#### The nature & scope of some whole genome analyses in US beef cattle

Dorian Garrick Lush Chair in Animal Breeding & Genetics National Beef Cattle Evaluation Consortium



#### Fools Gold or Real Gold

- What is the value of genomic selection technology ?
- How good will the genomic predictions be ?
  For which traits ?
  - For which breeds ?
  - From what size of panel ?
- How much will it cost?

#### **Three Phases**

- Training or Discovery Phase
  - Using Illumina 50k SNP chip to estimate g-EPD associated with various chromosome fragments
- Validation Phase
  - Determine the accuracy of prediction of all (or subsets of) fragments in "independent" datasets and/or subpopulations
- Commercialization Phase
  - Marketing of panel(s) that predict g-EPD or MBV
  - Incorporation of g-EPD into National Cattle Evaluation

#### **Training Phase**

- Requires 1,000's of animals with DNA samples plus individual and/or offspring performance relative to their cohorts
  - Genotyping will cost \$250-300,000 per 1,000 animals
     Phenotyping costs will vary according to the nature of the traits – easily exceed the cost of genotyping

#### Phenotyping Options (1 of 2)

- Use existing EPD (e.g. on current & historical AI sires)
  - Limits the trait options to those available for the breed
     Growth
  - BWD, WWD, WWM, YWD, perhaps later or mature wt Carcass
  - Fat/yield grade, marbling/quality grade, ribeye area
  - Reproduction
  - Calving ease (D & M), heifer pregnancy, stayability Feed requirements
  - Maintenance energy
  - Other
  - Docility

#### **Principal Datasets**

- Collection of some 2,000 Angus AI bulls put together by Jerry Taylor at University of Missouri and Merial Smaller collections of other breeds (eg Limousin)
- US MARC collection of some 2,000 recent Al bulls including 16 breeds
  - Angus, Beefmaster, Brahman, Brangus, Braunvieh, Charolais, Chiangus, Gelbvieh, Hereford, Limousin, Maine Anjou, Red Angus, Salers, Santa Gertrudis, Shorthorn, and Simmental
  - Validation rather than training population

#### Phenotyping Options (2 of 2)

Collect individual phenotypes for specific traits beyond those used in national cattle evaluation Enables use of non seedstock animals

#### **Principal Datasets**

#### Reproduction

- Funded by USDA-NRI
- Led by Dr Milt Thomas, New Mexico State University
- 800 Brangus heifers from Camp Cooley
- Growth and reproductive measures
- Collection of DNA & phenotypes from Rex Ranch Sisters of feedlot health project
- Facilitated by NBCEC (Drs Pollak CU & Spangler UNL)

#### **Principal Datasets**

- Beef Healthfulness (healthy Beef)
  - Facilitated by NBCEC
  - Funded by Pfizer Animal Genetics
  - led by Dr James Reecy at Iowa State University (ISU)
  - 2,200 Angus from Jack Cowley, Don Smith & ISU
  - Field work by Dr Alison Van Eenennaam UC Davis
  - Meat traits & taste panel Oklahoma State University
- Growth, carcass, meat traits, fatty acid, vitamin and mineral concentrations of ribeye

Reecy, 2009 BIF GPW

#### **Principal Datasets**

- Feedlot Health
  - Facilitated by NBCEC
  - Funded by Pfizer Animal Genetics
  - Led by Dr Mark Enns at Colorado State University
  - 2,900 Angus steers from Rex Ranch, fed in Colorado
  - Feedlot growth and performance
  - Flight speed, chute score, stress measures
  - Visual indicators of sickness, BVD, lung lesions

Enns, 2009 BIF GPW

### **Principal Datasets**

- Comprehensive phenotypes US MARC Drs Thallman, Snelling, Kuehn, Keele, Bennett etc
- Cycle VII offspring of Angus, Hereford and MARC III cows mated to Angus, Charolais, Gelbvieh, Hereford, Limousin, Red Angus, or Simmental sires
- Nebraska environment
- Growth, carcass, reproduction, feed intake, disease
- Comprehensive phenotypes Texas A&M
- F2 Angus-Nellore Dr Clare Gill
- Includes feed intake, temperament & behavior

## Principal Datasets

Feed Intake (efficiency, RFI etc)

- Portfolio of datasets championed by various researchers
  - University of Alberta, Dr Stephen Moore
  - Circle A/University of Missouri
  - University of Guelph
  - US MARC, Dr Cal Ferrell
  - Texas A&M, Dr Clare Gill

#### Scope of Training

- Many breeds represented but principally Angus Bos indicus poorly represented
- National environments represented for EPDs but a very limited number of local environments represented for other traits
- Carcass traits well represented
- Disease and fertility traits weakly represented (except respiratory/feedlot)

#### Nature of Training

- Least Squares Single SNP & stepwise analyses
  - Find single most informative marker
- Then second most informative marker etc
   Tends to overestimate effects, esp. in small studies
- Can't use all SNP at once
- Bayesian whole genome analyses
- Fit a portfolio of SNP all at once
- Repeatedly test models to see if they can be improved
- by adding a new SNP or removing an existing SNP
- Generates g-EPD for every SNP

#### Validation

- Almost always SNP that spuriously fit the data well
   Having a model that fits the training data well provides relatively little information about how good the prediction will be in new data
  - Many world-changing research discoveries are announced in news releases and then never-to-be-heard-of-again
- Training & Validation can be done together to quantify the likely confidence in predictions







5	50k within-breed predictions						
				-			
	Angus Al bulls Trait	Train 2 & 3 Predict 1	Train 1 & 3 Predict 2	Train 2 & 3 Predict 3	Overall		
	BFat	0.71	0.64	0.73	0.69		

Angus Al bulls	Train 2 & 3 Predict 1	Train 1 & 3 Predict 2	Train 2 & 3 Predict 3	Overall
BFat	0.71	0.64	0.73	0.69
CED	0.65	0.47	0.65	0.59
CEM	0.58	0.56	0.62	0.53
Marb	0.72	0.73	0.64	0.70
REA	0.63	0.63	0.60	0.62
SC	0.60	0.57	0.50	0.55
WWD	0.65	0.44	0.66	0.52
YWT	0.69	0.51	0.72	0.56

#### 50k within-breed predictions

- These predictions are characterized by correlations between genomic merit and realized performance from 0.5 to 0.7
  - They will account for 25 (0.5<sup>2</sup>) to 50% (0.7<sup>2</sup>) genetic variation
  - Compared to a trait with heritability of 25%, the genomic predictions would be equivalent to observing
    6 to 15 offspring in a progeny test
- Correlations of 0.7 are similar to the performance of genomic predictions in dairy cattle

#### 50k within-breed predictions

- These predictions are not as highly accurate as can be achieved in a well designed and managed progeny test, say with 100 or more offspring
- However, for many traits they are much more reliable for animals of a young age (eg prior to first selection) than is currently achievable from individual performance

#### Across-breed prediction

- Refers to the process of predicting performance for a breed or cross that was not in the training dataset
- Critical interest to those selecting breeds that are not well represented in the training populations
- May not be as reliable as within-breed predictions due to complexities associated with non-additive genetic effects (dominance and epistasis)
- Potential can be assessed by simulating the effects of major genes using real SNP genotypes on various populations



	50K	SI	VP D	atase	ets
	MB Populatio	on (N=	<u>924)</u>	PB Popul	ation (N=1,086)
	Angus	239	1 mil	Angus	1,086
	Brahman	10			
	Charolais	183			
1-1-	Hereford	78			
<b>Property</b>	Limousin	45			
Et.	Maine-Anjou	137			
	Shorthorn	97			
2.4	South Devon	135		Kizilkaya et	al, ASAS, 2009





S	ata	henotypes/	real 50k	
	<ul> <li>Correlation bet</li> </ul>	ween genomic &	true merit	
	50 QTL	Train in Multibreed Validate in Purebreed	Train in Purebreed Validate in Multibreed	
	Just QTL	0.95	0.96	

Effect of number	r of available mar	kers
50 QTL	Train in Multibreed Validate in Purebreed	Train in Purebreed Validate in Multibreed
Just QTL	0.95	0.96
QTL + Best markers	0.93	0.94
OTL + 50k	0.77	0.84

#### Simulated Phenotypes/real 50k Data

#### Effect of number of available markers

50 QTL	Train in Multibreed Validate in Purebreed	Train in Purebreed Validate in Multibreed
Just QTL	0.95	0.96
QTL + Best markers	0.93	0.94
QTL + 50k	0.77	0.84
Just Best markers	0.57	0.49
50k w/o QTL (real life)	0.39	0.42
	٢	Kizilkaya et al, ASAS, 200

Simulated P	henotypes/r	eal 50k
Dala		
Effect of numbe	r of QTL	
50k w/o QTL	Train in Multibreed Validate in Purebreed	Train in Purebreed Validate in Multibreed
50 QTL	0.39	0.42
100 QTL	0.29	0.31

Simulated P	henotypes/r	eal 50k
Data		
<ul> <li>Effect of number</li> </ul>	r of QTL	
50k w/o QTL	Train in Multibreed Validate in Purebreed	Train in Purebreed Validate in Multibreed
50 QTL	0.39	0.42
100 QTL	0.29	0.31
250 QTL	0.25	0.28
500 QTL	0.20	0.30
<ul> <li>These correlation</li> <li>Genes and &lt;10%</li> </ul>	ns account for < 20 6 variation if 500 g	0% variation if 50

Less informative than 1 observation on individual

### 50k across-breed prediction

- Current prospects are not promising
- Better results will be achievable with a higher density SNP panel (e.g. 500k rather 50k chip)
- More animals in the training analyses may also improve results
- Impact of dominance and epistasis may reduce reallife compared to simulated performance



# Reduced panel within-breed selection

- Two-stage Bayesian analysis
  - Run all 50k markers
  - in each of the three training sets (2&3, 1&3, 1&2)
  - Select the best 600 markers on model frequency and genomic coverage
  - Rerun the training and validation analyses using only the markers on the 600 marker panel



	000	
k vers	SUS 600	markers
Angus Al bulls	50k panel	600 markers
Trait	Overall	Overall
BEat	0.69	0.63
CFD	0.59	0.61
CEM	0.53	0.55
Marh	0.33	0.55
	0.70	0.07
REA	0.62	0.56
SC	0.55	0.51
WWD	0.52	0.49
YWT	0.56	0.55

#### **384 SNP Panels**

- Panels of 600 markers per trait for 8 traits would require a single panel of 4,800 markers
- Technology is moving such that larger panels are costing the same as smaller panels used to, rather than reducing the cost of smaller panels
- Significantly cheaper panels are currently limited to 384 (or less) SNP
  - Allow 100 or so of the best SNP for 3-4 key traits

Validation in 698 steers with carcass phenotypes           50         100         150         200         384           Trait
50         100         150         200         384           Trait
Trait              Marb         0.28         0.29         0.39         0.43         0.49           REA            0.43         0.43
Marb         0.28         0.29         0.39         0.43         0.49           REA             0.43
REA 0.43

# Validation in New Al Bulls

Validation	3-v	vay	275	
BFat	0.69	0.63	0.32	
Marb	0.70	0.67	0.59	
REA	0.62	0.56	0.58	
YWT	0.56	0.55	0.35	
CCWT			0.44	
HP			0.39	

#### Summary

- Breeds mostly Angus
- Traits mostly growth & carcass
- Environments mostly "not yours"
- 50k within breed (like 5-15 progeny)
- 50k across breed (like 1 individual record or 5 progeny)
- Reduced panel within breed (varies up to 50k accuracy)

## Acknowledgements

- Dr Rohan Fernando, Iowa State University
- Dr Kadir Kizilkaya, NBCEC-funded post-doc
- Dr Jerry Taylor, University of Missouri
- Igenity/Merial
- Pfizer Animal Genetics
- American Angus Association
- USDA-NRI Funding