A rapid method for the identification of epistatic ‘dormant’ SNPs

Epistasis = When a gene masks another gene
= Gene x Gene
= \( N^2 \) Comparisons

Epistasis against whole genome
= When the genome masks a gene
= \( N \) Comparisons

**AIM:** Identify epistatic SNP based on those SNP with significant yet opposed effect depending on the genetic background
Epistasis against whole genome

Identifying Quantitative Trait Locus by Genetic Background Interactions in Association Studies

Jean-Luc Jannink

J. Dairy Sci. 94:1597–1600
doi:10.3168/jds.2010-3834

Short communication: Evidence for a major gene by polygene interaction for milk production traits in German Holstein dairy cattle

M. Streit,* N. Neugebauer,* T. H. E. Meuwissen,† and J. Bennnewitz‡

Detecting epistasis with the marginal epistasis test in genetic mapping studies of quantitative traits

Lorin Crawford,1,2,* Ping Zeng,4,5, Sayan Mukherjee,6,7,8,9, Xiang Zhou4,5,*
Epistasis against whole genome

I’m SHORT ➔ There’s a gene trying to make me taller

She’s TALL ➔ The same gene is trying to make her shorter
METHOD 1: Bin-based “Mechanical Heuristics”:
(Reverter & Henshall, 2017. Poster at Gordon Conference, Galveston, TX)

**NO** – Run GBLUP and Rank individuals from lowest to highest GEBV.

**OS** – Create 5 equally-sized bins with BIN1 containing the 20% individuals with lowest GEBV, BIN2 the next 20%, ...and so on until BIN5 with 20% with highest GEBV.

**ES** – Within each bin, perform a GWAS (ie. Regression of phenotype on SNP genotype).

**ATRO** – Call epistatic SNP those with significant yet opposed effect in BIN1 and BIN5, and a monotonic pattern of effects from BIN1 to BIN5 (ie. “negative to positive” or “positive to negative”).

**NCO** – Confirm the SNP collected are not significant in GWAS using the entire population.
METHOD 2: Regression of Residuals on GEBV:

The quantity of interest is the regression of \( y \) on \( Z_i u \), which can be approximated as follows:

**UNO** – Run GBLUP and extract Residuals (\( \hat{e} \)) and GEBV (\( \hat{u} \))

**DOS** – For each SNP in \( i \):

a. Multiply \( \hat{u} \) by centred gene content: \( Z_i \hat{u} \)

b. Run a single-marker regression:
\[
\hat{e} = \mathbf{1}\mu + (\alpha\alpha)_i Z_i \hat{u} + \epsilon
\]

c. Obtain a t-test and associated \( P \)-value for \( (\alpha\alpha)_i \)

NB: This approximate method is VERY FAST, but ignores the uncertainty in the estimation of \( \hat{e} \) and \( \hat{u} \). It may be used for a fast screening followed by a REML analysis for a subset.
Data & Methods:

(1) PHENOTYPE: Yearling Weight in 2,111 Brahman cattle.
(2) GENOTYPE: 651,253 SNP with MAF > 1%.
(3) GENOMIC RELATIONSHIP MATRIX (GRM): \[ G = \frac{M M^T}{2 \sum p_i (1 - p_i)} \]
(4) GBLUP: \[ y = X \beta + Zu + e \]
\[ V(u) = G \sigma^2_u \text{ and } V(e) = I \sigma^2_e \]
(5) GWAS: \[ y = X \beta + Zu + S_i a_i + e \]

We used the Qxpak5 software [1] for GBLUP and GWAS.
Results:

Manhattan Plot of the GWAS for Yearling Weight using the Entire Population: Highlighted is the region in BTA14 around PLAG1, a known QTL for growth in cattle [2].

Negative Control?
**METHOD 1:** Bin-based “Mechanical Heuristics”:

- Rank individuals from lowest to highest GEBV.

- Create 5 equally-sized bins.

- Within each bin, perform a GWAS.

**TRO** – Call epistatic SNP those with significant yet opposed effect in BIN1 and BIN5.
<table>
<thead>
<tr>
<th>BIN 1</th>
<th>BIN 2</th>
<th>BIN 3</th>
<th>BIN 4</th>
<th>BIN 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>N Animals</td>
<td>422</td>
<td>422</td>
<td>422</td>
<td>422</td>
</tr>
<tr>
<td>GEBV Range</td>
<td>-4.1 ; -9.3</td>
<td>-9.3 ; 2.4</td>
<td>-2.3 ; 2.8</td>
<td>2.8 ; 9.7</td>
</tr>
<tr>
<td>Phenotype Range</td>
<td>157.8 ; 272.0</td>
<td>211.4 ; 291.0</td>
<td>224.9 ; 303.0</td>
<td>233.3 ; 323.6</td>
</tr>
</tbody>
</table>

- Look for SNP going like this
- or for SNP going like this

- 113 Negative (BIN1) to Positive (BIN5)
- 130 Positive (BIN1) to Negative (BIN5)
Estimated SNP effects within BINs and in the whole population

Two examples of “Negative to Positive” and “Positive to Negative” pattern as well as for a SNP in the PLAG1 coding region. Asterisks indicate significance at P < 0.001.

<table>
<thead>
<tr>
<th>SNP chr:Mb (Gene)</th>
<th>BIN1</th>
<th>BIN2</th>
<th>BIN3</th>
<th>BIN4</th>
<th>BIN5</th>
<th>Whole</th>
</tr>
</thead>
<tbody>
<tr>
<td>18:56.5 (CPT1C)</td>
<td>-7.58*</td>
<td>-1.38</td>
<td>-0.86</td>
<td>2.67</td>
<td>4.84*</td>
<td>0.60</td>
</tr>
<tr>
<td>28:23.3 (CTNNA3)</td>
<td>-8.00*</td>
<td>-3.05</td>
<td>-1.93</td>
<td>1.55</td>
<td>5.06*</td>
<td>-0.05</td>
</tr>
<tr>
<td>27:1.1 (CSMD1)</td>
<td>4.53*</td>
<td>1.18</td>
<td>-0.66</td>
<td>-0.69</td>
<td>-2.37*</td>
<td>0.58</td>
</tr>
<tr>
<td>5:54.9 (LRIG3)</td>
<td>4.86*</td>
<td>-0.30</td>
<td>-0.79</td>
<td>-2.05</td>
<td>-3.94*</td>
<td>1.18</td>
</tr>
<tr>
<td>14:25.0 (PLAG1)</td>
<td>0.74</td>
<td>2.07</td>
<td>3.46*</td>
<td>2.15</td>
<td>2.02</td>
<td>4.71*</td>
</tr>
</tbody>
</table>
Estimated SNP effects within BINs and in the whole population

Two examples of “Negative to Positive” and “Positive to Negative” pattern as well as for a SNP in the PLAG1 coding region. Asterisks indicate significance at $P < 0.001$.

<table>
<thead>
<tr>
<th>SNP chr:Mb (Gene)</th>
<th>BIN1</th>
<th>BIN2</th>
<th>BIN3</th>
<th>BIN4</th>
<th>BIN5</th>
<th>Whole</th>
<th>$R(\hat{e}.on.\hat{u})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>18:56.5 (CPT1C)</td>
<td>-7.58*</td>
<td>-1.38</td>
<td>-0.86</td>
<td>2.67</td>
<td>4.84*</td>
<td>0.60</td>
<td>4.48*</td>
</tr>
<tr>
<td>28:23.3 (CTNNA3)</td>
<td>-8.00*</td>
<td>-3.05</td>
<td>-1.93</td>
<td>1.55</td>
<td>5.06*</td>
<td>-0.05</td>
<td>3.33*</td>
</tr>
<tr>
<td>27:1.1 (CSMD1)</td>
<td>4.53*</td>
<td>1.18</td>
<td>-0.66</td>
<td>-0.69</td>
<td>-2.37*</td>
<td>0.58</td>
<td>-3.24*</td>
</tr>
<tr>
<td>5:54.9 (LRIG3)</td>
<td>4.86*</td>
<td>-0.30</td>
<td>-0.79</td>
<td>-2.05</td>
<td>-3.94*</td>
<td>1.18</td>
<td>-4.44*</td>
</tr>
<tr>
<td>14:25.0 (PLAG1)</td>
<td>0.74</td>
<td>2.07</td>
<td>3.46*</td>
<td>2.15</td>
<td>2.02</td>
<td>4.71*</td>
<td>1.02</td>
</tr>
<tr>
<td>BIN1</td>
<td>BIN2</td>
<td>BIN3</td>
<td>BIN4</td>
<td>BIN5</td>
<td>Whole</td>
<td>Re.r.u.</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>-------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>-7.58</td>
<td>-1.38</td>
<td>-0.86</td>
<td>2.67</td>
<td>4.84*</td>
<td>0.60</td>
<td>4.48*</td>
<td></td>
</tr>
<tr>
<td>-8.00</td>
<td>-3.05</td>
<td>-1.93</td>
<td>1.55</td>
<td>5.06*</td>
<td>-0.05</td>
<td>3.33*</td>
<td></td>
</tr>
<tr>
<td>4.53</td>
<td>1.18</td>
<td>-0.66</td>
<td>-0.69</td>
<td>-2.37*</td>
<td>0.58</td>
<td>-3.24*</td>
<td></td>
</tr>
<tr>
<td>4.86</td>
<td>-0.30</td>
<td>-0.79</td>
<td>-2.05</td>
<td>-3.94*</td>
<td>1.18</td>
<td>-4.44*</td>
<td></td>
</tr>
<tr>
<td>0.74</td>
<td>2.07</td>
<td>3.46*</td>
<td>2.15</td>
<td>2.02</td>
<td>4.71*</td>
<td>1.02</td>
<td></td>
</tr>
</tbody>
</table>

\[ \frac{(t_{BIN5} - t_{BIN1})}{2} \sim \text{Regr}(\hat{e}, \hat{u}) \]
### Candidate Genes

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>BINs</th>
<th>Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>BovineHD0500015637</td>
<td>LRIG3</td>
<td>-4.403</td>
<td>-4.439</td>
</tr>
<tr>
<td>BovineHD2000011089</td>
<td>PRLR</td>
<td>3.581</td>
<td>4.913</td>
</tr>
<tr>
<td>BovineHD0600010671</td>
<td>LAP3</td>
<td>-2.917</td>
<td>-4.460</td>
</tr>
<tr>
<td>BovineHD0200001836</td>
<td>MSTN</td>
<td>5.172</td>
<td>3.079</td>
</tr>
<tr>
<td>BovineHD0500005343</td>
<td>KITLG</td>
<td>-4.961</td>
<td>-2.991</td>
</tr>
<tr>
<td>BovineHD0700004860</td>
<td>INSR</td>
<td>2.234</td>
<td>3.986</td>
</tr>
</tbody>
</table>

\[
\frac{(t_{\text{BIN5}} - t_{\text{BIN1}})}{2} \sim \text{Regr}(\hat{\theta}, \hat{\mu})
\]
CONCLUSIONS: We regard these SNPs as being ‘dormant’ with an effect waiting to be ‘released’ when selection moves the population to either tail of the distribution. Further, these SNPs could provide an answer to the long-standing paradox by which genetic variation does not diminish with selection as fast as theory would anticipate [3].

“Conversion” of epistatic into additive genetic variance in finite populations and possible impact on long-term selection response

W.G. Hill


sampling and therefore has a potential indirect role in medium and long-term selection response, with superficial similarity to and hard to distinguish from mutation. Whilst predictions of response require knowledge of genetic parameters, an infinitesimal model provides some analytic results. Otherwise there is little quantitative information relevant to animal populations on which to judge this potential role of epistasis and reach firm conclusions.

FUTURE/CURRENTLY
1. More data
2. “Proper” REML
3. Impact of $h^2$
4. Role of epistasis in selection response
5. Suggestions welcome
Thank you!