Fine-mapping of QTL using high-density SNP genotypes

J. B. Cole, P. M. VanRaden, D. J. Null, T. S. Sonstegard, M. McClure, C. P. Van Tassell, H. A. Adams, and H. A. Lewin

1Animal Improvement Programs and 2Bovine Functional Genomics Laboratories, Agricultural Research Service, USDA, Beltsville, MD, USA
3Department of Animal Sciences & Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA
4Department of Evolution and Ecology, University of California, Davis, Davis, CA 95616, USA

john.cole@ars.usda.gov
Illumina genotyping arrays

BovineSNP50
- 54,001 SNPs (version 1)
- 54,609 SNPs (version 2)

BovineHD
- 777,962 SNPs

BovineLD
- 6,909 SNPs

Allows for additional SNPs (e.g., GeneSeek Genomic Profiler)
Bovine High-Density Bead Chip (HD)

- **778K** SNP chosen to
  - Be evenly spaced
  - Include some Y-specific SNP
  - Include mitochondrial SNP

- Utilize across-breed information
- Fine mapping of QTL
- Enhanced performance in Zebu cattle
Estimation of marker effects

Predict traditional PTA using phenotypes and pedigree

Compute SNP effects using deregressed PTA weighted by reliability

Bayes A-type model

Regress small effects to mean

Allow large effects to grow
We got right to the point…

...and found some interesting things

Is luck better than skill?

ARS-BFGL-NGS-109285 at 57,895,121 Mb on BTA18

Signal consistent in HD data

Intronic to a putative CD33-related Siglec gene

This region is gene-rich

Many Siglecs involved in leptin signaling

Also affects gestation length

50k SNP are good for regions, not genes

Simple strategy

Compute SNP effects for trait of interest

Look for peaks

Perform bioinformatics on regions under interesting peaks

NCBI/Ensembl

Bovine Gene Atlas

Bovine QTLdb
Can we look at every peak?
Are 777k SNP better than 50k?

Case studies

Identification of causal variants associated with two haplotypes related to fertility

Discovery and validation of HH1 in U.S. Holstein cattle

Discovery and validation of JH1 in U.S. Jersey cattle

Fine-mapping of the Weaver locus
Recessive defect discovery

Check for homozygous haplotypes

- 7 to 90 expected but none observed
- 5 of top 11 are potentially lethal
- 3.1% to 3.7% lower conception rates
- Some slightly higher stillbirth rates

Confirmed Brachyspina same way
### Novel haplotypes affecting fertility

<table>
<thead>
<tr>
<th>Name</th>
<th>Chromosome</th>
<th>Location</th>
<th>Carrier Freq</th>
<th>Earliest Known Ancestors</th>
</tr>
</thead>
<tbody>
<tr>
<td>HH1</td>
<td>5</td>
<td>62-68</td>
<td>4.5</td>
<td>Pawnee Farm Arlinda Chief</td>
</tr>
<tr>
<td>HH2</td>
<td>1</td>
<td>93-98</td>
<td>4.6</td>
<td>Willowholme Mark Anthony</td>
</tr>
<tr>
<td>HH3</td>
<td>8</td>
<td>92-97</td>
<td>4.7</td>
<td>Glendell Arlinda Chief, Gray View Skyliner</td>
</tr>
<tr>
<td>JH1</td>
<td>15</td>
<td>11-16</td>
<td><strong>23.4</strong></td>
<td>Observer Chocolate Soldier</td>
</tr>
<tr>
<td>BH1</td>
<td>7</td>
<td>42-47</td>
<td>14.0</td>
<td>West Lawn Stretch Improver</td>
</tr>
</tbody>
</table>

Economics of Next Generation Sequencing

Bovine genome = 2.85 billion bases
30X Coverage = about 90 billion bases
1 run on HiSeq2000 = 600 billion bases
20 animals sequenced to 30X coverage in 9 days (2 x 100 bp reads)
Cost about €19,600
HH1 in Holstein cattle

75 SNPs spanning 7 Mbp on *Bos taurus* chromosome 5

Traced to Pawnee Farm Arlinda Chief

Very popular bull

- >16,000 daughters
- >500,000 granddaughters
- >2 million recorded great-granddaughters
Sequencing and haplotype reconstruction

- Eight bulls in study derived from purchased semen - ~30X seq. coverage
- Previously sequenced-6X on 454 Roche

= Four bulls used to identify mutation causative for HH1
Locating HH1 causal variant

Original 75 SNP HH1 haplotype

Refined 38 SNP haplotype

Reference genome

C/T
Sequence analysis of HH1 SNP

The table below includes ANNOVAR results for Ensembl genes:

<table>
<thead>
<tr>
<th>Chr 5 coordinates 66,870,000 – 67,990,000]</th>
<th>SNPs</th>
<th>Ensembl</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,582 SNPs in region</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Downstream</td>
<td>2</td>
<td>A6QP25[2]</td>
<td></td>
</tr>
<tr>
<td>Exonic</td>
<td>3</td>
<td>APAF1[2], ENSB_17385[1]</td>
<td></td>
</tr>
<tr>
<td>Intergenic</td>
<td>1,221</td>
<td>ENSB_38223, THPS, MPCI, IKIR, ENSB_17385, A6QP25, ENSB_40364</td>
<td></td>
</tr>
<tr>
<td>Intrinsic</td>
<td>354</td>
<td>THPS[17], MPCI[2], IKIR_BOVIN[9], APAF1[88], ENSB_17385[1]</td>
<td></td>
</tr>
<tr>
<td>UTR3</td>
<td>2</td>
<td>THPS[2]</td>
<td></td>
</tr>
</tbody>
</table>

* SNPs indicate the number of SNPs identified within each region
* Brackets indicate the number of SNPs associated with the given gene

### Gene Information

<table>
<thead>
<tr>
<th>Gene</th>
<th>Location (BTau4.0)</th>
<th>Exon</th>
<th>Ref</th>
<th>Alt</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>APAF1</td>
<td>67,577,782</td>
<td>11</td>
<td>C</td>
<td>T</td>
<td>Stopgain</td>
</tr>
<tr>
<td>APAF1</td>
<td>67,626,029</td>
<td>25</td>
<td>C</td>
<td>T</td>
<td>Synonymous</td>
</tr>
<tr>
<td>ENSBTAG00000017385</td>
<td>67,636,761</td>
<td>2</td>
<td>C</td>
<td>T</td>
<td>Synonymous</td>
</tr>
</tbody>
</table>
Annotation of mutated gene

**APAF1** - *Bos taurus* apoptotic peptidase activating factor 1

ATP binding factor

Gene expression for *APAF1* in murine development begins between 7 and 9 d in heart, mesenchyme, periderm, and primitive intestine (Muller et al., 2005)

Gene knockout of *APAF1* in mice leads to embryonic lethality (Muller et al., 2005)

Proteins required for this pathway/cascade are important for neural tube closure *in vivo*
SNP validation in HH1 interval

Designed Sequenome 24 assays for 12 SNP in the HH1 interval

Encompasses all SNP for coding, 3’ UTR, and downstream regions – and 5 other SNP covering all other genes in the interval

Tested wide diversity of haplotypes for HH1 region – use SNP50 DNA archive
Concordance of genotype to HH1 status

1.2% false positives at intron SNP

Genome Coordinates UMD3.1
HH1 - Conclusions

HH1 Haplotype 12 100% concordant with heterozygosity at one exonic SNP location (N=486)

HH1 Haplotype 32 100% concordant with het status at same exonic SNP location (N=11)

Other 256 non-carrier animals for HH1 interval were homozygous normal at this exonic SNP

One intronic SNP was 98.8% concordant to HH1 carrier status

No other individual SNP were viable candidates

Based on collaborative work with Harris Lewin
NGS sequencing of JH1 carrier animals

- Sequencing run included Chocolate Soldier (JH1 founder) and Oman (for HH3)
- 30X coverage of whole genome per animal
- Funded by American Jersey Cattle Association
- Found 38 candidate SNP in 492 kbp interval
SNP Validation in JH1 Interval

Designed 30 Sequenom assays for 15 unique SNP in the JH1 interval

Only 1 SNP in a gene

- Stopgain mutation in \( CWC15 \) spliceosome-associated protein
- Not expressed in every bovine tissue

Test all JH1 carriers in the SNP50 repository

185 normal, 546 carriers
Spliceosome structure and *CWC15*

JH1 - SNP Validation Results

- JH1 Haplotype 99.3% concordant with \textit{CWC15} stopgain mutation
- 5 non-concordant samples
  - False negative JH1 haplotypes (2 cases)
  - DNA misplating? Retest samples
- 2 other SNP in complete LD with JH1-\textit{CWC15} stopgain
- Working with Jersey to complete an independent validation
Annotation is a big problem

Function of *CWC15* in the bovine is unknown

*CWC15* ortholog in mice, *MED1*, expressed in early embryos (Duan et al., 2010)

No analogous studies in large mammals

Funding

Time

Perceived impact
Preliminary fine-mapping of Weavers

35,353 SNP on BTA4

69 Brown Swiss bulls with HD genotypes

20 cases and 49 controls

No affected animals!

Microsatellite-mapped to the interval 43.2–51.2 cM

Phenotype based on name
Preliminary sliding-window analysis

BTA4_43-60Mb

-log10p

Mb
HD analysis and NGS

GWAS with Bovine HD

20 carrier and 50 controls

Collaboration with Italian consortium

Refined historic interval from 46-56Mb to 48-53Mb on BTA4

Weaver similar to ALS in humans

18 annotated genes in interval

– 7 of them interact with major gene responsible for heritable ALS

NGS performed on a pool of 10 Normal and 10 Carriers resulting in ~30x coverage

117 SNV in the Weaver locus,

– 1 synonymous, 3 non-synonymous

– Test all SNP
Validation of Weavers

5 Sequenom multiplexes tested 114 SNV

715 Brown Swiss, 26 Carora, 4 Angus, 4 Holstein, 4 Jersey, 3 Senepol, 3 Herefords

Phenotype mapping refined locus to 35 SNP

The test is accurate

5 Carora carriers based upon the 35 SNP

Multiple ‘assumed normal’ BS are carriers/affected

A few BS diagnosed normals were carriers

Found 1 Holstein heterozygous for the right 30 SNP and homozygous normal for the left 5 SNP

- Related (6.25%) to BSUSA000000183023 (MEADOW VIEW MATT ALEX) who is a confirmed Weaver carrier
Researchers Catalog Loss-of-Function Variants in Human Protein-Coding Genes

NEW YORK (GenomeWeb News) – A Wellcome Trust Sanger Institute and Yale University-led team has sifted through data from three 1000 Genomes Project pilot efforts to find a set of authentic loss-of-function variants in the human genome.

"Each of us can be walking around with at least 20 genes basically inactivated," the study's first author, Daniel MacArthur, told GenomeWeb Daily News.

If this is true for humans, then is it true for cattle???
Conclusions

Targeted resequencing is effective

Genotyping-by-sequencing can identify loss-of-function mutations

Functional studies of protein function *in vivo* are needed

New precision mating strategies should be developed
Acknowledgements

• AIPL, Beltsville, MD (USDA-ARS)
  - George Wiggans and Tabatha Cooper

• BFGL, Beltsville, MD (USDA-ARS)
  - Larry Shade
  - George Liu, Derek Bickhart, Rueben Anderson, Sasho A.
  - Steve Schroeder
  - Alicia Beavers

• University of Illinois, Urbana-Champaign
  - Denis Larkin

• Meat Animal Research Center, Clay Center, NE (USDA-ARS)
  - Tim Smith, Tara McDaneld, John Keele