ABSTRACT

OGOREK, BENJAMIN ALEXANDER. Orthology-Based Multilevel Modeling of Differentially Expressed Mouse and Human Gene Pairs. (Under the direction of Professor Leonard Allen Stefanski.)

There is great interest in finding human genes expressed through pharmaceutical intervention, thus opening a genomic window into benefit and side-effect profiles of a drug. Human insight gained from FDA-required animal experiments has historically been limited, but in the case of gene expression measurements, proposed biological orthologies between mouse and human genes provide a foothold for animal-to-human extrapolation. We have investigated a five-component, multilevel, bivariate normal mixture model that incorporates mouse, as well as human, gene expression data. The goal is two-fold: to increase human differential gene-finding power; and to find a subclass of gene pairs for which there is a direct exploitable relationship between animal and human genes. In simulation studies, the dual-species model boasted impressive gains in differential gene-finding power over a related marginal model using only human data. Bias in parameter estimation was problematic, however, and occasionally led to failures in control of the false discovery rate. Though it was considerably more difficult to find species-extrapolative gene-pairs (than differentially-expressed human genes), simulation experiments deemed it to be possible, especially when traditional FDR controls are relaxed and under hypothetical parameter configurations.

Orthology-Based Multilevel Modeling of Differentially Expressed Mouse and Human
Gene Pairs

by
Benjamin Alexander Ogorek

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial satisfaction of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina

2008

APPROVED BY:

Dr. Jason Osborne

Mr. Alan Menius

Dr. Leonard Stefanski
Chair of Advisory Committee

Dr. Jackie Hughes-Oliver

Dr. David Dickey

DEDICATION

To “Pop” ...

BIOGRAPHY

After graduating from Butler High School in Charlotte, NC, Benjamin enrolled at North Carolina State University in the Fall of 1999. Owing to the outcome of a high school AP Statistics course, he was invited by Dr. William Swallow into the Statistics 101 Honors section at North Carolina State University. Within a semester, Ben declared Statistics as a major. He received a Bachelor's Degree in Statistics in 2003 and, exploiting a departmental option to complete first-year graduate courses as a senior, graduated with a Master's of Statistics in 2004. At this time, Ben decided to remain at North Carolina State to pursue a doctorate. From the summer of 2005 to 2008, he enjoyed a Graduate Industrial Traineeship (GIT) at GlaxoSmithKline, where under the guidance of his thesis advisor, Dr. Leonard Stefanski, and GSK's Global Director of Discovery Analytics, Alan Menius, he completed his doctoral research on the simultaneous analysis of differential gene expression in mice and humans. Following his defense in August 2008, Ben is pursuing a career in industry.

ACKNOWLEDGEMENTS

- **Lei Zhu**, for statistical advice and introduction to the “Shrinkage T” methodology;
- **Kwan Li**, for many helpful suggestions and an emphasis on practicality;
- **Yvonne Napier**, for inspiration and help throughout daily life at GSK;
- **Gary Merrill**, for guidance, strategy, perspective, and technical expertise;
- **Natalie Doe**, for technical help;
- **Jeffery Painter**, for help with L^AT_EX and programming in general;
- **Terry Speed**, for his generous time explaining his perspectives on microarray data transformation;
- **Adrian Blue**, who helped me schedule my preliminary exam under extraordinary circumstances;
- **Dr. Dickey**, for his insightful comments, fresh perspective, and meticulous analysis of this work. He is also responsible for teaching me a few things.
- **Dr. Hughes-Oliver**, who originally suggested computing the species extrapolation probabilities in Section 2.3.2;
- **Dr. Osborne**, for lending his particular expertise and insight into the analysis of microarray data;
- **Alan Menius**, for inspiring this research and help with writing, presentation skills, and focus;
- **Mrs. Ziemba**, for introducing me to Statistics and sparking my interest in further studies.
- **Dr. William Swallow**, who brought me onboard at NC State, and who has been a constant encouragement throughout my academic career.
- My Advisor, **Dr. Leonard Stefanski**, for guiding this research, helping to make me a more disciplined statistical researcher and writer, and for the countless amounts of knowledge I have absorbed.

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	xi
1 Background	1
1.1 Introduction	1
1.2 Multiple Testing and the Search for Differentially Expressed Genes	4
1.2.1 Frequentist Multiple Testing	4
1.2.2 Bayesian Multiple Testing	7
1.3 Multi-Species Genomic Data and Mammalian Orthology	11
1.4 GlaxoSmithKline’s Integrative Biology Initiative	12
1.4.1 Parallel Analysis of Mouse and Human Experiments	13
1.4.2 Description of the Data	15
1.5 Joint Modeling	16
2 Joint Multilevel Modeling and Species Extrapolation	19
2.1 The I-Ortholog Hypothesis	20
2.2 Latent Modeling	21
2.2.1 Latent Model Assumptions	22
2.2.2 Joint Model for Latent Treatment Effects	24
2.3 Estimated Treatment Effects	26
2.3.1 Distribution of Binary Latent Variables	28
2.3.2 The Species Extrapolation	29
2.4 Estimation	30
2.4.1 Computational Issues	32
2.4.2 Estimation Summary	34
2.5 Decision Rules for Active Genes and Orthologs	35
2.6 An Example Application	36
2.6.1 Discovery and Methodology Comparisons	39
3 Simulation	42
3.1 A Description of the Simulated Data	43
3.1.1 Factors	47
3.2 Simulation Implementation	48
3.2.1 Generation of (β_{Mi}, β_{Hi})	48
3.2.2 Generating (O_i, A_{Mi}, A_{Hi})	50
3.2.3 Generating σ_{Mi} and σ_{Hi}	50
3.2.4 Generating α_{Mi}, α_{Hi} and $\gamma_{1i} - \gamma_{5i}$	52
3.2.5 Setting the $n \times 5$ human covariate matrix $\mathbf{Z}_i$, $i = 1, \dots, N$	52
3.2.6 Generating $\mathbf{Y}_H$ and $\mathbf{Y}_M$	53
3.3 Simulation Experimental Design	53
3.3.1 Infinite Sample Size Experiment	53
3.3.2 Finite Sample Size Experiment I	54
3.3.3 Finite Sample Size Experiment II	55

3.3.4	Finite Sample Size Experiment III	55
3.4	Simulation Study Results	56
3.4.1	Infinite Sample Size Experiment	56
3.4.2	Outliers in the Finite Sample Size Experiments	58
3.4.3	Finding Differentially Expressed Human Genes	59
3.4.4	Finding I-Orthologs the Finite Sample Size Experiments	62
3.4.5	Estimation in the Finite Sample Size Experiments	63
4	Discussion and Conclusions	70
4.1	Differential Gene Finding	70
4.2	Bias	71
4.3	I-Orthologs and Species Extrapolation	72
4.4	Mixture Modeling	72
4.5	Reflections and Future Outlook	74
	Bibliography	75
	Appendices	80
	Appendix A Key Model Derivations	81
	A.1 Derivation of the Distribution of $(\hat{\beta}_{Mi}, \hat{\beta}_{Hi})^T$	81
	A.2 Derivation of $\Pr(A_{Hi} = 1)$ and $\Pr(O_i = 1)$	84
	Appendix B Identifiability	86
	B.1 Estimability of Gene Level Parameters	87
	B.2 Asymptotic Estimability of Marginal Parameters	88
	B.3 Asymptotic Estimability of Dependence Parameters	90
	Appendix C Microarray Data Transformation	94
	Appendix D Conditional Expectations of Treatment Coefficients	99
	Appendix E Applications of the EM Algorithm	104
	E.1 A Custom EM Algorithm	104
	E.2 Performance	107
	Appendix F Gene Pooling Based Variance Estimation	111
	F.1 Inverse Gamma Priors	112
	F.2 Lognormal Priors	113
	F.3 Nonparametric Shrinkage	114
	F.4 Discussion	115
	Appendix G Simulation Figures and Tables	117

LIST OF TABLES

Table 1.1	The frequentist multiple testing framework.	5
Table 2.1	Distributional Models for $(\beta_{Mi}, \beta_{Hi} O_i, A_{Hi}, A_{Mi})$ and (O_i, A_{Hi}, A_{Mi})	25
Table 2.2	Estimates and Wald Confidence intervals for model (2.18).....	37
Table 2.3	Estimates and Wald Confidence intervals for marginal models (2.20) and (2.21).....	38
Table 2.4	Numbers of Discovered Genes for the various methodologies, where “Joint” refers to key model (2.18), “Marginal” refers to human marginal model (2.21), and “BH” refers to the method of Benjamini and Hochberg (1995).	40
Table 3.1	The following are $\hat{\sigma}_{Hi}^{-2}$ -weighted means of $\hat{\gamma}_{1i} - \hat{\gamma}_{5i}$ from the GlaxoSmithKline data, used to set the simulation parameters $\gamma_{1i} - \gamma_{5i}, i = 1, \dots, N$	47
Table 3.2	From FSSE I, the factorial makeup of the 74 out of 8059 experimental runs flagged as outliers when $\hat{\sigma}_H < 3$	59
Table 3.3	From FSSE II, the factorial makeup of the 524 out of 8000 experimental runs flagged as outliers when $\hat{\sigma}_H < 2.55$	60
Table 3.4	From FSSE III, the factorial makeup of the 8 out of 2120 experimental runs flagged as outliers when $\hat{\sigma}_H < 2.4$	60
Table A.1	The distribution of $(T_{1i}, T_{2i}, T_{3i}, T_{4i}, T_{5i})$	82
Table C.1	Human expression data from the gene SCEL with and without log transformation.	96
Table C.2	Human expression data from the gene RPS23 with and without the log transformation.	96
Table E.1	The relative performance of the EM algorithm to the Quasi-Newton algorithm for the first eight (of sixteen) starting value sets.	109
Table E.2	The relative performance of the EM algorithm to the Quasi-Newton algorithm for the second eight (of sixteen) starting value sets.	110

Table G.1	From the 504 replicates of the Finite Sample Size Experiment I (FSSE I), this table contains ratios of the mean information matrix standard error estimate to the Monte Carlo estimate of standard error, based on 504 replicates. Delete-1 Jackknife standard error estimates for the ratios are included in parentheses....	124
Table G.2	From the 500 replicates of the Finite Sample Size Experiment II (FSSE II), this table contains ratios of the mean information matrix standard error estimate to the Monte Carlo estimate of standard error. Delete-1 Jackknife standard error estimates for the ratios are included in parentheses.....	125
Table G.3	An estimation summary of $\hat{\mu}_M$ from the 504 replicates of Finite Sample Size Experiment I (FSSE I).....	126
Table G.4	An estimation summary of $\hat{\mu}_H$ from the 504 replicates of Finite Sample Size Experiment I (FSSE I).....	127
Table G.5	An estimation summary of $\hat{\sigma}_M$ from the 504 replicates of Finite Sample Size Experiment I (FSSE I).....	128
Table G.6	An estimation summary of $\hat{\sigma}_H$ from the 504 replicates of Finite Sample Size Experiment I (FSSE I).....	129
Table G.7	An estimation summary of $\hat{\rho}$ from the 504 replicates of Finite Sample Size Experiment I (FSSE I).....	130
Table G.8	An estimation summary of $\hat{\eta}$ from the 504 replicates of Finite Sample Size Experiment I (FSSE I).....	131
Table G.9	An estimation summary of $\hat{\theta}$ from the 504 replicates of Finite Sample Size Experiment I (FSSE I).....	132
Table G.10	An estimation summary of $\hat{\mu}_M$ from the 500 replicates of Finite Sample Size Experiment II (FSSE II).....	133
Table G.11	An estimation summary of $\hat{\mu}_H$ from the 500 replicates of Finite Sample Size Experiment II (FSSE II).....	134
Table G.12	An estimation summary of $\hat{\sigma}_M$ from the 500 replicates of Finite Sample Size Experiment II (FSSE II).....	135
Table G.13	An estimation summary of $\hat{\sigma}_H$ from the 500 replicates of Finite Sample Size Experiment II (FSSE II).....	136
Table G.14	An estimation summary of $\hat{\rho}$ from the 500 replicates of Finite Sample Size Experiment II (FSSE II).....	137

Table G.15	An estimation summary of $\hat{\eta}$ from the 500 replicates of Finite Sample Size Experiment II (FSSE II).	138
Table G.16	An estimation summary of $\hat{\theta}$ from the 500 replicates of Finite Sample Size Experiment II (FSSE II).	139
Table G.17	An estimation summary of $\hat{\mu}_M$ from the 530 replicates of Finite Sample Size Experiment III (FSSE III).	140
Table G.18	An estimation summary of $\hat{\mu}_H$ from the 530 replicates of Finite Sample Size Experiment III (FSSE III).	140
Table G.19	An estimation summary of $\hat{\sigma}_M$ from the 530 replicates of Finite Sample Size Experiment III (FSSE III).	140
Table G.20	An estimation summary of $\hat{\sigma}_H$ from the 530 replicates of Finite Sample Size Experiment III (FSSE III).	141
Table G.21	An estimation summary of $\hat{\rho}$ from the 530 replicates of Finite Sample Size Experiment III (FSSE III).	141
Table G.22	An estimation summary of $\hat{\eta}$ from the 530 replicates of Finite Sample Size Experiment III (FSSE III).	141
Table G.23	An estimation summary of $\hat{\theta}$ from the 530 replicates of Finite Sample Size Experiment III (FSSE III).	141
Table G.24	An estimation summary of $\hat{\mu}_H$ from the human marginal model (2.21). Includes results from both Finite Sample Size Experiments I and II.	142
Table G.25	An estimation summary of $\hat{\sigma}_H$ from the human marginal model (2.21). Includes results from both Finite Sample Size Experiments I and II.	142
Table G.26	An estimation summary of $\hat{\eta}_H$ from the human marginal model (2.21). Includes results from both Finite Sample Size Experiments I and II.	142
Table G.27	From FSSE I, the empirical powers and false discovery rates for detecting differentially expressed human genes at nominal FDR .05.	143
Table G.28	From FSSE I, the empirical powers and false discovery rates for detecting differentially expressed human genes at nominal FDR .20.	144
Table G.29	From FSSE I, the empirical powers and false discovery rates for detecting I-Orthologs.	145
Table G.30	From FSSE II, the empirical powers and false discovery rates for detecting differentially expressed human genes at nominal FDR .05.	146

Table G.31	From FSSE II, the empirical powers and false discovery rates for detecting differentially expressed human genes at nominal FDR .20.....	147
Table G.32	From FSSE II, the empirical powers and false discovery rates for detecting I-Orthologs.	148
Table G.33	From FSSE III, the empirical powers and false discovery rates for detecting differentially expressed human genes at nominal FDR .05.....	149
Table G.34	From FSSE III, the empirical powers and false discovery rates for detecting differentially expressed human genes at nominal FDR .20.....	150
Table G.35	From FSSE III, the empirical powers and false discovery rates for detecting I-Orthologs.	151
Table G.36	Effects on the distribution of human differential expression posterior probabilities (the p_i s) when one parameter is modulated from its true value. This report is based on a single simulated dataset from joint model (2.18) with $N = 10,000$, $\eta = .5$, $\theta = .4$, and the remaining parameters set to values as estimated from the GlaxoSmithKline Data (Table 2.2).....	152
Table G.37	Effects on the local false discovery rates for differentially expressed human genes (the p_i s for which $\beta_{Hi} = 0$) when one parameter is modulated from its true value. This report is based on a single simulated dataset from joint model (2.18) with $N = 10,000$, $\eta = .5$, $\theta = .4$, and the remaining parameters set to values as estimated from the GlaxoSmithKline Data (Table 2.2).	153

LIST OF FIGURES

Figure 2.1	Population mean model for the relationship among I-Orthologs.....	21
Figure 2.2	A visual representation of the five components of model (2.16). The top left graph represents the gene pairs in which neither mouse or human gene was differentially expressed. The top center and top right graphs represent the cases where the human gene is affected and the mouse gene is not, and vice versa. The bottom left figure represents the differentially expressed I-Orthologs. The differentially expressed non-I-Orthologs are depicted at the bottom center. Finally, the bottom right figure is an overlay representation of all five components.	26
Figure 2.3	Mouse and human treatment effect statistics from the GlaxoSmithKline data.....	36
Figure 2.4	Pairs $(\hat{\beta}_{Hi}, \hat{\beta}_{Mi})$ and $(\hat{\beta}_{Hi}/\hat{\sigma}_{2i}, \hat{\beta}_{Mi}/\hat{\sigma}_{1i})$ are color coded by $\hat{p}_i$, the posterior probability of treatment effect as follows: red for $\hat{p}_i > .886$ (all .05 FDR discoveries), yellow for $.643 \leq \hat{p}_i < .886$ (all additional .20 FDR discoveries) and green otherwise.....	39
Figure 2.5	Pairs $(\hat{\beta}_{Hi}, \hat{\beta}_{Mi})$ and $(\hat{\beta}_{Hi}/\hat{\sigma}_{2i}, \hat{\beta}_{Mi}/\hat{\sigma}_{1i})$ are color coded by $\hat{q}_i$, the posterior probability of I-Orthology as follows: red for $\hat{q}_i > .329$ (twice $\hat{\theta}$), yellow for $.165 \leq \hat{q}_i < .329$ (at least $\hat{\theta}$) and green otherwise.....	40
Figure 2.6	Venn diagrams showing the relationships between the various methodologies in terms of human genes deemed differentially expressed.....	41
Figure 3.1	Histograms of $\log \hat{\sigma}_{Mi}^2$ and $\log \hat{\sigma}_{Hi}^2$ and Estimated Normal Densities (via Least Squares).....	44
Figure 3.2	Boxplots of $\log \hat{\sigma}_{Mi}^2$ and $\log \hat{\sigma}_{Hi}^2$ for ordered groupings of $\hat{\beta}_{Mi}$ and $\hat{\beta}_{Hi}$, respectively, with 300 observations per bin.....	45
Figure 3.3	The baseline expression vs. variability relationship for both mice and humans.....	46
Figure 3.4	Estimates of σ_H and ρ in FSSE I, over all factorial combinations. The dotted lines indicate the 1 st and 99 th percentiles, respectively.....	58
Figure 3.5	Estimates of σ_H and ρ in FSSE I, where 74 out of 8059 runs with $\hat{\sigma}_H < 3$ were removed. The dotted lines indicate the 1 st and 99 th percentiles, respectively.	59

- Figure C.1 An informal search for a Behrens-Fisher problem in the two most significant mouse and human genes (with respect to treatment). The above graphs display the raw gene expression data by treatment group. 98
- Figure G.1 From the Infinite Sample Size Experiment (ISSE), boxplots were constructed to display the distribution of the maximum likelihood estimates for μ_M , μ_H , σ_M^{Tr} , σ_H^{Tr} at various levels of N . The plots include estimates from 497 replicate data sets with $(\beta_{Mi}, \beta_{Hi}), i = 1, \dots, N$, generated from model (2.16). 118
- Figure G.2 From the Infinite Sample Size Experiment (ISSE), boxplots were constructed to display the distribution of the maximum likelihood estimates for ρ^{Tr} , η , θ (with ρ reparameterized as in Section 2.4.1) at various levels of N . The plots include estimates from 497 replicate data sets with $(\beta_{Mi}, \beta_{Hi}), i = 1, \dots, N$, generated from model (2.16). 119
- Figure G.3 From the Infinite Sample Size Experiment (ISSE), these overlay plots display mean information matrix standard errors (blue) alongside Monte Carlo standard errors (red) for μ_M , μ_H , σ_M^{Tr} , σ_H^{Tr} (using the reparameterizations from Section 2.4.1) with number of genes = N and 497 replicate data sets. 120
- Figure G.4 From the Infinite Sample Size Experiment (ISSE), these overlay plots display mean information matrix standard errors (blue) alongside Monte Carlo standard errors (red) for $\hat{\rho}^{\text{Tr}}$, $\hat{\eta}^{\text{Tr}}$, $\hat{\theta}^{\text{Tr}}$ (using the reparameterizations from Section 2.4.1) with number of genes = N and 497 replicate data sets. 121
- Figure G.5 From the Infinite Sample Size Experiment (ISSE), empirical false discovery rates of I-Ortholog discovery are plotted as a function of the number of gene pairs, N , for nominal FDRs .05 and .20 and based on 497 replicate data sets. 122
- Figure G.6 From the Infinite Sample Size Experiment (ISSE), powers to detect I-Orthologs are plotted as a function of the number of gene pairs, N , for nominal FDRs .05 and .20 and based on 497 replicate data sets. 123

Chapter 1

Background

1.1 Introduction

The attrition rate of drug compounds entering phase I clinical trials is greater than 90% (Ganter et al, 2005) and the FDA has acknowledged a problem with the entire industry pipeline. There is no shortage of spending; Cuatrecasas (2006) notes that pharmaceuticals collectively exceed the NIH budget of \$28 billion a year and blames the industry's problems on modern management practices and short-term strategies appealing to shareholders, investment bankers and analysts. In an FDA white paper addressing the pipeline problem, authors discuss various technological deficiencies for the prediction of safety and efficacy (Frantz, 2004). In particular, Deputy FDA Commissioner Janet Woodcock claims that the tools used for determining human toxicology and safety are "decades old."

A brief history of drug discovery, inspired by Drews (2000), is that new drugs once arose from the imaginations of chemists. This was replaced with a dialogue between biologists and chemists, and molecular biology eventually integrated genetic information into the chemical realm. Today, drug discovery is based upon technologies such as combinatorial chemistry and automated high-throughput screening (HTS), testing large numbers of potential drug candidates against large numbers of hypothetical therapeutic targets.

In the present practices of drug discovery, quality often gives way to the "magic of large numbers," argues Drews (2000). Ulrich and Friend (2002) explain that the high throughput nature of the modern drug discovery paradigm has too often sent ill-understood compounds into development only to become costly failures. Fortunately, the advent of high-throughput drug screening technologies has coincided with the advent of high-throughput

genomics technologies which promise to produce biological knowledge on pace with the increasing number of drug candidates. In particular, *gene expression microarrays* allow researchers to measure the simultaneous expression activity of thousands of genes for a given cell sample. With respect to a drug’s administration, microarray experiments may elucidate the set of differentially expressed genes, the drug’s “gene expression fingerprint,” which Debouck and Goodfellow (1999), Orphanides (2003), and Ulrich and Friend (2002) believe will prove useful in predicting adverse events. Ulrich and Friend (2002) even suggest that a compound’s gene expression profile is representative of its “global biological effect,” including efficacy, toxicity, and all side effects. Orphanides (2003) believes that deciphering a compound’s “fingerprint” may enable determination of the long-term effects of a compound from short-term studies.

While microarray technology has not yet had a substantial impact on drug discovery, many studies give weight to the “fingerprint” hypothesis. Notably, Orphanides (2003) cites van ’t Veer et al. (2002) for identifying a “fingerprint” capable of distinguishing metastatic and non-metastatic breast tumors, and Hughes et al. (2000) for generating a gene expression database for yeast capable of predicting the biological effects of exogenous compounds. Ulrich and Friend (2002) cite Waring et al. (2002) for demonstrating that the toxic mechanism for an unexplored drug can be determined by comparison against a small gene-expression reference collection. Fielden et al. (2005) claim to have discovered a gene expression signature predictive of drug induced renal tubular toxicity in rats. Additionally, a number of toxicogenomic databases are arising that may begin to provide researchers with toxicogenomic catalogues. Notably, the private DrugMatrix® database boasts of gene-expression profiling in over 4100 “drug-dose-time-tissue” combinations and boasts of the utility of its Drug Signatures® gene biomarkers (Ganter et al, 2006).

Finding the genes whose expression levels are associated with an experimental condition is a difficult statistical problem involving the simultaneous testing of

$$\begin{aligned} H_{0i} : & \text{ gene } i \text{ was not affected by the drug compound,} \\ H_{1i} : & \text{ gene } i \text{ was affected by the drug compound,} \end{aligned}$$

for $i = 1, \dots, N$, where N is typically in the thousands. Finding a compound’s expression profile encompasses statistical multiple testing problems, an area of research addressing the failure of error control in individual testing to generalize to simultaneous testing (with enough hypotheses it becomes probable that at least one null hypothesis will be rejected

incorrectly). Despite an extensive body of literature, multiple testing in microarray settings is a difficult problem due to high measurement error, large numbers of simultaneous tests, and small subject sample sizes.

If the promises of microarray technologies are fulfilled, then significant financial gains will certainly result from enhanced prediction of toxicity and other side effects. Unclear, however, is the impact that microarray technology may have on preclinical (animal testing) phases of development, though the FDA (2004) estimates that enhanced preclinical predication of drug failures could save \$100 million dollars per drug. Aside from ethical controversies, the utility of animal experimentation has seen the gamut of viewpoints, from being largely worthless (Freedman and Zeisel, 1988) to having variable utility (depending on the particular animal and biological system, Greaves et al, 2004) to being irreplaceable, useful and “essential” to drug discovery (Thomas, 2006). Microarrays are designed for animals as well as humans, and Greaves et al. (2004) recognize these technologies as “exciting new developments in the preclinical study of drug toxicity,” but also acknowledge that “there is very little evidence that they have had a major and direct impact on the safety assessment which supports the very first studies of a new drug in humans.” Indeed, interpretation of preclinical microarray data is especially difficult as a single amino acid difference between animal and human proteins can change whether a particular drug compound has an effect (Debouck and Goodfellow, 1999). Though describing gene profiling in animal model systems as “critical” to future discovery and validation efforts, Debouck and Goodfellow (1999) warn that gene sequencing and gene function “are not always faithfully conserved between species” and thus “relevance to human disease needs to be carefully gauged.”

This research seeks to combine the analysis of preclinical and clinical gene expression data via a novel statistical platform (a model) that increases power to find differentially expressed human genes and enables future animal-to-human gene expression extrapolations. Section 1.2 is a survey of the multiple hypothesis testing literature especially relevant to microarray data. Section 1.3 introduces the biological machinery necessary to simultaneously analyze genomic data from multiple species. Section 1.4 describes a unique contribution from GlaxoSmithKline, a multi-species genomic dataset from its Integrative Biology Initiative. Finally, Section 1.5 introduces the methodology that will be the subject of further investigation throughout this paper.

1.2 Multiple Testing and the Search for Differentially Expressed Genes

Estimating the set of differentially expressed genes requires a thoughtful approach to the simultaneous testing of numerous hypotheses. For every individual hypothesis tested, there are two types of errors. Type I errors are incorrect decisions made to declare a gene differentially expressed. Type II errors are failures to declare a differentially expressed gene significant. Additionally, a test's *power* is the complement of its Type II error probability, i.e. the probability of detecting a particular differentially expressed gene.

This section provides an overview of multiple hypothesis testing literature that is particularly relevant in the context of microarray experiments. Section 1.2.1 describes frequentist approaches, where adjustments are made to p-values or test-statistics in order to meet a generalized version of Type I error. Section 1.2.2 describes methods typically called “Bayesian” or “empirical Bayes,” which are based on posterior probabilities of differential expression but may in actuality employ frequentist estimation. In the *The Interplay of Bayesian and Frequentist Analysis*, Bayarri and Berger (2004) classify multiple testing as one of three listed “areas of current disagreement” among frequentist and Bayesian approaches, calling the resulting methodologies “competing and quite distinct.”

1.2.1 Frequentist Multiple Testing

Dudoit et al. (2002) go as far as crediting the field of genomics for reviving interest in multiple testing procedures and report on the performance of various methodologies in the context of a microarray experimental framework. This section is largely a summary of their presentation. Though describing Bayesian approaches as an “important class of methods for the identification of differentially expressed genes,” the study examined only frequentist techniques. Table (1.1) below (originally appearing in Benjamini and Hochberg, 1995) summarizes multiple testing situation from the frequentist perspective, where m is the number of hypotheses (specified in advance), m_0 and $m_1 = m - m_0$ are the numbers of true and false null hypotheses, and R and $\{S, T, U, V\}$ are observable and unobservable random variables, respectively.

Based on Table (1.1), the following are the “most standard” generalizations of Type I error to the multiple testing situation (Dudoit et al. citing Shaffer 1995):

Table 1.1: The frequentist multiple testing framework.

	# not rejected	# rejected	
# true null hypotheses	U	V	m_0
# non-true null hypotheses	T	S	m_1
	$m - R$	R	m

- *Per-comparison error rate (PCER)*: $E(V)/m$, the expected ratio of Type I errors to the number of hypotheses;
- *Per-family error rate (PFER)*: $E(V)$, the expected number of Type I errors;
- *Family-wise error rate (FWER)*: $\Pr(V \geq 1)$, the probability of at least one Type I error;
- *False discovery rate (FDR)*: $E(Q)$, where $Q = \begin{cases} V/R & \text{if } R > 0, \\ 0 & \text{if } R = 0, \end{cases}$ (Benjamini and Hochberg, 1995). Thus the FDR is the expected proportion of false discoveries relative to the rejected hypotheses.

Additionally, Dudoit et al. report on three common definitions of power from Shaffer (1995):

- The probability of rejecting *at least one* null hypothesis, $\Pr(S \geq 1)$;
- *Average power*, the average probability of rejecting a false null hypothesis, $E(S)/m_1$;
- The probability of rejecting *all* false null hypotheses $\Pr(S = m_1)$.

Dudoit et al. (2002) point out that the probabilities and expectations defined above depend on the particular concept chosen for the generalized null hypothesis for multiple testing. The *complete null hypothesis* states that all null hypotheses are true, in contrast to the more general composite null hypothesis allowing any combination of true and false individual null hypotheses. When the Type I error rate is controlled conditional on the complete null hypothesis, the control is called *weak*. Alternatively, when the Type I error rate is controlled conditional on the more general global null hypothesis, the control is

called *strong*. Dudoit et al. (2002) argue that for microarray data, strong control of the Type I error rate is desirable since it is especially unrealistic that no genes are differentially expressed.

For control of the FWER, Dudoit et al. (2002) endorse the step-down maxT procedure of Westfall and Young (1993) but note that (for fixed nominal level $\alpha = .05$) the FDR procedures investigated provided “substantial” increases in power. The authors furthermore present FDR control as “promising, ...[since] one may be willing to bear a few false positives as long as their number is small in comparison to the number of rejected hypotheses.” In addition to the potential increase in power, FDR control is also appealing in toxicogenomic and drug discovery contexts where falsely discovering genes may at worst lead to inappropriate warnings subject to further investigation.

The current flagship procedure for control of the FDR was presented by Benjamini and Hochberg (1995). When testing hypotheses $H_1, H_2, \dots, H_m$ with ordered p-values $p_{r_1} \leq p_{r_2} \leq \dots \leq p_{r_m}$, the procedure is to reject hypotheses $H_{r_1}, \dots, H_{r_k}$ where $k = \max\{i : p_{r_i} \leq i\alpha/m\}$. The corresponding adjusted p-values are

$$\tilde{p}_{r_i} = \min_{j=i, \dots, m} \left\{ \min \left(\left\{ \frac{m}{j} p_{r_j}, 1 \right\} \right) \right\}, \quad (1.1)$$

where α is the nominal FDR. This procedure was demonstrated to control FDR at level α for independent test statistics (Benjamini and Hochberg, 1995) and various other dependence structures (Benjamini and Yekutieli, 2001). Even in the independent case, Dudoit et al. (2002) argue that the Benjamini and Hochberg (1995) procedure may still be conservative, as a consequence of the procedure is that $E(Q) \leq \alpha m_0/m$. Additionally, Benjamini and Yekutieli (2001) presented a conservative procedure which controls FDR for arbitrary dependence structures and Yekutieli and Benjamini (1999) proposed resampling based FDR adjusted p-values which allows for certain dependency structures (with additional assumptions).

In simulations by Dudoit et al. (2002), the Benjamini and Yekutieli (2001) procedure was substantially more conservative than that of Benjamini and Hochberg (1995), with power discrepancies of up to 30%. Notably, with correlated test statistics, the Benjamini and Hochberg (1995) procedure still controlled FDR at the nominal level. In addition, permutation-adjusted p-values for the Benjamini and Hochberg (1995) procedure were also investigated (not to be confused with that of Benjamini and Yekutieli, 2001) but the discreteness of the permutation unadjusted p-values led to a substantial increase in

conservatism.

An additional frequentist methodology that has received much attention in the literature is the Significance Analysis of Microarrays (SAM) method of Tusher et al. (2001), a multiple testing method developed specifically for the purpose of detecting differentially expressed genes in microarray experiments. At the foundation of the methodology lie gene-specific “scores,” closely resembling pooled two-sample t -statistics:

$$t_i^* = \frac{\bar{x}_{Ti} - \bar{x}_{Ci}}{s_0 + \sqrt{S_{pooled_i}^2}}, \quad (1.2)$$

where i is a gene index, $\bar{x}_{Ti}$, $\bar{x}_{Ci}$ are treatment and control group means, S_{pooled}^2 is the pooled estimate of variance, and s_0 is a small positive constant (s_0 is added to the denominator to keep genes with small variances from inflating t_i^*). The next step in the SAM procedure is to define (more or less arbitrary) thresholds for t_i^* yielding groups of “potentially significant” genes. For every choice of threshold, the FDR of the resulting group is estimated by permutation methodology. Tusher et al. (2001) admit that SAM only provides weak control of the FDR but suggest that it might approximately control FDR in the strong sense.

1.2.2 Bayesian Multiple Testing

Opgen-Rhein and Strimmer (2007) view Bayesian methodology as the gold standard in differential gene expression problems because of the information sharing *across genes* they naturally allow for. This information sharing is especially important when sample sizes are small (as is typical for microarray experiments). Furthermore, there is a connection between the posterior probabilities of individual hypotheses and the frequentist concept of False Discovery Rate (Efron et al, 2001). On the other hand, Bayesian methodology is often computationally expensive and tends to rely on a heavy assumptional framework (Opgen-Rhein and Strimmer, 2007).

As one of the earlier efforts to address Bayesian differential gene expression, Efron et al. (2001) choose to minimize the assumptional framework with a nonparametric empirical Bayes model. At the heart of the procedure are gene level test statistics Z_i (where i is the gene index) for differential expression due to an experimental condition. The Z_i s are essentially t -statistics with a positive constant added to the denominator (similar to the SAM method of Tusher et al, 2001). Avoiding any distribution assumptions, Efron et al.

(2001) introduce the models

p_1 = probability that a gene is affected,

p_0 = $1 - p_1$ = probability unaffected,

and

$f_1(z)$ = the density of Z for affected genes,

$f_0(z)$ = the density of Z for unaffected genes,

leading to the mixture density

$$f(z) = p_0 f_0(z) + p_1 f_1(z). \quad (1.3)$$

Application of Bayes rule to (1.3) leads to the updating equations

$$\Pr(\text{Gene } i \text{ was affected by treatment} | Z_i) = 1 - p_0 f_0(Z_i) / f(Z_i),$$

and

$$\Pr(\text{Gene } i \text{ was unaffected by treatment} | Z_i) = p_0 f_0(Z_i) / f(Z_i). \quad (1.4)$$

The null distribution $f_0(z)$ is estimated by permutation methods and the ratio f_0/f is estimated using logistic regression. While p_0 and p_1 are not identifiable, Efron et al. (2001) provide probabilistic upper bounds for p_0 that allow the analyst to choose a conservative prior without resorting to $p_0 = 1$.

More generally than the method itself, Efron et al. (2001) discuss the connection of posterior probabilities of differential expression to the FDR of Benjamini and Hochberg (1995). In one sense, the connection could not be any more straightforward: while FDR is the expected proportion of false discoveries among a collection of simultaneous hypotheses, Efron et al. (2001) define the *local false discovery rate* to be the posterior probability of no treatment effect,

$$\text{fdr}(T_i) = \Pr(\text{Gene } i \text{ was unaffected by treatment} | T_i), \quad (1.5)$$

where T_i is a test statistic for differential expression (Z_i in the context of Efron et al, 2001). For an intuitive explanation, consider the situation where there exists a group of genes with posterior probabilities of differential gene expression at or very near .95. Based on this high probability of differential expression, the decision is made to declare this entire group

of genes “significant.” Using a frequentist interpretation of probability, an approximate proportion of .95 of these “discoveries” will be true and .05 (the posterior probability of no treatment effect) will be false. Of course the posterior probabilities of differential gene expression will greatly vary among genes, and although Efron et al. (2001) discuss a “global definition of FDR,” a more general discussion for model-based procedures exists in Newton et al. (2004). For identifying as many differentially expressed genes as possible while bounding the rate of false discoveries at α , Newton et al. (2004) use the so-called *direct posterior probability approach*. For control of the FDR at level α , conclude all genes with $\text{fdr}(T_i) < \kappa^*$ significant, where

$$\kappa^* = \max \left\{ \kappa : \frac{\sum_{i \in G} \text{fdr}(T_i) \mathbf{I}(\text{fdr}(T_i) < \kappa)}{\sum_{i \in G} \mathbf{I}(\text{fdr}(T_i) < \kappa)} \leq \alpha \right\}, \quad (1.6)$$

where $\mathbf{I}(\cdot)$ is the indicator function and G is the index set of genes. Thus, the method chooses the largest set of genes with an *average local FDR* of less than α . Newton et al. (2004) warn however that this method of FDR control is model-based and thus susceptible to inadequacies in the modeling.

Do et al. (2005) extend the framework of Efron et al. (2001) by modeling the unknown densities f_0 and f_1 with the Dirichlet process, which the authors describe as “by far the most popular nonparametric Bayesian model.” The Dirichlet process requires a scalar parameter and a base probability measure. For the base probability measure, Do et al. (2005) use a normal probability measure, for the null scores, and a mixture of two normals, for the differentially expressed scores (with component means of $-b_1$ and b_1 and equal variances). Rather than using a probabilistic bound for the estimation of p_0 , here a prior is used, $p_0 \sim \text{Uniform}(0.05, 1)$. Markov chain Monte Carlo (MCMC) sampling is used for inference, and the authors conclude that the method is “effective” as it alleviates the posterior probability bias of the method of Efron et al. (2001).

Despite the reluctance by Efron et al. (2001) and Do et al. (2005) to make distributional assumptions, Newton et al. (2004) embrace a parametric formulation. Consider treatment and control group of gene expression values with unknown population means of μ_{Ti} and μ_{Ci} , for the i^{th} gene. Based on the following three hypotheses,

$$\begin{aligned} H_{i,0} : \mu_{Ti} &= \mu_{Ci}, \\ H_{i,1} : \mu_{Ti} &< \mu_{Ci}, \\ H_{i,2} : \mu_{Ti} &> \mu_{Ci}, \end{aligned}$$

Newton et al. (2004) propose the following latent treatment effects model:

$$\begin{aligned} f_{\mu_{Ci}, \mu_{Ti}}(\mu_{Ci}, \mu_{Ti}) &= p_0 \pi(\mu_{Ci}) I(\mu_{Ci} = \mu_{Ti}) + 2p_1 \pi(\mu_{Ci}) \pi(\mu_{Ti}) I(\mu_{Ci} < \mu_{Ti}) \\ &\quad + 2p_2 \pi(\mu_{Ci}) \pi(\mu_{Ti}) I(\mu_{Ci} > \mu_{Ti}), \end{aligned} \quad (1.7)$$

where $\pi(\cdot)$ is the pdf of a gamma distribution with shape parameter a_0 and “rate” parameter $a_0 x_0$. Model (1.7) is fit to data using full maximum likelihood via the EM algorithm. In simulation studies, Newton et al. (2004) remark that (1.7) captured “substantial” sources of variation. They also acknowledge the advantages of working with a full probability model, which provides gene-specific probabilities and error rates for various collections of genes.

At the opposite end of the spectrum relative to Efron et al. (2001) lies fully parametric Bayesian modeling, in which an extensive distributional framework for gene expression phenomena is presupposed. Lewin et al. (2006) model log-gene expression y_{ijk} for the i^{th} gene, treatment condition $j = 0, 1$, and replicate k as

$$\begin{aligned} y_{i0k} &\sim N\left(\alpha_i - \frac{1}{2}\beta_i + \delta_{i0k}, \sigma_{i0}^2\right), \\ y_{i1k} &\sim N\left(\alpha_i + \frac{1}{2}\beta_i + \delta_{i1k}, \sigma_{i1}^2\right), \end{aligned} \quad (1.8)$$

where α_i is a baseline level of expression, β_i is the differential effect between conditions, and σ_{ij}^2 is the condition-specific variance for gene i (i.e., an *exchangeable* variance model). The modeling integrates what is typically thought of as data preprocessing, as δ_{ijk} is the additive affect associated with the individual array (modeled with a quadratic spline that incorporates α_i).

Every parameter in Lewin et al.’s (2006) construction is assigned a prior. Two interesting assignments are

$$\beta_i \sim N(0, 100^2), \quad (1.9)$$

$$\sigma_{ij}^2 \sim \text{logNorm}(\mu_j, \eta_j^2). \quad (1.10)$$

The standard deviation of 100 in (1.9) reflects the typically large variation among genes in microarray data (see Section 2.6), despite the common scale of measurement. The lognormal model for σ_{ij}^2 in (1.10) was suggested from the data over one and two parameter gamma models, and the exchangeable variance model was deemed necessary.

Lewin et al. (2006) also use model-based FDR control to search for differentially expressed genes, but the posterior probabilities are based on an arbitrary notion of equiva-

lence. Specifically, Lewin et al. (2006) define

$$p_i = \Pr(\beta_i > \beta_{cut} | \text{data}), \quad (1.11)$$

and declare genes significant if $p_i > p_{cut}$, where β_{cut} is an equivalence limit (chosen by expert opinion or biological significance) and p_{cut} is chosen in a manner that controls FDR (the authors also investigate a similar decision rule that incorporates the baseline α_i). Lewin et al.’s (2006) full Bayesian model, implemented using WinBUGS software, yields posterior distributions for all model parameters.

To assess the differences between a full and empirical Bayesian formulation, Blangiardo et al. (2006) compare both types of methodology in an actual data analysis context. The full Bayesian model studied is remarkably similar to that of Lewin et al. (2006) and the empirical Bayes model is from Tseng et al. (2001). The same rule of declaring genes significant from the treatment effect posterior distributions was applied to both methods. Though differences may be due to a variety of factors, Blangiardo et al. conclude that:

- the full Bayesian model was more conservative than the empirical Bayes model;
- the full Bayesian model was more robust to different normalization procedures;
- while both models reveal a shrinkage effect, gene-level variance estimates are larger in the full Bayesian model.

1.3 Multi-Species Genomic Data and Mammalian Orthology

The past decade has seen development of quantitative methodologies that incorporate genetic information from two or more species. For example, Batzoglou et al. (2000) and Taher et al. (2004) combined the genomic sequences of multiple species in order to gain sensitivity in gene-finding algorithms. Sandelin et al. (2004) used DNA sequences from multiple species to locate regions of functional importance.

Joint modeling of animal and human gene expression data however poses some unique challenges. The simple reality that animals and humans have different genes would seem to limit analysis to well-understood model systems. Assistance arises, however, from the field of comparative and evolutionary genomics, where there has been recent interest in the evolutionary relationship known as “orthology” (Sonnhammer and Koonin, 2002). According to Sonnhammer and Koonin, the descriptor “orthologs” is often misused to

mean genes in different organisms that have the same function, whereas a more correct definition is “two genes from two different species that derive from a single gene in the last common ancestor of the species.” Regardless, orthology arguments abound for a variety of genes and species. In particular, the Mouse Genome Informatics (MGI) group warehouses a continuously-updated database of thousands of proposed human and mouse orthologs (www.informatics.jax.org), with annotations of supporting evidence for each gene pair.

In the context of gene expression, orthology concepts have proven remarkably useful. Kaski et al.’s (2005) cross-clustering algorithm found groups of “highly conserved” mouse and human orthologs. Working with yeast, Jiménez et al. (2002) concluded that gene expression was “preserved” within so-called Clusters of Orthologous Groups (COGs) and suggested that these categories provide a “scaffold for the comparison of expression data between different organisms.” Adjaye et al. (2004) investigated hybridizing bovine brain tissue to human cDNA microarrays and achieved a gene expression correlation of .94 when focusing on 349 human/bovine orthologs.

1.4 GlaxoSmithKline’s Integrative Biology Initiative

In 2000, GlaxoSmithKline launched the Integrative Biology (IB) Initiative, an ambitious undertaking which collected high-throughput bioinformatic data from multiple platforms to evaluate mechanisms of human disease and treatment. In addition to human data, IB included an equally rigorous preclinical component to investigate applications of animal data to human target validation and biomarker discovery.

Type-II Diabetes Mellitus was the first disease project undertaken by Integrative Biology, where both preclinical and clinical experiments were implemented to investigate the treatment rosiglitazone. The preclinical experiment used C57BL/6J laboratory mice fed either a high-fat diet (58% k/cal) or a low-fat diet (11% k/cal), such that the mice on the high-fat diet developed symptoms consistent with Type-II Diabetes Mellitus. Dosing for the treatment mice was implemented at levels of 1, 3 and 10 mg per kilogram of body weight, administered in 10 ml/kg doses twice a day. The clinical study, on the other hand, used only humans that were suffering from symptoms of Type-II Diabetes Mellitus. Treatment humans received a dosing regimen of 4 to 8 mg total daily dose of rosiglitazone via twice-a-day treatment application. The variations in human dosing are due to up and down titrations implemented to maintain glycemic control and are ignored for the purpose of

analysis in this thesis.

Common to both clinical and preclinical experiments was collection of gene expression data using Affymetrix GeneChip® oligonucleotide microarrays (discussed in Section 1.4.1). Specifically, the human experiment uses the Human Genome U133 Plus 2.0 Array and the mouse experiment uses the MOE430A Array. In both mice and humans, subcutaneous fat cells are used in the expression analysis and are the target of the following investigation.

1.4.1 Parallel Analysis of Mouse and Human Experiments

The experimental studies on mice and humans, though complimentary, fail on several counts to align in a conceptually one-to-one fashion. Though “throwing away” data is generally not regarded as good practice, some deletions are made in order to minimize complication and retain the form of “future data” prescribed by this research. For the investigative purposes of this research, we believe little harm is incurred by the following data decisions:

The “Healthy” Group

A difference between the two studies is the inclusion of a “healthy” mouse group in addition to a (laboratory induced) “disease” group. In the human experiment, only patients suffering from Type II diabetes mellitus were studied. To eliminate complication resulting from a potential disease status $\times$ treatment interaction, the decision was made to discard the healthy mouse data and keep only the data from the disease group.

Different Dosing Regimens

Another key difference between the two studies was that, while humans received only one dose level (varying to maintain glycemic control), the mice received three (1, 3 and 10 mg/kg). The 10 mg/kg dosing level was associated with a substantially higher signal-to-noise ratio in the data and Greaves et al. (2004) report that single, high-dose studies in animals is often sufficient for detecting adverse events in humans. Thus the decision was made to keep this dosing level (along with the control group), discarding the others.

Gene Level Summarization

In oligonucleotide microarrays, the lowest level of data collection occurs at the “probe level.” Probes are short DNA strands consisting of 25 base pairs that are built onto the microarray chip. When a prepared cell sample is washed over the chip, fluorescent-dyed RNA strands hybridize to matching probes, indicating expression of the corresponding gene. Additionally, for every probe design there is a so-called “mismatch” counterpart with a slight alteration in the base pair sequence. Individual probes (both “perfect match” and “mismatch”) are replicated on the chip, thus forming a *probe set*. In microarray analysis, the concept of “summarization” ubiquitously refers to the process by which probe-level data is transformed into a one-number probe set summary (see Li and Wong, 2001). Though an active area of research, for our purposes, the preprocessing algorithm of the Affymetrix MAS 5.0 software is trusted to produce quality probe set summaries.

Nearly all analyses involving oligonucleotide data is carried out at the probe set level, presenting a number of issues. Most notably, multiple probe sets are often annotated to a single gene, and there is no guarantee that these probe sets will behave in a consistent manner. A number of factors can influence hybridization to probes, and some probes are better designed than others (Affymetrix support). Though Affymetrix provides a mapping of probe sets from the (mouse) MOE430A Array to the Human Genome U133 Plus 2.0 Array, the mappings are based on sequence similarity alone. The richer biological analysis of the MGI ortholog database is more appealing, but the mapping exists at the gene level. Consequently a gene-level summary is necessary for the cases where multiple probe sets are annotated to a single gene. Averaging probe sets across a gene however is not recommended for “various reasons,” including cases where lack of knowledge about transcript orientation compelled Affymetrix designers to create multiple probe sets for each orientation when only one will work theoretically (Affymetrix support).

For this thesis, the chosen method of probe-set to gene level summary combines the use of limited bioinformatic information with a rule concerning maximal significance. The nomenclature of probe sets has bioinformatic significance, indicated from the following suffixes:

- **_at**: all the probes hit one known transcript;
- **_a**: all probes in the set hit alternate transcripts from the same gene;

- `_s`: all probes in the set hit transcripts from different genes;
- `_x`: some probes hit transcripts from different genes.

While not particularly intuitive, assistance from Affymetrix support has confirmed that “`_at`” is the best, “`_a`” and “`_s`” is the next best, and “`_x`” is the worst in terms of the quality in which these probe sets represent the genes they are annotated to. For the human data, the number of probe sets per gene was reduced by keeping only the highest quality probe sets available for a given gene. That is, if a gene is associated with one or more “`_at`” probe sets, then all “`_a`,” “`_s`,” and “`_x`” probe sets are discarded. Otherwise, if a gene has no “`_x`” probe sets but “`_a`,” “`_s`,” and “`_x`” probe sets, then the “`_x`” probe sets will be discarded. For the mouse data, only genes with uniquely mapped “`_at`” probe sets are kept and the rest discarded in order to keep only the highest quality animal transcripts for the purposes of extrapolation.

For both species, when there are still multiple probe sets for a gene after all reductions, the probe set with the *most significant differential expression* (using the regression modeled described in Section 1.5) is chosen to represent the gene. With this practice, the chosen probe set to represent each gene is the most likely to be functional, and the bias associated with the choice is no different than the common practice of analyzing data at the probe set level and interpreting discoveries at the gene level. Though annoying, technologies of the future will likely render this problem obsolete, as emerging “exon arrays” are more amenable to gene-level summarization.

1.4.2 Description of the Data

For the clinical experiment, data are available for 24 humans; 13 given the treatment rosiglitazone and 11 given a placebo. Every human had gene expression data collected at two time points, one before treatment (or placebo) administration, and one at the study’s completion (after eight weeks). The U133 Plus 2.0 Array provided usable expression measurements for 20,090 genes. The $20,090 \times 1$ vectors of gene expression measurements at “week 8” are viewed as the response, whereas the $20,090 \times 1$ vectors of gene expression measurements taken pre-study are viewed as covariates. In addition to treatment status, there are two other covariates measured for each human: *prior therapy* status (four categorical levels) and *concomitant medication* status (two categorical levels).

For the preclinical experiment, data are available for 19 mice; 9 given the treatment rosiglitazone, and 10 given a placebo. Gene expression measurements are collected at the end of the study, and the MOE430A Array provides usable expression measurements for 11,398 genes. This $11,398 \times 1$ vector of gene expression measurements constitutes the response vector for the mice.

As of November 3, 2007, the MGI database listed 16,928 pairs of mouse and human genes hypothesized to be orthologous, based on various evidences. The intersection of the usable mouse and human gene expression data from the GlaxoSmithKline experiments, with the information from the MGI database (as accessed), yields 9,683 purported orthologs with bivariate (mouse, human) gene expression data.

1.5 Joint Modeling

For the purposes of this research, we restrict attention to mouse/human gene pairs in the MGI ortholog database (Section 1.3), allowing a single gene index i to represent both mouse and human genes. Let y_{Mij} and y_{Hik} denote relative gene expression measurements of the i^{th} gene pair for the j^{th} mouse and k^{th} human. The following linear models parsimoniously describe the GlaxoSmithKline data:

$$y_{Mij} = \alpha_{Mi} + \beta_{Mi}T_{Mj} + \sigma_{Mi}\epsilon_{Mij}, \quad j = 1, \dots, 19,$$

and

$$y_{Hik} = \alpha_{Hi} + \beta_{Hi}T_{Hk} + \gamma_i^T \mathbf{z}_{ik} + \sigma_{Hi}\epsilon_{Hik}, \quad k = 1, \dots, 24,$$

for $i = 1, \dots, 9683$, where $\mathbf{z}_{ik}$ is a 5×1 vector of human covariates (week-0 expression, prior therapy and concomitant medication indicators), T_{Mj} and T_{Hk} are $\{0,1\}$ treatment indicators for rosiglitazone, and ϵ_{Mij} and ϵ_{Hik} are independent $N(0,1)$ random variables.

The least squares estimates $\hat{\beta}_{Mi}$ and $\hat{\beta}_{Hi}$ and estimated standard errors $\hat{\sigma}_{1i}$ and $\hat{\sigma}_{2i}$ (see Appendix B for more detail) comprise the essential ingredients of test-statistics (or “scores”) for differential expression. Rather than modeling “ t -like” statistics as in Efron et al. (2001) or latent means as in Newton et al. (2004), we proceed in this research to develop models for $\hat{\beta}_{Mi}$ and $\hat{\beta}_{Hi}$. Specifically, we investigate the marginal models

$$\hat{\beta}_{Mi} \sim \eta_M N(\mu_M, \sigma_{Mi}^2 + \sigma_{1i}^2) + (1 - \eta_M) N(0, \sigma_{1i}^2), \quad (1.12)$$

$$\hat{\beta}_{Hi} \sim \eta_H N(\mu_H, \sigma_{Hi}^2 + \sigma_{2i}^2) + (1 - \eta_H) N(0, \sigma_{2i}^2), \quad (1.13)$$

and the joint model

$$\begin{aligned}
\begin{pmatrix} \hat{\beta}_{Mi} \\ \hat{\beta}_{Hi} \end{pmatrix} &\sim \theta \eta N \left\{ \begin{pmatrix} \mu_M \\ \mu_H \end{pmatrix}, \begin{pmatrix} \sigma_M^2 + \sigma_{1i}^2 & \rho \sigma_H \sigma_M \\ \rho \sigma_H \sigma_M & \sigma_H^2 + \sigma_{2i}^2 \end{pmatrix} \right\} \\
&+ (1 - \theta) \eta (1 - \eta) N \left\{ \begin{pmatrix} \mu_M \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_M^2 + \sigma_{1i}^2 & 0 \\ 0 & \sigma_{2i}^2 \end{pmatrix} \right\} \\
&+ (1 - \theta) \eta (1 - \eta) N \left\{ \begin{pmatrix} 0 \\ \mu_H \end{pmatrix}, \begin{pmatrix} \sigma_{1i}^2 & 0 \\ 0 & \sigma_H^2 + \sigma_{2i}^2 \end{pmatrix} \right\} \\
&+ (1 - \theta) \eta^2 N \left\{ \begin{pmatrix} \mu_M \\ \mu_H \end{pmatrix}, \begin{pmatrix} \sigma_M^2 + \sigma_{1i}^2 & 0 \\ 0 & \sigma_H^2 + \sigma_{2i}^2 \end{pmatrix} \right\} \\
&+ \{ (1 - \theta)(1 - \eta)^2 + \theta(1 - \eta) \} N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{1i}^2 & 0 \\ 0 & \sigma_{2i}^2 \end{pmatrix} \right\},
\end{aligned} \tag{1.14}$$

for $i = 1, \dots, 9683$.

Note that marginal models (1.12) and (1.13) are two-component normal mixtures, where the latter component (in both models) has a zero mean and a variance equal to the estimation error to model the $\hat{\beta}_{Mi}$ s and $\hat{\beta}_{Hi}$ s without a treatment effect. The remaining component features a non-zero mean and additively-inflated variance to model β_{Mi} and β_{Hi} when there is a treatment effect. The means μ_M and μ_H describe the tendency of treatment to up or down regulate expression, and variances σ_M and σ_H describe the inherent variability of the latent treatment effects. Mixing parameters η_M and η_H are the proportion of mouse and human genes affected by the treatment.

The joint model (1.14) is in one sense an extension of marginal models (1.12) and (1.13), as the latter four distributional components follow from an assumption of independence between $\hat{\beta}_{Mi}$ and $\hat{\beta}_{Hi}$. The key difference is the inclusion of a hypothetical class of mouse and human gene pairs, herein called *I-Orthologs*, that share similar responses to a stimulus and are potentially useful extrapolators of gene expression. The proportion of I-Orthologs in a given dataset is described by mixing parameter θ , and the strength of the I-Ortholog relationship between β_{Mi} and β_{Hi} is given by correlation ρ . For a useful parameter reduction in the joint model, we make the assumption $\eta_M = \eta_H = \eta$.

After estimation of unknown parameters via likelihood-based methodology (Chapter 2), the following probabilities are computed:

- $\Pr(\text{mouse gene } i \text{ is affected by treatment} | \hat{\beta}_{Mi});$
- $\Pr(\text{human gene } i \text{ is affected by treatment} | \hat{\beta}_{Hi});$
- $\Pr(\text{human gene } i \text{ is affected by treatment} | \hat{\beta}_{Mi}, \hat{\beta}_{Hi}).$

Differentially expressed sets of genes are discovered using the direct posterior probability approach to FDR control (Section 1.2.2). The hope is that performance in detecting differentially expressed human genes will be enhanced by using posterior probabilities of differential expression based on both mouse and human data.

Also following estimation of unknown parameters, the posterior probability

$$\Pr(\text{mouse/human gene pair } i \text{ is an I-Ortholog} | \hat{\beta}_{Mi}, \hat{\beta}_{Hi})$$

is computed, allowing quantification of the amount of evidence that a particular mouse/human ortholog from the MGI database is an I-Ortholog and thus a useful cross-species predictor.

The remaining chapters of this thesis concentrate on the motivation, development and performance of joint model (1.14). Chapter 2 builds (1.14) from the ground up, discusses applications of the model, and includes an analysis of the GlaxoSmithKline data. Chapter 3 investigates the sampling and testing properties associated with (1.14) via simulation studies. Additionally, appendices are included that discuss model identifiability, microarray data transformations, the modeling of σ_{1i} and σ_{2i} , and other relevant topics.

Chapter 2

Joint Multilevel Modeling and Species Extrapolation

From the GlaxoSmithKline data (Section 1.4), let y_{Mij} and y_{Hik} denote gene-expression measurements from the i^{th} MGI ortholog gene pair (Section 1.3) for the j^{th} mouse and k^{th} human. The following linear models parsimoniously describe the data:

$$y_{Mij} = \alpha_{Mi} + \beta_{Mi}T_{Mj} + \sigma_{Mi}\epsilon_{Mij}, \quad j = 1, \dots, 19, \quad (2.1)$$

$$y_{Hik} = \alpha_{Hi} + \beta_{Hi}T_{Hk} + \boldsymbol{\gamma}_i^T \mathbf{z}_{ik} + \sigma_{Hi}\epsilon_{Hik}, \quad k = 1, \dots, 24, \quad (2.2)$$

for $i = 1, \dots, 9683$, where $\mathbf{z}_{ik}$ is a 5×1 vector of human covariates (week-0 expression, prior therapy and concomitant medication indicators), T_{Mj} and T_{Hk} are $\{0,1\}$ treatment indicators for rosiglitazone, and ϵ_{Mij} and ϵ_{Hik} are independent $N(0,1)$ random variables. The decision not to log-transform the positive and continuous y_{Mij} and y_{Hik} is discussed in Appendix C.

For the i^{th} mouse and human gene, α_{Mi} and α_{Hi} are baseline expression levels and σ_{Mi} and σ_{Hi} describe among-subject expression variability. The human parameter vector $\boldsymbol{\gamma}_i^T$ describes the linear relationship of the covariates to the mean expression, and is included in the model to account for otherwise unexplained variability. Most importantly, β_{Mi} and β_{Hi} describe the differential expression of the respective mouse and human genes due to a treatment intervention. Thus, the multiple hypotheses described in Section 1.2 can be

rewritten as

$$H_{0i} : \beta_{Hi} = 0,$$

$$H_{1i} : \beta_{Hi} \neq 0,$$

for the $i = 1, \dots, 9683$ human genes, and

$$H_{0i} : \beta_{Mi} = 0,$$

$$H_{1i} : \beta_{Mi} \neq 0,$$

for the $i = 1, \dots, 9683$ mouse genes.

Though perhaps β_{Mi} and β_{Hi} are best thought of as fixed, unknown quantities, there are compelling reasons to adopt a random modeling framework. The key is to model (β_{Mi}, β_{Hi}) so that something is gained by the pooling of human and animal gene expression data, and probabilistic models provide two helpful tools: (i) correlation, and (ii) mixture distributions to deal with hypothetical subgroups of genes.

2.1 The I-Ortholog Hypothesis

The MGI gene pairs focused on in this research (Section 1.3) are purported orthologs hypothesized to have been conserved from a single gene in a common ancestor. While Sonhammer and Koonin (2002) clarify that orthology is not equivalent to functional similarity, such evolutionary conservation suggests that orthologs serve similar purposes in their respective organisms. Thus in a search for mouse genes that are human predictors, the MGI orthologs are a logical place to start.

As this research is concerned with finding differentially expressed human genes, some additional machinery is necessary to benefit from the available mouse expression data. Thus, we define a new class of mouse and human gene pairs, herein called *I-Orthologs*. A mouse and human gene pair is an I-Ortholog *if the response of the mouse gene to an arbitrary treatment intervention is informative about the response of the human gene to the same stimulus, and vice versa*. We assume that the set of I-Orthologs is non-empty, though anticipate that only a subset of purported orthologs warehoused at MGI are in fact I-Orthologs. Among thousands of orthology hypotheses, some may be incorrect and others may not be useful for species extrapolation.

Figure 2.1: Population mean model for the relationship among I-Orthologs.

In the definition of I-Ortholog, that a mouse gene is “informative” of human response is left intentionally vague as we wish to separate the I-Orthology concept from models of the concept. Here, for concreteness, we focus on the GlaxoSmithKline data and the differential expression parameters β_{Mi} and β_{Hi} . If a gene pair is an I-Ortholog, then intuitively (β_{Mi}, β_{Hi}) would either lie in the first quadrant (the treatment increases the expression of both mouse and human genes), the third quadrant (the treatment decreases the expression of both mouse and human genes), or the origin (the treatment had no effect on either gene).

In yet a greater leap, we hypothesize a population mean-model for the I-Orthologs,

$$E(\beta_{Hi}) = \delta\beta_{Mi}, \quad (2.3)$$

as illustrated in Figure (2.1). Model (2.3) specifies that the latent differential expression coefficients β_{Mi} and β_{Hi} co-vary about a line passing through the origin. In other words, there is a common multiplier relating differential expression of an animal ortholog to its human counterpart. If this construction proves useful in practice and generalizes to a population of potential treatments, then I-Ortholog animal genes will serve as useful predictors of human differential expression in the preclinical testing of future drug compounds.

2.2 Latent Modeling

In this section we develop a latent bivariate model for the differential expression vector (β_{Mi}, β_{Hi}) , $i = 1, \dots, 9683$, from the GlaxoSmithKline data (Section 1.4) and models

(2.1) and (2.2). First, we define the following latent binary variables:

- $O_i = 1$ (0) if gene pair i is (is not) an I-Ortholog;
- $A_{Mi} = 1$ (0) if treatment did (did not) affect gene i among mice;
- $A_{Hi} = 1$ (0) if treatment did (did not) affect gene i among humans.

Section 2.2.1 discusses the modeling assumptions used in this research, and Section 2.2.2 introduces the resulting latent model for (β_{Mi}, β_{Hi}) .

2.2.1 Latent Model Assumptions

We define the probability that an arbitrary gene pair in the MGI database is an I-Ortholog:

$$\Pr(O_i = 1) = \theta. \quad (2.4)$$

Assuming similar pharmaceutical interventions in humans and mice differentially express the same *proportion* of genes on average, we furthermore define

$$\Pr(A_{Mi} = 1) = \Pr(A_{Hi} = 1) = \eta, \quad (2.5)$$

allowing for a useful reduction of parameters and a pooled estimate of η .

Additionally, we assume

$$A_{Mi} \perp O_i \text{ and } A_{Hi} \perp O_i, \quad (2.6)$$

$$A_{Mi} \perp A_{Hi} | O_i = 0, \quad (2.7)$$

$$\Pr(A_{Mi} = 0 | A_{Hi} = 0, O_i = 1) = 1, \quad (2.8)$$

where $\perp$ is the independence relation. Assumption (2.6) states that treatment indiscriminately affects I-Orthologs and non-I-Orthologs alike. From Assumption (2.7), it follows that for non-I-Orthologs, mouse differential expression status is uninformative of human differential expression status. Assumption (2.8), from which $\Pr(A_{Hi} = 0 | A_{Mi} = 0, O_i = 1) = 1$ can be derived, essentially says that I-Ortholog gene pairs always agree with respect to differential expression.

The distribution of (O_i, A_{Mi}, A_{Hi}) follows from (2.4)-(2.8), as

$$\begin{aligned}
\Pr(A_{Mi} = j, A_{Hi} = k, O_i = 0) \\
&= \Pr(A_{Mi} = j, A_{Hi} = k | O_i = 0) \Pr(O_i = 0) \\
&= \Pr(A_{Mi} = j | O_i = 0) \Pr(A_{Hi} = k | O_i = 0) (1 - \theta) \\
&= \Pr(A_{Mi} = j) \Pr(A_{Hi} = k) (1 - \theta),
\end{aligned} \tag{2.9}$$

where the second equality follows from (2.4) and (2.7) and the last equality follows from (2.6). Similarly,

$$\begin{aligned}
\Pr(A_{Mi} = j, A_{Hi} = k, O_i = 1) \\
&= \Pr(A_{Mi} = j | A_{Hi} = k, O_i = 1) \Pr(A_{Hi} = k | O_i = 1) \Pr(O_i = 1) \\
&= \Pr(A_{Mi} = j | A_{Hi} = k, O_i = 1) \Pr(A_{Hi} = k) \theta,
\end{aligned} \tag{2.10}$$

where the first equality follows from (2.4) and (2.7). Furthermore, note that $\Pr(A_{Mi} = j | A_{Hi} = k, O_i = 1)$ takes either the values of 0 or 1 according to (2.8).

Conditional on a treatment effect having occurred for the i^{th} mouse and/or human gene, we model the latent differential expression with the normal distribution, largely for convenience:

$$\begin{aligned}
\beta_{Mi} | A_{Mi} = 1 &\sim N(\mu_M, \sigma_M^2); \\
\beta_{Hi} | A_{Hi} = 1 &\sim N(\mu_H, \sigma_H^2).
\end{aligned} \tag{2.11}$$

Blangiardo et al. (2005) and Lewin et al. (2006) also model differential expression parameters with the normal distribution, though in both cases the raw expression data was log-transformed.

The non-differentially expressed genes are assumed to have differential expression parameters of zero, modeled with the degenerate distributions:

$$\begin{aligned}
\beta_{Mi} | A_{Mi} = 0 &\sim N(0, 0); \\
\beta_{Hi} | A_{Hi} = 0 &\sim N(0, 0).
\end{aligned} \tag{2.12}$$

Efron et al. (2001) and Newton et al. (2004) model non-differentially expressed genes in similar manner.

For differentially expressed non-I-Ortholog gene pairs, we assume

$$\beta_{Mi} \perp \beta_{Hi} | O_i = 0, \tag{2.13}$$

and thus $(\beta_{Mi}, \beta_{Hi})|O_i = 0$ can be represented as a bivariate normal with zero correlation. For differentially expressed I-Orthologs, a bivariate normal distribution with correlation ρ is used:

$$(\beta_{Mi}, \beta_{Hi})|O_i = 1, A_{Mi} = 1, A_{Hi} = 1 \sim N \left\{ \begin{pmatrix} \mu_M \\ \mu_H \end{pmatrix}, \begin{pmatrix} \sigma_M^2 & \rho\sigma_H\sigma_M \\ \rho\sigma_H\sigma_M & \sigma_H^2 \end{pmatrix} \right\}. \quad (2.14)$$

Note that (2.14) is actually a generalization of I-Ortholog model (2.3) discussed in Section 2.1, as $E(\beta_{Hi}|\beta_{Mi})$ is not forced through the origin. Though the parameter constraint $\mu_H - \rho\sigma_H/\sigma_M = 0$ would accomplish this, no such constraints are applied in this paper.

Finally, this research operates under the *working independence assumption*,

$$(\beta_{Mi}, \beta_{Hi}) \perp (\beta_{Mj}, \beta_{Hj}) \text{ for } j \neq i. \quad (2.15)$$

Though genes are known to interact and the change in expression of one gene may affect another (Stuart et al, 2003), it is generally not practical to take gene dependence structure into account when modeling expression. For instance, Newton et al. (2004) note that their presented methodology “deals only with the two-group comparison and [ignores] dependence among genes.” Do et al. (2005) state that the independence approximation is “still useful to determine whether expression level differences are significant solely on a gene-by-gene basis.”

2.2.2 Joint Model for Latent Treatment Effects

Table 2.1 completely describes the distributional models of $(\beta_{Hi}, \beta_{Mi}|O_i, A_{Hi}, A_{Mi})$ and (O_i, A_{Hi}, A_{Mi}) that follow from the assumptions given in Sec-

Table 2.1: Distributional Models for $(\beta_{Mi}, \beta_{Hi} | O_i, A_{Hi}, A_{Mi})$ and (O_i, A_{Hi}, A_{Mi})

Case	(O_i, A_{Hi}, A_{Mi})	$\begin{pmatrix} \beta_{Mi} \\ \beta_{Hi} \end{pmatrix} (O_i, A_{Hi}, A_{Mi}) \sim$	$\Pr(O_i, A_{Hi}, A_{Mi})$
1	$O_i = 0, A_{Hi} = 0, A_{Mi} = 0$	$N(\mathbf{0}, \mathbf{0})$	$(1 - \theta)(1 - \eta)^2$
2	$O_i = 0, A_{Mi} = 0, A_{Hi} = 1$	$N \left\{ \begin{pmatrix} 0 \\ \mu_H \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & \sigma_H^2 \end{pmatrix} \right\}$	$(1 - \theta)\eta(1 - \eta)$
3	$O_i = 0, A_{Mi} = 1, A_{Hi} = 0$	$N \left\{ \begin{pmatrix} \mu_M \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_M^2 & 0 \\ 0 & 0 \end{pmatrix} \right\}$	$(1 - \theta)\eta(1 - \eta)$
4	$O_i = 0, A_{Mi} = 1, A_{Hi} = 1$	$N \left\{ \begin{pmatrix} \mu_M \\ \mu_H \end{pmatrix}, \begin{pmatrix} \sigma_M^2 & 0 \\ 0 & \sigma_H^2 \end{pmatrix} \right\}$	$(1 - \theta)\eta^2$
5	$O_i = 1, A_{Mi} = 0, A_{Hi} = 0$	$N(\mathbf{0}, \mathbf{0})$	$\theta(1 - \eta)$
6	$O_i = 1, A_{Mi} = 0, A_{Hi} = 1$	NA	0
7	$O_i = 1, A_{Mi} = 1, A_{Hi} = 0$	NA	0
8	$O_i = 1, A_{Mi} = 1, A_{Hi} = 1$	$N \left\{ \begin{pmatrix} \mu_M \\ \mu_H \end{pmatrix}, \begin{pmatrix} \sigma_M^2 & \rho \sigma_H \sigma_M \\ \rho \sigma_H \sigma_M & \sigma_H^2 \end{pmatrix} \right\}$	$\theta\eta$

tion 2.2.1. Hence, the following mixture model results for the latent vector (β_{Hi}, β_{Mi}) :

$$\begin{aligned}
\begin{pmatrix} \beta_{Mi} \\ \beta_{Hi} \end{pmatrix} &\sim \theta \eta N \left\{ \begin{pmatrix} \mu_M \\ \mu_H \end{pmatrix}, \begin{pmatrix} \sigma_M^2 & \rho \sigma_H \sigma_M \\ \rho \sigma_H \sigma_M & \sigma_H^2 \end{pmatrix} \right\} \\
&+ (1 - \theta)\eta(1 - \eta) N \left\{ \begin{pmatrix} \mu_M \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_M^2 & 0 \\ 0 & 0 \end{pmatrix} \right\} \\
&+ (1 - \theta)\eta(1 - \eta) N \left\{ \begin{pmatrix} 0 \\ \mu_H \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & \sigma_H^2 \end{pmatrix} \right\} \\
&+ (1 - \theta)\eta^2 N \left\{ \begin{pmatrix} \mu_M \\ \mu_H \end{pmatrix}, \begin{pmatrix} \sigma_M^2 & 0 \\ 0 & \sigma_H^2 \end{pmatrix} \right\} \\
&+ \{(1 - \theta)(1 - \eta)^2 + \theta(1 - \eta)\} N(\mathbf{0}, \mathbf{0}).
\end{aligned} \tag{2.16}$$

The eight possible values of (O_i, A_{Hi}, A_{Mi}) collapse into five distributional components in (2.16); Cases 6 and 7 have zero probability of occurrence and Cases 1 and 5 yield identical distributional models, their contributions collapsing into the last component. As there is no distributional distinction between Cases 1 and 5, it will not be possible to

Figure 2.2: A visual representation of the five components of model (2.16). The top left graph represents the gene pairs in which neither mouse or human gene was differentially expressed. The top center and top right graphs represent the cases where the human gene is affected and the mouse gene is not, and vice versa. The bottom left figure represents the differentially expressed I-Orthologs. The differentially expressed non-I-Orthologs are depicted at the bottom center. Finally, the bottom right figure is an overlay representation of all five components.

separate I-Orthologs from other gene pairs which are unaffected by treatment. Furthermore, note that (2.16) does not have a density on the real plane because of the degenerate component $N(\mathbf{0}, \mathbf{0})$.

2.3 Estimated Treatment Effects

The latent quantities β_{Mi} and β_{Hi} are unobserved but lie at the bottom of a hierarchical model with estimable parameters. We observe gene expression data for mice and humans that, by hypothesis, follow linear models (2.1) and (2.2). Separate *least squares* estimation using mouse and human datasets yields estimates $\hat{\beta}_{Mi}$ and $\hat{\beta}_{Hi}$, with the familiar property

$$\begin{pmatrix} \hat{\beta}_{Mi} \\ \hat{\beta}_{Hi} \end{pmatrix} \mid \begin{pmatrix} \beta_{Mi} \\ \beta_{Hi} \end{pmatrix} \sim N \left[\begin{pmatrix} \beta_{Mi} \\ \beta_{Hi} \end{pmatrix}, \begin{pmatrix} \sigma_{1i}^2 & 0 \\ 0 & \sigma_{2i}^2 \end{pmatrix} \right], \quad (2.17)$$

where σ_{1i}^2 and σ_{2i}^2 are the conditional variances of $\hat{\beta}_{Mi}$ and $\hat{\beta}_{Hi}$ given β_{Mi} and β_{Hi} , respectively. The joint distribution of the observable vector $(\hat{\beta}_{Mi}, \hat{\beta}_{Hi})^T$ is derived from (2.16) and (2.17) in Appendix A.1, resulting in:

$$\begin{aligned} \begin{pmatrix} \hat{\beta}_{Mi} \\ \hat{\beta}_{Hi} \end{pmatrix} &\sim \theta \eta N \left\{ \begin{pmatrix} \mu_M \\ \mu_H \end{pmatrix}, \begin{pmatrix} \sigma_M^2 + \sigma_{1i}^2 & \rho \sigma_H \sigma_M \\ \rho \sigma_H \sigma_M & \sigma_H^2 + \sigma_{2i}^2 \end{pmatrix} \right\} \\ &+ (1 - \theta) \eta (1 - \eta) N \left\{ \begin{pmatrix} \mu_M \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_M^2 + \sigma_{1i}^2 & 0 \\ 0 & \sigma_{2i}^2 \end{pmatrix} \right\} \\ &+ (1 - \theta) \eta (1 - \eta) N \left\{ \begin{pmatrix} 0 \\ \mu_H \end{pmatrix}, \begin{pmatrix} \sigma_{1i}^2 & 0 \\ 0 & \sigma_H^2 + \sigma_{2i}^2 \end{pmatrix} \right\} \\ &+ (1 - \theta) \eta^2 N \left\{ \begin{pmatrix} \mu_M \\ \mu_H \end{pmatrix}, \begin{pmatrix} \sigma_M^2 + \sigma_{1i}^2 & 0 \\ 0 & \sigma_H^2 + \sigma_{2i}^2 \end{pmatrix} \right\} \\ &+ \{ (1 - \theta)(1 - \eta)^2 + \theta(1 - \eta) \} N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{1i}^2 & 0 \\ 0 & \sigma_{2i}^2 \end{pmatrix} \right\}. \end{aligned} \quad (2.18)$$

Because $\sigma_{1i}^2 > 0$ and $\sigma_{2i}^2 > 0$, the distribution of $(\hat{\beta}_{Mi}, \hat{\beta}_{Hi})$ has a density on $\mathbb{R}^2$,

$$\begin{aligned} f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}}(x, y) &= f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}}(x, y; \theta, \eta, \rho, \mu_M, \mu_H, \sigma_M^2, \sigma_H^2, \sigma_{1i}^2, \sigma_{2i}^2) \\ &= \theta \eta f(x, y; \mu_M, \mu_H, \sigma_M^2 + \sigma_{1i}^2, \sigma_H^2 + \sigma_{2i}^2, \rho) \\ &+ (1 - \theta) \eta (1 - \eta) f(x, y; \mu_M, 0, \sigma_M^2 + \sigma_{1i}^2, \sigma_{2i}^2, 0) \\ &+ (1 - \theta) \eta (1 - \eta) f(x, y; 0, \mu_H, \sigma_{1i}^2, \sigma_H^2 + \sigma_{2i}^2, 0) \\ &+ (1 - \theta) \eta^2 f(x, y; \mu_M, \mu_H, \sigma_M^2 + \sigma_{1i}^2, \sigma_H^2 + \sigma_{2i}^2, 0) \\ &+ \{ (1 - \theta)(1 - \eta)^2 + \theta(1 - \eta) \} f(x, y; 0, 0, \sigma_{1i}^2, \sigma_{2i}^2, 0), \end{aligned} \quad (2.19)$$

where $f(x, y; \mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$ is the bivariate normal density function with argument $(x, y)^T$, mean vector $(\mu_X, \mu_Y)^T$, variances σ_X^2 and σ_Y^2 , and correlation ρ .

Marginal Models for $\hat{\beta}_{Mi}$ and $\hat{\beta}_{Hi}$ resulting from (2.18) are less formidable,

$$\hat{\beta}_{Mi} \sim \eta N(\mu_M, \sigma_M^2 + \sigma_{1i}^2) + (1 - \eta) N(0, \sigma_{1i}^2), \quad (2.20)$$

$$\hat{\beta}_{Hi} \sim \eta N(\mu_H, \sigma_H^2 + \sigma_{2i}^2) + (1 - \eta) N(0, \sigma_{2i}^2), \quad (2.21)$$

with densities

$$\begin{aligned} f_{\hat{\beta}_{Mi}}(x) &= f_{\hat{\beta}_{Mi}}(x; \eta, \mu_M, \sigma_M^2, \sigma_{1i}^2) \\ &= \eta g(x; \mu_M, \sigma_M^2 + \sigma_{1i}^2) + (1 - \eta)g(x; 0, \sigma_{1i}^2), \end{aligned} \quad (2.22)$$

$$\begin{aligned} f_{\hat{\beta}_{Hi}}(y) &= f_{\hat{\beta}_{Hi}}(y; \eta, \mu_H, \sigma_H^2, \sigma_{2i}^2) \\ &= \eta g(y; \mu_H, \sigma_H^2 + \sigma_{2i}^2) + (1 - \eta)g(y; 0, \sigma_{2i}^2), \end{aligned} \quad (2.23)$$

where $g(z; \mu_Z, \sigma_Z^2)$ is the univariate normal density with argument z , mean μ_Z and variance σ_Z^2 . Relaxing assumption (2.5), we may refer to η in (2.20) and (2.21) as η_M and η_H , respectively. Note that θ and ρ do not appear in (2.20)-(2.23), as these parameters only have relevance to joint modeling of $\hat{\beta}_{Mi}$ and $\hat{\beta}_{Hi}$.

2.3.1 Distribution of Binary Latent Variables

The primary and secondary interests of this research are the discovery of differentially expressed human genes, using a combination of animal and human data, and the identification of useful animal genes for the prediction of human differential expression, respectively. Thus the probabilities

1. $\Pr(A_{Hi} = 1 | \hat{\beta}_{Mi}, \hat{\beta}_{Hi})$,
2. $\Pr(O_i = 1 | \hat{\beta}_{Mi}, \hat{\beta}_{Hi})$,

are of special importance. In the context of our model, $\Pr(A_{Hi} = 1 | \hat{\beta}_{Mi}, \hat{\beta}_{Hi})$ is the conditional probability that human gene i is affected by treatment (differentially expressed), given $\hat{\beta}_{Mi}$ and $\hat{\beta}_{Hi}$. Similarly, $\Pr(O_i = 1 | \hat{\beta}_{Mi}, \hat{\beta}_{Hi})$ is the conditional probability that gene pair i is an I-Ortholog, given $\hat{\beta}_{Mi}$ and $\hat{\beta}_{Hi}$.

With f denoting the bivariate normal density functions as in (2.19), routine applications of Bayes' theorem (illustrated in Appendix A.2) show that

$$\begin{aligned} \Pr(A_{Hi} = 1 | \hat{\beta}_{Mi} = x, \hat{\beta}_{Hi} = y) &= \{ (1 - \theta) \eta (1 - \eta) f(x, y; 0, \mu_H, \sigma_{1i}^2, \sigma_H^2 + \sigma_{2i}^2, 0) \\ &\quad + (1 - \theta) \eta^2 f(x, y; \mu_M, \mu_H, \sigma_M^2 + \sigma_{1i}^2, \sigma_H^2 + \sigma_{2i}^2, 0) \\ &\quad + \theta \eta f(x, y; \mu_M, \mu_H, \sigma_M^2 + \sigma_{1i}^2, \sigma_H^2 + \sigma_{2i}^2, \rho) \} / f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}}(x, y) \end{aligned} \quad (2.24)$$

and

$$\begin{aligned} \Pr(O_i = 1 | \hat{\beta}_{Mi} = x, \hat{\beta}_{Hi} = y) &= \{\theta (1 - \eta) f(x, y; 0, 0, \sigma_{1i}^2, \sigma_{2i}^2, 0) \\ &+ \theta \eta f(x, y; \mu_M, \mu_H, \sigma_M^2 + \sigma_{1i}^2, \sigma_H^2 + \sigma_{2i}^2, \rho)\} / f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}}(x, y). \end{aligned} \quad (2.25)$$

For comparative purposes, we are also interested in the conditional probability

$$\Pr(A_{Hi} = 1 | \hat{\beta}_{Hi} = y) = \frac{\eta g(y; \mu_H, \sigma_H^2 + \sigma_{2i}^2)}{f_{\hat{\beta}_{Hi}}(y)}, \quad (2.26)$$

where g denotes the univariate normal density as in (2.23). The functions (2.24-2.25) are known modulo parameters that can be estimated from the data, the topic of Section 2.4.

2.3.2 The Species Extrapolation

Here we address model (2.18) with respect to full-fledged species extrapolation, i.e., using *only* animal data to predict a treatments effect on human genes. Based on the definition of I-Ortholog (Section 2.1), when $O_i = 1$, it follows that $A_{Mi} = 1 \iff A_{Hi} = 1$ and $A_{Mi} = 0 \iff A_{Hi} = 0$. Thus if one is able to restrict attention to I-Ortholog gene pairs, only mice are necessary to learn about a treatment effect on the respective human genes. Since I-Ortholog status is uncertain and there is error associated with estimation of β_{Mi} , we are interested in the probability of human treatment effect, conditional only on mouse data:

$$\Pr(A_{Hi} = 1 | \hat{\beta}_{Mi}). \quad (2.27)$$

Before addressing (2.27), for simplicity we first work under the infinite subject sample size model (2.16), where no error is involved in estimation of β_{Mi} and (2.27) becomes $\Pr(A_{Hi} = 1 | \beta_{Mi})$. Then

$$\begin{aligned} \Pr(A_{Hi} = 1 | \beta_{Mi} = 0) &= \Pr(O_i = 1) \Pr(A_{Hi} = 1 | \beta_{Mi} = 0, O_i = 1) \\ &\quad + \Pr(O_i = 0) \Pr(A_{Hi} = 1 | \beta_{Mi} = 0, O_i = 0) \\ &= (1 - \theta)\eta, \end{aligned} \quad (2.28)$$

as $\Pr(A_{Hi} = 1 | \beta_{Mi} = 0, O_i = 1) = 0$ (Assumption 2.8) and $\Pr(A_{Hi} = 1 | \beta_{Mi} = 0, O_i = 0) = \Pr(A_{Hi} = 1 | O_i = 0) = \Pr(A_{Hi} = 1) = \eta$ (Assumption 2.5). Similarly,

$$\begin{aligned} \Pr(A_{Hi} = 1 | \beta_{Mi} \neq 0) &= \Pr(O_i = 1) \Pr(A_{Hi} = 1 | \beta_{Mi} \neq 0, O_i = 1) \\ &\quad + \Pr(O_i = 0) \Pr(A_{Hi} = 1 | \beta_{Mi} \neq 0, O_i = 0) \\ &= \theta + (1 - \theta)\eta, \end{aligned} \quad (2.29)$$

as $\Pr(A_{Hi} = 1 | \beta_{Mi} \neq 0, O_i = 1) = 1$ (Assumption 2.8).

Remaining in the optimal scenario of an infinite mouse sample size, we interpret (2.28) and (2.29). In the absence of additional information, $\Pr(A_{Hi} = 1) = \eta$. On the other hand, if there is knowledge that respective mouse gene is not affected by treatment, as in (2.28), then the conditional probability of a human treatment effect is attenuated proportional to the percentage of I-Orthologs. If there is knowledge that the respective mouse gene is affected by treatment, as in (2.29), then the updated probability of human treatment effect is increased proportional to the percentage of I-Orthologs. Note that if $\theta = 1$, then $\beta_{Mi} = 0 \Rightarrow \beta_{Hi} = 0$ and $\beta_{Mi} = 1 \Rightarrow \beta_{Hi} = 1$. Thus, as far as the modeling in this chapter is concerned, a perfect species extrapolation is achieved in the case of an infinite mouse sample size and the presence of only I-Ortholog gene pairs.

Since mouse sample sizes will be finite in practice, we return to (2.27). As $\hat{\beta}_{Mi}$ is simply β_{Mi} plus error, $\Pr(\beta_{Hi} \neq 0 | \beta_{Mi}) = \Pr(\beta_{Hi} \neq 0 | \beta_{Mi}, \hat{\beta}_{Mi})$. Thus

$$\begin{aligned}
 \Pr(\beta_{Hi} \neq 0 | \hat{\beta}_{Mi}) &= \Pr(\beta_{Hi} \neq 0, \beta_{Mi} = 0 | \hat{\beta}_{Mi}) + \Pr(\beta_{Hi} \neq 0, \beta_{Mi} \neq 0 | \hat{\beta}_{Mi}) \\
 &= \Pr(\beta_{Hi} \neq 0 | \beta_{Mi} = 0, \hat{\beta}_{Mi}) \Pr(\beta_{Mi} = 0 | \hat{\beta}_{Mi}) \\
 &\quad + \Pr(\beta_{Hi} \neq 0 | \beta_{Mi} \neq 0, \hat{\beta}_{Mi}) \Pr(\beta_{Mi} \neq 0 | \hat{\beta}_{Mi}) \\
 &= \Pr(\beta_{Mi} = 0 | \hat{\beta}_{Mi}) (1 - \theta) \eta + \Pr(\beta_{Mi} \neq 0 | \hat{\beta}_{Mi}) \{ \theta + (1 - \theta) \eta \}.
 \end{aligned} \tag{2.30}$$

Thus (2.30) is essentially an average of (2.28) and (2.29) weighted by $\Pr(\beta_{Mi} = 0 | \hat{\beta}_{Mi})$ and $\Pr(\beta_{Mi} \neq 0 | \hat{\beta}_{Mi})$, measures of uncertainty regarding mouse treatment effect status.

2.4 Estimation

In Section 2.3, the parametric model (2.16) for $(\hat{\beta}_{Mi}, \hat{\beta}_{Hi})^T$ yields bivariate mixture density

$$f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}}(x, y; \theta, \eta, \mu_M, \mu_H, \sigma_M, \sigma_H, \rho, \sigma_{1i}, \sigma_{2i}),$$

defined in (2.19). Generalizing the number of gene pairs to N , we proceed in the style of Lindsay (1988) to construct the *composite log-likelihood*

$$l(\theta, \eta, \mu_M, \mu_H, \sigma_M, \sigma_H, \rho, \sigma_{1i}, \sigma_{2i}) = \sum_{i=1}^p \log f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}}(x, y; \theta, \eta, \mu_M, \mu_H, \sigma_M, \sigma_H, \rho, \sigma_{1i}, \sigma_{2i}). \tag{2.31}$$

Composite log-likelihoods are composed of true log-likelihoods from marginal or conditional events, a property Lindsay exploits to demonstrate consistency of resulting estimators. Since dependence is likely to exist among genes, information matrix standard errors should only be accepted as approximate.

Even with the flexibility of composite likelihood methods, maximization of (2.31) is not computationally feasible given the dimensionality resulting from the gene level parameters σ_{1i} and σ_{2i} , $i = 1, \dots, N$. Consequently, we employ the pseudo maximum likelihood estimation technique of Gong and Samaniego (1981), replacing σ_{1i} and σ_{2i} in (2.31) with their least squares estimates $\hat{\sigma}_{1i}$ and $\hat{\sigma}_{2i}$ (see Appendix F for a discussion of alternate Bayesian and shrinkage estimators). Thus we wish to obtain estimates of $\theta, \eta, \mu_M, \mu_H, \sigma_M, \sigma_H, \rho$ by maximizing the pseudo composite log-likelihood

$$l^*(\theta, \eta, \mu_M, \mu_H, \sigma_M, \sigma_H, \rho) = \sum_{i=1}^N \log f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}}(x, y; \theta, \eta, \mu_M, \mu_H, \sigma_M, \sigma_H, \rho, \hat{\sigma}_{1i}, \hat{\sigma}_{2i}). \quad (2.32)$$

Gong and Samaniego report that pseudo maximum likelihood techniques produce consistent, though possibly less efficient, estimates under “fairly standard” regularity conditions.

That (2.18) is a mixture of densities yields its own set of complications. Titterton et al. (1985) warn against directly maximizing likelihoods of unrestricted mixtures of normals due to unboundedness (every data point is eligible for its own degenerate distribution, with infinite likelihood). In this context, the choice to use pseudo maximum likelihood estimation has an interesting consequence. Since $\hat{\sigma}_{1i}$ and $\hat{\sigma}_{2i}$ are guaranteed to be positive (save the case where all expression measurements within a gene are equal), degenerate components are not possible.

Regardless of likelihood unboundedness, Titterton et al. (1985) acknowledge that Newton-based updating schemes are worth considering over the popular EM algorithm due to the potential increase in speed. Based on simulations, Redner and Walker (1984) claim that an “exorbitantly large” number of EM iterations may be necessary for a very accurate numerical approximation in the case that the mixture is poorly separated. Since the current methodology looks for mixtures of (otherwise identical) distributions with and without correlation, such poor separation is likely. Redner and Walker also describe quasi-Newton methods, which do not possess the full quadratic convergence of Newton’s method but are less computationally intensive at each iteration. Proc NLP in SAS 9.1.3 has a quasi-Newton algorithm (QUANEW) that has worked quite well for maximizing (2.32) and

is the default maximization procedure for Proc NLMIXED. Stopping rules used in Proc NLP are based on a set of termination criteria as described in SAS OnlineDoc® 9.1.3.

Development and discussion of an EM algorithm for the maximization of (2.32), with comparisons to the QUANEW algorithm and applications to the GlaxoSmithKline data, is given in Appendix E. For the starting values described in Section 2.4.1, the EM algorithm reached a consistent maximum which the QUANEW algorithm often failed to attain. We offer two comments about this disagreement. First, when the QUANEW algorithm is initiated at the consistent likelihood maximum found by the EM algorithm, there is negligible change in the resulting parameter estimates. Second, no numerical maximum produced by the QUANEW algorithm ever exceeded that of the EM algorithm, but in six cases (out of sixteen) there was agreement. Since the maximization protocol of this research specifies running algorithms at each set of starting values and using the largest achieved maximum, the two algorithms effectively agree on the likelihood maximum and associated parameter estimates for the GlaxoSmithKline data. The speed of the QUANEW over the EM algorithm (often over 100 times faster with the tolerances specified in Appendix E) led to its use in the simulations of Chapter 3. In applications to actual data, however, we recommend taking the time to implement both algorithms.

2.4.1 Computational Issues

Reparameterizations

Instead of enforcing constraints on a maximization procedure, we choose to reparameterize so that numerical maximization may be carried out unconstrained. The following is a list of 1-1 transformations that were used to aid in the maximization of (2.32):

1. $\sigma_M^{\text{Tr}} = \log(\sigma_M)$
2. $\sigma_H^{\text{Tr}} = \log(\sigma_H)$
3. $\rho^{\text{Tr}} := \log(\frac{1+\rho}{1-\rho})$
4. $\eta^{\text{Tr}} := \log(\frac{\eta}{1-\eta})$
5. $\theta^{\text{Tr}} := \log(\frac{\theta}{1-\theta})$

Starting Values

This section describes calculation of starting values necessary for a numerical algorithm seeking to maximize pseudo-composite likelihood (2.32). We assume that the least squares estimates $\hat{\beta}_{Mi}$ and $\hat{\beta}_{Hi}$, $i = 1, \dots, N$, have been calculated as well as their estimated standard errors $\hat{\sigma}_{1i}$ and $\hat{\sigma}_{2i}$.

The starting values described in this section are guided by moment equations resulting from (2.18). In particular,

$$E(\hat{\beta}_{Mi}) = \eta\mu_M, \quad (2.33)$$

$$E(\hat{\beta}_{Hi}) = \eta\mu_H, \quad (2.34)$$

$$E(\hat{\beta}_{Mi}^2) = \sigma_{1i}^2 + \eta(\mu_M^2 + \sigma_M^2), \quad (2.35)$$

$$E(\hat{\beta}_{Hi}^2) = \sigma_{2i}^2 + \eta(\mu_H^2 + \sigma_H^2), \quad (2.36)$$

$$E(\hat{\beta}_{Mi}\hat{\beta}_{Hi}) = \{\theta\eta + (1 - \theta)\eta^2\}\mu_M\mu_H + \theta\eta\sigma_M\sigma_H\rho, \quad (2.37)$$

are useful for our purposes.

Starting values for η and θ . The starting values for η and θ are the sets,

$$\eta^{(0)} \in \{.1, .3, .5, .7\}, \quad (2.38)$$

$$\theta^{(0)} \in \{.1, .3, .5, .7\}. \quad (2.39)$$

Numerical maximization techniques are implemented for each combination of $\eta^{(0)}$ and $\theta^{(0)}$.

Starting values for μ_M and μ_H . Owing to moment equations (2.33) and (2.34) and starting values for η , we arrive at

$$\mu_M^{(0)} = (\eta^{(0)})^{-1} \frac{1}{N} \sum_{i=1}^N \hat{\beta}_{Mi}, \quad (2.40)$$

$$\mu_H^{(0)} = (\eta^{(0)})^{-1} \frac{1}{N} \sum_{i=1}^N \hat{\beta}_{Hi}. \quad (2.41)$$

Starting values for σ_M^2 and σ_H^2 . Moment equations (2.35) and (2.36) suggest using

$$\sigma_M^{2*} = (\eta^{(0)})^{-1} \frac{1}{N} \sum_{i=1}^N (\hat{\beta}_{Mi}^2 - \hat{\sigma}_{1i}^2) - (\mu_M^{(0)})^2, \quad (2.42)$$

$$\sigma_H^{2*} = (\eta^{(0)})^{-1} \frac{1}{N} \sum_{i=1}^N (\hat{\beta}_{Hi}^2 - \hat{\sigma}_{2i}^2) - (\mu_H^{(0)})^2, \quad (2.43)$$

as starting values, but σ_{Mi}^* and σ_{Hi}^* could potentially take negative values. Define

$$\hat{\sigma}_M^{2 \text{ Naïve}} = \frac{1}{N} \sum_{i=1}^N \left(\hat{\beta}_{Mi} - \frac{1}{N} \sum_{i=1}^N \hat{\beta}_{Mi} \right)^2, \quad (2.44)$$

$$\hat{\sigma}_H^{2 \text{ Naïve}} = \frac{1}{N} \sum_{i=1}^N \left(\hat{\beta}_{Hi} - \frac{1}{N} \sum_{i=1}^N \hat{\beta}_{Hi} \right)^2, \quad (2.45)$$

which are always positive, though less theoretically justifiable. Thus for starting values, we use

$$\sigma_M^{2(0)} = \begin{cases} \sigma_M^{2*} & \text{if } \sigma_M^{2*} > 0, \\ \hat{\sigma}_M^{2 \text{ Naïve}} & \text{if } \sigma_M^{2*} \leq 0, \end{cases} \quad (2.46)$$

and

$$\sigma_H^{2(0)} = \begin{cases} \sigma_H^{2*} & \text{if } \sigma_H^{2*} > 0, \\ \hat{\sigma}_H^{2 \text{ Naïve}} & \text{if } \sigma_H^{2*} \leq 0. \end{cases} \quad (2.47)$$

Starting values for ρ . Moment equation (2.37) suggests using

$$\rho^* = \left(\theta^{(0)} \eta^{(0)} \sqrt{\sigma_M^{2(0)} \sigma_H^{2(0)}} \right)^{-1} \left[\frac{1}{N} \sum_{i=1}^N \hat{\beta}_{Mi} \hat{\beta}_{Hi} - \left\{ \theta^{(0)} \eta^{(0)} + (1 - \theta^{(0)}) (\eta^{(0)})^2 \right\} \mu_M^{(0)} \mu_H^{(0)} \right], \quad (2.48)$$

as a starting value, but there is no guarantee that $\rho^* \in (-1, 1)$. Thus for a starting value we use

$$\rho^{(0)} = \begin{cases} -.95 & \text{if } \rho^* < -.95, \\ \rho^* & \text{if } -.95 \leq \rho^* \leq .95, \\ .95 & \text{if } \rho^* > .95. \end{cases}$$

This completes the discussion on starting values.

2.4.2 Estimation Summary

The following list describes the estimation process on a step-by-step basis.

1. Compute least squares estimates of $\hat{\beta}_{Mi}, \hat{\beta}_{Hi}, \hat{\sigma}_{1i}^2, \hat{\sigma}_{2i}^2, i = 1, \dots, N$, from mouse and human gene expression data.
2. Compute starting values for $(\mu_M, \mu_H, \sigma_M, \sigma_H, \eta, \theta, \rho)$ using the rules presented in (2.4.1).

3. Using starting values from Step 2, numerically maximize (2.32) in terms of the transformed parameters described in (2.4.1).
4. Transform parameter estimates back to their original scale.

2.5 Decision Rules for Active Genes and Orthologs

Following the parameter estimation described in Section 2.4 and referring to the conditional distributions of A_{Hi} and O_i given the data (discussed in Section 2.3.1), consider the random triple $(A_{Hi}, \hat{\beta}_{Mi}, \hat{\beta}_{Hi})$. Define

$$\begin{aligned} p_i &= \Pr(A_{Hi} = 1 | \hat{\beta}_{Mi}, \hat{\beta}_{Hi}) \\ &= p_i(\hat{\beta}_{Mi}, \hat{\beta}_{Hi}, \mu_M, \mu_H, \sigma_M, \sigma_H, \eta, \theta, \rho, \sigma_{1i}, \sigma_{2i}) \end{aligned} \quad (2.49)$$

and let

$$\hat{p}_i = p_i(\hat{\beta}_{Mi}, \hat{\beta}_{Hi}, \hat{\mu}_M, \hat{\mu}_H, \hat{\sigma}_M, \hat{\sigma}_H, \hat{\eta}, \hat{\theta}, \hat{\rho}, \hat{\sigma}_{1i}, \hat{\sigma}_{2i}). \quad (2.50)$$

Similarly, define

$$\begin{aligned} q_i &= \Pr(O_i = 1 | \hat{\beta}_{Mi}, \hat{\beta}_{Hi}) \\ &= q_i(\hat{\beta}_{Mi}, \hat{\beta}_{Hi}, \mu_M, \mu_H, \sigma_M, \sigma_H, \eta, \theta, \rho, \sigma_{1i}, \sigma_{2i}) \end{aligned} \quad (2.51)$$

and let

$$\hat{q}_i = q_i(\hat{\beta}_{Mi}, \hat{\beta}_{Hi}, \hat{\mu}_M, \hat{\mu}_H, \hat{\sigma}_M, \hat{\sigma}_H, \hat{\eta}, \hat{\theta}, \hat{\rho}, \hat{\sigma}_{1i}, \hat{\sigma}_{2i}). \quad (2.52)$$

Thus, (2.50) is the estimated probability that human gene i is a differentially expressed human gene, and (2.52) is the estimated probability that gene pair i is an I-Ortholog, given the mouse and human gene expression data.

Refer to Section 1.2.2 for the *direct posterior probability approach* to false discovery rate (FDR) control. It is possible to make decisions regarding the hypotheses

$$\begin{aligned} H_{0i}^A : A_{Hi} &= 0, \\ H_{0i}^O : O_i &= 0, \end{aligned}$$

versus the alternatives

$$\begin{aligned} H_{1i}^A : A_{Hi} &= 1, \\ H_{1i}^O : O_i &= 1, \end{aligned}$$

which approximately control the FDR at α . Define

$$\kappa_A = \max \left\{ \kappa : \frac{\sum_{i=1,\dots,p} (1 - \hat{p}_i) \mathbf{I}\{(1 - \hat{p}_i) < \kappa\}}{\sum_{i=1,\dots,p} \mathbf{I}\{(1 - \hat{p}_i) < \kappa\}} \leq \alpha \right\}, \quad (2.53)$$

and

$$\kappa_O = \max \left\{ \kappa : \frac{\sum_{i=1,\dots,p} (1 - \hat{q}_i) \mathbf{I}\{(1 - \hat{q}_i) < \kappa\}}{\sum_{i=1,\dots,p} \mathbf{I}\{(1 - \hat{q}_i) < \kappa\}} \leq \alpha \right\}. \quad (2.54)$$

For approximate (model-based) control of the FDR at level α ,

- reject H_{0i}^A in favor of H_{1i}^A for all genes with $(1 - \hat{p}_i) < \kappa_A$,
- reject H_{0i}^O in favor of H_{1i}^O for all genes with $(1 - \hat{q}_i) < \kappa_O$.

2.6 An Example Application

From the GlaxoSmithKline data, the least squares estimates $(\hat{\beta}_{Hi}, \hat{\beta}_{Mi})$ and the t -statistics $(\hat{\beta}_{Hi}/\hat{\sigma}_{2i}, \hat{\beta}_{Mi}/\hat{\sigma}_{1i})$, $i = 1, \dots, 9683$, are shown in Figure 2.3. Note that the mouse

Figure 2.3: Mouse and human treatment effect statistics from the GlaxoSmithKline data

t -statistics tend to be larger than those of the humans, indicating the higher signal-to-noise ratio of the preclinical experiment. In the plot of $(\hat{\beta}_{Hi}, \hat{\beta}_{Mi})$, a limited number of extreme observations receive disproportionate visual weight, and in later plots the window is altered (see Figures 2.4, 2.6).

Joint Model (2.18) and marginal models (2.20) and (2.21) were fit to the Glaxo-SmithKline data via the estimation methodology described in Section 2.4. For a thorough

description of the numerical maximization of (2.32), including the relative performance of the quasi-Newton and EM algorithms, refer to Appendix E. For all three models, estimates along with their 95% Wald confidence intervals are presented in Tables 2.2-2.3, though recall that the composite likelihood method provides only approximate standard errors and confidence intervals. Browne and Draper (2006) have furthermore warned against using Wald intervals for variance components in multilevel models.

Table 2.2: Estimates and Wald Confidence intervals for model (2.18).

Parameter	Joint Model	
	estimate	95% C.I.
μ_M	-2.71	(-4.55, -.86)
μ_H	2.83	(2.44, 3.21)
σ_M	55.28	(52.67, 58.02)
σ_H	6.77	(6.27, 7.30)
ρ	.934	(.697, .987)
η	.446	(.431, .461)
θ	.165	(.114, .231)

The higher average signal-to-noise ratio in the preclinical experiment is arguably manifested in $\hat{\sigma}_M$, which is over eight times larger than $\hat{\sigma}_H$. Since μ_M (μ_H) is small, a higher value of σ_M (σ_H) corresponds to situations where β_{Mi} (β_{Hi}) takes larger values with higher probability, and thus σ_M and σ_H are themselves measures of signal.

The estimate of correlation $\hat{\rho}$ is close to one (though the 95% confidence interval cannot rule out values as low as .7) and $\hat{\theta} = .165$, a suggestion that approximately seventeen percent of the proposed orthologs in the Mouse Genome Informatics (MGI) database are I-Orthologs whose latent differential-expression coefficients tightly covary about a line. Combined with the joint estimate of η at .446, approximately 7.4% of gene pairs are estimated to be differentially expressed I-Orthologs ($.446 \times .165 = .074$). These genes are the most interesting as they are expected to have increased detection probabilities due to I-Orthology-based species extrapolation.

While the mean and variance estimates from marginal models (2.20) and (2.21) are remarkably consistent with the joint estimates from model (2.18), the mixing parameter estimates $\hat{\eta}_M$ and $\hat{\eta}_H$ deserve discussion. Referring to the t -statistics from Figure 2.3, it is remarkable that the greater signal-to-noise ratio of the mouse data over human data is reflected in estimates of σ_M and σ_H , instead of the inherently signal-relevant η_H and η_M .

Table 2.3: Estimates and Wald Confidence intervals for marginal models (2.20) and (2.21).

Marginal Mouse			Marginal Human		
Parameter	estimate	95% CI	Parameter	estimate	95% CI
μ_M	-2.75	(-4.70, -.81)	μ_H	2.48	(2.16, 2.81)
σ_M	57.07	(54.30, 59.98)	σ_H	5.50	(5.08, 5.95)
η_M	.428	(.413, .444)	η_H	.560	(.514, .605)

The estimate of η_H has a larger confidence interval than η_M , suggesting that the former may be an inferior estimate and explaining why the joint estimate of η is influenced so little by the human data.

Given the estimates from Table 2.2, the estimated posterior probabilities of human differential expression, $\hat{p}_i$, and I-Orthology, $\hat{q}_i$ (Section 2.3.1), may be computed for $i = 1, \dots, 9683$. The $\hat{p}_i$ s have a 5-number summary (minimum, 25th percentile, median, 75th percentile, maximum) of (.01, .36, .44, .53, 1.00) and are generally much larger than the $\hat{q}_i$ s, with a 5-number summary of (.00, .15, .17, .18, .78). In fact, as the maximum probability of I-Orthology is less than .8, no I-Ortholog discoveries can be made even with a false discovery rate of .20.

In Figure 2.4, the $(\hat{\beta}_{Hi}, \hat{\beta}_{Mi})$ and $(\hat{\beta}_{Hi}/\hat{\sigma}_{2i}, \hat{\beta}_{Mi}/\hat{\sigma}_{1i})$ pairs are colored red for human differential expression discoveries with .05 FDR control ($\hat{p}_i > .886$), yellow for additional discoveries with .20 FDR control ($.643 \leq \hat{p}_i < .886$) and green otherwise. While the $\hat{p}_i$ s are consistent with traditional insight based on the t -statistics $(\hat{\beta}_{Hi}/\hat{\sigma}_{2i}, \hat{\beta}_{Mi}/\hat{\sigma}_{1i})$, the lack of a coherent pattern in the $\hat{p}_i$ -coded plot of $(\hat{\beta}_{Hi}, \hat{\beta}_{Mi})$ is indicative of the wide range of variances associated with estimates of human differential expression. Often the largest values of $\hat{\beta}_{Hi}$ ($\hat{\beta}_{Mi}$) also correspond to the largest values of $\hat{\sigma}_{1i}$ ($\hat{\sigma}_{2i}$), resulting in low t -statistics for differential expression.

In Figure 2.6, the $(\hat{\beta}_{Hi}, \hat{\beta}_{Mi})$ and $(\hat{\beta}_{Hi}/\hat{\sigma}_{2i}, \hat{\beta}_{Mi}/\hat{\sigma}_{1i})$ pairs are colored red when $\hat{q}_i > 2\hat{\theta}$, yellow when $\hat{\theta} \leq \hat{q}_i < 2\hat{\theta}$, and green otherwise. The red points in the plot of $(\hat{\beta}_{Hi}, \hat{\beta}_{Mi})$, the highest observed probabilities of I-Orthology, appear to highlight the hypothesized relationship shown in Figure 2.1. The yellow points, the observed probabilities at least greater than the prior estimate, $\hat{\theta}$, appear to comprise all points where mouse and human treatment effects behaved consistently at all, i.e. the points in quadrants I and III. The plot of color coded t -statistics has an obvious interpretation: higher values of $\hat{q}_i$ exist where there is more evidence of dual treatment effect, but also there are a few red points

Figure 2.4: Pairs $(\hat{\beta}_{Hi}, \hat{\beta}_{Mi})$ and $(\hat{\beta}_{Hi}/\hat{\sigma}_{2i}, \hat{\beta}_{Mi}/\hat{\sigma}_{1i})$ are color coded by $\hat{p}_i$, the posterior probability of treatment effect as follows: red for $\hat{p}_i > .886$ (all .05 FDR discoveries), yellow for $.643 \leq \hat{p}_i < .886$ (all additional .20 FDR discoveries) and green otherwise.

that do not conform to the overall pattern, indicating the impact lying on the “line” in the plot of $(\hat{\beta}_{Hi}, \hat{\beta}_{Mi})$.

2.6.1 Discovery and Methodology Comparisons

In this section we reflect on the differentially expressed genes found using joint model (2.18), human marginal model (2.21), and the procedure of Benjamini and Hochberg (1995). One important change was made in use of the human marginal model: *when estimating the marginal differential expression probabilities based on (2.26), the estimate of η from the joint model, $\hat{\eta} = .446$, is used instead of the marginally estimated $\hat{\eta}_H = .560$.* Increases in mixing parameter η (or η_H) will uniformly raise posterior probabilities, but gene rankings remain unchanged. Since the goodness of $\hat{\eta}_H$ is in question, we decided to use a single estimate of η for computation of joint and marginal probabilities, in order to more easily compare procedures.

Table 2.4 reports the numbers of differentially expressed human genes discovered using competing methodologies. Note that the popular frequentist procedure of Benjamini and Hochberg (1995) yields far fewer discoveries than the two posterior probability-based procedures. The joint model, presumably because of information obtained from the animal data, discovers more genes than its marginal counterpart.

Figure 2.5: Pairs $(\hat{\beta}_{Hi}, \hat{\beta}_{Mi})$ and $(\hat{\beta}_{Hi}/\hat{\sigma}_{2i}, \hat{\beta}_{Mi}/\hat{\sigma}_{1i})$ are color coded by $\hat{q}_i$, the posterior probability of I-Orthology as follows: red for $\hat{q}_i > .329$ (twice $\hat{\theta}$), yellow for $.165 \leq \hat{q}_i < .329$ (at least $\hat{\theta}$) and green otherwise.

Table 2.4: Numbers of Discovered Genes for the various methodologies, where “Joint” refers to key model (2.18), “Marginal” refers to human marginal model (2.21), and “BH” refers to the method of Benjamini and Hochberg (1995).

FDR	Joint	Marginal	BH
.05	287	259	2
.20	1056	968	179

From Figure 2.6, various agreements and disagreements between the competing procedures are apparent. Discounting number of discoveries, differentially expressed genes found by the procedure of Benjamini and Hochberg are generally found by the posterior probability-based methodologies. The joint and marginal model also agree for a vast majority of discovered genes, but both procedures are associated with sizeable numbers of genes not discovered by the other. The genes discovered by the joint model and not the marginal model are especially interesting: at nominal FDR .05 (.20), over 68% (75%) of these genes have posterior I-Ortholog probabilities greater than the estimated prior value of $\hat{\theta} = .165$ and over 45% (22%) have probabilities greater than twice $\hat{\theta}$. For comparison, recall that the 75th percentile of I-Ortholog probabilities ($\hat{q}_i$) in the GlaxoSmithKline data is .18. We refer to power increases due to relatively high I-Ortholog probabilities as *I-Ortholog lifts*.

Figure 2.6: Venn diagrams showing the relationships between the various methodologies in terms of human genes deemed differentially expressed.

Chapter 3

Simulation

In this chapter we describe a simulation study that investigates the sampling properties of $\hat{\mu}_M, \hat{\mu}_H, \hat{\sigma}_M, \hat{\sigma}_H, \hat{\theta}, \hat{\eta}, \hat{\rho}$, assesses the power and error-control of the dual-species joint model (2.18) for detecting differentially expressed human genes, and determines the feasibility of finding I-Ortholog mouse and human gene pairs (Section 2.1). Each of these endpoints is observed relative to various parameter and sample size modulations. The primary objective of this experiment is to learn about the differential gene-finding performance of the joint model relative to competing procedures which use only the human data, in the hope that the addition of animal gene expression data allows for meaningful improvements. The ability to find I-Orthologs is a secondary but exciting endpoint; succeeding in this task may illuminate useful and unknown genetic extrapolators from animals to humans.

This chapter discusses the design, implementation, and analysis of a simulation experiment designed to investigate the performance of the methodology developed in this research. The layout of the chapter is as follows. Section 3.1 reviews our modeling of the GlaxoSmithKline data and discusses other features to be emulated in the simulation data. Section 3.2 describes the process of generating the simulated data in steps amenable to standard statistical packages. Section 3.3 describes the design of four separate sub-experiments designed to answer various questions of interest. Section 3.4, in conjunction with tables and figures from Appendix G, presents the results of the four experiments.

3.1 A Description of the Simulated Data

Owing to the unique nature of GlaxoSmithKline's Integrative Biology Initiative, data from the experiment described in Section 1.4 will serve as a template for the data generated in this simulation study. Generalizing the numbers of humans, mice and proposed orthologs, our simulation data consists of

- $\mathbf{Y}_H$, a $N \times n$ human gene expression matrix consisting of gene expression measurements of N genes from each of n humans,
- $\mathbf{T}_H$, an $n \times 1$ vector consisting of human treatment indicator variables,
- $\mathbf{Z}_k$, a $N \times 5$ matrix of covariate data for each of $k = 1, \dots, n$ humans,
- $\mathbf{Y}_M$, a $N \times m$ a mouse gene expression matrix that consists of gene expression measurements of N genes from each of m mice, and
- $\mathbf{T}_M$, an $m \times 1$ a vector of mouse treatment indicator variables.

The gene expression responses have the following distributional properties:

$$Y_{Hik} \mid T_{Hk}, \mathbf{Z}_{ik}, \alpha_{Hi}, \beta_{Hi}, \sigma_{Hi}, \gamma_i \sim N(\alpha_{Hi} + \beta_{Hi}T_{Hk} + \mathbf{Z}_{ik}\gamma_i, \sigma_{Hi}^2),$$

for $i = 1, \dots, N$ and $k = 1, \dots, n$, and

$$Y_{Mij} \mid T_{Mj}, \alpha_{Mi}, \beta_{Mi}, \sigma_{Mi} \sim N(\alpha_{Mi} + \beta_{Mi}T_{Mj}, \sigma_{Mi}^2),$$

for $i = 1, \dots, N$ and $j = 1, \dots, m$. In Section 2.2.2, the distributional model (2.16) was developed for the latent parameter vector (β_{Mi}, β_{Hi}) and is reprinted below:

$$\begin{aligned} \begin{pmatrix} \beta_{Mi} \\ \beta_{Hi} \end{pmatrix} &\sim \theta \eta N \left\{ \begin{pmatrix} \mu_M \\ \mu_H \end{pmatrix}, \begin{pmatrix} \sigma_M^2 & p \sigma_H \sigma_M \\ p \sigma_H \sigma_M & \sigma_H^2 \end{pmatrix} \right\} \\ &+ (1 - \theta) \eta (1 - \eta) N \left\{ \begin{pmatrix} \mu_M \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_M^2 & 0 \\ 0 & 0 \end{pmatrix} \right\} \\ &+ (1 - \theta) \eta (1 - \eta) N \left\{ \begin{pmatrix} 0 \\ \mu_H \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & \sigma_H^2 \end{pmatrix} \right\} \\ &+ (1 - \theta) \eta^2 N \left\{ \begin{pmatrix} \mu_M \\ \mu_H \end{pmatrix}, \begin{pmatrix} \sigma_M^2 & 0 \\ 0 & \sigma_H^2 \end{pmatrix} \right\} \\ &+ \{(1 - \theta)(1 - \eta)^2 + \theta(1 - \eta)\} N(\mathbf{0}, \mathbf{0}), \end{aligned}$$

for $i = 1, \dots, N$.

From linear models (2.1) and (2.2) describing mouse and human gene expression data at its most basic level, the latent differential gene expression parameters (β_{Mi}, β_{Hi}) , $i = 1, \dots, N$, are of primary importance. The remaining gene-level parameters α_{Mi} , σ_{Mi} , α_{Hi} , σ_{Hi} , γ_i , $i = 1, \dots, N$, are still required for data generation despite the fact that they have not been assigned distributional models. This second class of parameters has interesting characteristics in the GlaxoSmithKline data that, when emulated, greatly influence the simulated data.

Recall from Section 1.2.2 that Lewin et al. (2006) model parameters analogous to σ_{Mi}^2 and σ_{Hi}^2 with the lognormal distribution. The least squares estimates $\hat{\sigma}_{Mi}^2$ and $\hat{\sigma}_{Hi}^2$ from the GlaxoSmithKline data also loosely follow this model (see Figure 3.1).

Figure 3.1: Histograms of $\log \hat{\sigma}_{Mi}^2$ and $\log \hat{\sigma}_{Hi}^2$ and Estimated Normal Densities (via Least Squares)

Generating (β_{Mi}, β_{Hi}) from (2.16) and $(\sigma_{Mi}^2, \sigma_{Hi}^2)$ from independent lognormal models, however, yields data that are grossly inconsistent with the GlaxoSmithKline data. This is due to cases where β_{Mi} (β_{Hi}) is large (in an absolute sense) and σ_{Mi}^2 (or σ_{Hi}^2) is close to zero. In these cases, the observed t -statistics for tests of $H_0 : \beta_{Mi} = 0$ ($H_0 : \beta_{Hi} = 0$) stretch into the hundreds or even thousands. This is quite unlike the GlaxoSmithKline data, where the t -statistics are rarely larger than 10.

Solving the large- β /small- σ problem means incorporating the dependence between expression variability (measurement or otherwise) and magnitude of treatment effect. The

claim, based on observations in the GlaxoSmithKline data, is that genes with higher natural expression variabilities tend to experience greater absolute changes in mean expression when there is a treatment effect. In the following parallel boxplots of $\log \hat{\sigma}_{Mi}^2$ and $\log \hat{\sigma}_{Hi}^2$ over ordered bins of $\hat{\beta}_{Mi}$ and $\hat{\beta}_{Hi}$, (Figure 3.2, a technique chosen because of the high data density), the bins with the smallest median values of $\log \hat{\sigma}_{Mi}^2$ and $\log \hat{\sigma}_{Hi}^2$ contain the values of $\hat{\beta}_{Mi}$ and $\hat{\beta}_{Hi}$ closest to zero.

Figure 3.2: Boxplots of $\log \hat{\sigma}_{Mi}^2$ and $\log \hat{\sigma}_{Hi}^2$ for ordered groupings of $\hat{\beta}_{Mi}$ and $\hat{\beta}_{Hi}$, respectively, with 300 observations per bin.

A crude but effective methodology was chosen to reproduce relationships between σ_{Mi}^2 and β_{Mi} (and σ_{Hi}^2 and β_{Hi}) as observed in the GlaxoSmithKline data. After generating β_{Mi} and β_{Hi} , described in Section 3.2.1, values of σ_{Mi}^2 and σ_{Hi}^2 are generated from an approximate empirical distribution of $\sigma_{Mi}^2|\beta_{Mi}$ and $\sigma_{Hi}^2|\beta_{Hi}$, respectively, described in Section 3.2.3.

Additionally, a relationship between a baseline expression, α_{Mi} and α_{Hi} , and expression variability, σ_{Mi} and σ_{Hi} , was observed in the GlaxoSmithKline data. The relationship between the estimates is roughly loglinear (see Figure 3.3), and given generated values

Figure 3.3: The baseline expression vs. variability relationship for both mice and humans.

of σ_{Mi}^2 and σ_{Hi}^2 , α_{Mi} and α_{Hi} are created according to:

$$\alpha_{Mi} = \exp\{1.10918 + 0.56964 * \log(\sigma_{Mi}^2)\}, \quad (3.1)$$

$$\alpha_{Hi} = \exp\{0.57881 + 0.57797 * \log(\sigma_{Hi}^2)\}. \quad (3.2)$$

For simplicity, the regression coefficients $\gamma_{1i} - \gamma_{5i}$ for the human covariates are set to constants defined by $\hat{\sigma}_{Hi}^{-2}$ -weighted means of $\hat{\gamma}_{1i} - \hat{\gamma}_{5i}$ (appearing in Table 3.1). Furthermore, simulation covariates $Z_{ik1} - Z_{ik5}$ come directly from human subjects in the GlaxoSmithKline data. This is accomplished via sampling with replacement, described fully in Section 3.2.5.

Table 3.1: The following are $\hat{\sigma}_{Hi}^{-2}$ -weighted means of $\hat{\gamma}_{1i} - \hat{\gamma}_{5i}$ from the GlaxoSmithKline data, used to set the simulation parameters $\gamma_{1i} - \gamma_{5i}, i = 1, \dots, N$.

GSK Data Estimate	$\hat{\gamma}_{1i}$	$\hat{\gamma}_{2i}$	$\hat{\gamma}_{3i}$	$\hat{\gamma}_{4i}$	$\hat{\gamma}_{5i}$
Weighted Mean	-0.0328488	0.5269950	-0.1330379	-0.1162620	-0.2716631

3.1.1 Factors

The simulated data is subject to the modulation of the following quantities:

1. N : The number of ortholog pairs;
2. n_t : The human treatment sample size;
3. n_p : The human placebo sample size;
4. m_t : The mouse treatment sample size;
5. m_p : The mouse placebo sample size;
6. θ : The proportion of I-Orthologs;
7. η : The proportion of genes affected by the treatment;
8. μ_M : Mean of the non-zero mouse treatment coefficients;
9. μ_H : Mean of the non-zero human treatment coefficients;
10. σ_M : Standard deviation of the non-zero mouse treatment coefficients;
11. σ_H : Standard deviation of the non-zero human treatment coefficients;
12. ρ : The correlation between treatment coefficients for the orthologs.

The parameters θ, ρ and η are arguably the most meaningful. Gains in the power to detect differentially expressed human genes from the presence of animal data logically depend on both θ , the proportion of mouse-exploitable I-Orthologs, and ρ , the strength of the linear relationship to be exploited. The mixing parameter η may be an equally important parameter, as it answers the question of how many differentially expressed genes are available to be found. In both the marginal and joint cases, larger values of η correspond to larger posterior probabilities of differential expression and thus more genes deemed significant. Furthermore, larger values of η logically correspond to greater powers to detect

I-Orthologs, since such gene pairs can only be identified in the presence of a treatment effect.

Of secondary interest are parameters μ_M , μ_H , σ_M and σ_H . The latent means μ_M and μ_H describe a treatment's tendency to up or down regulate genes. Standard deviations σ_M and σ_H of the latent quantities β_{Mi} and β_{Hi} are somewhat intriguing, as they simultaneously function as descriptors of error and signal. When μ_M and μ_H are close to zero (as might be expected), higher variabilities of β_{Mi} and/or β_{Hi} suggest that there are more genes with larger absolute values of β_{Mi} and/or β_{Hi} which (all other factors equal) correspond to more easily found differentially expressed genes.

Finally, there is interest in how the sample dimensions N, n_t, n_p, m_t, m_p relate to performance in the estimation of unknown parameters and the discovery of differentially expressed genes and I-Orthologs. This is complicated by the fact that there are two distinct categories of sample sizes: N , the gene pair sample size, and n_t, n_p, m_t, m_p , the subject sample sizes. While the subject sample sizes may be increased indefinitely in practice, N ($\approx 10,000$) depends on current biological knowledge of mouse and human orthology and is more or less fixed at any given point in time (though the analyst need not be restricted to any particular set of gene pairs). Regardless, N is a valid sample size of models (2.16) and (2.18) and will be modulated alongside n_t, n_p, m_t, m_p in the simulation experiments.

3.2 Simulation Implementation

This section was created to provide step-by-step instructions for creating the simulated data described in this chapter. Aside from a standard statistical package, this simulation requires inputs from the GlaxoSmithKline data (Section 1.4).

3.2.1 Generation of (β_{Mi}, β_{Hi})

The user-defined factors μ_M , μ_H , σ_M , σ_H , ρ , η , θ and gene pair sample size N (Section 3.1.1) comprise all necessary information to generate the latent quantities (β_{Mi}, β_{Hi}) , $i = 1, \dots, N$. For convenience, consider the following random variable representation of model (2.16),

$$\begin{pmatrix} \beta_{Mi} \\ \beta_{Hi} \end{pmatrix} \sim \sum_{j=1}^5 \pi_j N(\boldsymbol{\mu}_j, \mathbf{V}_j), \quad (3.3)$$

where $j = 1, \dots, 5$. The 2×1 vectors $\boldsymbol{\mu}_j$, the 2×2 matrices $\mathbf{V}_j$ and the scalars π_j are the mean vectors, covariance matrices and mixing probabilities, respectfully. Furthermore, a random variable representation of (3.3) is

$$\begin{pmatrix} \beta_{Mi} \\ \beta_{Hi} \end{pmatrix} = \sum_{j=1}^5 T_{ij} \mathbf{W}_{ij}, \quad (3.4)$$

where $(T_{i1}, \dots, T_{i5}) \sim \text{Multinomial}(n = 1, \pi_1, \dots, \pi_5)$, with $n = 1$ independent trial, and $W_{ij} \sim N(\boldsymbol{\mu}_j, \mathbf{V}_j)$.

For notational convenience, the subscript i will be left out of the following expressions and discussion is based upon an arbitrary gene pair. For $j = 1, \dots, 5$, the normal random variable W_j is generated as

$$W_j = \boldsymbol{\mu}_j + \begin{pmatrix} \sqrt{v_{j11}} & 0 \\ a_j & b_j \end{pmatrix} \begin{pmatrix} \epsilon_{j1} \\ \epsilon_{j2} \end{pmatrix}, \quad (3.5)$$

where $\epsilon_{j1}, \epsilon_{j2}$ are i.i.d. $N(0, 1)$ random variables and v_{jkl} is the kl^{th} element of $\mathbf{V}_j$, and

$$a_j = \begin{cases} 0 & \text{if } v_{j11} \leq 0, \\ \frac{v_{j12}}{\sqrt{v_{j11}}} & \text{if } v_{j11} > 0, \end{cases}$$

$$b_j = \sqrt{v_{j22} - \frac{v_{j12}^2}{v_{j11}}}.$$

Generate a uniform random variable u . Defining $\pi_0 = 0$, let

$$T_j = I\left(\sum_{k=0}^{j-1} \pi_k \leq u < \sum_K^j \pi_k\right), \quad j = 1, \dots, 5,$$

where $I(\cdot)$ is the indicator function. Since T_j will be zero for all but a single j , let j^* be the value of j for which $T_j = 1$. Generate i.i.d. $N(0, 1)$ random variables $\epsilon_{j^*1}, \epsilon_{j^*2}$ and let

$$\mathbf{W}_{j^*} = \boldsymbol{\mu}_{j^*} + \begin{pmatrix} \sqrt{v_{j^*11}} & 0 \\ a_{j^*} & b_{j^*} \end{pmatrix} \begin{pmatrix} \epsilon_{j^*1} \\ \epsilon_{j^*2} \end{pmatrix}.$$

Following from (3.4), let $(\beta_{Mi}, \beta_{Hi})^T = \mathbf{W}_{j^*}$.

3.2.2 Generating (O_i, A_{Mi}, A_{Hi})

While not part of the observed data, we include the values of O_i , A_{Mi} and A_{Hi} in our simulation's output dataset, as knowledge of whether $A_{Hi} = 1$ and $O_i = 1$ is necessary for determining correct classification of differentially expressed human genes and I-orthologs. The vector (O_i, A_{Mi}, A_{Hi}) takes six values with positive probability (Table 2.1), four of which are completely determined by the multinomial random vector $\mathbf{T}_i = (T_{i1}, \dots, T_{i5})$, defined in Section 3.2.1. The mapping is illustrated below:

$$(O_i, A_{Mi}, A_{Hi}) = \begin{cases} (1, 1, 1) & \text{if } T_{i1} = 1, \\ (0, 1, 0) & \text{if } T_{i2} = 1, \\ (0, 0, 1) & \text{if } T_{i3} = 1, \\ (0, 1, 1) & \text{if } T_{i4} = 1. \end{cases}$$

Recall from Section 2.1 that when $T_{i5} = 1$ ($A_{Mi} = A_{Hi} = 0$), O_i is indeterminate. Thus for the cases where $T_{i5} = 1$, generate O_i from a Bernoulli(κ) distribution, where

$$\begin{aligned} \kappa = \Pr(O_i = 1 | A_{Mi} = 0, A_{Hi} = 0) &= \frac{\Pr(O_i = 1, A_{Mi} = 0, A_{Hi} = 0)}{\sum_{k=0}^1 \Pr(O_i = k, A_{Mi} = 0, A_{Hi} = 0)} \\ &= \frac{\theta(1 - \eta)}{(1 - \theta)(1 - \eta)^2 + \theta(1 - \eta)}. \end{aligned} \quad (3.6)$$

That is, generate the uniform random variable u_{κ_i} and define $(O_i, A_{Mi}, A_{Hi}) = (1, 0, 0)$ if $u_{\kappa_i} < \kappa$ and $(O_i, A_{Mi}, A_{Hi}) = (0, 0, 0)$ otherwise. This concludes the generation of the latent variables $O_i, A_{Mi}, A_{Hi}, i = 1, \dots, N$.

3.2.3 Generating σ_{Mi} and σ_{Hi}

Appealing to Figure 3.2, there are relationships between σ_{Mi} and β_{Mi} and σ_{Hi} and β_{Hi} that must be addressed in order to produce realistic differential expression t -statistics for mouse and human genes. Rather than modeling conditional means and variances, σ_{Mi} and σ_{Hi} are simulated by essentially taking random draws from the boxplots of Figure 3.2, with minor differences. First, the differential expression estimates $\hat{\beta}_{Mi}$ and $\hat{\beta}_{Hi}$ from the GlaxoSmithKline data are ordered in ascension and systematically grouped into bins of 100. The final (97th) bin includes only 83 gene pairs as the sample size, 9683, is not divisible by 100. With a simulated β_{Mi} and β_{Hi} in hand (from Section 3.2.1), attention is restricted to the particular GlaxoSmithKline bin of 100 (or 83) genes (separately for both

species). In these restricted subsets of data, σ_{Mi} and σ_{Hi} are randomly drawn from an equal-probability mixture of four uniform distributions, with ranges based on the following bin-restricted percentiles of $\hat{\sigma}_{Mi}$ and $\hat{\sigma}_{Hi}$: $[5^{th}, 25^{th}]$, $[25^{th}, 50^{th}]$, $[50^{th}, 75^{th}]$, $[75^{th}, 95^{th}]$. The 5th and 95th percentiles are used instead of the minimum and maximum in an attempt to avoid simulating the very large variances that sometimes occur in the GlaxoSmithKline data.

The following is a more detailed description of the generation of σ_{Mi} and σ_{Hi} . With $\hat{\beta}_{Mi}$ and $\hat{\beta}_{Hi}$ from the GlaxoSmithKline data grouped in an ascending manner into the 97 bins, compute the 5th, 25th, 50th, 75th and 95th percentiles of $\hat{\sigma}_{Mi}$ and $\hat{\sigma}_{Hi}$, denoted $p_{M05j}, p_{M25j}, p_{M50j}, p_{M75j}, p_{M95j}$ and $p_{H05j}, p_{H25j}, p_{H50j}, p_{H75j}, p_{H95j}$, for all bins $j = 1, \dots, 97$. For a given i , and simulated β_{Mi} and β_{Hi} , define

$$j_{Mi}^* = \begin{cases} 1 & \text{if } \beta_{Mi} < \min\{\hat{\beta}_{Mi} \in \text{bin } 1\}, \\ j & \text{if } \min\{\hat{\beta}_{Mi} \in \text{bin } j\} < \beta_{Mi} \leq \min\{\hat{\beta}_{Mi} \in \text{bin } j+1\}, \\ 97 & \text{if } \beta_{Mi} \geq \min\{\hat{\beta}_{Mi} \in \text{bin } 97\}, \end{cases}$$

and similarly,

$$j_{Hi}^* = \begin{cases} 1 & \text{if } \beta_{Hi} < \min\{\hat{\beta}_{Hi} \in \text{bin } 1\}, \\ j & \text{if } \min\{\hat{\beta}_{Hi} \in \text{bin } j\} < \beta_{Hi} \leq \min\{\hat{\beta}_{Hi} \in \text{bin } j+1\}, \\ 97 & \text{if } \beta_{Hi} \geq \min\{\hat{\beta}_{Hi} \in \text{bin } 97\}. \end{cases}$$

Generate four i.i.d. uniform random variables u_{M1i}, u_{M2i} and u_{H1i}, u_{H2i} . Let

$$\sigma_{Mi} = \begin{cases} p_{M05j_{Mi}^*} + (p_{M25j_{Mi}^*} - p_{M05j_{Mi}^*}) * u_{M2i} & \text{if } u_{M1i} \leq .25, \\ p_{M25j_{Mi}^*} + (p_{M50j_{Mi}^*} - p_{M25j_{Mi}^*}) * u_{M2i} & \text{if } .25 < u_{M1i} \leq .50, \\ p_{M50j_{Mi}^*} + (p_{M75j_{Mi}^*} - p_{M50j_{Mi}^*}) * u_{M2i} & \text{if } .50 < u_{M1i} \leq .75, \\ p_{M75j_{Mi}^*} + (p_{M95j_{Mi}^*} - p_{M75j_{Mi}^*}) * u_{M2i} & \text{if } u_{M1i} > .75, \end{cases}$$

and similarly

$$\sigma_{Hi} = \begin{cases} p_{H05j_{Hi}^*} + (p_{H25j_{Hi}^*} - p_{H05j_{Hi}^*}) * u_{H2i} & \text{if } u_{H1i} \leq .25, \\ p_{H25j_{Hi}^*} + (p_{H50j_{Hi}^*} - p_{H25j_{Hi}^*}) * u_{H2i} & \text{if } .25 < u_{H1i} \leq .50, \\ p_{H50j_{Hi}^*} + (p_{H75j_{Hi}^*} - p_{H50j_{Hi}^*}) * u_{H2i} & \text{if } .50 < u_{H1i} \leq .75, \\ p_{H75j_{Hi}^*} + (p_{H95j_{Hi}^*} - p_{H75j_{Hi}^*}) * u_{H2i} & \text{if } u_{H1i} > .75. \end{cases}$$

3.2.4 Generating α_{Mi}, α_{Hi} and $\gamma_{1i} - \gamma_{5i}$

As discussed in Section 3.1, let

$$\begin{aligned}\alpha_{Mi} &= \exp\{1.10918 + 0.56964 * \log(\sigma_{Mi}^2)\}, \\ \alpha_{Hi} &= \exp\{0.57881 + 0.57797 * \log(\sigma_{Hi}^2)\},\end{aligned}$$

after obtaining σ_{Mi} and σ_{Hi} from the methodology described in Section 3.2.3. Also discussed Section 3.1, $\gamma_{1i} - \gamma_{5i}$ are set to be the σ_{Hi}^{-2} -weighted means of $\hat{\gamma}_{1i} - \hat{\gamma}_{5i}$ from Table 3.1. Thus for all i , define

$$\begin{aligned}\gamma_{1i} &= -0.0328488; \\ \gamma_{2i} &= 0.5269950; \\ \gamma_{3i} &= -0.1330379; \\ \gamma_{4i} &= -0.1162620; \\ \gamma_{5i} &= -0.2716631.\end{aligned}$$

3.2.5 Setting the $n \times 5$ human covariate matrix $\mathbf{Z}_i$, $i = 1, \dots, N$

For obtaining human covariates in the simulation, sampling with replacement is carried out with actual subjects and human genes from the GlaxoSmithKline experiment. These human subjects and genes are randomly assigned to simulation human and gene labels. The human labels in the simulation are assigned to the prior therapy and concomitant medication statuses of the real human subjects, and the gene-level covariates are assigned to those of from an actual human subject and gene combinations.

The following description is for a single gene pair i and must be repeated for all $i = 1, \dots, N$. In order to maintain an association between $\mathbf{z}_{1i}$ and β_{Hi} , we restrict attention to the human genes in bin j_{Hi}^* from Section 3.2.3 when performing the random assignment of real genes to gene labels in the simulation. For the 100 (or 83) GlaxoSmithKline gene indices in bin j_{Hi}^* , $\{i_1, \dots, i_{100}\}$ (or $\{i_1, \dots, i_{83}\}$), randomly select a single index i_k . Next, take a discrete uniform sample $r_1, \dots, r_n$ of $\{1, \dots, 24\}$, and define $\mathbf{e}_{r_j}$ as a 1×24 vector with a

1 in the r_j^{th} position and 0 elsewhere. Finally, set the $n \times 5$ human covariate matrix $\mathbf{Z}_i$ to

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{e}_{r_1} \\ \mathbf{e}_{r_2} \\ \vdots \\ \mathbf{e}_{r_n} \end{pmatrix} \times \begin{pmatrix} \mathbf{z}_{1i_k}^* & \mathbf{z}_{2i_k}^* & \mathbf{z}_{3i_k}^* & \mathbf{z}_{4i_k}^* & \mathbf{z}_{5i_k}^* \end{pmatrix}, \quad (3.7)$$

where $\mathbf{z}_{1i}^* - \mathbf{z}_{5i}^*$ are the actual 24×1 covariate vectors for gene i in the GlaxoSmithKline data.

3.2.6 Generating $\mathbf{Y}_H$ and $\mathbf{Y}_M$

The treatment indicator vectors $\mathbf{T}_M$ and $\mathbf{T}_H$ are determined by the sample sizes m_p, m_t, n_p and n_t . Let $\mathbf{T}_M = (\mathbf{0}_{m_p}^T, \mathbf{1}_{m_t}^T)^T$ and $\mathbf{T}_H = (\mathbf{0}_{n_p}^T, \mathbf{1}_{n_t}^T)^T$ where $\mathbf{0}_r$ and $\mathbf{1}_s$ are column vectors consisting of r zeros and s ones, respectively.

With all necessary parameters and covariates available, the simulated raw human gene expression data, $\mathbf{y}_H = \{y_{Hik}\}, i = 1, \dots, N; k = 1, \dots, n$, are generated according to the linear model

$$y_{Hik} = \alpha_{Hi} + \beta_{Hi}T_{Hk} + \mathbf{Z}_{ik}^T\boldsymbol{\gamma}_i + \sigma_{Hi}\epsilon_{Hik},$$

upon generation of i.i.d. $N(0, 1)$ random variables ϵ_{Hik} for all i, k . Similarly, the simulated raw mouse gene expression data, $\mathbf{y}_M = \{y_{Mij}\}, i = 1, \dots, N; j = 1, \dots, m$, are generated according to

$$y_{Mij} = \alpha_{Mj} + \beta_{Mj}T_{Mj} + \mathbf{Z}_{ij}^T\boldsymbol{\gamma}_i + \sigma_{Mi}\epsilon_{Mij},$$

upon generation of i.i.d. $N(0, 1)$ random variables ϵ_{Mik} for all i, j .

3.3 Simulation Experimental Design

3.3.1 Infinite Sample Size Experiment

The first experiment undertaken, the Infinite Sample Size Experiment (herein referred to as the ISSE), addresses the optimal case where there are an infinite number of mouse and human subjects, i.e., $\sigma_{1i} = \sigma_{2i} = 0$, $\hat{\beta}_{Mi} = \beta_{Mi}$, and $\hat{\beta}_{Hi} = \beta_{Hi}$. In this case, the latent treatment coefficients are observed, eliminating the distinction between models (2.16) and (2.18). Furthermore, the vector of sample sizes is reduced to one dimension, the number of gene pairs, N .

The goals of this experiment were to

- empirically demonstrate local consistency of parameter estimates and standard errors;
- isolate and determine the impact of gene pair sample size N on estimation performance;
- gauge the performance of the model to find I-Ortholog gene pairs in this optimal setting;
- question whether the unboundedness of the likelihood surface (variances are no longer bounded above zero) manifests itself in the estimation process.

No additional techniques, other than those provided in the chapter, were necessary for carrying out the ISSE. The gene pair sample size, N , was the only factor modulated in this experiment, at levels 250, 500, 1,000, 2,000, 4,000, 8,000, 16,000, 32,000, and 64,000. Although $N = 32,000$ and $N = 64,000$ seem unrealistically large, investigating model performance in these scenarios is relevant from a theoretical standpoint. The parameters $\mu_M, \sigma_M, \mu_H, \sigma_H, \rho, \eta, \theta$ were set to the values estimated from the GlaxoSmithKline data, as listed in Table 2.2.

The ISSE was planned to have 500 complete replications, but three replications were missing (370, 402, 403) and four replications were incomplete (13 missing runs combined) in the output dataset, due to less-than-careful stopping and starting of the simulation. The 497 replications resulted in 4460 total runs ($4460 = 9 \times 497 - 13$). The analysis of the ISSE is presented in Section 3.4.1.

3.3.2 Finite Sample Size Experiment I

A logical followup to the ISSE is an investigation of joint model (2.18) when subject sample sizes are finite but relatively large. The next experimental undertaking, known as Finite Sample Size Experiment I (herein referred to as FSSE I) involved a factorial modulation of four factors while maintaining large subject sample sizes. These factors are

- N , the number of gene pairs, taking values 4,000 and 16,000;
- m_t and m_p , the mouse subject sample sizes for treatment and control, taking values 40 and 80;

- η , the proportion of differentially expressed genes, taking values .25 and .5;
- θ , the proportion of I-Orthologs, taking values .2 and .4.

The human subject sample sizes n_t and n_p were kept constant at 40 each (80 total humans), and the remaining parameters $\mu_M, \sigma_M, \mu_H, \sigma_H, \rho$ were set to the values estimated from the GlaxoSmithKline data, as displayed in Table 2.2.

The primary goal of FSSE I is to learn about performance of the joint modeling methodology, in the presence of large subject sample sizes, regarding detection of differentially expressed human genes relative to competing single-species methodologies (i.e. Benjamini and Hochberg (1995) and the human marginal model). After obtaining knowledge about what is possible in the ISSE, the degradation of parameter estimation and I-Ortholog discovery is also of interest. Finally, the ways in which the modulated experimental factors impact the results are thoroughly investigated.

Finite Sample Size Experiment I was planned to have 500 replications, but the simulation ran until 504 replications and 8059 Runs. One replicate was incomplete, containing data on only 11 of 16 factorial combinations and resulting in five missing runs ($8059 = 16 \times 504 - 5$). The analysis of FSSE I is presented in Section 3.4.

3.3.3 Finite Sample Size Experiment II

Finite Sample Size Experiment II (herein referred to as FSSE II) is identical to its predecessor except for considerably smaller subject sample sizes. The mouse subject sample sizes m_t and m_p are factors taking values 10 and 20 (rather than 40 and 80 as in FSSE I), and the human sample sizes n_t and n_p are fixed at 10 (rather than 40 as in FSSE I). These smaller sample sizes are in fact much closer to that of the GlaxoSmithKline experiment (with 19 total mice and 24 total humans). In addition to the goals outlined in the previous section, this experiment is intended to provide insight to the analysis of the GlaxoSmithKline data in Section 2.6.

Finite Sample Size Experiment II was carried out with 500 complete replications, resulting in 8000 Runs. The analysis is presented in Section 3.4.

3.3.4 Finite Sample Size Experiment III

With approximately 16 percent of the 10,000 gene pairs from the GlaxoSmithKline data estimated to be I-Orthologs, it is reasonable to envision a situation (or a future

time, perhaps) where the $.16 \times 10,000 = 1600$ I-Orthologs comprise 80% of (a better understood) 2,000 gene pairs. Thus with θ fixed at .8 and N fixed at 2,000, Finite Sample Size Experiment III (herein referred to as FSSE III) modulates only two factors:

- m_t and m_p , the mouse subject sample sizes, taking values 10 and 20 as in FSSE II;
- η , taking values .25 and .5 as in FSSE I and II.

The small subject sample sizes from FSSE II are used, with human sample sizes n_t and n_p again fixed at 10. The remaining parameters $\mu_M, \sigma_M, \mu_H, \sigma_H, \rho$ are again set to the values estimated from the GlaxoSmithKline data, presented in Table 2.2. In this situation with a high proportion I-Orthologs, there is special interest in estimation and testing performance as a function of the mouse sample size.

Finite Sample Size Experiment III was carried out with 530 complete replications, resulting in 2120 Runs. The analysis is presented in Section 3.4.

3.4 Simulation Study Results

3.4.1 Infinite Sample Size Experiment

The ISSE seeks empirical validation of the identifiability arguments in Appendix B, as well as insight into estimation and testing performance when there are an infinite number of mouse and human subjects. The ISSE results are displayed graphically (Figures G.1-G.6) to efficiently convey meaning.

From the boxplots in Figures G.1-G.2, there is evidence that estimates of $\mu_M, \sigma_M, \mu_H, \sigma_H, \rho, \eta$, and θ are locally consistent at the true parameter values (Table 2.2). Very little bias is visible at any level of N (except perhaps for $\hat{\rho}$ at $N = 250$) and variability appears to decrease exponentially with the 2-fold increases in N .

The higher signal observed in the mouse data does not translate into improved estimates of $\hat{\mu}_M$ and $\hat{\sigma}_M$, as these exhibit higher levels of variability than their human counterparts $\hat{\mu}_H$ and $\hat{\sigma}_H$ (though $\hat{\sigma}_M$ and $\hat{\sigma}_H$ are comparable in terms of relative error). Estimation of ρ is plagued by exorbitant variability at $N = 250$, where in 12 cases $\hat{\rho} < 0$, but estimation quality is high when $N \geq 500$ and there are no further cases of negative estimates. The I-Ortholog mixing parameter θ is also estimated poorly at $N = 250$, with more gradual improvements from increasing N but fairly accurate by $N = 8,000$. Estimation of η also

experiences a gradual reduction in variability, but is accurate to within a few percentage points by $N = 1,000$.

Information matrix-based standard error estimates were typically very similar to the Monte Carlo estimates (see Figures G.3-G.4). Minor discrepancies were observed for $\hat{\mu}_M$ at $N = 250$ and $N = 500$. For $\hat{\rho}$, the information matrix standard errors grossly underestimated the true variability when $N = 250$, but when $N \geq 1000$ the two methods closely agree. Note however that model (2.16), from which the ISSE data are generated, assumes independence among genes, an assumption that is certainly violated in real data (Section 2.2.1). Hence this discussion is concerned with model standard errors, and not necessarily standard errors associated with actual gene expression data.

Though it is unnecessary to search for differentially expressed genes when there are an infinite number of subjects, the ISSE illuminates properties of I-Ortholog discovery in an optimal subject sample size situation. Refer to Figures G.5-G.6, and note that standard errors span .001-.02 (.0001-.012) for empirical FDR and .0002-.0036 (.0000-.0001) for empirical power, at nominal FDR .05 (.20).

Empirical FDRs exceed their nominal levels of .05 and .20 by approximately 3.5 and 7 percentage points, respectively, when $N = 250$. By $N = 500$, the exceedence in nominal FDR is on the order of a single percentage point for both nominal levels. At $N \geq 500$ and nominal FDR .20, the empirical FDR is relatively settled at nominal level of .20. For nominal FDR .05, there is a slight dip of the empirical FDR to below nominal before climbing to stabilize around the nominal level, at around 8,000-16,000 gene pairs.

Power rates for both nominal FDRs largely share the same trends as their empirical FDR counterparts (see Figure G.6), starting higher before stabilizing at a lower level. Even in this infinite subject sample size configuration, with well understood simulated data, the search for I-Orthologs is a difficult problem: at nominal FDRs .05 and .20, only about 1% and 7% of I-Orthologs, respectively, are discovered at large values of N .

Finally, despite the potential for degeneracy at $\hat{\sigma}_M = 0$ and $\hat{\sigma}_H = 0$, there was no sign of problems from spikes in the likelihood surface. Though $\hat{\rho}$ occasionally took negative values when $N = 250$, in no case did $\hat{\sigma}_M$, $\hat{\sigma}_H$, $\hat{\theta}$ or $\hat{\eta}$ take zero values.

Figure 3.4: Estimates of σ_H and ρ in FSSE I, over all factorial combinations. The dotted lines indicate the 1st and 99th percentiles, respectively.

3.4.2 Outliers in the Finite Sample Size Experiments

From the ISSE analysis (Section 3.4.1), recall that 12 runs, all occurring at $N = 250$, suffered extremely poor estimation of $\rho = .934$. These 12 cases greatly impacted the $\hat{\rho}^T$ boxplot in Figure (G.2). Though arguably outliers, these runs remained in the analysis due to their isolation at $N = 250$ and the exploratory nature of the ISSE.

In FSSEs I-III, however, a particular phenomenon occurred prompting the flagging and removal of a very small subset of experimental observations. This phenomenon can be seen graphically in Figure 3.4, where a small fraction of runs in FSSE I resulted in estimates of $\hat{\sigma}_H$ that were unusually close to zero and remarkably far away from the larger distribution of values. No other parameter in FSSE I had such isolated estimates of unusual values, though $\hat{\rho}$ can be seen in Figure 3.4 with a long tail of values extending to -1.0. Interestingly, when $\hat{\sigma}_H$ -outliers were removed (based on the rule $\hat{\sigma}_H < 3$ in FSSE I), so were a striking proportion of the poor estimates of ρ (see Figure 3.5). Similar phenomena occurred in FSSEs II and III and outliers based on unusually small $\hat{\sigma}_H$ values were removed. See Tables 3.2-3.4 for more details about these outliers in the specific contexts of FSSE I-FSSE III.

Experimental runs deemed $\hat{\sigma}_H$ -outliers are not parameter independent, and some factorial combinations are more likely to have these outliers than others. Tables 3.2-3.4 show the distribution of outliers over the factorial combinations in FSSEs I-III. Outliers in FSSE I were more common at 80 mice over 40 mice per condition, $\theta = .2$ over $\theta = .4$, and

Figure 3.5: Estimates of σ_H and ρ in FSSE I, where 74 out of 8059 runs with $\hat{\sigma}_H < 3$ were removed. The dotted lines indicate the 1st and 99th percentiles, respectively.

Table 3.2: From FSSE I, the factorial makeup of the 74 out of 8059 experimental runs flagged as outliers when $\hat{\sigma}_H < 3$.

Number of Genes	$\theta =$	40 mice		80 mice	
		.20	.40	.20	.40
N=4,000	$\eta = .25$	5	1	20	1
	$\eta = .50$	0	0	0	0
N=16,000	$\eta = .25$	13	4	21	5
	$\eta = .50$	0	0	2	2

overwhelmingly $\eta = .25$ over $\eta = .5$. Though $\hat{\sigma}_H$ -outliers were more common in FSSE II than FSSE I, the distribution is similar; outliers were more common at 20 mice than 10 mice per condition, $\theta = .2$ than $\theta = .4$, and $\eta = .25$ than $\eta = .5$. The eight outliers in FSSE III are evenly spread out among the four factorial combinations.

3.4.3 Finding Differentially Expressed Human Genes

This section summarizes the results from FSSEs I-III regarding use of joint model (2.18), marginal model (2.21), and the procedure of Benjamini and Hochberg (1995) to find differentially expressed human genes. The full results are given in Appendix G, starting with Table G.27. Metrics for comparison include estimates of power and FDR (herein referred to as *empirical power* and *empirical FDR*). Standard errors for empirical power are generally

Table 3.3: From FSSE II, the factorial makeup of the 524 out of 8000 experimental runs flagged as outliers when $\hat{\sigma}_H < 2.55$.

Number of Genes	$\theta =$	10 mice		20 mice	
		.20	.40	.20	.40
N=4,000	$\eta = .25$	40	16	81	51
	$\eta = .50$	17	9	31	24
N=16,000	$\eta = .25$	27	21	99	46
	$\eta = .50$	12	2	35	13

Table 3.4: From FSSE III, the factorial makeup of the 8 out of 2120 experimental runs flagged as outliers when $\hat{\sigma}_H < 2.4$.

Number of Genes		$\theta = .8$	
		10 mice	20 mice
N=2,000	$\eta = .25$	2	2
	$\eta = .50$	2	2

negligible for the purposes of the following discussion: the maximum observed standard error for across all procedures and factorial combinations is .002, with the exception of the joint model in FSSE III (where it is .004 and still negligible relative to the empirical powers observed in this experiment). The maximum observed standard errors for empirical FDR were .002 in FSSE I, .010 in FSSE II, and .011 in FSSE III. Though possessing larger standard errors than empirical power, empirical FDRs are estimated precisely enough to freely discuss whether approximate FDR control is achieved.

FSSE I. In FSSE I, nominal false discovery rate control was either achieved or approximately achieved under the joint model; for nominal FDR .05, empirical FDRs spanned .057-.077 when $\eta = .25$ and .038-.047 when $\eta = .5$; for nominal FDR .20, empirical FDRs spanned .113-.124 when $\eta = .25$ and .137-.161 when $\eta = .5$. Meaningful reductions of empirical FDRs always occurred with more I-Orthologs ($\theta = .4$ versus $\theta = .2$). An equally consistent but less substantial trend was larger empirical FDRs with more mice (up to .004 at nominal FDR .05 and .010 at nominal FDR .20). There is some evidence of lower empirical FDRs with higher gene pair sample sizes, as for $\eta = .25$, 16,000 versus 4,000 gene pairs were associated with small decreases in FDR for all eight factorial combinations (there was less consistency at $\eta = .5$). For both nominal levels of FDR, the human marginal model

(2.21) and the procedure of Benjamini and Hochberg (1995) were somewhat conservative.

At nominal FDR .05, empirical human differential gene-finding powers of the joint model ranged from .081-.168, compared to .051-.083 with the marginal model and .025-.048 with the procedure of Benjamini and Hochberg. At nominal FDR .20, empirical joint model powers ranged from .171-.438, compared to .100-.172 with the marginal model and .094-.160 with the procedure of Benjamini and Hochberg. Only small increases in power were observed from doubling the number of mice (up to a .01 increase), though the associated increases were always proportionally greater than the corresponding increases in FDR (except in the case of $\theta = .2$ and $\eta = .25$). This effect of doubling the number of mice, however, is not illustrative of the overall species' contribution, as increases in the percentage of I-Orthologs or the complete removal of mice from analysis (i.e., the marginal model) had a more substantial impact. For example, in the case of 16,000 genes, 40 mice per condition, $\eta = .5$, and nominal FDR .05 (.20), power increased from .123 to .159 (.285 to .401) as θ doubled from .2 to .4. When mice were not included in the analysis, this power dropped to .082 (.171).

FSSE II. In FSSE II, both joint model (2.18) and marginal model (2.21) experienced failures to meet nominal FDRs (see Tables G.30-G.31). At nominal level .05, empirical FDRs ranged from .118-.295 in the joint model and .114-.264 in the marginal model. At nominal level .20, empirical FDRs ranged from .144-.341 in the joint model and .147-.303 in the marginal model. Empirical FDRs at $\eta = .5$ were roughly half their $\eta = .25$ counterparts and all fell under the controlled limit at nominal FDR .20. Increased proportions of I-Orthologs ($\theta = .4$ versus $\theta = .2$) were associated with substantial reductions in empirical FDR (up to eight percentage points). Unlike FSSE I, no even marginally significant effect of gene pair sample size, N , is apparent.

At nominal FDR .05, empirical human differential gene-finding powers of the joint model ranged from .028-.062, compared to .020-.031 with the marginal model and .0002-.0010 with the procedure of Benjamini and Hochberg. At nominal FDR .20, empirical joint model power ranged from .063-.200, compared to .039-.068 with the marginal model and .0021-.0110 with the procedure of Benjamini and Hochberg. The joint model considerably outperformed its marginal counterpart, often with power increases of two to three times (especially at nominal FDR .20) and smaller or similar empirical FDRs. The power gain of doubling the number of mice was more pronounced than in FSSE I, up to an estimated increase of .05. At $\eta = .5$, empirical power is consistently higher and empirical FDR

consistently lower than for $\eta = .25$, and FDR control is successful at nominal level .20 and $\eta = .5$. Also, I-Ortholog gene pairs were especially helpful in FSSE II, as the doubling of θ from .2 to .4 uniformly increased powers (up to .07). While the Benjamini and Hochberg (1995) procedure consistently provided nominal false discovery rate control, its power to find differentially expressed human genes is weak at best: a mere 1 in 1,000 genes discovered at nominal FDR .05 and 1 in 100 at nominal FDR .20.

FSSE III. Despite possessing the small subject sample sizes of FSSE II, empirical FDRs of the joint model in FSSE III were closer to those of FSSE I. For nominal FDR .05, these FDRs spanned .078-.081 when $\eta = .5$ and .040-.041 when $\eta = .5$. For nominal FDR .20, the joint model was more conservative than the procedure of Benjamini and Hochberg (1995), with empirical FDRs spanning .091-.110 in the former and .101-.171 in the latter. The human marginal model, on the other hand, experienced failures similar to in FSSE II: empirical FDRs ranged from .12-.27 for nominal FDR .05 and .15-.32 for nominal FDR .20.

In this optimistic scenario where 80% of gene pairs are I-Orthologs, the relative power of the joint model was impressive. The joint model, with empirical FDRs of only a fraction of the marginal model, boasted power increases of approximately 5-7 and 8-13 times those of the marginal model at nominal FDRs .05 and .20, respectively. Furthermore, the joint model enjoyed 41-60 times the empirical power of the Benjamini and Hochberg (1995) procedure at nominal FDR .20, where the former is more conservative. Increases in the number of mice correspond to substantially higher empirical power and brought empirical FDRs closer to their nominal levels. At the factorial combination with the highest performance (20 mice per condition, $\eta = .50$, and nominal FDR .20), 73.2% of all differentially expressed genes were discovered using the joint model.

3.4.4 Finding I-Orthologs the Finite Sample Size Experiments

In the search for I-Orthologs, the maximum standard error for empirical FDR over all finite sample size experimental combinations was .014. Similarly, maximum standard errors for empirical power, *expressed in units of percent*, were .062 in FSSE I, .039 in FSSE II and 1.42 in FSSE III. For the full results on I-Ortholog discovery in the finite sample size experiments, see Tables G.29, G.32, and G.35.

Empirical FDRs for I-Ortholog discovery often significantly fell both above and below their nominal levels, depending on the particular factorial combination. In FSSE I,

empirical FDRs ranged from .014-.090 for nominal level .05 and .183-.266 for nominal level .20. The greatest exceedences of the nominal FDR in FSSE I occurred at $\theta = .4$ for FDR .05 and $\theta = .2$ for FDR .20. Empirical FDRs in FSSE II spanned a wider range throughout the various factorial configurations: .004-.115 at nominal level .05 and .184-.338 at nominal level .20. As in FSSE I, smaller empirical FDRs were associated with $\theta = .2$ for nominal FDR .05 and $\theta = .4$ for nominal FDR .20. At the opposite configuration ($\theta = .4$, FDR .05 and $\theta = .2$, FDR .20), increasing the gene pair sample size, N , significantly increased empirical FDR. Empirical FDRs in FSSE III were relatively well behaved, only slightly exceeding nominal level .05 (.052-.061) and always falling under nominal level .20.

Powers to find I-Orthologs were typically low and are given in either units of percent or rates of detection. In FSSE I, the power to find I-Orthologs ranged from .01 to .41 *percent* for nominal FDR .05 and 0.6 to 5.4 *percent* for nominal FDR .20. In FSSE II, empirical detection rates spanned 8 in 10,000 to a mere 3 in 100,000 at nominal FDR .05, and 1.2 in 1,000 to 1.75 in 100 for nominal FDR .20. For the above ranges of empirical powers in FSSEs I and II, the highest observed values are crucially dependent on $\theta = .4$ versus $\theta = .2$. In these situations with the highest empirical powers ($\theta = .4$, FDR .20), $\eta = .5$ versus $\eta = .25$ is associated with greater empirical power. For factorial combinations with relatively low powers, however, this trend is reversed. In FSSE III, with 80% of all gene pairs as I-Orthologs, the four-fold increase in the nominal FDR (.05 to .20) translated to power increases of up to 28 times: powers ranged from 1.8-2.6 *percent* for nominal FDR .05 and 24.0-72.5 *percent* for nominal FDR .20.

3.4.5 Estimation in the Finite Sample Size Experiments

Point Estimation

Though not a primary focus of this research, the quality of parameter estimation in joint model (2.18) has relevance to differential expression and I-Ortholog discovery. This section summarizes the performance of estimators of μ_M , μ_H , σ_M , σ_H , ρ , η , θ (Section 2.4) in the finite sample size settings of FSSEs I-III (Sections 3.3.2-3.3.4). Recall that in the Infinite Sample Size Experiment (ISSE), estimates were largely bias free (Section 3.4.1), thus the occurrence of finite sample size bias in FSSEs I-III is the primary focus of this discussion. For the complete results including Mean Squared Errors, refer to Tables G.3-G.26 in Appendix G.

Despite complications from three distinct sample sizes (gene pair, mouse, and human), one familiar property of standard maximum likelihood theory was observed: when the gene pair sample size quadruples (from $N = 4,000$ to $N = 16,000$ in FSSEs I and II) the standard deviations of parameter estimates were generally halved in all factorial combinations. In contrast, the effect of N on finite sample size bias is insignificant across all parameter and experimental combinations.

The following paragraphs discuss separately the estimators for μ_M , μ_H , σ_M , σ_H , ρ , η , and θ in FSSEs I-III. Standard errors for various estimated quantities are often non-negligible, but care has been taken to make statements which are statistically meaningful, including use of Bonferroni adjustments and weighted linear models for the analysis of bias. Trends in an estimator's variability, based on computed standard deviations over experimental repetitions, were considered using sign-test p-values.

Estimators of $\mu_M = -2.71$ (Tables G.3, G.10 and G.17) are biased downwards, up to -0.37 in FSSE I, -0.7 in FSSE II, and -0.38 in FSSE III. The only exception is the case of $\eta = .5$ and $m_t = m_p = 10$ in FSSE III, where the bias is 0.28 (with standard error .101). In FSSEs I and II, there is significantly less bias and variability when $\eta = .5$ over $\eta = .25$. Also in FSSE I, $\theta = .4$ versus $\theta = .2$ is a bias reducer. The effects of doubling the mouse sample size on bias are insignificant in FSSE I, but in two of eight cases in FSSE II, bias significantly increased. In the one case of positive initial bias in FSSE III, the doubling of mice significantly decreased bias. Despite occasional bias increases, there was a significant association of lower variability with the doubling of mice in FSSEs I-III.

Unlike its mouse counterpart, estimators of $\mu_H = 2.83$ (Tables G.4, G.11 and G.18) more often experience gentler, positive biases. Modulating η from .25 to .5 has a marginally significant bias reduction effect in FSSE II and reduces estimation variability. Also, modulating θ from .2 to .4 significantly reduces bias in FSSE I and both bias and variability in FSSE II. A very significant drop in bias in FSSEs I-III, in addition to a modest reduction in variability, was associated with doubling of the mouse sample size. In FSSE II, the bias is absolutely reduced, but appears to have "overshot" to negative values when there are more mice.

Estimators of $\sigma_M = 55.28$ (Tables G.5, G.12 and G.19) tend to experience positive bias, of up to 1.04 in FSSE I and to over 5 in FSSEs II and III. Significant and substantial decreases in bias are associated with increases in mouse sample size, especially in FSSEs II and III. In FSSE I, these bias reductions occurred where there were large positive biases

(> .80) to begin with; for cases with smaller biases (< .20), bias reduction was less substantial (in one case the bias was significantly greater). Furthermore, in FSSEs I-III, a decrease in bias was associated with $\theta = .4$ versus $\theta = .2$, and estimation variability of $\hat{\sigma}_M$ and RMSE was substantially less when $\eta = .5$ versus $\eta = .25$.

Estimators of $\sigma_H = 6.77$ (Tables G.6, G.13 and G.20) are biased downward in FSSEs I and II, up to -1.19 in the former and -1.03 in the latter. In three of four combinations of FSSE III, however, $\hat{\sigma}_H$ exhibits positive bias, up to .80. In FSSEs I and II, $\eta = .25$ is associated with less bias (albeit higher variability) than $\eta = .5$, and $\theta = .4$ corresponds to less bias than $\theta = .2$. In FSSEs I and II, an increase in the number of mice was significantly associated with greater bias, and in FSSE III, doubling the number of mice significantly alleviates bias when $\eta = .25$, but not when $\eta = .5$.

Estimators of $\rho = .934$ (Tables G.7, G.14 and G.21) are biased downward, up to -.01 in FSSE I, -.11 in FSSE II, and -.016 in FSSE III. In FSSEs I and II, less bias and variability is associated with $\eta = .5$ and $\theta = .4$ than $\eta = .25$ and $\theta = .2$. Increases in the number of mice are often statistically related to increases in bias, though the differences in means are small (up to .017).

In all three experiments, estimators of η suffer from precise but inaccurate estimation (Tables G.8, G.15 and G.22). There is uniformly downward bias: up to -.10 in FSSE I, -.20 in FSSE II, and -.17 in FSSE III. The bias is larger for $\eta = .5$ than $\eta = .25$ in an absolute (though not necessarily relative) sense. In FSSE I and II, the bias is less when $\theta = .4$ versus $\theta = .2$. Significant and meaningful reductions in bias are noted in all finite sample size experiments as the number of mice is increased, and FSSEs II and III, with their smaller subject sample sizes, showcase the most spectacular improvements.

Estimators of the I-Ortholog parameter θ also tend to be biased downward, but fare much better than $\hat{\eta}$, facing biases of up to -.03 in FSSE I, -.08 in FSSE II, and -.07 in FSSE III. In all three experiments, $\eta = .5$ is associated with less bias and variability than $\eta = .25$. In FSSE I, $\theta = .4$ is associated with less bias than $\theta = .2$, however this trend is not significant in FSSE II. The effects from doubling the number of mice are insignificant in FSSE I, though in FSSEs II and III, increasing the number of mice often significantly decreases bias.

As a competitor to joint model (2.18), human marginal model (2.21) was fit to the simulated data in FSSEs I-III, where posterior probabilities of differential expression were constructed conditional on only human data. Tables G.24-G.26 detail the marginal

estimation of μ_H , σ_H and η_H in FSSEs I and II, and here we give a brief summary of these results.

- Marginal estimators of μ_H exhibit larger bias and variability than their joint model counterparts, with RMSEs up to 12 times larger in the marginal model. Bias in FSSE I is significantly less than in FSSE II when $\eta = .5$, but greater when $\eta = .25$.
- Both bias and variability of $\hat{\sigma}_H$ are remarkably attenuated from FSSE II to FSSE I. In FSSE I, the bias of $\hat{\sigma}_H$ is worse in the joint model than in the marginal model, though in FSSE II, both bias and variability are substantially reduced in the joint model relative to the marginal model.
- Marginal estimators of η (i.e., η_H) have more severe downward biases. When $\eta = .25$, biases are approximately -.185 in FSSE I and -.204 in FSSE II. When $\eta = .5$, biases are approximately -.307 in FSSE I and -.378 in FSSE II. In both experiments, estimation variability is negligible compared to the bias. An increase in the number of humans is associated with a bias correction of .01 when $\eta = .25$ and .07 when $\eta = .5$.

Point Estimation and Differential Gene Expression

Given the failures of FDR control in FSSE II (Section 3.4.3), there is a need to understand how goodness of parameter estimation is related to testing performance, namely control of the nominal false discovery rate (FDR). This section summarizes the results of an investigation into how probabilities of discovery (and false discovery) are impacted by incorrect estimation of the parameters of model (2.18).

Similar to a subconfiguration of FSSE II, a 10,000 gene pair dataset was simulated with $\eta = .5$, $\theta = .4$, 10 mice and humans per treatment group, and all other parameters from the GlaxoSmithKline data (Table 2.2). Posterior probabilities of treatment effect (the p_i s from Section 2.3.1) were computed using, (i) the true parameter values, (ii) the true parameter values except for one parameter with a larger value, (iii) the true parameter values except for one parameter with a smaller value. The incorrect parameter values are denoted with plus and minus superscripts, indicating the direction of the error (for instance, the true value of θ is $\theta = .4$, the incorrectly small value is $\theta^- = .3$, and the incorrectly large value is $\theta^+ = .5$). The incorrect values of each parameter were chosen somewhat arbitrarily, but with consideration to the errors of estimation in FSSEs I and II and thus realistically

bad estimates of each parameter. Since a change in any one parameter's value may not affect all genes consistently, Tables G.36 and G present the effects on the population of p_i s. Table G.36 describes the effect on all p_i s, while Table G describes the effects on *false discovery* p_i s, the probabilities of differential expression conditional on $\beta_{Hi} = 0$. The following is a discussion on the findings from this sensitivity analysis and their abilities to explain the mishaps in FSSE II.

- The effect of over- or under-estimation of mouse marginal parameters μ_M and σ_M on human differential expression probabilities is particularly minimal, as 95% of p_i s are increased by .01 or less in both cases.
- Under- or over-estimation of μ_H appears to have relatively symmetric effects on the p_i s, increasing some probabilities and decreasing others. In terms of false discoveries, 5% of non-differentially expressed genes have false discovery p_i s that are increased by .06-.13 when $\mu_H^- = 1$ (versus 2.83). Another (non-exclusive) 5% experience increases of .04-.11 when $\mu_H^+ = 5$.
- The under-estimation of σ_H may have substantial consequences, as over 75% of the false discovery p_i s were increased, the greatest 5% of increases spanning .09-.25. Over-estimation also posed a threat, as the greatest 5% of increases spanned .01 – .18.
- Modulation of the correlation ρ both increased and decreased probabilities of differential expression. Notably, the greatest 5% of false discovery p_i increases were .05-.15 when $\rho^- = .60$ and .01-.15 when $\rho^+ = .99$.
- Mixing proportion η is the one parameter with completely predictable effects on the p_i s and false discovery p_i s. Smaller (larger) values of η monotonically decrease (increase) probabilities of differential expression.
- Finally, poor estimation of θ may also impact FDR control. When $\theta^- = .3$ (versus .4), 50% of false discovery p_i s are increased by .02 – .06.

Besides generally poorer estimation of parameters in FSSE II, simple explanations of why nominal FDRs were exceeded are not apparent. The sensitivity analysis, in fact, suggests that the similar performance degradation of the marginal and joint models are due to different reasons. In the marginal model, it seems that the large errors in the estimation

of σ_H in FSSE II relative to FSSE I would be the best one-parameter explanation. In the joint model, it seems likely that inferior estimation of μ_H , ρ , and θ in FSSE II relative to FSSE I all contributed to the increase in model-based false discoveries. Circumstantial evidence points especially to ρ , the one parameter that was estimated substantially better in FSSE III than FSSE II (Section 3.4.3).

Standard Errors

This section describes the performance of information matrix-based standard errors in FSSEs I and II when all modeling assumptions are correct. Parameter estimation via maximization of composite pseudolikelihood (2.32) (Section 2.4) is based upon a working assumption of gene pair independence, an assumption that is almost certainly violated in actual gene expression data (Section 2.2.1). Consequently, as in the ISSE, the performance of standard error estimates is more a model curiosity than an indicator of standard error reliability in real data.

Recall that in the ISSE, information matrix-based standard errors were quite accurate for $N \geq 1,000$ (see Section 3.4.1). The results of this evaluation are presented in Tables G.1 and G.2, which contain ratios, for every parameter and factorial combination of FSSE I and II, of the average information matrix-based standard error to the Monte Carlo estimate of standard error. A ratio greater than one (less than one) indicates that the information matrix-based standard error overestimates (underestimates) the true sampling variability of the parameter. Delete-1 Jackknife standard errors are computed for each ratio and included in parentheses. The following is a summary of results from this performance evaluation.

- Information matrix-based standard errors for $\hat{\mu}_M$ perform well in all factorial combinations of FSSEs I and II.
- Information matrix-based standard errors for $\hat{\mu}_H$ underestimate the true sampling variability. The underestimation is more severe with the addition of mice and gene pairs, but less severe in FSSE I than FSSE II. This latter fact indicates the importance of human subjects for quality estimation of the information-based standard error of $\hat{\mu}_H$.
- There is a slight tendency of information matrix-based standard errors to overestimate

the variability in $\hat{\sigma}_M$. This tendency is greater in FSSE II than FSSE I and is somewhat alleviated at 16,000 gene pairs and 80 mice.

- The information matrix-based standard errors for $\hat{\sigma}_H$ underestimate the true sampling variability. Similar to the case of $\hat{\mu}_H$, this tendency is relieved when there are more human subjects (FSSE I versus FSSE II), but may worsen with more mouse subjects and/or gene pairs.
- There is wildly high variability in the ratio of standard errors of $\hat{\rho}$ in FSSE I and (especially) FSSE II. When this variability subsides in the presence of more mice, the information matrix-based standard errors tend to underestimate the true variability.
- Information matrix-based standard errors for $\hat{\eta}$ are quite good in FSSE I. In FSSE II there is a slight tendency of these standard errors to underestimate the true sampling variability.
- Information matrix-based standard errors for $\hat{\theta}$ are also quite good in both FSSE I and II. In one case in FSSE II, however (4,000 gene pairs, 20 mice, $\eta = .25, \theta = .2$), the ratio estimate had an unusually high standard error.

Chapter 4

Discussion and Conclusions

4.1 Differential Gene Finding

Reflecting on early aspirations for the dual-species gene expression model (2.18), this research has both validated and challenged preconceived notions. One motivating vision was quickly discarded: that model (2.18) is a tool enabling endless gains in human inferential power from indefinite increases of the mouse sample size. In the simulation studies of Chapter 3, the increase in differential gene-finding power from doubling the number of mice was sometimes impressive (up to 20 percentage points in Finite Sample Size Experiment III) but more often increases of a few percentage points or less were observed. In Finite Sample Size Experiment I (FSSE I), doubling the number of mice at best increases differential gene-finding power by a single percentage point, but this metric ignores the larger enjoyed contribution of the mouse data, as removing the mice altogether (i.e., the human marginal model (2.21) in FSSE I) resulted in power decreases of up to 50%. The potential impact of the mouse data is also illustrated when the percentage of species-extrapolative I-Orthologs, θ , is doubled from .2 to .4, as differential gene finding power increases by up to four percentage points (four times the increase from doubling the mouse sample size).

In its most impressive scenarios, the gains from using joint model (2.18) versus competing procedures are remarkable. In Finite Sample Size Experiment II (FSSE II), where 80% of gene pairs are I-Orthologs, the joint model is 5-13 times more powerful than its marginal counterpart, despite substantially lower empirical false discovery rates, and 41-60 times more powerful than the procedure of Benjamini and Hochberg (1995), while more conservative at nominal level .20. On the other hand, the joint model (2.18) grossly failed to

control FDR in FSSE II, the simulation experiment most resembling the GlaxoSmithKline data (this should justifiably lead the reader to be wary of the set of discovered differentially-expressed genes due to rosiglitazone in Section 2.6). The best explanation for the poor performance of the joint model is a downward bias in ρ not seen in FSSEs I or III. For the human marginal model, poor estimation of σ_H is the most reasonable explanation. Despite the failures of FDR control in FSSE II, the smaller subject sample sizes were associated with great improvements of the joint model over the human marginal model, with powers increases of up to three times. Furthermore, doubling the number of mice resulted in power increases of up to 4.5 percentage points (over four times that of FSSE I) with similar or smaller empirical FDRs.

4.2 Bias

Finite sample size bias is a subject of intense curiosity regarding models (2.18) and (2.21). With ρ as a slight exception, estimates from the Infinite Sample Size Experiment (ISSE) were practically bias-free, and hence finite subject sample sizes are responsible for bias in parameter estimation in FSSEs I-III. As expected, increases in the mouse and human sample sizes usually correspond to less biased estimates, but this is not always the case. For certain parameters, such as σ_H , estimators are more biased when the number of mice is doubled, whereas estimators such as $\hat{\eta}$ always experience bias reductions when the mouse sample size is doubled. This should not be thought of as problematic. In a half-optimal scenario of an infinite number of mice and a finite number of humans, there can still be finite sample size bias of certain estimators (such as $\hat{\sigma}_H$), and in FSSEs I-III, only mouse sample sizes are modulated. Thus, the observed finite sample size biases do not necessarily tend towards zero bias, but towards an asymptotic finite sample size bias that occurs when the mouse sample size approaches infinity. With simultaneous increases in both subject sample sizes, however, bias in estimators will theoretically tend towards zero. Acknowledging the impact of bias on the posterior probabilities of differential expression, investigation of bias correction in the parameter estimation of model (2.18), perhaps using the jackknife, would be an interesting topic of future research.

4.3 I-Orthologs and Species Extrapolation

As species extrapolation exists at the heart of this research, we reflect on the concept of I-Orthology, which compromises the machinery to make such extrapolations. The hope is that the single linear relationship specified in this research will largely capture the essence of the inter-species agreement. Herein lies another opportunity for refinement of the proposed methodology: to relax the nature of the relationship between human and mouse treatment coefficients or to allow this relationship to differ among genes. If successful, then not only could extrapolative success improve, but the whole concept of I-Orthology would be altered and the increased flexibility would likely lead to larger estimates of (a suitably redefined) θ .

Unfortunately, the task of discovering I-Orthologs is difficult, especially near the one parameter configuration estimated from actual data. In the GlaxoSmithKline data, there was one close miss ($q_i = .78$) but no I-Ortholog discoveries at nominal FDR .20. Considering the results of FSSE II with a similar parameter configuration (depending on the quality of estimation in the GlaxoSmithKline example) we would have expected to find ten or so I-Orthologs in the GlaxoSmithKline data. As performance in discovering I-Orthologs increased dramatically at nominal FDR .20 versus .05 in simulations, we would encourage the analyst to consider tolerating a larger proportion of false discoveries in exchange for discovering a set of likely inter-species extrapolators. In the GlaxoSmithKline data, discovering gene pairs with posterior I-Ortholog probabilities greater than half ($q_i > .5$) results in 18 discoveries and an estimated FDR of .42 (with standard error .02). Here we would like to suggest the potential impact that a meta analysis of many dual-species experiments (with a variety of treatments) might have on I-Ortholog discovery. A theoretical formulation in the context of model (2.18) would presumably be another interesting area of future research.

4.4 Mixture Modeling

Perhaps an under-addressed aspect of this research, the use of mixture models to describe β_{Mi} and β_{Hi} has presented its own set of challenges. Unlike the method of Lewin et al. (2006), our methodology, based on a mixture of distributions, needs no arbitrary equivalence limit (i.e., differential expression only when $|\beta_{Hi}| > \delta$, for an analyst-chosen δ). While mixing parameter η , estimated from the data, essentially determines how many

genes *will* be discovered, it does not appear to be an accurate indicator of the amount of genes that *should* be discovered. For instance, in the GlaxoSmithKline data, the mouse gene expression data undeniably possessed more signal than its human counterpart (Section 2.6), however in the marginal models, $\hat{\eta}_H > \hat{\eta}_M$ (though it is difficult to formalize this difference due to the lack of theoretical standard errors available). Instead, the amount of signal in the mouse and human datasets seemed to be embedded in the latent differential expression variabilities σ_M^2 and σ_H^2 . Alternative estimates for η that prioritize its role as a descriptor of signal are worth considering.

An unusual feature of this particular mixture model application is that there are fewer mixing parameters than distributional components, as θ and η completely determine the mixing proportions of the five bivariate normal components in models (2.16) and (2.18). Though this could be achieved under a more standard framework via linear constraints applied to the mixing proportions, we did not succeed in finding any third-party software specifically for fitting mixture models with such constraints (or the constraints within the distributional components). Considering the lack of degeneracy problems in this research, especially when fitting model (2.16) in the ISSE (Section 3.4.1), perhaps similar constraints would be useful in wider mixture model applications.

Before discovering the large body of literature on Bayesian multiple testing (Section 1.2.2), it was somewhat difficult for these authors, favoring frequentist methodology, to subscribe to the notion that a certain percentage of null hypotheses were in fact “true.” Although intuitively reasonable enough (it is easy to believe that some genes will not be affected by a given treatment), there were earlier (unreported) efforts made to find evidence that a mixture of bivariate distributions truly exists in the GlaxoSmithKline data. Though realization of widespread application of the Bayesian methodology diminished this concern, the question of whether there are truly mixtures to be found (for either differential expression or I-Orthology) remains. Notably, when the GlaxoSmithKline gene expression data are log-transformed, the differential expression mixture ceases to exist, thus rendering all such conclusions meaningless (see Appendix C). This scenario needs to be addressed, perhaps with alternative estimators for η and θ that encourage a mixture.

4.5 Reflections and Future Outlook

Even if there was a crystal ball that gave researchers complete knowledge about human differential expression, the quite independent problem of biological interpretability would remain. Whether it is presently economically advisable to carry out gene expression experiments in animals for the purposes of human extrapolation, this research presents a novel strategy with possible applications in many areas. If I-Orthology concepts fail in the context of gene expression, perhaps they will succeed in other animal-to-human biological mappings (for instance, metabolites). Regardless of how difficult the animal-to-human extrapolation problem is, or has been, emerging high-throughput bioinformatic technologies provide hope that monumental advances are near, ushering in a new era of preclinical testing for human safety and efficacy.

Bibliography

- Adjaye, J. et al. (2004), Cross-species hybridisation of human and bovine orthologous genes on high density cDNA microarrays. *BMC Genomics* **5**(83).
- Affymetrix (2001*a*), *Data Analysis Fundamentals*. Affymetrix, Santa Clara, CA.
- Affymetrix (2001*b*), *Microarray Suite User Guide, Version 5.0*. Affymetrix, Santa Clara, CA.
- Affymetrix (2003), *User's Guide to Product Comparison Spreadsheets*. Affymetrix, Santa Clara, CA.
- Affymetrix (2005–2007). Technical support via www.affymetrix.com and email correspondence.
- Batzoglou, S., Pachter, L., Mesirov, J., Berger, B. & Lander, E. (2000), Human and mouse gene structure: Comparative analysis and application to exon prediction. *Genome Res.* **10**(7), 950–958.
- Bayarri, M. & Berger, J. (2004), The interplay of Bayesian and frequentist analysis. *Statistical Science* **19**(1), 58–80.
- Benjamini, Y. & Hochberg, Y. (1995), Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Statistical Science* **57**(1), 289–300.
- Benjamini, Y. & Yekutieli, D. (2001), The control of the false discovery rate in multiple hypothesis testing under dependency. *Annals of Statistics* **29**(4), 1165–1188.
- Blake, J. et al. (2003), MGD: the mouse genome database. *Nucleic Acids Res.* **31**(1), 193–195. MGD is a core component of the Mouse Genome Informatics (MGI) database

- resource (<http://www.informatics.jax.org>) hosted at The Jackson Laboratory (<http://www.jax.org>).
- Blangiardo, M. et al. (2006), Using a calibration experiment to assess gene-specific information: full Bayesian and empirical Bayesian models for two-channel microarray data. *Bioinformatics* **22**(1), 50–57.
- Brown, W. & Draper, D. (2006), A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis* **1**(3), 473–514.
- Casella, G. & Berger, R. (2002). *Statistical Inference*, second edn, Duxbury, Pacific Grove, CA.
- Chudin, E. et al. (2001), Assessment of the relationship between signal intensities and transcript concentration for Affymetrix GeneChip® arrays. *Genome Biology* **3**(1), 1–10.
- Cuatrecasas, P. (2006), Drug discovery in jeopardy. *J. Clin. Invest.* **116**(11), 2837–2842.
- Davidian, M. (2005), Nonlinear models for univariate and multivariate response. ST 762 Course Notes, North Carolina State University.
- Debouck, C. & Goodfellow, P. (1999), DNA microarrays in drug discovery and development. *Nature Genetics* **21**, 48–50.
- Do, K., Peter, M. & Tang, F. (2005), A Bayesian mixture model for differential gene expression. *Appl. Statist.* **54**(3), 627–644.
- Drews, J. (2000), Drug discovery: A historical perspective. *Science* **287**(5460), 1960–1964.
- Dudoit, S., Shaffer, J. P. & Boldrick, J. C. (2002), Multiple hypothesis testing in microarray experiments. Working Paper 110, Published 2003 in *Statistical Science*, 18: 71–103.
- Durbin, B., Hardin, J., Hawkins, D. & Rocke, D. (2002), A variance-stabilizing transformation for gene-expression microarray data. *Bioinformatics* **18**, S105–S110. Suppl. 1.
- Efron, B., Tibshirani, R., Storey, J. & Tusher, V. (2001), Empirical Bayes analysis of a microarray experiment. *Journal of the American Statistical Association* **96**(456), 1151–1160.

- FDA (2004), Challenge and opportunity on the critical path to new medical products. FDA White Paper.
- Frantz, S. (2004), FDA publishes analysis of the pipeline problem. *Nature Reviews Drug Discovery* **3**(5), 379.
- Freedman, D. A. & Zeisel, H. (1988), From mouse-to-man: The quantitative assessment of cancer risks. *Statistical Science* **3**(1), 3–28.
- Ganter, B., Snyder, R., Halbert, D. & Lee, M. (2006), Toxicogenomics in drug discovery and development: mechanistic analysis of compound/class-dependent effects using the DrugMatrix® database. *Pharmacogenomics* **7**(7), 1025–1044.
- Ganter, B. et al. (2005), Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. *Journal of Biotechnology* **119**(3), 219–244.
- Gong, G. & Samaniego, F. (1981), Pseudo maximum likelihood estimation: theory and applications. *The Annals of Statistics* **9**(4), 861–869.
- Greaves, P., Williams, A. & Eve, M. (2004), First dose of potential new medicines to humans: How animals help. *Nature Reviews Drug Discovery* **3**, 226–236.
- Hughes, T. et al. (2000), Functional discovery via a compendium of expression profiles. *Science* **102**(1), 109–126.
- Jiménez, J., Mitchell, M. & Sgouros, J. (2002), Microarray analysis of orthologous genes: conservation of the translational machinery across species at the sequence and expression level. *Genome Biology* **4**(1), 1–28. Artical R4.
- Kaski, A., Nikkila, J., Sinkkonen, J., Lahti, L., Knuuttila, J. & Roos, C. (2004), Associative clustering for exploring dependencies between functional genomics data sets. *Computational Biology and Bioinformatics* **2**(3), 203–216.
- Keene, O. (1995), The log transformation is special. *Stat Med.* **14**(8), 811–819.
- Lewin, A., Richardson, S., Marshall, C., Glazier, A. & Aitman, T. (2006), Bayesian modeling of differential gene expression. *Biometrics* **62**, 1–9.

- Li, C. & Wong, H. (2001), Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. *Genome Biology* **2**(8), 1–11.
- Linsay, B. (1988), Composite likelihood methods. *Contemporary Mathematics* **80**, 221–239.
- Newton, M., Noueiry, A., Sarkar, D. & Ahlquist, P. (2004), Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics* **5**(2), 155–176.
- Opgen-Rhein & Strimmer (2007), Accurate ranking of differentially expressed genes by a distribution-free shrinkage approach. *Statistical Applications in Genetics and Molecular Biology* **6**(1). Article 9, available at: <http://www.bepress.com/sagmb/vol6/iss1/art9>.
- Orphanides, G. (2003), Toxicogenomics: challenges and opportunities. *Toxicology Letters* **140–141**, 145–148.
- Rawlings, J. O., Pantula, S. & Dickey, D. A. (2001), *Applied Regression Analysis*, second edn, Springer, New York, NY.
- Redner, R. & Walker, H. (1984), Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review* **26**(2), 195–239.
- Rocke, D. & Durbin, B. (2001), A model for measurement error for gene expression arrays. *Journal of Computational Biology* **8**(6), 557–569.
- Sandelin, A., Wasserman, W. & Lenhard, B. (2004), Consite: web-based prediction of regulatory elements using cross-species comparison. *Nucleic Acids Research* **32**, 249–252.
- SAS (2002–2007), Sas OnlineDoc® 9.1.3. <http://support.sas.com/onlinedoc/913>.
- Shaffer, J. (1995), Multiple hypothesis testing. *Annu. Rev. Psychol.* **46**, 561–584.
- Smyth, G. K. (2004), Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* **3**(1), 1–26. Article 3.
- Sonnhammer, E. & Koonin, E. (2002), Orthology, paralogy and proposed classification for paralog subtypes. *Trends in Genetics* **18**(12), 619–620.

- Speed, T. (2000), Always log spot intensities and ratios. The Speed Group Microarray Page: Hints and Prejudices.
- Speed, T. (2007). Email Correspondence.
- Stuart, J., Segal, E., Koller, D. & Kim, S. (2003), A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302**(5643), 249–255.
- Taher, L., Rinner, O., Garg, S., Sczyrba, A. & Morgenstern, B. (2004), Agenda: gene prediction by cross-species sequence comparison. *Nucleic Acids Research* **32**, 305–308.
- Thomas, S. (2006), Animal research and the search for understanding. *Nature Genetics* **38**, 497–498.
- Titterton, D. M., Smith, A. F. M. & Markov, U. E. (1985), *Statistical Analysis of Finite Mixture Distributions*, John Wiley & Sons.
- Tseng, G. et al. (2001), Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Research* **29**(12), 2549–2557.
- Tusher, V. G., Tibshirani, R. & Chu, G. (2001), Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci.* **98**, 5116–5121.
- Ulrich, R. & Friend, S. H. (2002), Toxicogenomics and drug discovery: will new technologies help us produce better drugs? *Nature Reviews Drug Discovery* **1**, 84–88.
- van 't Veer, L. et al. (2002), Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536.
- Waring, J. et al. (2002), Identifying toxic mechanisms using DNA microarrays: evidence that an experimental inhibitor of cell adhesion molecule expression signals through the aryl hydrocarbon nuclear receptor. *Toxicology* **181–182**(1), 537–550.
- Westfall, P. & Young, S. (1993), *Resampling-based multiple testing: Examples and methods for p-value adjustment*, John Wiley & Sons.
- Yekutieli, D. & Benjamini, Y. (1999), Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *Journal of Statistical Planning and Inference* **82**, 171–196.

Appendices

Appendix A

Key Model Derivations

This Appendix provides two key derivations omitted from Chapter 2. Section A.1 includes a derivation of model (2.18) for $(\hat{\beta}_{Mi}, \hat{\beta}_{Hi})$ from the hierarchy of the latent model (2.16) and the errors from linear models (2.1) and (2.2). Section A.2 provides a derivation of the posterior probabilities $\Pr(A_{Hi} = 1)$ and $\Pr(O_i = 1)$, based on model (2.18) and the application of Bayes' rule.

A.1 Derivation of the Distribution of $(\hat{\beta}_{Mi}, \hat{\beta}_{Hi})^T$

Model (2.18) is a hierarchical (multistage) model for the observed treatment coefficient vector $(\hat{\beta}_{Mi}, \hat{\beta}_{Hi})$, with levels corresponding to:

1. the latent treatment coefficients (β_{Mi}, β_{Hi}) , with model (2.16);
2. the least squares error of estimation $(\epsilon_{Mij}, \epsilon_{Hik})$, from models (2.1) and (2.2).

At the first level in the hierarchy, model (2.16) for $(\beta_{Mi}, \beta_{Hi})^T$ is a bivariate normal mixture with five components. The following random variable representation is a complete description of this distribution:

$$\begin{pmatrix} \beta_{Mi} \\ \beta_{Hi} \end{pmatrix} = T_{1i} \begin{pmatrix} x_{1i} \\ y_{1i} \end{pmatrix} + T_{2i} \begin{pmatrix} x_{2i} \\ y_{2i} \end{pmatrix} + T_{3i} \begin{pmatrix} x_{3i} \\ y_{3i} \end{pmatrix} + T_{4i} \begin{pmatrix} x_{4i} \\ y_{4i} \end{pmatrix} + T_{5i} \begin{pmatrix} x_{5i} \\ y_{5i} \end{pmatrix}, \quad (\text{A.1})$$

where $(T_{1i}, T_{2i}, T_{3i}, T_{4i}, T_{5i})$ is a multinomial random vector, of sample size 1, with a distribution described in Table A.1.

Table A.1: The distribution of $(T_{1i}, T_{2i}, T_{3i}, T_{4i}, T_{5i})$

$(T_{1i}, T_{2i}, T_{3i}, T_{4i}, T_{5i})$	$\Pr\{(T_{1i}, T_{2i}, T_{3i}, T_{4i}, T_{5i})\}$
$(1, 0, 0, 0, 0)$	$\theta\eta$
$(0, 1, 0, 0, 0)$	$(1 - \theta)\eta(1 - \eta)$
$(0, 0, 1, 0, 0)$	$(1 - \theta)\eta(1 - \eta)$
$(0, 0, 0, 1, 0)$	$(1 - \theta)\eta^2$
$(0, 0, 0, 0, 1)$	$\{(1 - \theta)(1 - \eta)^2 + (1 - \theta)(1 - \eta)\}$

Furthermore,

$$\begin{aligned}
\begin{pmatrix} x_{1i} \\ y_{1i} \end{pmatrix} &\sim N \left\{ \begin{pmatrix} \mu_M \\ \mu_H \end{pmatrix}, \begin{pmatrix} \sigma_M^2 & \rho\sigma_M\sigma_H \\ \rho\sigma_M\sigma_H & \sigma_H^2 \end{pmatrix} \right\}; \\
\begin{pmatrix} x_{2i} \\ y_{2i} \end{pmatrix} &\sim N \left\{ \begin{pmatrix} \mu_M \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_M^2 & 0 \\ 0 & 0 \end{pmatrix} \right\}; \\
\begin{pmatrix} x_{3i} \\ y_{3i} \end{pmatrix} &\sim N \left\{ \begin{pmatrix} 0 \\ \mu_H \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & \sigma_H^2 \end{pmatrix} \right\}; \\
\begin{pmatrix} x_{4i} \\ y_{4i} \end{pmatrix} &\sim N \left\{ \begin{pmatrix} \mu_M \\ \mu_H \end{pmatrix}, \begin{pmatrix} \sigma_M^2 & 0 \\ 0 & \sigma_H^2 \end{pmatrix} \right\}; \\
\begin{pmatrix} x_{5i} \\ y_{5i} \end{pmatrix} &\sim N(\mathbf{0}, \mathbf{0}).
\end{aligned}$$

A property of the least squares estimators $\hat{\beta}_{Mi}$ and $\hat{\beta}_{Hi}$ is that

$$\begin{pmatrix} \hat{\beta}_{Mi} \\ \hat{\beta}_{Hi} \end{pmatrix} \mid \begin{pmatrix} \beta_{Mi} \\ \beta_{Hi} \end{pmatrix} \sim N \left[\begin{pmatrix} \beta_{Mi} \\ \beta_{Hi} \end{pmatrix}, \begin{pmatrix} \sigma_{1i}^2 & 0 \\ 0 & \sigma_{2i}^2 \end{pmatrix} \right],$$

or in terms of random variables,

$$\begin{pmatrix} \hat{\beta}_{Mi} \\ \hat{\beta}_{Hi} \end{pmatrix} = \begin{pmatrix} \beta_{Mi} \\ \beta_{Hi} \end{pmatrix} + \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix}.$$

The random variable representation of $(\beta_{Mi}, \beta_{Hi})^T$ facilitates

$$\begin{pmatrix} \hat{\beta}_{Mi} \\ \hat{\beta}_{Hi} \end{pmatrix} = T_{1i} \left\{ \begin{pmatrix} x_{1i} \\ y_{1i} \end{pmatrix} + \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \right\} + T_{2i} \left\{ \begin{pmatrix} x_{2i} \\ y_{2i} \end{pmatrix} + \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \right\} + \\ T_{3i} \left\{ \begin{pmatrix} x_{3i} \\ y_{3i} \end{pmatrix} + \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \right\} + T_{4i} \left\{ \begin{pmatrix} x_{4i} \\ y_{4i} \end{pmatrix} + \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \right\} + \\ T_{5i} \left\{ \begin{pmatrix} x_{5i} \\ y_{5i} \end{pmatrix} + \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \right\}.$$

Since $(x_{ji}, y_{ji})^T$ is assumed independent of $(\epsilon_{1i}, \epsilon_{2i})^T$ for $j = 1, \dots, 5$, and these vectors are bivariate normal, the distribution of any sum is easily computed:

$$\begin{aligned} \left\{ \begin{pmatrix} x_{1i} \\ y_{1i} \end{pmatrix} + \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \right\} &\sim N \left\{ \begin{pmatrix} \mu_M \\ \mu_H \end{pmatrix}, \begin{pmatrix} \sigma_M^2 + \sigma_{1i}^2 & \rho\sigma_M\sigma_H \\ \rho\sigma_M\sigma_H & \sigma_H^2 + \sigma_{2i}^2 \end{pmatrix} \right\}; \\ \left\{ \begin{pmatrix} x_{2i} \\ y_{2i} \end{pmatrix} + \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \right\} &\sim N \left\{ \begin{pmatrix} \mu_M \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_M^2 + \sigma_{1i}^2 & 0 \\ 0 & \sigma_{2i}^2 \end{pmatrix} \right\}; \\ \left\{ \begin{pmatrix} x_{3i} \\ y_{3i} \end{pmatrix} + \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \right\} &\sim N \left\{ \begin{pmatrix} 0 \\ \mu_H \end{pmatrix}, \begin{pmatrix} \sigma_{1i}^2 & 0 \\ 0 & \sigma_H^2 + \sigma_{2i}^2 \end{pmatrix} \right\}; \\ \left\{ \begin{pmatrix} x_{4i} \\ y_{4i} \end{pmatrix} + \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \right\} &\sim N \left\{ \begin{pmatrix} \mu_M \\ \mu_H \end{pmatrix}, \begin{pmatrix} \sigma_M^2 + \sigma_{1i}^2 & 0 \\ 0 & \sigma_H^2 + \sigma_{2i}^2 \end{pmatrix} \right\}; \\ \left\{ \begin{pmatrix} x_{5i} \\ y_{5i} \end{pmatrix} + \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \right\} &\sim N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{1i}^2 & 0 \\ 0 & \sigma_{2i}^2 \end{pmatrix} \right\}. \end{aligned}$$

Thus the distribution of $(\hat{\beta}_{Mi} \ \hat{\beta}_{Hi})^T$ is

$$\begin{aligned}
\begin{pmatrix} \hat{\beta}_{Mi} \\ \hat{\beta}_{Hi} \end{pmatrix} &\sim \Pr(T_{1i} = 1) N \left\{ \begin{pmatrix} \mu_M \\ \mu_H \end{pmatrix}, \begin{pmatrix} \sigma_M^2 + \sigma_{1i}^2 & p \sigma_H \sigma_M \\ p \sigma_H \sigma_M & \sigma_H^2 + \sigma_{2i}^2 \end{pmatrix} \right\} \\
&+ \Pr(T_{2i} = 1) N \left\{ \begin{pmatrix} \mu_M \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_M^2 + \sigma_{1i}^2 & 0 \\ 0 & \sigma_{2i}^2 \end{pmatrix} \right\} \\
&+ \Pr(T_{3i} = 1) N \left\{ \begin{pmatrix} 0 \\ \mu_H \end{pmatrix}, \begin{pmatrix} \sigma_{1i}^2 & 0 \\ 0 & \sigma_H^2 + \sigma_{2i}^2 \end{pmatrix} \right\} \\
&+ \Pr(T_{4i} = 1) N \left\{ \begin{pmatrix} \mu_M \\ \mu_H \end{pmatrix}, \begin{pmatrix} \sigma_M^2 + \sigma_{1i}^2 & 0 \\ 0 & \sigma_H^2 + \sigma_{2i}^2 \end{pmatrix} \right\} \\
&+ \Pr(T_{5i} = 1) N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{1i}^2 & 0 \\ 0 & \sigma_{2i}^2 \end{pmatrix} \right\}.
\end{aligned} \tag{A.2}$$

The values $\Pr(T_{ji} = 1)$, for $j = 1, \dots, 5$, are included in Table A.1, and this completes the derivation.

A.2 Derivation of $\Pr(A_{Hi} = 1)$ and $\Pr(O_i = 1)$

In this section we derive two key posterior probabilities using Bayes' rule:

1. $\Pr(A_{Hi} = 1 | \hat{\beta}_{Mi}, \hat{\beta}_{Hi}) = \frac{f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}, A_{Hi}}(x_i, y_i, 1)}{f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}}(x_i, y_i)}$;
2. $\Pr(O_i = 1 | \hat{\beta}_{Mi}, \hat{\beta}_{Hi}) = \frac{f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}, O_i}(x_i, y_i, 1)}{f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}}(x_i, y_i)}$.

The denominator density $f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}}(x_i, y_i)$ is given in (2.19) and discussed in Section 2.3.

Furthermore, refer to Sections 2.1 and 2.2. for modeling descriptions of

$$f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi} | O_i, A_{Mi}, A_{Hi}}(x_i, y_i | O_i = j, A_{Mi} = k, A_{Hi} = l) \text{ for } j, k, l \in \{0, 1\},$$

and

$$\Pr(O_i = j, A_{Mi} = k, A_{Hi} = l) \text{ for } j, k, l \in \{0, 1\},$$

respectively.

Let $A_{Hi} = 1$. Then

$$\begin{aligned}
& f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}, A_{Hi}}(x_i, y_i, 1) \\
&= \sum_{j=0}^1 \sum_{k=0}^1 f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}, O_i, A_{Mi}, A_{Hi}}(x_i, y_i, j, k, 1) \\
&= \sum_{j=0}^1 \sum_{k=0}^1 f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi} | O_i, A_{Mi}, A_{Hi}}(x_i, y_i | O_i = j, A_{Mi} = k, A_{Hi} = 1) \times \\
&\quad \Pr(O_i = j, A_{Mi} = k, A_{Hi} = 1).
\end{aligned}$$

Since $\Pr(O_i = 1, A_{Mi} = 0, A_{Hi} = 1) = 0$,

$$\begin{aligned}
& f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}, A_{Hi}}(x_i, y_i, 1) \\
&= f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi} | O_i, A_{Mi}, A_{Hi}}(x_i, y_i | O_i = 0, A_{Mi} = 0, A_{Hi} = 1) \Pr(O_i = 0, A_{Mi} = 0, A_{Hi} = 1) \\
&\quad + f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi} | O_i, A_{Mi}, A_{Hi}}(x_i, y_i | O_i = 0, A_{Mi} = 1, A_{Hi} = 1) \Pr(O_i = 0, A_{Mi} = 1, A_{Hi} = 1) \\
&\quad + f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi} | O_i, A_{Mi}, A_{Hi}}(x_i, y_i | O_i = 1, A_{Mi} = 1, A_{Hi} = 1) \Pr(O_i = 1, A_{Mi} = 1, A_{Hi} = 1),
\end{aligned}$$

completing the first part of the derivation. Similarly, let $O_i = 1$. Then

$$\begin{aligned}
& f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}, O_i}(x_i, y_i, 1) \\
&= \sum_{k=0}^1 \sum_{l=0}^1 f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}, O_i, A_{Mi}, A_{Hi}}(x_i, y_i, 1, k, l) \\
&= \sum_{k=0}^1 \sum_{l=0}^1 f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi} | O_i, A_{Mi}, A_{Hi}}(x_i, y_i | O_i = 1, A_{Mi} = k, A_{Hi} = l) \times \\
&\quad \Pr(O_i = 1, A_{Mi} = k, A_{Hi} = l).
\end{aligned}$$

Also, since $\Pr(O_i = 1, A_{Mi} = 0, A_{Hi} = 1) = \Pr(O_i = 1, A_{Mi} = 1, A_{Hi} = 1) = 0$,

$$\begin{aligned}
& f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}, O_i}(x_i, y_i, 1) \\
&= f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi} | O_i, A_{Mi}, A_{Hi}}(x_i, y_i | O_i = 0, A_{Mi} = 0, A_{Hi} = 0) \Pr(O_i = 0, A_{Mi} = 0, A_{Hi} = 0) \\
&\quad + f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi} | O_i, A_{Mi}, A_{Hi}}(x_i, y_i | O_i = 1, A_{Mi} = 1, A_{Hi} = 1) \Pr(O_i = 1, A_{Mi} = 1, A_{Hi} = 1),
\end{aligned}$$

completing the derivation.

Appendix B

Identifiability

This appendix contains theoretical arguments that seek to establish the identifiability of model (2.18). The strategy for establishing this identifiability, based on Theorem B.0.1, is to prove the existence of closed-form, asymptotically-unbiased estimates for every parameter in (2.18). Section B.1 discusses the usual least-squares estimators of σ_{1i} and σ_{2i} . Section B.2 derives moment estimators of μ_M , σ_M , μ_H , σ_H and η , which can be obtained from the mouse and human datasets in isolation. Finally, asymptotically-unbiased estimators for the joint dependence parameters ρ and θ are introduced in Section B.3.

The following theorem establishes the link between asymptotic estimability and identifiability.

Theorem B.0.1 *Let $\{F(x; \theta) : \theta \in \Theta\}$ be a parametric family and $x_1, \dots, x_n \stackrel{i.i.d.}{\sim} F(x; \theta)$. If there exists a function $S(x_1, \dots, x_n)$ such that $\lim_{n \rightarrow \infty} E\{S(x_1, \dots, x_n)\} = \theta$ for all $\theta \in \Theta$, then $\{F(x; \theta) : \theta \in \Theta\}$ is an identifiable parametric family.*

Proof Suppose there exists θ_1 and θ_2 such that $\theta_1 \neq \theta_2$ and $F(x, \theta_1) = F(x, \theta_2)$ for all $x \in \chi$. By assumption and using $F(x, \theta_1) = F(x, \theta_2)$,

$$\begin{aligned} \lim_{n \rightarrow \infty} E\{S(x_1, \dots, x_n)\} &= \lim_{n \rightarrow \infty} \int_{\chi^n} S(x_1, \dots, x_n) d\left\{\prod_{i=1}^n F(x_i, \theta_1)\right\} = \theta_1, \\ \lim_{n \rightarrow \infty} E\{S(x_1, \dots, x_n)\} &= \lim_{n \rightarrow \infty} \int_{\chi^n} S(x_1, \dots, x_n) d\left\{\prod_{i=1}^n F(x_i, \theta_2)\right\} = \theta_2, \end{aligned}$$

resulting in $0 = \theta_1 - \theta_2$. By assumption, $\theta_1 \neq \theta_2$ and hence a contradiction. Consequently, if $\theta_1 \neq \theta_2$, then $F(x, \theta_1) \neq F(x, \theta_2)$ for some x . In other words, $\{F(x; \theta) : \theta \in \Theta\}$ is an identifiable parametric family.

Thus, Theorem B.0.1 states that a parametric model is identifiable if it is possible to find asymptotically-unbiased (not necessarily zero-variance) estimates for all model parameters.

B.1 Estimability of Gene Level Parameters

Though the gene level parameters σ_{1i} and σ_{2i} are numerous, they are easily handled with least squares regression methods. Recall the linear models for y_{Mij} and y_{Hik} used in this research,

$$y_{Mij} = \alpha_{Mi} + \beta_{Mi}T_{Mj} + \sigma_{Mi}\epsilon_{ij}, \quad j = 1, \dots, 19, \quad (\text{B.1})$$

and

$$y_{Hik} = \alpha_{Hi} + \beta_{Hi}T_{Hk} + \gamma_i^T \mathbf{z}_{ik} + \sigma_{Hi}\epsilon_{ik}, \quad k = 1, \dots, 24, \quad (\text{B.2})$$

for the i^{th} gene, the j^{th} mouse and k^{th} human, where $i = 1, \dots, 7410$, $\mathbf{z}_{ik}$ is a 5×1 vector of human covariates, T_{Mj} and T_{Hk} are $\{0,1\}$ treatment indicators, and ϵ_{ij} and ϵ_{ik} are independent $N(0,1)$ random variables.

Focusing on the design matrices of (B.1) and (B.2), let $\mathbf{1}_m$ be an $m \times 1$ vector consisting of ones and let $\mathbf{T}_M$ be the $m \times 1$ vector of treatment indicator variables for the mice. The design matrix for the mouse model is

$$\mathbf{X}_M = (\mathbf{1}_m, \mathbf{T}_M).$$

Furthermore, let $\mathbf{1}_n$ be an $n \times 1$ vector consisting of ones, $\mathbf{T}_H$ be the $n \times 1$ vector of treatment indicator variables for the humans and $\mathbf{Z}_i$ be a $n \times 5$ matrix such that the k^{th} row of $\mathbf{Z}_i$ is $\mathbf{z}_{ik}^T$. The design matrix for the human model is

$$\mathbf{X}_{Hi} = (\mathbf{1}_n, \mathbf{T}_H, \mathbf{Z}_i).$$

and the variance of $\hat{\beta}_{Mi}$ is

$$\begin{aligned} \sigma_{1i}^2 &= \left[(\mathbf{X}_M^T \mathbf{X}_M)^{-1} \right]_{(2,2)} \sigma_{Mi}^2 \\ &= \left(\frac{m_p + m_t}{m_p m_t} \right) \sigma_{Mi}^2, \end{aligned}$$

where m_p is the number of placebo mice and m_t is the number of treatment mice. The variance of $\hat{\beta}_{Hi}$ is

$$\sigma_{2i}^2 = \left[(\mathbf{X}_{Hi}^T \mathbf{X}_{Hi})^{-1} \right]_{(2,2)} \sigma_{Hi}^2.$$

Finally,

$$\hat{\sigma}_{Mi}^2 = \left(\frac{1}{m-2} \right) \mathbf{Y}_{Mi}^T \left[\mathbf{I}_m - (\mathbf{X}_M^T \mathbf{X}_M)^{-1} \mathbf{X}_M^T \right] \mathbf{Y}_{Mi} \quad (\text{B.3})$$

and

$$\hat{\sigma}_{Hi}^2 = \left(\frac{1}{n-7} \right) \mathbf{Y}_{Hi}^T \left[\mathbf{I}_n - (\mathbf{X}_{Hi}^T \mathbf{X}_{Hi})^{-1} \mathbf{X}_{Hi}^T \right] \mathbf{Y}_{Hi} \quad (\text{B.4})$$

have the property that $E(\hat{\sigma}_{Mi}^2) = \sigma_{Mi}^2$ and $E(\hat{\sigma}_{Hi}^2) = \sigma_{Hi}^2$ (Rawlings et al, 2001), where $\mathbf{Y}_{Mi}$ is an $m \times 1$ vector consisting of Y_{Mij} , $j = 1, \dots, m$, $\mathbf{Y}_{Hi}$ is an $n \times 1$ vector consisting of Y_{Hik} , $k = 1, \dots, n$, and $\mathbf{I}_s$ is an $s \times s$ identity matrix.

B.2 Asymptotic Estimability of Marginal Parameters

For a single gene pair i , consider the construction of an approximate $(1 - \alpha)100\%$ confidence interval for β_{Mi} ,

$$\hat{\beta}_{Mi} \pm \hat{\sigma}_{1i} z_{\alpha/2}, \quad (\text{B.5})$$

where $z_{\alpha/2}$ satisfies $\Pr(Z > z_{\alpha/2}) = \frac{\alpha}{2}$ and $Z \sim N(0, 1)$. Additionally, consider the indicator

$$C_{Mi} = I(\hat{\beta}_{Mi} - \hat{\sigma}_{1i} z_{\alpha/2} \leq 0 \leq \hat{\beta}_{Mi} + \hat{\sigma}_{1i} z_{\alpha/2}), \quad (\text{B.6})$$

which equals 1 if $\{0\}$ is in confidence interval (B.5) and 0 otherwise. Suppose there exists an asymptotically-unbiased estimate $\hat{p}_{(0)}$ of $p_{(0)} = \Pr\{\beta_{Mi} = 0\}$, the proportion of treatment-unaffected genes (i.e., genes for which $\beta_{Mi} = 0$). Then a consistent estimate of η is

$$\hat{\eta} = 1 - \hat{p}_{(0)}. \quad (\text{B.7})$$

The indicator A_{Mi} takes the value 1 when $\beta_{Mi} \neq 0$ and 0 when $\beta_{Mi} = 0$. Thus

$$\begin{aligned} E(C_{Mi}) &= \Pr\{C_{Mi} = 1\} \\ &= \Pr\{C_{Mi} = 1 | A_{Mi} = 1\} \{1 - p_{(0)}\} + \Pr\{C_{Mi} = 1 | A_{Mi} = 0\} p_{(0)}. \end{aligned} \quad (\text{B.8})$$

Suppose $A_{Mi} = 1$ and define $r_i = \beta_{Mi}$. By assumption, $r_i \neq 0$ w.p.1. Since $\hat{\beta}_{Mi} | \{\beta_{Mi} = r_i\} \sim N(r_i, \sigma_{1i})$,

$$\begin{aligned} \Pr\{C_{Mi} = 1 | \beta_{Mi} = r_i\} &= \Pr\{-\hat{\sigma}_{1i} z_{\alpha/2} \leq \hat{\beta}_{Mi} \leq \hat{\sigma}_{1i} z_{\alpha/2} | \beta_{Mi} = r_i\} \\ &= \Pr\left\{ \frac{-r_i - \hat{\sigma}_{1i} z_{\alpha/2}}{\sigma_{1i}} \leq \frac{\hat{\beta}_{Mi} - r_i}{\sigma_{1i}} \leq \frac{-r_i - \hat{\sigma}_{1i} z_{\alpha/2}}{\sigma_{1i}} \right\} \\ &= \Phi\left(-\frac{r_i}{\sigma_{1i}} + \frac{\hat{\sigma}_{1i}}{\sigma_{1i}} z_{\alpha/2}\right) - \Phi\left(-\frac{r_i}{\sigma_{1i}} - \frac{\hat{\sigma}_{1i}}{\sigma_{1i}} z_{\alpha/2}\right), \end{aligned}$$

where $\Phi(\cdot)$ is the c.d.f. of an $N(0, 1)$ random variable. As $m \rightarrow \infty$,

$$\hat{\sigma}_{1i} \rightarrow \sigma_{1i} \text{ in probability,} \quad (\text{B.9})$$

$$\sigma_{1i} \rightarrow 0, \quad (\text{B.10})$$

$$\left(-\frac{r_i}{\sigma_{1i}} \right) \rightarrow \begin{cases} \infty & \text{if } r_i < 0 \\ -\infty & \text{if } r_i > 0. \end{cases} \quad (\text{B.11})$$

Suppose that $r_i < 0$. Then

$$\begin{aligned} \lim_{m \rightarrow \infty} \Pr\{C_{Mi} = 1 | \beta_{Mi} = r_i\} &= \Phi(\infty) - \Phi(\infty) \\ &= 1 - 1 \\ &= 0 \end{aligned}$$

Suppose alternately that $r_i > 0$. Then

$$\begin{aligned} \lim_{m \rightarrow \infty} \Pr\{C_{Mi} = 1 | \beta_{Mi} = r_i\} &= \Phi(-\infty) - \Phi(-\infty) \\ &= 0 - 0 \\ &= 0 \end{aligned}$$

Thus

$$\lim_{m \rightarrow \infty} \Pr\{C_{Mi} = 1 | \beta_{Mi} = r_i\} = 0,$$

for $r_i \neq 0$, and therefore

$$\lim_{m \rightarrow \infty} \Pr\{C_{Mi} = 1 | A_{Mi} = 1\} = 0.$$

Now consider the case where $\beta_{Mi} = 0$. Then

$$\Pr\{C_{Mi} = 1 | A_{Mi} = 0\} = \left\{ \Phi\left(\frac{\hat{\sigma}_{1i}}{\sigma_{1i}} z_{\alpha/2}\right) - \Phi\left(-\frac{\hat{\sigma}_{1i}}{\sigma_{1i}} z_{\alpha/2}\right) \right\}.$$

Taking limits,

$$\begin{aligned} \lim_{m \rightarrow \infty} \Pr\{C_{Mi} = 1 | A_{Mi} = 0\} &= \Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2}) \\ &= (1 - \alpha). \end{aligned}$$

Returning to (B.8), the limiting expression becomes

$$\lim_{m \rightarrow \infty} E(C_{Mi}) = (1 - \alpha)p_{(0)}, \quad (\text{B.12})$$

suggesting the estimate

$$\hat{p}_{(0)} = \frac{1}{(1-\alpha)} \frac{1}{N} \sum_{i=1}^N C_{Mi}. \quad (\text{B.13})$$

As the number of genes become large, $N^{-1} \sum_{i=1}^N C_{Mi}$ converges to $E(C_{Mi})$ almost surely. Thus $\hat{p}_{(0)}$ is asymptotically-unbiased for $p_{(0)}$, completing the estimation of η .

The following moment equations resulting from (2.18) provide simple moment estimates for parameters $\mu_M, \mu_H, \sigma_M^2, \sigma_H^2$, given an estimate of η :

$$E(\hat{\beta}_{Mi}) = \eta^{-1} \mu_M, \quad (\text{B.14})$$

$$E(\hat{\beta}_{Hi}) = \eta^{-1} \mu_H, \quad (\text{B.15})$$

$$E(\hat{\beta}_{Mi}^2) = \sigma_{1i}^2 + \eta(\mu_M^2 + \sigma_M^2), \quad (\text{B.16})$$

$$E(\hat{\beta}_{Hi}^2) = \sigma_{2i}^2 + \eta(\mu_H^2 + \sigma_H^2). \quad (\text{B.17})$$

These moment equations lead to the following set of asymptotically-unbiased estimates, using $\hat{\eta}$ from (B.7):

$$\hat{\mu}_M = (\hat{\eta})^{-1} \frac{1}{N} \sum_{i=1}^N \hat{\beta}_{Mi}, \quad (\text{B.18})$$

$$\hat{\mu}_H = (\hat{\eta})^{-1} \frac{1}{N} \sum_{i=1}^N \hat{\beta}_{Hi}, \quad (\text{B.19})$$

$$\hat{\sigma}_M^2 = (\hat{\eta})^{-1} \frac{1}{N} \sum_{i=1}^N (\hat{\beta}_{Mi}^2 - \hat{\sigma}_{1i}^2) - (\hat{\mu}_M)^2, \quad (\text{B.20})$$

$$\hat{\sigma}_H^2 = (\hat{\eta})^{-1} \frac{1}{N} \sum_{i=1}^N (\hat{\beta}_{Hi}^2 - \hat{\sigma}_{2i}^2) - (\hat{\mu}_H)^2. \quad (\text{B.21})$$

B.3 Asymptotic Estimability of Dependence Parameters

For a single gene pair i , consider the construction of two approximate $(1-\alpha)100\%$ confidence intervals for β_{Mi} and β_{Hi} ,

$$\hat{\beta}_{Mi} \pm \hat{\sigma}_{1i} z_{\alpha/2}, \quad (\text{B.22})$$

$$\hat{\beta}_{Hi} \pm \hat{\sigma}_{2i} z_{\alpha/2}, \quad (\text{B.23})$$

where $z_{\alpha/2}$ satisfies $\Pr(Z > z_{\alpha/2}) = \alpha/2$ and $Z \sim N(0, 1)$. Furthermore, consider the indicators

$$C_{Mi} = I(\hat{\beta}_{Mi} - \hat{\sigma}_{1i}z_{\alpha/2} \leq 0 \leq \hat{\beta}_{Mi} + \hat{\sigma}_{1i}z_{\alpha/2}), \quad (\text{B.24})$$

$$C_{Hi} = I(\hat{\beta}_{Hi} - \hat{\sigma}_{2i}z_{\alpha/2} \leq 0 \leq \hat{\beta}_{Hi} + \hat{\sigma}_{2i}z_{\alpha/2}). \quad (\text{B.25})$$

Note that the quantity $C_{Mi}C_{Hi}$ is 1 when $(\beta_{Mi}, \beta_{Hi}) = (0, 0)$ is in the confidence rectangle formed by (B.22) and (B.23), and 0 otherwise.

If there exists an asymptotically unbiased estimate, $\hat{p}_{(0,0)}$, of

$$p_{(0,0)} = \Pr\{(\beta_{Mi}, \beta_{Hi}) = (0, 0)\},$$

then (2.16) provides an asymptotically unbiased estimate of $[(1 - \theta)(1 - \eta)^2 + \theta(1 - \eta)]$. Given $\hat{\eta}$ from Section B.2,

$$\hat{\theta} = \frac{\hat{p}_{(0,0)} - (1 - \hat{\eta})^2}{\hat{\eta}(1 - \hat{\eta})} \quad (\text{B.26})$$

is an asymptotically unbiased estimate of θ .

Recall that $A_{Mi} = I(\beta_{Mi} \neq 0)$ and $A_{Hi} = I(\beta_{Hi} \neq 0)$. Then

$$\begin{aligned} E(C_{Mi}C_{Hi}) &= \Pr\{C_{Hi} = 1, C_{Mi} = 1\} \\ &= \Pr\{C_{Hi} = 1, C_{Mi} = 1 | (A_{Mi}, A_{Hi}) = (0, 0)\} p_{(0,0)} + \\ &+ \Pr\{C_{Hi} = 1, C_{Mi} = 1 | (A_{Mi}, A_{Hi}) = (0, 1)\} \Pr\{(A_{Mi}, A_{Hi}) = (0, 1)\} \\ &+ \Pr\{C_{Hi} = 1, C_{Mi} = 1 | (A_{Mi}, A_{Hi}) = (1, 0)\} \Pr\{(A_{Mi}, A_{Hi}) = (1, 0)\} \\ &+ \Pr\{C_{Hi} = 1, C_{Mi} = 1 | (A_{Mi}, A_{Hi}) = (1, 1)\} \Pr\{(A_{Mi}, A_{Hi}) = (1, 1)\}. \end{aligned}$$

Suppose $A_{Mi} = 1$ and define $r_i = \beta_{Mi}$. By assumption, $r_i \neq 0$ w.p.1. Since $\hat{\beta}_{Mi} | \{\beta_{Mi} = r_i\} \sim N(r_i, \sigma_{1i})$,

$$\begin{aligned} \Pr\{C_{Mi} = 1 | \beta_{Mi} = r_i\} &= \Pr\{-\hat{\sigma}_{1i}z_{\alpha/2} \leq \hat{\beta}_{Mi} \leq \hat{\sigma}_{1i}z_{\alpha/2} | \beta_{Mi} = r_i\} \\ &= \Pr\left\{\frac{-r_i - \hat{\sigma}_{1i}z_{\alpha/2}}{\sigma_{1i}} \leq \frac{\hat{\beta}_{Mi} - r_i}{\sigma_{1i}} \leq \frac{-r_i - \hat{\sigma}_{1i}z_{\alpha/2}}{\sigma_{1i}}\right\} \\ &= \Phi\left(-\frac{r_i}{\sigma_{1i}} + \frac{\hat{\sigma}_{1i}}{\sigma_{1i}}z_{\alpha/2}\right) - \Phi\left(-\frac{r_i}{\sigma_{1i}} - \frac{\hat{\sigma}_{1i}}{\sigma_{1i}}z_{\alpha/2}\right), \end{aligned}$$

where $\Phi(\cdot)$ is the c.d.f. of a $N(0, 1)$ random variable. As $m \rightarrow \infty$,

$$\hat{\sigma}_{1i} \rightarrow \sigma_{1i} \text{ in probability,} \quad (\text{B.27})$$

$$\sigma_{1i} \rightarrow 0, \quad (\text{B.28})$$

$$\left(-\frac{r_i}{\sigma_{1i}}\right) \rightarrow \begin{cases} \infty & \text{if } r_i < 0, \\ -\infty & \text{if } r_i > 0. \end{cases} \quad (\text{B.29})$$

Suppose that $r_i < 0$. Then

$$\begin{aligned} \lim_{m \rightarrow \infty} \Pr\{C_{Mi} = 1 | \beta_{Mi} = r_i\} &= \Phi(\infty) - \Phi(\infty) \\ &= 1 - 1 \\ &= 0. \end{aligned}$$

Suppose alternately that $r_i > 0$. Then

$$\begin{aligned} \lim_{m \rightarrow \infty} \Pr\{C_{Mi} = 1 | \beta_{Mi} = r_i\} &= \Phi(-\infty) - \Phi(-\infty) \\ &= 0 - 0 \\ &= 0. \end{aligned}$$

Thus

$$\lim_{m \rightarrow \infty} \Pr\{C_{Mi} = 1 | \beta_{Mi} = r_i\} = 0.$$

Since C_{Mi} is conditionally independent of β_{Hi} given β_{Mi} ,

$$\begin{aligned} \Pr\{C_{Mi} = 1 | \beta_{Mi} = r_i\} &= \Pr\{C_{Mi} = 1 | \beta_{Mi} = r_i, \beta_{Hi} = s_i\} \\ &\geq \Pr\{C_{Mi} = 1, C_{Hi} = 1 | \beta_{Mi} = r_i, \beta_{Hi} = s_i\}, \end{aligned}$$

for arbitrary $s_i \in \mathbb{R}$. The arbitrary nature of s_i implies that

$$\begin{aligned} \lim_{m \rightarrow \infty} \Pr\{C_{Mi} = 1, C_{Hi} = 1 | A_{Mi} = 1, A_{Hi} = 1\} &= 0, \\ \lim_{m \rightarrow \infty} \Pr\{C_{Mi} = 1, C_{Hi} = 1 | A_{Mi} = 1, A_{Hi} = 0\} &= 0, \end{aligned}$$

by the so-called “squeeze theorem.” Similarly, starting the above argument by supposing $A_{Hi} = 1$ leads to

$$\lim_{m \rightarrow \infty} \Pr\{C_{Mi} = 1, C_{Hi} = 1 | A_{Mi} = 0, A_{Hi} = 1\} = 0.$$

The case where $(\beta_{Mi}, \beta_{Hi}) = (0, 0)$ is of a different nature.

$$\begin{aligned}
& \Pr\{C_{Hi} = 1, C_{Mi} = 1 | (A_{Mi}, A_{Hi}) = (0, 0)\} \\
&= \Pr\{C_{Mi} = 1 | \beta_{Mi} = 0\} \Pr\{C_{Hi} = 1 | \beta_{Hi} = 0\} \\
&= \left\{ \Phi\left(\frac{\hat{\sigma}_{1i}}{\sigma_{1i}} z_{\alpha/2}\right) - \Phi\left(-\frac{\hat{\sigma}_{1i}}{\sigma_{1i}} z_{\alpha/2}\right) \right\} \left\{ \Phi\left(\frac{\hat{\sigma}_{2i}}{\sigma_{2i}} z_{\alpha/2}\right) - \Phi\left(-\frac{\hat{\sigma}_{2i}}{\sigma_{2i}} z_{\alpha/2}\right) \right\}.
\end{aligned}$$

Taking limits,

$$\begin{aligned}
\lim_{m \rightarrow \infty} \Pr\{C_{Hi} = 1, C_{Mi} = 1 | (A_{Mi}, A_{Hi}) = (0, 0)\} &= \left\{ \Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2}) \right\}^2 \\
&= (1 - \alpha)^2.
\end{aligned}$$

We finally arrive at the limiting expression

$$\lim_{m \rightarrow \infty} E(C_{Mi} C_{Hi}) = (1 - \alpha)^2 p_{(0,0)}, \quad (\text{B.30})$$

suggesting the estimate

$$\hat{p}_{(0,0)} = \frac{1}{(1 - \alpha)^2} \frac{1}{N} \sum_{i=1}^N C_{Mi} C_{Hi}. \quad (\text{B.31})$$

As the number of genes become large, $N^{-1} \sum_{i=1}^N C_{Mi} C_{Hi}$ converges to $E(C_{Mi} C_{Hi})$ almost surely. Thus $\hat{p}_{(0,0)}$ is asymptotically-unbiased for $p_{(0,0)}$ and this completes the estimation of θ .

The last parameter is ρ , the correlation between the latent β_{Mi} and β_{Hi} of I-Ortholog gene pairs affected by treatment. Consider the expectation of $\hat{\beta}_{Mi} \hat{\beta}_{Hi}$,

$$\begin{aligned}
E(\hat{\beta}_{Mi} \hat{\beta}_{Hi}) &= \theta \eta (\mu_M \mu_H + \rho \sigma_M \sigma_H) + (1 - \theta) \eta^2 \mu_M \mu_H \\
&= \theta \eta \sigma_M \sigma_H \rho + \{\theta \eta + (1 - \theta) \eta^2\} \mu_M \mu_H.
\end{aligned}$$

Thus, an asymptotically unbiased estimating equation for ρ is

$$\hat{\rho}^{MOM} = \frac{1}{\sigma_H \sigma_M \theta \eta} \left\{ \frac{1}{N} \sum_{i=1}^N \hat{\beta}_{Mi} \hat{\beta}_{Hi} - \mu_M \mu_H [\theta \eta + (1 - \theta) \eta^2] \right\}. \quad (\text{B.32})$$

With the estimates for $\mu_M, \mu_H, \sigma_M, \sigma_H, \eta$ and θ as presented in this Appendix, we arrive at the following asymptotically-unbiased estimate of ρ ,

$$\hat{\rho} = \frac{1}{\hat{\sigma}_H \hat{\sigma}_M \hat{\theta} \hat{\eta}} \left\{ \frac{1}{N} \sum_{i=1}^N \hat{\beta}_{Mi} \hat{\beta}_{Hi} - \hat{\mu}_M \hat{\mu}_H [\hat{\theta} \hat{\eta} + (1 - \hat{\theta}) \hat{\eta}^2] \right\}. \quad (\text{B.33})$$

This completes the discussion on the identifiability of model (2.18).

Appendix C

Microarray Data Transformation

Gene expression measurements y_{Mij} and y_{Hik} (Section 1.5) are positive, continuous data, and thus the question of whether to log-transform these measurements naturally arises. Affymetrix’s *Data Analysis Fundamentals*, citing Speed (2000), recommends log-transforming gene expression data on the grounds that

- signal intensity is proportional to error magnitude,
- the log transformation symmetrises data,
- the log transformation reduces the effect of outliers.

In a personal email correspondence with Terry Speed in October, 2007, Speed reasoned that the “fold-change” interpretation provided by the log scale is more intuitive than a difference in means, and gave two microarray-specific reasons for use of the log transformation. First, the log transformation largely alleviates the so-called Behrens-Fisher problem of evaluating t -statistics with unequal variances. Such t -statistics would be approximately i.i.d. under the null hypothesis for log-transformed gene expression data, greatly simplifying analysis. Second, data preprocessing steps (such as those implemented by the Affymetrix MAS 5.0 software in the case of the GlaxoSmithKline data) typically use multiplicative models to remove “probe affinities,” so the resulting signal intensities have already gone in and come out of a logarithmic model.

Despite convincing arguments from Speed (2000 and 2007), the choice of whether to use the log transformation is not straightforward. Durbin et al. (2002) state that the use of log transformations for microarray data “is subject to a number of problems,” including

substantial variance inflation in genes with small average expression levels. They conclude that “though log transformations approximately stabilize the variance of data expressed at high levels, their performance on data at or near background leave much to be desired.” In a laboratory experiment involving “spike-in” hybridizations and four prokaryotic transcripts, Chudin et al. (2001) note that magnitudes of fold changes are “not always linearly related to target concentration and [are] extremely variable for low transcript concentrations,” and additionally fault the log transformation for giving a “misleading impression” in a visual of among-sample uniformity at regions of low gene expression. It is important to note that Durbin et al. (2002) developed a modified log transformation based on an earlier linear/loglinear “combo” model (Rocke and Durbin, 2001), but this method poses unique challenges in parameter estimation.

While Keene (1995) justifiably warns against letting data “decide” the transformation, we consider some particulars of the GlaxoSmithKline data. It has been argued that the log transformation fails in lesser ranges of gene expression data, however we observe a full-fledged paradox regarding the estimation of σ_{1i}^2 and σ_{2i}^2 . Refer to the human gene expression data in Tables C.1-C.2. Table C.1 contains (non-missing) expression levels for human gene SCEL, a gene with relatively small variability, $\hat{\sigma}_{2i}^2 = .95$, falling at around the 4th percentile of human genes. On the logarithmic scale, however, $\hat{\sigma}_{2i}^{2 \log} = .6065$, the maximum standard error on the log scale! Alternatively, Table C.2 contains (non-missing) expression values for human gene RPS23, where the level of variability on the untransformed intensity scale is $\hat{\sigma}_{2i}^2 = 627,314$, the second largest among all genes. On the logarithmic scale, however, $\hat{\sigma}_{2i}^{2 \log} = .0033$, falling at around the 11th percentile of all human genes. Tables C.1-C.2 largely tell the story: genes that have lesser expressions are more likely to incur larger ratio fluctuations in ratio than genes with larger expressions.

Also in the GlaxoSmithKline data, we question the extent to which Behrens-Fisher problem exists. Figure C displays the raw gene-expression data from the mouse and human genes with the most significant up and down regulated treatment effect. These genes are among those most likely to display evidence of the Berens-Fisher problem, should it exist. Visually, however, heterogeneity of variance is not obvious for any of these four genes, even when the mean expression is considerably different among the treatment groups. These plots are not meant to suggest that the Behrens-Fisher problem does not occur in Glaxo-SmithKline dataset, only that the assumption of equal treatment-group variances may often be reasonable on the untransformed scale.

Table C.1: Human expression data from the gene SCEL with and without log transformation.

y_{Hi}	$\log y_{Hi}$
0.1	-2.30259
0.4	-0.91629
1.3	0.26236
3.0	1.09861
6.9	1.93152
0.7	-0.35667
2.7	0.99325
0.7	-0.35667
0.2	-1.60944
3.2	1.16315
3.6	1.28093
0.1	-2.30259
5.6	1.72277
3.5	1.25276
4.1	1.41099
0.3	-1.20397
0.3	-1.20397
2.7	0.99325
2.5	0.91629
0.1	-2.30259
0.1	-2.30259
0.2	-1.60944
0.3	-1.20397

Table C.2: Human expression data from the gene RPS23 with and without the log transformation.

y_{Hi}	$\log y_{Hi}$
11,141.4	9.31842
11,690.3	9.36651
12,253.4	9.41356
14969.1	9.61374
15,255.6	9.63270
11,924.3	9.38633
10,243.9	9.23444
17,508.1	9.77042
12,436.9	9.42842
11,298.5	9.33243
13,066.5	9.47781
15,211.0	9.62977
14,669.9	9.59355
12,390.1	9.42465
13,424.0	9.50480
17,203.3	9.75286
13,402.9	9.50323
14,388.8	9.57421
12,545.3	9.43710
12,427.4	9.42766
14,014.8	9.54787
14,737.8	9.59817

Finally, there is the issue of fitting model (2.18) to the log-transformed GlaxoSmithKline data. The estimation methodology presented in Section 2.4, albeit with log-transformed expression measurements, yields the parameter estimates: $\hat{\mu}_M = .0092$, $\hat{\mu}_H = .0208$, $\hat{\sigma}_M = .2271$, $\hat{\sigma}_H = .0955$, $\hat{\rho} = .772$, $\hat{\theta} = .234$, and $\hat{\eta} = 1$. Very unfortunately, $\hat{\eta} = 1$, indicating that there is no mixture of distributions for treatment effect. The interpretation would be that all genes are affected by the treatment to greater and lesser extents, and the regression coefficients approximately follow a bivariate normal distribution (an approach taken by Lewin et al, 2006). This may or may not be true, though there are clear benefits to assuming a mixture, including the avoidance of arbitrary significance cutoffs (as in Lewin et al, 2006). Even unrealistic null hypotheses may be useful, and such treatment-effect mixture-models frequently appear in the Bayesian multiple testing literature (Section 1.2.2).

Neither mouse or human data alone is to blame for $\hat{\eta} = 1$ in model (2.18). Marginal models (2.20 and 2.21) were also fit to the log-transformed GlaxoSmithKline data. For mice, the marginal estimates are $\hat{\mu}_M = .0085$, $\hat{\sigma}_M = .2267$, and $\hat{\eta}_M = 1$; for humans, the marginal estimates are $\hat{\mu}_H = .0224$, $\hat{\sigma}_H = .1000$, and $\hat{\eta}_H = .922$. While the estimate of η drops slightly below unity in the human data, it is close enough to suggest primarily one distribution of treatment effects.

Despite the potential advantages of using the log transformation, the variance inflation of less-expressed genes and the denial of a treatment-effect mixture in the GlaxoSmithKline data leads this research not to log-transform the data. Also, microarray-specific problems such as unequal treatment-group variances are not overwhelmingly evident in the GlaxoSmithKline dataset. Other transformations, such as suggested by Durbin et al. (2002), may be useful and require further investigation. Despite the decision not to log-transform the GlaxoSmithKline data, the authors of this research agree with Speed (2007) on the following: “not logging is not the end of the world...everything is approximate, and plenty of useful, correct answers will be obtained on the (adjusted) intensity scale.”

Figure C.1: An informal search for a Behrens-Fisher problem in the two most significant mouse and human genes (with respect to treatment). The above graphs display the raw gene expression data by treatment group.

Appendix D

Conditional Expectations of Treatment Coefficients

Section 2.3.1 presented the posterior probabilities $\Pr(\beta_{Hi} = 0 | \hat{\beta}_{Mi}, \hat{\beta}_{Hi})$, $i = 1, \dots, N$, which have been the posterior metrics for deciding human differential expression in this research. Other attributes of posterior distribution $f_{\beta_{Hi} | \hat{\beta}_{Mi}, \hat{\beta}_{Hi}}$, such as the mode or expectation, may also be useful in the analysis of human differential expression. While the posterior mode is analytically intractable, this Appendix provides derivations of the posterior expectations

$$E(\beta_{Mi} | \hat{\beta}_{Mi}, \hat{\beta}_{Hi}), \quad (\text{D.1})$$

$$E(\beta_{Hi} | \hat{\beta}_{Mi}, \hat{\beta}_{Hi}). \quad (\text{D.2})$$

The Law of Total Probability yields the expansions

$$E(\beta_{Hi} | \hat{\beta}_{Mi}, \hat{\beta}_{Hi}) = \sum_{O_i, A_{Mi}, A_{Hi}} E(\beta_{Hi} | \hat{\beta}_{Mi}, \hat{\beta}_{Hi}, O_i, A_{Mi}, A_{Hi}) \Pr(O_i, A_{Mi}, A_{Hi} | \hat{\beta}_{Mi}, \hat{\beta}_{Hi}), \quad (\text{D.3})$$

$$E(\beta_{Mi} | \hat{\beta}_{Mi}, \hat{\beta}_{Hi}) = \sum_{O_i, A_{Mi}, A_{Hi}} E(\beta_{Mi} | \hat{\beta}_{Mi}, \hat{\beta}_{Hi}, O_i, A_{Mi}, A_{Hi}) \Pr(O_i, A_{Mi}, A_{Hi} | \hat{\beta}_{Mi}, \hat{\beta}_{Hi}), \quad (\text{D.4})$$

which ease computation by allowing one to address D.1 and D.2 conditionally on the Cases 1-8 from Table 2.1. The following is a case-by-case evaluation of

$$E(\beta_{Mi} | \hat{\beta}_{Mi}, \hat{\beta}_{Hi}, O_i, A_{Mi}, A_{Hi})$$

and

$$E(\beta_{Hi}|\hat{\beta}_{Mi}, \hat{\beta}_{Hi}, O_i, A_{Mi}, A_{Hi})$$

from (D.3)-(D.4).

Cases 1 and 5 : $O_i = 0, A_{Mi} = 0, A_{Hi} = 0$ and $O_i = 1, A_{Mi} = 0, A_{Hi} = 0$. Conditional on the simultaneous absence of treatment effect for both species,

$$E(\beta_{Mi}|\hat{\beta}_{Mi}, \hat{\beta}_{Hi}) = 0, \quad (D.5)$$

$$E(\beta_{Hi}|\hat{\beta}_{Mi}, \hat{\beta}_{Hi}) = 0, \quad (D.6)$$

follow trivially.

Case 2 : $O_i = 0, A_{Mi} = 0, A_{Hi} = 1$. As there is no treatment effect for mice,

$$E(\beta_{Mi}|\hat{\beta}_{Mi}, \hat{\beta}_{Hi}) = 0. \quad (D.7)$$

For humans, where there is a treatment effect independent of the mice, $E(\beta_{Hi}|\hat{\beta}_{Mi}, \hat{\beta}_{Hi}) = E(\beta_{Hi}|\hat{\beta}_{Hi})$. This reduces the problem to well-understood normal Bayes estimation, as $\hat{\beta}_{Hi}|\beta_{Hi} \sim N(\beta_{Hi}, \sigma_{2i}^2)$ and $\beta_{Hi} \sim N(\mu_H, \sigma_H^2)$. Casella and Berger (2002) present the posterior mean in such a situation (example 7.2.16, page 326), and here this result is adapted to give

$$E(\beta_{Hi}|\hat{\beta}_{Mi}, \hat{\beta}_{Hi}) = E(\beta_{Hi}|\hat{\beta}_{Hi}) = \left(\frac{\sigma_{2i}^2}{\sigma_H^2 + \sigma_{2i}^2} \right) \mu_H + \left(\frac{\sigma_H^2}{\sigma_H^2 + \sigma_{2i}^2} \right) \hat{\beta}_{Hi}. \quad (D.8)$$

Case 3 : $O_i = 0, A_{Mi} = 1, A_{Hi} = 0$. The opposite of Case 2,

$$E(\beta_{Mi}|\hat{\beta}_{Mi}, \hat{\beta}_{Hi}) = E(\beta_{Mi}|\hat{\beta}_{Mi}) = \left(\frac{\sigma_{1i}^2}{\sigma_M^2 + \sigma_{1i}^2} \right) \mu_M + \left(\frac{\sigma_M^2}{\sigma_M^2 + \sigma_{1i}^2} \right) \hat{\beta}_{Mi}, \quad (D.9)$$

and

$$E(\beta_{Hi}|\hat{\beta}_{Mi}, \hat{\beta}_{Hi}) = 0. \quad (D.10)$$

Case 4 : $O_i = 0, A_{Mi} = 1, A_{Hi} = 1$. Using the reasoning of Cases 2 and 3, Case 4 describes the situation where both mice and humans exhibit independent treatment effects. Thus,

$$E(\beta_{Mi}|\hat{\beta}_{Mi}, \hat{\beta}_{Hi}) = E(\beta_{Mi}|\hat{\beta}_{Mi}) = \left(\frac{\sigma_{1i}^2}{\sigma_M^2 + \sigma_{1i}^2} \right) \mu_M + \left(\frac{\sigma_M^2}{\sigma_M^2 + \sigma_{1i}^2} \right) \hat{\beta}_{Mi}, \quad (D.11)$$

and

$$E(\beta_{Hi}|\hat{\beta}_{Mi}, \hat{\beta}_{Hi}) = E(\beta_{Hi}|\hat{\beta}_{Hi}) = \left(\frac{\sigma_{2i}^2}{\sigma_H^2 + \sigma_{2i}^2} \right) \mu_H + \left(\frac{\sigma_H^2}{\sigma_H^2 + \sigma_{2i}^2} \right) \hat{\beta}_{Hi}. \quad (\text{D.12})$$

Cases 6 and 7 : $O_i = 1, A_{Mi} = 0, A_{Hi} = 1$ and $O_i = 1, A_{Mi} = 1, A_{Hi} = 0$. Recall from Section 2.2 that these cases have zero probability of occurrence and do not contribute to the conditional expectations.

Case 8 : $O_i = 1, A_{Mi} = 1, A_{Hi} = 1$.

The case of I-Orthologs affected by treatment is by far the most involved. We rely on the following identity,

Theorem D.0.1

$$\int_{-\infty}^{\infty} x e^{-\frac{1}{2}(ax^2+bx)} dx = -\sqrt{\frac{\pi}{2}} \left(\frac{b}{a\sqrt{a}} \right) e^{\frac{b^2}{8a}}, \quad (\text{D.13})$$

for computational and notational convenience. Conditional on $O_i = 1, A_{Mi} = 1, A_{Hi} = 1$, define $\kappa_M = E(\beta_{Mi}|\hat{\beta}_{Mi}, \hat{\beta}_{Hi})$. Then

$$\begin{aligned} \kappa_M &= (2\pi)^{-3/2} [(1-\rho^2)\sigma_H^2 + \sigma_{2i}^2]^{-1/2} (\sigma_M \sigma_{1i})^{-1} \\ &\times \exp\left\{-\frac{1}{2} \left[\frac{(\hat{\beta}_{Hi}-\mu_H)^2}{(1-\rho^2)\sigma_H^2 + \sigma_{2i}^2} + \frac{2\rho\sigma_H(\hat{\beta}_{Hi}-\mu_H)\mu_M}{\sigma_M[(1-\rho^2)\sigma_H^2 + \sigma_{2i}^2]} + \frac{\rho^2\sigma_H^2\mu_M^2}{\sigma_M^2[(1-\rho^2)\sigma_H^2 + \sigma_{2i}^2]} + \frac{\hat{\beta}_{Mi}}{\sigma_{1i}^2} + \frac{\mu_M^2}{\sigma_M^2} \right] \right\} \\ &\times \int_{-\infty}^{\infty} \beta_{Mi} \exp\left\{-\frac{1}{2} \left[\beta_{Mi}^2 \left(\frac{1}{\sigma_{1i}^2} + \frac{1}{\sigma_M^2} + \frac{\sigma_H^2\rho^2}{\sigma_M^2[(1-\rho^2)\sigma_H^2 + \sigma_{2i}^2]} \right) \right] \right\} \\ &\times \beta_{Mi} \left(-\frac{2\rho\sigma_H(\hat{\beta}_{Hi}-\mu_H)}{\sigma_M[(1-\rho^2)\sigma_H^2 + \sigma_{2i}^2]} - \frac{2\rho^2\sigma_H^2\mu_M}{\sigma_M^2[(1-\rho^2)\sigma_H^2 + \sigma_{2i}^2]} - \frac{2\hat{\beta}_{Mi}}{\sigma_{1i}^2} - \frac{2\mu_M}{\sigma_M^2} \right) d\beta_{Mi}. \end{aligned}$$

Applying Theorem D.13, we arrive at

$$\begin{aligned} \kappa_M &= -\frac{1}{4\pi} [(1-\rho^2)\sigma_H^2 + \sigma_{2i}^2]^{-1/2} (\sigma_M \sigma_{1i})^{-1} \\ &\times \exp\left\{-\frac{1}{2} \left[\frac{(\hat{\beta}_{Hi}-\mu_H)^2}{(1-\rho^2)\sigma_H^2 + \sigma_{2i}^2} + \frac{2\rho\sigma_H(\hat{\beta}_{Hi}-\mu_H)\mu_M}{\sigma_M[(1-\rho^2)\sigma_H^2 + \sigma_{2i}^2]} + \frac{\rho^2\sigma_H^2\mu_M^2}{\sigma_M^2[(1-\rho^2)\sigma_H^2 + \sigma_{2i}^2]} + \frac{\hat{\beta}_{Mi}}{\sigma_{1i}^2} + \frac{\mu_M^2}{\sigma_M^2} \right] \right\} \\ &\times a_M^{-3/2} b_M \exp \frac{b_M^2}{8a_M}, \end{aligned}$$

where

$$\begin{aligned} a_M &= \frac{1}{\sigma_{1i}^2} + \frac{1}{\sigma_M^2} + \frac{\sigma_H^2\rho^2}{\sigma_M^2[(1-\rho^2)\sigma_H^2 + \sigma_{2i}^2]}, \\ b_M &= -\frac{2\rho\sigma_H(\hat{\beta}_{Hi}-\mu_H)}{\sigma_M[(1-\rho^2)\sigma_H^2 + \sigma_{2i}^2]} - \frac{2\rho^2\sigma_H^2\mu_M}{\sigma_M^2[(1-\rho^2)\sigma_H^2 + \sigma_{2i}^2]} - \frac{2\hat{\beta}_{Mi}}{\sigma_{1i}^2} - \frac{2\mu_M}{\sigma_M^2}. \end{aligned}$$

Similarly, define $\kappa_H = E(\beta_{Hi}|\hat{\beta}_{Mi}, \hat{\beta}_{Hi})$. Then

$$\begin{aligned} \kappa_H &= (2\pi)^{-3/2} [(1-\rho^2)\sigma_M^2 + \sigma_{1i}^2]^{-1/2} (\sigma_H \sigma_{2i})^{-1} \\ &\times \exp\left\{-\frac{1}{2} \left[\frac{(\hat{\beta}_{Mi} - \mu_M)^2}{(1-\rho^2)\sigma_M^2 + \sigma_{1i}^2} + \frac{2\rho\sigma_M(\hat{\beta}_{Mi} - \mu_M)\mu_H}{\sigma_H[(1-\rho^2)\sigma_M^2 + \sigma_{1i}^2]} + \frac{\rho^2\sigma_M^2\mu_H^2}{\sigma_H^2[(1-\rho^2)\sigma_M^2 + \sigma_{1i}^2]} + \frac{\hat{\beta}_{Hi}}{\sigma_{2i}^2} + \frac{\mu_H^2}{\sigma_H^2} \right]\right\} \\ &\times \int_{-\infty}^{\infty} \beta_{Hi} \exp\left\{-\frac{1}{2} \left[\beta_{Hi}^2 \left(\frac{1}{\sigma_{2i}^2} + \frac{1}{\sigma_H^2} + \frac{\sigma_M^2\rho^2}{\sigma_H^2[(1-\rho^2)\sigma_M^2 + \sigma_{1i}^2]} \right) \right]\right\} \\ &\times \beta_{Hi} \left(-\frac{2\rho\sigma_M(\hat{\beta}_{Mi} - \mu_M)}{\sigma_H[(1-\rho^2)\sigma_M^2 + \sigma_{1i}^2]} - \frac{2\rho^2\sigma_M^2\mu_H}{\sigma_H^2[(1-\rho^2)\sigma_M^2 + \sigma_{1i}^2]} - \frac{2\hat{\beta}_{Hi}}{\sigma_{2i}^2} - \frac{2\mu_H}{\sigma_H^2} \right) d\beta_{Hi}. \end{aligned}$$

Applying Theorem D.13,

$$\begin{aligned} \kappa_H &= -\frac{1}{4\pi} [(1-\rho^2)\sigma_M^2 + \sigma_{1i}^2]^{-1/2} (\sigma_H \sigma_{2i})^{-1} \\ &\times \exp\left\{-\frac{1}{2} \left[\frac{(\hat{\beta}_{Mi} - \mu_M)^2}{(1-\rho^2)\sigma_M^2 + \sigma_{1i}^2} + \frac{2\rho\sigma_M(\hat{\beta}_{Mi} - \mu_M)\mu_H}{\sigma_H[(1-\rho^2)\sigma_M^2 + \sigma_{1i}^2]} + \frac{\rho^2\sigma_M^2\mu_H^2}{\sigma_H^2[(1-\rho^2)\sigma_M^2 + \sigma_{1i}^2]} + \frac{\hat{\beta}_{Hi}}{\sigma_{2i}^2} + \frac{\mu_H^2}{\sigma_H^2} \right]\right\} \\ &\times a_H^{-3/2} b_H \exp \frac{b_H^2}{8a_H}, \end{aligned}$$

where

$$\begin{aligned} a_H &= \frac{1}{\sigma_{2i}^2} + \frac{1}{\sigma_H^2} + \frac{\sigma_M^2\rho^2}{\sigma_H^2[(1-\rho^2)\sigma_M^2 + \sigma_{1i}^2]}, \\ b_H &= -\frac{2\rho\sigma_M(\hat{\beta}_{Mi} - \mu_M)}{\sigma_H[(1-\rho^2)\sigma_M^2 + \sigma_{1i}^2]} - \frac{2\rho^2\sigma_M^2\mu_H}{\sigma_H^2[(1-\rho^2)\sigma_M^2 + \sigma_{1i}^2]} - \frac{2\hat{\beta}_{Hi}}{\sigma_{2i}^2} - \frac{2\mu_H}{\sigma_H^2}. \end{aligned}$$

The final task is the computation of $\Pr(O_i, A_{Mi}, A_{Hi}|\hat{\beta}_{Mi}, \hat{\beta}_{Hi})$ in D.3-D.4. Application of Bayes rule leads to

$$f_{O_i, A_{Mi}, A_{Hi}|\hat{\beta}_{Mi}, \hat{\beta}_{Hi}} = \frac{f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}|O_i, A_{Mi}, A_{Hi}} f_{O_i, A_{Mi}, A_{Hi}}}{f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}}}. \quad (D.14)$$

The p.d.f $f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}|O_i, A_{Mi}, A_{Hi}}$ and p.m.f. $f_{O_i, A_{Mi}, A_{Hi}}$ are given in Table 2.1. The joint density $f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}}$ of distributional model (2.18) is given in Section 2.3. Thus, we arrive at

$$\begin{aligned} E(\beta_{Mi}|\hat{\beta}_{Mi} = x, \hat{\beta}_{Hi} = y) &= \\ &\left(\frac{f(x, y; \mu_M, 0, \sigma_M^2 + \sigma_{1i}^2, \sigma_{2i}^2, 0) + f(x, y; \mu_M, \mu_H, \sigma_M^2 + \sigma_{1i}^2, \sigma_H^2 + \sigma_{2i}^2, 0)}{f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}}} \right) \\ &\times \left\{ \left(\frac{\sigma_{1i}^2}{\sigma_M^2 + \sigma_{1i}^2} \right) \mu_M + \left(\frac{\sigma_M^2}{\sigma_M^2 + \sigma_{1i}^2} \right) \hat{\beta}_{Mi} \right\} \\ &+ \left(\frac{f(x, y; \mu_M, \mu_H, \sigma_M^2 + \sigma_{1i}^2, \sigma_H^2 + \sigma_{2i}^2, \rho)}{f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}}} \right) \kappa_M \end{aligned}$$

and

$$\begin{aligned}
E(\beta_{Hi} | \hat{\beta}_{Mi} = x, \hat{\beta}_{Hi} = y) = & \\
& \left(\frac{f(x, y; 0, \mu_H, \sigma_{1i}^2, \sigma_H^2 + \sigma_{2i}^2, 0) + f(x, y; \mu_M, \mu_H, \sigma_M^2 + \sigma_{1i}^2, \sigma_H^2 + \sigma_{2i}^2, 0)}{f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}}} \right) \\
& \times \left\{ \left(\frac{\sigma_{2i}^2}{\sigma_H^2 + \sigma_{2i}^2} \right) \mu_H + \left(\frac{\sigma_H^2}{\sigma_H^2 + \sigma_{2i}^2} \right) \hat{\beta}_{Hi} \right\} \\
& + \left(\frac{f(x, y; \mu_M, \mu_H, \sigma_M^2 + \sigma_{1i}^2, \sigma_H^2 + \sigma_{2i}^2, \rho)}{f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}}} \right) \kappa_H,
\end{aligned}$$

where $f(x, y; \mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$ is the bivariate normal density function with argument $(x, y)^T$, mean vector $(\mu_X, \mu_Y)^T$, variances σ_X^2 and σ_Y^2 , and correlation ρ . This completes the derivation.

Appendix E

Applications of the EM Algorithm

A variant of the EM Algorithm was developed for the fitting of model (2.18), with the additional purposes of:

1. providing an alternative to Newton-based updating schemes for maximizing the composite pseudo-likelihood (2.32);
2. bolstering computational integrity by demonstrating agreement between EM and Newton-based updating when applied to real data.

In this Appendix, Section E.1 contains the derivation of the EM variant and Section E.2 addresses its performance relative to the Quasi-Newton algorithm discussed in Section 2.4.

E.1 A Custom EM Algorithm

This section includes the derivation of an EM algorithm specifically for fitting model (2.18) to data. This algorithm, like the Quasi-Newton algorithm from Section 2.4, requires starting values for $\mu_M, \mu_H, \sigma_M, \sigma_H, \rho, \eta, \theta$, denoted $\mu_M^{(0)}, \mu_H^{(0)}, \sigma_M^{(0)}, \sigma_H^{(0)}, \rho^{(0)}, \eta^{(0)}, \theta^{(0)}$.

The expectation step yields the function $g(\cdot)$:

$$\begin{aligned}
g(\theta, \eta, \mu_M, \mu_H, \sigma_M, \sigma_H, \rho; \hat{\beta}_{Mi} = x, \hat{\beta}_{Hi} = y) = \\
& a_i \log f(x, y; \mu_M, \mu_H, \sigma_M^2 + \sigma_{1i}^2, \sigma_H^2 + \sigma_{2i}^2, \rho) + a_i \log(\theta\eta) \\
& + b_i \log f(x, y; \mu_M, 0, \sigma_M^2 + \sigma_{1i}^2, \sigma_{2i}^2, 0) + b_i \log\{(1 - \theta)\eta(1 - \eta)\} \\
& + c_i \log f(x, y; 0, \mu_H, \sigma_{1i}^2, \sigma_H^2 + \sigma_{2i}^2, 0) + c_i \log\{(1 - \theta)\eta(1 - \eta)\} \\
& + d_i \log f(x, y; \mu_M, \mu_H, \sigma_M^2 + \sigma_{1i}^2, \sigma_H^2 + \sigma_{2i}^2, 0) + d_i \log\{(1 - \theta)\eta^2\} \\
& + e_i \log f(x, y; 0, 0, \sigma_{1i}^2, \sigma_{2i}^2, 0) + e_i \log\{(1 - \theta)(1 - \eta)^2 + \theta(1 - \eta)\},
\end{aligned} \tag{E.1}$$

where $f(x, y; \mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$ is the bivariate normal density function with argument $(x, y)^T$, mean vector $(\mu_X, \mu_Y)^T$, variances σ_X^2 and σ_Y^2 , and correlation ρ and

$$\begin{aligned}
a_i &= \frac{f(x, y; \mu_M^{(0)}, \mu_H^{(0)}, \sigma_M^{2(0)} + \sigma_{1i}^2, \sigma_H^{2(0)} + \sigma_{2i}^2, \rho^{(0)})}{f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}}(x, y; \theta^{(0)}, \eta^{(0)}, \rho^{(0)}, \mu_M^{(0)}, \mu_H^{(0)}, \sigma_M^{2(0)}, \sigma_H^{2(0)}, \sigma_{1i}^2, \sigma_{2i}^2)}, \\
b_i &= \frac{f(x, y; \mu_M^{(0)}, 0, \sigma_M^{2(0)} + \sigma_{1i}^2, \sigma_{2i}^2, 0)}{f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}}(x, y; \theta^{(0)}, \eta^{(0)}, \rho^{(0)}, \mu_M^{(0)}, \mu_H^{(0)}, \sigma_M^{2(0)}, \sigma_H^{2(0)}, \sigma_{1i}^2, \sigma_{2i}^2)}, \\
c_i &= \frac{f(x, y; 0, \mu_H^{(0)}, \sigma_{1i}^2, \sigma_H^{2(0)} + \sigma_{2i}^2, 0)}{f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}}(x, y; \theta^{(0)}, \eta^{(0)}, \rho^{(0)}, \mu_M^{(0)}, \mu_H^{(0)}, \sigma_M^{2(0)}, \sigma_H^{2(0)}, \sigma_{1i}^2, \sigma_{2i}^2)}, \\
d_i &= \frac{f(x, y; \mu_M^{(0)}, \mu_H^{(0)}, \sigma_M^{2(0)} + \sigma_{1i}^2, \sigma_H^{2(0)} + \sigma_{2i}^2, 0)}{f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}}(x, y; \theta^{(0)}, \eta^{(0)}, \rho^{(0)}, \mu_M^{(0)}, \mu_H^{(0)}, \sigma_M^{2(0)}, \sigma_H^{2(0)}, \sigma_{1i}^2, \sigma_{2i}^2)}, \\
e_i &= \frac{f(x, y; 0, 0, \sigma_{1i}^2, \sigma_{2i}^2, 0)}{f_{\hat{\beta}_{Mi}, \hat{\beta}_{Hi}}(x, y; \theta^{(0)}, \eta^{(0)}, \rho^{(0)}, \mu_M^{(0)}, \mu_H^{(0)}, \sigma_M^{2(0)}, \sigma_H^{2(0)}, \sigma_{1i}^2, \sigma_{2i}^2)}.
\end{aligned}$$

Despite lengthy algebraic manipulation, maximization of

$$\sum_{i=1}^N g(\theta, \eta, \mu_M, \mu_H, \sigma_M, \sigma_H, \rho; \hat{\beta}_{Mi} = x, \hat{\beta}_{Hi} = y) \tag{E.2}$$

from (E.1) with respect to θ and η may be obtained analytically. Setting the partial derivatives of (E.2) to zero and solving, we obtain the solution set

$$\theta = \frac{-\theta_b \pm \sqrt{\theta_b^2 - 4\theta_a\theta_c}}{2\theta_a}, \tag{E.3}$$

$$\eta = \frac{v/\theta + x/(\theta - 1)}{v/\theta - n}, \tag{E.4}$$

where

$$\begin{aligned}
\theta_a &= -yu^2 + uxy + uyv - zu^2 + wvu + wux, \\
\theta_b &= yu^2 - 2yxu + yx^2 - yuv + xyv + uzv + zu^2 - zux \\
&\quad + wv^2 + 2wxv - 3wuv + wxv + 2wx^2 - 2uxw, \\
\theta_c &= -zuv + xzv - 2xwv - wv^2 + 2wuv,
\end{aligned}$$

and

$$\begin{aligned}
u &= \sum_{i=1}^N (a_i + b_i + c_i + d_i + e_i), \\
v &= \sum_{i=1}^N a_i, \\
w &= \sum_{i=1}^N (b_i + c_i + e_i), \\
x &= \sum_{i=1}^N (b_i + c_i + d_i), \\
y &= \sum_{i=1}^N (a_i + 2d_i), \\
z &= \sum_{i=1}^N (b_i + c_i).
\end{aligned}$$

In practice, the upper solution to (E.3) yielded values of θ outside of the parametric boundaries, leaving only one viable solution. If two viable solutions exist for (η, θ) , a simple numerical remedy is to evaluate (E.2) at both solutions, and defining the solution that corresponds to the larger value of (E.2) as $(\eta^{(k+1)}, \theta^{(k+1)})$.

With the updates $\theta^{(k+1)}$ and $\eta^{(k+1)}$ in hand, maximizing (E.2) simplifies the max-

imization of

$$\begin{aligned}
 \sum_{i=1}^N h(\mu_M, \mu_H, \sigma_M, \sigma_H, \rho; x = \hat{\beta}_{Mi}, y = \hat{\beta}_{Hi}) = \\
 \sum_{i=1}^N a_i \log f(x, y; \mu_M, \mu_H, \sigma_M^2 + \sigma_{1i}^2, \sigma_H^2 + \sigma_{2i}^2, \rho) \\
 + \sum_{i=1}^N b_i \log f(x, y; \mu_M, 0, \sigma_M^2 + \sigma_{1i}^2, \sigma_{2i}^2, 0) \\
 + \sum_{i=1}^N c_i \log f(x, y; 0, \mu_H, \sigma_{1i}^2, \sigma_H^2 + \sigma_{2i}^2, 0) \\
 + \sum_{i=1}^N d_i \log f(x, y; \mu_M, \mu_H, \sigma_M^2 + \sigma_{1i}^2, \sigma_H^2 + \sigma_{2i}^2, 0)
 \end{aligned} \tag{E.5}$$

with respect to $\mu_M, \mu_H, \sigma_M, \sigma_H$, and ρ .

In a departure from more common applications of the EM algorithm, analytic maximization of (E.5) is not possible due to the fact that the bivariate-normal variances depend on i . Numerical maximization using Quasi-Newton updating, with starting values $\mu_M^{(k)}, \mu_H^{(k)}, \sigma_M^{(k)}, \sigma_H^{(k)}, \rho^{(k)}$, is used to complete the updating. This maximization provides $\mu_M^{(k+1)}, \mu_H^{(k+1)}, \sigma_M^{(k+1)}, \sigma_H^{(k+1)}, \rho^{(k+1)}$ in addition to the analytically obtained $\eta^{(k+1)}$ and $\theta^{(k+1)}$.

E.2 Performance

Both the Quasi-Newton algorithm, with the default stopping criteria from Proc NLP in SAS, and the EM algorithm, which stops on the iteration where both the change in the loglikelihood *and* the maximum parameter estimate change are less than 1×10^{-5} , were implemented to fit model (2.18) to the GlaxoSmithKline data (Section 1.4). The algorithms were initialized at the 16 sets of starting values provided by the methodology of Section 2.4.1, and the results corresponding to each starting value set are provided in Tables E.1-E.2.

Arguably the most important difference between the two algorithms is the amount of computer time taken, with the EM algorithm typically taking over 100 (and sometimes close to 200) times as long as the Quasi-Newton algorithm. This comparison is not entirely

straightforward, however, as the stopping rules for both algorithms are adjustable. Despite the longer computational time, the EM Algorithm always reaches the same maximum (up to rounding error), whereas the Quasi-Newton algorithm stops at different numerical maximums. In no case does a numerical maximum from the Quasi-Newton algorithm exceed the common value reached by the EM algorithm, but in six cases this maximum is reached, with two other cases only slightly falling short. Since the estimation methodology outlined in Section 2.4 specifies using results from the starting value set with the largest numerical maximum, the two algorithms technically agree on the fit of model (2.18) to the Glaxo-SmithKline data. When there were problems with the Quasi-Newton maximization, they tended to involve situations where $\hat{\rho} \approx 1$, $\hat{\rho} \approx -1$ and/or $\hat{\theta} \approx 0$. Most other parameter estimates did not differ considerably between the two algorithms. Though nowhere near the outlier phenomenon observed in the Finite Sample Size Experiments (Section 3.4.2), $\hat{\sigma}_H$ was smaller in every situation where the common maximum from the EM algorithm failed to be reached.

Table E.1: The relative performance of the EM algorithm to the Quasi-Newton algorithm for the first eight (of sixteen) starting value sets.

Starting Values							Estimates							Objective	Time	
$\mu_M^{(0)}$	$\sigma_M^{(0)}$	$\mu_H^{(0)}$	$\sigma_H^{(0)}$	$\rho^{(0)}$	$\eta^{(0)}$	$\theta^{(0)}$	Method	$\hat{\mu}_M$	$\hat{\sigma}_M$	$\hat{\mu}_H$	$\hat{\sigma}_H$	$\hat{\rho}$	$\hat{\eta}$	$\hat{\theta}$	Function	(sec)
56.17	367.15	102.94	284.86	.95	.10	.10	EM	-2.71	55.28	2.83	6.77	.934	.446	.165	-87704.63	1281.55
							QUANEW	-2.65	55.23	2.80	6.75	1.00	.446	.139	-87706.26	10.16
56.17	367.15	102.94	284.86	.94	.10	.30	EM	-2.71	55.28	2.83	6.77	.934	.446	.165	-87704.63	1343.95
							QUANEW	-2.65	55.23	2.80	6.75	1.00	.446	.139	-87706.26	10.73
56.17	367.15	102.94	284.86	.54	.10	.50	EM	-2.71	55.28	2.83	6.77	.934	.446	.165	-87704.63	1426.58
							QUANEW	-2.71	55.28	2.83	6.77	.934	.446	.165	-87704.63	11.59
56.17	367.15	102.94	284.86	.37	.10	.70	EM	-2.71	55.28	2.83	6.77	.934	.446	.165	-87704.63	1444.77
							QUANEW	-2.78	55.73	2.94	6.19	.877	.443	0.00	-87723.16	9.58
18.72	213.62	34.31	171.47	.95	.30	.10	EM	-2.71	55.28	2.83	6.77	.934	.446	.165	-87704.63	1078.02
							QUANEW	-2.71	55.75	2.96	6.23	-1.00	.444	0.03	-87722.03	10.33
18.72	213.62	34.31	171.47	.93	.30	.30	EM	-2.71	55.28	2.83	6.77	.934	.446	.165	-87704.63	1404.87
							QUANEW	-2.71	55.75	2.96	6.23	-1.00	.444	0.03	-87722.03	10.58
18.72	213.62	34.31	171.47	.55	.30	.50	EM	-2.71	55.28	2.83	6.77	.934	.446	.165	-87704.63	1414.15
							QUANEW	-2.65	55.23	2.80	6.75	1.00	.446	0.139	-87706.26	8.64
18.72	213.62	34.31	171.47	.39	.30	.70	EM	-2.71	55.28	2.83	6.77	.934	.446	.165	-87704.63	1483.32
							QUANEW	-2.83	55.72	2.93	6.16	1.00	.444	0.00	-87723.29	8.48

Table E.2: The relative performance of the EM algorithm to the Quasi-Newton algorithm for the second eight (of sixteen) starting value sets.

$\mu_M^{(0)}$	$\sigma_M^{(0)}$	Starting Values					Method	Estimates							Objective Function	Time (sec)
		$\mu_H^{(0)}$	$\sigma_H^{(0)}$	$\rho^{(0)}$	$\eta^{(0)}$	$\theta^{(0)}$		$\hat{\mu}_M$	$\hat{\sigma}_M$	$\hat{\mu}_H$	$\hat{\sigma}_H$	$\hat{\rho}$	$\hat{\eta}$	$\hat{\theta}$		
11.23	165.73	20.59	133.88	.95	.50	.10	EM	-2.71	55.28	2.83	6.76	.934	.446	.165	-87704.63	804.75
							QUANEW	-2.73	55.28	2.83	6.76	.927	.446	.166	-87704.64	10.05
11.23	165.73	20.59	133.88	.93	.50	.30	EM	-2.71	55.28	2.83	6.77	.934	.446	.165	-87704.63	1288.35
							QUANEW	-2.70	55.31	2.82	6.77	.944	.446	.163	-87704.65	8.03
11.23	165.73	20.59	133.88	.55	.50	.50	EM	-2.71	55.28	2.83	6.77	.934	.446	.165	-87704.63	1455.17
							QUANEW	-2.67	55.25	2.80	6.76	.996	.446	.140	-87706.04	7.72
11.23	165.73	20.59	133.88	.39	.50	.70	EM	-2.71	55.28	2.83	6.77	.934	.446	.165	-87704.63	1525.74
							QUANEW	-2.71	55.28	2.83	6.76	.934	.446	.165	-87704.63	8.97
8.02	140.16	14.71	113.53	.95	.70	.10	EM	-2.71	55.28	2.83	6.77	.934	.446	.165	-87704.63	764.85
							QUANEW	-2.71	55.28	2.83	6.77	.934	.446	.165	-87704.63	9.27
8.02	140.16	14.71	113.53	.93	.70	.30	EM	-2.71	55.28	2.83	6.77	.934	.446	.165	-87704.63	1326.74
							QUANEW	-2.70	55.28	2.83	6.76	.934	.446	.164	-87704.63	8.36
8.02	140.16	14.71	113.53	.56	.70	.50	EM	-2.71	55.28	2.83	6.77	.934	.446	.165	-87704.63	1366.79
							QUANEW	-2.71	55.28	2.83	6.77	.934	.446	.165	-87704.63	8.16
8.02	140.16	14.71	113.53	.40	.70	.70	EM	-2.71	55.28	2.83	6.77	.934	.446	.165	-87704.63	1526.00
							QUANEW	-2.71	55.28	2.83	6.77	.934	.446	.165	-87704.63	8.24

Appendix F

Gene Pooling Based Variance Estimation

Regarding gene-level variance estimation, Smyth (2004) remarks that “given the large number of gene-wise linear model fits arising from a microarray experiment, there is a pressing need to take advantage of the parallel structure whereby the same model is fitted to each gene.” This section evaluates Smyth’s assertion as it relates to the fit of models (1.12) and (1.13) to the GlaxoSmithKline data (Sections 2.6). The possibility of improved estimation of σ_{1i}^2 and σ_{2i}^2 via Bayesian or other shrinkage methodology is investigated and discussed. Section F.1 reports on the method of Smyth (2004), based on Bayesian estimation with inverse gamma priors. Section F.2 derives and evaluates an approximate Bayesian estimation method using lognormal priors. Section F.3 reports on a non-parametric shrinkage technique, the Shrinkage t of Opgen-Rhein and Strimmer (2007). Finally, Section F.4 is a discussion of the alternative estimators for σ_{1i}^2 and σ_{2i}^2 relative to least squares.

For convenience, the following discussions will use only the mouse gene expression model

$$y_{Mij} = \alpha_{Mi} + \beta_{Mi}T_j + \sigma_{Mi}^2\epsilon_{ij} \quad j = 1, \dots, 19, \quad (\text{F.1})$$

where T_j are $\{0,1\}$ treatment indicators and ϵ_{ij} are standard normal random variables, for $i = 1, \dots, 9683$ mouse genes in the GlaxoSmithKline data. Furthermore, since $\sigma_{1i}^2 = d\sigma_{Mi}$, where d is a design constant, we will discuss estimation of σ_{Mi} and σ_{1i}^2 interchangeably. Denoting $\hat{\sigma}_{Mi}^2$ as the least squares estimate of σ_{Mi}^2 ,

$$\hat{\sigma}_{Mi}^2 | \sigma_{Mi}^2 \sim \frac{\sigma_{Mi}}{v_i} \chi_{v_i}^2 \quad (\text{F.2})$$

follows from least squares assumptions, where v_i is the error degrees of freedom for gene i (typically 17 in the GlaxoSmithKline data, though less when there is missing data).

F.1 Inverse Gamma Priors

Smyth (2004) suggests modeling prior information for σ_{Mi}^2 in a way that seems to algebraically rearrange (2):

$$\frac{1}{\sigma_{Mi}^2} \sim \frac{1}{v_0 s_0^2} \chi_{v_0}^2, \quad (\text{F.3})$$

where s_0^2 is a “prior estimator” with v_0 degrees of freedom. In actuality, (F.3) is a reparameterization of the usual inverse gamma prior and equivalent to

$$\sigma_{Mi}^2 \sim \text{InvGam}\left(\frac{v_0}{2}, \frac{v_0 s_0^2}{2}\right). \quad (\text{F.4})$$

Smyth constructs estimates of s_0^2 and v_0 using the moments of $\log(\hat{\sigma}_{Mi}^2)$. The logarithmic transformation is used to ensure finite moments for all degrees of freedom and to increase efficiency due to approximate normality.

Using Smyth’s (2004) methodology to estimate v_0 and s_0^2 (for which we refer the readers to Smyth’s 2004 paper), and the GlaxoSmithKline data, we arrive at $\hat{v}_0 = 0.86$ and $\hat{s}_0^2 = 47.26$ for the mouse data and $\hat{v}_0 = 0.84$ and $\hat{s}_0 = 92.91$ for the human data. Given Smyth’s interpretation of v_0 as the degrees-of-freedom gained from borrowed among-gene information, not much is achieved from this methodology: in both mouse and human datasets, $\hat{v}_0 < 1$, meaning that the information gained from Smyth’s Bayesian procedure is less than the addition of a single mouse or human subject to the experiment.

The impact of v_0 can be clearly seen in the updating formula reported by Smyth (2004),

$$\tilde{\sigma}_{Mi}^2 = \text{E}(\sigma_{Mi}^2 | \hat{\sigma}_{Mi}^2) = \frac{v_0 s_0^2 + v_i \hat{\sigma}_{Mi}^2}{v_0 + v_i}, \quad (\text{F.5})$$

where $\tilde{\sigma}_{Mi}^2 \rightarrow \hat{\sigma}_{Mi}^2$ as $v_0 \rightarrow 0$. We believe that (F.5) is incorrect however, and that Smyth (2004) meant to derive

$$\tilde{\sigma}_{Mi}^2 = \text{E}(\sigma_{Mi}^2 | \hat{\sigma}_{Mi}^2) = \frac{v_0 s_0^2 + v_i \hat{\sigma}_{Mi}^2}{v_0 + v_i - 2}, \quad (\text{F.6})$$

though this is disputed by Smyth himself (via e-mail correspondence). Regardless, the point remains that very little shrinkage is achieved when $\hat{v}_0$ is small, as did occur for both species in the GlaxoSmithKline data.

F.2 Lognormal Priors

In a full Bayesian approach, Lewin et al. (2006) report success in modeling gene-level variances with lognormal models. Despite a general lack of analytical tractability, lognormal models have particular appeal in this situation because of their relative success in modeling $\hat{\sigma}_{Mi}^2$ and $\hat{\sigma}_{Hi}^2$ in the GlaxoSmithKline data (see Figure 3.1). The multiplicative error structure of $\hat{\sigma}_{Mi}^2$ and $\hat{\sigma}_{Hi}^2$ is suggestive when taking logarithms of both sides of (F.3),

$$\log \hat{\sigma}_{Mi}^2 = \log \sigma_{Mi}^2 - \log v_i + \log X_i, \quad (\text{F.7})$$

where $X_i \sim \chi_{v_i}^2$. Since $\hat{\sigma}_{Mi}^2$ is successfully modeled by a lognormal distribution, it follows that *if* X_i could be approximated by a lognormal, then σ_{Mi}^2 would itself be *approximately* lognormal. In this case, parameter estimation is straightforward. As an approximation, assume $X_i \sim \text{lognormal}(\mu_{v_i}, \sigma_{v_i}^2)$, where μ_{v_i} and $\sigma_{v_i}^2$ are chosen to optimize the lognormal approximation to a χ_{v_i} distribution. Then, approximately,

$$\begin{aligned} \log \hat{\sigma}_{Mi}^2 &\sim N(\mu_h, \sigma_h^2), \\ \log \sigma_{Mi}^2 &\sim N(\kappa, \tau^2), \\ \log X_i &\sim N(\mu_{v_i}, \sigma_{v_i}^2). \end{aligned}$$

Matching moments of the right and left sides of (F.7),

$$\begin{aligned} \kappa &= \mu_h - \log v_i + \mu_{v_i}, \\ \tau^2 &= \sigma_h^2 - \sigma_{v_i}^2, \end{aligned}$$

for $i = 1, \dots, 9683$. Estimates of μ_h and σ_h are easily obtained by taking means and variances of the observed $\log \hat{\sigma}_{Mi}^2$:

$$\begin{aligned} \hat{\mu}_h &= \frac{1}{9683} \sum_{i=1}^{9683} \log \hat{\sigma}_{Mi}^2, \\ \hat{\sigma}_h^2 &= \frac{1}{9683-1} \sum_{i=1}^{9683} (\log \hat{\sigma}_{Mi}^2 - \hat{\mu}_h)^2. \end{aligned}$$

To estimate τ and κ however, we must find μ_{v_i} and $\sigma_{v_i}^2$ that make the lognormal approximation to a chi-square useful. The first-order Taylor approximation, $\mu_v \approx \log(v_i)$ and $\sigma_{v_i}^2 \approx \frac{2}{v_i}$ is particularly convenient as $\log v_i$ centers $\log X_i$ in (F.7),

$$\log \hat{\sigma}_{Mi}^2 = \log \sigma_{Mi}^2 + \log \epsilon_i, \quad (\text{F.8})$$

where $\log \epsilon_i \sim N(0, 2/v)$ and thus $\kappa \approx \mu_h$. Defining $\bar{\sigma}_v^2 = 2/9683 \sum_{i=1}^{9683} \frac{1}{v_i}$, we arrive at the estimates

$$\begin{aligned}\hat{\kappa} &= \hat{\mu}_h \\ \hat{\tau}^2 &= \hat{\sigma}_h^2 - \bar{\sigma}_v^2.\end{aligned}$$

The Normal Bayes estimator, presented in Example 7.2.16 (page 326) of Casella and Berger (2002), coupled with the fact that

$$\log \hat{\sigma}_{Mi}^2 | \log \sigma_{Mi}^2 \sim N(\log \sigma_{Mi}^2, \sigma_{v_i}^2), \quad (\text{F.9})$$

leads to

$$\log \sigma_{Mi}^2 | \log \hat{\sigma}_{Mi}^2 \sim N\left(\frac{\tau^2}{\tau^2 + \sigma_{v_i}^2} \log \hat{\sigma}_{Mi}^2 + \frac{\sigma_{v_i}^2}{\tau^2 + \sigma_{v_i}^2} \kappa, \frac{\sigma_{v_i}^2 \tau^2}{\tau^2 + \sigma_{v_i}^2}\right), \quad (\text{F.10})$$

and the estimate

$$\tilde{\sigma}_{Mi}^2 = \exp\left(\frac{\hat{\tau}^2}{\hat{\tau}^2 + \sigma_{v_i}^2} \log \hat{\sigma}_{Mi}^2 + \frac{\sigma_{v_i}^2}{\hat{\tau}^2 + \sigma_{v_i}^2} \hat{\kappa}\right).$$

In the mouse dataset, $\hat{\tau}^2 = 6.381$ and $\hat{\kappa} = 5.303$; in the human dataset, $\hat{\tau}^2 = 6.737$ and $\hat{\kappa} = 6.029$. For both species, $v_i = 17$ for genes with complete data (and thus σ_{v_i} is generally $2/17$). A few points are worth noting. First, since $\hat{\kappa} = \hat{\mu}_h$, the shrinkage is in the direction of the mean on the log scale. The extent of the shrinkage, however, depends on the ratio of $\sigma_{v_i}^2$ to $\hat{\tau}^2$, with larger ratios corresponding to more shrinkage. For both mice and humans, this ratio is very small: 1:54 and 1:51, respectively (when $v_i = 17$). This corresponds to less than 2% shrinkage towards the mean (on the log scale) for both humans and mice. Thus, this approximate Bayesian methodology, using lognormal priors, agrees with the methodology of Symth (2004) on the point that there should only be minimal shrinkage of the least-squares variance estimates.

F.3 Nonparametric Shrinkage

In an attempt to reap the benefits of Bayesian hierarchical methodology while maintaining computational simplicity, Opgein-Rhein and Strimmer (2007) construct the Shrinkage t statistic, based on a simple non-parametric variance shrinking formula. In the context of mouse model F.1, the shrinkage variance estimate is

$$\tilde{\sigma}_{Mi}^2 = \hat{\lambda} \sigma_M^2 \text{median} + (1 - \hat{\lambda}) \hat{\sigma}_{Mi}^2, \quad (\text{F.11})$$

where $\sigma_M^{2 \text{ median}}$ is the median of $\{\hat{\sigma}_{M1}^2, \dots, \hat{\sigma}_{M9683}^2\}$ and

$$\hat{\lambda} = \min \left(1, \frac{\sum_{i=1}^{9683} \widehat{\text{Var}}(\hat{\sigma}_{Mi}^2)}{\sum_{i=1}^{9683} (\hat{\sigma}_{Mi}^2 - \sigma_M^{2 \text{ median}})^2} \right). \quad (\text{F.12})$$

There are two formulas for $\widehat{\text{Var}}(\hat{\sigma}_{Mi}^2)$ that we will examine. The first, given by Opgen-Rhein and Strimmer (2007), is completely non-parametric. Define

$$\begin{aligned} \bar{y}_{Mi.}^{\text{TRT}} &= \frac{1}{17} \sum_{i=1}^{17} T_j y_{Mij}, & \bar{y}_{Mi.}^{\text{PLA}} &= \frac{1}{17} \sum_{i=1}^{17} (1 - T_j) y_{Mij}, \\ w_{ij} &= \begin{cases} (y_{Mij} - \bar{y}_{Mi.}^{\text{TRT}})^2 & \text{if } T_j = 1, \\ (y_{Mij} - \bar{y}_{Mi.}^{\text{PLA}})^2 & \text{if } T_j = 0, \end{cases} \\ \bar{w}_i &= \frac{1}{17} \sum_{i=1}^{17} w_{ij}. \end{aligned}$$

Then

$$\widehat{\text{Var}}(\hat{\sigma}_{Mi}^2) = \frac{9863}{(9863 - 1)^3} \sum_{i=1}^{9683} (w_{ij} - \bar{w}_i)^2. \quad (\text{F.13})$$

Our least-squares assumptions lend somewhat simpler formulas for $\widehat{\text{Var}}(\hat{\sigma}_{Mi}^2)$:

$$\widehat{\text{Var}}(\hat{\sigma}_{Mi}^2)^{\text{MODEL}} = \frac{2\hat{\sigma}_{Mi}^4}{v_{Mi}}. \quad (\text{F.14})$$

Additionally, in the human data, the ability of covariates to explain variation is accounted for.

The nonparametric variance estimates lead to $\hat{\lambda} = .1347$ in the mouse dataset and $\hat{\lambda} = .1933$ for the human dataset, while the model-based variance estimates lead to $\hat{\lambda} = .1184$ in the mouse dataset and $\hat{\lambda} = .1328$ for the human dataset. Thus, in both cases slight to moderate shrinkage is suggested and achieved using shrinkage t methodology. An interesting interpretation of $\hat{\lambda}$, given by Opgen-Rhein and Strimmer (2007), is that small values correspond to situations where the empirical variances σ_{Mi}^2 (or σ_{Hi}^2) can be “reliably determined from the data” without shrinkage, whereas larger values indicate that gains can be achieved from the pooling of information (i.e., shrinkage).

F.4 Discussion

The parameters σ_{1i}^2 and σ_{2i}^2 are the overwhelmingly most abundant parameters in model (2.18), and quality estimation of these parameters is of paramount importance.

The least squares estimates, as competitors to their Bayesian and shrinkage counterparts, should be taken seriously, especially in light of the computational convenience. Two of three procedures discussed in this section yield almost no shrinkage to the mouse and human variance estimates, while the third Shrinkage t procedure suggests slight to moderate shrinkage. While this is by no means an exhaustive investigation into pooling-based estimation of gene-level variances, the choice was made not to address the impact of these estimates on the performance of model (2.18), as doing so would increase complexity of an already complicated simulation. Thus, the verdict remains out on the best estimation technique for σ_{1i}^2 and σ_{2i}^2 for use with model (2.18).

Appendix G

Simulation Figures and Tables

Figure G.1: From the Infinite Sample Size Experiment (ISSE), boxplots were constructed to display the distribution of the maximum likelihood estimates for μ_M , μ_H , σ_M^{Tr} , σ_H^{Tr} at various levels of N . The plots include estimates from 497 replicate data sets with $(\beta_{Mi}, \beta_{Hi}), i = 1, \dots, N$, generated from model (2.16).

Figure G.2: From the Infinite Sample Size Experiment (ISSE), boxplots were constructed to display the distribution of the maximum likelihood estimates for ρ^{Tr} , η , θ (with ρ reparameterized as in Section 2.4.1) at various levels of N . The plots include estimates from 497 replicate data sets with $(\beta_{Mi}, \beta_{Hi}), i = 1, \dots, N$, generated from model (2.16).

Figure G.3: From the Infinite Sample Size Experiment (ISSE), these overlay plots display mean information matrix standard errors (blue) alongside Monte Carlo standard errors (red) for μ_M , μ_H , σ_M^{Tr} , σ_H^{Tr} (using the reparameterizations from Section 2.4.1) with number of genes = N and 497 replicate data sets.

Figure G.4: From the Infinite Sample Size Experiment (ISSE), these overlay plots display mean information matrix standard errors (blue) alongside Monte Carlo standard errors (red) for $\hat{\rho}^{\text{Tr}}$, $\hat{\eta}^{\text{Tr}}$, $\hat{\theta}^{\text{Tr}}$ (using the reparameterizations from Section 2.4.1) with number of genes = N and 497 replicate data sets.

Figure G.5: From the Infinite Sample Size Experiment (ISSE), empirical false discovery rates of I-Ortholog discovery are plotted as a function of the number of gene pairs, N , for nominal FDRs .05 and .20 and based on 497 replicate data sets.

Figure G.6: From the Infinite Sample Size Experiment (ISSE), powers to detect I-Orthologs are plotted as a function of the number of gene pairs, N , for nominal FDRs .05 and .20 and based on 497 replicate data sets.

Table G.1: From the 504 replicates of the Finite Sample Size Experiment I (FSSE I), this table contains ratios of the mean information matrix standard error estimate to the Monte Carlo estimate of standard error, based on 504 replicates. Delete-1 Jackknife standard error estimates for the ratios are included in parentheses.

Parm	Number of Genes	$\theta =$	40 mice		80 mice	
			.20	.40	.20	.40
μ_M	N=4,000	$\eta = .25$	0.95 (.029)	0.96 (.031)	1.01 (.031)	1.01 (.033)
		$\eta = .50$	1.03 (.031)	0.97 (.029)	0.99 (.029)	0.94 (.028)
	N=16,000	$\eta = .25$	1.00 (.030)	1.02 (.034)	1.03 (.035)	0.96 (.029)
		$\eta = .50$	1.03 (.032)	0.99 (.032)	1.00 (.032)	0.99 (.031)
μ_H	N=4,000	$\eta = .25$	.91 (.028)	.93 (.031)	.90 (.029)	.91 (.030)
		$\eta = .50$	.92 (.030)	.93 (.033)	.91 (.032)	.92 (.032)
	N=16,000	$\eta = .25$	.88 (.028)	.88 (.029)	.81 (.028)	.86 (.030)
		$\eta = .50$	.72 (.032)	.84 (.036)	.74 (.029)	.80 (.032)
σ_M^T	N=4,000	$\eta = .25$	1.05 (.031)	1.12 (.037)	1.04 (.033)	1.08 (.034)
		$\eta = .50$	1.10 (.032)	1.07 (.033)	1.03 (.035)	1.07 (.035)
	N=16,000	$\eta = .25$	1.03 (.034)	1.07 (.034)	1.04 (.037)	1.07 (.036)
		$\eta = .50$	0.99 (.031)	1.01 (.031)	1.02 (.030)	0.95 (.033)
σ_H^T	N=4,000	$\eta = .25$	.75 (.026)	.76 (.025)	.74 (.028)	.70 (.027)
		$\eta = .50$	.70 (.030)	.72 (.028)	.67 (.037)	.70 (.030)
	N=16,000	$\eta = .25$	.70 (.024)	.66 (.028)	.59 (.021)	.60 (.026)
		$\eta = .50$	.45 (.030)	.59 (.026)	.43 (.027)	.50 (.021)
ρ^T	N=4,000	$\eta = .25$	27.36 (13.66)	1.66 (.547)	25.28 (9.78)	8.28 (4.64)
		$\eta = .50$	22.06 (9.95)	0.84 (.576)	6.65 (2.42)	0.88 (.05)
	N=16,000	$\eta = .25$	1.43 (.284)	0.89 (.039)	0.93 (.110)	0.94 (.034)
		$\eta = .50$	0.82 (.044)	0.97 (.031)	0.87 (.107)	0.87 (.029)
η^T	N=4,000	$\eta = .25$	1.05 (.036)	1.00 (.029)	0.96 (.031)	0.99 (.030)
		$\eta = .50$	0.97 (.029)	0.99 (.032)	1.01 (.032)	1.01 (.033)
	N=16,000	$\eta = .25$	1.01 (.032)	1.01 (.030)	0.99 (.033)	1.05 (.033)
		$\eta = .50$	0.97 (.031)	1.00 (.031)	0.97 (.033)	0.94 (.029)
θ^T	N=4,000	$\eta = .25$	0.99 (.029)	0.95 (.032)	0.98 (.036)	1.03 (.032)
		$\eta = .50$	1.02 (.033)	0.97 (.030)	1.00 (.031)	0.96 (.033)
	N=16,000	$\eta = .25$	0.98 (.033)	1.02 (.029)	0.97 (.034)	1.04 (.033)
		$\eta = .50$	0.96 (.031)	0.95 (.029)	0.99 (.030)	0.98 (.030)

Table G.2: From the 500 replicates of the Finite Sample Size Experiment II (FSSE II), this table contains ratios of the mean information matrix standard error estimate to the Monte Carlo estimate of standard error. Delete-1 Jackknife standard error estimates for the ratios are included in parentheses.

Parm	Number of Genes	$\theta =$	10 mice		20 mice	
			.20	.40	.20	.40
μ_M	N=4,000	$\eta = .25$	1.02 (.034)	1.01 (.036)	1.08 (.036)	1.01 (.032)
		$\eta = .50$	1.04 (.032)	0.95 (.029)	1.01 (.032)	0.97 (.031)
	N=16,000	$\eta = .25$	1.01 (.034)	1.02 (.036)	1.03 (.037)	0.99 (.033)
		$\eta = .50$	1.01 (.031)	1.01 (.030)	1.02 (.034)	0.96 (.030)
μ_H	N=4,000	$\eta = .25$	.87 (.029)	.84 (.026)	.82 (.029)	.85 (.027)
		$\eta = .50$	.83 (.027)	.82 (.027)	.73 (.025)	.81 (.025)
	N=16,000	$\eta = .25$	.73 (.024)	.77 (.025)	.79 (.025)	.71 (.026)
		$\eta = .50$	.63 (.023)	.63 (.029)	.55 (.018)	.54 (.021)
σ_M^T	N=4,000	$\eta = .25$	1.12 (.042)	1.09 (.033)	1.06 (.037)	1.09 (.036)
		$\eta = .50$	1.03 (.031)	1.06 (.037)	1.04 (.033)	1.11 (.039)
	N=16,000	$\eta = .25$	1.11 (.036)	1.10 (.033)	1.08 (.035)	1.11 (.034)
		$\eta = .50$	1.14 (.038)	1.09 (.035)	1.06 (.035)	1.04 (.036)
σ_H^T	N=4,000	$\eta = .25$	.67 (.022)	.70 (.027)	.69 (.024)	.64 (.028)
		$\eta = .50$	.73 (.025)	.64 (.026)	.56 (.023)	.62 (.022)
	N=16,000	$\eta = .25$	.62 (.042)	.63 (.027)	.62 (.020)	.61 (.021)
		$\eta = .50$	.55 (.021)	.54 (.027)	.43 (.017)	.42 (.019)
ρ^T	N=4,000	$\eta = .25$	43.52 (12.00)	17.81 (6.29)	8.48 (2.11)	19.70 (5.11)
		$\eta = .50$	15.79 (5.47)	54.08 (18.67)	23.64 (9.91)	13.34 (10.71)
	N=16,000	$\eta = .25$	3.30 (0.81)	0.99 (.365)	1.74 (.263)	0.74 (.043)
		$\eta = .50$	3.19 (.744)	0.88 (.100)	4.55 (3.11)	0.72 (.024)
η^T	N=4,000	$\eta = .25$	0.98 (.034)	1.04 (.034)	0.93 (.031)	0.95 (.033)
		$\eta = .50$	1.02 (.033)	0.92 (.030)	0.88 (.029)	0.94 (.031)
	N=16,000	$\eta = .25$	.96 (.034)	.98 (.035)	.97 (.034)	.96 (.034)
		$\eta = .50$	.92 (.029)	.94 (.037)	.87 (.031)	.93 (.032)
θ^T	N=4,000	$\eta = .25$	1.01 (.034)	0.95 (.032)	17.45 (16.45)	0.98 (.034)
		$\eta = .50$	1.01 (.031)	0.97 (.030)	1.02 (.036)	0.90 (.030)
	N=16,000	$\eta = .25$	0.97 (.034)	0.92 (.034)	0.99 (.042)	1.01 (.033)
		$\eta = .50$	0.97 (.029)	0.94 (.041)	0.99 (.033)	0.90 (.030)

Table G.3: An estimation summary of $\hat{\mu}_M$ from the 504 replicates of Finite Sample Size Experiment I (FSSE I).

Number of Genes	Number of Mice	θ	η	Bias	Standard Deviation	RMSE
N=4,000	40	.2	.25	-.34 (.094)	2.09	2.11 (.066)
			.50	-.21 (.058)	1.30	1.31 (.038)
		.4	.25	-.18 (.089)	2.00	2.00 (.063)
			.50	-.10 (.060)	1.35	1.35 (.041)
	80	.2	.25	-.30 (.083)	1.82	1.84 (.057)
			.50	-.11 (.057)	1.28	1.29 (.038)
		.4	.25	-.37 (.080)	1.79	1.83 (.059)
			.50	-.02 (.060)	1.34	1.34 (.039)
N=16,000	40	.2	.25	-.35 (.044)	0.98	1.08 (.064)
			.50	-.12 (.029)	0.65	0.66 (.021)
		.4	.25	-.29 (.042)	0.94	0.99 (.032)
			.50	-.12 (.029)	0.66	0.67 (.021)
	80	.2	.25	-.35 (.040)	0.87	0.95 (.031)
			.50	-.26 (.042)	0.64	0.66 (.021)
		.4	.25	-.26 (.042)	0.94	0.97 (.030)
			.50	-.12 (.028)	0.63	0.64 (.020)

Table G.4: An estimation summary of $\hat{\mu}_H$ from the 504 replicates of Finite Sample Size Experiment I (FSSE I).

Number of Genes	Number of Mice	θ	η	Bias	Standard Deviation	RMSE
N=4,000	40	.2	.25	.21 (.018)	.40	.45 (.013)
			.50	.16 (.011)	.25	.30 (.009)
		.4	.25	.14 (.016)	.35	.38 (.012)
			.50	.09 (.010)	.23	.24 (.009)
	80	.2	.25	.09 (.017)	.38	.39 (.012)
			.50	.04 (.011)	.24	.24 (.008)
		.4	.25	-.04 (.015)	.34	.34 (.011)
			.50	-.03 (.003)	.22	.22 (.008)
N=16,000	40	.2	.25	.24 (.009)	.21	.32 (.008)
			.50	.15 (.007)	.16	.22 (.006)
		.4	.25	.14 (.008)	.19	.23 (.007)
			.50	.09 (.006)	.13	.15 (.005)
	80	.2	.25	.04 (.010)	.21	.22 (.007)
			.50	.01 (.007)	.15	.15 (.006)
		.4	.25	-.02 (.008)	.18	.18 (.007)
			.50	-.06 (.005)	.12	.14 (.006)

Table G.5: An estimation summary of $\hat{\sigma}_M$ from the 504 replicates of Finite Sample Size Experiment I (FSSE I).

Number of Genes	Number of Mice	θ	η	Bias	Standard Deviation	RMSE
N=4,000	40	.2	.25	1.06 (.064)	1.42	1.77 (.053)
			.50	0.18 (.041)	0.92	0.94 (.028)
		.4	.25	0.80 (.058)	1.31	1.53 (.046)
			.50	-0.07 (.042)	0.94	0.94 (.029)
	80	.2	.25	0.16 (.060)	1.31	1.32 (.041)
			.50	-0.15 (.041)	0.92	0.93 (.031)
		.4	.25	0.15 (.056)	1.25	1.25 (.039)
			.50	-0.33 (.039)	0.87	0.93 (.031)
N=16,000	40	.2	.25	1.04 (.033)	0.72	1.27 (.031)
			.50	0.16 (.023)	0.51	0.54 (.016)
		.4	.25	0.83 (.031)	0.69	1.07 (.028)
			.50	-0.02 (.022)	0.50	0.50 (.015)
	80	.2	.25	0.29 (.030)	0.65	0.72 (.026)
			.50	-0.13 (.021)	0.47	0.48 (.014)
		.4	.25	0.16 (.028)	0.63	0.65 (.022)
			.50	-0.27 (.022)	0.49	0.56 (.017)

Table G.6: An estimation summary of $\hat{\sigma}_H$ from the 504 replicates of Finite Sample Size Experiment I (FSSE I).

Number of Genes	Number of Mice	θ	η	Bias	Standard Deviation	RMSE
N=4,000	40	.2	.25	-0.97 (.021)	.47	1.08 (.021)
			.50	-1.06 (.014)	.32	1.11 (.015)
		.4	.25	-0.76 (.020)	.44	0.88 (.019)
			.50	-0.90 (.013)	.30	0.95 (.014)
	80	.2	.25	-1.12 (.021)	.46	1.21 (.021)
			.50	-1.19 (.014)	.32	1.23 (.016)
		.4	.25	-0.95 (.020)	.45	1.05 (.021)
			.50	-1.00 (.013)	.29	1.05 (.014)
N=16,000	40	.2	.25	-0.92 (.012)	.26	0.95 (.012)
			.50	-1.06 (.011)	.24	1.09 (.013)
		.4	.25	-0.76 (.011)	.25	0.80 (.012)
			.50	-0.89 (.008)	.18	0.91 (.009)
	80	.2	.25	-1.12 (.013)	.29	1.16 (.014)
			.50	-1.19 (.011)	.24	1.22 (.013)
		.4	.25	-0.92 (.012)	.27	0.96 (.013)
			.50	-1.02 (.009)	.20	1.04 (.010)

Table G.7: An estimation summary of $\hat{\rho}$ from the 504 replicates of Finite Sample Size Experiment I (FSSE I).

Number of Genes	Number of Mice	θ	η	Bias	Standard Deviation	RMSE
N=4,000	40	.2	.25	-.013 (.0034)	.077	.078 (.0044)
			.50	-.007 (.0026)	.059	.059 (.0027)
		.4	.25	.001 (.0017)	.038	.038 (.0016)
			.50	.001 (.0013)	.029	.029 (.0010)
	80	.2	.25	-.006 (.0034)	.074	.074 (.0039)
			.50	-.005 (.0024)	.053	.053 (.0021)
		.4	.25	-.007 (.0017)	.038	.039 (.0015)
			.50	-.003 (.0013)	.028	.029 (.0013)
N=16,000	40	.2	.25	-.003 (.0017)	.038	.039 (.0015)
			.50	-.001 (.0014)	.032	.032 (.0016)
		.4	.25	-.001 (.0008)	.019	.019 (.0006)
			.50	.000 (.0006)	.014	.014 (.0005)
	80	.2	.25	-.012 (.0017)	.038	.040 (.0017)
			.50	-.004 (.0012)	.027	.028 (.0013)
		.4	.25	-.004 (.0008)	.018	.019 (.0009)
			.50	-.003 (.0007)	.015	.015 (.0006)

Table G.8: An estimation summary of $\hat{\eta}$ from the 504 replicates of Finite Sample Size Experiment I (FSSE I).

Number of Genes	Number of Mice	θ	η	Bias	Standard Deviation	RMSE
N=4,000	40	.2	.25	-.067 (.0003)	.006	.067 (.0003)
			.50	-.099 (.0004)	.008	.099 (.0004)
		.4	.25	-.061 (.0003)	.006	.061 (.0003)
			.50	-.090 (.0004)	.008	.090 (.0004)
	80	.2	.25	-.051 (.0003)	.006	.052 (.0003)
			.50	-.076 (.0003)	.008	.076 (.0003)
		.4	.25	-.045 (.0003)	.006	.045 (.0003)
			.50	-.066 (.0004)	.008	.066 (.0004)
N=16,000	40	.2	.25	-.067 (.0001)	.003	.067 (.0001)
			.50	-.099 (.0002)	.004	.099 (.0002)
		.4	.25	-.061 (.0001)	.003	.061 (.0001)
			.50	-.090 (.0002)	.004	.090 (.0002)
	80	.2	.25	-.051 (.0001)	.003	.051 (.0001)
			.50	-.076 (.0002)	.004	.076 (.0002)
		.4	.25	-.045 (.0001)	.003	.045 (.0001)
			.50	-.066 (.0002)	.004	.066 (.0002)

Table G.9: An estimation summary of $\hat{\theta}$ from the 504 replicates of Finite Sample Size Experiment I (FSSE I).

Number of Genes	Number of Mice	θ	η	Bias	Standard Deviation	RMSE
N=4,000	40	.2	.25	-.022 (.0014)	.032	.039 (.0011)
			.50	-.020 (.0011)	.025	.032 (.0009)
		.4	.25	-.013 (.0016)	.036	.038 (.0013)
			.50	-.013 (.0012)	.027	.030 (.0009)
	80	.2	.25	-.024 (.0014)	.031	.039 (.0012)
			.50	-.022 (.0011)	.025	.033 (.0009)
		.4	.25	-.012 (.0014)	.032	.034 (.0010)
			.50	-.014 (.0012)	.027	.030 (.0009)
N=16,000	40	.2	.25	-.027 (.0007)	.016	.031 (.0007)
			.50	-.022 (.0006)	.014	.026 (.0006)
		.4	.25	-.014 (.0008)	.017	.022 (.0006)
			.50	-.013 (.0006)	.014	.019 (.0005)
	80	.2	.25	-.026 (.0007)	.016	.030 (.0007)
			.50	-.022 (.0006)	.013	.025 (.0005)
		.4	.25	-.014 (.0007)	.016	.021 (.0006)
			.50	-.012 (.0006)	.013	.018 (.0005)

Table G.10: An estimation summary of $\hat{\mu}_M$ from the 500 replicates of Finite Sample Size Experiment II (FSSE II).

Number of Genes	Number of Mice	θ	η	Bias	Standard Deviation	RMSE
N=4,000	10	.2	.25	-.25 (.134)	2.88	2.89 (.093)
			.50	-.12 (.077)	1.69	1.69 (.052)
		.4	.25	-.41 (.127)	2.79	2.81 (.096)
			.50	-.26 (.082)	1.82	1.84 (.055)
	20	.2	.25	-.46 (.104)	2.14	2.19 (.073)
			.50	-.32 (.068)	1.46	1.49 (.047)
		.4	.25	-.70 (.105)	2.22	2.32 (.072)
			.50	-.28 (.069)	1.51	1.54 (.049)
N=16,000	10	.2	.25	-.39 (.065)	1.41	1.47 (.048)
			.50	-.25 (.039)	0.87	0.90 (.027)
		.4	.25	-.36 (.063)	1.37	1.41 (.050)
			.50	-.10 (.038)	0.85	0.85 (.025)
	20	.2	.25	-.45 (.055)	1.09	1.18 (.042)
			.50	-.25 (.034)	0.72	0.77 (.024)
		.4	.25	-.67 (.053)	1.13	1.31 (.039)
			.50	-.27 (.035)	0.76	0.81 (.026)

Table G.11: An estimation summary of $\hat{\mu}_H$ from the 500 replicates of Finite Sample Size Experiment II (FSSE II).

Number of Genes	Number of Mice	θ	η	Bias	Standard Deviation	RMSE
N=4,000	10	.2	.25	.34 (.032)	.69	.77 (.025)
			.50	.40 (.019)	.43	.58 (.015)
		.4	.25	.35 (.030)	.66	.75 (.021)
			.50	.33 (.018)	.40	.52 (.014)
	20	.2	.25	-.25 (.029)	.58	.64 (.023)
			.50	-.15 (.019)	.40	.43 (.015)
		.4	.25	-.13 (.025)	.53	.54 (.017)
			.50	-.11 (.016)	.34	.36 (.012)
N=16,000	10	.2	.25	.30 (.019)	.41	.51 (.015)
			.50	.39 (.013)	.28	.48 (.010)
		.4	.25	.32 (.017)	.37	.48 (.013)
			.50	.36 (.012)	.27	.45 (.008)
	20	.2	.25	-.23 (.015)	.31	.40 (.013)
			.50	-.11 (.012)	.27	.29 (.011)
		.4	.25	-.18 (.015)	.32	.37 (.013)
			.50	-.12 (.012)	.26	.28 (.013)

Table G.12: An estimation summary of $\hat{\sigma}_M$ from the 500 replicates of Finite Sample Size Experiment II (FSSE II).

Number of Genes	Number of Mice	θ	η	Bias	Standard Deviation	RMSE
N=4,000	10	.2	.25	5.83 (.094)	2.01	6.16 (.089)
			.50	3.02 (.063)	1.38	3.32 (.061)
		.4	.25	5.71 (.094)	2.06	6.06 (.091)
			.50	2.87 (.060)	1.33	3.17 (.059)
	20	.2	.25	2.74 (.081)	1.66	3.20 (.079)
			.50	1.01 (.051)	1.11	1.50 (.042)
		.4	.25	2.62 (.075)	1.60	3.07 (.072)
			.50	0.87 (.048)	1.04	1.35 (.041)
N=16,000	10	.2	.25	5.75 (.047)	1.01	5.83 (.046)
			.50	2.98 (.028)	0.63	3.05 (.028)
		.4	.25	5.63 (.046)	1.01	5.72 (.045)
			.50	2.80 (.029)	0.64	2.87 (.028)
	20	.2	.25	2.69 (.040)	0.81	2.81 (.040)
			.50	1.01 (.025)	0.54	1.15 (.026)
		.4	.25	2.50 (.037)	0.78	2.62 (.036)
			.50	1.09 (.048)	0.55	1.05 (.023)

Table G.13: An estimation summary of $\hat{\sigma}_H$ from the 500 replicates of Finite Sample Size Experiment II (FSSE II).

Number of Genes	Number of Mice	θ	η	Bias	Standard Deviation	RMSE
N=4,000	10	.2	.25	0.13 (.043)	.92	0.92 (.029)
			.50	-0.36 (.023)	.51	0.62 (.021)
		.4	.25	0.40 (.038)	.84	0.93 (.030)
			.50	-0.25 (.025)	.55	0.61 (.025)
	20	.2	.25	-0.67 (.037)	.76	1.01 (.032)
			.50	-1.05 (.027)	.58	1.20 (.029)
		.4	.25	-0.43 (.037)	.78	0.89 (.034)
			.50	-0.85 (.024)	.51	0.99 (.024)
N=16,000	10	.2	.25	0.12 (.023)	.49	0.51 (.023)
			.50	-0.38 (.015)	.34	0.51 (.017)
		.4	.25	0.33 (.021)	.47	0.57 (.016)
			.50	-0.21 (.015)	.33	0.39 (.019)
	20	.2	.25	-0.71 (.021)	.43	0.83 (.020)
			.50	-1.03 (.017)	.37	1.10 (.019)
		.4	.25	-0.48 (.020)	.42	0.64 (.019)
			.50	0.40 (.025)	.37	0.95 (.020)

Table G.14: An estimation summary of $\hat{\rho}$ from the 500 replicates of Finite Sample Size Experiment II (FSSE II).

Number of Genes	Number of Mice	θ	η	Bias	Standard Deviation	RMSE
N=4,000	10	.2	.25	-.072 (.0073)	.16	.173 (.0084)
			.50	-.065 (.0061)	.13	.149 (.0066)
		.4	.25	-.032 (.0040)	.09	.094 (.0043)
			.50	-.032 (.0034)	.07	.081 (.0039)
	20	.2	.25	-.110 (.0087)	.18	.208 (.0117)
			.50	-.074 (.0060)	.13	.150 (.0070)
		.4	.25	-.049 (.0042)	.09	.102 (.0049)
			.50	-.048 (.0033)	.07	.087 (.0034)
N=16,000	10	.2	.25	-.069 (.0046)	.10	.122 (.0061)
			.50	-.035 (.0035)	.08	.085 (.0036)
		.4	.25	-.032 (.0022)	.05	.058 (.0023)
			.50	-.026 (.0018)	.04	.049 (.0025)
	20	.2	.25	-.081 (.0047)	.09	.124 (.0045)
			.50	-.049 (.0035)	.08	.090 (.0035)
		.4	.25	-.050 (.0023)	.05	.070 (.0026)
			.50	-.043 (.0020)	.04	.061 (.0025)

Table G.15: An estimation summary of $\hat{\eta}$ from the 500 replicates of Finite Sample Size Experiment II (FSSE II).

Number of Genes	Number of Mice	θ	η	Bias	Standard Deviation	RMSE
N=4,000	10	.2	.25	-.128 (.0003)	.006	.128 (.0003)
			.50	-.204 (.0004)	.008	.204 (.0004)
		.4	.25	-.124 (.0003)	.006	.124 (.0003)
			.50	-.197 (.0004)	.009	.197 (.0004)
	20	.2	.25	-.090 (.0003)	.009	.091 (.0003)
			.50	-.139 (.0004)	.006	.139 (.0004)
		.4	.25	-.086 (.0003)	.009	.086 (.0003)
			.50	-.131 (.0004)	.007	.131 (.0004)
N=16,000	10	.2	.25	-.128 (.0001)	.009	.128 (.0001)
			.50	-.204 (.0002)	.004	.204 (.0002)
		.4	.25	-.125 (.0001)	.004	.125 (.0001)
			.50	-.196 (.0002)	.003	.196 (.0002)
	20	.2	.25	-.090 (.0002)	.005	.091 (.0002)
			.50	-.139 (.0002)	.003	.139 (.0002)
		.4	.25	-.086 (.0002)	.005	.086 (.0002)
			.50	-.131 (.0002)	.003	.131 (.0002)

Table G.16: An estimation summary of $\hat{\theta}$ from the 500 replicates of Finite Sample Size Experiment II (FSSE II).

Number of Genes	Number of Mice	θ	η	Bias	Standard Deviation	RMSE
N=4,000	10	.2	.25	-.054 (.0021)	.044	.070 (.0017)
			.50	-.037 (.0017)	.037	.052 (.0012)
		.4	.25	-.070 (.0024)	.052	.087 (.0021)
			.50	-.040 (.0018)	.040	.057 (.0015)
	20	.2	.25	-.045 (.0021)	.043	.063 (.0018)
			.50	-.033 (.0016)	.034	.047 (.0013)
		.4	.25	-.060 (.0021)	.045	.074 (.0020)
			.50	-.028 (.0019)	.040	.049 (.0014)
N=16,000	10	.2	.25	-.060 (.0011)	.023	.064 (.0010)
			.50	-.044 (.0009)	.019	.048 (.0008)
		.4	.25	-.076 (.0013)	.027	.080 (.0012)
			.50	-.043 (.0009)	.021	.047 (.0010)
	20	.2	.25	-.054 (.0011)	.021	.058 (.0010)
			.50	-.040 (.0008)	.018	.044 (.0008)
		.4	.25	-.058 (.0011)	.023	.063 (.0010)
			.50	-.032 (.0009)	.020	.038 (.0008)

Table G.17: An estimation summary of $\hat{\mu}_M$ from the 530 replicates of Finite Sample Size Experiment III (FSSE III).

Number of Genes	Number of Mice	θ	η	Bias	Standard Deviation	RMSE
N=2,000	10	.8	.25	-.08 (.168)	3.87	3.86 (.111)
			.50	.28 (.101)	2.32	2.34 (.069)
	20	.8	.25	-.38 (.138)	3.16	3.18 (.094)
			.50	-.14 (.088)	2.03	2.04 (.065)

Table G.18: An estimation summary of $\hat{\mu}_H$ from the 530 replicates of Finite Sample Size Experiment III (FSSE III).

Number of Genes	Number of Mice	θ	η	Bias	Standard Deviation	RMSE
N=2,000	10	.8	.25	.43 (.030)	.69	0.81 (.024)
			.50	.38 (.018)	.40	0.55 (.015)
	20	.8	.25	.08 (.023)	.60	0.60 (.019)
			.50	.03 (.017)	.38	0.38 (.013)

Table G.19: An estimation summary of $\hat{\sigma}_M$ from the 530 replicates of Finite Sample Size Experiment III (FSSE III).

Number of Genes	Number of Mice	θ	η	Bias	Standard Deviation	RMSE
N=2,000	10	.8	.25	5.28 (.125)	2.88	6.01 (.120)
			.50	2.64 (.079)	1.81	3.20 (.077)
	20	.8	.25	2.32 (.097)	2.23	3.22 (.088)
			.50	0.70 (.062)	1.42	1.58 (.047)

Table G.20: An estimation summary of $\hat{\sigma}_H$ from the 530 replicates of Finite Sample Size Experiment III (FSSE III).

Number of Genes	Number of Mice	θ	η	Bias	Standard Deviation	RMSE
N=2,000	10	.8	.25	.80 (.039)	.90	1.20 (.035)
			.50	.25 (.025)	.56	0.62 (.021)
	20	.8	.25	.12 (.035)	.81	0.82 (.035)
			.50	-.30 (.022)	.51	0.60 (.023)

Table G.21: An estimation summary of $\hat{\rho}$ from the 530 replicates of Finite Sample Size Experiment III (FSSE III).

Number of Genes	Number of Mice	θ	η	Bias	Standard Deviation	RMSE
N=2,000	10	.8	.25	-.006 (.0021)	.049	.049 (.0021)
			.50	-.008 (.0016)	.036	.037 (.0015)
	20	.8	.25	-.015 (.0024)	.054	.056 (.0031)
			.50	-.016 (.0017)	.040	.042 (.0030)

Table G.22: An estimation summary of $\hat{\eta}$ from the 530 replicates of Finite Sample Size Experiment III (FSSE III).

Number of Genes	Number of Mice	θ	η	Bias	Standard Deviation	RMSE
N=2,000	10	.8	.25	-.112 (.0004)	.009	.112 (.0004)
			.50	-.173 (.0006)	.013	.173 (.0006)
	20	.8	.25	-.072 (.0004)	.009	.073 (.0004)
			.50	-.108 (.0005)	.013	.109 (.0005)

Table G.23: An estimation summary of $\hat{\theta}$ from the 530 replicates of Finite Sample Size Experiment III (FSSE III).

Number of Genes	Number of Mice	θ	η	Bias	Standard Deviation	RMSE
N=2,000	10	.8	.25	-.066 (.0026)	.059	.089 (.0022)
			.50	-.043 (.0020)	.047	.063 (.0019)
	20	.8	.25	-.068 (.0023)	.054	.086 (.0024)
			.50	-.033 (.0017)	.040	.052 (.0015)

Table G.24: An estimation summary of $\hat{\mu}_H$ from the human marginal model (2.21). Includes results from both Finite Sample Size Experiments I and II.

Number of Genes	Experiment	η	Bias	Standard Deviation	RMSE
N=4,000	FSSE I	.25	3.34 (.027)	1.12	3.56 (.033)
		.50	2.28 (.013)	0.60	2.35 (.016)
	FSSE II	.25	2.79 (.033)	1.42	3.13 (.038)
		.50	3.10 (.021)	0.93	3.24 (.023)
N=16,000	FSSE I	.25	3.18 (.015)	0.68	3.26 (.018)
		.50	2.24 (.010)	0.44	2.29 (.011)
	FSSE II	.25	2.70 (.016)	0.67	2.78 (.016)
		.50	3.10 (.013)	0.58	3.15 (.014)

Table G.25: An estimation summary of $\hat{\sigma}_H$ from the human marginal model (2.21). Includes results from both Finite Sample Size Experiments I and II.

Number of Genes	Experiment	η	Bias	Standard Deviation	RMSE
N=4,000	FSSE I	.25	0.28 (.022)	0.98	1.02 (.024)
		.50	0.09 (.010)	0.43	0.44 (.012)
	FSSE II	.25	2.86 (.041)	1.72	3.34 (.054)
		.50	1.99 (.026)	1.13	2.29 (.032)
N=16,000	FSSE I	.25	0.44 (.010)	0.45	0.63 (.011)
		.50	0.11 (.005)	0.22	0.24 (.005)
	FSSE II	.25	2.85 (.024)	1.01	3.02 (.028)
		.50	1.98 (.019)	0.83	4.60 (.100)

Table G.26: An estimation summary of $\hat{\eta}_H$ from the human marginal model (2.21). Includes results from both Finite Sample Size Experiments I and II.

Number of Genes	Experiment	η	Bias	Standard Deviation	RMSE
N=4,000	FSSE I	.25	-.19 (.000)	.013	.19 (.000)
		.50	-.31 (.001)	.025	.31 (.001)
	FSSE II	.25	-.20 (.000)	.010	.20 (.000)
		.50	-.38 (.001)	.024	.38 (.001)
N=16,000	FSSE I	.25	-.19 (.000)	.010	.19 (.000)
		.50	-.31 (.000)	.023	.31 (.001)
	FSSE II	.25	-.20 (.000)	.008	.20 (.000)
		.50	-.38 (.001)	.021	.38 (.001)

Table G.27: From FSSE I, the empirical powers and false discovery rates for detecting differentially expressed human genes at nominal FDR .05.

Number of Genes	Number of Mice	Joint				Marginal		BH	
		θ	η	Power	FDR	Power	FDR	Power	FDR
N=4,000	40	.2	.25	.081	.073	.053	.043	.027	.037
			.50	.125	.045	.082	.028	.048	.025
		.4	.25	.109	.059	.052	.046	.026	.041
			.50	.159	.038	.083	.027	.048	.026
	80	.2	.25	.083	.077	.052	.045	.027	.039
			.50	.129	.046	.082	.027	.047	.025
		.4	.25	.113	.060	.051	.045	.025	.036
			.50	.167	.041	.082	.028	.048	.026
N=16,000	40	.2	.25	.081	.071	.053	.044	.026	.037
			.50	.123	.045	.082	.027	.047	.025
		.4	.25	.109	.057	.052	.045	.025	.038
			.50	.159	.038	.083	.028	.047	.026
	80	.2	.25	.083	.074	.052	.044	.025	.038
			.50	.129	.047	.082	.027	.047	.025
		.4	.25	.114	.059	.051	.044	.025	.038
			.50	.168	.040	.083	.027	.047	.025
minimum standard error				.00072	.00029	.00057	.00032	.00050	.00039
maximum standard error				.00128	.00145	.00094	.00138	.00087	.00199

Table G.28: From FSSE I, the empirical powers and false discovery rates for detecting differentially expressed human genes at nominal FDR .20.

Number of Genes	Number of Mice	Joint				Marginal		BH	
		θ	η	Power	FDR	Power	FDR	Power	FDR
N=4,000	40	.2	.25	.171	.155	.103	.092	.098	.152
			.50	.288	.116	.171	.066	.160	.099
		.4	.25	.245	.141	.102	.096	.097	.148
			.50	.403	.113	.172	.065	.160	.100
	80	.2	.25	.177	.165	.103	.095	.098	.153
			.50	.304	.124	.171	.065	.159	.100
		.4	.25	.260	.145	.100	.094	.095	.149
			.50	.438	.121	.170	.067	.159	.101
N=16,000	40	.2	.25	.171	.154	.103	.093	.098	.150
			.50	.285	.116	.169	.066	.158	.100
		.4	.25	.243	.137	.101	.093	.095	.150
			.50	.401	.113	.171	.066	.160	.100
	80	.2	.25	.177	.161	.102	.094	.097	.151
			.50	.303	.123	.170	.066	.159	.100
		.4	.25	.259	.143	.100	.094	.094	.150
			.50	.438	.120	.170	.066	.159	.100
minimum standard error				.0011	.0003	.0010	.0003	.0011	.0004
maximum standard error				.0020	.0013	.0016	.0013	.0017	.0018

Table G.29: From FSSE I, the empirical powers and false discovery rates for detecting I-Orthologs.

Number of Genes	Number of Mice	Nominal Fdr =		.05		.20	
		θ	η	Power $\times 100$	FDR	Power $\times 100$	FDR
N=4,000	40	.2	.25	.038 (.0041)	.014 (.0050)	1.111 (.0369)	.236 (.0075)
			.50	.013 (.0020)	.017 (.0057)	0.605 (.0203)	.247 (.0099)
		.4	.25	.383 (.0117)	.070 (.0051)	4.971 (.0615)	.185 (.0019)
			.50	.186 (.0072)	.074 (.0081)	5.424 (.0731)	.213 (.0019)
	80	.2	.25	.034 (.0036)	.015 (.0052)	1.110 (.0346)	.242 (.0073)
			.50	.013 (.0019)	.013 (.0049)	0.620 (.0209)	.255 (.0097)
		.4	.25	.390 (.0119)	.077 (.0057)	5.142 (.0600)	.183 (.0019)
			.50	.191 (.0073)	.076 (.0069)	5.293 (.0719)	.212 (.0019)
N=16,000	40	.2	.25	.030 (.0019)	.034 (.0062)	1.027 (.0174)	.237 (.0032)
			.50	.016 (.0013)	.046 (.0084)	0.623 (.0120)	.266 (.0044)
		.4	.25	.388 (.0065)	.074 (.0025)	4.858 (.0389)	.183 (.0009)
			.50	.215 (.0041)	.090 (.0042)	5.317 (.0482)	.215 (.0010)
	80	.2	.25	.031 (.0018)	.055 (.0080)	1.005 (.0188)	.232 (.0036)
			.50	.015 (.0013)	.036 (.0071)	0.650 (.0127)	.266 (.0046)
		.4	.25	.406 (.0064)	.076 (.0024)	5.076 (.0396)	.182 (.0010)
			.50	.238 (.0045)	.086 (.0037)	5.380 (.0511)	.212 (.0010)

Table G.30: From FSSE II, the empirical powers and false discovery rates for detecting differentially expressed human genes at nominal FDR .05.

Number of Genes	Number of Mice	Joint				Marginal		BH	
		θ	η	Power	FDR	Power	FDR	Power	FDR
N=4,000	10	.2	.25	.028	.286	.020	.253	.0007	.033
			.50	.043	.145	.030	.118	.0010	.034
		.4	.25	.036	.233	.020	.266	.0010	.049
			.50	.052	.121	.029	.123	.0013	.046
	20	.2	.25	.031	.293	.021	.240	.0006	.032
			.50	.047	.153	.030	.114	.0009	.033
		.4	.25	.041	.230	.020	.254	.0008	.046
			.50	.062	.119	.029	.118	.0009	.044
N=16,000	10	.2	.25	.028	.287	.020	.253	.0006	.051
			.50	.042	.145	.029	.119	.0004	.029
		.4	.25	.035	.236	.019	.264	.0004	.042
			.50	.053	.118	.029	.122	.0007	.031
	20	.2	.25	.031	.295	.021	.248	.0002	.036
			.50	.048	.151	.031	.114	.0005	.026
		.4	.25	.041	.228	.020	.253	.0002	.028
			.50	.061	.118	.029	.118	.0007	.030
minimum standard error				.0005	.0014	.0003	.0013	.0000	.0053
maximum standard error				.0008	.0046	.0005	.0052	.0003	.0089

Table G.31: From FSSE II, the empirical powers and false discovery rates for detecting differentially expressed human genes at nominal FDR .20.

Number of Genes	Number of Mice	Joint				Marginal		BH	
		θ	η	Power	FDR	Power	FDR	Power	FDR
N=4,000	10	.2	.25	.063	.333	.041	.297	.0040	.143
			.50	.113	.181	.066	.149	.0110	.105
		.4	.25	.090	.258	.041	.302	.0045	.165
			.50	.151	.147	.064	.150	.0103	.112
	20	.2	.25	.074	.340	.043	.292	.0043	.136
			.50	.133	.191	.067	.144	.0107	.107
		.4	.25	.111	.260	.042	.295	.0040	.165
			.50	.200	.149	.066	.147	.0098	.103
N=16,000	10	.2	.25	.063	.331	.041	.296	.0026	.169
			.50	.111	.181	.065	.148	.0081	.010
		.4	.25	.086	.262	.039	.303	.0022	.157
			.50	.152	.144	.065	.150	.0088	.105
	20	.2	.25	.073	.341	.043	.291	.0024	.161
			.50	.134	.189	.068	.145	.0094	.101
		.4	.25	.110	.254	.041	.293	.0021	.144
			.50	.197	.148	.065	.147	.0088	.102
minimum standard error				.0008	.0007	.0005	.0010	.0001	.0042
maximum standard error				.0018	.0034	.0010	.0039	.0005	.0099

Table G.32: From FSSE II, the empirical powers and false discovery rates for detecting I-Orthologs.

Number of Genes	Number of Mice	Nominal Fdr =		.05		.20	
		θ	η	Power $\times 100$	FDR	Power $\times 100$	FDR
N=4,000	10	.2	.25	.0062 (.0016)	.014 (.0053)	0.282 (.0170)	.241 (.0138)
			.50	.0025 (.0008)	.004 (.0029)	0.145 (.0091)	.205 (.0140)
		.4	.25	.0804 (.0052)	.047 (.0070)	1.427 (.0316)	.226 (.0049)
			.50	.0369 (.0031)	.044 (.0083)	1.680 (.0418)	.268 (.0043)
	20	.2	.25	.0026 (.0009)	.008 (.0043)	0.251 (.0155)	.228 (.0141)
			.50	.0021 (.0007)	.011 (.0048)	0.118 (.0080)	.184 (.0140)
		.4	.25	.0721 (.0046)	.057 (.0080)	1.645 (.0389)	.234 (.0044)
			.50	.0350 (.0029)	.055 (.0090)	1.752 (.0409)	.272 (.0042)
N=16,000	10	.2	.25	.0039 (.0005)	.017 (.0057)	0.233 (.0082)	.303 (.0092)
			.50	.0019 (.0004)	.016 (.0055)	0.136 (.0054)	.334 (.0120)
		.4	.25	.0818 (.0030)	.115 (.0077)	1.320 (.0232)	.227 (.0022)
			.50	.0423 (.0018)	.099 (.0080)	1.649 (.0280)	.267 (.0021)
	20	.2	.25	.0039 (.0006)	.031 (.0084)	0.222 (.0084)	.293 (.0100)
			.50	.0025 (.0004)	.007 (.0037)	0.138 (.0051)	.338 (.0120)
		.4	.25	.0775 (.0028)	.107 (.0081)	1.621 (.0253)	.226 (.0021)
			.50	.0433 (.0018)	.102 (.0085)	1.674 (.0288)	.267 (.0022)

Table G.33: From FSSE III, the empirical powers and false discovery rates for detecting differentially expressed human genes at nominal FDR .05.

Number of Genes	Number of Mice			Joint		Marginal		BH	
		θ	η	Power	FDR	Power	FDR	Power	FDR
N=2,000	10	.8	.25	.096	.081	.018	.268	.0014	.033
			.50	.142	.041	.029	.121	.0016	.035
	20	.8	.25	.126	.078	.019	.271	.0024	.059
			.50	.198	.040	.029	.124	.0023	.035
	minimum standard error			.0014	.0007	.0004	.0031	.0002	.0062
	maximum standard error			.0023	.0024	.0005	.0076	.0006	.0092

Table G.34: From FSSE III, the empirical powers and false discovery rates for detecting differentially expressed human genes at nominal FDR .20.

Number of Genes	Number of Mice			Joint		Marginal		BH	
		θ	η	Power	FDR	Power	FDR	Power	FDR
N=2,000	10	.8	.25	.343	.110	.039	.306	.0057	.145
			.50	.530	.077	.066	.149	.0128	.101
	20	.8	.25	.503	.112	.039	.321	.0069	.171
			.50	.732	.091	.064	.151	.0123	.110
	minimum standard error			.0018	.0005	.0008	.0024	.0004	.0058
	maximum standard error			.0041	.0010	.0010	.0060	.0007	.0108

Table G.35: From FSSE III, the empirical powers and false discovery rates for detecting I-Orthologs.

Number of Genes	Number of Mice	Nominal Fdr =		.05		.20	
		θ	η	Power $\times 100$	FDR	Power $\times 100$	FDR
N=2,000	10	.8	.25	1.762 (.0421)	.052 (.0022)	23.95 (1.276)	.133 (.0019)
			.50	2.531 (.0729)	.061 (.0021)	59.46 (1.420)	.157 (.0016)
	20	.8	.25	1.948 (.0455)	.053 (.0026)	30.35 (1.295)	.137 (.0018)
			.50	2.626 (.0615)	.061 (.0021)	72.50 (1.182)	.166 (.0013)

Table G.36: Effects on the distribution of human differential expression posterior probabilities (the p_i s) when one parameter is modulated from its true value. This report is based on a single simulated dataset from joint model (2.18) with $N = 10,000$, $\eta = .5, \theta = .4$, and the remaining parameters set to values as estimated from the GlaxoSmithKline Data (Table 2.2).

Percentile	p_i	$p_i^{\mu_M^-} - p_i$	$p_i^{\mu_M^+} - p_i$	$p_i^{\sigma_M^-} - p_i$	$p_i^{\sigma_M^+} - p_i$	$p_i^{\mu_H^-} - p_i$	$p_i^{\mu_H^+} - p_i$	$p_i^{\sigma_H^-} - p_i$	$p_i^{\sigma_H^+} - p_i$
5	.04	-.01	-.01	-.01	-.01	-.05	-.09	-.07	-.06
25	.21	-.00	-.00	.00	-.01	-.01	-.03	-.00	-.03
50	.34	-.00	.00	.00	-.00	.00	-.01	.01	-.01
75	.60	.00	.00	.00	-.00	.02	.01	.04	-.00
95	.87	.01	.01	.01	.01	.06	.04	.09	.03
alternatives		μ_M^-	μ_M^+	σ_M^-	σ_M^+	μ_H^-	μ_H^+	σ_H^-	σ_H^+
		-6	1	50	65	1	5	4	9

Percentile	p_i	$p_i^{\rho^-} - p_i$	$p_i^{\rho^+} - p_i$	$p_i^{\eta^-} - p_i$	$p_i^{\eta^+} - p_i$	$p_i^{\theta^-} - p_i$	$p_i^{\theta^+} - p_i$
5	.04	-.03	-.02	-.20	.02	-.06	-.05
25	.21	-.01	-.00	-.15	.09	-.02	-.04
50	.34	-.00	-.00	-.12	.13	.01	-.01
75	.60	.00	.00	-.08	.16	.04	-.02
95	.87	.05	.01	-.02	.20	.05	.05
alternatives		ρ^-	ρ^+	η^-	η^+	θ^-	θ^+
		.60	.99	.30	.70	.3	.5

Table G.37: Effects on the local false discovery rates for differentially expressed human genes (the p_i s for which $\beta_{Hi} = 0$) when one parameter is modulated from its true value. This report is based on a single simulated dataset from joint model (2.18) with $N = 10,000$, $\eta = .5$, $\theta = .4$, and the remaining parameters set to values as estimated from the GlaxoSmithKline Data (Table 2.2).

Percentile	p_i	$p_i^{\mu_M^-} - p_i$	$p_i^{\mu_M^+} - p_i$	$p_i^{\sigma_M^-} - p_i$	$p_i^{\sigma_M^+} - p_i$	$p_i^{\mu_H^-} - p_i$	$p_i^{\mu_H^+} - p_i$	$p_i^{\sigma_H^-} - p_i$	$p_i^{\sigma_H^+} - p_i$
5	.03	-.01	-.01	-.01	-.01	-.04	-.08	-.04	-.05
25	.10	-.00	.00	.00	-.00	-.00	-.03	.01	-.03
50	.24	-.00	.00	.00	-.00	.01	-.01	.02	-.01
75	.36	-.00	.00	.00	-.00	.02	-.00	.04	-.00
95	.70	.00	.01	.01	.01	.06	.04	.09	.01
100	1	.02	.03	.02	.07	.13	.11	.25	.18
alternatives		μ_M^-	μ_M^+	σ_M^-	σ_M^+	μ_H^-	μ_H^+	σ_H^-	σ_H^+
		-6	1	50	65	1	5	4	9

Percentile	p_i	$p_i^{\rho^-} - p_i$	$p_i^{\rho^+} - p_i$	$p_i^{\eta^-} - p_i$	$p_i^{\eta^+} - p_i$	$p_i^{\theta^-} - p_i$	$p_i^{\theta^+} - p_i$
5	.03	-.03	-.02	-.18	.02	-.05	-.05
25	.10	-.00	-.00	-.13	.07	.00	-.04
50	.24	-.00	-.00	-.10	.12	.02	-.02
75	.36	.00	.00	-.05	.15	.04	-.00
95	.70	.05	.01	-.01	.20	.05	.05
100	1	.15	.15	-.00	.22	.06	.08
alternatives		ρ^-	ρ^+	η^-	η^+	θ^-	θ^+
		.60	.99	.30	.70	.3	.5