GGSB 2015 Prelim

Charles Czysz

September 2015

Required

1-8: General Genetic Principles
9-12: Mapping
39-50: Study Design and Statistical Data Analysis

Choice between

13-17: Genetic Architecture of Human Phenotypes

or

18-28: Population and Evolutionary Genetics

29-32: Molecular Mechanisms and Model Organisms in Human Genetics

or

33-38: Gene Regulation and Human Phenotypes

A good answer would show in escalating order:

- Basic understanding via descriptions and definition of basic terms and concepts
- Knowledge of biology/literature via empirical examples of concepts in action,
- Engagement of critical thinking by highlighting well-known limitations or novel critiques of a concept or its common application
- Recognition of open problems and novel research opportunities.

Contents

49. Describe the principles underlying the use of False Discovery Rates in Genomic studies. 41

50. Describe Bayes Theorem and be familiar with its application in genetics. 42

General Genetic Principles

1. Explain the distinction between allelic heterogeneity, genetic (locus) heterogeneity, and clinical heterogeneity. Give examples of each.

Great review of this topic:
Genetic Heterogeneity in Human Diseases

Overview
Each of the following topics have implications in the type of studies which can or cannot be used. Overall, heterogeneity ensures that large-scale association tests or case-control studies will be poorly powered to detect causal variants or genes. See the review linked above for more detail.

Definition Allelic Heterogeneity - *In a given population, different mutations in the same gene result in a similar phenotype.*

Example

Cystic Fibrosis is caused by defective cystic fibrosis transmembrane conductance regulator proteins (CFTR). Many mutations in the CFTR gene can give rise to non-functioning proteins, which all lead to the same CF phenotype.

Two-thirds of all CF mutations are a 3bp deletion at position 508, resulting in a loss of phenylalanine. 1,500 other mutations also exist which lead to CF. However, this disease is haplosufficient..

Unknown alleleic heterogeneity can affect GWA results[1] when LD methods are used.

[1] http://hmg.oxfordjournals.org/content/11/20/2417.short

Sources
http://hmg.oxfordjournals.org/content/20/20/4082.short

Definition Genetic (locus) heterogeneity - *Mutations in different genes result in a similar phenotype.*

Example - *The BRCA1 and BRCA2 genes are a good example of how mutations in different genes lead to the same phenotype.*

Definition Clinical Heterogeneity - *Variability in clinical manifestations, or phenotypes, with the same underlying mutation/genetic disorder.*

Mutations in several genes can lead to familial hypercholesterolemia, high level of LDL cholesterol. Mutations in LDLR (LDL receptor), Apolipoprotein B, proprotein convertase subtilisin/kexin type 9 (PCSK9), and the ARH/LDLRAP1 genes can all lead to familial hypercholesterolemia.

2. What is the relationship between the inbreeding coefficient, kinship coefficient, and coefficient of relatedness? How are they calculated in pedigrees? Can they be estimated in the absence of pedigree information?

Definition Identity By Descent (IBD) - *Two alleles at the same locus that are descended from the same ancestral allele within the recent past. Can be 0, 1, or 2 depending on how many ancestral alleles shared between individuals.*

Definition Coefficient of Kinship - f_{xy}*: The probability that two alleles, one from X and the other from Y, are IBD.*

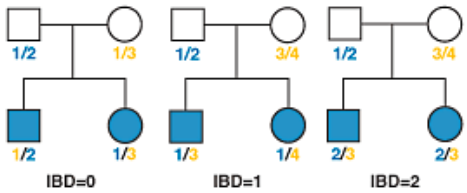
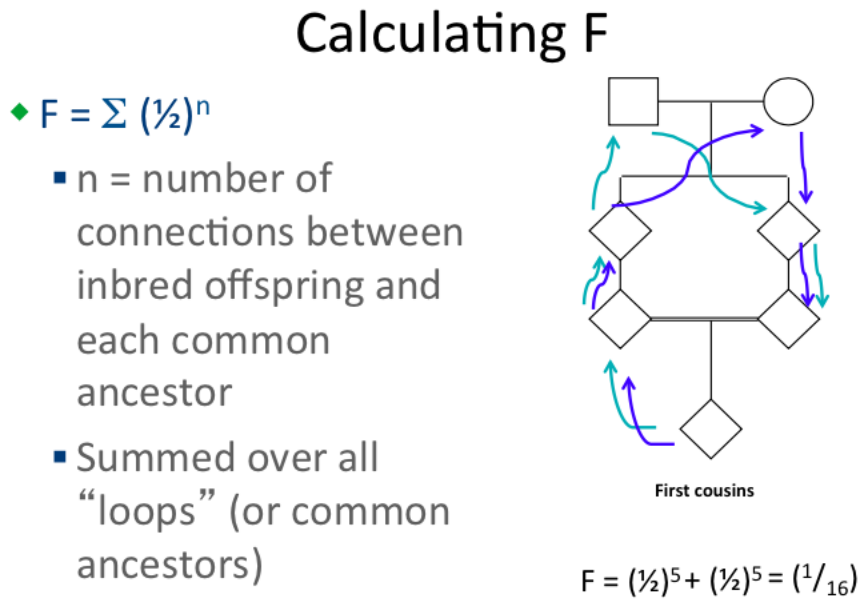


Figure 1: IBD Pedigree Example



Definition Coefficient of Relatedness - R_n*, the probability of sharing* $n \in \{0, 1, 2\}$ *IBD alleles. Mean relatedness,* $\bar{r} = 0 \times r_0 + 1 \times r_1 + 2 \times r_2$.

Relationship between Coefficients

$$f_{xy} = \frac{1}{2}\bar{r}$$

Estimation without Pedigrees

Genotype data can be used to estimate the inbreeding coefficient between individuals using computational tools such as PLINK.

3. What are the key distinguishing characteristics of pedigrees segregating autosomal dominant, autosomal recessive, X-linked, Y-linked, and mitochondrial diseases?

Autosomal Dominant

Both sexes are equally affected. Offspring have a 50% chance of inheriting the disease from an affected parent.

Autosomal Recessive

If both parents are carriers, then $\frac{1}{4}$th of offspring will show a disease phenotype. May appear to skip generations residing in a heterozygous carrier state instead of homozygous affected.

X-linked Dominant

Affected men always pass disease to daughters. Affected women have $\frac{1}{2}$ths chance of passing disease to offspring.

X-linked Recessive

Women can be either affected or carriers. Men always affected. Colorblindness as an example.

Y-linked

Affected fathers always pass to sons, never to daughters.

Mitochondrial

Passed on only by mothers to all offspring.

4. Explain the "non-Mendelian" concepts of uniparental disomy and imprinting. How would these be manifested in pedigrees and how are they demonstrated at the cellular or molecular levels?

Definition Mendelian Genetics - *Involves the "Laws" of random segregation, independent assortment, and dominance. Non-Mendelian genetics violate at least one of these assumptions.*

Definition Uniparental Disomy - *Offspring inherits a chromosome, or part of one chromosome, from one parent only. These offspring are euploid, but with an unequal contribution of genetic material from the parents.*

Uniparental disomy can manifest in two ways: inheriting both homologs from one parent (heterodisomy), or duplication of a single inherited chromosome (isodisomy).

Heterodisomy

Offspring inherits homologs from one parent only. Can occur during a nondisjunction error in Meiosis I, where one daughter cell inherits both sister chromatids while the other inherits none. This will lead to trisomy after fertilization, and heterodisomy if trisomy rescue occurs and results in the loss of the allele from the other parent.

Heterodisomy can also result from crossover events. In the parental generation, a cross between a balanced translocation carrier and a normal carrier can result in partial heterodisomy if the offspring inherits a crossover product and non-crossover product. However, this is an unbalanced gene load and is less prone to survival (10-15% incidence versus theoretical 50%).

In a Robertsonian translocation, acrocentric chromosomes fuse to create a single fusion chromosome. If this translocation occurs between homologs and is passed down to offspring, trisomy occurs (eg Trisomy 21). If, however, the non-fusion chromosome is lost by trisomy rescue, then heterodisomy occurs.

In a pedigree analysis, heterodisomies result in a parent passing their recessive condition to an offspring. Additionally, the parents of the affected parent are carriers.

Isodisomy

Isodisomy occurs when a single parental homolog is inherited and duplicated. This can occur after nondisjunction events in both Meiosis I and II. While the Meiosis I error gives 2 gametes which can result in heterodisomy, the other two gametes result in isodisomy if chromosomal duplication occurs. In this case, the other parental chromosome is duplicated.

During a meiosis II error, the sister chromatids fail to separate, leaving one product with two copies of the same chromosome in one gamete and none in the other. For the former gamete, trisomy rescue and loss of the other parent's chromosome leads to isodisomy. For the latter gamete, duplication of the other parental chromosome leads to isodisomy.

In a pedigree analysis, isodisomy should be suspected when an offspring shows a recessive disease when only one parent is a carrier.

This suggests that the recessive allele was inherited and duplicated.

Discovering Uniparental Disomy

Discrepancies between karyotypes of placenta and fetus.

Mosaicism of normal and trisomic cells (rescue occurs after first cell division).

Abnormal chromosome structure giving evidence of inheritance of Robertsonian product or unbalanced products. Usually seen by FISH.

Other DNA tests:

Test for STR (short tandem repeats) of known high heterozygosity, genotyping parents and offspring.

Definition Imprinting - *Imprinting results from the inheritance of epigenetic marks silencing an allele in a parent-of-origin specific manner.*

With imprinting, only one allele is expressed, and which allele depends on its parent-of-origin. When uniparental disomy of an imprinted allele occurs, the resulting expression can either go to 0x or 2x depending on if the silenced allele is the one doubled or not.

Sources

http://www.nature.com/gim/journal/v3/n3/full/gim200144a.html

5. What evidence is there for the presence of modifier loci? How is this related to the concept of epistasis and how is it distinct (or not) from polygenic and other models of inheritance?

Definition Modifier loci - *A loci or gene which alters the expression of another gene, usually through epistatic interactions.*

Modifier loci usually considered part of the genetic background. Thus, some modifier loci can have no effect on wildtype phenotypes but can affect mutant phenotypes.

In epistatic interactions, modifier loci do not lead to binary traits (present or absent, etc), but result in a gradient of phenotypes (coat color in dogs and horses). Can enhance, suppress, normalize, and enhance pleiotropy. These modifier loci usually act on or within the same biological pathway as the gene they affect.

Evidence

Clinical variability in disease manifestation despite identical genotypes at the risk loci. Many exist in mice, but fewer found in humans due to difficulty to study.

Variation in phenotype without variation in genotype.

Mode of Inheritance

The inheritance of modifier loci cannot be accurately described by Mendelian or polygenetic modes of inheritance. Modifier alleles not strictly dominant or recessive, but depend on other alleles. For example, a dominant negative modifier loci only destroys a phenotype when inherited with a mutant allele in the same pathway.

In the polygenic model of inheritance, alleles affecting a phenotype act independently and with small, additive effect sizes. Thus only the total allele count matters. With modifier loci, the presence or absence of a particular allele determines the phenotype. Not only do alleles affecting a phenotype matter, but alleles affecting those alleles do also.

6. What are distinctions among the concepts linkage, linkage disequilibrium, and association? Under what circumstances would each be preferable for genetic mapping? Consider both sample composition and types of diseases.

Definition Linkage - *Two loci which tend to be inherited together due to low probability of recombination.*

Using data from many recombinations, linkage maps were created to map loci relative to the proportion of recombinations between them. In units of map units **m.u.** or centiMorgans. 1 mu is defined as the distance between genes with 1/100 meiosis products as recombinant.

Parametric Linkage Mapping

Useful when samples are from large and informative pedigrees. Informative indicates that each allele can be assigned to a specific parent.

Calculate LOD scores by finding the MLE of θ

$$\text{LOD} = \log \frac{(1-\theta)^{NR}\theta^{R}}{0.5^{NR+R}}$$

Nonparametric Linkage Mapping

Used when the underlying model is unknown. Use affected relatives in order to leverage degree of allele-sharing. Estimate IBD between affected relatives. Look for alleles shared between affected individuals more often than chance.

Linkage mapping usefulness

Linkage mapping is useful with large and informative pedigrees. Mapping also requires easily genotyped markers providing good resolution of genome cover. Linkage mapping also best identifies Mendelian genes with high penetrance.

Definition Linkage Disequilibrium (LD) - *Non-random association of alleles at multiple loci.*

LD is measured using a variety of statistics, the most common is the difference between expected heterozygosity and observed heterozygosity.

LD decreases each generation due to recombination. Including LD information in linkage analyses can increase power.

Limitations

LD models assume constant population size and drift-recombination equilibrium. Population structure can artificially inflate LD estimates.

Definition Association Studies - *Utilizing a large sample of unrelated cases and controls to determine which, if any, tag SNPs are more associated with a disease than not.*

The most common, now, association study is GWAS. In GWAS, measure associations with common (MAF > 5%) SNPs and a phenotype of interest. Using a large cohort of people in the same population increases power to detect associations.

7. Define epistasis. Describe approaches that allow epistasis to be detected or quantified. Describe some biological mechanisms that can produce epistasis. Discuss the implications of epistasis for efforts to map the genetic causes of phenotypes. Discuss the potential implications of epistasis for the evolutionary process.

Multiple types of epistasis observed:

- **Functional** - Genes in the same pathway
- **Compositional** -
- **Statistical**

Effect of epistasis on mapping

An epistatic interaction might prevent identifying an association with one loci and a phenotype depending on

8. Define heritability. Describe methods used to quantify the heritability of a phenotype. Discuss the value and limitations of heritability as a descriptor of the extent to which a phenotype has genetic causes. Describe the "missing heritability problem" and its potential explanations.

A measurable phenotypic trait varies within a population. Beak length in finches is a good example: there exists a mean length, but the values can be higher or lower. This variance is denoted by σ_P^2.

For most traits, the phenotype is made up of contributions from genes and environment, and similarly for the variance.

$$\sigma_P^2 = \sigma_G^2 + \sigma_E^2$$

The relationship between genetic variability and phenotypic variability is deemed heritability.

$$h^2 = \frac{\sigma_a^2}{\sigma_P^2}$$

Where σ_a^2 is the additive genetic variance (as opposed to dominance effects).

An estimate of h^2 can be found by plotting the phenotypic value of offspring versus the midparent value and fitting a linear regression. The slope of this regression estimates h^2, how well the parent phenotype determines the offspring phenotype.

Twin Method

Twin studies allow the inflation of heritability by common environment to be minimized. Since twins are presumed to have experienced the same environment growing up, the environmental variance is the same. Comparing monozygotic twins to dizygotic twins will allow a better estimation of heritability.

Considerations

Calculations of heritability can be influenced by shared environment. Thus, heritability is usually specific to a single population or region.

GCTA

Missing heritability

A heritability of 1 indicates that a genetic variant or variants can explain all of the phenotypic variance. As association studies and GWA studies have increased sample size and tagged variants, the heritability measures of measured traits has not increased as quickly.

Mapping

9. What is the difference between an odds ratio and relative risk? When would you use each and how might these relate to the concept of heritability in a genetic study?

Definition Relative Risk - *λ_R is the risk that an affected person acquires a trait divided by the risk for the general population.*

Consider a 2x2 contingency table testing a response (yes or no) based on group (placebo versus medicine). π_1 is defined as the probability of a "yes" response for the placebo group. π_2 is the probability of a "yes" response for the medicine group.

Group	Diseased	Not Diseased	Total
Placebo	π_1	$1-\pi_1$	1
Medicine	π_2	$1-\pi_2$	1

The relative risk for this table is given by

$$\text{relative risk} = \frac{\pi_1}{\pi_2}$$

Example Data:

Group	Diseased	Not Diseased	Total
Placebo	189	10,845	11,034
Medicine	104	10,933	11,037

Table 1: Adapted from Table 2.3 from *Introduction to Categorical Data Analysis* by Alan Agresti

In this example, $\hat{\pi_1} = \frac{189}{11,034}$ and $\hat{\pi_2} = \frac{104}{11,037}$. A small difference, but a relative risk of 1.83, meaning the placebo group is 83% more at risk than the medicine group.

Definition Odds - *Expected ratio of successes. Define p as probability of success, $odds_1 = \frac{p}{1-p}$. Thus, an odds of 3 indicates a probability of success of 0.75 and of failure 0.25.*

For the previous table, we can define the odds ratios for each group, giving $\text{odds}_1 = \frac{\pi_1}{1-\pi_1}$ and $\text{odds}_2 = \frac{\pi_2}{1-\pi_2}$, respectively.

Definition Odds Ratio - *A ratio of odds. For a 2x2 contingency table, $OR = \theta = \frac{odds_1}{odds_2}$.*

Odds ratios of 1 describe the null hypothesis that the group and response are independent. When $\theta > 1$, the odds of success for group 1 are θ times higher than for group 2. When $\theta < 1$, the odds of success for group 2 are $\frac{1}{\theta}$ times higher than for group 1.

Definition Log Odds - *As a fraction of positive values, the odds ratio is only lower-bounded by 0. Thus, the distribution is very skewed. Taking the log odds removes this skewedness since it is symmetric around its null value of 0.*

Relative Risk versus Odds Ratio

Recall that

$$\text{RR} = \frac{\pi_1}{\pi_2}$$

and

$$\text{OR} = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}$$

Thus,

$$OR = RR \times \frac{1-\pi_2}{1-\pi_1}$$

and when the probabilities of success are both small, the RR and the OR are very similar.

Odds Ratio and Heritability

In GWAS, many variants show low effect sizes with OR between 1 and 1.3. However, GWAS might miss variants with very low OR (e.g 1.05), thus contributing to the missing heritability problem.

Sources

Lander and Schork. "Genetic Dissection of Complex Traits" 1994. *Science*

10. What are the differences between quantitative and qualitative trait mapping, correlated phenotypes, multi-trait mapping?

Qualitative versus Quantitative Traits

Qualitative traits are those of categorical, rather than numeric, values. Coat color in rats is one example (black, white, or brown). Quantitative traits, on the other hand, are phenotypes which can be measured (tail length, in centimeters). Qualitative trait mapping is useful for Mendelian inheritance while quantitative trait mapping is better for traits with multiple underlying causative loci.

Trait Mapping

Trait mapping allows the interrogation of the genetic region associated with a phenotypic trait.

Qualitative trait mapping requires

- Separate phenotypic categories
- Many genetic markers
- Inbred line for crosses

11. Describe the impact of using subjects with varied degrees of relatedness in an association study. How can this be addressed?

Using related individuals

One reason to include related individuals in a study is to increase the power to detect rare variants passed down through a family.

Related genomes can be used to correct for sequencing errors (differentiating a SNP or indel) and to allow imputation to cheaply genotype others after deeply genotyping a few.

Family-based association tests are performed using the transmission distortion test.

Kinship Matrix

The relatedness between subjects is calculated in a kinship matrix, an nxn matrix of correlations between genotyped markers.

12. What are the strategies used to identify causal variants within an LD block associated with a trait? How do you assign a target gene as connected to a noncoding variant associated with a trait? How do you prioritize variants as likely causal/ how do you experimentally test them?

Due to LD, many variants can be shown to associate with a particular phenotype. Also due to LD, the most associated variant may not be the causal variant, but in high LD with the causal variant. This usually arises from relatively low genotyping coverage (not all variants will be typed).

Fine Mapping When a candidate region is found in GWAS, fine mapping studies are used to determine the causal variant. Fine-mapping studies rely on both a high number of genotypic markers and high recombination rates. These two factors allow precise identification of markers segregating with a phenotype of interest.

Connecting noncoding variants to genes

Non-coding variation can still play a role in gene expression and activity by affecting how it is transcribed. Variation in promoters, enhancers, introns, and other regions can cause differential gene regulation, affecting a trait of interest.

Functional genomic techniques can be used to assign variants to functional elements in the genome. If a variant overlaps or is near an element, the most likely mechanism of action for that variant is explained by the regulatory element.

Prioritizing variants as causal

Functional annotations - Overlap variants with functional elements to see any correlations with regulatory elements or other mechanisms of affecting genes

Nucleotide-level analysis - Determine if the variant affects a motif.

Conservation - Compare variants to related species to determine conservation. Areas under selective constraint relative to surrounding regions can be assumed biologically important

Network analysis -

Genetic Architecture of Human Phenotypes

13) What is "genetic architecture" in the context of diseases and traits? What methods do we have for understanding genetic architecture and how would you characterize the relative power and resolution among these tools? What practical reasons are there for having a good understanding of it? What if any value is there in knowing general features of architecture in the absence of complete and detailed knowledge?

14) What are the expected relationships between allele frequency spectrum, penetrance, and prevalence of genetic disease?

Definition Allele frequency spectrum - *AFS is defined as the frequency of all alleles of a given gene.*

The frequency of allele a is given by

$$\frac{a}{N * n}$$

for a population of N individuals with poloidy n and a copies of the a allele.

Definition Prevalence - *In a population, the prevalence of a disease is the proportion of individuals affected.*

Definition Penetrance - *For pedigrees, penetrance is the proportion of affected individuals given a particular genotype.*

15) What are the assumptions of the polygenic model of quantitative traits and of binary (disease) traits? What are alternative models?

The polygenic model of quantitative traits assumes a polygenic phenotype is due to additive interactions of alleles at many loci. Each loci serves as a binomial random variable, and the sum of multiple loci result in a normal distribution describing the distribution of number of additive alleles in a population.

Assumptions

- Alleles have small, equal effects
- Effects are additive
- No dominance
- No epistatic or gene by gene interactions
- Loci are independent, i.e. no linkage
- Environmental variance is minimal

16. Discuss and contrast the common versus rare allele hypotheses to explain the genetic architecture of human diseases. What are the implications of each hypothesis for mapping strategies, including association studies and large-scale sequencing efforts?

17. How noncoding variants lead to phenotypic variation? What do we know about the role of enhancers in Mendelian disease? What's the evidence for enhancer variation in complex disease? Do enhancer variants have small or large effects n disease phenotypes?

Population and Evolutionary Genetics

27. Genome structure: Describe major features of the structure of mammalian genomes (e.g. genome size, nucleotide composition, gene families, repetitive elements, small noncoding RNAs, centromeres and telomeres) and hypothesized mechanisms affecting their evolution.

Definition Telomeres - *In humans, a 6bp (TTAGGG) repeated region at the ends of chromosomes. This repeat allows the chromosome ends to form a T-loop to prevent recognition as a break.*

Telomeres are shortened after each DNA replication event due to inability for DNA Pol to start at the ends. Telomerase allows the lengthening of telomeres in the 3' direction by use of an RNA template. After one strand is lengthened enough, DNA Pol can fill in the opposite direction.

In humans, telomerase is suppressed in somatic tissues. The prevaling hypothesis is that as organisms grew larger, telomere shortening was used as a way to prevent cell proliferation. There is a negative correlation between body mass and telomerase activity.

Genome Structure

As humans, chimps, and other apes evolved from their common ancestor, the structure of their chromosomes also changed. Due to duplications, deletions, and fusions, modern apes can have very different chromosome structure. In humans, chromosome 2 is the result of a fusion of two ancestral chromosomes still present in other apes.

Centromeres

Definition Point Centromeres - *Point centromeres are a small, specific sequence to which mitotic machinery bind with high affinity. Observed in yeast.*

Definition Regional Centromeres - *Human centromeres are regional, defined by large, repetitive sequences.*

Definition Holocentromeres - *In* C. elegans*, the spindle fibers can attach anywhere along the chromosome.*

Genome Size

In general, the size of the genome does not track with measures of "complexity". Important to consider is the proportion and possible function of non-coding DNA.

GC Content

The percentage of bases which are G or C. Varies across different species, and could be the result of evolving to conserve nitrogen or change genome stability. There's a slight negative correlation between body mass and percent GC content.

Gene Families

Genes which share common ancestors, one example being the globin family of genes.

Gene Regulation and Human Phenotypes

33. Critically review the evidence and mechanisms of transgenerational epigenetic inheritance.

34. Discuss the theoretical impact of epigenetic phenomena on association studies.

35. Define eQTL mapping. Discuss the insight we gained from eQTL mapping on gene regulation, on the genetic architecture of phenotypes. Discuss their utility for understanding the genetic bases of human diseases.

36. Histone modifications: what we know about their role in gene regulation.

37. Network motifs in biology. Examples of regulatory circuits and their function.

38. Robustness and redundancy in biological functions. How do they evolve and how are they being maintained?

Study Design and Statistical Data Analysis

39. Normalization of genomic data: why it is needed and what assumptions are required.

Standardization and normalization are important when considering data from the same biological source across different technical replicates. Technical variation can introduce unwanted noise onto an underlying biological distribution, so removing that noise is the goal.

Normalization casts data into a predefined range while standardization casts the data to have mean 0 and variance 1. Applying both yields the standard normal distribution used in Z-tests.

Normalization is necessary for allowing comparisons of biological and technical replicates. Despite coming from the same source type, random variation can alter the distribution of the measurements without reflecting differences in biology.

Assumptions

- All samples share the same underlying distribution
- Experimental treatments affect a small number of genes
- Noise is additive

40. Accounting for confounding factors (e.g., batch effect) in genomic data: methods to test and correct for such effects.

Definition Batch - *Variables differing between samples where the difference is not due to measured variables. Examples include sequencing machine, technician, etc.*

Batches can introduce bias in measurements, and must be minimized in experimental design. However, if there is a batch effect in the data, identifying and correcting are important.

Identifying

The easiest way to identify batches is by keeping records of all sample preparation and including them as cofactors in the linear model. Computationally, PCA and clustering methods can identify batches to include as factors.

41. Effective design for genomic studies (how to replicate? how to account for confounding factors? when and how to pool samples? variance vs. bias).

42. Be familiar with the use of Principal Components Analysis (and related dimension reduction techniques) in genetic applications.

43. Be familiar with methods for identifying population structure and admixture.

44. Genotype imputation: how does it work, how is it used?

45. What is a Hidden Markov Model?

46. Describe the principles behind statistical methods for inferring haplotypic phase from simple pedigrees (eg parent-child trios), and samples of unrelated individuals.

47. Describe how to fit a linear model and test for significance of effects. What does a p value of 0.05 (or 0.01, or any other value) mean?

48. What issues affect your assessment of whether an observed association is likely to be real?

49. Describe the principles underlying the use of False Discovery Rates in Genomic studies.

FDR is a method used to correct for multiple testings, which inflate the Type I error rate as the number of tested hypotheses increases.

In genetic tests, the likelihood of seeing a significant p-value increases as the number of tests increases (i.e. associations for 1,000,000 SNPs or 20,000 genes).

The Benjamini-Hochberg method is the most widely-used. For independent null hypotheses $H_1 ... H_m$ and corresponding p-values $P_1 ... P_m$,

- For an FDR α, find the largest k s.t. $P_k \leq \frac{k}{m}\alpha$
- Then, reject all hypotheses H_i for $i = 1, ..., k$

In this way, the FDR represents the number of significant hits which are false, instead of the Type I proportion of false positives in all tests.

	Do not reject	Reject	Total
H_0 True	U	V	m_0
H_1 True	T	S	m_1
Total	W	R	m

A p-value significance cutoff measures Type I error as V/m, while q-values set the FDR at V/R.

Comparison with Family-wise error rate correction

The most common FWER correction is the Bonferroni correction, defined as $\alpha* = \frac{\alpha}{m}$, where α is the initial p-value significance and m is the number of hypothesis tests. This cutoff is much more conservative than FDR, in that fewer tests will be found significant, inflating the Type II error rate. FDR also gives tests greater power than Bonferroni corrections.

50. Describe Bayes Theorem and be familiar with its application in genetics.

Relationship	Kinship coefficient	Coeffcient of relatedness
Self	0.5000	1.000
Monozygotic twins	0.5000	1.000
Parent-child	0.2500	0.500
Full siblings	0.2500	0.500
Half siblings	0.1250	0.250
First cousins	0.0625	0.1250
Unrelated	0.0000	0.0000

Figure 2: Relationship between Kinship and Relatedness coefficients