

<title>

Benjamin Yu Hang Bai

2020-07-30 19:16:50+01:00



# Contents

<b>List of Figures</b>	<b>v</b>
<b>List of Tables</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Structure and diversity of the human genome . . . . .	1
1.2 Genetic association studies for complex traits . . . . .	2
1.2.1 Principles of genetic association . . . . .	2
1.2.2 Lessons from the past 15 years . . . . .	3
1.2.3 From complex trait to locus . . . . .	4
1.2.4 From locus to causal variant . . . . .	5
1.3 Gene expression as an intermediate phenotype . . . . .	6
1.3.1 From causal variant to target gene . . . . .	6
1.3.2 Expression is a complex trait . . . . .	7
1.3.3 Genetic effects on expression: environment is key . . .	8
1.3.4 Response expression quantitative trait loci in the immune system . . . . .	10
1.4 Immune phenotypes are a complex trait . . . . .	12
1.4.1 Response to vaccination is a complex trait . . . . .	13
1.4.2 Response to anti-TNF therapy is a complex trait . . .	14
1.5 Thesis overview . . . . .	15
<b>Bibliography</b>	<b>19</b>
<b>List of Abbreviations</b>	<b>27</b>

*CONTENTS*

---

*CONTENTS*

# List of Figures

1.1	The genomic mosaic: block-like linkage disequilibrium (LD) structure of the genome . . . . .	3
1.2	The reach of GWAS. OR vs MAF ala tam2019BenefitsLimitationsGenomewide, extended by imputation, sample size, WGS based genotypes, but may be indistinguishable from noise at the limits . . . . .	5
1.3	Mediation of genetic effect to phenotype, through the biological system . . . . .	9
1.4	eqtl mech models: magnify, dampen, flip . . . . .	10



# List of Tables



# Chapter 1

## Introduction

### 1.1 Structure and diversity of the human genome

- The human genome is almost three billion **base pairs (bps)** in length, containing 20000-25000 protein-coding genes [1, 2] that span 1-3% of its length, with the remainder being non-coding. Each diploid human cell contains two copies of the genome; 46 chromosomes comprised of 23 maternal-parental pairs: 22 pairs of homologous autosomes and one pair of sex chromosomes.
- Variation in the genome between individuals in a population exists in the form of **single nucleotide polymorphisms (SNPs)**, short indels, and structural variants—the vast majority of common variants (**MAF** > 1 – 5%) are **SNPs** and short indels (> 99.9%) [2]. On average, a pair of human genomes differs by one **SNP** per 1000-2000 **bp** [3]. Each version of a variant is called an allele; an individual has a maternal and parental allele at each variant.
- The many variants in a population are inherited in a smaller number of haplotypes: contiguous stretches of the genome passed through generations via meiotic segregation. The fundamental sources of genetic diversity are mutation and meiotic recombination, generating new alleles and breaking apart haplotypes into shorter ones over time. Variants at locations on a chromosome (loci) that are physically close are less likely to flank a recombination event, hence more likely to cosegregate on the same haplotype, referred to as genetic linkage.

consider moving awkward  
defs to margin notes, in  
the style of nature re-  
views

LD decay just takes a  
really really long time,  
but there are evo forces  
at work too that maintain  
LD

## 1.2. GENETIC ASSOCIATION STUDIES

Genetic linkage is one source of **linkage disequilibrium (LD)**: the non-random association of alleles at two loci, differing from expectation based on their frequencies and the law of independent assortment [4]. LD is often quantified within a population by  $r^2$ , the squared correlation coefficient between alleles [4].

Recombination events are not distributed uniformly throughout the genome. The genome is a mosaic of blocks delimited by recombination hotspots, characterised by strong LD within blocks, and little LD between blocks [5, 6] (Fig. 1.1). The structure of correlated haplotypes reflects a population's unique evolutionary history, and can be used to trace the demography of human populations back through time [7].

Heard it's good for the reader's attention span to have figures in intro. Unless it's ok to use figures from papers, I only want to spend the time making the min that are necessary though.

can i use published figures?

add something sweeping about utility here or elsewhere: e.g. insights into trait biology and clinical translational potential for disease traits, genetically support drug target identification

### 1.2 Genetic association studies for complex traits

#### 1.2.1 Principles of genetic association

- Variation in human traits arises from an interplay between genetics and environment. Traits for which genetic variation explains a non-zero fraction of phenotypic variation are heritable. Many measurable human traits are heritable and twin studies provide upper bounds on this heritability <https://www.nature.com/articles/ng.3285>. Discovering the specific genetic variants that contribute to heritability, through association of variants and phenotypes measured from the same individual, is a mainstay of the field of human genetics. Barring somatic mutation, an individual's genome is fixed at conception, providing a causally upstream anchor. Genetic association studies have intrinsic resistance to some back-door path effects such as reverse causality, which permeate observational studies of the causes of human phenotypes.
- Under the central dogma, information flows from gene to RNA to protein to phenotype via transcription and translation, thus it is assumed that genetic variants at loci in the genome affect phenotype by impacting on the function or regulation of target genes. How genetic variation contributes to any heritable trait defines its genetic architecture: the number of genes affecting that trait; along with the allele frequencies, effect sizes, and interactions of trait-associated variants

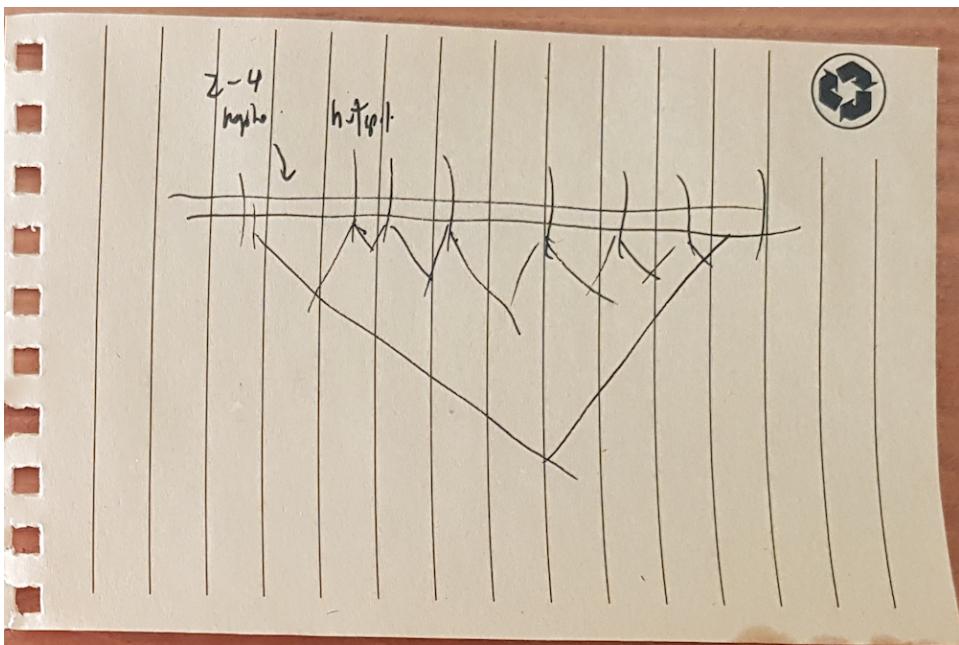


Figure 1.1: The genomic mosaic: block-like LD structure of the genome

[8]. The number of genes defines a spectrum of traits from monogenic (where inheritance follows simple Mendelian patterns) to polygenic (where inheritance is complex). Many architectures have been proposed for complex traits; all have in common that the number of genes that affect a complex trait is large (ranging from dozens to many thousands), thus the average effect of each trait-associated loci is small [9, 10] <https://www.pnas.org/content/106/23/9362>.

### 1.2.2 Lessons from the past 15 years

- For decades, linkage analysis had been successfully applied to map loci affecting Mendelian traits by tracing their cosegregation with the trait through pedigrees [11]. Small-scale genetic association studies were also performed, focusing on variants in or near candidate genes selected on the basis of prior biological knowledge [12]. These approaches were not successful for complex traits, as small effect sizes lead to low penetrance in pedigrees [11] and poor power at the sample sizes typically used in early candidate gene studies [13].
- Genome-wide association studies (GWAS) systematically test common

## 1.2. GENETIC ASSOCIATION STUDIES AND FROM TRAIT TO Locus

variants selected in a comparatively hypothesis-free manner across the genome for association with a trait (Fig. 1.2). Using large sample sizes to overcome small effects and large multiple testing burden, thousands of associations have been discovered for complex traits and disease, many robustly replicated across populations [11, 14]. Most genetic variance is explained by additive effects, the contribution of epistatic interactions is small [8], and pleiotropy is widespread [11]. Sample sizes in the millions are increasingly commonplace, and discovery of new associations with increasing sample size shows no sign of plateauing [15]. It is now appreciated that most heritable organism-level phenotypes are complex, and have remarkable polygenicity, with many hundreds or thousands of associated loci.

- In general, the more organism level a phenotype, the more polygenic, but even molecular traits are very polygenic

### **1.2.3 From complex trait to locus**

**GWAS** rely on the tendency of common variants on the same haplotype to be in strong **LD**. As the number of haplotypes is comparatively few, it is possible to select a subset of tag variants such that all other known common variants are within a certain **LD** threshold of that subset. In practice, there is enough redundancy that the number of variants measured on a modern genotyping array (in the order of  $10^5$  to  $10^6$ ) is sufficient to tag almost all common variants [16, 17]. Associations with unmeasured variants are indirectly detected through their strong correlation with a tag variant. Furthermore, as unrelated individuals still share short ancestral haplotypes, study samples can be assigned haplotypes from a panel of haplotypes derived from reference samples by matching on the directly genotyped variants. This process of genotype imputation allows ascertainment of many more variants not directly genotyped [18], but helps to recover rarer variants that are poorly-tagged [14]. Modern imputation panels enable cost-effective **GWAS** including tens of millions of variants down to frequencies of  $\sim 0.01\%$  <https://www.biorxiv.org/content/10.1101/563866v1>.

seems like there is some connection to be made between the tagability of common variation and the feasibility of imputation both being enabled by the relatively small number of common haplotypes compared to variants

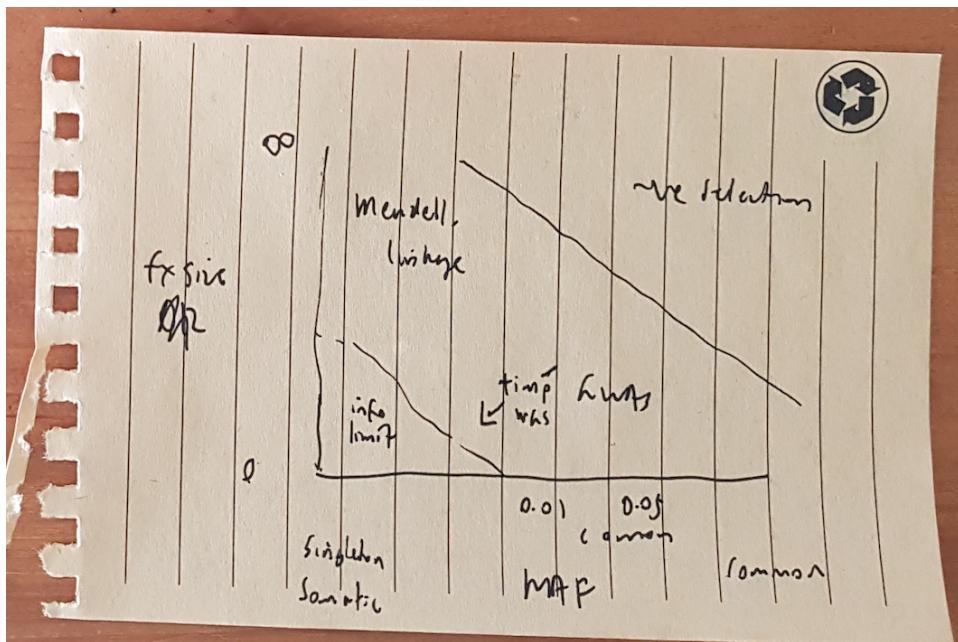


Figure 1.2: The reach of GWAS. OR vs MAF ala tam2019BenefitsLimitationsGenomewide, extended by imputation, sample size, WGS based genotypes, but may be indistinguishable from noise at the limits

Testing such large numbers of variants incurs a massive multiple testing burden, but acknowledging the correlation between variants due to **LD**, there are only the equivalent of  $\sim 10^6$  independent tests in the European genome, regardless of the number of tests actually performed [19]. The field has thus converged on a fixed discovery threshold of  $0.05/10^6 = 5 \times 10^{-8}$  for genome-wide significance in European populations [20], akin\* to controlling the family-wise type I error rate at using the Bonferroni correction.

#### 1.2.4 From locus to causal variant

- By design, a significantly-associated variant from a **GWAS** needs not be a variant that causally affects the trait, and may only tag a causal variant.
  - Fine-mapping is the process of determining which of the many

\*The Bonferroni procedure makes no assumptions about the dependence structure of the  $p$ -values, and is conservative (i.e. controls the **family-wise error rate (FWER)** at a stricter level than the chosen  $\alpha$ ) even for independent tests. In fact it is always conservative unless the  $p$ -values have strong negative correlations [21].

### 1.3. GENE EXPRESSION AS AN INTERMEDIATE PHENOTYPE

correlated variants at a **GWAS** locus are causal.

- State-of-the-art methods (e.g. PAINTOR, CAVIARBF, FINEMAP <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6050137/>, SuSiE) provide Bayesian posterior probabilities that associated variants are causal, and some methods can consider the presence of multiple causal variants at the same locus [22].
- Even if a single causal variant cannot be assigned, a credible set can.
- Power: to separate causal and tag variants depends on **LD** and sample size [14]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6050137/>
- Resolution: Naturally, these methods assign probabilities assuming the causal variant is in the set of variants observed.
- The causal variant must either be genotyped or confidently imputed. Denser genotyping e.g. by WGS, and larger imputation panels will help.

## 1.3 Gene expression as an intermediate phenotype

### 1.3.1 From causal variant to target gene

- For coding variants, there is a reasonable prior as to the target gene.
- Unlike for Mendelian traits where most causal variants are coding <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4573249/>, over 90% of **GWAS** loci fall in non-coding regions of the genome [23], and often too far from the nearest gene to be in **LD** <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5291268/>. Thus even if the causal variant at a locus is fine-mapped, it may not be obvious how to find the target genes through which that variant affects the trait.
- Rather than directly impacting the coding sequence of a gene, many non-coding GWAS loci are thought to affect traits by affecting the regulation of target gene expression [23]. **GWAS** loci are enriched in regulatory

elements annotated by functional genomics studies, such as regions of open chromatin, DNase I hypersensitive sites, splice sites, UTRs, histone binding sites, **transcription factor (TF)** binding motifs, and enhancers [23, 24] <https://genome.cshlp.org/content/22/9/1748.full>.

- For complex diseases, enrichment is observed in disease-relevant tissues [14].
- These enrichments put forth expression as an important molecular phenotype linking non-coding **GWAS** variants to their associated traits (Fig. 1.3).

### 1.3.2 Expression is a complex trait

- Studies of the genetic architecture of expression have further reinforced this hypothesis.
  - Molecular phenotypes like expression are heritable complex traits [25]
  - Expression can be assayed by e.g. array or RNAseq
  - The variants associated with expression are called **expression quantitative trait loci (eQTLs)**.
  - eQTLs can also be *cis*- or *trans*- to their target gene [26].
  - Their effect size declines with distance to the TSS, so the most readily detectable eQTLs are *cis*, and within 1Mb [27]
- GWAS variants are enriched for eQTLs <https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1000888>
  - So GWAS variants that are also eQTL naturally prioritise target genes.
  - Is it a narrow view to assume that the effect of GWAS loci on complex traits not only act through a target gene, but are specifically mediated by eQTL effects?
  - Over many complex traits, a median of 11% heritability could be explained by mediation of GWAS loci by common ( $MAF > 0.01$ ) *cis*-eQTL, and this proportion does not include *trans* or post-transcriptional effects.

### 1.3. GENE EXPRESSION AS AN INTERMEDIATE PHENOTYPE

- With increasing sample size, most genes (60-80%) have a detectable eQTL [27]. Assuming that a locus on the genome is associated with both a complex trait and an eQTL, how can we separate the scenario where one variant affects both trait and expression (pleiotropy), from coincidental overlap between distinct causal variants that may possibly in LD? Bayesian probabilistic colocalisation methods (e.g. eCAVIAR, Sherlock, coloc [28]) address this by estimating the posterior probability that the same causal variant is associated with both phenotypes. distinguishing pleiotropy from linkage, but not vertical pleiotropy (mediation) from horizontal pleiotropy (independent effects on trait and expression) [29]. As colocalisation of a GWAS loci with eQTLs is necessary but not sufficient for mediation, it should be supported by complementary lines of evidence from other methods that integrate intermediate phenotypes (TWAS, MR, mediation analysis etc.) [29] to help untangle the multiplex of possible causal pathways from variant to trait.

add uses other vars

#### 1.3.3 Genetic effects on expression: environment is key

- The effects of eQTLs (and molecular quantitative trait loci (QTLs) in general) are incredibly context-dependent [26, 27].
  - This represents genotype-environment interactions at those eQTL.
  - A non-exhaustive list of environments that eQTLs have been found to interact with:
    - \* sex, age <https://academic.oup.com/hmg/article/23/7/1947/655184>
    - \* ancestry [30–32]
    - \* tissue [33, 34]
    - \* cell type composition in bulk samples [35–38]
    - \* individual cell type [30, 38–41]
    - \* disease status [40],
    - \* and experimental stimulation (see subsection 1.3.4).
- Given the effect of an eQTL can be starkly different between environments, it is difficult to determine the appropriate eQTL dataset to use for target gene prioritisation at GWAS loci.

## CHAPTER 1: GENOME EXPRESSION AS AN INTERMEDIATE PHENOTYPE

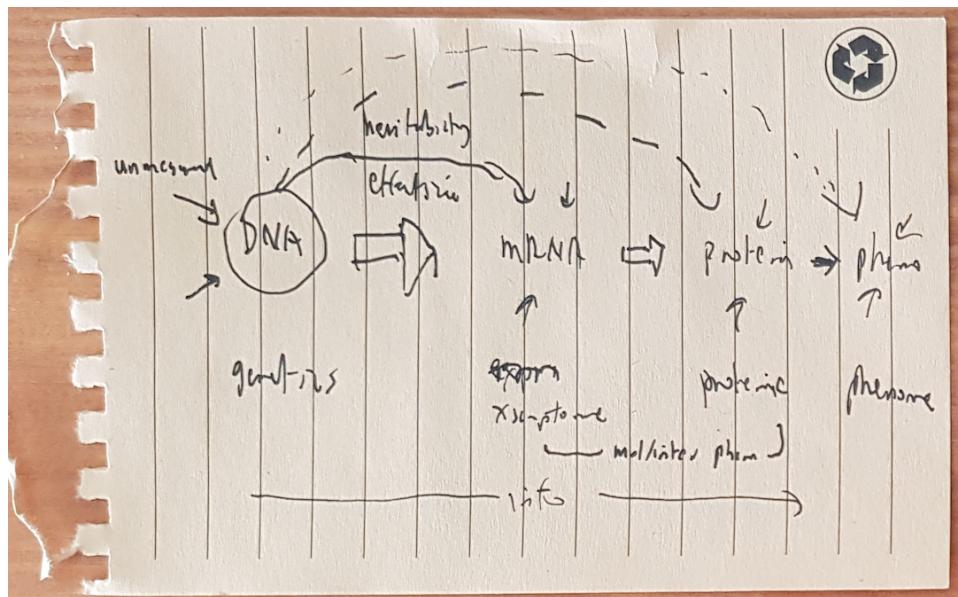


Figure 1.3: Mediation of genetic effect to phenotype, through the biological system

- It has already been shown that use of cell-type specific eQTLs increases coloc rates with GWAS hits [38] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4498151/> <https://www.biorxiv.org/content/10.1101/2020.01.15.907436v1>
- Successful colocalisation of GWAS loci with coloc may prioritise not only the target gene, but the specific environments most relevant to a trait.
- What molecular mechanisms might facilitate genotype-environment interactions at **eQTLs**?
  - [42]: defines static, conditional, dynamic eQTLs
  - Fu *et al.* [43]: proposes TF-based mechanisms for cis-eQTL (here, define mag, damp, flip) (**Fig. 1.4**)
  - Gaffney [25] and Rotival [44]: suggests info on more regulatory layers will help break down transcriptional and post-transcriptional
    - \* also, priming

### 1.3. GENE EXPRESSION AS AN INTERMEDIATE PHENOTYPE

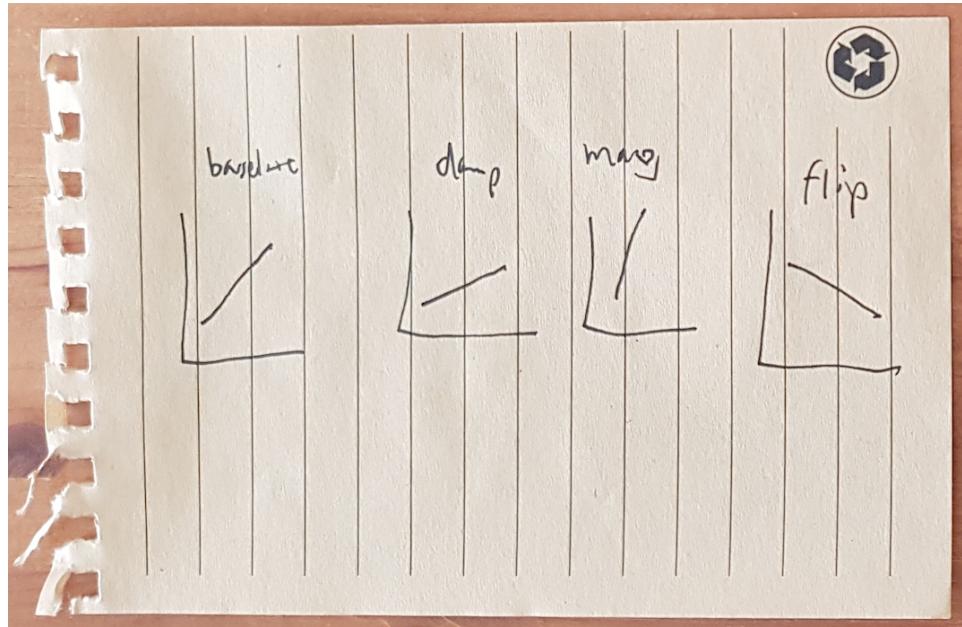


Figure 1.4: eqtl mech models: magnify, dampen, flip

#### **1.3.4 Response expression quantitative trait loci in the immune system**

- An important subclass of context-dependent eQTL are **response expression quantitative trait locus (reQTL)**, where the interacting environment is experimental stimulation [27, 45]. Most reQTL studies to date have been conducted on immune cells *in vitro*, not only because the immune system is specialised for responding to environmental exposures, but due to the abundance of immune cells easily accessible in peripheral blood, and amenable to separation (e.g. FACS) and stimulation.
  - *In vitro*, potential interacting variables such as cell type, and the nature, length, and intensity of stimulation can be precisely controlled.
- A seminal early study was conducted by [46], where eQTLs were mapped separately in monocyte-derived dendritic cells before and after 18h infection with *Mycobacterium tuberculosis*.
  - reQTLs were detected for 198 genes, 102 specific to the uninfected state, and 96 specific to the infected state.

- Since then, *in vitro* immune reQTL studies have been conducted for a variety of cell types (e.g. primary CD14+ monocytes [47]) and stimulations (IFN $\gamma$  and LPS [47]).
- A complementary approach is *in vivo* reQTL mapping
  - There are pros to *in vivo* stimulation.
    - \* the innumerable interactions in the immune system that are absent *in vitro*
    - \* ability to get whole organism phenotypes
    - \* ability to get repeated measures: can reason about change in expression over time
  - Major disadvantages: the choice of stimulation must be ethical *in vivo*, and many environmental factors (e.g. diet, lifestyle, immune exposures) cannot be controlled, leading to greater experimental noise (?), and more complex interpretations.
  - There are few published *in vivo* reQTL studies.
    - \* [49]: seasonal trivalent inactivated influenza vaccine (TIV), whole blood, antigen processing and intracellular trafficking genes, attempted mediation for Ab titres, but concluded they were underpowered
    - \* [50]: fold-change expression after inactivated vaccinia vaccine, focus was on pairwise epistatic interactions, apoptosis pathways
    - \* [51]: whole blood, IFN status and anti-IL6 drug exposure, reQTL driven by ISRE and IRF4 motifs
- <why care about immune reQTLs>
  - Exposes differences in regulatory architecture between conditions, but does not automatically reveal the mechanisms behind those differences
  - Immune *in vitro* reQTL have been shown to be enriched more so than non-reQTL among GWAS loci for immune-related phenotypes such as susceptibility to infectious [46, 52] and immune-mediated diseases [52, 53].

list a few more types and  
stims from [47] until [48]

## 1.4. IMMUNE PHENOTYPES ARE A ~~CHAPTER~~ INTRODUCTION

- Not yet clear whether *in vivo* reQTL have any utility on top of *in vitro* reQTL for interpreting GWAS loci: not that many studies, and complex interpretations.
- Nevertheless, as the number of cell types systems and stimulations both *in vitro* and *in vivo* increases, the number of known reQTLs continues to grow.

### 1.4 Immune phenotypes are a complex trait

- Heritability of immune phenotypes is not only restricted to the expression phenotypes discussed above.

not sure if right order.  
Since most reQTL studies are immune, I went context-specific -> reQTL  
-> immune rather than context-specific -> immune -> reQTL

- Studies of interindividual variation in the healthy immune system shows many aspects of the immune system are heritable and complex.

- Immune parameters are influenced by age, sex, seasonality, and chronic infection [54–58] <https://www.nature.com/articles/ncomms8000>, but most individuals have a healthy baseline immune state that is individual-specific, and relatively stable over time [55, 56, 59].

stable, yet varies by age?  
respecify scale of stability

- Overall estimates of the heritability of many immune parameters, such as cell composition and serum protein levels, lies between 20-40% [55–58]
- Genetic regulation is more important for the innate immune system than the adaptive immune system [57].

- A central goal of systems immunology is to establish causal relationships between the many components of the immune system

- Natural genetic variation represents small scale perturbation that is causally anchored [60, 61]
- But as discussed in the context section above, specific effects may not be apparent in the baseline state, stimulation is required
- Studies of natural infection are complicated by e.g. determining exposure.

## CHAPTER 1.1 INTRODUCTION: HOW GENOTYPES ARE A COMPLEX TRAIT

- As in the immune in vivo reQTL studies, vaccines and drugs used as controlled immune perturbations to study the activated immune system
  - \* reQTL may have utility for interpretation of these immune-related complex traits too, not just IMIDs/infectious disease discussed in reQTL section above

### **1.4.1 Response to vaccination is a complex trait**

- Vaccination has enormous impact on global health [62]
  - <quick vaccine bio, specific flu vaccine goes in ch2>
    - \* 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.
    - \* <...>
  - A vaccine that is highly efficacious in one human population may have significantly lower efficacy in other populations. Particularly challenging populations for vaccination include the infants and elderly, pregnant, immuno compromised patients, ethnically-diverse populations, and developing countries.
    - \* <1 example statistic on vaccine efficacy differences e.g. rotavirus>
    - \* e.g. <https://www.sciencedirect.com/science/article/pii/S1473309918304900>
  - Traditional vaccine dev is empirical (classical "isolate, inactivate, inject" paradigm), often successful vaccine dev does not offer insights into the mechanisms of efficacy
  - The immunological mechanisms that underpin a specific vaccine's success or failure in a given individual are often poorly understood.
- A sub-discipline of systems immunology is systems vaccinology.

## 1.4. IMMUNE PHENOTYPES ARE A ~~CHAPTER~~ INTRODUCTION

- Systems vaccinology is the application of -omics technologies to provide a systems-level characterisation of the human immune system after vaccine-perturbation.
- 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.
  - \* <add more to list of what vaccines have been studied, pull out of sysvacc\_review\_docx>
- Sysvacc informs more mechanism-based and cost-effective design (rational paradigm), and the move towards personalised vaccinology.
- Sysvacc has revealed many influences on vaccine response (age, sex, dose, adjuvants, expression signatures, microbiome, strain etc.)
- Studies of impact of host genetics is underrepresented [63]

- Like for other complex traits, from twin studies it's known that vaccine Ab responses are heritable.

- Moving out of the candidate gene era (e.g. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3570049/>) into GWAS.
- [64] has heritability estimates
- Many loci have been implicated by GWAS e.g. HLA [63–68]

Overall, systems vacc studies that include genetics are nowhere near as mature compared to the trait to gene pipeline described in above e.g. for immune-mediated disease

### **1.4.2 Response to anti-TNF therapy is a complex trait**

not sure about scope  
of the subsection, cur-  
rently some overlap with  
PANTS chapter intro.

- <quick anti-tnf summary, specific ADA/IFX biology goes in ch4>
  - anti-TNFs (or TNF inhibitors), are drugs that suppress the activity of the TNF signalling pathway of the immune system

- they are used to treat immune-mediated inflammatory diseases  
e.g. RA, IBD ...
- response has many definitions
  - \* subphenotypes that may be complex traits: immunogenicity, primary response, loss of response, remission rate, adverse fx
- <expression signatures of response to anti-TNFs>
  - have been detected e.g. for RA "Validation study of existing gene expression signatures for anti-TNF treatment in patients with rheumatoid arthritis" <https://pubmed.ncbi.nlm.nih.gov/22457743/>
  - also done for IBD (described in ch4)
  - most detected in small cohorts, requires validation
- <genetics of anti-TNF response>
  - pharmacogenomics is the study of the role of genetics in beneficial and adverse effects of drugs and therapeutics [https://doi.org/10.1016/S0140-6736\(19\)31276-0](https://doi.org/10.1016/S0140-6736(19)31276-0)
  - some implementation in clinic already e.g. screening for certain allele-drug combos <https://www.nature.com/articles/nature15817> <https://academic.oup.com/bmb/article/124/1/65/4430783>
  - GWAS in the pharmacogenomics field <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3003940/> <https://www.futuremedicine.com/doi/full/10.2217/pgs-2018-0204>
  - GWAS studies of anti-TNF response in RA <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6614444/>
    - \* a few validation studies attempted e.g. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5937760/>
  - also done for IBD (described in ch4)

## 1.5 Thesis overview

- <By chapter context-content-conclusion overview.>

- <ch 2: systems vaccinology study of Pandemrix>
  - \* context: existing Sobolev study of expression differences between pandemic flu vaccine R/NR had small sample size and binary phenotype
  - \* content: meta-analysis of existing array with new RNAseq data and continuous phenotype
  - \* conclusion: distinct innate and adaptive expression response at d1 and d7; heterogeneity between array and RNAseq. significant expression differences between R/NR in meta-analysis at the gene set level
- <ch 3: in vivo reQTL study of Pandemrix>
  - \* context: relatively few studies have assessed the impact of genetic variation on expression response to flu vaccine
  - \* content: reQTL analysis for flu vaccine at d0, d1, d7. many reQTLs including sign flips. no particular gene set enrichments. evidence of cell type interactions at top hits.
  - \* conclusion: difficult to separate out modifying effect of cell composition. this may be a fundamental flaw in the study design
- <ch 4: systems immunology and reQTL study of response to anti-TNF treatment in CD>
  - \* context: studies on expression signatures of anti-TNF PNR have been small
  - \* content: R/NR comparison with larger n, at baseline, w14, and over time. reQTL analysis over 4 timepoints.
  - \* conclusion: a few hits for PNR at baseline. much stronger expression differences stronger at w14, then maintained until w54. Weak evidence for reQTLs, probably due to smaller magnitude of cell proportion changes over time vs the previous chapter.
- <discussion: limitations, future outlook>
  - \* main themes and parallels tying together the thesis
  - \* shared set of limitations permeating all chapters
  - \* recommendations for future analyses and study design

- \* future outlook for the fields of vaccinogenomics and pharmacogenomics



# Bibliography

1. The ENCODE Project Consortium. An Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature* **489**, 57–74. doi:[10.1038/nature11247](https://doi.org/10.1038/nature11247) (2012).
2. 1000 Genomes Project Consortium *et al.* A Global Reference for Human Genetic Variation. *Nature* **526**, 68–74. doi:[10.1038/nature15393](https://doi.org/10.1038/nature15393) (2015).
3. The International SNP Map Working Group. A Map of Human Genome Sequence Variation Containing 1.42 Million Single Nucleotide Polymorphisms. *Nature* **409**, 928–933. doi:[10.1038/35057149](https://doi.org/10.1038/35057149) (2001).
4. Slatkin, M. Linkage Disequilibrium — Understanding the Evolutionary Past and Mapping the Medical Future. *Nature Reviews Genetics* **9**, 477–485. doi:[10.1038/nrg2361](https://doi.org/10.1038/nrg2361) (2008).
5. Wall, J. D. & Pritchard, J. K. Haplotype Blocks and Linkage Disequilibrium in the Human Genome. *Nature Reviews Genetics* **4**, 587–597. doi:[10.1038/nrg1123](https://doi.org/10.1038/nrg1123) (2003).
6. The International HapMap Consortium. A Second Generation Human Haplotype Map of over 3.1 Million SNPs. *Nature* **449**, 851–861. doi:[10.1038/nature06258](https://doi.org/10.1038/nature06258) (2007).
7. Karczewski, K. J. & Martin, A. R. Analytic and Translational Genetics. *Annual Review of Biomedical Data Science* **3**. doi:[10.1146/annurev-biodatasci-072018-021148](https://doi.org/10.1146/annurev-biodatasci-072018-021148) (2020).
8. Visscher, P. M. & Goddard, M. E. From R.A. Fisher's 1918 Paper to GWAS a Century Later. *Genetics* **211**, 1125–1130. doi:[10.1534/genetics.118.301594](https://doi.org/10.1534/genetics.118.301594) (2019).
9. Gibson, G. Rare and Common Variants: Twenty Arguments. *Nature reviews. Genetics* **13**, 135–145. doi:[10.1038/nrg3118](https://doi.org/10.1038/nrg3118) (2011).

---

**APPENDIX . BIBLIOGRAPHY**

---

10. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnipathogenic. *Cell* **169**, 1177–1186. doi:[10.1016/j.cell.2017.05.038](https://doi.org/10.1016/j.cell.2017.05.038) (2017).
11. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five Years of GWAS Discovery. *The American Journal of Human Genetics* **90**, 7–24. doi:[10.1016/j.ajhg.2011.11.029](https://doi.org/10.1016/j.ajhg.2011.11.029) (2012).
12. Hirschhorn, J. N., Lohmueller, K., Byrne, E. & Hirschhorn, K. A Comprehensive Review of Genetic Association Studies. *Genetics in Medicine* **4**, 45–61. doi:[10.1097/00125817-200203000-00002](https://doi.org/10.1097/00125817-200203000-00002) (2002).
13. Border, R. *et al.* No Support for Historical Candidate Gene or Candidate Gene-by-Interaction Hypotheses for Major Depression Across Multiple Large Samples. *American Journal of Psychiatry* **176**, 376–387. doi:[10.1176/appi.ajp.2018.18070881](https://doi.org/10.1176/appi.ajp.2018.18070881) (2019).
14. Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *The American Journal of Human Genetics* **101**, 5–22. doi:[10.1016/j.ajhg.2017.06.005](https://doi.org/10.1016/j.ajhg.2017.06.005) (2017).
15. Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G. & Meyre, D. Benefits and Limitations of Genome-Wide Association Studies. *Nature Reviews Genetics*. doi:[10.1038/s41576-019-0127-1](https://doi.org/10.1038/s41576-019-0127-1) (2019).
16. The International HapMap Consortium. A Haplotype Map of the Human Genome. *Nature* **437**, 1299–1320. doi:[10.1038/nature04226](https://doi.org/10.1038/nature04226) (2005).
17. Barrett, J. C. & Cardon, L. R. Evaluating Coverage of Genome-Wide Association Studies. *Nature Genetics* **38**, 659–662. doi:[10.1038/ng1801](https://doi.org/10.1038/ng1801) (2006).
18. Das, S., Abecasis, G. R. & Browning, B. L. Genotype Imputation from Large Reference Panels. *Annual Review of Genomics and Human Genetics* **19**, 73–96. doi:[10.1146/annurev-genom-083117-021602](https://doi.org/10.1146/annurev-genom-083117-021602) (2018).
19. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M. J. Estimation of the Multiple Testing Burden for Genomewide Association Studies of Nearly All Common Variants. *Genetic Epidemiology* **32**, 381–385. doi:[10.1002/gepi.20303](https://doi.org/10.1002/gepi.20303) (2008).

## APPENDIX . BIBLIOGRAPHY

---

20. Jannet, A.-S., Ehret, G. & Perneger, T.  $P < 5 \times 10^{-8}$  Has Emerged as a Standard of Statistical Significance for Genome-Wide Association Studies. *Journal of Clinical Epidemiology* **68**, 460–465. doi:[10.1016/j.jclinepi.2015.01.001](https://doi.org/10.1016/j.jclinepi.2015.01.001) (2015).
21. Goeman, J. J. & Solari, A. Multiple Hypothesis Testing in Genomics. *Statistics in Medicine* **33**, 1946–1978. doi:[10.1002/sim.6082](https://doi.org/10.1002/sim.6082) (2014).
22. Schaid, D. J., Chen, W. & Larson, N. B. From Genome-Wide Associations to Candidate Causal Variants by Statistical Fine-Mapping. *Nature Reviews Genetics* **19**, 491–504. doi:[10.1038/s41576-018-0016-z](https://doi.org/10.1038/s41576-018-0016-z) (2018).
23. Gallagher, M. D. & Chen-Plotkin, A. S. The Post-GWAS Era: From Association to Function. *The American Journal of Human Genetics* **102**, 717–730. doi:[10.1016/j.ajhg.2018.04.002](https://doi.org/10.1016/j.ajhg.2018.04.002) (2018).
24. Trynka, G. *et al.* Disentangling the Effects of Colocalizing Genomic Annotations to Functionally Prioritize Non-Coding Variants within Complex-Trait Loci. *The American Journal of Human Genetics* **97**, 139–152. doi:[10.1016/j.ajhg.2015.05.016](https://doi.org/10.1016/j.ajhg.2015.05.016) (2015).
25. Gaffney, D. J. Global Properties and Functional Complexity of Human Gene Regulatory Variation. *PLoS Genetics* **9** (ed Abecasis, G. R.) e1003501. doi:[10.1371/journal.pgen.1003501](https://doi.org/10.1371/journal.pgen.1003501) (2013).
26. Albert, F. W. & Kruglyak, L. The Role of Regulatory Variation in Complex Traits and Disease. *Nature Reviews Genetics* **16**, 197–212. doi:[10.1038/nrg3891](https://doi.org/10.1038/nrg3891) (2015).
27. Vandiedonck, C. Genetic Association of Molecular Traits: A Help to Identify Causative Variants in Complex Diseases. *Clinical Genetics*. doi:[10.1111/cge.13187](https://doi.org/10.1111/cge.13187) (2017).
28. Wallace, C. Eliciting Priors and Relaxing the Single Causal Variant Assumption in Colocalisation Analyses. *PLOS Genetics* **16** (ed Epstein, M. P.) e1008720. doi:[10.1371/journal.pgen.1008720](https://doi.org/10.1371/journal.pgen.1008720) (2020).
29. Hemani, G., Bowden, J. & Davey Smith, G. Evaluating the Potential Role of Pleiotropy in Mendelian Randomization Studies. *Human Molecular Genetics* **27**, R195–R208. doi:[10.1093/hmg/ddy163](https://doi.org/10.1093/hmg/ddy163) (2018).

---

**APPENDIX . BIBLIOGRAPHY**

---

30. De Jager, P. L., Hacohen, N., Mathis, D., Regev, A., Stranger, B. E. & Benoist, C. ImmVar Project: Insights and Design Considerations for Future Studies of “Healthy” Immune Variation. *Seminars in Immunology* **27**, 51–57. doi:[10.1016/j.smim.2015.03.003](https://doi.org/10.1016/j.smim.2015.03.003) (2015).
31. Nédélec, Y. *et al.* Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell* **167**, 657–669.e21. doi:[10.1016/j.cell.2016.09.025](https://doi.org/10.1016/j.cell.2016.09.025) (2016).
32. Quach, H. & Quintana-Murci, L. Living in an Adaptive World: Genomic Dissection of the Genus Homo and Its Immune Response. *Journal of Experimental Medicine* **214**, 877–894. doi:[10.1084/jem.20161942](https://doi.org/10.1084/jem.20161942) (2017).
33. Nica, A. C. *et al.* The Architecture of Gene Regulatory Variation across Multiple Human Tissues: The MuTHER Study. *PLoS Genetics* **7** (ed Barsh, G.) e1002003. doi:[10.1371/journal.pgen.1002003](https://doi.org/10.1371/journal.pgen.1002003) (2011).
34. Aguet, F. *et al.* Genetic Effects on Gene Expression across Human Tissues. *Nature* **550**, 204–213. doi:[10.1038/nature24277](https://doi.org/10.1038/nature24277) (2017).
35. Westra, H.-J. *et al.* Cell Specific eQTL Analysis without Sorting Cells. *PLOS Genetics* **11** (ed Pastinen, T.) e1005223. doi:[10.1371/journal.pgen.1005223](https://doi.org/10.1371/journal.pgen.1005223) (2015).
36. Zhernakova, D. V. *et al.* Identification of Context-Dependent Expression Quantitative Trait Loci in Whole Blood. *Nature Genetics* **49**, 139–145. doi:[10.1038/ng.3737](https://doi.org/10.1038/ng.3737) (2017).
37. Glastonbury, C. A., Couto Alves, A., El-Sayed Moustafa, J. S. & Small, K. S. Cell-Type Heterogeneity in Adipose Tissue Is Associated with Complex Traits and Reveals Disease-Relevant Cell-Specific eQTLs. *The American Journal of Human Genetics* **104**, 1013–1024. doi:[10.1016/j.ajhg.2019.03.025](https://doi.org/10.1016/j.ajhg.2019.03.025) (2019).
38. Kim-Hellmuth, S. *et al.* Cell Type Specific Genetic Regulation of Gene Expression across Human Tissues. *bioRxiv*. doi:[10.1101/806117](https://doi.org/10.1101/806117) (2019).
39. Dimas, A. S. *et al.* Common Regulatory Variation Impacts Gene Expression in a Cell Type-Dependent Manner. *Science* **325**, 1246–1250. doi:[10.1126/science.1174148](https://doi.org/10.1126/science.1174148) (2009).

## APPENDIX . BIBLIOGRAPHY

---

40. Peters, J. E. *et al.* Insight into Genotype-Phenotype Associations through eQTL Mapping in Multiple Cell Types in Health and Immune-Mediated Disease. *PLOS Genetics* **12** (ed Plagnol, V.) e1005908. doi:[10.1371/journal.pgen.1005908](https://doi.org/10.1371/journal.pgen.1005908) (2016).
41. Chen, L. *et al.* Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell* **167**, 1398–1414.e24. doi:[10.1016/j.cell.2016.10.026](https://doi.org/10.1016/j.cell.2016.10.026) (2016).
42. Ackermann, M., Sikora-Wohlfeld, W. & Beyer, A. Impact of Natural Genetic Variation on Gene Expression Dynamics. *PLoS Genetics* **9** (ed Wells, C. A.) e1003514. doi:[10.1371/journal.pgen.1003514](https://doi.org/10.1371/journal.pgen.1003514) (2013).
43. Fu, J. *et al.* Unraveling the Regulatory Mechanisms Underlying Tissue-Dependent Genetic Variation of Gene Expression. *PLoS Genetics* **8** (ed Gibson, G.) e1002431. doi:[10.1371/journal.pgen.1002431](https://doi.org/10.1371/journal.pgen.1002431) (2012).
44. Rotival, M. Characterising the Genetic Basis of Immune Response Variation to Identify Causal Mechanisms Underlying Disease Susceptibility. *HLA* **94**, 275–284. doi:[10.1111/tan.13598](https://doi.org/10.1111/tan.13598) (2019).
45. Huang, Q. *The Genetics of Gene Expression: From Simulations to the Early-Life Origins of Immune Diseases* (2019).
46. Barreiro, L. B., Tailleux, L., Pai, A. A., Gicquel, B., Marioni, J. C. & Gilad, Y. Deciphering the Genetic Architecture of Variation in the Immune Response to Mycobacterium Tuberculosis Infection. *Proceedings of the National Academy of Sciences* **109**, 1204–1209. doi:[10.1073/pnas.11115761109](https://doi.org/10.1073/pnas.11115761109) (2012).
47. Fairfax, B. P. *et al.* Innate Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene Expression. *Science* **343**, 1246949–1246949. doi:[10.1126/science.1246949](https://doi.org/10.1126/science.1246949) (2014).
48. Alasoo, K., Rodrigues, J., Danesh, J., Freitag, D. F., Paul, D. S. & Gaffney, D. J. Genetic Effects on Promoter Usage Are Highly Context-Specific and Contribute to Complex Traits. *eLife* **8**. doi:[10.7554/eLife.41673](https://doi.org/10.7554/eLife.41673) (2019).
49. Franco, L. M. *et al.* Integrative Genomic Analysis of the Human Immune Response to Influenza Vaccination. *eLife* **2**, e00299. doi:[10.7554/eLife.00299](https://doi.org/10.7554/eLife.00299) (2013).

---

**APPENDIX . BIBLIOGRAPHY**

50. Lareau, C. A., White, B. C., Oberg, A. L., Kennedy, R. B., Poland, G. A. & McKinney, B. A. An Interaction Quantitative Trait Loci Tool Implicates Epistatic Functional Variants in an Apoptosis Pathway in Smallpox Vaccine eQTL Data. *Genes & Immunity* **17**, 244–250. doi:[10.1038/gene.2016.15](https://doi.org/10.1038/gene.2016.15) (2016).
51. Davenport, E. E. *et al.* Discovering in Vivo Cytokine-eQTL Interactions from a Lupus Clinical Trial. *Genome Biology* **19**. doi:[10.1186/s13059-018-1560-8](https://doi.org/10.1186/s13059-018-1560-8) (2018).
52. Manry, J. *et al.* Deciphering the Genetic Control of Gene Expression Following *Mycobacterium Leprae* Antigen Stimulation. *PLOS Genetics* **13** (ed Sirugo, G.) e1006952. doi:[10.1371/journal.pgen.1006952](https://doi.org/10.1371/journal.pgen.1006952) (2017).
53. Kim-Hellmuth, S. *et al.* Genetic Regulatory Effects Modified by Immune Activation Contribute to Autoimmune Disease Associations. *Nature Communications* **8**. doi:[10.1038/s41467-017-00366-1](https://doi.org/10.1038/s41467-017-00366-1) (2017).
54. Brodin, P. *et al.* Variation in the Human Immune System Is Largely Driven by Non-Heritable Influences. *Cell* **160**, 37–47. doi:[10.1016/j.cell.2014.12.020](https://doi.org/10.1016/j.cell.2014.12.020) (2015).
55. Liston, A., Carr, E. J. & Linterman, M. A. Shaping Variation in the Human Immune System. *Trends in Immunology* **37**, 637–646. doi:[10.1016/j.it.2016.08.002](https://doi.org/10.1016/j.it.2016.08.002) (2016).
56. Brodin, P. & Davis, M. M. Human Immune System Variation. *Nature Reviews Immunology* **17**, 21–29. doi:[10.1038/nri.2016.125](https://doi.org/10.1038/nri.2016.125) (2017).
57. Patin, E. *et al.* Natural Variation in the Parameters of Innate Immune Cells Is Preferentially Driven by Genetic Factors. *Nature Immunology*. doi:[10.1038/s41590-018-0049-7](https://doi.org/10.1038/s41590-018-0049-7) (2018).
58. Liston, A. & Goris, A. The Origins of Diversity in Human Immunity. *Nature Immunology* **19**, 209–210. doi:[10.1038/s41590-018-0047-9](https://doi.org/10.1038/s41590-018-0047-9) (2018).
59. Lakshmikanth, T. *et al.* Human Immune System Variation during 1 Year. *Cell Reports* **32**, 107923. doi:[10.1016/j.celrep.2020.107923](https://doi.org/10.1016/j.celrep.2020.107923) (2020).

## **APPENDIX . BIBLIOGRAPHY**

---

60. Tsang, J. S. Utilizing Population Variation, Vaccination, and Systems Biology to Study Human Immunology. *Trends in Immunology* **36**, 479–493. doi:[10.1016/j.it.2015.06.005](https://doi.org/10.1016/j.it.2015.06.005) (2015).
61. Villani, A.-C., Sarkizova, S. & Hacohen, N. Systems Immunology: Learning the Rules of the Immune System. *Annual Review of Immunology* **36**, 813–842. doi:[10.1146/annurev-immunol-042617-053035](https://doi.org/10.1146/annurev-immunol-042617-053035) (2018).
62. Greenwood, B. The Contribution of Vaccination to Global Health: Past, Present and Future. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 20130433. doi:[10.1098/rstb.2013.0433](https://doi.org/10.1098/rstb.2013.0433) (2014).
63. Linnik, J. E. & Egli, A. Impact of Host Genetic Polymorphisms on Vaccine Induced Antibody Response. *Human Vaccines & Immunotherapeutics* **12**, 907–915. doi:[10.1080/21645515.2015.1119345](https://doi.org/10.1080/21645515.2015.1119345) (2016).
64. O'Connor, D. & Pollard, A. J. Characterizing Vaccine Responses Using Host Genomic and Transcriptomic Analysis. *Clinical Infectious Diseases* **57**, 860–869. doi:[10.1093/cid/cit373](https://doi.org/10.1093/cid/cit373) (2013).
65. Mooney, M., McWeeney, S. & Sékaly, R.-P. Systems Immunogenetics of Vaccines. *Seminars in Immunology* **25**, 124–129. doi:[10.1016/j.smim.2013.06.003](https://doi.org/10.1016/j.smim.2013.06.003) (2013).
66. Mentzer, A. J., O'Connor, D., Pollard, A. J. & Hill, A. V. S. Searching for the Human Genetic Factors Standing in the Way of Universally Effective Vaccines. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**, 20140341–20140341. doi:[10.1098/rstb.2014.0341](https://doi.org/10.1098/rstb.2014.0341) (2015).
67. Scepanovic, P. *et al.* Human Genetic Variants and Age Are the Strongest Predictors of Humoral Immune Responses to Common Pathogens and Vaccines. *Genome Medicine* **10**. doi:[10.1186/s13073-018-0568-8](https://doi.org/10.1186/s13073-018-0568-8) (2018).
68. Dhakal, S. & Klein, S. L. Host Factors Impact Vaccine Efficacy: Implications for Seasonal and Universal Influenza Vaccine Programs. *Journal of Virology* **93** (ed Coyne, C. B.) doi:[10.1128/JVI.00797-19](https://doi.org/10.1128/JVI.00797-19) (2019).

*APPENDIX . BIBLIOGRAPHY*

---

# List of Abbreviations

**bp** base pair

**eQTL** expression quantitative trait locus

**FWER** family-wise error rate

**GWAS** genome-wide association study

**LD** linkage disequilibrium

**MAF** minor allele frequency

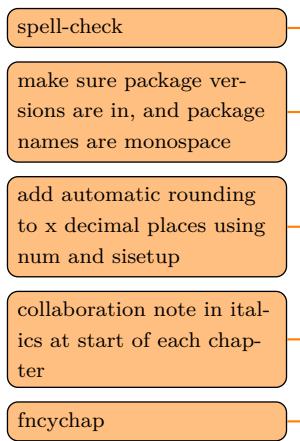
**QTL** quantitative trait locus

**reQTL** response expression quantitative trait locus

**SNP** single nucleotide polymorphism

**TF** transcription factor

**TIV** trivalent inactivated influenza vaccine



# Todo list

consider moving awkward defs to margin notes, in the style of nature reviews . . . . .	1
LD decay just takes a really really long time, but there are evo forces at work too that maintain LD . . . . .	1
Heard it's good for the reader's attention span to have figures in intro. Unless it's ok to use figures from papers, I only want to spend the time making the min that are necessary though. . . . .	2
can i use published figures? . . . . .	2
add something sweeping about utility here or elsewhere: e.g. insights into trait biology and clinical translational potential for disease traits, genetically support drug target identification . . . . .	2
seems like there is some connection to be made between the tagability of common variation and the feasibility of imputation both being enabled by the relatively small number of common haplotypes compared to variants . . . . .	4
add uses other vars . . . . .	8
list a few more types and stims from [47] until [48] . . . . .	11
not sure if right order. Since most reQTL studies are immune, I went context-specific -> reQTL -> immune rather than context-specific -> immune -> reQTL . . . . .	12
stable, yet varies by age? respecify scale of stability . . . . .	12
define what a signature is . . . . .	14
find best GWAS ref, probably mooney2013SystemsImmunogeneticsVaccines, then prune and reassign these citations . . . . .	14
not sure about scope of the subsection, currently some overlap with PANTS chapter intro. . . . .	14
spell-check . . . . .	28

make sure package versions are in, and package names are monospace . . . . .	28
add automatic rounding to x decimal places using num and sisetup . . . . .	28
collaboration note in italics at start of each chapter . . . . .	28
fncychap . . . . .	28