

# **Variance Heterogeneity in Genetic Mapping**

Robert Wallace Corty

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Computer Science.

Chapel Hill  
2018

Approved by:

Fernando Pardo Manuel de Villena, Ph.D.

James Evans, M.D., Ph.D.

Yun Li, Ph.D.

Lisa Tarantino, Ph.D.

William Valdar, Ph.D.

©2018  
Robert Wallace Corty  
ALL RIGHTS RESERVED

## **ABSTRACT**

Robert Wallace Corty: Variance Heterogeneity in Genetic Mapping  
(Under the direction of William Valdar)

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

## TABLE OF CONTENTS

LIST OF TABLES .....	vi
LIST OF FIGURES .....	vii
LIST OF ABBREVIATIONS .....	viii
1 Introduction .....	1
<b>I Exchangeable Populations</b>	<b>2</b>
2 Mean-Variance QTL Mapping on a Background of Variance Heterogeneity .....	3
3 Mean-Variance QTL Mapping Identifies Novel QTL for Circadian Activity and Exploratory Behavior in Mice .....	5
4 vqtl: An R package for Mean-Variance QTL Mapping .....	7
<b>II Non-Exchangeable Populations</b>	<b>9</b>
5 Genetic Mapping in Human Cohorts with the Double Generalized Linear Model .....	10
6 The Heteroscedastic Linear Mixed Model .....	11
6.1 A simpler, but equivalent, parametrization .....	13
6.2 Given $h^2$ , the LMM Reduces to Multiple Linear Regression .....	13
6.3 Rotation to Independence .....	14
6.4 Rotation for the Homoscedastic Model .....	15
6.5 Rotation for the Heteroscedastic Model .....	15
6.6 Simulation Results .....	17
6.7 Software .....	17

7	Conclusion and Future Directions .....	18
7.1	QTL Mapping with Causally Ambiguous Covariates .....	18
	BIBLIOGRAPHY .....	20

## **LIST OF TABLES**

## **LIST OF FIGURES**

## LIST OF ABBREVIATIONS

QTL	quantitative trait locus
mQTL	mean-controlling quantitative trait loci
vQTL	variance-controlling quantitative trait loci
mvQTL	mean or variance controlling quantitative trait loci
LMM	Linear Mixed Model
GLS	Generalized Least Squares
SLM	Standard Linear Model
DGLM	Double Generalized Linear Model
LRT	Likelihood Ratio Test
ML	Maximum Likelihood
AYO	Add Your Own in alphabetic order. . .



## CHAPTER 1

# Introduction

Describe exchangeable and non-exchangeable populations

In an exchangeable pop, knowing

	accommodate variance hetero- geneity	detect variance heterogeneity	handle differen- tial relatedness
LM	no	no	no
DGLM	yes	yes	no
LMM	no	no	yes
wLMM	yes	no	yes

## **Part I**

# **Exchangeable Populations**

## CHAPTER 2

# Mean-Variance QTL Mapping on a Background of Variance Heterogeneity

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec

bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

## CHAPTER 3

# Mean-Variance QTL Mapping Identifies Novel QTL for Circadian Activity and Exploratory Behavior in Mice

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim.

Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

## CHAPTER 4

# **vqtl: An R package for Mean-Variance QTL Mapping**

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec

bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.



## **Part II**

# **Non-Exchangeable Populations**

## CHAPTER 5

# Genetic Mapping in Human Cohorts with the Double Generalized Linear Model

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

## CHAPTER 6

# The Heteroscedastic Linear Mixed Model

Chapter 2 described a statistical approach that stands on the assumption that all individuals in the mapping population are equally related. This assumption is appropriate in the analysis of experimental crosses because it is nearly true. It has been shown that a more general approach that accounts for the slight differential relatedness in an F2 intercross delivers little or no additional benefit in terms of mapping power or precision (Parker et al., 2014). Chapter 5 described a statistical approach to QTL mapping that accounts for differential relatedness in the mapping population by including principle components of the genetic variation as regressors. This approach is heuristic, correcting for large-scale relationships, but not finer grained ones. This chapter describes the linear mixed model (LMM), which can accommodate any pattern of differential relatedness a mapping population may have, at both large and fine scales.

The LMM models an observed phenotype,  $y$ , as,

$$y = \mathbf{1}\mu + \mathbf{X}\beta + \mathbf{G}\alpha + \mathbf{a} + \mathbf{e} \quad (6.1)$$

where  $\mathbf{1}$  is a column vector of ones,  $\mathbf{X}$  is the design matrix of covariates,  $\mathbf{G}$  is the design matrix of focal genetic effects, and  $\mu$ ,  $\beta$ , and  $\alpha$  are unconstrained parameters that can be referred to as the population mean, the effect(s) of the covariate(s), and the effect(s) of the focal genetic factor(s), respectively.

$\mathbf{a}$  and  $\mathbf{e}$  are so-called “random effects”, estimated in the process of model fitting, but with constraints. Specifically, they are modeled hierarchically as

$$\mathbf{a} \sim N(0, \mathbf{K}\tau^2), \quad (6.2)$$

$$\mathbf{e} \sim N(0, \mathbf{D}\sigma^2) \quad (6.3)$$

where  $\mathbf{K}$  is a known, positive semi-definite genomic similarity matrix, and  $\mathbf{D}$  is a known diagonal residual variance matrix. The scale parameters,  $\tau^2$  and  $\sigma^2$ , are constrained only to be non-negative.

In the context of genetic mapping, the this model is fit to each putative QTL, using  $\mathbf{G}$  to encode the locus design matrix and testing whether  $\alpha = \mathbf{0}$ . If  $\alpha \neq \mathbf{0}$ , the locus is a QTL.

Two statistical tests used to determine whether  $\alpha = \mathbf{0}$ , the  $t$  test and the likelihood ratio test (LRT), require computation of the values of the parameters that maximize the likelihood of the data. Henderson (1984) described a suite of procedures for computing these maximum likelihood parameter values in a variety of situations.

However, Henderson’s methods are of limited use in QTL mapping. His focus was on estimation of breeding values ( $\mathbf{a}$  in Equation 6.1) and therefore the model only needed to be fit once, to one design matrix, and therefore speed was not a primary concern. Therefore, Henderson’s methods are not of great use in the context of QTL mapping, where each putative QTL demands its own maximum likelihood model fit.

## 6.1 A simpler, but equivalent, parametrization

The LMM as specified in Equation 6.1 is equivalent to:

$$\mathbf{y} \sim \mathcal{N}(\mathbf{X}_c \boldsymbol{\beta}_c, \boldsymbol{\Sigma}) \quad (6.4)$$

where fixed effects design matrices are collapsed into  $\mathbf{X}$  and the variance-covariance matrices of the random effects are combined into  $\boldsymbol{\Sigma}$ . Specifically,

$$\mathbf{X}_c = [1 \ \mathbf{X} \ \mathbf{G}] \quad (6.5)$$

$$\boldsymbol{\beta}_c = [\mu \ \boldsymbol{\beta}^T \boldsymbol{\alpha}^T]^T \quad (6.6)$$

$$\boldsymbol{\Sigma} = \mathbf{K}\tau^2 + \mathbf{D}\sigma^2 \quad (6.7)$$

Going forward, we refer to  $\mathbf{X}_c$  and  $\boldsymbol{\beta}_c$  as simply  $\mathbf{X}$  and  $\boldsymbol{\beta}$  for simplicity.

This model can be re-parametrized to directly use the well-known narrow sense heritability, increasing its interpretability. Specifically,  $h^2 = \frac{\tau^2}{\tau^2 + \sigma^2}$  and  $\lambda = \tau^2 + \sigma^2$ .

$$\boldsymbol{\Sigma} = \left( \mathbf{K} \frac{\tau^2}{\tau^2 + \sigma^2} + \mathbf{D} \frac{\sigma^2}{\tau^2 + \sigma^2} \right) (\tau^2 + \sigma^2) \quad (6.8)$$

$$= \mathbf{K}h^2 + \mathbf{D}(1 - h^2)\lambda \quad (6.9)$$

## 6.2 Given $h^2$ , the LMM Reduces to Multiple Linear Regression

[transition] Thus, for any given value of  $h^2$ , we can define  $\mathbf{V} = \mathbf{K}h^2 + \mathbf{D}(1 - h^2)$ , and the LMM reduces to a standard linear model (SLM):

$$\mathbf{y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}\lambda) \quad (6.10)$$

with log likelihood:

$$\ell(\boldsymbol{\beta}, \lambda; \mathbf{y}, \mathbf{X}, \mathbf{V}) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{V}\lambda| - \frac{1}{2\lambda} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (6.11)$$

$$= -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{V}| - \frac{n}{2} \log \lambda - \frac{1}{2\lambda} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (6.12)$$

And there is a known, closed-form, maximum likelihood estimator for its two parameters:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y} \quad (6.13)$$

$$\hat{\lambda} = \left\| (\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y})^T (\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y}) \right\|_2 \quad (6.14)$$

In the context of a genome scan, we want to calculate these maximum likelihood estimators for a single phenotype,  $\mathbf{y}$ , genomic similarity,  $\mathbf{K}$  residual variance,  $\mathbf{D}$ , and many different values of  $\mathbf{X}$  — recall that the genetic variant(s) being investigated are in  $\mathbf{G}$ . Also recall that, though we have described a parameterization that recapitulates a standard linear regression problem when  $h^2$  is known, we do not, in general know the true value of  $h^2$ , so we will additionally want to optimize over all possible values of  $h^2 \in [0, 1]$ .

We now describe a matrix algebra trick that allows us to increase the up-front computational cost of fitting the linear regression model Equation 6.10, but drastically decreases the computational cost for each new value of  $\mathbf{X}$ .

### 6.3 Rotation to Independence

Imagine we knew of a multiplier matrix,  $\mathbf{M}$  such that

$$\mathbf{M}^T \mathbf{M} = \mathbf{V}^{-1} \quad (6.15)$$

Let us now consider a new, rotated phenotype vector,  $\mathbf{y}_r = \mathbf{M}\mathbf{y}$ , a new covariate matrix,  $\mathbf{X}_r = \mathbf{M}\mathbf{X}$ , and a new SLM.

$$\mathbf{y}_r \sim \mathcal{N}(\mathbf{X}_r \boldsymbol{\beta}_r, \mathbf{I} \lambda_r) \quad (6.16)$$

The log likelihood of this SLM is:

$$\ell(\boldsymbol{\beta}_r, \lambda_r; \mathbf{X}_r, \mathbf{y}_r) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{I}| - \frac{n}{2} \log \lambda_r - \frac{1}{2\lambda_r} (\mathbf{y}_r - \mathbf{X}_r \boldsymbol{\beta}_r)^T (\mathbf{y}_r - \mathbf{X}_r \boldsymbol{\beta}_r) \quad (6.17)$$

$$= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \lambda_r - \frac{1}{2\lambda_r} (\mathbf{M}\mathbf{y} - \mathbf{M}\mathbf{X}\boldsymbol{\beta})^T (\mathbf{M}\mathbf{y} - \mathbf{M}\mathbf{X}\boldsymbol{\beta}) \quad (6.18)$$

$$= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \lambda_r - \frac{1}{2\lambda_r} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{M}^T \mathbf{M} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (6.19)$$

$$= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \lambda_r - \frac{1}{2\lambda_r} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (6.20)$$

and its ML estimators are:

$$\widehat{\boldsymbol{\beta}}_r = (\mathbf{X}_r^T \mathbf{X}_r)^{-1} \mathbf{X}_r^T \mathbf{y}_r \quad (6.21)$$

$$\widehat{\lambda}_r = \left\| (\mathbf{X}_r \widehat{\boldsymbol{\beta}}_r - \mathbf{y}_r)^T (\mathbf{X}_r \widehat{\boldsymbol{\beta}}_r - \mathbf{y}_r) \right\|_2 \quad (6.22)$$

Thus we have established that, given  $\mathbf{M}$  we can fit the model

## 6.4 Rotation for the Homoscedastic Model

[D = I] Kang proposed the following [cite EMMA].

## 6.5 Rotation for the Heteroscedastic Model

The above  $\mathbf{M}$  relied on [point out where D=I is required] [in step X]. Generally, however, some phenotypes are known with more certainty than others. Want to have a diagonal D, but not necessarily I.

### 6.5.1 Proposal

$$\mathbf{M} = (\boldsymbol{\Lambda}_L + \delta \mathbf{I})^{-\frac{1}{2}} \mathbf{U}_L^T \mathbf{D}^{-\frac{1}{2}} \quad (6.23)$$

where

$$\mathbf{L} = \mathbf{D}^{-\frac{1}{2}} \mathbf{K} \mathbf{D}^{-\frac{1}{2}} \quad (6.24)$$

and

$$\mathbf{L} = \mathbf{U}_L \mathbf{\Lambda}_L \mathbf{U}_L^T \quad (6.25)$$

is its eigen decomposition

### 6.5.2 Validity

To be a valid multiplier matrix,  $\mathbf{M}$  must have the property:

$$\mathbf{M}^T \mathbf{M} = \mathbf{V}^{-1} \quad (6.26)$$

*Proof.* First, derive a useful form of  $\mathbf{V}$ .

$$\mathbf{V} = \mathbf{K} + \delta \mathbf{D} \quad \text{definition} \quad (6.27)$$

$$= \mathbf{D}^{\frac{1}{2}} \mathbf{D}^{-\frac{1}{2}} (\mathbf{K} + \delta \mathbf{D}) \quad \text{pre-multiply by } \mathbf{D}^{\frac{1}{2}} \mathbf{D}^{-\frac{1}{2}} = \mathbf{I} \quad (6.28)$$

$$= \mathbf{D}^{\frac{1}{2}} \mathbf{D}^{-\frac{1}{2}} (\mathbf{K} + \delta \mathbf{D}) \mathbf{D}^{-\frac{1}{2}} \mathbf{D}^{\frac{1}{2}} \quad \text{post-multiply by } \mathbf{D}^{-\frac{1}{2}} \mathbf{D}^{\frac{1}{2}} = \mathbf{I} \quad (6.29)$$

$$= \mathbf{D}^{\frac{1}{2}} (\mathbf{D}^{-\frac{1}{2}} \mathbf{K} \mathbf{D}^{-\frac{1}{2}} + \delta \mathbf{D}^{-\frac{1}{2}} \mathbf{D} \mathbf{D}^{-\frac{1}{2}}) \mathbf{D}^{\frac{1}{2}} \quad \text{distribute } \mathbf{D}^{-\frac{1}{2}} \text{ in} \quad (6.30)$$

$$= \mathbf{D}^{\frac{1}{2}} (\mathbf{D}^{-\frac{1}{2}} \mathbf{K} \mathbf{D}^{-\frac{1}{2}} + \delta \mathbf{I}) \mathbf{D}^{\frac{1}{2}} \quad \text{definition of root inverse} \quad (6.31)$$

$$= \mathbf{D}^{\frac{1}{2}} (\mathbf{L} + \delta \mathbf{I}) \mathbf{D}^{\frac{1}{2}} \quad \text{define: } \mathbf{L} = \mathbf{D}^{-\frac{1}{2}} \mathbf{K} \mathbf{D}^{-\frac{1}{2}} \quad (6.32)$$

$$= \mathbf{D}^{\frac{1}{2}} (\mathbf{U}_L \mathbf{\Lambda}_L \mathbf{U}_L^T + \delta \mathbf{I}) \mathbf{D}^{\frac{1}{2}} \quad \text{eigen decomposition of } \mathbf{L} \quad (6.33)$$

$$= \mathbf{D}^{\frac{1}{2}} (\mathbf{U}_L \mathbf{\Lambda}_L \mathbf{U}_L^T + \delta \mathbf{U}_L \mathbf{U}_L^T) \mathbf{D}^{\frac{1}{2}} \quad \text{property of eigen vectors} \quad (6.34)$$

$$= \mathbf{D}^{\frac{1}{2}} \mathbf{U}_L (\mathbf{\Lambda}_L + \delta \mathbf{I}) \mathbf{U}_L^T \mathbf{D}^{\frac{1}{2}} \quad \text{distributive property} \quad (6.35)$$



and invert it

$$\mathbf{V}^{-1} = \left( \mathbf{D}^{\frac{1}{2}} \mathbf{U}_L (\mathbf{\Lambda}_L + \delta \mathbf{I}) \mathbf{U}_L^T \mathbf{D}^{\frac{1}{2}} \right)^{-1} \quad \text{definition} \quad (6.36)$$

$$= \left( \mathbf{D}^{\frac{1}{2}} \right)^{-1} \left( \mathbf{U}_L (\mathbf{\Lambda}_L + \delta \mathbf{I}) \mathbf{U}_L^T \right)^{-1} \left( \mathbf{D}^{\frac{1}{2}} \right)^{-1} \quad \text{inverse of product} \quad (6.37)$$

$$= \mathbf{D}^{-\frac{1}{2}} \left( \mathbf{U}_L (\mathbf{\Lambda}_L + \delta \mathbf{I}) \mathbf{U}_L^T \right)^{-1} \mathbf{D}^{-\frac{1}{2}} \quad \text{inverse of diagonal matrix} \quad (6.38)$$

$$= \mathbf{D}^{-\frac{1}{2}} \mathbf{U}_L (\mathbf{\Lambda}_L + \delta \mathbf{I})^{-1} \mathbf{U}_L^T \mathbf{D}^{-\frac{1}{2}} \quad \text{inverse of eigen decomposition} \quad (6.39)$$

and compare to  $\mathbf{M}^T \mathbf{M}$

$$\mathbf{M}^T \mathbf{M} = \left( (\mathbf{\Lambda}_L + \delta \mathbf{I})^{-\frac{1}{2}} \mathbf{U}_L^T \mathbf{D}^{-\frac{1}{2}} \right)^T \left( (\mathbf{\Lambda}_L + \delta \mathbf{I})^{-\frac{1}{2}} \mathbf{U}_L^T \mathbf{D}^{-\frac{1}{2}} \right) \quad \text{definition} \quad (6.40)$$

$$= \left( \mathbf{D}^{-\frac{1}{2}} \mathbf{U}_L (\mathbf{\Lambda}_L + \delta \mathbf{I})^{-\frac{1}{2}} \right) \left( (\mathbf{\Lambda}_L + \delta \mathbf{I})^{-\frac{1}{2}} \mathbf{U}_L^T \mathbf{D}^{-\frac{1}{2}} \right) \quad \text{transpose of product} \quad (6.41)$$

$$= \mathbf{D}^{-\frac{1}{2}} \mathbf{U}_L \left( (\mathbf{\Lambda}_L + \delta \mathbf{I})^{-\frac{1}{2}} (\mathbf{\Lambda}_L + \delta \mathbf{I})^{-\frac{1}{2}} \right) \mathbf{U}_L^T \mathbf{D}^{-\frac{1}{2}} \quad \text{associative property} \quad (6.42)$$

$$= \mathbf{D}^{-\frac{1}{2}} \mathbf{U}_L (\mathbf{\Lambda}_L + \delta \mathbf{I})^{-1} \mathbf{U}_L^T \mathbf{D}^{-\frac{1}{2}} \quad \text{definition of root inverse} \quad (6.43)$$

□

### 6.5.3 Calculation of Necessary Quantities

#### 6.5.3.1 M itself

#### 6.5.3.2 something else

#### 6.5.3.3 Its determinant

## 6.6 Simulation Results

## 6.7 Software

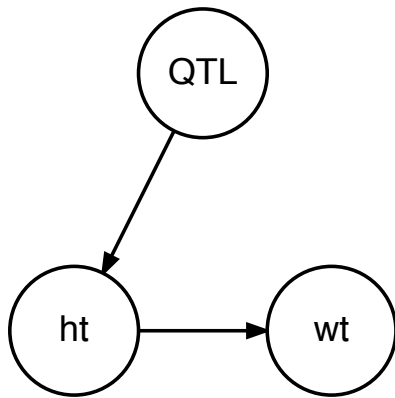
## CHAPTER 7

# Conclusion and Future Directions

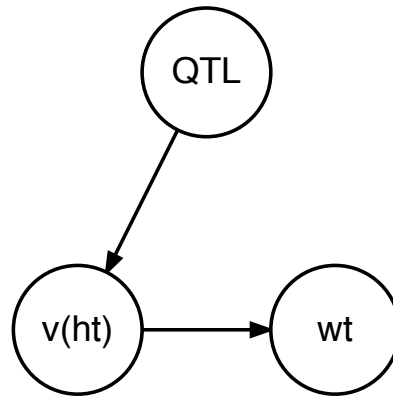
### 7.1 QTL Mapping with Causally Ambiguous Covariates

In chapter 2, when using B as a mean or variance covariate for mapping A, we always assumed A was causally downstream of B, that no information flowed from B to A. For example, no phenotype influences which batch, housing, or sex a mouse is. In this setting, the value of including these covariates was clear.

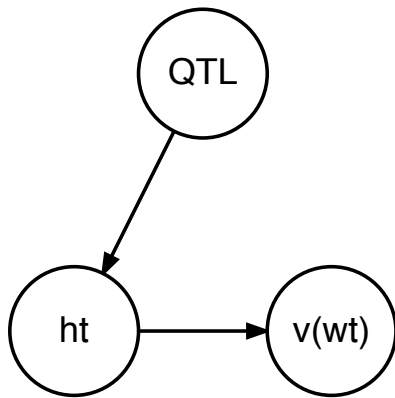
Now, let's consider cases where the flow of causality is ambiguous. What is the value in using B as a mean or variance covariate for mQTL and vQTL mapping? WV and RC agree that using it as a variance covariate for mQTL mapping is probably a good thing – gives weights that are more reflective of the true extent of residual variation. But what about the other 3 possibilities? There should be loads of literature on using it as a mean covariate for mQTL mapping – trying to find some of that will be the next step for this section.



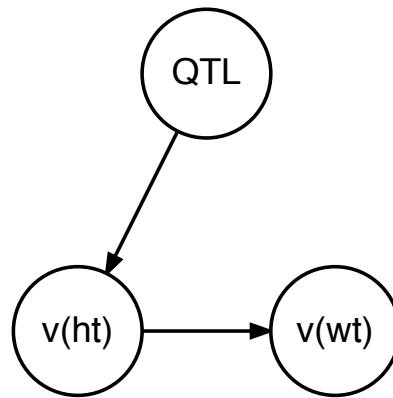
(a) Model MM: the QTL influences the mean of height and height influences the mean of weight.



(b) Model VM: the QTL influences the variance of height and variance of height influences the mean of weight.



(c) Model MV: the QTL influences the mean of height and height influences the variance of weight.



(d) Model VV: the QTL influences the variance of height and variance of height influences the variance of weight.

## BIBLIOGRAPHY

- Henderson, C. R. (1984). Applications of Linear Models in Animal Breeding Models. *Univ. Guelph*, page 384.
- Parker, C. C., Carbonetto, P., Sokoloff, G., Park, Y. J., Abney, M., and Palmer, A. A. (2014). High-Resolution genetic mapping of complex traits from a combined analysis of F2 and advanced intercross mice. *Genetics*, 198(1):103–116.