

## State of Genomics including an update on NBCEC projects

Dorian Garrick  
dorian@iastate.edu

## National Beef Cattle Evaluation Consortium

- Mission: Develop and implement improved predictions so selection can enhance economic viability of US beef cattle producers
  - Establish & co-ordinate priorities for prediction
  - Consolidate research efforts
  - Streamline the development and adoption of new genetic evaluation methodologies
  - Identify new traits & technologies
  - Create decision-making tools

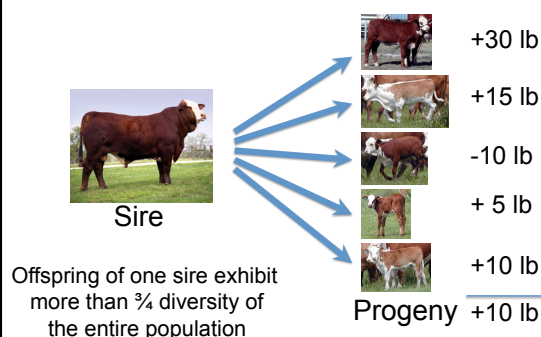
## NBCEC

- Funded for 10 years (to 2012) by a Special Grant to Cornell University on behalf of 4 core institutions
  - Colorado State, Iowa State, Univ of Georgia
  - Increasingly included other institutions
- Evolved to represent the consortium of researchers pursuing the NBCEC mission through USDA, other competitive grants, and breed association funding

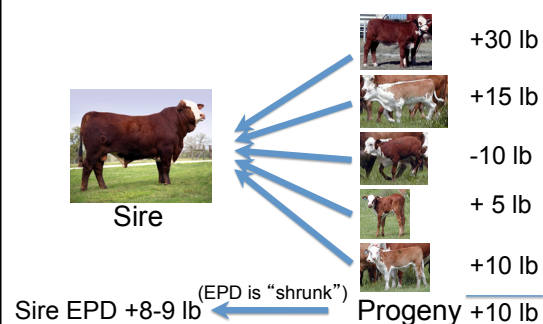
## Implementation of Genomic Prediction

- Major focus of NBCEC activity
- Other activities included extending the range of economically relevant traits beyond growth and carcass
  - Animal Health
  - Healthfulness of beef
  - Feed Efficiency
  - Reproduction
- All of which likely require genomic prediction

## Performance of the Progeny



## We learn about Parents from Progeny



EPDs on widely-used old sires are accurate



Sire

With enough progeny, this is usually close to the bulls true EPD (not surprisingly!)

Sire EPD +8-9 lb

Chromosomes are a sequence of base pairs

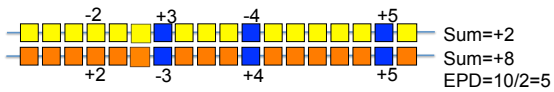
Part of 1 pair of chromosomes



Cattle usually have 30 pairs of chromosomes  
One member of each pair was inherited from the sire, one from the dam  
Each chromosome has about 100 million base pairs (A, G, T or C)  
About 3 billion describe the animal

- Blue base pairs represent genes
- Yellow represents the strand inherited from the sire
- Orange represents the strand inherited from the dam

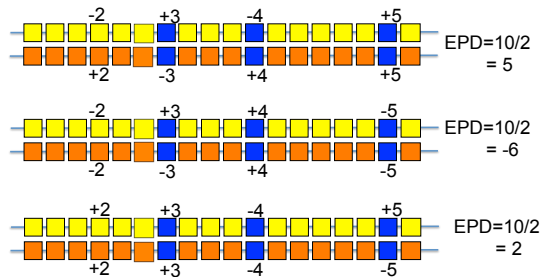
EPD is half sum of the gene effects



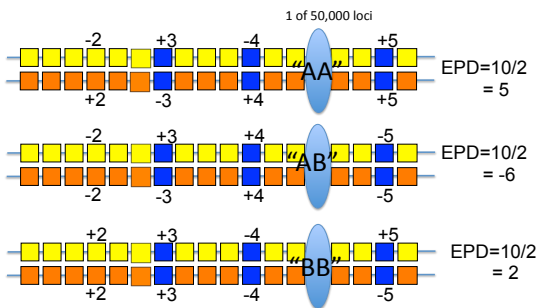
Blue base pairs represent genes

The EPD is half the sum of all these genetic values (half because offspring inherit a random half sample of each parents chromosomes)

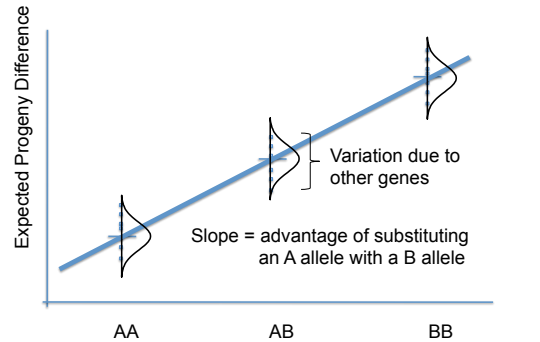
Consider 3 Bulls



SNP Genotyping the Bulls



Regress EPD on SNP genotype



## Genomic Analysis

- First (Training)
  - Collect 50k genotypes on 1,000 or more animals with EPDs
  - Compute the substitution effects for all 50k loci fitted at the same time
- Second (Validation or Implementation)
  - Genotype target animal and multiply the number of B alleles at each locus by the substitution effect at that locus

## Problems

- The training analysis will virtually perfectly predict the merit of all the animals in the training
  - But how reliable are new animals?
- The training analysis would be good if the markers were the causal variants
  - Many markers maybe should not have any effect
  - Those that are physically near causal variants may or may not be closely correlated (linkage disequilibrium or LD measures the strength of association)

## Cross Validation

- Partition our available data into training and validation subsets
  - Train in a large and validate in a small subset
  - Repeat so that every animal is in a validation
  - Form the subsets so close relatives are always together and not separated across training and validation (using clustering methods like k-means)

## Generally good predictions in Angus

Trait	Angus		
	(3,500)	Igenity	Pfizer
BirthWt	0.64	0.57	0.51
WeanWt	0.67	0.45	0.52
YearlingWt	0.75	0.34	0.64
Milk	0.51	0.24	0.32
Fat	0.70	0.50	0.56
REA	0.75	0.58	0.60
Marbling	0.80	0.65	0.57
CalvEase (D)	0.69		
CalvEase (M)	0.73		
Scrotal Circ	0.71		

## Angus predictions no good in Red Angus

Trait	Validating in American Angus	Validating in Red Angus
BirthWt	0.64	0.27
WeanWt	0.67	0.28
YearlingWt	0.75	0.23
Fat	0.70	0.21
REA	0.75	0.29
Marbling	0.80	0.21
CalvEase (D)	0.69	0.14
CalvEase (M)	0.73	0.18

Red Angus is more closely related to Angus than is Hereford

*Angus = ASREML 5-fold validation Red Angus = correlation Training on de-regressed EPDs Saatchi et al (GSE)*

## Angus predictions no good in Hereford

Raw Correlations standardized for EPD accuracy

Trait	Genetic Correlations	Validation in Hereford
Birthweight		0.18
Weaning wt		0.14
Yearling wt		0.17
Milk		0.02
Calving Ease D		0.10
Calving Ease M		0.19
Fat		0.07
Marbling		0.16
Ribeye Area		0.06
Scrotal Circum		0.03

Cannot predict across-breed using the 50k procedures that are adequate within breed

### Hereford predictions

Trait	Genetic Correlations	ASREML 4-fold x validn
Birthweight	0.43	0.43
Weaning wt	0.32	0.32
Yearling wt	0.30	0.30
Milk	0.22	0.22
Calving Ease D	0.43	0.43
Calving Ease M	0.18	0.18
Fat	0.40	0.40
Marbling	0.27	0.27
Ribeye Area	0.36	0.36
Scrotal Circum	0.28	0.28

Accuracy from 800 Herefords  
 Poorer than accuracy from 3,500 Angus  
 Size of the training population is important

### International Hereford Comparisons

Trait	Genetic Correlations	ASREML 4-fold x validn	Raw Corr 99 URG bulls	Raw Corr 75 CDN bulls	Raw Corr 59 ARG bulls (unrelated)	Raw Corr 41 ARG bulls (US-like)
Birthweight	0.43	0.22	0.40	0.15	0.24	
Weaning wt	0.32	0.13	0.07	-0.35	0.23	
Yearling wt	0.30	0.03	0.12	-0.27	0.32	
Milk	0.22	0.23	0.09	0.15	-0.03	
Calving Ease D	0.43		0.25			
Calving Ease M	0.18		0.44			
Fat	0.40	0.22	0.39	0.02	0.10	
Marbling	0.27	0.25	0.29	0.17	0.35	
Ribeye Area	0.36	0.01	-0.01	-0.11	0.18	
Scrotal Circum	0.28	0.36	0.19	-0.03	0.10	

Combined PanAmerican International Evaluation

### Training and validation:

- Weight Trait Project (WTP) as validation

Breed	Training	Validation
Black Angus	2,359	845
Limousin	1,655	530
Hereford	891	232
Simmental	102	190
Red Angus	86	90

- Training populations: Black Angus, Limousin, Hereford, 3 breed (A, L, H) or 5 breed mixes
- Simple correlation use to reflect accuracy

### The Accuracies of GEBV: Black Angus

Training	Black Angus	Limousin	Hereford	MIX-5B
Birth Weight	0.66	0.05	0.22	0.65
Weaning Weight	0.47	0.01	0.18	0.47
Yearling Weight	0.48	0.03	0.22	0.45
Maternal Milk	0.42	-0.06	0.05	0.43
Fat Thickness	0.44	*	0.11	0.45
Calving Ease Maternal	0.34	-0.04	0.06	0.34
Calving Ease Direct	0.58	0.03	0.03	0.55
Marbling	0.58	-0.03	-0.06	0.56
Rib Eye Muscle Area	0.45	-0.06	0.10	0.45
Scrotal Circumference	0.51	-0.04	-0.01	0.50

22

### The Accuracies of GEBV: Simmental

Training	Black Angus	Limousin	Hereford	MIX-5B	MIX-3B
Birth Weight	0.13	0.08	0.16	0.30	0.23
Weaning Weight	0.01	0.07	0.16	0.19	0.17
Yearling Weight	0.01	-0.06	0.15	0.31	0.19
Maternal Milk	0.16	-0.07	0.12	0.19	0.09
Fat Thickness	0.15	*	0.10	0.13	0.12
Calving Ease Maternal	0.06	0.02	-0.04	-0.11	-0.09
Calving Ease Direct	-0.09	0.01	0.09	0.10	-0.00
Marbling	0.08	0.05	0.18	0.26	0.18
Rib Eye Muscle Area	0.12	0.01	0.08	0.24	0.15
Scrotal Circumference	*	*	*	*	*

23

### The Accuracies of GEBV: Red Angus

Training	Black Angus	Limousin	Hereford	MIX-5B	MIX-3B
Birth Weight	0.22	0.16	0.31	0.53	0.35
Weaning Weight	0.18	0.22	0.22	0.41	0.30
Yearling Weight	0.23	0.33	0.22	0.47	0.39
Maternal Milk	*	*	*	*	*
Fat Thickness	0.13	*	0.00	0.16	0.01
Calving Ease Maternal	0.17	-0.17	-0.12	0.48	0.08
Calving Ease Direct	0.08	-0.06	0.32	0.50	0.21
Marbling	0.29	0.16	-0.10	0.33	0.24
Rib Eye Muscle Area	0.18	-0.25	0.03	0.26	0.10
Scrotal Circumference	*	*	*	*	*

24

## Conclusions – pooling US breeds

- Each breed needs its own training data
- Pooling breeds does not improve accuracy unless the breed has inadequate training data

## 700k will improve prediction

## Holstein-Friesian Results

Correlations	50k		700k		3-400k*
	BayesC.95	BayesB.975	BayesC.999	BayesB.999	BayesB.998
Milk	0.71	0.72	0.71	0.71	0.70
Fat	0.55	0.53	0.58	0.58	0.58
Protein	0.54	0.53	0.57	0.57	0.57

Training in a mixed breed dataset. Validating in separate breeds from next generation

\*SNPs in window with <0.01% variance removed  
For fat, all SNP never accepted also rejected

## Jersey Results

Correlations	50k		700k		3-400k
	BayesC.95	BayesB.975	BayesC.999	BayesB.999	BayesB.998
Milk	0.66	0.66	0.66	0.65	0.64
Fat	0.62	0.64	0.59	0.59	0.59
Protein	0.62	0.61	0.58	0.57	0.59

## HFXJ composite Results

Correlations	50k		700k		3-400k
	BayesC.95	BayesB.975	BayesC.999	BayesB.999	BayesB.998
Milk - corr	0.76	0.75	0.75	0.74	0.73
Fat - corr	0.57	0.58	0.59	0.59	0.59
Protein - corr	0.67	0.66	0.69	0.69	0.68

700k panels do not improve within-breed prediction  
Composites tend to have higher predictive ability than purebreds

## 50k Results: Milk Volume

R	50k	M	F Fries	F Jers	F HFXJ
Friesians	0.66	0.69	0.47	0.58	
Jerseys	0.65	0.45	0.56	0.47	
HFXJ	0.75	0.66	0.43	0.60	
TrainObs	3,606	5,718	1,287	3,575	

Training in males from the breed designated by the row  
Validating within breed in next generation females represented by the column

### 700k Results: Milk Volume

R_50k	M	F Fries	F Jers	F HFXJ
Friesians	0.66	0.69	0.47	0.58
Jerseys	0.65	0.45	0.56	0.47
HFXJ	0.75	0.66	0.43	0.60
TrainObs	3,606	5,718	1,287	3,575

R_700k	M	F Fries	F Jers	F HFXJ
Friesians	0.68	0.70	0.18	0.59
Jerseys	0.65	0.39	0.59	0.50
HFXJ	0.74	0.65	0.43	0.62

700k panels do not improve across-breed prediction

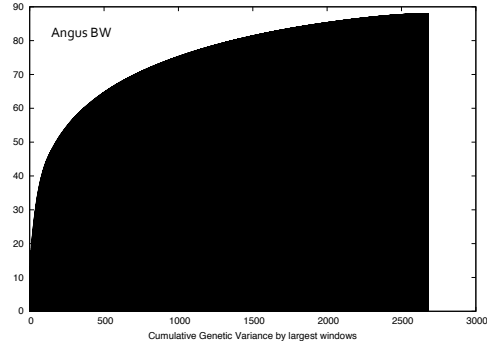
### QTL Detection

- Probably going to rely on validated QTL (ideally QTV) in order to predict across breed

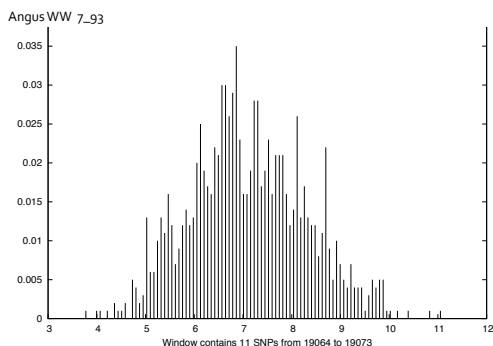
### 1 mb Window Results

Window	#SNPs	%Var	Cum%Var	p>0	p>Average	map_pos
876	11	7.10	7.10	1.000	1.000	7_93
1974	28	3.70	10.80	1.000	0.999	20_4
1480	22	1.34	12.14	0.990	0.852	13_58
2370	22	1.23	13.37	0.987	0.832	26_34
692	9	0.92	14.29	0.727	0.564	6_29
493	25	0.89	16.09	0.806	0.610	4_75
532	26	0.79	16.88	0.901	0.569	4_114
280	23	0.65	17.53	0.947	0.446	2_121
1894	17	0.61	18.14	0.835	0.467	18_55
984	25	0.60	18.74	0.873	0.406	8_88
2268	29	0.59	19.33	0.894	0.405	24_38
1975	29	0.55	19.88	0.836	0.357	20_5
1321	28	0.54	20.43	0.803	0.370	11_100
1221	27	0.48	20.91	0.819	0.320	11_0
1136	24	0.45	21.83	0.764	0.293	10_20
1977	29	0.45	22.28	0.704	0.299	20_7
1531	21	0.42	22.70	0.735	0.262	14_25
2089	19	0.42	23.12	0.586	0.303	21_47
858	19	0.42	23.53	0.713	0.264	7_75

### 1 mb Window Results



### 1 mb Window Results



### Birth Weight QTL

Chr_mb	Percentage of Genetic Variance			
	Angus	Hereford	Limousin	Simmental
7_93	7.10	5.85	0.02	0.18
6_38	0.39	7.07	5.58	13.78
20_4	3.70	7.99	0.07	1.53
6_39	0.08	1.41	0.32	2.52
6_104	0.00	0.42	1.25	0.27
14_25	0.42	0.01	0.71	3.05 (14_26)

Windows accounting for >0.3 genetic variance in >1 breed

### Body Weight QTL

ChrMb	%g var	Angus			Hereford			Simmental		
		B	W	Y	B	W	Y	B	W	Y
7_93	7.1	1.27	1.24	5.85			0.18	1.96	3.12	
6_38	0.39			7.07			13.78	7.92	19.21	
20_4	3.70	1.87	2.62	7.99			1.53	2.24	3.35	
6_39				1.41			2.52	5.64	9.26	
6_104				0.42			0.27			
14_25	0.42						3.05	0.25	0.52	
20_63					12.9					
4_24					3.92	0.90				
8_77					2.98	1.00				
11_49					1.08					
2_6										
10_79		0.66	0.44							
11_54		0.62	0.54							

Note more similarity Angus & Hereford for B than for W weight

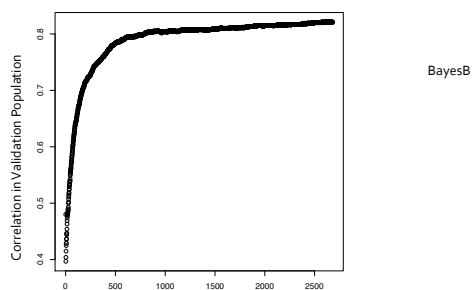
### Precision of 1 mb windows

- Simulation using 1,000 bovine 50k genotypes:
  - Significant QTL are almost always real
  - QTV may be 1 or even 2 Mb up or downstream
- Real data with >2,000 animals with 50k
  - Among 4 windows for fatty acids, 3 contain genes known to be involved in FA metabolism FAS, SCD

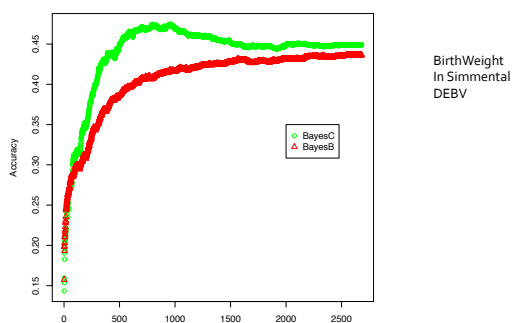
### QTL mapping: Angus Heifer Pregnancy Rate

Window	%Var	P>0
24_25	1.38	0.704
24_26	0.76	0.717
24_28	0.4	0.882
24_27	0.27	0.709
2_135	0.21	0.950
2_134	0.15	0.882
25_31	0.14	0.897
10_95	0.14	0.840
29_25	0.13	0.877
29_29	0.13	0.850

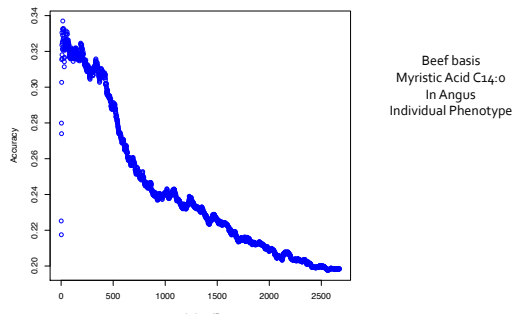
### Genomic Architecture



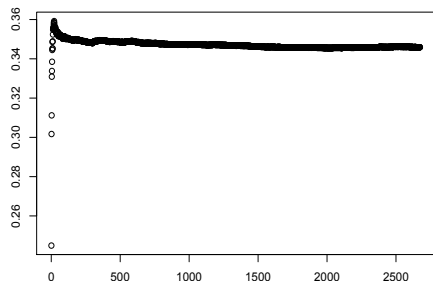
### Genomic Architecture



### Genomic Architecture



### New Method: BayesN (nested)



### Need SNP in higher LD with QTL

- Higher density genotyping
- Using haplotypes rather than genotypes
- Imputation of positional candidates
  - Published variants or QTV from sequencing

### QTL mapping results: Birth weight

Method	Rank	%Var	Marker*	Effect	EffectVar	Model Freq	Window Freq	Gene Freq	GenVar
Single	2	3.70	42843	1.09E+00	4.86E-01	0.9999	1	0.448	5.91E-01
Haplotype (wo org)	2	5.91	428471	9.56E-04	3.20E-04	0.0111	0.9727	0.059	1.02E-07
			428472	-7.14E-04	2.87E-04	0.0097	0.9725	0.201	1.64E-07
			428473	-1.74E-04	2.97E-04	0.0092	0.9726	0.014	8.43E-10
			428474	-2.97E-04	2.22E-04	0.0083	0.9728	0.114	1.79E-08
			428475	3.04E-04	2.81E-04	0.0096	0.9727	0.007	1.26E-09
			428476	-3.74E-04	3.43E-04	0.0101	0.9982	0.021	5.84E-09
			428477	-1.41E-04	2.12E-04	0.0081	0.9976	0.206	6.53E-09
		428478	1.48E+00	7.88E-01	0.9689	0.9722	0.377	1.02E+00	
Haplotype (wi org)	2	5.90	428470	2.08E-04	2.10E-04	0.0092	0.101	0.611	2.06E-08
			428471	9.88E-04	3.23E-04	0.0116	0.1025	0.059	1.09E-07
			428472	-3.72E-04	2.31E-04	0.0101	0.1009	0.201	4.46E-08
			428473	-2.91E-04	2.07E-04	0.0098	0.9861	0.014	2.37E-09
			428474	-2.96E-04	2.15E-04	0.0094	0.9861	0.114	1.78E-08
			428475	1.63E-04	2.57E-04	0.01	0.986	0.007	3.62E-10
			428476	-3.13E-04	2.39E-04	0.0102	0.9862	0.021	4.11E-09
		428477	-2.12E-04	1.82E-04	0.0084	0.9862	0.206	1.47E-08	
		428478	1.50E+00	7.93E-01	0.9847	0.9863	0.377	1.06E+00	

Chr_mb	Angus	Hereford	Limousin	Simmental
20_4	3.70	7.99	0.07	1.53

### QTL mapping results: Birth weight

Method	Rank	%Var	Marker*	Effect	EffectVar	Model Freq	Window Freq	Gene Freq	GenVar
Single	154	0.08	15154	4.30E-02	1.53E-02	0.1093	0.1897	0.349	8.40E-04
Haplotype (wo org)	3	3.39	151581	-7.36E-04	3.15E-04	0.0106	1	0.04	4.12E-08
			151582	3.23E+00	3.58E+00	1	1	0.029	5.94E-01
			151583	-2.87E-04	2.48E-04	0.0095	1	0.114	1.67E-08
			151584	1.15E-03	3.19E-04	0.0114	1	0.618	6.29E-07
			151585	-8.24E-05	2.15E-04	0.0081	1	0.143	1.66E-09
			151586	-5.95E-04	3.11E-04	0.0102	1	0.051	3.66E-08
			151587	-1.61E-04	3.89E-04	0.0108	1	0.004	2.11E-10
Haplotype (wi org)	3	3.20	151580	2.69E-04	2.22E-04	0.0096	0.9983	0.931	9.29E-09
			151581	-6.12E-04	2.46E-04	0.0098	0.9983	0.04	2.85E-08
			151582	3.23E+00	3.57E+00	0.9982	0.9983	0.029	5.94E-01
			151583	-3.44E-04	2.32E-04	0.0086	0.9983	0.114	2.40E-08
			151584	1.05E-03	2.91E-04	0.0118	0.9983	0.618	5.20E-07
			151585	-1.74E-04	1.83E-04	0.0084	0.9984	0.143	7.37E-09
			151586	-5.52E-04	2.73E-04	0.0098	0.9984	0.051	2.98E-08
		151587	-1.27E-04	2.42E-04	0.0099	0.9983	0.004	1.31E-10	

Chr_mb	Angus	Hereford	Limousin	Simmental
6_38/6_39	0.39/0.08	7.07	5.58	13.78

### Angus BWT QTL using haplotypes

Rank	BayesB		BayesN	
	Window	%Var	Window	%Var
1	7_93	6.18	7_93	7.39
2	20_4	5.91	20_4	6.23
3	6_39	3.39	6_39	4.51
4	26_34	1.40	26_34	1.84
5	10_20	0.57	10_20	0.92

Chr_mb	Angus
7_93	7.10
6_38	0.39
20_4	3.70
6_39	0.08
6_104	0.00

Compare BayesB on 50k SNP

### Future

- To get improved predictive ability, we need the volume of data to keep growing!
- More animals with SNP genotypes would be better than more SNP per animal
- Need sequencing and imputation to get causal mutations in regions of interest



Trait	Breed		
	Limousin	Simmental	
Birth weight	0.91	1.08	
Calving ease direct	0.81	0.76	
Calving ease maternal	0.81	0.55	
Carcass weight	0.90	1.11	
Docility	0.66		50k predictions alone have similar predictive ability to parent average
Fat thickness		0.53	
Marbling	1.05	1.15	But they are independent sources and can be blended
Rib eye muscle area	1.02	1.10	
Scrotal circumference	0.82		
Shear force		1.42	
Stayability	0.67	0.92	
Weaning weight direct	0.94	0.89	
Weaning weight maternal	0.75	0.57	
Yield grade	1.09	1.24	Saatchi et al (submitted GSE)
Yearling weight	1.30	0.77	

## Implementation

- ASA has now released genomic enhanced EPD using 50k genotypes
  - Incorporated into their national evaluation
- AHA will be releasing genomic enhanced EPD using 50k genotypes in May
- Some other breeds are actively growing their training populations with plans to implement

## Other NBCEC Projects

- California Commercial Ranch Project
  - Friday Genetic Prediction Committee (Dr Van E)
- Feed Efficiency Project
  - About to analyze first results from 5,000 animals
  - Includes what was the Weight Trait Project
- Several collaborators applied for reproduction

## Acknowledgements

- Dr Mahdi Saatchi, NBCEC postdoc who did the analyses
- Drs Rohan Fernando and Kadir Kizilkaya who helped with development of methodology
- Dr Jerry Taylor, Univ of Missouri, for collaboration over most of the Angus and Limousin samples, and analyses
- Drs Matt Spangler and Steve Kachman, Univ Nebraska-Lincoln, for Weight Trait Project Samples and analyses
- All the Breed Associations for providing samples and data for analysis
- Dr Richard Spelman for access to dairy samples