High resolution copy number variable regions in Brown Swiss dairy cattle and their value as markers

MA Dolezal¹,²,³, K Schlangen¹,², F Panitz⁴, L Pellegrino¹, M Soller⁵, E Santus⁶, M Jaritz³, A Bagnato¹,⁷

1 University of Milan
2 Vetmeduni Vienna
3 FH Campus Vienna
4 Aarhus University
5 Hebrew University Jerusalem
6 ANARAB
7 G&B platform Filarete Foundation - UNIMI

This presentation represents the views of the Authors, not the EC. The EC is not liable for any use that may be made of the information.
CNVs are duplications, insertions and deletions of chromosomal segments in comparison to a reference genome.

- CNVs recognized as substantial source of genetic variation
- CNVs are summarized to copy number variable regions (CNVRs) at the population level
Objective

- evaluate potential contribution of CNVs as genetic markers for GWAS & GS in cattle

  - Polymorphic information content (PIC) of CNV loci
  - correlations between SNPs residing in CNVRs and their underlying CNVs
  - LD of SNPs residing in CNVRs with SNPs surrounding the CNVRs (adjacent SNPs)
Data for SNP array based CNV detection

- 192 BS bulls genotyped with Illumina HD chip
- Log R Ratios (LRR) - total signal intensities
- B allele frequencies (BAF) - allelic intensity ratio values

- LRR & BAF values for 735,239 SNPs on UMD3.1 autosome
CNV calling in 164 stringently quality filtered bulls
- PennCNV\(^1\) & genoCN\(^2\)
- reliable CNV calls ≥3 consecutive SNPs of the same type

1 Wang et al. (2007) doi: 10.1101/gr.6861907
2 Sun et al. (2009) doi: 10.1093/nar/gkp493
Material & Methods – 2

definition of CNVRs

- within each algorithm summarisation of CNVs to CNVRs
  - union set of CNVs\(^1\)

- high confidence set of CNVRs for population genetic analysis
  - intersection of overlapping CNVRs of same type\(^2\) across algorithms

1 Redon et al. (2006) doi: 10.1038/nature05329
2 Wain et al. (2009) doi:10.1371/journal.pone.0008175
Material & Methods – 3
identification of “real“ alleles

- genoCN\(^1\) employs a 3 copy number state model
  - 0-1-2 copies per haploid
  - possible alleles: 0, A, B, AA, BB and AB
  - total allelic content with highest posterior probability
    - eg. cn=3 AAB , possible alleles AA,B or AB,A
    - not equivalent to knowing the real alleles

- allele calling & phasing with polyHap\(^2\) v2.0

1 Sun et al. (2009) doi: 10.1093/nar/gkp493
Material & Methods – 4

population genetic characterization

\[ PIC = 1 - \sum_{i=1}^{n} p_i^2 \]

- LD between SNPs residing within CNVRs and their underlying CNV
  - standard metrics incorrect\(^1\)
  - \( r_c^2 \) correctly quantifies covariance\(^1\)

Global LD between SNPs in CNVRs & neighbouring SNPs: $W_n$ (Cramer’s $V^{1,2}$)

$$W_n = \left[ \sum_{i=1}^{I} \sum_{j=1}^{J} D_{ij}^2 \frac{p_{ij}}{\min(I-1,J-1)} \right]^{1/2} = \left[ \frac{X_{LD}^2}{2N \min(I-1,J-1)} \right]^{1/2}$$

1 Cramer (1946) Mathematical Models of Statistics
Results - 1
number of alleles in CNVRs
Results - 2
Polymorphic Information Content

PIC

cn=2: copy number normal
cn≠2: copy number variable
Results - 3
LD between SNPs in CNVRs & underlying CNV

$r_c^2$
Results – 4 Global LD between neighbouring SNPs and SNPs in CNVRs

<table>
<thead>
<tr>
<th>U10</th>
<th>U9</th>
<th>U8</th>
<th>U7</th>
<th>U6</th>
<th>U5</th>
<th>U4</th>
<th>U3</th>
<th>U2</th>
<th>U1</th>
<th>R1</th>
<th>R..</th>
<th>Rn</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>D5</th>
<th>D6</th>
<th>D7</th>
<th>D8</th>
<th>D9</th>
<th>D10</th>
</tr>
</thead>
<tbody>
<tr>
<td>far</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>close</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>close</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>far</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Wn

**loss**

**gain**

**complex**
Conclusions

- CNVs are valuable genetic markers
  - high PIC
  - not sufficiently tagged by SNPs on HD chip

Thank you for your attention!
Quality filtering
Distances
Results – 4

LD

loss CNVRs

gain CNVRs

complex CNVRs
Figure 1 from Kato et al. (2011) doi: 10.1534/g3.111.000174