Chip technology for use in genomic evaluation and new developments

André Eggen
Agrigenomics Specialist, Europe Illumina, Inc.
Mission
Innovating for the Future of Genetic Analysis

Food and feed are central to human health. Illumina is advancing the use of genetic variation-based approaches to crop and livestock markets.
Chip Technology for use in Genomic Evaluation and New Developments

- The genomic era: a reality in animal breeding
- The technology behind the change of paradigm
- Illumina’s genomic portfolio for Cattle
  - Bovine HD, BovineSNP50, Bovine3K, ...
  - Whole Genome Sequencing
- Conclusions
THE AGRIGENOMIC ERA
The AgriGenomic Era ...

Sequencing of livestock genomes has discovered 100s of thousands of Single Nucleotide Polymorphism (SNP) markers
The big promises of genomic selection

General idea

• With enough markers one can follow the transmission (segregation) of the entire genome and not only a set of specific regions of interest

• Parental relationship is no more used to explain similar performances in animal

• Similar performances are explain by the fact that animals are sharing identical chromosome fragments
Why is genomics so attractive?

- Genetic evaluation can be performed as soon as DNA is available
- Allows accurate selection in both genders early in life

Some direct consequences
- Genetic progress could be doubled
- Generation interval reduced
- Cost for proving bulls could be reduced
- Low heritability traits could be selected
- ...
Genomic Selection for Males

<table>
<thead>
<tr>
<th>Age in years</th>
<th>0</th>
<th>2</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genomic Evaluation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Milk</td>
<td>62%</td>
<td>62%</td>
<td>70%</td>
</tr>
<tr>
<td>Fert</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td><strong>Traditional Evaluation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Milk</td>
<td>30%</td>
<td>35%</td>
<td>70%</td>
</tr>
<tr>
<td>Fert</td>
<td>-</td>
<td>-</td>
<td>20%</td>
</tr>
</tbody>
</table>

As shown by E. Verrier, November 2009
# Genomic Selection for Females

As shown by E. Verrier, November 2009

<table>
<thead>
<tr>
<th>Age in years</th>
<th>0</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genomic Evaluation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Milk</td>
<td>62%</td>
<td>62%</td>
<td>70%</td>
</tr>
<tr>
<td>Fert</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td><strong>Traditional Evaluation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Milk</td>
<td>30%</td>
<td>35%</td>
<td>45%</td>
</tr>
<tr>
<td>Fert</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
THE TECHNOLOGY OF THE NEW PARADIGM
BeadArray™ Technology: Array format generation

BeadChip

- photo-resist
- silicon wafer
- plasma etching
- cleaning
GoldenGate® Custom Genotyping

Assay Plex:

48

1536

GoldenGate BeadArray:

- 96 to 1536-plex
- DNA input ~ 250ng
- iScan system
- LIMS/Automation
- BeadChips
- ~288 samples/day

GoldenGate Veracode:

- 48 to 384-plex
- DNA input ~ 250ng
- BeadXpress system
- Automation
- VeraCode
- ~288 samples/day

*Dependent on equipment and workflow*
Infinium iSelect® HD Custom Genotyping

- 12 Samples per Chip
- Fully customizable content
- Target Any Species, Any SNP
- Low DNA input (200ng)

Potential content could include:
- CNV, Rare variants
- Any species
- Candidate gene coverage
Consortia-developed Focused Genotyping Panels

Canine SNP20  Bovine SNP50  Equine SNP50  Porcine SNP60  Ovine SNP50  Canine HD  Maize SNP50  Bovine HD


……
<table>
<thead>
<tr>
<th>BREED</th>
<th>SAMPLES</th>
<th>POLYMORPHIC LOCI*</th>
<th>MEAN MAF†</th>
<th>MEDIAN MAF†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angus</td>
<td>60</td>
<td>41,491</td>
<td>0.21</td>
<td>0.21</td>
</tr>
<tr>
<td>Beefmaster</td>
<td>24</td>
<td>42,925</td>
<td>0.22</td>
<td>0.21</td>
</tr>
<tr>
<td>Bos indicus</td>
<td>24</td>
<td>23,971</td>
<td>0.11</td>
<td>0.02</td>
</tr>
<tr>
<td>Gir</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bos indicus</td>
<td>21</td>
<td>25,814</td>
<td>0.11</td>
<td>0.02</td>
</tr>
<tr>
<td>Nelore</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brahman</td>
<td>25</td>
<td>30,284</td>
<td>0.13</td>
<td>0.08</td>
</tr>
<tr>
<td>Brown Swiss</td>
<td>24</td>
<td>36,347</td>
<td>0.19</td>
<td>0.17</td>
</tr>
<tr>
<td>Charolais</td>
<td>26</td>
<td>42,589</td>
<td>0.22</td>
<td>0.21</td>
</tr>
<tr>
<td>Guernsey</td>
<td>21</td>
<td>38,632</td>
<td>0.19</td>
<td>0.17</td>
</tr>
<tr>
<td>Hereford</td>
<td>32</td>
<td>42,992</td>
<td>0.20</td>
<td>0.23</td>
</tr>
<tr>
<td>Holstein</td>
<td>64</td>
<td>42,730</td>
<td>0.22</td>
<td>0.22</td>
</tr>
<tr>
<td>Jersey</td>
<td>28</td>
<td>35,976</td>
<td>0.18</td>
<td>0.14</td>
</tr>
<tr>
<td>Limousin</td>
<td>45</td>
<td>42,821</td>
<td>0.22</td>
<td>0.22</td>
</tr>
<tr>
<td>N'Dama</td>
<td>25</td>
<td>29,049</td>
<td>0.14</td>
<td>0.08</td>
</tr>
<tr>
<td>Norwegian Red</td>
<td>21</td>
<td>42,782</td>
<td>0.22</td>
<td>0.21</td>
</tr>
<tr>
<td>Piedmontese</td>
<td>24</td>
<td>42,185</td>
<td>0.22</td>
<td>0.21</td>
</tr>
<tr>
<td>Red Angus</td>
<td>15</td>
<td>40,188</td>
<td>0.21</td>
<td>0.20</td>
</tr>
<tr>
<td>Romagnola</td>
<td>24</td>
<td>38,830</td>
<td>0.20</td>
<td>0.19</td>
</tr>
<tr>
<td>Santa Gertrudis</td>
<td>24</td>
<td>42,064</td>
<td>0.22</td>
<td>0.21</td>
</tr>
<tr>
<td>Sheko</td>
<td>20</td>
<td>35,726</td>
<td>0.17</td>
<td>0.12</td>
</tr>
<tr>
<td>Outgroup†</td>
<td>18</td>
<td>11,206</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td>Overall</td>
<td>565</td>
<td>47,545</td>
<td><strong>0.25</strong></td>
<td><strong>0.24</strong></td>
</tr>
</tbody>
</table>

*MAF > 0.05
†Across all 54,001 loci
††Bos bison, Bos gaurus, Bos grunniens, Bos javanicus, Bubalus depressicornis, and Syncerus caffer.

**BovineSNP50**

SNPs also informative in many breeds
The current trend in animal breeding (especially Dairy Cattle)

- Make use of the genomic information by genotyping to predict the genetic/genomic value of animals!
- Marker Assisted Breeding is changing the face of animal health, welfare and productivity

Towards Genomic Selection

- Cost savings in measuring the genetic merit of an animal have been forecast.
- A very large impact in a short time on the structure of the breeding industry

Illumina BovineSNP50 has become the standard whole-genome genotyping tool for research and industry in Cattle
THE BOVINE GENOMIC PORTFOLIO AT ILLUMINA
The current Bovine Genomic Portfolio

W H O L E  
G E N O M E  
S E Q U E N C I N G

B O V I N E H D

B O V I N E S N P 5 0

B O V I N E 3 K

P A R E N T A G E  
P A N E L
From BovineSNP50 v1 to v2

Illumina BovineSNP50 has become the standard whole-genome genotyping tool for research and industry in Cattle
BovineSNP50 has been resynthesized on the 2µm 24-sample BeadChip Format

- Faster sample throughput
- Faster scan times per sample
- Availability of 24-sample tip guide for more accurate loading

Due to the re-synthesis of beadpool onto 2µm beads,

the content of v2 is slightly different than v1

- 54,609 loci are on BovineSNP50v2
- 52,340 loci are common to v1 and v2
- 1661 loci on v1 are not on v2
- 2269 loci are new on v2

The concordance of genotypes between v1 and v2 is 99.72%
**BovineHD**

The Next Generation High Density Bovine BeadChip

- Whole genome research array to complement the current BovineSNP50 for the identification of agriculturally important genes and for a deeper implementation of genomic selection in cattle

- Developed in collaboration
  - USDA-ARS
  - Pfizer Animal Genetics
  - University of Missouri
  - UNCEIA, INRA
  - Other high-profile agricultural organizations

*BovineHD : Groups, institutions and companies involved in the development*

**USA**
- USDA
- George Mason University
- University of Missouri
- University of Illinois
- Pfizer

**Brasil**
- UNESP, Aracatuba
- Embrapa, Brasil

**France**
- UNCEIA-INRA

**Netherlands**
- CRV

**Republic of Korea**
- Chungbuk National University
- Yeungnam University, Gyeongsan

**Danemark**
- University of Aarhus, Danemark

**Germany**
- Technical University of Munich

**Italy**
- University of Piacenza
- University of Milano

**UK**
- Roslin Institute
Assay Design Results

- **All of BovineSNP50 SNPs** will be attempted
- ~800,000 bead types after 60,800 beads of BovineSNP50 positioned
- 795,000 SNP positioned on BTA1-29,X
- 5,000 beads represent unknown contigs, BTA Y, and mitochondrial SNP

Breed groups used:
- Holstein, Angus, Nelore, Taurine dairy, Taurine beef, Indicine, tropically adapted Taurine

- 852,645 total gaps
  - 850,816 <20kb
  - 1795 >20kb, < 100kb
  - 34 > 100 kb

*C. Van Tassell*, PAG, January 2010
Intervals between adjacent SNPs under 20,000 bp

C. Van Tassell, PAG, January 2010
Bovine 3K panel : 3072 SNPs (GoldenGate Technology)

- Designed in collaboration with the USDA and Dairy Cattle Breed Associations
- Enables broader access of cattle herds to genetic tests for the evaluation of net merit
- Content derived from BovineSNP50 and designed to accurately impute back to BovineSNP50
- Includes > 100 parentage SNPs (Heaton’s panel, USDA)
- Less sensitive to DNA quality (esp. Sheared DNA)
In Silico Genome Re-Sequencing

Use Low-Density SNP Chip

Modified from Hayes, 2009
In Silico Genome Re-Sequencing

Use Low-Density SNP Chip to infer Sequence data

Modified from Hayes, 2009
Selection of SNPs (*for low-plex genotyping*)

- Dense genotypic data available for
  - Both parents
  - Only sires

Selection of SNPs

- Magnitude of the estimated marker effects
- Expected contributions to the genetic variance
  - In previous generation
  - In the next generation
  - *Evenly spacing on chromosome*

Some concerns about the *accuracy of Genomic Evaluation*

- Across populations
- Over generations

*Time consuming* (computing involved)
Whole Genome (re-)Sequencing

Research

Whole genome sequencing of a single Bos taurus animal for single nucleotide polymorphism discovery
Sebastian H Eck**, Anna Benet-Pagés**, Krzysztof Fisikowski*, Thomas Meitinger**, Ruedi Fries* and Tim M Strom***

Addresses: *Institute of Systems Genetics, Julius Maximilian University, German Research Center for Environment Studies, Garching, Germany; **Institute of Human Genetics, University of Wuerzburg, Wuerzburg, Germany; ***Institute of Human Genetics, Julius Maximilian University, Wuerzburg, Germany; reit@imq.uni-wuerzburg.de

* These authors contributed equally to this work.

Published: 6 August 2009
Received: 21 April 2009
Accepted: 6 August 2009

Open Access

Abstract

Background: The majority of the 2 million bovine single nucleotide polymorphisms (SNPs) currently available in dbSNP have been identified in a single breed, Hereford cattle, during the bovine genome project. In an attempt to evaluate the variance of a second breed, we have produced a whole genome sequence at low coverage of a single Fleckvieh bull.

Results: We generated 24 gigabases of sequence, mainly using 36 bp paired-end reads, resulting in an average 7.4-fold sequence depth. This coverage was sufficient to identify 2.44 million SNPs, 62% of which were previously unknown, and 115,000 genotypes of the same animal, generated on a 50 k chip of 74% and 26% for homozygous and heterozygous SNPs determined by comparison with genotypes determined approximately 1.1%. We further determined the allelic status and Ruminant breeds. 95% of the SNPs were polymorphic of 24.5% and with 83% of the SNPs having a minor allele frequency.

Conclusions: This work provides the first single cow whole genome sequencing. The chosen approach - low to medium coverage results in SNP calling accuracy and quality as well as high efficiency in SNP discovery and genotyping in the Fleckvieh breed.
Using Whole Genome Sequence Data in Genetic Evaluation *(Meuwissen & Goddard, 2010)*

- Accuracies of prediction of genetic values increased by more than 40% relative to the use of dense ~30K SNP chips.
- Predictions of genetic values remained accurate even when the training and evaluation data were 10 generations apart.
- Similar accuracies to those where test and training data come from the same generation.

**These results suggest that with a combination of genome sequence data, large sample sizes and a statistical method that detects the polymorphisms that are informative, high accuracy is attainable.**

- Not possible for a cost point of view.
- The costs may be reduced substantially by the whole-genome sequencing of a limited subset of the individual.
CONCLUSIONS
Optimization of Genomic Breeding

Genomic Tools

- Whole Seq.
- BovineHD
- BovineSNP50
- Bovine3K
- Custom Sol.
- Not genotyped
## Optimization of Genomic Breeding

*Different combinations of genomic tools*

### Actions, tools & Costs

<table>
<thead>
<tr>
<th>Biological sample</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomic tool</td>
<td>Bovine3K</td>
</tr>
<tr>
<td>Genotyping</td>
<td>1</td>
</tr>
<tr>
<td>Imputation</td>
<td>1</td>
</tr>
<tr>
<td>Genomic Breeding Value</td>
<td>1</td>
</tr>
</tbody>
</table>

- **Imputation needed**
- **Imputation Error**
- **Accuracy of GE**
Optimization of Genomic Breeding

Different combinations of genomic tools

<table>
<thead>
<tr>
<th>Actions, tools &amp; Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biological sample</td>
</tr>
<tr>
<td>Genomic tool</td>
</tr>
<tr>
<td>Genotyping</td>
</tr>
<tr>
<td>Imputation</td>
</tr>
<tr>
<td>Genomic Breeding Value</td>
</tr>
</tbody>
</table>

Imputation needed; Error due to Imputation reduced; Accuracy of GE
Optimization of Genomic Breeding

Different combinations of genomic tools

<table>
<thead>
<tr>
<th>Actions, tools &amp; Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biological sample</td>
</tr>
<tr>
<td>Genomic tool</td>
</tr>
<tr>
<td>Genotyping</td>
</tr>
<tr>
<td>Imputation</td>
</tr>
<tr>
<td>Genomic Breeding Value</td>
</tr>
</tbody>
</table>

No imputation; no bias; individuals to be integrated in future reference population.
Towards « The genome-assisted Barnyard »

*Nature Biotechnology 27, 487 (2009)*

- Livestock genomes are providing
  - Hugely valuable biological information
  - *Immediate benefits to the way livestock breeders go about their business*

- « In contrast to the slow translation of human genome information into medicine, animal genomics is likely to have a rapid and tangible impact on Agriculture »
Economical optimization of genomic breeding:

Genomic evaluation is one important step but not the only one.

<table>
<thead>
<tr>
<th>Price (DNA + Genomic Tool + Analysis)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information &amp; Accuracy (Genomic Information)</td>
</tr>
</tbody>
</table>

Getting the biological sample

Blood, sperm, hair roots, ear punches, nasal ...

Genotyping using the genomic tool of choice

BovineSNP50, BovineHD, 3K SNP, ...

Computing the genomic breeding value

Prediction equation

Genomic breeding value = \( w_1x_1 + w_2x_2 + w_3x_3 \ldots \)
Discover. Validate. Screen

Thank you for your attention!

aeggen@illumina.com

A new evolution in agriculture genetic analysis.