 1991)
    - Also expected to be 1 (>1 under-prediction; <1 over-)
  - Helped to diagnose and correct scale problems between DGV and NCE EPDs

Example (GE-EPD on EPD)



Further diagnostics

- Current 'gold standard' in literature is prediction of progeny performance on parent EPD/EBV
  - Used heavily in dairy, swine, poultry
  - Progeny phenotypes are not allowed to contribute to GE-EPD prediction
    - Resulting regressions and SE measure bias and accuracy of genomic-enhancement
    - Can be compared to other methods of genomic-enhancement (or none)

Predicting progeny performance

- Problem in beef NCE
  - Many traits are difficult to collect younger generations (at least since genotyping began)
  - Examples
    - Maternal weaning weight
    - Actual carcass traits
    - Feed efficiency traits
    - Lifetime measures (e.g., stayability)

Possible solution

- External data
  - Should be independent from EPD
  - Can use similar diagnostic measures
    - Regression of performance on external EPD
    - Correlations of internal/external EPD
  - Research trials
    - USMARC
    - USDA grant project
- Commercial data

Example – Single Step Validation

- GPE data from across-breed EPD adjustment derivations used to validate
  - EBV derived from this program as a byproduct
    - Traits

Birth weight	Carcass weight
Weaning weight	Marbling score
Yearling weight	Ribeye area
Maternal milk (weaning wt)	Backfat depth

    - Single trait analyses with breed genetic groups
    - Single Step (SS), Multi-Step (MS), Non-genomic (NG) EPDs from American Angus Association

Statistics

- EBV correlated to EBV from MS, SS, NG analysis from AGI
  - Higher correlations indicate better prediction of genetic merit
- Also produced a breed-specific regression of performance on sire EPD from MS, SS, and NG using GPE performance data
  - Indication of bias in NCE relative to GPE data
  - Used two data sets:
    - All GPE and progeny born after 1999 (genotyped sires)

Results

	All bulls in GPE			Bulls with genotypes		
Trait	SS	MS				NG
Birth weight	0.60	0.				0.63
Weaning weight	0.43	0.				0.38
Yearling weight	0.52	0.				0.38
Maternal milk	0.40	0.				0.26
Carcass weight	0.35	0.				0.26
Marbling score	0.51	0.				0.55
Ribeye area	0.27	0.24	0.22	0.28	0.24	0.22
Backfat depth	0.42	0.24	0.38	0.44	0.25	0.40

Single-step correlations consistently higher for all traits, especially carcass

Multi-step correlations often lower than non-genomically enhanced

- AGI EBV correlations were > 0.98 for weight traits
- For carcass trait, AGI correlations were > 0.97 between SS and NG but < 0.91 for both with MS

Research data

- USMARC is an option
  - Certainly open to performing the same process with other breeds
  - Some data limitations
    - Not all breeds represented with high numbers of bulls sampled/progeny
    - Continues to improve but takes time
- Commercial data may be more viable

Commercial data

- Far more data could be available than we're collecting right now from seedstock sources
  - Commercial cow/calf
    - Cow weaning rate, cow fertility/longevity, days to calving – with commercial management
  - Feedlot
    - Pen feed consumption, treatment rates, mortality, morbidity, daily gains (pen/animal), days on feed
  - Abattoir
    - Quality grade, yield grade, hot carcass weight, dressing %, several possible camera predictions

Commercial data challenges

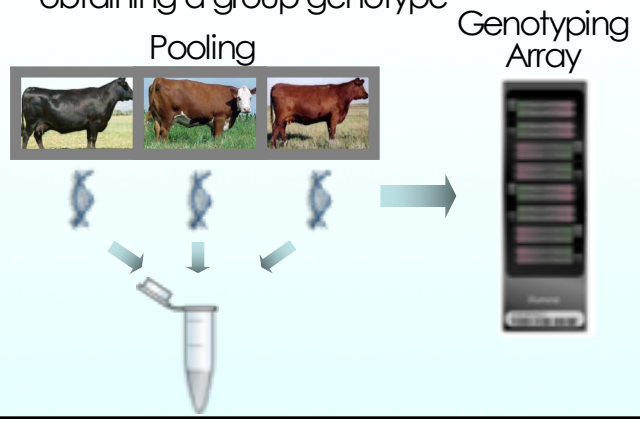
- Can be difficult to capture data
  - Not recorded, not linked to animal ID
  - Data may not be free/accessible
    - May require agreements/fees/animal tracking
- Animals are almost never genotyped and generally don't have known parentage
  - Not tied to NCE
  - Individual genotyping cost prohibitive

Possible solutions

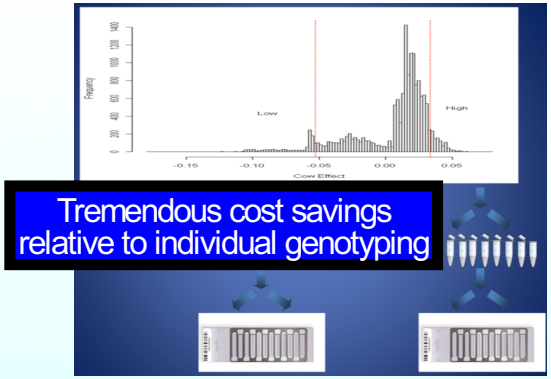
- Data capture
  - Demonstrate possibilities for commercial sector to utilize data from seedstock and genomic testing
  - More on that later
- DNA pooling
  - Use lower cost methods to genotype groups rather than individual genotyping

DNA pooling

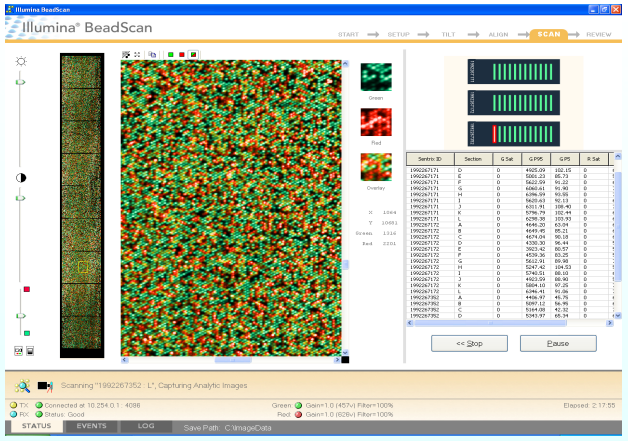
- Combine samples from multiple animals to obtaining a group genotype



DNA pooling – GWAS



Genotype calling



DNA pooling

- Derive allele frequencies based on color intensity
  - With individual genotyping red dye and green dye generally correspond to SNP alleles
  - Pooling use Pooling Allele Frequency (PAF)

    PAF = red / (red + green)
  - For research trials, markers with different frequencies between 'high' and 'low' phenotype pools may be predictive

Next step – DNA pooling and NCE

- Most beef breeds are currently using a single step procedure using a relationship matrix based on genomic (SNP) data in their National Cattle Evaluation (NCE)
  - Some applied as correlations among animals and others estimating marker effects
  - Genomic relationships between pools and genotyped animals or marker effects can be derived in a similar fashion
    - Bell et al. (2017) were able to detect sire contributions to pools using a modified genomic relationship matrix

Genomic relationships

- Can detect genetic potential of groups of cattle (pools) by identifying bulls
  - Complement to commercial marketing programs
  - Could further adjust/predict means using tools like across-breed EPD
  - Weighted average of GE-EPD based on sire contribution could provide a validation tool using pool means

DNA pooling validation

- Regressions of pool means for performance on weighted EPD/GE-EPD would provide a similar validation tool to individually genotyped/pedigreed populations mentioned earlier
  - Cost effective
    - 50-100 animal pools equivalent cost to 2-5 genotyped animals
  - Takes advantage of large industry data sets

Challenges

- Need to develop procedures to collect tissue samples and at least tie group means
  - Several tissue samplers are available but cost can be prohibitive
- Continued research on variation in contribution to pool
  - Larger pool size = less variation in contribution but also less variation in group means

Challenges – beyond validation

- Assuming genomic relationships to pools can be developed, pool averages can be used in NCE to inform GE-EPDs
  - We are actively researching this area to measure the impact of different source of error (e.g., animal contribution) on incorporating pool means into NCE
  - Not surprisingly, sires of animals in pools will see the most benefit
  - Designing prototypes for research trials

Challenges

- Need to get buy-in from the commercial entities to collect tissue, record/release group means, and potentially genotype the pools
- What is the value proposition for using information from DNA pooling in commercial management?
  - Breed composition
  - Management strategies

Breed composition

- Primary buying criteria for feedlot operations
  - Often use proxy indicators
    - Hair color
    - Ear length
  - These physical characteristics often drive weaned calf prices
- Emphasis on breed warranted based on research data

Breed differences

TABLE 2: BREED OF SIRE MEANS FOR 2015 BORN ANIMALS UNDER CONDITIONS SIMILAR TO USMARC

Breed	Birth Wt. (lb)	Weaning Wt. (lb)	Yearling Wt. (lb)	Maternal Milk (lb)	Marbling Score[a]	Ribeye Area (in^2)	Fat Thickness (in)	Carcass Wt.(lb)
Angus	86.1	567.2	1061.4	553.9	5.66	13.65	0.657	931.4
Hereford	89.6	548.5	1011.1	539.1	4.90	13.43	0.577	885.0
Red Angus	85.7	546.3	1025.5	557.3	5.40	13.36	0.623	899.8
Shorthorn	91.0	528.6	1000.5	551.6	5.04	13.77	0.500	886.1
South Devon	89.2	529.7	1001.2	570.1	5.04	14.05	0.437	858.2
Beefmaster	89.7	562.1	1014.1	549.8				
Brahman	97.2	583.7	1016.1	555.7	4.48	13.27	0.477	864.5
Brangus	89.0	556.9	1027.0	552.1				
Santa Gertrudis	89.7	559.7	1018.0	549.4	4.64	13.24	0.562	891.7
Braunvieh	89.7	537.3	998.1	570.3	5.13	14.62	0.451	870.1
Charolais	92.0	576.5	1045.8	545.3	4.90	14.70	0.448	921.3
Chiangus	89.8	539.9	1004.2	547.2	5.02	14.09	0.501	887.7
Gelbvieh	88.0	559.9	1036.3	562.9	4.93	14.45	0.496	902.9
Limousin	88.5	556.8	1011.3	549.8	4.65	14.77	0.476	897.7
Maine-Anjou	88.8	528.7	978.9	542.4	4.68	14.40	0.414	870.0
Salers	87.2	544.5	1010.5	558.8	5.33	14.23	0.468	872.6
Simmental	89.6	570.4	1049.5	555.7	5.04	14.47	0.482	920.5
Tarentaise	88.7	550.3	988.7	552.0				

[a]Marbling score units: 4.00 = Sl^{00}; 5.00 = Sm^{00}

Kuehn and Thallman, 2017

Breeds - Feed Intake and Gain

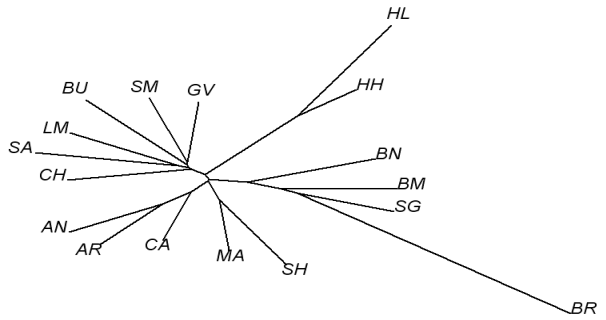
Breed	Steer ADFI (g)	Steer TESTADG (g)	Heifer ADFI (g)	Heifer TESTADG (g)
Angus	0	0	0	0
Hereford	-788 (286)	-35 (55)	-962 (266)	-21 (44)
Red Angus	-310 (275)	-66 (52)	-684 (255)	-86 (42)
Shorthorn	-997 (320)	-100 (61)	-1021 (298)	-98 (49)
South Devon	-1856 (666)	-274 (134)	-1576 (641)	13 (109)
Beefmaster	-771 (346)	72 (68)	-1556 (334)	-91 (56)
Brahman	-1321 (350)	-124 (68)	-1351 (319)	-185 (53)
Brangus	-173 (335)	-31 (65)	-585 (317)	-120 (53)
Santa Gertrudis	-569 (334)	22 (63)	-1039 (306)	-113 (50)
Braunvieh	-1488 (351)	-180 (68)	-1841 (305)	-299 (50)
Charolais	-521 (289)	-18 (55)	-876 (270)	-75 (45)
Chiangus	-1245 (334)	-81 (64)	-1049 (296)	-118 (49)
Gelbvieh	-1051 (278)	-72 (53)	-723 (253)	-114 (42)
Limousin	-1238 (281)	-5 (53)	-1471 (255)	-160 (42)
Maine Anjou	-1646 (334)	-150 (64)	-1101 (302)	-102 (50)
Salers	-1211 (333)	-136 (63)	-1176 (306)	-139 (51)
Simmental	-43 (288)	-19 (55)	-530 (275)	-68 (45)
Tarentaise	-1178 (678)	-150 (136)	-1926 (566)	-312 (96)

Retallick et al., 2017

Predicting breed/genetic factors

- Can substantially improve basic visual appraisal (color, etc.) using available genomic tools
  - High-throughput genotyping
- Use of genomic tools could extend to predictions of performance beyond breed identification

Breed Identification



Predicting breed

- Breed compositions are generally predicted reasonably
  - Really need far less than 50,000 markers
- Have been able to improve prediction accuracy since using markers that are more variable or fixed across breeds
- Can detect breed composition of a lot (as a pool) as well as individuals

Commercial application

- Knowing breed alone can facilitate management decisions
  - Endpoint differences
  - Growth potential
  - Intake differences
  - Ration/days on feed/selling criteria
  - Marketing grid
  - Implanting/feed additive decisions
- Individually genotyping animals cost prohibitive

Application

- Feedlot buyer obtains lot of 100 animals
  - Unknown origin (sale barn)
  - Obtain blood or ear sample from each animal
  - Cost: < $200 for DNA extraction
    < $100 for genotyping
  - Can $300 ($3/hd) be recovered?

Return on investment

- Value:

Scenario 1: ½ Charolais, ¼ Limousin, ¼ Angus

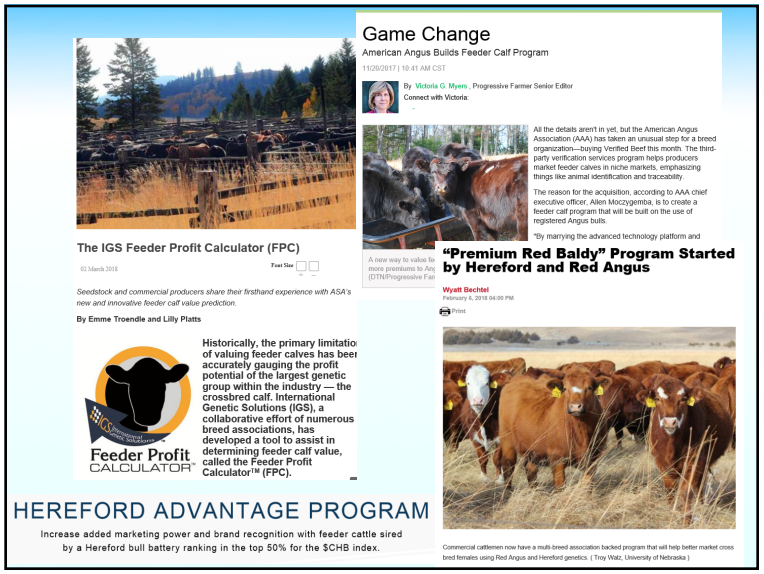
- High yield potential, carcass weight
- Lower quality grade opportunity
- 10 days less on feed, decreased feed and implant risk
  - Greater than $10/hd return from changing strategy

Scenario 2: ¾ Angus, ¼ Hereford

- High quality potential
- Carcass quality grid
- Higher feed cost, $10-$25 more per cwt
  - Target ration to increased marbling potential

Beyond breed

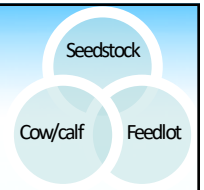
- Focus on breed as a tool for management is warranted
  - Breed is likely one of the largest sources of variation
  - However, there are extensive genetic differences beyond breed that could be utilized
- Need to capitalize on genomic variation in addition to that from breed differences
  - DNA pooling to establish sire contributions or genomic relationships



Genomic enhancement

- Improve management using genomic relationships to performance databases
  - Genomically enhanced predictions of group
  - Databases:
    - National cattle evaluations
      - Could work with current commercial programs
      - More difficult with crossbred pools
      - May require fee structures/collaborative agreements
    - Commercial producers (e.g., feedlots)
      - Commercial databases – record pool performance
      - Could predict future pools using own data
      - Tie together time and space to increase accuracy

Databases



- Most optimal solution (my opinion)
  - Develop agreements to share data across as many databases as possible
  - Synergistic relationship
    - Data from commercial sources would inform seedstock selection decisions
    - Seedstock genomic information and infrastructure would inform decisions in commercial sector
    - Data gathering in current structure of beef sector could be improved dramatically

Conclusions

- DNA pooling is a viable options for increasing data available to NCE for validation and, eventually, for performance databases
- Need to find ways to bring commercial segments 'into the fold' to implement these programs

Conclusions

- Breed prediction is both feasible and useful to commercial cattle programs and design of research programs
- While DNA pooling is still a developing research area, the potential for cheap genomic information in commercial application is tremendous
  - Breed prediction is possible now
  - Genomic prediction on the horizon

Conclusions

- Current commercial marketing programs would benefit from utilizing genomic relationships to performance databases
  - Trace back to sires that contribute to groups
  - Eventually genomically enhanced performance prediction
- Synergistic agreements would be highly beneficial and should be explored
- Similar tools could inform design and analysis of applied research programs

Questions

- Mention of a trade name, proprietary product, or specific equipment does not constitute a guarantee or warranty by the USDA and does not imply approval to the exclusion of other products that may be suitable.