Analytical and statistical consideration on the use of the ICAR-SNP panel for parentage control, genotyped with the Illumina bead chip technology, exemplified on the German Holstein population.

Prof. Dr. Ekkehard Schütz, MD, PhD  
Prof. Dr. Dr. Bertram Brenig, DVM, PhD  

Center for Molecular Diagnostics  
Institute of Veterinary Medicine  

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN  

Germany
Background

- **Paternity control is moving from STR to SNP based systems**
  - Cost savings if SNP-genomic selection is done
  - Excl. power in specific livestock populations needs to be defined

- **Genomic Selection based on thousands of SNP loci (LM)**
  - Tolerant for low confidence calls

- **Pat. Cntrl. based on hundreds (high effect of single values)**
  - Not that tolerant for GC calls
  - Ways of addressing this are desirable
Exclusion Power

- **STRs have shifted over time in German HF**
  - Example for two STRs (~20,000/ year)
Minor Allelic Frequencies

- **SNP genotyping based on Illumina’s Bead-Chip technology**
  - Bov50Kv2 & EuroG 10k
- **EP for SNP based on MAF (0 -0.5)**
  - MAF calculated for 200 ISAG SNPs in 25,000 German HF
Exclusion Power
Exclusion Power (SNP vs. STR)

- EP of STRs and (100 core) SNPs compared
SNP genotyping
What are we talking about technically?
SNP genotyping
What are we talking about technically?
DNA was purified from these cells and was then genotyped at over a million loci using Illumina’s Omni1-Quad Beadchip. Because our assay is particularly sensitive to false positive homozygous SNP calls for the recipient, we limited our focus to Beadchip SNPs with high GenCall and Cluster Separation scores, which yielded ~150,000 usable loci in each case.

To minimize false positive homozygous recipient SNP calls, only SNP loci with a GenCall score >0.70 and a Cluster Separation score of 1.00 were considered....

From: Snyder TM, Kush KK, Valentine HA and Quake SR. PNAS 2011: 108; 6229-34
........ during the standard clustering process. This is set by the GenCall threshold for clustering. Illumina recommends that you use a GenCall Score cut-off of 0.15 for Infinium arrays.

From: Illumina Technical Support; email in April 2014
Goodness of Gene Calls

Using SNP with GenCall score >0.70 – an option?
Bead Chip Technology Limitations

Anything we can do about it?
Limitations in Molecular Diagnostic

High GC-Score average

Low GC-Score average
Limitations in Molecular Diagnostic

Very Low GC-Score average (<6)

<table>
<thead>
<tr>
<th>Description</th>
<th>Max score</th>
<th>Total score</th>
<th>Query cover</th>
<th>E value</th>
<th>Ident</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bos taurus breed Hereford chromosome 12, alternate assembly Blau 4.6.1</td>
<td>99.6</td>
<td>303</td>
<td>100%</td>
<td>1e-19</td>
<td>100%</td>
<td>NC_007310.5</td>
</tr>
<tr>
<td>Bos taurus breed Hereford chromosome 16, alternate assembly Blau 4.6.1</td>
<td>32.2</td>
<td>399</td>
<td>98%</td>
<td>19</td>
<td>100%</td>
<td>NC_007314.4</td>
</tr>
<tr>
<td>Bos taurus breed Hereford chromosome 5, alternate assembly Blau 4.6.1</td>
<td>30.2</td>
<td>227</td>
<td>90%</td>
<td>74</td>
<td>95%</td>
<td>NC_007331.4</td>
</tr>
<tr>
<td>Bos taurus breed Hereford chromosome 13, alternate assembly Blau 4.6.1</td>
<td>30.2</td>
<td>284</td>
<td>92%</td>
<td>74</td>
<td>100%</td>
<td>NC_007315.4</td>
</tr>
<tr>
<td>Bos taurus breed Hereford chromosome 5, alternate assembly Blau 4.6.1</td>
<td>32.2</td>
<td>443</td>
<td>86%</td>
<td>19</td>
<td>100%</td>
<td>NC_007303.5</td>
</tr>
<tr>
<td>Bos taurus breed Hereford chromosome 9, alternate assembly Blau 4.6.1</td>
<td>32.2</td>
<td>403</td>
<td>86%</td>
<td>19</td>
<td>100%</td>
<td>NC_007302.5</td>
</tr>
<tr>
<td>Bos taurus breed Hereford chromosome 6, alternate assembly Blau 4.6.1</td>
<td>30.2</td>
<td>516</td>
<td>86%</td>
<td>74</td>
<td>100%</td>
<td>NC_007304.5</td>
</tr>
<tr>
<td>Bos taurus breed Hereford chromosome 20, alternate assembly Blau 4.6.1</td>
<td>30.2</td>
<td>284</td>
<td>86%</td>
<td>74</td>
<td>100%</td>
<td>NC_007310.5</td>
</tr>
<tr>
<td>Bos taurus breed Hereford chromosome 14, alternate assembly Blau 4.6.1</td>
<td>32.2</td>
<td>467</td>
<td>84%</td>
<td>19</td>
<td>95%</td>
<td>NC_007312.5</td>
</tr>
<tr>
<td>Bos taurus breed Hereford chromosome 8, alternate assembly Blau 4.6.1</td>
<td>30.2</td>
<td>371</td>
<td>84%</td>
<td>74</td>
<td>91%</td>
<td>NC_007306.5</td>
</tr>
<tr>
<td>Bos taurus breed Hereford chromosome 10, alternate assembly Blau 4.6.1</td>
<td>36.2</td>
<td>437</td>
<td>80%</td>
<td>1.2</td>
<td>100%</td>
<td>NC_007306.5</td>
</tr>
<tr>
<td>Bos taurus breed Hereford chromosome 19, alternate assembly Blau 4.6.1</td>
<td>32.2</td>
<td>290</td>
<td>80%</td>
<td>19</td>
<td>100%</td>
<td>NC_007317.5</td>
</tr>
<tr>
<td>Bos taurus breed Hereford chromosome 2, alternate assembly Blau 4.6.1</td>
<td>30.2</td>
<td>603</td>
<td>80%</td>
<td>74</td>
<td>100%</td>
<td>NC_007300.5</td>
</tr>
<tr>
<td>Bos taurus breed Hereford chromosome 22, alternate assembly Blau 4.6.1</td>
<td>30.2</td>
<td>256</td>
<td>78%</td>
<td>74</td>
<td>100%</td>
<td>NC_007302.5</td>
</tr>
<tr>
<td>Bos taurus breed Hereford chromosome 27, alternate assembly Blau 4.6.1</td>
<td>30.2</td>
<td>227</td>
<td>78%</td>
<td>74</td>
<td>100%</td>
<td>NC_007328.4</td>
</tr>
<tr>
<td>Bos taurus breed Hereford chromosome 1, alternate assembly Blau 4.6.1</td>
<td>32.2</td>
<td>605</td>
<td>76%</td>
<td>19</td>
<td>100%</td>
<td>NC_007289.5</td>
</tr>
<tr>
<td>Bos taurus breed Hereford chromosome 15, alternate assembly Blau 4.6.1</td>
<td>32.2</td>
<td>262</td>
<td>74%</td>
<td>19</td>
<td>100%</td>
<td>NC_007315.5</td>
</tr>
<tr>
<td>Bos taurus breed Hereford chromosome 17, alternate assembly Blau 4.6.1</td>
<td>32.2</td>
<td>362</td>
<td>74%</td>
<td>19</td>
<td>95%</td>
<td>NC_007315.5</td>
</tr>
<tr>
<td>Bos taurus breed Hereford chromosome 11, alternate assembly Blau 4.6.1</td>
<td>34.2</td>
<td>661</td>
<td>72%</td>
<td>4.7</td>
<td>100%</td>
<td>NC_007309.5</td>
</tr>
</tbody>
</table>
GUIDELINES FOR PARENTAGE VERIFICATION BASED ON SNP MARKERS are,
SNP profile: Minimum number of SNPs in panel: 100; Minimum number of SNPs available in profile: 95
(if less than 95 SNPs can be scored, retest the sample or request a new sample).

If mismatches occur in a supposed parentage, the general rule first is to retest the samples involved or request new samples to confirm the determined genotypes. If the genotypes are confirmed the following guidelines are suggested.

<table>
<thead>
<tr>
<th>Minimum number of SNPs</th>
<th>one parent</th>
<th>both parents</th>
<th>Certify:</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
<td>85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of mismatches*:</td>
<td>0 - 1</td>
<td>0 - 2</td>
<td></td>
</tr>
<tr>
<td>parentage accepted</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of mismatches*:</td>
<td>2 - 3</td>
<td>3 - 4</td>
<td></td>
</tr>
<tr>
<td>parentage doubtful, use additional panel</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of mismatches*:</td>
<td>≥ 4</td>
<td>≥ 5</td>
<td></td>
</tr>
<tr>
<td>parentage excluded</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Example: offspring = GG, sire = AA, offspring = AG, sire = AA, dam = AA

**: When the parentage is doubtful, first genotype the samples with both panels (ISAG panel and the backup) and then, if results remain doubtful, ask customer for other possible parents, if there are no other then qualify the parents. In any case, it is recommended that samples be retested if there is a parentage exclusion with ISAG panel and/or back up panel.
SNP Parentage
- the effect of errors-

[Chart showing bar graph with categories: <2, 2-3, 2-3 (Backup), >3, excluded (backup).]
SNP Genotyping based Parentage - process error control-
• Using each individual GC-Score as censor:
  
  • Calculating average and standard deviation for 200 panel SNPs
    • 20,000 bov50kV2
    • 7,000 EuroG 10k
  
  • Defining the error for a given individual SNP call
    • Cross-platform (50k; EuroG 10k) / cross-version
    • Transform to Z-Values (per SNP): \( \frac{\text{GC}_i - \overline{\text{GC}_s}}{\sigma(\text{GC}_s)} \)
  
  • Defining the error for a given parentage (per SNP)
  
  • Additive EP: \( z_{\text{total}} = \sqrt{\sum Z^2} \); \( Z > 1 \) set to 0
  
  • Test for effect in simulations
  • Test for performance in randomly chosen field data set
Simulated loss by censoring

The graph shows the percent combinations censored for different values of Z ranging from 2 to 5.33. The bars are color-coded with blue representing LD simulated and red representing All simulated. The y-axis represents the percent combinations censored, ranging from 0% to 14%. The x-axis lists the values of Z, starting from Z-2 to Z-5.33.
SNP Parentage
- process error control optimization -
SNP-based PC has high EP for one and both parents
  - 2.5 and 6 logs better than STRs

Widely used Illumina Bead-Chips error prone (as any system)
  - e.g. false homozygous calls

A process error control is proposed that:
  - Minimizes doubtful results (>90%)
  - Avoids false parentage exclusions (100%)
  - Harboring only low censoring frequency (<3%)
    - A priori vs a posteriori application?
Discussion points

- Include a goodness of GC measure in DNA certificates?
- Rethinking the 200 SNP Panel
  - Avoid high GC% loci
  - Avoid loci within or adjacent to Repetitive Elements
- Rethinking the Recommendations
  - Do we need 90 / 85 SNPs for Pat. Cntrl. given the EP?
Acknowledgements

- Team of IVM:
  - Melanie Scharfenstein, Susann Loos, Louisa Jüttner

- German Holstein Association (DHV)
  - RUW
  - OHG
  - ZBH
  - VOST
  - WEU
  - WEU
  - RBW
  - SRB