An International Perspective on Genomics

Dorian Garrick
dorian@iastate.edu
Sire Progeny Means

Average profit overall 492
Independent Culling Levels

Average profit selected 658

Average profit overall 492
Index Selection

Average profit index selected 677

Average profit overall 492
Suppose we generate 100 progeny on 1 bull
Performance of the Progeny

Offspring of one sire exhibit more than \( \frac{3}{4} \) diversity of the entire population.

Sire

Progeny

+30 kg

+15 kg

-10 kg

+5 kg

+10 kg

+10 kg
We learn about parents from progeny

Sire EBV +16-18 kg

(EBV is “shrunk”)

<2x progeny difference

Progeny +10 kg

+30 kg

+15 kg

-10 kg

+5 kg

+10 kg
Suppose we generate new progeny

Sire

Sire EBV +16-18 kg

Progeny

Expect them to be 8-9 kg heavier than those from an average sire

Some will be more others will be less but we can’t tell which are better without “buying” more information
Chromosomes are a sequence of base pairs

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/exons
- Yellow represents the strand inherited from the sire
- Orange represents the strand inherited from the dam
Errors in duplication
- Most are repaired
- Some will be transmitted
- Some of those may influence performance
  - Some will be beneficial, others harmful

Inspection of whole genome sequence
- Demonstrate historical errors
- And occasional new (de novo) mutations

A common error is the substitution of one base pair for another
Single Nucleotide Polymorphism (SNP)
Leptin Receptor

Prokop et al, Peptides, 2012
Joining the two

Prokop et al, Peptides, 2012
Leptin and its Receptor Across Species

Prokop et al, Peptides, 2012
Breeding Merit is sum of average gene effects

Blue base pairs represent genes/exons

Sum = +2
Sum = +8
EBV = 10
Consider 3 Bulls

Below-average bulls will have some above-average alleles and vice versa!
At any 1 locus there are 3 genotypes -4Qq+4

Contribution of this QTL

QQ +8

qq -8
Regress BV on QTL genotype

$QTL = \text{Quantitative Trait Locus}$

Variation due to other genes

Slope = average effect of allele
Illumina Bovine 770k, 50k (v2), 3k

700k (HD)  
50k (Several versions)  
3k (LD)
SNP Genotyping the Bulls

1 of 50,000 loci = 50k

EBV = 10

EBV = -6

EBV = 2
Linkage Disequilibrium (LD)

LD occurs when genotypes at one locus are predictive of genotypes at another.
Practice – EBV on SNP

Use SNP genotypes at locus 1 (in high LD) as surrogates for QTL

True Breeding Value

A_1A_1  A_1B_1  B_1B_1
Use SNP genotypes at locus 2 (in low LD) as surrogates for QTL

In practice fitting all SNP simultaneously
Meuwissen, Hayes and Goddard (2001)
Decreased Risk

<table>
<thead>
<tr>
<th>Name</th>
<th>Confidence</th>
<th>Your Risk</th>
<th>Avg. Risk</th>
<th>Compared to Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alzheimer's Disease</td>
<td>⭐⭐⭐⭐</td>
<td>4.9%</td>
<td>7.2%</td>
<td>0.69x</td>
</tr>
</tbody>
</table>

**Gene or region: APOE**

<table>
<thead>
<tr>
<th></th>
<th>SNPs used</th>
<th>Genotype</th>
<th>Allele</th>
<th>Adjusted Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dorian Garrick</td>
<td>rs7412, rs429358</td>
<td>CC, TT</td>
<td>ε3/ε3</td>
<td>European: 0.67</td>
</tr>
</tbody>
</table>

Only significant, validated GWAS findings used in prediction
• Coronary Heart Disease

Each bar represents a different risk QTL allele (mouseover shows the allele and links to the research publications)

QTL=Quantitative Trait Locus

Only significant, validated GWAS findings used in prediction
Plant & Animal Perspective

• Typically more SNP loci than subjects
• Landmark concepts were suggested by Meuwissen, Hayes & Goddard (2001)
  – Could simply fit all the SNP together (regardless of “significance”) by treating as random effects
    • They referred to these methods as “BLUP” or “BayesA”
  – Or use a variable selection model to fit as random effects some subset of the most informative SNP
    • They proposed a method called “BayesB”
Theoretical Basis for Accuracy

Reliable prediction requires large training populations of genotyped and phenotyped individuals.

Predictive Ability = Accuracy (r) = correlation true & predicted merit

Heritability = 0.8

N_e = 100
like Holsteins & Jerseys

1,000 training animals
r = 0.43 20% genetic variance

3,000 training animals
r = 0.6 36% genetic variance

Goddard & Hayes (Nature Reviews Genetics, 2009)
Accuracy of Genomic Prediction

Validation in Offspring

Correlation(g, g-hat)
Early Selection Layers

Conventional pedigree relationships

Wolc et al 2010 9WCGALP
Accuracy of Genomic Prediction

Validation in Offspring

Correlation \( (g, \hat{g}) \)
Early Selection Layers

Superiority of prediction using genomic relationships

Wolc et al 9WCGALP
Accuracy of Genomic Prediction

Validation in Offspring

Correlation(g, g-hat)
Early Selection Layers

Extent genomic prediction captures Mendelian Sampling

Accuracy of Genomic Prediction

Validation in Offspring

Correlation(g, g-hat)
Early Selection Layers

Extent genomic prediction captures Mendelian Sampling
Impact on Accuracy--%GV=10%

Genetic correlation=0.3

Blending will not improve the accuracy of a bull that already has a reliable EBV

**Impact on Accuracy**

- **%GV=10%**
  - Genetic correlation=0.3

**Graph:**
- **Pedigree only**
- **Pedigree and genomic**

**Legend:**
- Blue line represents **Pedigree and genomic**
- Black line represents **Pedigree only**

**Blending will not improve the accuracy of a bull that already has a reliable EBV**
Impact on Accuracy--%GV=40%

Genetic correlation=0.64

Blended EBVs are equally likely to be better or worse than the preblended EBVs
Layer Hens – Dekkers scheme

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Traditional</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
</tr>
<tr>
<td># candidates with phenotype</td>
<td>1000</td>
</tr>
<tr>
<td># selected</td>
<td>60</td>
</tr>
<tr>
<td>Generation interval (months)</td>
<td>13</td>
</tr>
<tr>
<td>Information</td>
<td>Own Phenotype</td>
</tr>
</tbody>
</table>
Layer Hens – Dekkers scheme

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Traditional</th>
<th>GS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>#candidates with phenotype</td>
<td>1000</td>
<td>3000</td>
</tr>
<tr>
<td># selected</td>
<td>60</td>
<td>360</td>
</tr>
<tr>
<td>Generation interval (months)</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Information</td>
<td>Own Phenotype</td>
<td>Genotype+Phenotype</td>
</tr>
</tbody>
</table>

Halve the generation interval and reduce costs by (less phenotyping) to get same gain & same inbreeding
After 3 generations of conventional or 6 gens of genomic selection
Genomic selection was as good, if not better in terms of realized response
## Predictions in Beef Cattle Breeds

<table>
<thead>
<tr>
<th>Trait</th>
<th>RedAngus (6,412)</th>
<th>Angus (3,500)</th>
<th>Hereford (2,980)</th>
<th>Simmental (2,800)</th>
<th>Limousin (2,400)</th>
<th>Gelbvieh (1,321)+</th>
</tr>
</thead>
<tbody>
<tr>
<td>BirthWt</td>
<td>0.75</td>
<td>0.64</td>
<td>0.68</td>
<td>0.65</td>
<td>0.58</td>
<td>0.62</td>
</tr>
<tr>
<td>WeanWt</td>
<td>0.67</td>
<td>0.67</td>
<td>0.52</td>
<td>0.52</td>
<td>0.58</td>
<td>0.52</td>
</tr>
<tr>
<td>YlgWt</td>
<td>0.69</td>
<td>0.75</td>
<td>0.60</td>
<td>0.45</td>
<td>0.76</td>
<td>0.53</td>
</tr>
<tr>
<td>Milk</td>
<td>0.51</td>
<td>0.51</td>
<td>0.37</td>
<td>0.34</td>
<td>0.46</td>
<td>0.39</td>
</tr>
<tr>
<td>Fat</td>
<td>0.90</td>
<td>0.70</td>
<td>0.48</td>
<td>0.29</td>
<td></td>
<td>0.75</td>
</tr>
<tr>
<td>REA</td>
<td>0.75</td>
<td>0.75</td>
<td>0.49</td>
<td>0.59</td>
<td>0.63</td>
<td>0.61</td>
</tr>
<tr>
<td>Marbling</td>
<td>0.85</td>
<td>0.80</td>
<td>0.43</td>
<td>0.63</td>
<td>0.65</td>
<td>0.87</td>
</tr>
<tr>
<td>CED</td>
<td>0.60</td>
<td>0.69</td>
<td>0.68</td>
<td>0.45</td>
<td>0.52</td>
<td>0.47</td>
</tr>
<tr>
<td>CEM</td>
<td>0.32</td>
<td>0.73</td>
<td>0.51</td>
<td>0.32</td>
<td>0.51</td>
<td>0.62</td>
</tr>
<tr>
<td>SC</td>
<td></td>
<td>0.71</td>
<td>0.43</td>
<td></td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td><strong>0.67</strong></td>
<td><strong>0.69</strong></td>
<td><strong>0.52</strong></td>
<td><strong>0.47</strong></td>
<td><strong>0.57</strong></td>
<td><strong>0.56</strong></td>
</tr>
</tbody>
</table>

Genetic correlations from k-fold validation  Saatchi et al (GSE, 2011; 2012; J Anim Sc, 2013)
SNP Alleles are inherited in blocks

paternal

maternal

Chromosome pair
SNP Alleles are inherited in blocks

Occasionally (30%) one or other chromosome is passed on intact

paternal

maternal

Chromosome pair

e.g.
SNP Alleles are inherited in blocks

Typically (40%) one crossover produces a new recombinant gamete

Recombination can occur anywhere but there are “hot” spots and “cold” spots
SNP Alleles are inherited in blocks

- Paternal
- Maternal

Sometimes there may be two (20%) or more (10%) crossovers

Never close together
SNP Alleles are inherited in blocks

- Paternal
- Maternal

Interestingly the number of crossovers varies between sires and is heritable

On average, 1 crossover per chromosome per generation

Possible offspring chromosome inherited from one parent
SNP Alleles are inherited in blocks

Consider a small window of say 1% chromosome (1 Mb)
SNP Alleles are inherited in blocks

Offspring mostly (99%) segregate blue or red (about 1% are admixed)

“Blue” haplotype (eg sires paternal chromosome)

“Red” haplotype (eg sires maternal chromosome)
SNP Alleles are inherited in blocks

- Offspring mostly (99%) segregate blue or red (about 1% are admixed)

- “Blue” haplotype (e.g., sires paternal chromosome)
- “Red” haplotype (e.g., sires maternal chromosome)
Regress BV on haplotype dosage

Use multiple regression to simultaneously estimate dosage of all haplotypes (colours) in every 1 Mb window
Panel Comparison

Black = Illumina 50K
Panel Comparison

Black = Illumina 50K
Blue = Illumina HD (700K)
Panel Comparison

GeneSeek Genomic Profilers
- Low Density
  - Super GGP (20k) $45
- High Density
  - GGP HD (77k) $75

Orange = GGP-Super LD 19k
Green = GGP-HD (taurus) 70k
Black = Illumina 50k

GGP also include custom SNP

50k and GGP-HD share 28K
50k and GGP-Super LD share 8k

Need to genotype more individuals/yr
Need cheaper genotyping

There are multiple minor variants of all these panels!

Also a separate GGP-HD-I (Indicus)

No longer using Illumina 50k
## Lower Density Panels

<table>
<thead>
<tr>
<th>Trait</th>
<th>Actual</th>
<th>Imputed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth Weight</td>
<td>0.67</td>
<td>0.65</td>
</tr>
<tr>
<td>Calving Ease Direct</td>
<td>0.68</td>
<td>0.67</td>
</tr>
<tr>
<td>Calving Ease Maternal</td>
<td>0.51</td>
<td>0.50</td>
</tr>
<tr>
<td>Fat Thickness</td>
<td>0.47</td>
<td>0.46</td>
</tr>
<tr>
<td>Marbling</td>
<td>0.42</td>
<td>0.42</td>
</tr>
<tr>
<td>Mature cow weight</td>
<td>0.64</td>
<td>0.62</td>
</tr>
<tr>
<td>Rib Eye Muscle Area</td>
<td>0.49</td>
<td>0.46</td>
</tr>
<tr>
<td>Scrotal Circumference</td>
<td>0.43</td>
<td>0.42</td>
</tr>
<tr>
<td>Weaning Weight Direct</td>
<td>0.53</td>
<td>0.50</td>
</tr>
<tr>
<td>Weaning Weight Maternal</td>
<td>0.37</td>
<td>0.35</td>
</tr>
<tr>
<td>Yearling Weight</td>
<td>0.61</td>
<td>0.59</td>
</tr>
<tr>
<td>Mean</td>
<td>0.53</td>
<td>0.51</td>
</tr>
</tbody>
</table>

Actual $= 50k$  
Imputed $= 10k$  
(from GGP-LD)
Genomic Prediction Pipeline

Breeders

GeneSeek

Iowa State NBCEC

Prediction Equation

GeneSeek running the Beagle pipeline GGP to 50k then applying prediction equation

Hair/DNA

MBV and genotypes

AHA

ABRI Breedplan

Blend MBV & EPD
# Early 2014 Genotype Counts

<table>
<thead>
<tr>
<th>Breed</th>
<th>9k</th>
<th>GGP-LD</th>
<th>50k</th>
<th>GGP-HD</th>
<th>BOS-1</th>
<th>700k HD</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAN</td>
<td>911</td>
<td>13,409</td>
<td>787</td>
<td>947</td>
<td>16,054</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRG</td>
<td>1,128</td>
<td>173</td>
<td></td>
<td>243</td>
<td>1,544</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSH</td>
<td>325</td>
<td></td>
<td></td>
<td>136</td>
<td>461</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHA</td>
<td>1,617</td>
<td></td>
<td></td>
<td>525</td>
<td>2,142</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GVH</td>
<td>186</td>
<td>209</td>
<td>1,643</td>
<td>414</td>
<td>3,253</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HER</td>
<td>7,064</td>
<td>1,887</td>
<td>471</td>
<td>850</td>
<td>10,272</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIM</td>
<td>429</td>
<td>3,420</td>
<td>8</td>
<td>461</td>
<td>4,993</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEL</td>
<td></td>
<td></td>
<td></td>
<td>2,571</td>
<td>2,571</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAN</td>
<td>1,931</td>
<td>1,183</td>
<td>226</td>
<td></td>
<td>3,340</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RDP</td>
<td>1,394</td>
<td></td>
<td></td>
<td></td>
<td>1,394</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIM</td>
<td>5,223</td>
<td>7,026</td>
<td>6,501</td>
<td>1,347</td>
<td>2,601</td>
<td>674</td>
<td>22,372</td>
</tr>
<tr>
<td>TOTALS</td>
<td>5,409</td>
<td>8,575</td>
<td>38,432</td>
<td>5,756</td>
<td>3,173</td>
<td>7,051</td>
<td>68,396</td>
</tr>
</tbody>
</table>
## Major Regions for Birth Weight

Some of these same regions have big effects on one or more of weaning weight, yearling weight, marbling, ribeye area, calving ease.

<table>
<thead>
<tr>
<th>Chr_mb</th>
<th>Angus</th>
<th>Hereford</th>
<th>Shorthorn</th>
<th>Limousin</th>
<th>Simmental</th>
<th>Gelbvieh</th>
</tr>
</thead>
<tbody>
<tr>
<td>7_93</td>
<td>7.10</td>
<td>5.85</td>
<td>0.01</td>
<td>0.02</td>
<td>0.18</td>
<td>0.02</td>
</tr>
<tr>
<td>6_38-39</td>
<td>0.47</td>
<td>8.48</td>
<td>11.63</td>
<td>5.90</td>
<td>16.3</td>
<td>4.75</td>
</tr>
<tr>
<td>20_4</td>
<td>3.70</td>
<td>7.99</td>
<td>1.19</td>
<td>0.07</td>
<td>1.53</td>
<td>0.03</td>
</tr>
<tr>
<td>14_24-26</td>
<td>0.42</td>
<td>0.01</td>
<td>0.01</td>
<td>0.71</td>
<td>3.05</td>
<td>8.14</td>
</tr>
</tbody>
</table>

**Genetic Variance %**

Adding Haplotypes
- 3.20%
- 5.90%

Imputed 700k
- Collective 3 QTL
- 30% GV
## PLAG1 on Chromosome 14 @ 25 Mb

<table>
<thead>
<tr>
<th>Effect of 1 copy</th>
<th>Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birthweight</td>
<td>5 lb (10 lb for QQ – qq)</td>
</tr>
<tr>
<td>Weaning weight</td>
<td>10 lb</td>
</tr>
<tr>
<td>Feedlot on weight</td>
<td>16 lb</td>
</tr>
<tr>
<td>Feedlot off weight</td>
<td>24 lb</td>
</tr>
<tr>
<td>Carcass weight</td>
<td>14 lb</td>
</tr>
</tbody>
</table>
PLAG1 on Chromosome 14 @25 Mb

<table>
<thead>
<tr>
<th>Effect of 1 copy</th>
<th>Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birthweight</td>
<td>5 lb (10 lb for QQ – qq)</td>
</tr>
<tr>
<td>Weaning weight</td>
<td>10 lb</td>
</tr>
<tr>
<td>Feedlot on weight</td>
<td>16 lb</td>
</tr>
<tr>
<td>Feedlot off weight</td>
<td>24 lb</td>
</tr>
<tr>
<td>Carcass weight</td>
<td>14 lb</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect of 1 copy</th>
<th>Reproduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age CL (1\textsuperscript{st} Corpus Luteum)</td>
<td>38 days (76 days QQ – qq)</td>
</tr>
<tr>
<td>PPAI (post partum anoestrus)</td>
<td>15 days</td>
</tr>
<tr>
<td>Presence CL before weaning</td>
<td>-5%</td>
</tr>
<tr>
<td>Weight at CL</td>
<td>36 lb</td>
</tr>
<tr>
<td>Age at 26 cm Scrotal Circumf</td>
<td>19 days</td>
</tr>
</tbody>
</table>
Sequence

• Now sequencing individual sires
  – Identify loss-of-function alleles to compare to underrepresented haplotype alleles
  – Identify mutations that are perfectly concordant with haplotype allelic effect
    • More powerful across breed
Genomic Prediction

• Exploits advances in quantitative genetics, statistical genetics, computing, molecular biology, and bioinformatics

• Is the basis for some aspects of personalized medicine

• Will revolutionize plant and animal improvement programmes, but to different extents in different industries
Genomic Prediction

• Its application in humans, plants and animals is still an immature but maturing technology
• Its development will greatly benefit from collaborative activities with other researchers across the entire range of disciplines with interests in genomics
Acknowledgments

- Dr. Rohan Fernando
- Dr. Jack Dekkers
- Dr. Max Rothschild
- Dr. Ania Wolc
- Dr. Bruce Golden
- Dr. Mahdi Saatchi
- Dr. Kadir Kizilkaya
- Dr. David Habier
- Dr. Hailin Su
- Dr. Jungjae Lee
- Dr. Jingjing Yan
- Ziging Weng

- GeneSeek
- Beef Breed Associations
  - American Angus Assoc
  - American Hereford Assoc
  - American Simmental Assoc
  - American Gelbvieh Assoc
  - Red Angus Association
- Aviagen (Broilers)
- HyLine (Layers)
- Livestock Improvement Corp