Genome-wide Association Studies

IBS 746

10/22/2020

Jingjing Yang (jingjing.yang@emory.edu)

Outline

- Hypothesis Testing
  - Hardy-Weinberg Equilibrium (HWE)
- Linkage Disequilibrium
  - Origin
  - Calculation
- Single Variant Test
  - Dichotomous Trait
  - Quantitative Trait
- GWAS

Hypothesis Testing

- Null Hypothesis and Alternative Hypothesis
  - Two sample t-test : H_0: $\mu_1 = \mu_2$, H_a: $\mu_1 \neq \mu_2$
  - Two-Way tables and Chi-Square Test:
    - H_0 assumes that there is no association between the row and column variables (in other words, one variable does not vary according to the other variable).
    - H_a claims that some association does exist.
- Test Statistic
  - Two sample t-test with shared standard deviation σ

$$t = \frac{\widehat{\mu_1} - \widehat{\mu_2}}{\hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

  - Two-Way tables and Chi-Square Test

$$\mathrm{X}^2 = \sum_{All\ Cells} \frac{(Observed\ - Expected)^2}{Expected}$$

Hypothesis Testing

- Derive the distribution of test statistic under Null hypothesis H_0
- Calculate test statistic value with given data
- Obtain p-value based on test statistic distribution under H_0 and test statistic value with given data

Chi-Square test

- Let O_{ik} denote the count in the i^{th} row and k^{th} column in the contingency table, and E_{ik} denote the corresponding expected count
- Test statistic

$$X^2 = \sum_{i=1}^{I}\sum_{k=1}^{K}\frac{\left(O_{ik} - E_{ik}\right)^2}{E_{ik}} \sim \chi^2_{(I-1)(K-1)}$$

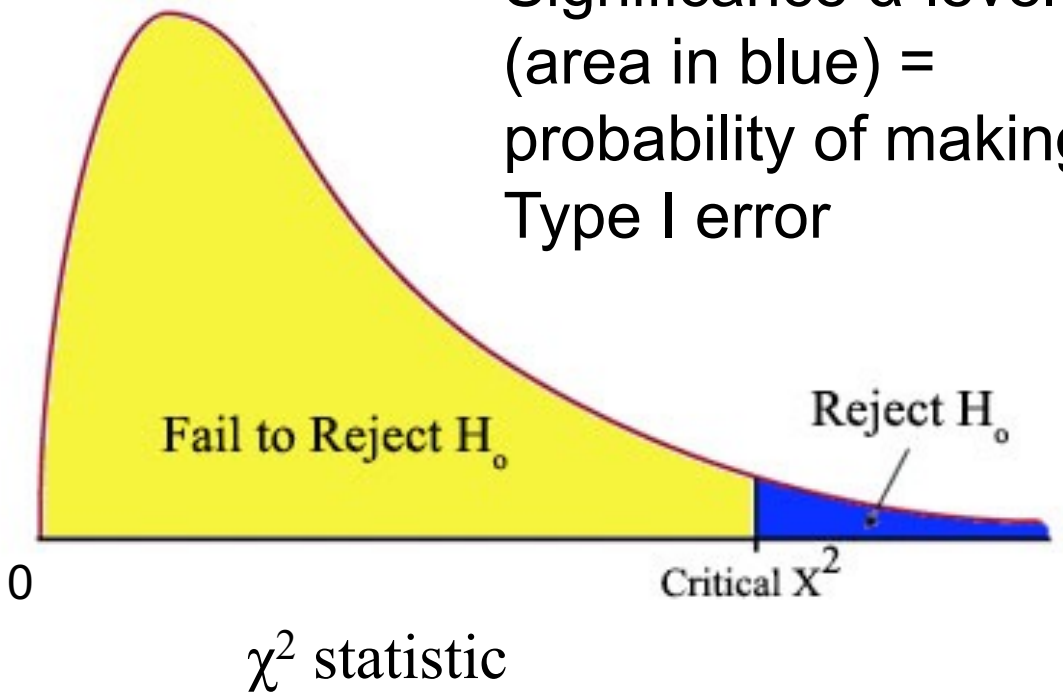
- If study sample size is large, the test statistic follows a chi-squared distribution with degrees of freedom (I - 1)(K - 1) under H_0, with I rows and K columns in the contingency table.

Hypothesis Testing

- Obtain p-value based on test statistic distribution under H_0 and test statistic value with given data
- Reject H_0 if p-value < significance level (0.05) or test statistic value exceed the critical values based on significance level
- What does significance level mean?
- How to determine the critical values?
- What is p-value and how to calculate it?

Chi-square (χ^2) distribution

Critical value of X^2 chosen to attain desired α-level



Significance α-level (area in blue) = probability of making a Type I error

What is a Type I error?

What is p-value?

Critical Values of the Chi-square Distribution

	P									
df	**0.995**	**0.975**	**0.9**	**0.5**	**0.1**	**0.05**	**0.025**	**0.01**	**0.005**	**df**
1	.000	.000	0.016	0.455	2.706	3.841	5.024	6.635	7.879	1
2	0.010	0.051	0.211	1.386	4.605	5.991	7.378	9.210	10.597	2
3	0.072	0.216	0.584	2.366	6.251	7.815	9.348	11.345	12.838	3
4	0.207	0.484	1.064	3.357	7.779	9.488	11.143	13.277	14.860	4
5	0.412	0.831	1.610	4.351	9.236	11.070	12.832	15.086	16.750	5
6	0.676	1.237	2.204	5.348	10.645	12.592	14.449	16.812	18.548	6

Hardy-Weinberg Equilibrium (HWE)

- HWE law: Allele and genotype frequencies in a population will remain constant from generation to generation in the absence of other evolutionary influences
- Consider two alleles: A and a
- Allele frequency for A is f(A) = p; allele frequency for a is f(a) = q = 1 – p
- The expected genotype frequencies under random mating are
  - $f(AA) = p^2$ for the AA homozygotes
  - $f(aa) = q^2$ for the aa homozygotes
  - f(Aa) = 2pq for the heterozygotes
- In the absence of selection, mutation, genetic drift, or other forces, allele frequencies p and q are constant between generations, so equilibrium is reached.

HWE Test

- Consider the following contingency table for N = $N_{AA} + N_{Aa} + N_{aa}$ samples
- $f(A) = p = \frac{2N_{AA}+N_{Aa}}{2N}$; $f(a) = q = 1 - p$

Genotype	AA	Aa	aa
Observed Count	N_{AA}	N_{Aa}	N_{aa}
Expected	p^2N	$2pqN$	q^2N

Test HWE

- Pearson's chi-square test:

$$\mathrm{X}^2 = \sum \frac{(O-E)^2}{E} = \frac{\left(N_{AA}-p^2N\right)^2}{p^2N} + \frac{(N_{Aa}-2pqN)^2}{2pqN} + \frac{\left(N_{aa}-q^2N\right)^2}{q^2N}$$

- Under the null hypothesis (HWE), X^2 follows a chi-square distribution with 1 degree of freedom

- The critical value for 0.05 significance level is 3.84

Test HWE

- Calculate the Pearson's chi-square p-value for testing the HWE with the following observation table

Genotype	AA	Aa	aa
Observed Count	1009	198	48

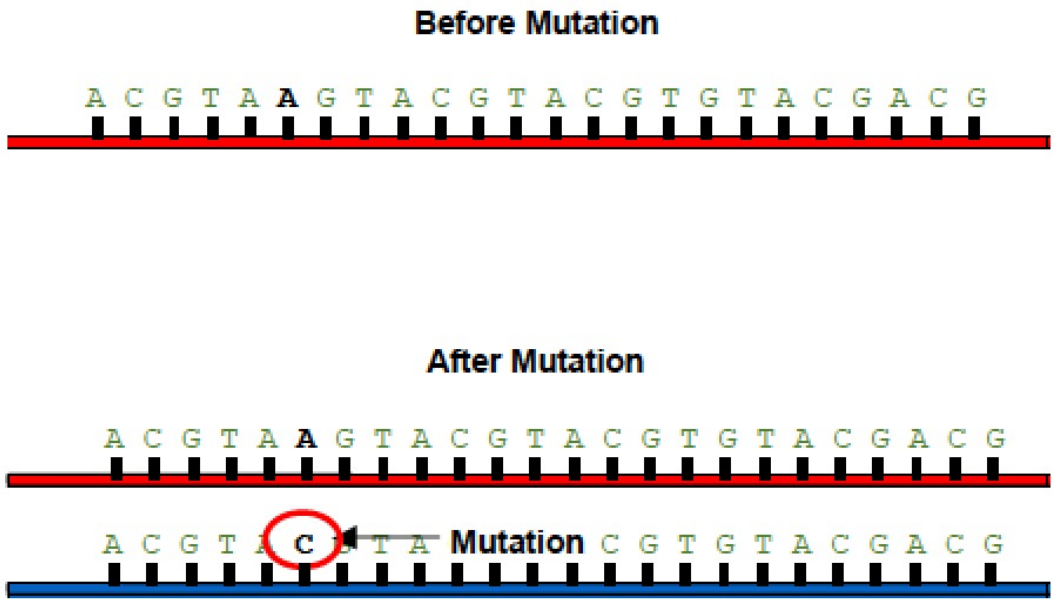
- Will you reject the null hypothesis (HWE)?
- What dose it mean if you reject the null hypothesis (HWE)?

Linkage Disequilibrium (LD)

- Definition in Population Genetics
  - **Linkage Disequilibrium (LD)** is the **non-random association** of alleles at different loci in a given population.
- Why nearby markers are likely to be correlated?
- The origin of LD?

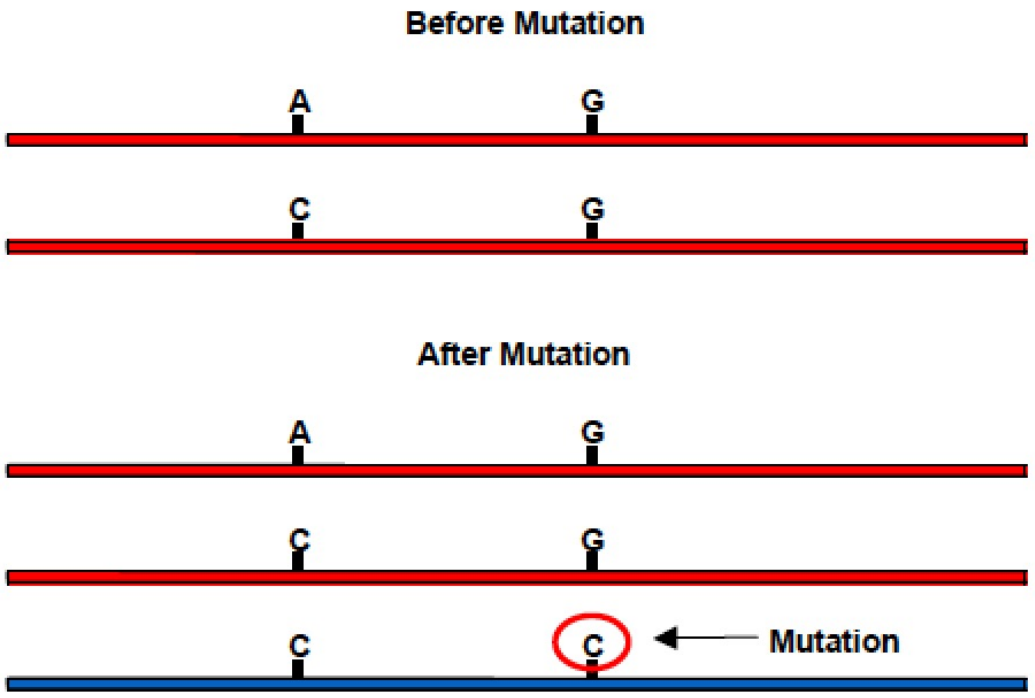
Linkage Disequilibrium (LD)

- Consider the history of two neighboring single nucleotide polymorphism (SNP)
- SNPs exist today arose through ancient mutation events…



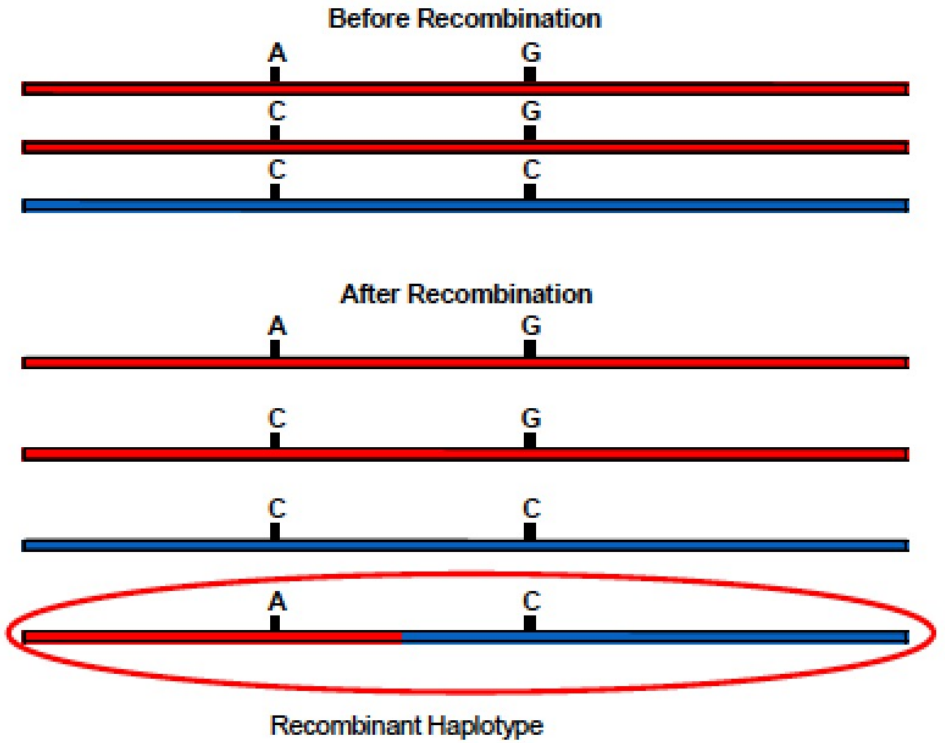
Linkage Disequilibrium (LD)

- One SNP arose first and then the other …



Linkage Disequilibrium (LD)

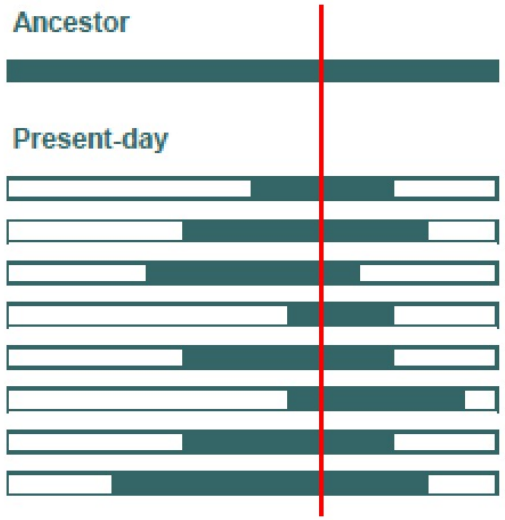
- Recombination generates new arrangements for the ancestral alleles



Linkage Disequilibrium (LD)

- Chromosomes are mosaics

- Extent and conservation of mosaic pieces depends on
  - Recombination rate
  - Mutation rate
  - Population size
  - Natural selection



- Combinations of alleles at very close markers reflect ancestral haplotypes

Quantify Linkage Disequilibrium (LD)

- **LD** is defined as the **difference** between the **observed frequency** of a particular combination of alleles at two loci and the **frequency expected** for random association.
- Allele frequency
  - $P_A, P_a, P_A + P_a = 1$
  - $P_B, P_b, P_B + P_b = 1$
  - $P_{AB} = P_A P_B$
    if and only if alleles A, B are independent
- Minor Allele Frequency (MAF)

		Locus B		Totals
		B	b	
Locus A	A	p_{AB}	p_{Ab}	p_A
	a	p_{aB}	p_{ab}	p_a
Totals		p_B	p_b	1.0

Linkage Equilibrium
Expected for Distant Loci

$$p_{AB} = p_A p_B$$

$$p_{Ab} = p_A p_b = p_A(1 - p_B)$$

$$p_{aB} = p_a p_B = (1 - p_A) p_B$$

$$p_{ab} = p_a p_b = (1 - p_A)(1 - p_B)$$

Linkage Disequilibrium
Expected for Nearby Loci

$$p_{AB} \neq p_A p_B$$

$$p_{Ab} \neq p_A p_b = p_A(1 - p_B)$$

$$p_{aB} \neq p_a p_B = (1 - p_A) p_B$$

$$p_{ab} \neq p_a p_b = (1 - p_A)(1 - p_B)$$

Disequilibrium Coefficient D_{AB}

$$D_{AB} = p_{AB} - p_A p_B$$

$$p_{AB} = p_A p_B + D_{AB}$$

$$p_{Ab} = p_A p_b - D_{AB}$$

$$p_{aB} = p_a p_B - D_{AB}$$

$$p_{ab} = p_a p_b + D_{AB}$$

D_{AB} is hard to interpret

- Sign is arbitrary …
  - A common convention is to set…
    - *A, B* as the common alleles
    - *a, b* as the rare allele
- Range depends on allele frequencies
  - Hard to compare between markers
- Can you see why the range of D_{AB} depends on allele frequencies?

Boundaries for D_{AB}

- By using the fact that $p_{AB} = P(AB)$ must be less than both $p_A = P(A)$ and $p_B = P(B)$, and that allele frequencies cannot be negative, the following relations can be obtained:
  - $0 \leqslant p_{AB} = p_A p_B + D_{AB} \leqslant p_A, p_B$
  - $0 \leqslant p_{aB} = p_a p_B - D_{AB} \leqslant p_a, p_B$
  - $0 \leqslant p_{Ab} = p_A p_b - D_{AB} \leqslant p_A, p_b$
  - $0 \leqslant p_{ab} = p_a p_b + D_{AB} \leqslant p_a, p_b$
- These inequalities lead to bounds for D_{AB} :

$$-p_A p_B, -p_a p_b \leqslant D_{AB} \leqslant p_a p_B, p_A p_b$$

Normalized Linkage Disequilibrium Coefficient

- The possible values of D depend on allele frequencies. This makes D difficult to interpret. For reporting purposes, the normalized linkage disequilibrium coefficient D' is often used.

$$D'_{AB} = \begin{cases} \frac{D_{AB}}{max(-p_A p_B, -p_a p_b)} & \text{if } D_{AB} < 0 \\ \frac{D_{AB}}{min(p_a p_B, p_A p_b)} & \text{if } D_{AB} > 0 \end{cases} \quad (1)$$

Estimate D_{AB}

- Suppose we have the N haplotypes for two loci on a chromosomes that have been sampled from a population of interest. The data might be arranged in a table such as:

	B	b	Total
A	n_{AB}	n_{Ab}	n_A
a	n_{aB}	n_{ab}	n_a
	n_B	n_b	N

- We would like to estimate D_{AB} from the data. The maximum likelihood estimate of D_{AB} is

$$\hat{D}_{AB} = \hat{p}_{AB} - \hat{p}_A \hat{p}_B$$

where $\hat{p}_{AB} = \frac{n_{AB}}{N}$, $\hat{p}_A = \frac{n_A}{N}$, and $\hat{p}_B = \frac{n_B}{N}$

- So the population frequencies are estimated by the sample frequencies

Measuring LD with r^2

- Define a random variable X_A to be 1 if the allele at the first locus is A and 0 if the allele is a.
- Define a random variable X_B to be 1 if the allele at the second locus is B and 0 if the allele is b.
- Then the correlation between these random variables is:

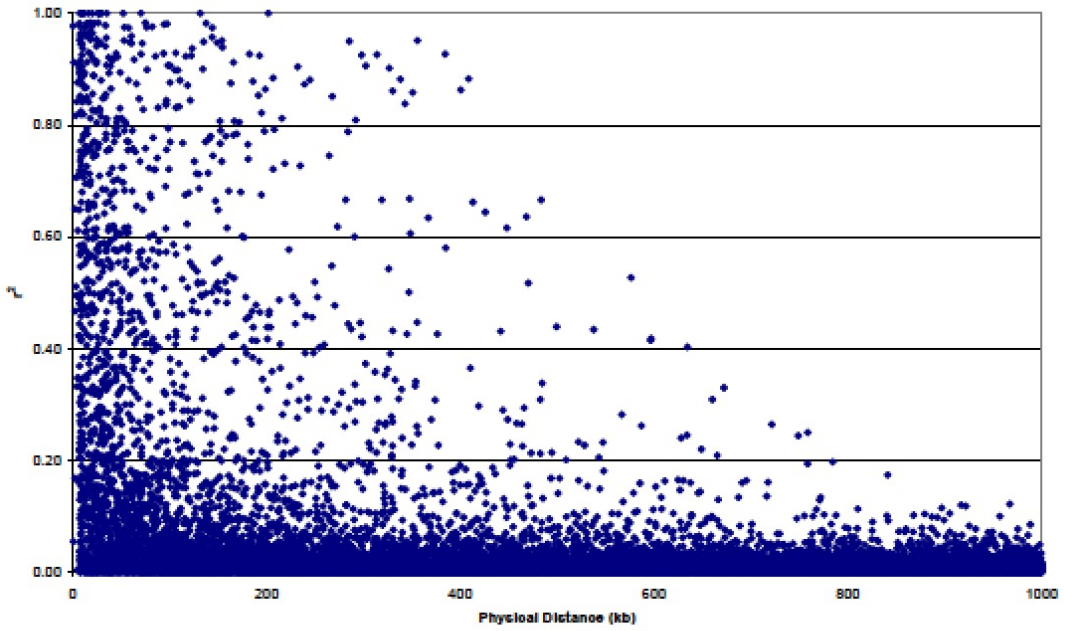
$$r_{AB} = \frac{COV(X_A, X_B)}{\sqrt{Var(X_A)Var(X_B)}} = \frac{D_{AB}}{\sqrt{p_A(1-p_A)p_B(1-p_B)}}$$

- It is usually more common to consider the r_{AB} value squared:

$$r^2_{AB} = \frac{D^2_{AB}}{p_A(1-p_A)p_B(1-p_B)}$$

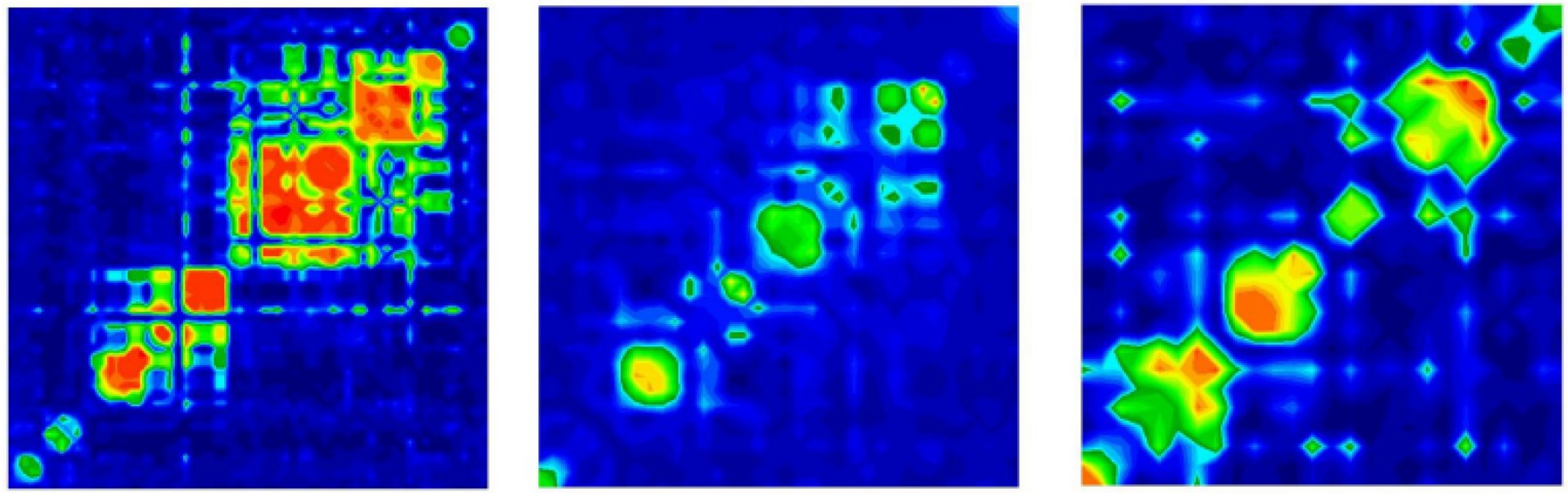
Properties for r^2

- Ranges between 0 and 1
  - 1 means two markers provide identical information, referred to as Perfect LD
  - 0 means two markers are in Perfect Equilibrium
- Raw r^2 from CHR22



Dawson et al, *Nature*, 2002

Linkage Disequilibrium in Three Regions



2q13
(63 markers)

13q13
(38 markers)

14q11
(26 markers)

Abecasis et al, *Am J Hum Genet*, 2001

What is Association Studies?

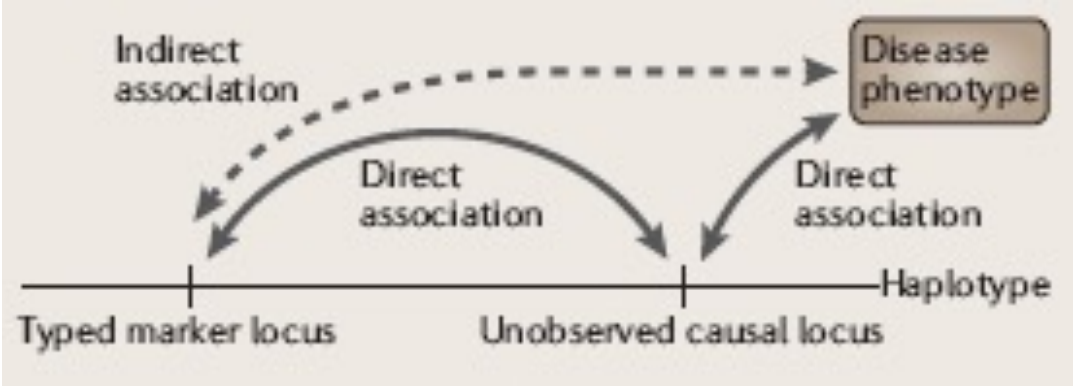
- Test associations between markers/SNPs/genes and the trait of interest

- Population Data
  - Contingency table tests (dichotomous trait, e.g., case/control)
  - Regression model based tests

- Family Data
  - Parametric Linkage Analysis

Why LD is Important for Association Studies?

- SNPs in strong LD with disease variant are good proxies for disease variant



Balding, 2006

- If testing (unobservable) disease variant for association would yield chi-squared statistic X^2, testing variant in LD yields r^2X^2
- Model LD in association studies based on a joint multivariate model to improve power

Association Studies

- Population-based vs. family-based
- Phenotype(s) of interest
  - Dichotomous trait, e.g., case/control
  - Quantitative trait
- Number of markers tested
  - May range from 1 to >10 million!
  - Candidate gene study
  - Genome-wide association study (GWAS)

Dichotomous Traits

- Compare frequencies of particular alleles, or genotypes, in set of cases and controls
  - Chi-square test using contingency tables
  - Logistic regression model based test

Chi-Square genotype test using contingency tables

- Test whether the trait and genotype are independent
- For example, we observe

	AA	Aa	aa
Control	n_{00}	n_{01}	n_{02}
Case	n_{10}	n_{11}	n_{12}

- Is the observation is significantly different from what we would expect if trait and the genotype are independent?
- H_0: No association between the trait and the genotype, i.e., sample with genotype AA, Aa, or aa have the same probability to develop disease (to be a Case)

Recall Chi-square test

- Let O_{ik} denote the count in the i^{th} row and k^{th} column in the contingency table, and E_{ik} denote the corresponding expected count
- Test statistic

$$X^2 = \sum_{i=1}^{I}\sum_{k=1}^{K}\frac{\left(O_{ik} - E_{ik}\right)^2}{E_{ik}} \sim \chi^2_{(I-1)(K-1)}$$

- If study sample size is large, the test statistic follows a chi-squared distribution with degrees of freedom (I - 1)(K - 1) under H_0 , with I rows and K columns in the contingency table.

How to Calculate E_{ik}?

- For example, we observe

	AA	Aa	aa	Row Total
Control	n_{00}	n_{01}	n_{02}	$n_{0.}$
Case	n_{10}	n_{11}	n_{12}	$n_{1.}$
Column Total	$n_{.0}$	$n_{.1}$	$n_{.2}$	n

- Observed number of samples per cell: $O_{ik} = n_{ik}$; with disease status (D) i = 0, 1; genotype (G) k = 0, 1, 2
- Expected number of samples per cell: $E_{ik} = n_{i.} n_{.k}/n$, under independence of disease status and genotype, $np_{D=i}p_{G=k} = n(n_{i.}/n)(n_{.k}/n)$
- Population genotype frequency: $p_{AA} = n_{.0}/n$; $p_{Aa} = n_{.1}/n$; $p_{aa} = n_{.2}/n$
- Expected number of samples per cell: $E_{i0}=n_{i.}\ p_{AA}$; $E_{i1}=n_{i.}\ p_{Aa}$; $E_{i2}=n_{i.}\ p_{aa}$;

Example of Genotypic Association Test

- TCF7L2 for Type 2 Diabetes in Finns
- SNP rs12255372 has alleles T and G

$$X_G^2 = \sum_{i=0,1} \sum_{j=0,1,2} (O_{ij} - E_{ij})^2 / E_{ij}$$

	GG	GT	TT	Total
Case	661	255	20	936
Control	724	354	50	1128
Total	1385	609	70	2064

$$X_G^2 = (661 - 628.08)^2/628.08 + \ldots \approx 14.08 \quad \sim \quad \chi^2,\ df = 2$$

$$p = .0009$$

Contingency Tables under Dominant (risk allele A) or Recessive (risk allele a) Disease Model

	AA or Aa	aa	Row Total
Control	$n_{00} + n_{01}$	n_{02}	$n_{0.}$
Case	$n_{10} + n_{11}$	n_{12}	$n_{1.}$
Column Total	$n_{.0} + n_{.1}$	$n_{.2}$	n

- H_0: No association between the trait (to be a Case or Control) and the genotype being AA/Aa or aa
- Chi-square test statistic:

$$X_D^2 = \sum_{i=0,1} \sum_{j=0,1} (O_{ij} - E_{ij})^2 / E_{ij} \quad \sim \quad \chi^2,\ df = 1$$

Example for Testing Dominant or Recessive Disease Model

- TCF7L2 for Type 2 Diabetes in Finns
- SNP rs12255372 has alleles T and G
- Allele T is dominant to G

$$X_D^2 = \sum_{i=0,1} \sum_{j=0,1} (O_{ij} - E_{ij})^2 / E_{ij} \quad \sim \quad \chi^2,\ df = 1$$

	GG	GT+TT	Total
Case	661	255+20=275	936
Control	724	354+50=404	1128
Total	1385	609+70=679	2064

$$X_D^2 \approx 9.60 \quad \sim \quad \chi^2,\ df = 1$$

$$p = .0019$$

Contingency Table for Allelic Association Test

	A	a	Row Total
Control	$n_{0A}=2n_{00}+n_{01}$	$n_{0a}=n_{01}+2n_{02}$	$2n_{0.}$
Case	$n_{1A}=2n_{10}+n_{11}$	$n_{1a}=n_{11}+2n_{12}$	$2n_{1.}$
Column Total	$n_{.A}=2n_{.0}+n_{.1}$	$n_{.a}=n_{.1}+2n_{.2}$	$2n$

- **Assume additive disease model**
- **Assume HWE**
- H_0: No association between the trait (sample to be a Case or Control) and the number of allele A in the sample genotype

$$X_L^2 = \sum_{i=0,1} \sum_{j=0,1} (O_{ij} - E_{ij})^2 / E_{ij} \quad \sim \quad \chi^2,\ df = 1$$

Example of Allelic Association Test

- TCF7L2 for Type 2 Diabetes in Finns
- SNP rs12255372 has alleles T and G

	G	T	Total
Case	1577	295	1872
Control	1802	454	2256
Total	3379	749	4128

$$X_L^2 \approx 13.13 \quad \sim \quad \chi^2,\ df = 1$$

$$p = .0003$$

Measure of Association Strength: Odds Ratio

	Exposed (E)	Not Exposed ($\bar{E}$)
Case (D)	a	b
Control ($\bar{D}$)	c	d

Odds ratio:

$$OR = \frac{P(D|E)/P(\bar{D}|E)}{P(D|\bar{E})/P(\bar{D}|\bar{E})} = \frac{P(E|D)/P(\bar{E}|D)}{P(E|\bar{D})/P(\bar{E}|\bar{D})} = ad/bc$$

- Exposed = carry certain genotype
- Counts pertain to individuals, not alleles.

Odds Ratio

Genotype Model ($\bar{E}$=aa)

	AA	Aa	aa
Case	n_{10}	n_{11}	n_{12}
Control	n_{00}	n_{01}	n_{02}

$$OR_{het} = (n_{11}n_{02})/(n_{01}n_{12})$$
$$OR_{hom} = (n_{10}n_{02})/(n_{00}n_{12})$$

Dominant Model ($\bar{E}$=aa)

	AA or Aa	aa
Case	$n_{10} + n_{11}$	n_{12}
Control	$n_{00} + n_{01}$	n_{02}

$$OR_D = [(n_{10} + n_{11})n_{02}]/[(n_{00} + n_{01})n_{12}]$$

Allele Model ($\bar{E}$=a)

	A	a
Case	$2n_{10} + n_{11}$	$n_{11} + 2n_{12}$
Control	$2n_{00} + n_{01}$	$n_{01} + 2n_{02}$

$$OR_L = [(2n_{10} + n_{11})(n_{01} + 2n_{02})]/[(2n_{00} + n_{01})(n_{11} + 2n_{12})]$$

Logistic Regression Model

- Y = dichotomous phenotype
- X = a coding for the genotype

Genotype	Codominant	Dominant	Recessive	Additive
AA	$X = (0,1)^{\mathrm{T}}$	$X = 1$	$X = 1$	$X = 2$
Aa	$X = (1,0)^{\mathrm{T}}$	$X = 1$	$X = 0$	$X = 1$
aa	$X = (0,0)^{\mathrm{T}}$	$X = 0$	$X = 0$	$X = 0$

Assume a logistic regression model:

$$\log\left[\frac{\Pr(Y=1|X)}{\Pr(Y=0|X)}\right] = \beta_0 + \alpha C + \beta_1 X$$

where β_0 is the intercept, α is the coefficient for covariates C, and β_1 is the genetic effect-size (i.e., log(Odds-Ratio)).

$$H_0 : \beta_1 = 0$$

$$H_a : \beta_1 \neq 0$$

Test Statistic

- Wald Test: $Z = \frac{\widehat{\beta_1}}{Standard\_Error(\widehat{\beta_1})} \sim N(0,1)$ under H_0
- Chi-square Test: $X^2 = \frac{\widehat{\beta_1}^2}{Var(\widehat{\beta_1})} \sim Chi\_Square$ with df=1 under H_0
- How to obtain p-value?

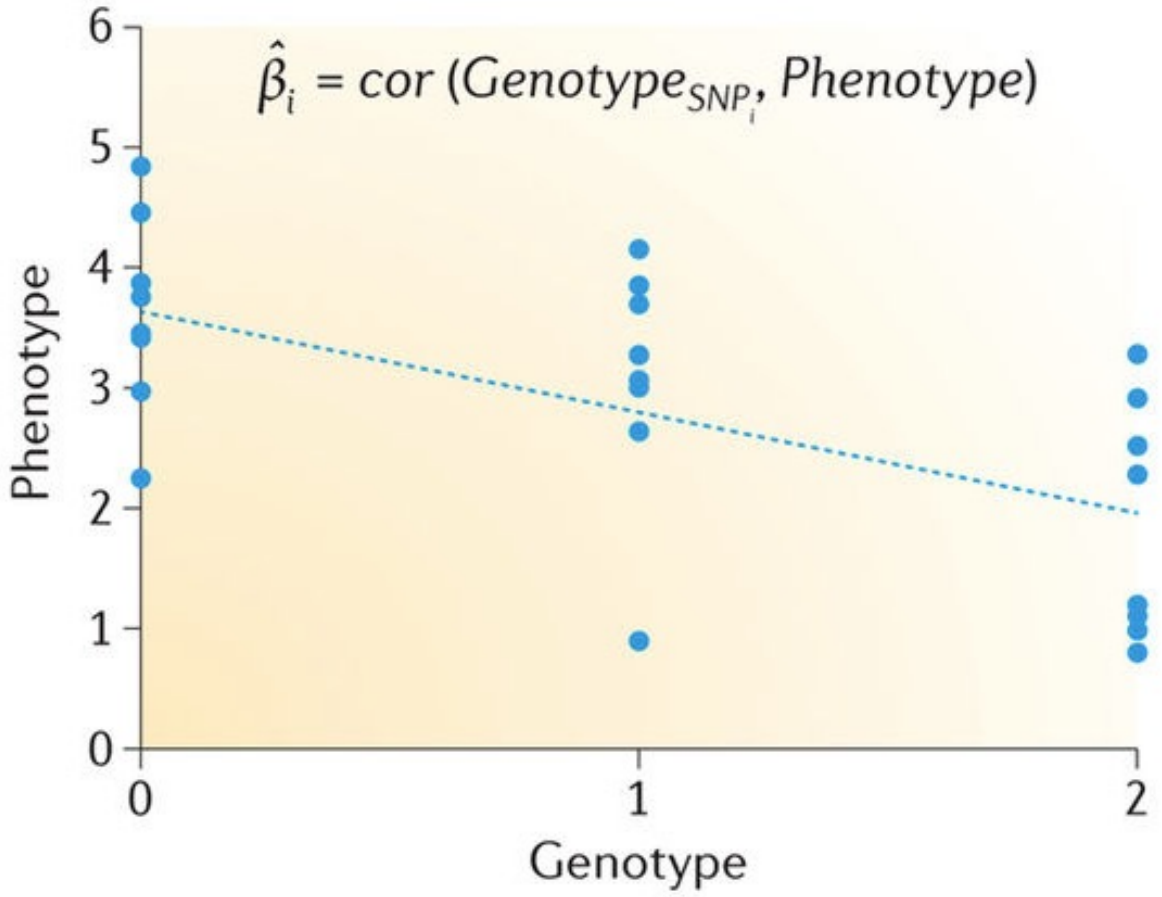
Advantages of Logistic Regression Model

- Account for confounding covariates (C), e.g., age, gender, BMI, smoking
- Flexible for various genetic models
- Flexible for testing multiple markers in the same model (modeling LD)
- Equivalent to the corresponding Chi-square test using contingency tables, if not modeling covariates
- Allow gene-environment interactions
- Without the assumption of HWE

Quantitative Trait

- Linear regression model
  - $Y = \beta_0 + \alpha C + \beta_1 X + \epsilon,\ \epsilon \sim N(0, \sigma^2)$
  - Y represents the quantitative trait values
  - X represents the genotype data (0, 1, 2) for additive genetic model
  - C represents the confounding covariates or other environmental variables
  - ϵ represents the error term, other unknown factors
- $H_0: \beta_1 = 0\ ; H_a: \beta_1 \neq 0$
- P-values can be obtained by Wald Test

Linear Regression Model



Genome-wide Association Study (GWAS)

GWAS: independent single-variant tests across all genome-wide variants

- Quality control (QC) of the study dataset
- Choose a model/test for the phenotype of interest (e.g., linear regression model for quantitative traits, logistic regression model for dichotomous traits, other association tests from previous lecture)
- Significance level $\alpha = 5 \times 10^{-8}$
- Report nearby genes of significant SNPs

Genotype Quality

Data quality is one of the key factors affecting the validity of findings.

Example factors affecting genotype quality:

- Quality of DNA samples, depending on the sample source (e.g., blood, buccal swab, spit kit)
- Handling and storage of the sample (e.g., sample contamination)
- Genotyping platforms/chips
- Sequence errors
- Variant calling

Quality Control

- Filter SNPs
  - Marker genotyping missing rate (e.g., $> 2\%$)
  - Mapping quality for sequence data (based on mapping quality scores)
  - Hardy-Weinberg Equilibrium (HWE) Testing (e.g., p-value $< 10^{-6}$)
  - MAF (e.g., $< 5\%$)
  - Control sample reproducibility
  - Mendelian Errors (e.g., $> 1\%$ families, or > 5 errors) for family-based studies
- Filter samples
  - Sex inconsistencies and chromosomal anomalies
  - Relatedness for population-based studies (how to quantify relatedness given genotype data?)
  - Ethnicity
  - Sample genotyping efficiency/call rate (e.g., $< 98\%$)

Additional Factors for GWAS: Batch effects, Population Stratification

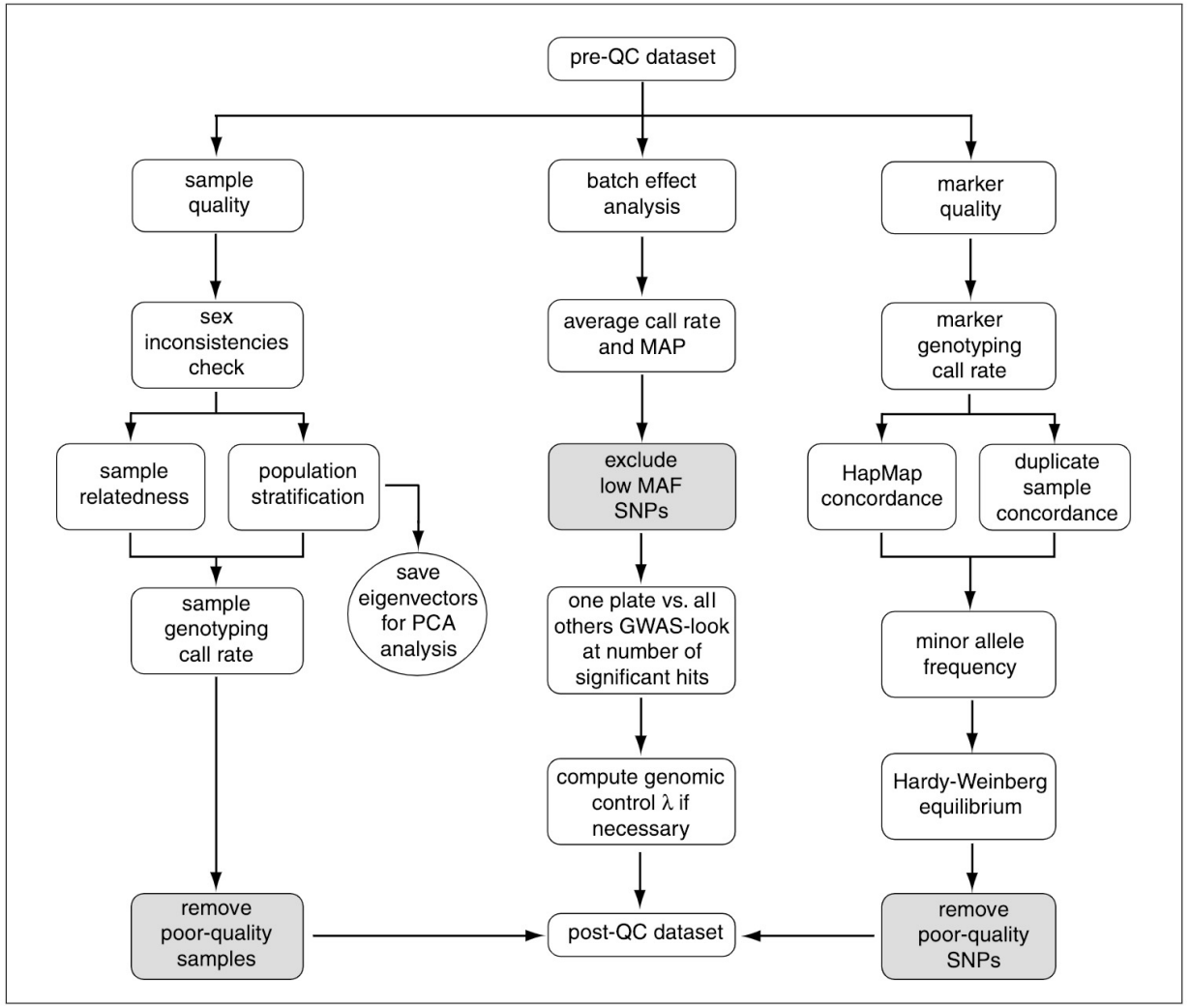
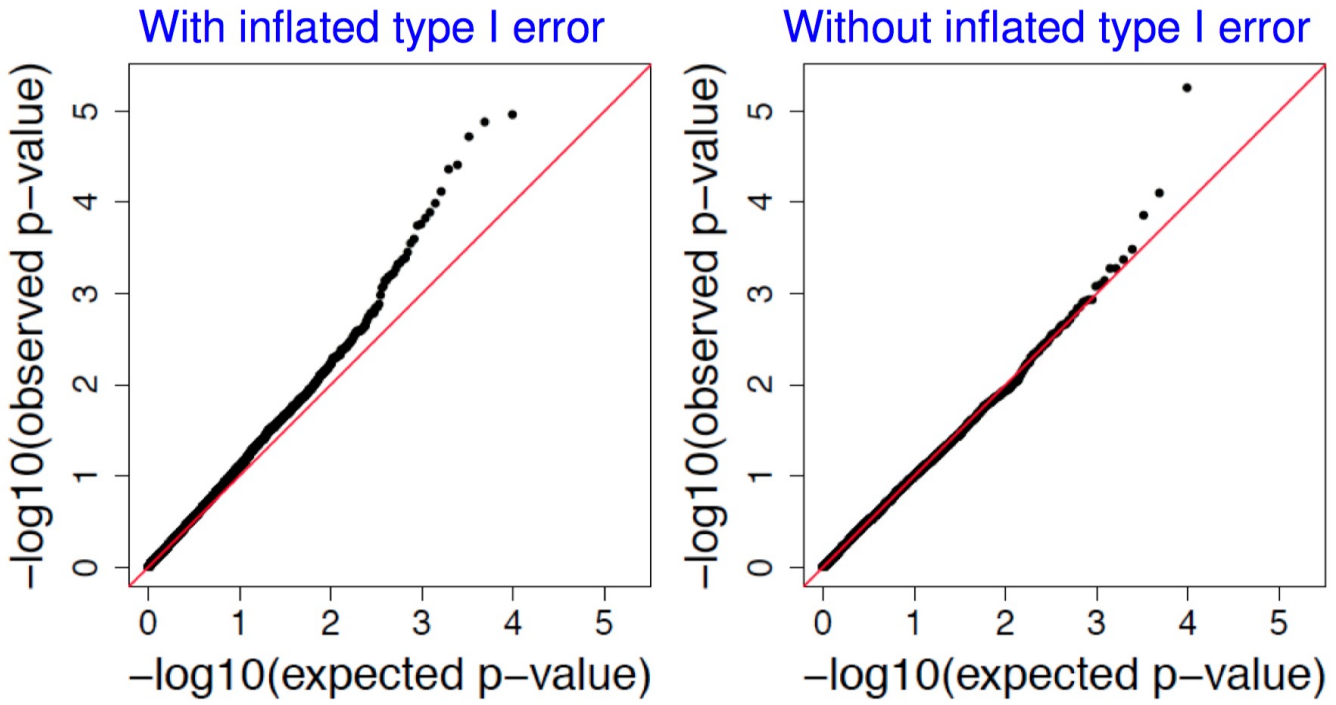


Figure 1.19.1 A flowchart overview of the entire GWAS QC process. Each topic is discussed in detail in the corresponding section in the text. Squares represent steps, ovals represent input or output data, and trapezoids represent filtering of data.

Check GWAS Results by Quantile-Quantile (QQ) Plot

– Obtained $-\log 10$(p-values) from GWAS
– Sort all $-\log 10$(p-values) from most significant to least
– Pair these with the expected values of order statistics of a Uniform$(0, 1)$ distribution
– Under NULL hypothesis (no association), p-values follow a Uniform$(0, 1)$ distribution

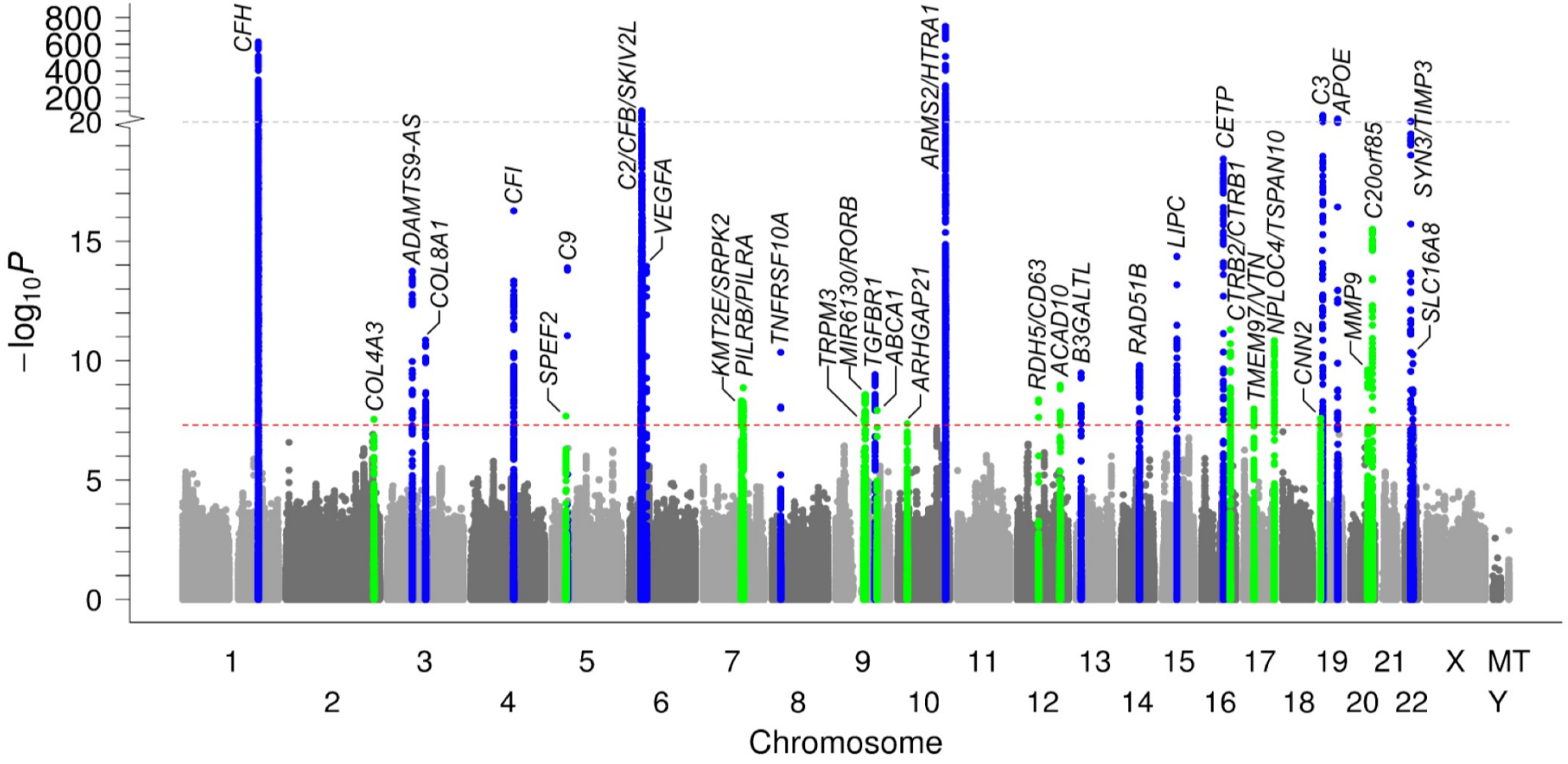


Visualize GWAS Results by Manhattan Plot

– Scatter plot of $-\log 10(\text{p-values})$ across all genome-wide variants

– Visualize signal peaks

GWAS Results

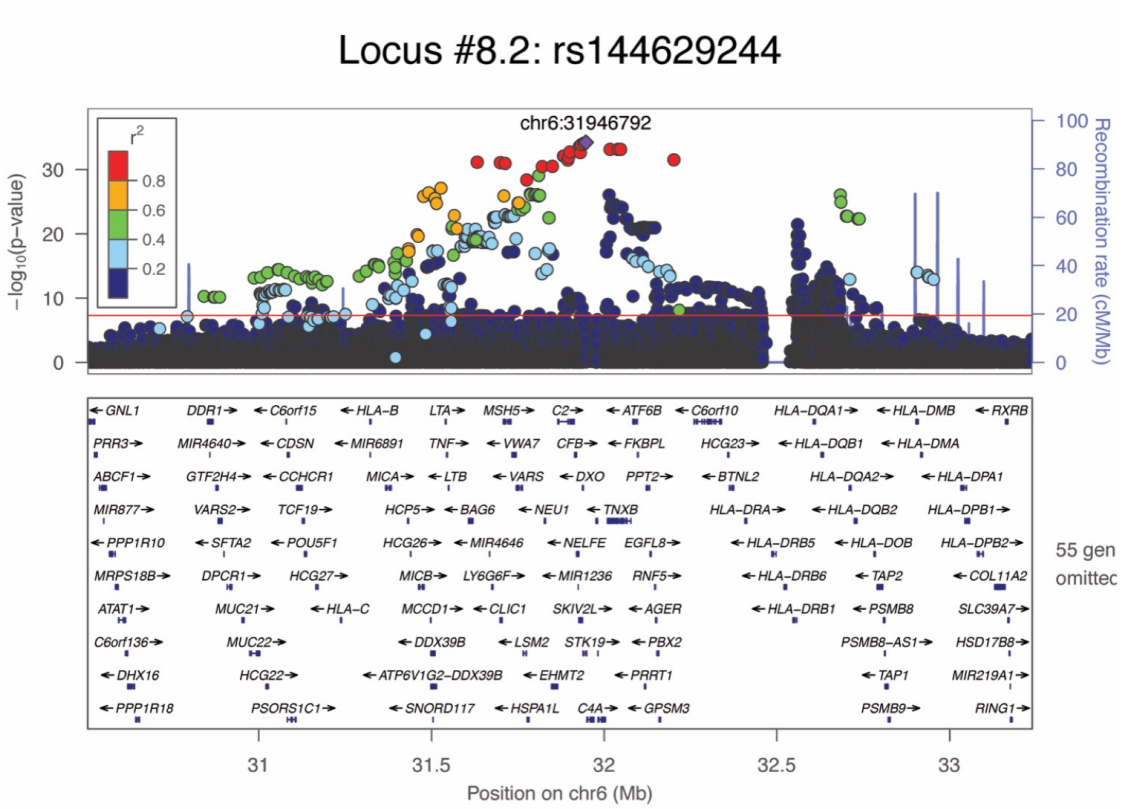


Fritsche L.G. et al. Nat Genet, 2016.

18 known AMD loci and **16 novel AMD loci**

Visualize GWAS Loci by Locus Zoom Plot

– Zoom into the peak region with gene annotations: http://locuszoom.org/
– Visualize r^2 between the specified significant (purple diamond) signal and its neighbor SNPs
– Visualize recombination rate



Visscher P.M. et al. AJHG 2017.

GWAS Catalogue Results

Example GWAS Discoveries

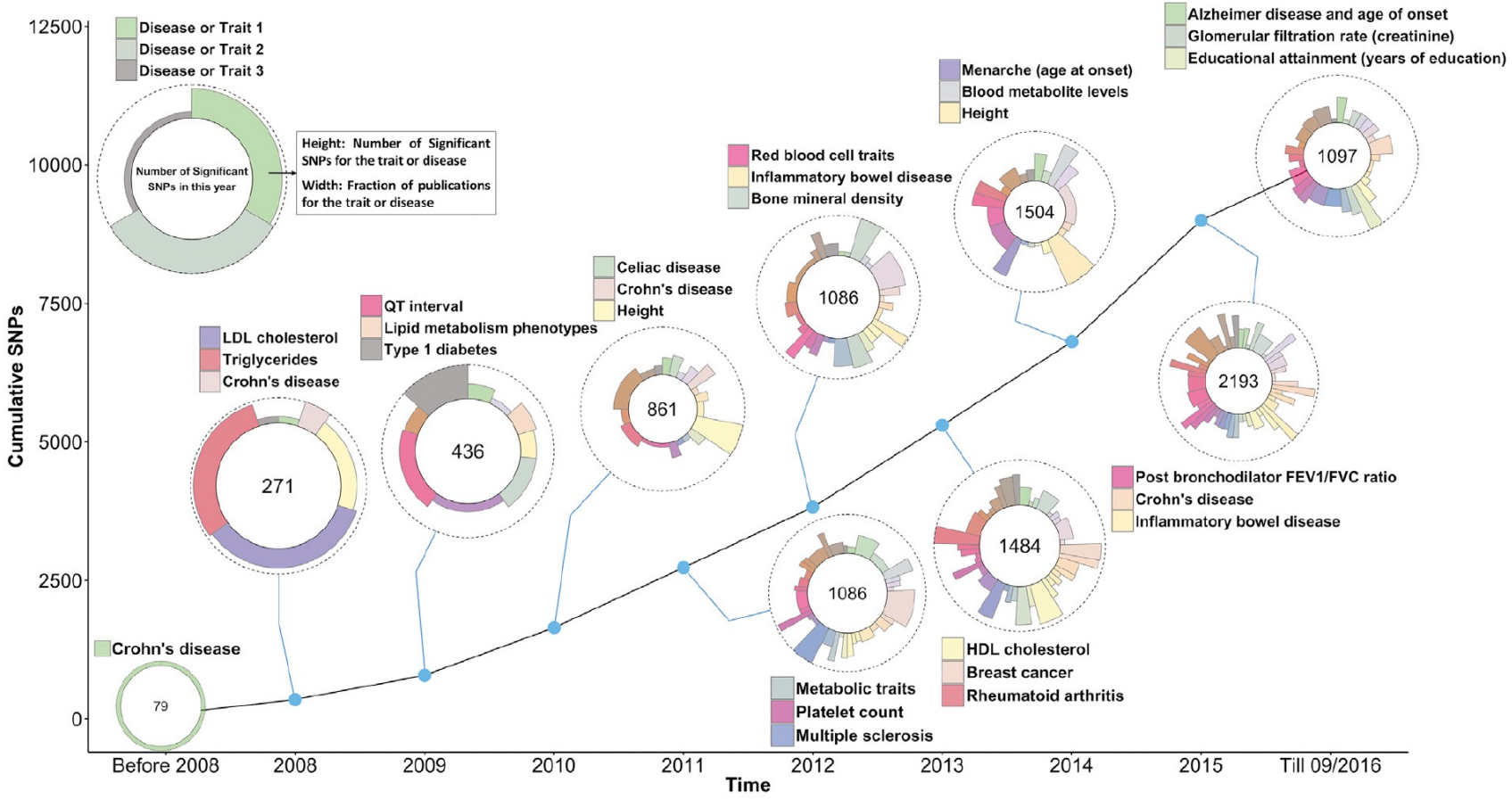
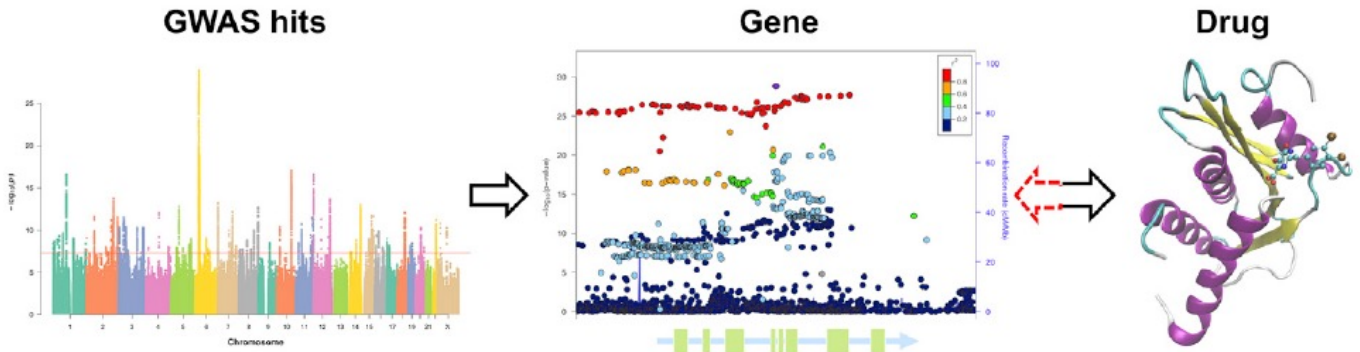


Figure 2. GWAS SNP-Trait Discovery Timeline

Data used for generating the graph were taken from the GWAS Catalogue.[10] SNPs and traits were selected according to the following filters. SNPs were selected with a p value $< 5 \times 10^{-8}$. For each trait with two or more selected SNPs, SNPs were removed if they had an LD $r^2 > 0.5$ (calculated from 1000 Genomes phase 3 data) with another selected SNPs and their p value was larger. For each year of discovery, only the top three traits and diseases with the largest number of SNPs are labeled in the circle.

Visscher P.M. et al. AJHG 2017.

Example links between GWAS discoveries and drug developments



Trait	Gene with GWAS hits	Known or candidate drug
Type 2 Diabetes	*SLC30A8/KCNJ11*	ZnT-8 antagonists/Glyburide
Rheumatoid Arthritis	*PADI4/IL6R*	BB-Cl-amidine/Tocilizumab
Ankylosing Spondylitis(AS)	*TNFR1/PTGER4/TYK2*	TNF-inhibitors/NSAIDs/fostamatinib
Psoriasis(Ps)	*IL23A*	Risankizumab
Osteoporosis	*RANKL/ESR1*	Denosumab/Raloxifene and HRT
Schizophrenia	*DRD2*	Anti-psychotics
LDL cholesterol	*HMGCR*	Pravastatin
AS, Ps, Psoriatic Arthritis	*IL12B*	Ustekinumab

Visscher P.M. et al. AJHG 2017.

Available Tools

- Association Study Tool
  - PLINK: https://www.cog-genomics.org/plink/2.0/
  - EPACTS: https://genome.sph.umich.edu/wiki/EPACTS
- LocusZoom Plot Tool
  - http://locuszoom.org/

Next Lecture

- **Homework Assignment**
  - Read the “Prioritizing GWAS Results: A Review of Statistical Methods and Recommendations for Their Application”, Cantor R.M. et al. AJHG, 2010 paper to answer given questions in your own words
  - Written answers (no more than 2 pages) are due by the beginning of next lecture
  - Email to jingjing.yang@emory.edu

- Topics for Next Lecture
  - Population Stratification
  - Meta-analysis
  - Family-based Association Test
  - Discuss homework questions