<title>

Benjamin Yu Hang Bai

November 12, 2019

<dedication>

Abstract

<thesis abstract>

Acknowledgements

pipelines

oucru team

research assistants/research managers

family friends cuams churchill MCR, various badminton

stackexchange publication quality dialogue, model for future peer review?

Contents

List of Figures **xi**

List of Tables **xiii**

1 Introduction **1**

1.1 A brief history of complex trait genetics 1

1.1.1 The era of GWAS 1

1.1.2 Post-GWAS: interpretation of genetic associations . . 1

1.2 Immunity is a complex trait 2

1.2.1 Genetic factors affecting the healthy immune system . 2

1.2.2 Genetic factors affecting immune response to challenge 2

1.3 Immune response to vaccination 2

1.3.1 Systems vaccinology: from empirical to rational vaccinology 3

1.3.2 Genetics factors affecting vaccine response 4

1.4 Immune response to biologic therapies 4

1.5 Thesis overview 4

2 Transcriptomic response to influenza A (H1N1)pdm09 vaccine (Pandemrix) **5**

2.1 Introduction 5

2.1.1 Systems vaccinology of influenza vaccines 6

2.1.2 Pandemic influenza and Pandemrix 6

2.1.3 The Human Immune Response Dynamics (HIRD) study 6

2.2 Methods 6

2.2.1 HIRD cohort 6

2.2.2 RNA-seq data generation 6

2.2.3 RNA-seq quality control 7

2.2.4 Computing TRI . . . 7
2.2.5 differential gene expression (DGE) . . . 7
2.2.6 DGE meta-analysis . . . 7
2.3 Gene set enrichment analysis . . . 10
2.4 Results . . . 10
2.4.1 Innate and adaptive immune response . . . 10
2.4.2 Expression associated with antibody response . . . 10
2.4.3 Identifying molecular signatures of response . . . 10
2.5 Discussion . . . 10
2.5.1 Comparison to Sobolev R vs. NR . . . 10

3 Genetic factors affecting Pandemrix vaccine response 13
3.1 Introduction . . . 13
3.1.1 Context-specific immune response QTLs . . . 13
3.1.2 *in vivo* response QTL mapping . . . 13
3.2 Methods . . . 14
3.2.1 Genotyping data generation . . . 14
3.2.2 Genotyping quality control . . . 14
3.2.3 Imputation . . . 14
3.2.4 Mapping cis-eQTLs with LMM . . . 14
3.2.5 eQTL meta-analysis . . . 16
3.2.6 Colocalization . . . 16
3.3 Results . . . 17
3.3.1 eQTLs at each timepoint . . . 17
3.3.2 Estimation of sharing . . . 17
3.3.3 Colocalisation of re-eQTLs with known context-specific immune QTLs . . . 17
3.3.4 Polygenic score to predict antibody response . . . 17
3.4 Discussion . . . 17

4 Response to live attenuated rotavirus vaccine (Rotarix) in Vietnamese infants 19
4.1 Introduction . . . 19
4.1.1 The genetics of vaccine response in early life . . . 19
4.1.2 Rotavirus and rotarix in Vietnam . . . 19
4.1.3 Known factors that affect rotavirus vaccine efficacy . . . 19
4.2 Methods . . . 19

4.2.1 RNA-seq data generation 19
4.2.2 Genotyping 20
4.3 Results 20
4.4 Discussion 20

5 multiPANTS **21**
5.1 Introduction 21
5.2 Methods 21
5.3 Results 21
5.4 Discussion 21

6 Discussion **23**

A Supplementary Materials **25**
A.1 Chapter 2 25
A.2 Chapter 3 25
A.3 Chapter 4 26

Bibliography **27**

List of Abbreviations **29**

List of Figures

List of Tables

Chapter 1

Introduction

Why study human genetics?

Human variation Nature vs nurture Causal anchors. Leveraging natural G variation.

1.1 A brief history of complex trait genetics

Prior to GWAS Mendelian genetics, family and linkage studies

Twin studies and heritability estimates.

Candidate gene studies (Border et al., 2019)

1.1.1 The era of GWAS

Common disease, common variant X years of GWAS

Missing heritability Rare variation, burden tests

Fine-mapping

Pathway analysis

TWAS PheWAS[1] MR

expanding into global populations polygenic scores

1.1.2 Post-GWAS: interpretation of genetic associations

Why? Understand mech. Drug target prioritisation for disease traits e.g. of successful GWAS -> drug target

Locus to gene problem, nc variation. Genome-wide association studies have successfully identified genetic variants associated with immune-mediated disease, the majority of which are non-coding[10 Years of GWAS Discovery].

Under the assumption that the mechanism by which non-coding associations affect disease risk is through their effect on gene expression, a successful way to link associations to their target gene is by statistical colocalisation with eQTL datasets, to determine if the GWAS and eQTL signal share the same causal variant[Co-localization of Conditional eQTL and GWAS Signatures in Schizophrenia].

molecular quantitative trait loci (QTLs) in context Coloc

Context is key

1.2 Immunity is a complex trait

Immune-mediated diseases Heritability of immune parameters and immune-mediated diseases

1.2.1 Genetic factors affecting the healthy immune system

Why study health? Factors affecting the healthy immune system.

1.2.2 Genetic factors affecting immune response to challenge

Given the genetic control of the healthy immune system, one can hypothesise that immune response to challenge may also be influenced by genetic factors.

The need for controlled immune challenge in trials. Studies of natural infection are complicated. Drug trials as an opportunity: Vaccines and biologics for controlled immune challenge.

1.3 Immune response to vaccination

Vaccination has enormous impact on global health [10.1098/rstb.2013.0433].

Vaccines stimulate the immune system with pathogen-derived antigens to induce effector responses (primarily antigen-specific antibodies) and immunological memory against the pathogen itself. These effector responses are then be rapidly reactivated in cases of future exposure to the pathogen, mediating long-term protection.

1.3.1 Systems vaccinology: from empirical to rational vaccinology

History of vaccine dev [summary of low-throughput immunology e.g. animal models]

However, a vaccine that is highly efficacious in one human population may have significantly lower efficacy in other populations. [1 statistic on vaccine efficiacy differences e.g. rotavirus] Particularly challenging populations for vaccination include the young and elderly, immunosuppressed patients, ethnically-diverse populations, and developing countries. For the majority of licensed vaccines, there is a lack of understanding regarding the molecular mechanisms that underpin this variation in host immune response. Immunological mechanisms that underpin a specific vaccine's success or failure in a given individual are often poorly understood[Immunological mechanisms of vaccination].

rational vacc, where the key is sys vacc Review of systems vaccinology (pull out of self_viva_copypasta) These systems vaccinology studies often consider longitudinal measurements of the transcriptomic, cellular, cytokine, and antibody immune responses following vaccination[Vaccinology in the era of high-throughput biology.]. Systems vaccinology is the application of -omics technologies to provide a systems-level characterisation of the human immune system after vaccine-perturbation. Measurements are taken at multiple molecular levels (e.g. genome, transcriptome, proteome), and molecular signatures that correlate with and predict vaccine-induced immunity are identified [http://dx.doi.org/10.1098/rstb.2014.0146]. Systems vaccinology has been successfully applied to a variety of licensed vaccines [yellow fever, influenza], and also to vaccine candidates against [HIV, malaria], resulting in the identification of early transcriptomic signatures that predict vaccine-induced antibody responses.

How to use sysvacc to inform better design (A systems framework for vaccine design Mooney2013), and how to move towards personalised vaccinology (https://doi.org/10.1016/j.vaccine.2017.07.062).

Overview, including pathogen-side factors

1.3.2 Genetics factors affecting vaccine response

Relatively few studies have assessed the impact of human genetic variation on responses[Franco, Lareau 2016].

This is despite evidence from genome-wide association studies suggesting such genetic variation influences immune response to vaccines and susceptibility to disease[Systems immunogenetics of vaccines.].

Search for "variation in vaccine response genetics GA Poland" in google scholar

Genetics of adverse events e.g. https://www.ncbi.nlm.nih.gov/pubmed/18454680

Results from vaccine-related twin studies e.g. in "TWIN STUDIES ON GENETIC VARIATIONS IN RESISTANCE TO TUBERCULOSIS", and (Defective T Memory Cell Differentiation after Varicella Zoster Vaccination in Older Individuals)

Review paper on GWAS for vaccines mooney2013SystemsImmunogeneticsVaccines

1.4 Immune response to biologic therapies

1.5 Thesis overview

Chapters 1 and 2. Chapter 3. Chapter 4. Chapter 5.

Chapter 2

Transcriptomic response to influenza A (H1N1)pdm09 vaccine (Pandemrix)

2.1 Introduction

[The influence of host genetics on vaccines response has also been explored] Vaccine-induced antibody response is a complex trait, with heritability estimates ranging from ... [e.g. seaonsal influenza 10.1016/j.vaccine.2008.07.065 Poland e.g. smallpox e.g. measeks 10.1080/21645515.2015.1119345.]

A potential mechanism through which genetic variation can affect vaccine response is through altering the expression of nearby genes (cis-eQTLs). In the case of inactivated trivalent influenza vaccine, genetic variation in membrane trafficking and antigen processing genes was associated with both transcriptomic and antibody responses in patients after vaccination [Franco]. [summary of Sobolev findings]

In this study, we model the influence of host genetics on longitudinal transcriptomic and antibody responses to Pandemrix, in vivo.

also, we have phenotype data, in vivo

[main aim: how much variation in response is genetic?] [other aims: assess differences to seasonal influenza vaccines] [summary of main results]

Knowns Sobolev: R vs NR, inconsistent variation in why people are NR

Prevacc signatures of Tri Using larger transcriptomic dataset Are they genetic

Good points Repeated measures in vivo perturbation

2.1.1 Systems vaccinology of influenza vaccines

2.1.2 Pandemic influenza and Pandemrix

Intro to the virus (Characteristics of Swine-Origin 2009 A(H1N1), DOI: 10.1126/science.1176225)

Pandemrix, as one of several vaccines licensed: https://www.ema.europa.eu/en/human-regulatory/overview/public-health-threats/pandemic-influenza/2009-h1n1-influenza-pandemic/medicines-authorised-during-pandemic

Relationship to seasonal H1N1. ...a single dose of monovalent 2009 H1N1 vaccine was recommended in adults, but young children were recommended to receive 2 doses (reviewed by [3••]). It is likely that a single dose was sufficient to induce immunity in adults because prior exposure to seasonal H1N1 viruses had immunologically primed the population. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3224079/ "Seasonal influenza vaccine provides priming for A/H1N1 immunization." https://www.ncbi.nlm.nih.gov/pubmed/20371459 Demonstration in a mouse model: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3024675/

- more variation will be explained by history of exposure rather than genetics, so may be harder to detect

Knowns about AS03

Narcolepsy controversy (evidence for genetics)

2.1.3 The Human Immune Response Dynamics (HIRD) study

https://www.nature.com/articles/ni.3372

2.2 Methods

2.2.1 HIRD cohort

Describe existing data types. Extend with RNAseq (disjoint) and genotyping.

2.2.2 RNA-seq data generation

Do we have enough reads for RNAseq analysis? https://www.ncbi.nlm.nih.gov/pubmed/24434847 and doi:10.1093/bioinformatics/btt688

2.2.3 RNA-seq quality control

2.2.4 Computing TRI

2.2.5 differential gene expression (DGE)

Batch effect correction (see batch effects tag) Combat is best here LM, LMM, Combat were comparable In some cases, Combat overcorrects But main issue is unbalanced design, which affects even 2-way anova. Rather than 2-step, Safest is to use a covariate, which seems to at least create appropriate conficence intervals (1e)

Why combine -7 and 0? See Sobolev: (a) Observed values of multivariate statistic t (m.v.t.) quantifying global PBMC gene-expression dissimilarity in comparison of two study days (red dots) to values expected when days are randomly assigned between groups.

Should we meta-analyse?

In conclusion, we found that underpowered studies play a very substantial role in meta-analyses reported by Cochrane reviews, since the majority of meta-analyses include no adequately powered studies. In meta-analyses including two or more adequately powered studies, the remaining underpowered studies often contributed little information to the combined results, and could be left out if a rapid review of the evidence is required.

2.2.6 DGE meta-analysis

Describe method here, put rest in appendix

Cross-platform meta-analysis methods

Whilst there is a slew of literature on metanalyasis of rnaseq adn array (e.g. metaMA), combining platforms is fraught with difficultiy. different probes, different tech -> diff stat models

Why expected het? platform effect (ratio compression, differences in preprocessing to genes). different sets of samples (more extreme in array)

exmplaes of e.g. random effects model of approx 24 datasets: e.g. sva: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3617154/

Alternative is: CorMotif first applies limma (Smyth, 2004) to each study separately. CorMotif for microarray data since it was motivated by the microarray analysis in the SHH study. However, the idea behind CorMotif is

general, and it should be straightforward to develop a similar framework for RNA-seq data.

Or MetaVolcano: vote counting, REM (note small k),

Sweeny tests diff ks: Methods to increase reproducibility in differential

or complicated R package, CBM ("Cross-platform Bayesian Model"), also see CBM paper for discussion of difficulties of combinding platform

cannot acually use CBM, as it operates on expressions, with a binary case vs control, so no covariates same limitation for cormotif, although it takes any number of groups

rankprod (focus on case/control design), mayday seasight

Choice of meta-analysis model

Two schools of thought for frequentist meta-analysis: fixed-effect, or in the presence of het, random-effects.

We have het, so def use random effects.

How to estimate het? Many methods to estimate het, but

The problem: we only have k=2, and MLE estimates of tau are not very good with k=2. Highly imprecise, and often: boundary estimate problems.

and We know 0 het is inappropriate.

Bayesian random-effects meta is attractive. But What priors should we use?

Prior for between-studies heterogeneity

Prior for tau.

A general rec is: Use distribution in the half-t family e.g. Cauchy (df=1) when the number of groups is small and in other settings where a weakly-informative prior is desired. In their 3-schools examples, choose a value of scale just higher than expected, this is to weakly constrain the posterior, and not to actually represent prior knowledge. - Warn against inverse-gamma(e, e), as it can influence the posterior mean.

But weak priors are not recommended, as k is small, so there is little information in the data.

We can get empirical distribution of many genes. fit a default reml model, exclude 0 ests. Advantage of getting the correct parameter scale for our data.

So use Empirical Bayes: aside: empirical bayes is popular for high dim data e.g. edgeR, DESeq2, limma-voom, combat (method of moments)

Papers that fit empirical datasets for tau2: Most of these are inverse-gamma/log-t family Fit inverse gamma distribution on method of moments estimates from 18 gastroenterology trials with similar endpoints. This paper has described the distribution of the between-study variance amongst Cochrane reviews published between 2008 and 2009, and investigating a binary outcome. A log-normal distribution incorporating the association between the between-study variance and the pooled effect size gave the best fit. Predictive distributions are presented for nine different settings, defined by type of outcome and type of intervention comparison. For example, for a planned meta-analysis comparing a pharmacological intervention against placebo or control with a subjectively measured outcome, the predictive distribution for heterogeneity is a log-normal (2.13, 1.582) distribution, which has a median value of 0.12. Model selection based on the deviance information criterion (DIC) [8] led to the choice of the log-t model for t2. (5df) The priors are derived as log-normal distributions for the between-study variance, applicable to meta-analyses of binary outcomes on the log odds-ratio scale.

We choose gamma: as Density at tau=0 is 0, but increases linearly from 0, so values close to 0 are still permitted if the data suggests it. For lognormal/inverse gamma, they have a derivative of 0 at tau=0, so they rule out small tau no matter what the data suggest. For The exponential and half-Cauchy families, for example, do not decline to zero at the boundary, so they do not rule out posterior mode estimates of zero.

Prior for DGE effect size

Prior for logFC

Not as much discussion in the lit: There is Typically enough data to estimate this to use a non informative prior. Even Friede Uses noninformative flat.

Two choices in bayesmeta are uniform and normal. We know Mean is 0: most genes are not DE. so flat prior makes no sense

To avoid overshrinking, could consider heavy-tailed priors (e.g. cauchy) for mu rather than normal, but this is not possible in bayesmeta. Cauchy 2.5 DEseq/apeglm: prior on logfc, cauchy with scale adapted.

But bayesmeta is normal. So weaken further to place more prior on larger values. This means less shrinkage.

Also: we will shrink again with ashr. which can fit a more complicated distr (mixture?)

So We use a very weak normal prior, scaled to each coef, as we still want some scaling based on parameter scales. Equiv to saying 95pc chance that effect is within log2FC of 20.

2.3 Gene set enrichment analysis

2.4 Results

2.4.1 Innate and adaptive immune response

2.4.2 Expression associated with antibody response

In clinical studies seroprotection is normally defined as a specific antibody titer or antibody titer increase (seroconversion).22

2.4.3 Identifying molecular signatures of response

2.5 Discussion

The reduced efficacy of vaccination has also been linked to excessive inflammation for influenza,31 yellow fever,32 tuberculosis,33 and hepatitis B34 vaccines.

2.5.1 Comparison to Sobolev R vs. NR

Stuff from 1st year report.

Current limitations Confounded by changes in immune cell proportions in bulk PBMCs

No conditional eQTL analysis to disentangle conditional effects Unclear connection to vaccine biology e.g. what genesets/pathways/cell types are driving the observed transcriptomic and eQTL response? Future work to address limitations Colocalisation with known associations Colocalisation is used to understand the molecular basis of GWAS associations (of a variety of human disease traits) (Giambartolome, 2014) Here the inverse: coloc is

used to understand the biological relevance of observed expression variation
Prediction of antibody response e.g. using expression of gene modules
adjuvant:

Chapter 3

Genetic factors affecting Pandemrix vaccine response

3.1 Introduction

Utility of genetics: allows coloc How does common genetic variation affect response to vaccine?

eQTL becomes more or less important after perturbation: Tells you something about the mechanism of perturbation. Either expression regulatory activation/repression (signalling cascade -> TFs, chromatin remodelling etc.)

3.1.1 Context-specific immune response QTLs

types of conditinoal QTL ackerman conditional vs dynamic

Review of stimulation condition QTL mapping, invitro and invivo what models used? did they use change scores for longitudinal?

QTLs can interact with sex and age

Mechanisms:

3.1.2 *in vivo* response QTL mapping

Review of in vivo mapping. Franco Lareau smallpox apoptosis Caliskan Rhinovirus Davenport

3.2 Methods

3.2.1 Genotyping data generation

3.2.2 Genotyping quality control

3.2.3 Imputation

why exclude x chrom? As is standard for imputation, we excluded all X-linked SNPs for the following reasons: (i) the X chromosome has to be treated differently from the autosomes; (ii) it cannot be predicted which allele is active on the X chromosome, (iii) testing males separately from females results in different sample sizes and power. Imputation of SNPs in the HapMap CEU population was performed using either MACH46 or IMPUTE47. All SNPs with a MAF <0.01 were excluded from analysis. In total, up to 2.11 million genotyped or imputed SNPs were analyzed.

3.2.4 Mapping cis-eQTLs with LMM

lmms: use a kinship matrix to scale the sample-sample genetic covariance see: 2018-11-16 notes in log

this is good background

Choice of lmm method for various methods, see 2018-03-05 and 2018-07-25 in log

for discussion of how lmm implementation doesn't matter (Eu-ahsunthornwattana et al., 2014)

Can also refer to previous notes in "2017_Book_SystemsGenetics"

why including known covariates: why not a two stage approach?

Why not mapping on deltas? (if we are interested in the direct question of G on change) ackermann: change scores are prone to increased noise from franco: "We attempted analyses with an approach similar to that proposed by the reviewers in the course of our work, but found that the approach that was ultimately chosen to explore the day differences was the most powerful. Specifically, utilizing a pairwise comparison (difference) between time points as the substrate for the eQTL analysis would lead to an increase in the technical variance of the phenotype, as the sum of two independent (technical) errors has twice the variance of an individual measurement. "

The final model:

Estimation of cell type abundances

decon eqtl requires full data i.e. it's an eqtl mapper

cell type interaction terms

Why impute for cell counts but not for eQTL? expression matricse are mostly complete, and we only exclude genes based on low expression in RNAseq we cannot drop whole panels so easily like we can drop genes

Kinship matrix computation

LDAK kinship matrix construction http://dougspeed.com/method-overview/

Note: can be negative https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6157025/ LDAK version 4.9 [3] and IBDLD version 3.33 [4] were used to derive 2 empirical kinship matrices based on the GAW20 genotype data. For LDAK, in principle, this kernel should correspond to a genetic relationship matrix; in practice, however, we observed that LDAK estimates of self-relatedness were widely spread around their expectation of 1 (Fig. 1a). For IBDLD the estimates of self-relatedness were closer to 1 (Fig. (Fig.1b).1b). The empirical kinship estimate matrices from LDAK and IBDLD were postprocessed to remove negative nonzero values and scaled to have a diagonal equal to 1.

Expression normalisation

Rank-based int: heavily used in genetics, Although criticised: "Rank-Based Inverse Normal Transformations are Increasingly Used, But are They Merited?"

PEER

Why RANKINT before PEER? "Many statistical tests rely on the assumption that the residuals of a model are normally distributed [1]. In genetic analyses of complex traits, the normality of residuals is largely determined by the normality of the dependent variable (phenotype) due to the very small effect size of individual genetic variants [2]. However, many traits do not follow a normal distribution." "applying rank-based INT to the dependent variable residuals after regressing out covariates re-introduces a linear correlation between the dependent variable and covariates, increasing type-I errors and reducing power."

PEER: expression PCs: if too many, will explain away the signal Not a problem with cis-eQTLs, but trans might have more global effects

why include genetic PCs see stegle 2012 PEER paper: if PCs are not included, they can be recapitulated in the factors

3.2.5 eQTL meta-analysis

Restricted to non-full bayesian methods. For small k, Sidik MVa or Ruhkin RBp recommended. Sidik-Jonkman estimator, also called the 'model error variance estimator', is implemented in metafor (SJ method).

Starts with an init estiamte of ri=sigma2i/tau2i i.e. ratio of study-specific and between-studies het variance, then updates.

They recommend using Hedges [1], to init, but this is bad???

We use mode of gamma as an apriori estiamte of tau.

Joint mapping

mashr smoothing

review: condition/Cell-type specific methods

Simple, mixed models, joint models, multilocus models; Ending with why we chose mashr

mashr beats out stuff it compared to in the paper e.g. metasoft

Definining response eQTLs: Sharing

3.2.6 Colocalization

Due to the increasingly abundant

For example, ran

Coloc and assumptions

Hypercoloc and assumptions

large numbers of traits

Confounding by multiple causal

Fine mapping

3.3 Results

3.3.1 eQTLs at each timepoint

3.3.2 Estimation of sharing

3.3.3 Colocalisation of re-eQTLs with known context-specific immune QTLs

3.3.4 Polygenic score to predict antibody response

3.4 Discussion

Chapter 4

Response to live attenuated rotavirus vaccine (Rotarix) in Vietnamese infants

4.1 Introduction

4.1.1 The genetics of vaccine response in early life

4.1.2 Rotavirus and rotarix in Vietnam

4.1.3 Known factors that affect rotavirus vaccine efficacy

4.2 Methods

4.2.1 RNA-seq data generation

Stranded RNAseq AUTO with Globin Depletion (>47 samples) uses the NEB Ultra II directional RNA library kit for the poly(A) pulldown, fragmentation, 1st and 2nd strand synthesis and the flowing cDNA library prep (with some minor tweaks e.g. at during the PCR we use kapa HiFi not NEB's Q5 polymerase). Between the poly (A) pulldown and the fragmentation we use a kapa globin depletion kit (it's very similar to their riboerase kit but the rRNA probes are swapped for globin ones).

4.2.2 Genotyping

4.3 Results

Transcriptomic response to rotavirus vaccination (pre- vs. post-, prime vs. boost dose, responders vs. non-responders)

Genetic contribution to transcriptomic response

4.4 Discussion

Chapter 5

multiPANTS

5.1 Introduction

5.2 Methods

5.3 Results

5.4 Discussion

Chapter 6

Discussion

Limitations, and the perfect study.

Era of single cell. 1st Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs https://www.nature.com/articles/s41588-018-0089-9

"Single-cell eQTLGen Consortium: a personalized understanding of disease" https://arxiv.org/abs/1909.12550

Optimal design of single-cell RNA sequencing experiments for cell-type-specific eQTL analysis https://www.biorxiv.org/content/biorxiv/early/2019/09/12/766972.full.pdf

Cost-effectiveness and clinical implementation

Appendix A

Supplementary Materials

A.1 Chapter 2

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

A.2 Chapter 3

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

luctus mauris.

A.3 Chapter 4

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

Bibliography

1. Verma, A. & Ritchie, M. D. Current Scope and Challenges in Phenome-Wide Association Studies. *Current Epidemiology Reports* **4,** 321–329. doi:10.1007/s40471-017-0127-7 (Dec. 2017).

List of Abbreviations

QTL quantitative trait locus

Todo list