Translational genomics: Progress in the adaptation of information derived from genome technologies for animal improvement

Alison Van Eenennaam
Animal Genomics and Biotechnology
Cooperative Extension Specialist
Department of Animal Science
University of California, Davis
alvaneenennaam@ucdavis.edu
(530) 752-7942

animalscience.ucdavis.edu/animalbiotech
“Translational genomics” is defined as the adaptation of information derived from genome technologies for animal improvement.

“We believe DNA marker profiles will become widely used in livestock in the near future as the cost decreases and the benefits increase. In fact, a major research objective may be to make best use of this DNA data in commercial animal production.”

“1954 version of what 'home computers' might look like in 50 years time (i.e. 2004)”
Wrong Expert Predictions

I think there's a world market for about five computers.
Thomas J. Watson, chairman of the board of IBM. 1943

There is no reason anyone would want a computer in their home.
Ken Olson, president of Digital Equipment Corp. 1977

The cost for a “large” genome scan (defined as 18 chromosomes* 7 markers (i.e. 126 markers!) * $4/marker) = $504
“what escaped their vision was that science might come up with new and different ways of commercializing and using new technologies.”
Overview

- What is working well
  - Parentage
  - Identification of recessive/single trait defects
  - Dairy genomic selection

- What is not working so well
  - Beef genomic selection

- What does the future hold?
  - Roadblocks to translational genomics
  - Some solutions and future prospects
Benefits of DNA-based parentage identification

- Correct pedigree errors thereby improving the rate of genetic gain
- Enables the use of multi-sire breeding pasture
  - Higher fertility
  - Elimination of sire failure
  - Tighter calving season
- Reduces the need for different breeding pastures
  - Allows for better pasture management
  - Less sorting and working of animals into different groups
- Reduces the need to disturb newborn animals
  - Labor savings so can focus on concentrate on offspring survival
  - Worker safety improvement
  - Better bonding of offspring with dam
- Enables the development of commercial-ranch genetic evaluations

Blood collected on FTA cards from 27 herd sires and 624 calves derived from a multiple-sire pasture

Analyzed using a 28 SNP panel in 2005

<table>
<thead>
<tr>
<th></th>
<th>28 SNP Panel – 27 sires 2005 (PE=95.5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>One sire assigned</td>
<td>175</td>
</tr>
<tr>
<td>More than one sire</td>
<td>420</td>
</tr>
<tr>
<td>All excluded</td>
<td>29</td>
</tr>
<tr>
<td>TOTAL</td>
<td>624</td>
</tr>
</tbody>
</table>
Translational considerations for migrating to SNPs for parentage

It is likely that SNP markers will replace alternatives (i.e. microsatellites) over the next 5 or so years

- How do you switch over from microsatellites to SNPS when a lot of historical information is stored as microsatellites?
- Which SNP genotyping platform should be used and how many and which SNP markers should be included in the panel?
- What should be the number of compared loci cutoff in the case of incomplete genotyping?
- How many exclusions (as a function of number of compared loci) should be allowed to account for genotyping errors – platform dependent?
- Which sample type works best for producers to collect and genotyping entities to run?
DNA-based parentage identification has shown significant uptake by deer, cattle, and sheep breeders in New Zealand.

Industry adoption looking at New Zealand as an example

- >200,000 dairy parentage tests out of a herd of about 4.6M cows. ~10% of the commercial tier. Most of these tests are SNP based.
- 20% of the ram, and 30% of the deer breeding industry (majority of stag breeders); mostly microsatellite-based parentage tests
- Emerging area in NZ is parentage testing in aquaculture species
- One of the reasons for the widespread adoption of this technology is the development of an integrated ID and collection system
- This is especially important for lower value animals such as sheep
- If DNA samples are already being collected for parentage verification or as part of a national animal identification scheme, then other DNA technologies can be introduced cost-effectively.

A key issue in commercial situations is ease of DNA sampling, tracking and quality of resultant DNA.
Reconcile of hair and meat samples (based on 427 records)

427 meat samples (3K genotyped) and hair samples (99 SNP parentage genotyped)

- 149 had no exclusions with hair (35%)
- 185 had 1 exclusion (43%)
- 31 had 2 exclusions (7%)
- 10 had 3 exclusions (2%)
- 31 had 4-10 exclusions (7%)
- 21 > 10 exclusions (5%)
Genetic Abnormalities
Images from an article by David S. Buchanan, Department of Animal Sciences, North Dakota State University
Compare dwarfism response in the 50s to the response to curly calf (AM)

An early '50s advertisement that superimposed a measuring stick in the picture of this bull who was nick-named "Short Snorter."

Based upon his height and age, he was less than a frame score 1.

Image from https://www.msu.edu/~ritchieh/historical/shortsnorter.jpg
Curly calf – Arthrogryposis multiplex

- From a scientific standpoint, AM is the complete deletion of a segment of DNA that encompasses two different genes.
- One of these genes is expressed at a crucial time in the development of nerve and muscle tissue. The mutation results in no protein being produced from this gene and therefore it is unable to carry out its normal function so homozygotes show phenotype.
From September 8 – November 3, 2008 identified genetic problem, developed test, and released carrier status of 736 bulls!

- In the 10 months following the release of the test, the AAA posted the results of tests for AM on about 90,000 cattle.

- These AM test costs less than $30.

- Of these, almost 5,000 bulls and more than 13,000 heifers have tested as carriers of AM. That leaves more than 22,000 bulls and more than 50,000 heifers which tested as free of AM.

Early extension education about dwarfism explaining carriers and inheritance

Image from Special CollectionsUniversity Libraries, Virginia Tech:
http://spec.lib.vt.edu/imagebase/agextension/boxseven/screen/AGR3618.jpg
If you breed a curly calf carrier cow (AMC) to an curly calf free bull (AMF), what is the chance that the offspring will be stillborn as a result of being curly calf?

1. 0
2. ¼ (25%)
3. ½ (50%)
4. 2/3 (66%)
5. ¾ (75%)
6. 1 (100%)

Results from a typical producer meeting
Moving onto Genomic Selection

Training 1: Old Progeny Tested Bulls

Training 2: Old & New Progeny Tested Bulls

Degree of genetic relationship between populations (ideally similar)

Validation: New Progeny Tested Bulls

Application: New Sire Candidates
Breeding value prediction in Dairy Sires

Young sire Parent Average

Young sire Progeny Test

Young sire Genomic Selection

Birth

5 years; >>>> cost

Birth; <<<< cost

Accuracy 0.20

Accuracy 0.80

Accuracy 0.65
Genomic selection can help breeders identify animals with superior breeding values at a young age.

\[ \Delta G = \text{intensity of selection} \times \text{accuracy of selection} \times \frac{\text{genetic variation in the population}}{\text{generation interval}} \]
Dairy industry suited to WGS

- High use of AI
- Clear selection goal
- One breed used extensively
- Large number of high accuracy A.I. sires for training
- Extensive, uniform collection of data on traits
- Central evaluation (AIPL) receiving genotypes
- Obvious way to increase rate of genetic gain
- AI companies funding the genotyping because they get a clear cost savings in terms of young sire program
Translational Questions for other animal industries

? How many phenotypic records are required in the initial experiment estimating the effect of chromosome segments?

? How many markers are needed—50K, 800K, whole genome?

? How does the relationship between the training population and the selection candidate affect accuracy?

? How often do chromosome segment effects need to be re-estimated?

? Do predictions work across breeds?

? What is the value generated by the increased accuracy?

? Does this technology change optimal breeding program design?
Accuracy of the prediction equation proportional to:

\[
\frac{Th^2}{NeL}
\]

T: total number of animals in the training population

\(h^2\): heritability of the trait

L: length of chromosomes (in Morgans)

Ne: effective population size

Also influenced by trait architecture, number of markers, availability of economically-relevant phenotypes, and relationship between animals in the training and target population.
Effect of number of animals on accuracy of prediction equation (for a $N_e$ of 100)

## Effective population size estimates for cattle

<table>
<thead>
<tr>
<th>Breed</th>
<th>$N_e$</th>
<th>Breed</th>
<th>$N_e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angus</td>
<td>136</td>
<td>Brown Swiss</td>
<td>61</td>
</tr>
<tr>
<td>Charolais</td>
<td>110</td>
<td>Guernsey</td>
<td>76</td>
</tr>
<tr>
<td>Hereford</td>
<td>97</td>
<td>Holstein</td>
<td>99</td>
</tr>
<tr>
<td>Limousin</td>
<td>174</td>
<td>Jersey</td>
<td>73</td>
</tr>
<tr>
<td>Red Angus</td>
<td>85</td>
<td>Norwegian Red</td>
<td>106</td>
</tr>
<tr>
<td>Brahman</td>
<td>115</td>
<td>Gir</td>
<td>133</td>
</tr>
<tr>
<td>Nelore</td>
<td>86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beef Master</td>
<td>106</td>
<td>Merino (sheep)</td>
<td>~ Big (&gt; 100)</td>
</tr>
<tr>
<td>Santa Gertrudis</td>
<td>107</td>
<td>Ben Hayes</td>
<td>(pers. comm.)</td>
</tr>
</tbody>
</table>

---

Effect of population size and heritability on the number of animals required in the training population (for an accuracy of 0.7)

There is also an effect of trait architecture

The accuracy of predicting genetic values is higher for traits with a proportion of large effects (e.g. proportion black and fat percentage) than for a trait with no loci of large effect (e.g. overall type), provided the method of analysis takes advantage of the distribution of loci effects.

Maybe R.A. Fisher was onto something?

If a nearly infinitesimal model is correct as seems to be the case for most quantitative traits; then large sample sizes will be needed to achieve high accuracy.
Markers can predict family relationships between animals, independently of LD between the markers and QTL (i.e. due to family relationships or linkage)

Additive-genetic relationships between training and validation animals was found to be a good indicator of accuracy.

Genomic Selection In Beef Cattle: Training And Validation In Multibreed Populations

Kristina Weber P514: Monday morning poster session

![Bar chart showing accuracy of GEBV for different scenarios.]

<table>
<thead>
<tr>
<th>Condition</th>
<th>BWT</th>
<th>WWT</th>
<th>YWT</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000_Bull-trained, with GPE sires</td>
<td>0.474</td>
<td>0.454</td>
<td>0.370</td>
</tr>
<tr>
<td>2000_Bull-trained, without GPE sires</td>
<td>0.428</td>
<td>0.260</td>
<td>0.237</td>
</tr>
<tr>
<td>USMARC GPE-trained, with GPE sires</td>
<td>0.369</td>
<td>0.217</td>
<td>0.322</td>
</tr>
<tr>
<td>USMARC GPE-trained, without GPE sires</td>
<td>0.321</td>
<td>0.181</td>
<td>0.279</td>
</tr>
</tbody>
</table>
Reduced SNP panels: Accuracy of direct genomic value (DGV) of dairy bulls using subsets of 5,000 or less of best SNP for each trait


Traits:
- Protein %
- ASI (Australian Selection Index)
- APR (Australian Profit Rank)

Predictions based on <1,000 SNP panels were very sensitive to the selection method and tended to be low accuracy
**Reduced SNP panels**: Percentage of the highest ranked SNP that are shared between sets of traits* for subsets including 500, 1,000, 5,000 or 10,000 SNP

* Dairy traits included:
1. Protein
2. Protein %
3. Survival
4. Fat %
5. Milk
6. Overall Type
7. APR (Australian Profit Rank)
8. ASI (Australian Selection Index)
9. Fat

Few SNPs were in common between the trait-specific subsets

In general accuracy is higher when:

- Small effective population size so small number of chromosome segments to track
- Small number of QTL effecting the trait so there is a marker associated with every QTL
- High density of markers
- Trait is highly heritable
- A large number of animals and high-quality phenotypic records available for training
- There are genetic relationships (linkage) between training and selection candidates
Other relevant interesting findings with practical implications

- If markers are picking up family relationships (linkage), then the accuracy of marker-based selection will decay over generations within breed.
- Prediction equations derived in one breed do not predict accurate GEBVs when applied to other breeds.
- To find markers that are in LD with QTL across diverged breeds, such as Holstein, Jersey, and Angus, will require high density SNPs (>300,000 markers).
- Combining breeds into one large multi-breed reference population gives reasonable accuracies in purebreds.
- Few of the “best” markers for one trait are common to another.
The Beef Cattle Industry

- Little use of AI
- Relatively few high accuracy sires for training
- Multiple competing selection goals – cow/calf, feedlot, processor – little data sharing between sectors
- Few/no records on many economically-relevant traits
- Many breeds, some small with limited resources
- Crossbreeding is important
- No centralized “national” cattle evaluation
What commercial products are out there for beef cattle producers?
The Power of the IGENITY® profile for Angus

The American Angus Association® through its subsidiary, Angus Genetics Inc.® (AGI), has a vision to provide Angus breeders with the most advanced solutions to their genetic selection and management needs.

Genomic-enhanced Expected Progeny Differences (EPDs) can now be calculated for your animals using the highly predictable American Angus Association database along with IGENITY® profile results to provide a more thorough characterization of economically important traits and improved accuracy on young animals.

Using the IGENITY profile for Angus, breeders receive comprehensive genomic results for multiple, economically important traits.

1. Dry Matter Intake
2. Birth Weight
3. Mature Height
4. Mature Weight
5. Milk
6. Scrotal Circumference
7. Weaning Weight
8. Yearling Weight
9. Marbling
10. Ribeye Area
11. Fat Thickness
12. Carcass Weight
13. Tenderness
14. Percent Choice (quality grade)
15. Heifer Pregnancy
16. Maternal Calving Ease
17. Direct Calving Ease
18. Docility
19. Average Daily Gain
20. Feed Efficiency
21. Yearling Height
Genetic correlations (r) between carcass traits and IGENITY® Angus Profile molecular breeding values (384 reduced-SNP panel) in Angus cattle

<table>
<thead>
<tr>
<th>TRAIT</th>
<th>Trait heritability</th>
<th>Accuracy (r)</th>
<th>% Genetic Variation (r²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carcass weight</td>
<td>0.39</td>
<td>.54</td>
<td>29%</td>
</tr>
<tr>
<td>Backfat thickness</td>
<td>0.36</td>
<td>.50</td>
<td>25%</td>
</tr>
<tr>
<td>Ribeye area</td>
<td>0.40</td>
<td>.58</td>
<td>34%</td>
</tr>
<tr>
<td>Marbling score</td>
<td>0.37</td>
<td>.65</td>
<td>42%</td>
</tr>
</tbody>
</table>

Lead Today with 50K

1. Birth weight
2. Weaning weight
3. Weaning maternal (milk)
4. Calving ease direct
5. Calving ease maternal
6. Marbling
7. Backfat thickness
8. Ribeye area
9. Carcass weight
10. Tenderness
11. Postweaning average daily gain
12. Daily feed intake
13. Feed efficiency (net feed intake)

50K SNP chip assays
50,000 SNPs spread throughout genome
<table>
<thead>
<tr>
<th>Trait</th>
<th>$h^2$</th>
<th>Number of animals in training population&lt;sup&gt;1&lt;/sup&gt;</th>
<th>% Genetic variation ($r^2$) Predicted from LD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Daily Gain</td>
<td>0.28</td>
<td>1254</td>
<td>7%</td>
</tr>
<tr>
<td>Net Feed Intake</td>
<td>0.39</td>
<td>1254</td>
<td>10%</td>
</tr>
<tr>
<td>Dry matter intake</td>
<td>0.39</td>
<td>1254</td>
<td>10%</td>
</tr>
<tr>
<td>Tenderness</td>
<td>0.37</td>
<td>1445</td>
<td>11%</td>
</tr>
<tr>
<td>Calving Ease (Direct)</td>
<td>0.1</td>
<td>1188</td>
<td>2%</td>
</tr>
<tr>
<td>Birth weight</td>
<td>0.31</td>
<td>1169</td>
<td>7%</td>
</tr>
<tr>
<td>Weaning Weight</td>
<td>0.25</td>
<td>1192</td>
<td>5%</td>
</tr>
<tr>
<td>Calving ease (maternal)</td>
<td>0.1</td>
<td>1177</td>
<td>2%</td>
</tr>
<tr>
<td>Milking Ability</td>
<td>0.25</td>
<td>1187</td>
<td>5%</td>
</tr>
<tr>
<td>Carcass weight</td>
<td>0.39</td>
<td>1100</td>
<td>9%</td>
</tr>
<tr>
<td>Backfat thickness</td>
<td>0.36</td>
<td>1097</td>
<td>8%</td>
</tr>
<tr>
<td>Ribeye area</td>
<td>0.4</td>
<td>1114</td>
<td>10%</td>
</tr>
<tr>
<td>Marbling score</td>
<td>0.37</td>
<td>1143</td>
<td>9%</td>
</tr>
</tbody>
</table>


Approx. cost of commercial tests
(estimates only!! - derived from web-sites or personal experience - not official quotes!!)

<table>
<thead>
<tr>
<th>Test</th>
<th>Species</th>
<th>Cost ($US)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parentage</td>
<td>Cattle</td>
<td>$13-25</td>
</tr>
<tr>
<td>Genetic Defects</td>
<td>Cattle</td>
<td>$15-150</td>
</tr>
<tr>
<td>3K (just the genotypes)</td>
<td>Cattle</td>
<td>$38</td>
</tr>
<tr>
<td>50K (just the genotypes)</td>
<td>Cattle</td>
<td>$150</td>
</tr>
<tr>
<td>800K (just the genotypes)</td>
<td>Cattle</td>
<td>$340</td>
</tr>
<tr>
<td>384 Angus Profile (Igenity US/AGI)</td>
<td>Beef Cattle</td>
<td>$65</td>
</tr>
<tr>
<td>3K (Pfizer US)</td>
<td>Dairy Cattle</td>
<td>$45</td>
</tr>
<tr>
<td>50K (Pfizer US/AGI)</td>
<td>Beef Cattle</td>
<td>$139</td>
</tr>
<tr>
<td>50K (Holstein Ass.)</td>
<td>Dairy Cattle</td>
<td>$150</td>
</tr>
<tr>
<td>800K (Holstein Ass.)</td>
<td>Dairy Cattle</td>
<td>$365</td>
</tr>
<tr>
<td>50K (Pfizer NZ)</td>
<td>Sheep</td>
<td>$756 (NZ$990)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unit</th>
<th>Accuracy of DNA test used</th>
<th>GRASS INDEX</th>
<th>FEEDLOT INDEX</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Terminal</td>
<td>Maternal</td>
</tr>
<tr>
<td>Improvement in selection response</td>
<td>%</td>
<td>Intermediate</td>
<td>29</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High</td>
<td>54</td>
<td>81</td>
</tr>
<tr>
<td>Increased value derived from $\Delta G$ in</td>
<td>$/</td>
<td>Intermediate</td>
<td>45</td>
<td>69</td>
</tr>
<tr>
<td>commercial sires</td>
<td>DNA test</td>
<td>High</td>
<td>83</td>
<td>124</td>
</tr>
<tr>
<td>Increased value derived from $\Delta G$ in</td>
<td>$/</td>
<td>Intermediate</td>
<td>160</td>
<td>203</td>
</tr>
<tr>
<td>stud sires</td>
<td>DNA test</td>
<td>High</td>
<td>297</td>
<td>366</td>
</tr>
<tr>
<td>Total value per test to seedstock operator</td>
<td>$/</td>
<td>Intermediate</td>
<td>$204</td>
<td>$272</td>
</tr>
<tr>
<td></td>
<td>DNA test</td>
<td>High</td>
<td>$380</td>
<td>$490</td>
</tr>
</tbody>
</table>

Van Eenennaam PAG 1/15/2011
Industry breakdown of ΔG value derived from increased accuracy from genomic selection


Van Eenennaam PAG 1/15/2011
The beef industry needs to share data and profit between sectors to most benefit from genomic selection.

CONCLUSION: Ramifications of genomic selection

- The benefits of genomic selection are best captured in well-structured industries that are already making significant genetic progress.
- May encourage more vertical integration to collect phenotypes to enable predictions for ERTs for all sectors.
- May see genetic evaluations developed for novel traits – if large enough populations can be amassed and data shared.
- May see breeds/countries start to share data – especially with HD chips and whole genome sequencing.
- Will beef follow the pig/poultry model of vertically-integrated breeding companies owning all sectors?
“This project is supported by National Research Initiative Grant no. 2009-55205-05057 from the USDA Cooperative State Research, Education, and Extension Service Animal Genome program.”
Come to Melbourne, Australia !!!
2-5 May, 2011

Date Claimer
Applied Genomics for Sustainable Livestock Breeding
2-5 May 2011
The Sebel Albert Park
Melbourne
www.smogenomics.org
genomics-conf@jkconnections.com.au
Questions?
Marker location relative to the gene of interest in two breeds when using the (A) 50K SNP chip assay (markers spaced at \( \sim 70 \) kb intervals), or (B) the high density 700 K SNP chip assay (markers spaced at \( \sim 5 \) kb intervals)
High density panels offer the opportunity to accelerate discovery of the causal mutations underlying genetic variation – especially if combined with full sequence data on key ancestors.