Two Reasons to Model Phenotype Mean and Variance in QTL Mapping

Robert Corty

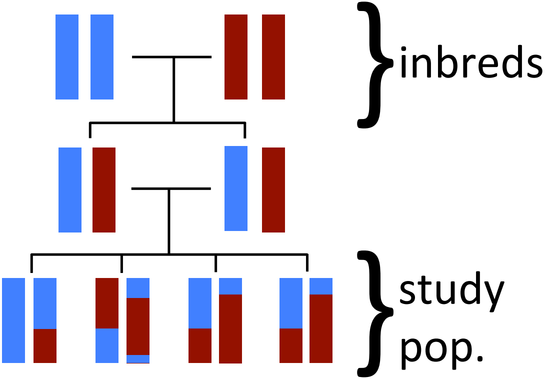
UNC Chapel Hill

Triangle Stat Gen
October 31, 2016

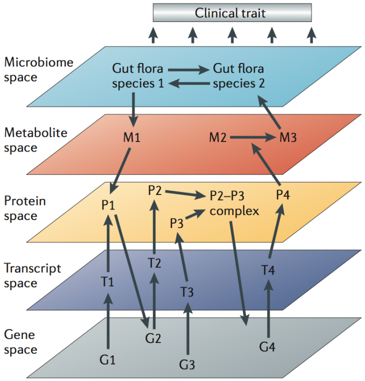
Overview

Background

F2 Intercross Mapping Population



Current Approach to QTL Mapping

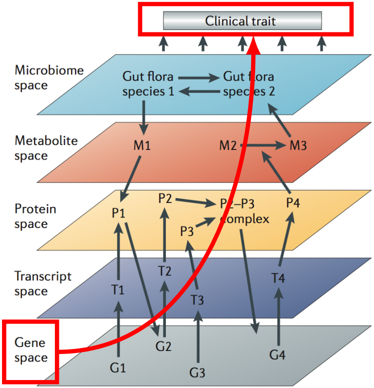


Civelek, 2014

Truth:

$$y_i = f(g_i, t_i, e_i, \ldots)$$

Current Approach to QTL Mapping



Civelek, 2014

Truth:

$$y_i = f(g_i, t_i, e_i, \ldots)$$

Statistical Model:

$$y_i = m_i + \epsilon_i$$

$$\epsilon_i \sim \mathrm{N}(0, \sigma^2)$$

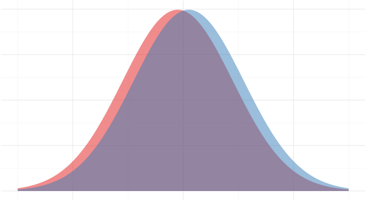
with

$$m_i = \boldsymbol{x}_i^\mathsf{T} \boldsymbol{\beta} + \boldsymbol{q}_i^\mathsf{T} \boldsymbol{\alpha}$$

DGLM Model

Constant variance
QTL Mapping Model:

$$y_i = m_i + \epsilon_i$$
$$\epsilon_i \sim \mathrm{N}(0, \sigma^2)$$
with
$$m_i = \boldsymbol{x}_i^{\mathsf{T}} \boldsymbol{\beta} + \boldsymbol{q}_i^{\mathsf{T}} \boldsymbol{\alpha}$$



DGLM Model

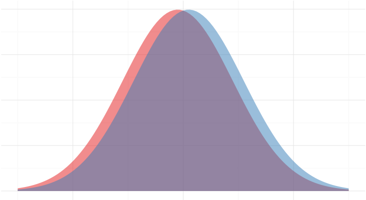
Constant variance
QTL Mapping Model:

$$y_i = m_i + \epsilon_i$$

$$\epsilon_i \sim \mathrm{N}(0, \sigma^2)$$

with

$$m_i = \boldsymbol{x}_i^\mathsf{T}\boldsymbol{\beta} + \boldsymbol{q}_i^\mathsf{T}\boldsymbol{\alpha}$$



Heterogeneous variance
QTL Mapping Model:

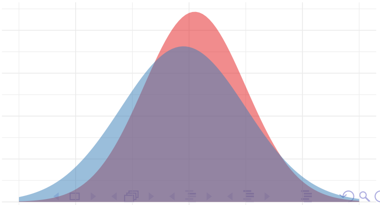
$$y_i = m_i + \epsilon_i$$

$$\epsilon_i \sim \mathrm{N}(0, \exp(v_i)^2)$$

with

$$m_i = \boldsymbol{x}_i^\mathsf{T}\boldsymbol{\beta} + \boldsymbol{q}_i^\mathsf{T}\boldsymbol{\alpha}$$

$$v_i = \boldsymbol{z}_i^\mathsf{T}\gamma + \boldsymbol{q}_i^\mathsf{T}\boldsymbol{\theta}$$



Credits

J. R. Statist. Soc. B (1989)
51, *No.* 1, *pp.* 47–60

Generalized Linear Models with Varying Dispersion

By GORDON K. SMYTH†

University of California, Santa Barbara, USA

DOI: 10.1534/genetics.111.127068

Detecting Major Genetic Loci Controlling Phenotypic Variability in Experimental Crosses

Lars Rönnegård*,[1] and William Valdar[†]

[*]Statistics Unit, Dalarna University, SE-781 70 Borlänge, Sweden and [†]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7265

Mean QTL Mapping

Nuisance Variance Heterogeneity

Could some levels of nuisance covariates yield more precise observations than others?

- Technician
- Day
- Apparatus
- Sex of model organism

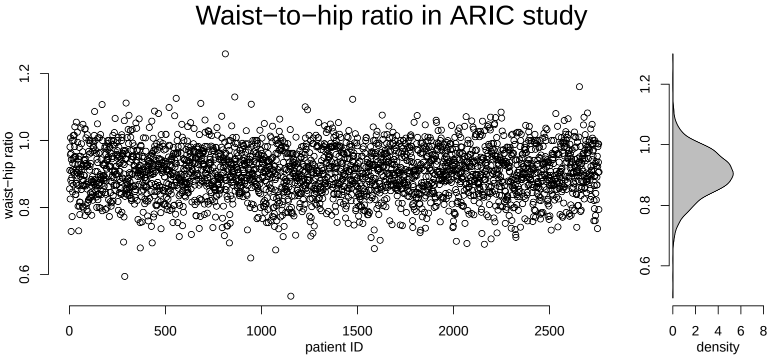
Up-weight those observations.
Down-weight the (otherwise) high leverage points.

Significance vs. Zero-ness

$p > 0.05 \not\Rightarrow \beta = 0$

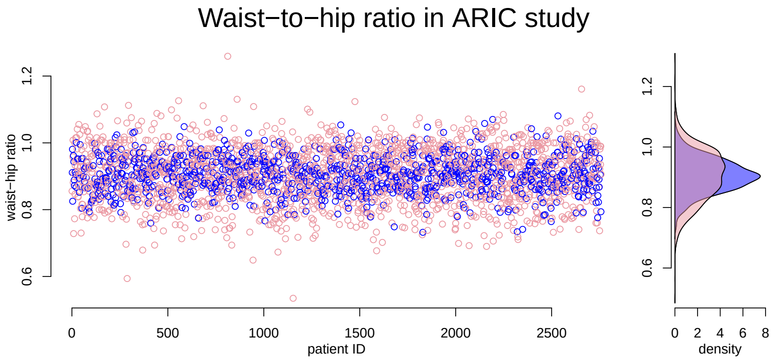
If we wouldn't trust a result that requires excluding some covariate, we should probably model it.

Example 1



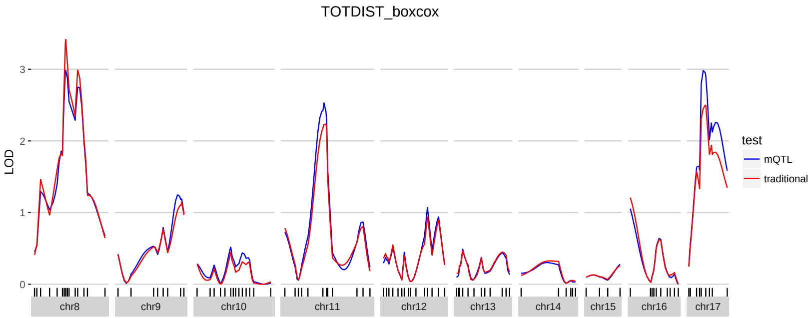
Implications for study design

Example 1



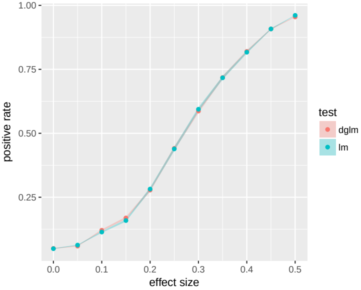
Implications for study design

Example 2

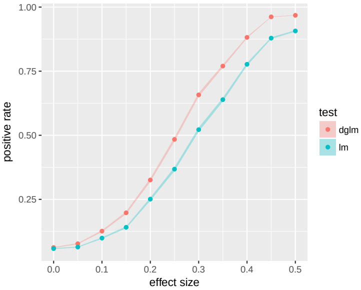

TOTDIST_boxcox
LOD
0
1
2
3
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
test
mQTL
traditional

Implications for Power

Homogenous variance scenario:

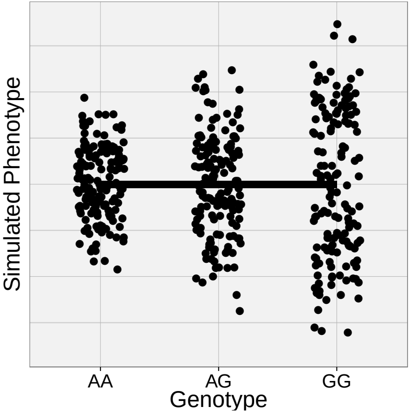


Heterogeneous variance scenario:

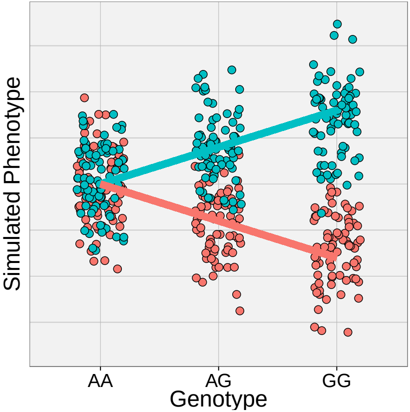


Variance QTL Mapping

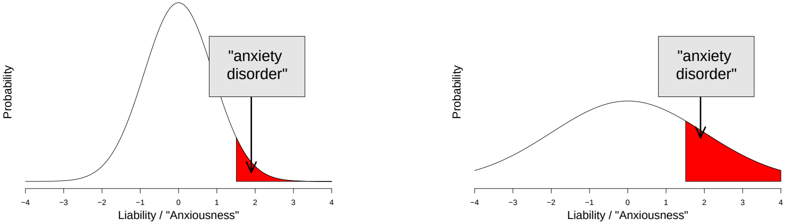
GxE



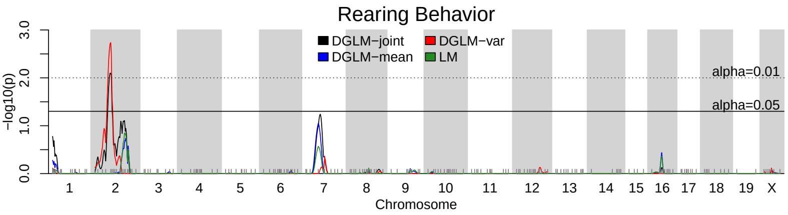
GxE



Liability-Threshold Model



Example 1


Rearing Behavior
DGLM−joint
DGLM−var
DGLM−mean
LM
alpha=0.01
alpha=0.05
−log10(p)
0.0
1.0
2.0
3.0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
X
Chromosome

Acknowledgements

Software: CRAN package vqtl
Slides: github.com/rcorty/TSG

Collaborators:
Lisa Tarantino
Ethan Lange
Leslie Lange
Laura Raffield

Valdar lab:
Will Valdar
Greg Keele
Dan Oreper
Paul Maurizio
Wes Crouse
Yanwei Cai