



A rapid method for the identification of epistatic '*dormant*' SNPs

A. Reverter, J. Henshall, L.R. Porto-Neto, F. Raidan, Y. Li, M. Naval-Sánchez,
S. Dominik, S.A. Lehnert, K. Meyer, Z. Vitezica & A. Legarra

CSIRO AGRICULTURE & FOOD
www.csiro.au



TRADITIONAL DEFINITION

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

ALTERNATIVE DEFINITION

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 Jannink¹

Copyright © 2007 by the Genetics Society of America
DOI: 10.1534/genetics.106.062992



J. Dairy Sci. 94:1597–1600

doi:10.3168/jds.2010-3834

© American Dairy Science Association®, 2011.

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. Bennewitz*¹

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



GENETICS

July 26, 2017

Lorin Crawford^{1,2,3*}, Ping Zeng^{4,5}, Sayan Mukherjee^{6,7,8,9}, Xiang Zhou^{4,5*}

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)

- O** – Run GBLUP and Rank individuals from lowest to highest GEBV.
- S** – 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”).
- ICO** – 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 \mathbf{y} on $\mathbf{Z}_i \mathbf{u}$, which can be approximated as follows:

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

DOS – For each SNP in i :

- Multiply $\hat{\mathbf{u}}$ by centred gene content: $\mathbf{Z}_i \hat{\mathbf{u}}$
- Run a single-marker regression: $\hat{\mathbf{e}} = \mathbf{1}\mu + (\alpha\alpha)_i \mathbf{Z}_i \hat{\mathbf{u}} + \boldsymbol{\varepsilon}$
- Obtain a t -test and associated P -value for $(\hat{\alpha}\hat{\alpha})_i$

NB: This approximate method is **VERY FAST**, but ignores the uncertainty in the estimation of $\hat{\mathbf{e}}$ and $\hat{\mathbf{u}}$. 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): $\mathbf{G} = \frac{\mathbf{MM}^T}{2\sum p_i(1-p_i)}$

(4) GBLUP: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$

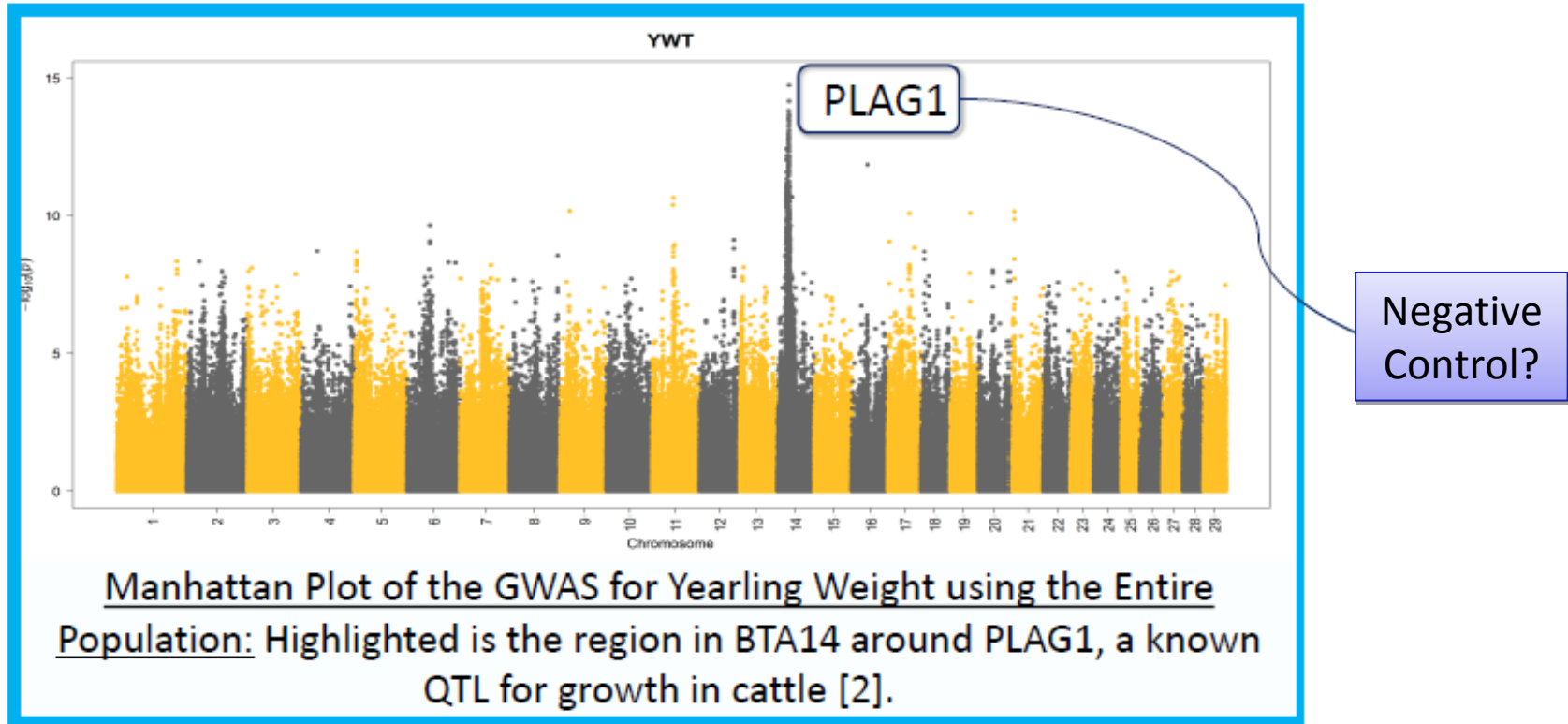
$$V(\mathbf{u}) = \mathbf{G}\sigma_u^2 \text{ and } V(\mathbf{e}) = \mathbf{I}\sigma_e^2$$

(5) GWAS: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{S}_i a_i + \mathbf{e}$

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



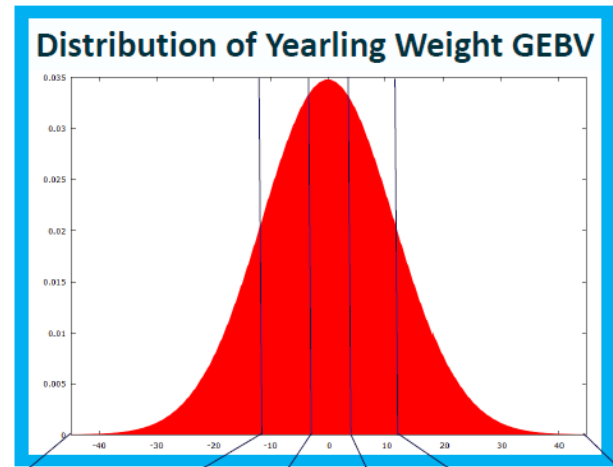
Results:























METHOD 1: Bin-based “Mechanical Heuristics”:

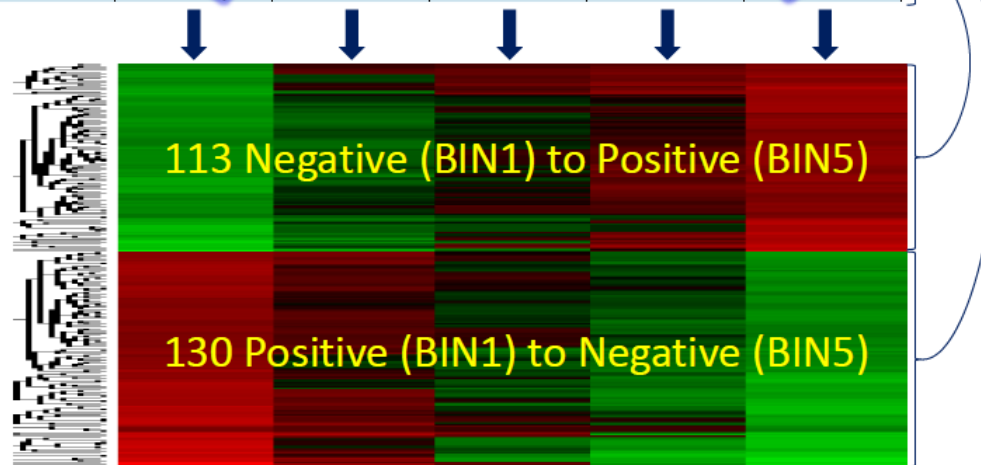
- Rank individuals from lowest to highest GEBV.
- Create 5 equally-sized bins.
- Within each bin, perform a GWAS.

PRO – Call epistatic SNP those with significant yet opposed effect in BIN1 and BIN5.



	BIN 1	BIN 2	BIN 3	BIN 4	BIN 5
N Animals	422	422	422	422	423
GEBV Range	-44.1 ; -9.3	-9.3 ; 2.4	-2.3 ; 2.8	2.8 ; 9.7	9.7 ; 41.1
Phenotype Range	157.8 ; 272.0	211.4 ; 291.0	224.9 ; 303.0	233.3 ; 323.6	250.4 ; 350.8
Look for SNP going like this					
...or for SNP going like this					

	BIN 1	BIN 2	BIN 3	BIN 4	BIN 5
N Animals	422	422	422	422	423
GEBV Range	-44.1 ; -9.3	-9.3 ; 2.4	-2.3 ; 2.8	2.8 ; 9.7	9.7 ; 41.1
Phenotype Range	157.8 ; 272.0	211.4 ; 291.0	224.9 ; 303.0	233.3 ; 323.6	250.4 ; 350.8
Look for SNP going like this					
...or for SNP going like this					



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 $P < 0.001$.

SNP chr:Mb (Gene)	BIN1	BIN2	BIN3	BIN4	BIN5	Whole
18:56.5 (CPT1C)	-7.58*	-1.38	-0.86	2.67	4.84*	0.60
28:23.3 (CTNNA3)	-8.00*	-3.05	-1.93	1.55	5.06*	-0.05
27:1.1 (CSMD1)	4.53*	1.18	-0.66	-0.69	-2.37*	0.58
5:54.9 (LRIG3)	4.86*	-0.30	-0.79	-2.05	-3.94*	1.18
14:25.0 (PLAG1)	0.74	2.07	3.46*	2.15	2.02	4.71*

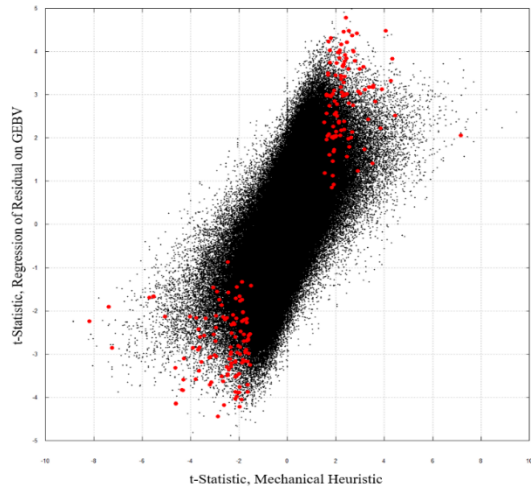
Estimated SNP effects within BINs and in the whole population

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 $P < 0.001$.

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

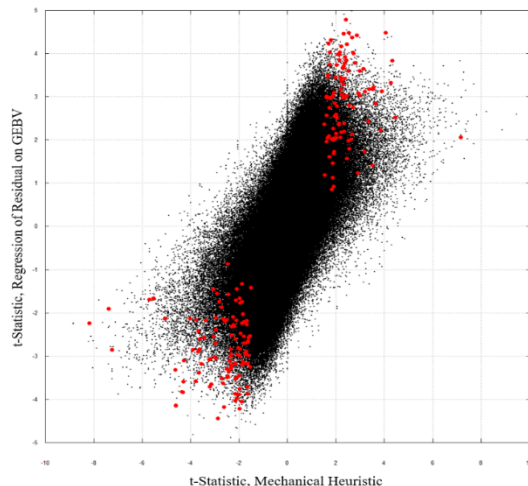
BIN1	BIN2	BIN3	BIN4	BIN5	Whole	R(e.on.u)
-7.58*	-1.38	-0.86	2.67	4.84*	0.60	4.48*
-8.00*	-3.05	-1.93	1.55	5.06*	-0.05	3.33*
4.53*	1.18	-0.66	-0.69	-2.37*	0.58	-3.24*
4.86*	-0.30	-0.79	-2.05	-3.94*	1.18	-4.44*
0.74	2.07	3.46*	2.15	2.02	4.71*	1.02

$$\frac{(t_{BIN5} - t_{BIN1})}{2} \sim Reqr(\hat{e}, \hat{u})$$



BIN1	BIN2	BIN3	BIN4	BIN5	Whole	R(e.on.u)
-7.58*	-1.38	-0.86	2.67	4.84*	0.60	4.48*
-8.00*	-3.05	-1.93	1.55	5.06*	-0.05	3.33*
4.53*	1.18	-0.66	-0.69	-2.37*	0.58	-3.24*
4.86*	-0.30	-0.79	-2.05	-3.94*	1.18	-4.44*
0.74	2.07	3.46*	2.15	2.02	4.71*	1.02

$$\frac{(t_{BIN5} - t_{BIN1})}{2} \sim Regr(\hat{e}, \hat{u})$$



Candidate Genes

SNP	Gene	Regression		
		(Bin5-Bin1)/2	t-stat	-Log(P)
BovineHD0500015637	LRIG3	-4.403	-4.439	5.345
BovineHD2000011089	PRLR	3.581	4.913	6.349
BovineHD0600010671	LAP3	-2.917	-4.460	5.388
BovineHD0200001836	MSTN	5.172	3.079	2.983
BovineHD0500005343	KITLG	-4.961	-2.991	2.857
BovineHD0700004860	INSR	2.234	3.986	4.474


A Meta-Assembly of Selection Signatures in Cattle

Imtiaz A. S. Randhawa  Mehār S. Khatkar, Peter C. Thomson, Herman W. Raadsma

Published: April 5, 2016 • <https://doi.org/10.1371/journal.pone.0153013>

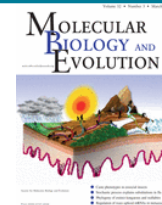


Genomic Signatures Reveal New Evidences for Selection of Important Traits in Domestic Cattle

Lingyang Xu, Derek M. Bickhart, John B. Cole, Steven G. Schroeder, Jiuzhou Song, Curtis P. Van Tassell, Tad S. Sonstegard, George E. Liu  [Author Notes](#)

Molecular Biology and Evolution, Volume 32, Issue 3, 1 March 2015, Pages 711–725,

Holstein, Angus, Charolais, Brahman, and N'Dama.




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].

ORIGINAL ARTICLE

WILEY Journal of
Animal Breeding and Genetics

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

W.G. Hill 

J Anim Breed Genet. 2017;134:196–201.

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!