Analysis of Genetic Association Studies

References:

Zintzaras E. The power of generalized odds ratio in assessing association in genetic studies with known mode of inheritance. J Appl. Stat. 2012. DOI:10.1080/02664763.2012.722611

Zintzaras E, Santos M. Performance of MAX test and degree of dominance index in predicting the mode of inheritance. Stat Appl Genet Mol Biol. 2012 Jun 27;11(4):Article 4. doi: 10.1515/1544-6115.1804. PubMed PMID: 22745022.

Zintzaras E, Santos M. Estimating the mode of inheritance in genetic association studies of qualitative traits based on the degree of dominance index. BMC Med Res Methodol. 2011 Dec 21;11(1):171. [Epub ahead of print] PubMed PMID: 22188898.

Zintzaras E, Santos M, Szathmary E. Selfishness versus functional cooperation in a stochastic protocell model. J Theor Biol. 2010 Dec 21;267(4):605-13. Epub 2010 Sep 17. PubMed PMID: 20837027.

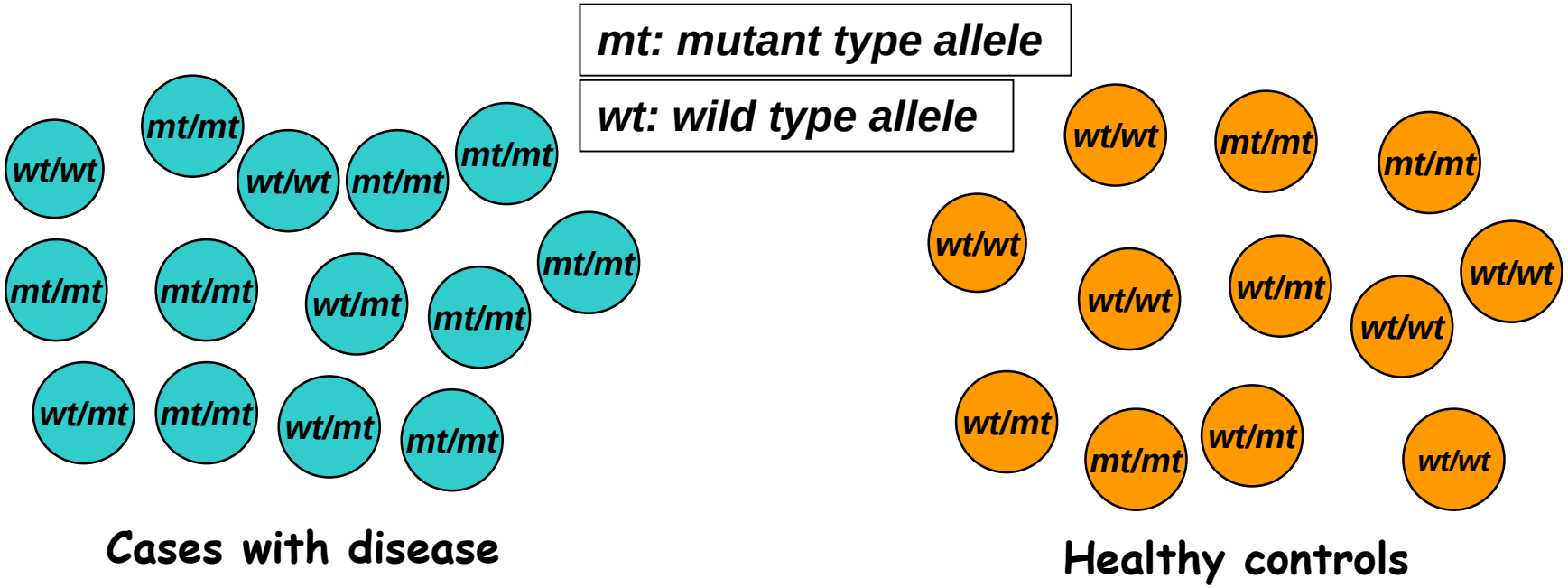
Zintzaras E. The generalized odds ratio as a measure of genetic risk effect in the analysis and meta-analysis of association studies. Stat Appl Genet Mol Biol 2010; Vol. 9: Iss. 1, Article 21.

Genetics background

- **Genetics is the science that studies the heredity of traits**
- **Genetic information is contained in DNA which consists of nucleotides**
- **Gene is a sequence of nucleotides that translates a protein**
- **Genes (through proteins) determine traits (phenotypes)**
- **A gene may have different forms called alleles**
- **An allele can be mutant type-*mt (change in nucleotudes*) or wild type-*wt***
- **For each gene there are two alleles due to diploidy of humans (homologous chromosome pairs)**
- **In an individual the genotype distribution of gene can be homozygous (*wtwt* or *mtmt*) or heterozygous (*wtmt*)**
- **The multiple alleles of a gene is called polymorphism or variant (preserved mutations), they usually expressing different phenotypes**

Genetic association studies (GAS)

The evaluation of possible associations between phenotypic traits (diseases) and genetic variants (gene polymorphisms) is carried out using GAS



In the case of a genetic variant with two alleles (mutant type-*mt* and wild type-wt), where *mt* is thought to be associated with a disease, GAS will collect information on the numbers of diseased subjects and control subjects with each of the three genotypes (*wt/wt*, *wt/mt*, *mt/mt*)

Study quality assessment

- **Prior to testing the association, the quality of a study should be assessed**
- **A study quality surrogate point is whether the controls conform with the HWE rule in the controls**
- **Lack of HWE implies: genotyping errors and/or structure in the population (i.e. non-unselected controls)**

Departure for HWE is tested using a x^2-test.

Genotype	Cases with CAD	Controls
mt/mt	1788	*874*
mt/wt	4145	*2165*
wt/wt	2328	*1335*

In HWE, the genotype distribution should follow the following rule:

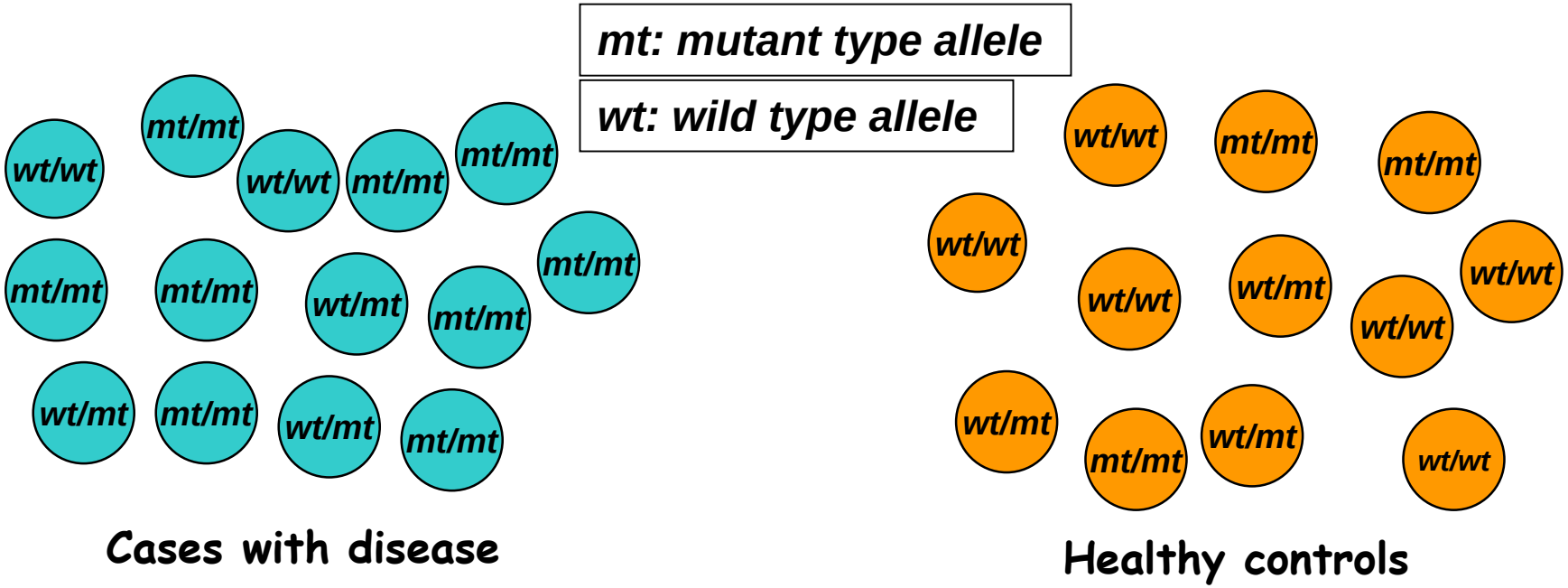
$$wt/wt : mt/wt : mt/mt = p^2 : 2pq : q^2$$

where p is the frequency of the wt alleles and
q=1-p is the frequency of the mt alleles.

P=0.944

P>0.05, thus the controls are in HWE.

The URL http://www.had2know.com/academics/hardy-weinberg-equilibrium-calculator-2-alleles.html **provides a calculator for testing HWE**



- **Association is tested using a x^2-test and the magnitude of association is expressed in terms of odds ratio (OR)**

- ***Example:*** **In a GAS with 8261/4374 cases/controls, the association between *ACE D/I* (*wt/mt*) and CAD investigated. The genotype distribution was as follows:**

Genotype	Cases with CAD	Controls
mt/mt	**1788**	***874***
mt/wt	**4145**	***2165***
wt/wt	**2328**	***1335***

The association between disease status and the genetic variant is tested using a chi-squared (x^2) test with (3-1)x(2-1)=2 df

- ***Example:*** **In a GAS with 8261/4374 cases/controls, the association between *ACE D/I* (*wt/mt*) and CAD investigated. The genotype distribution was as follows:**

Genotype	Cases with CAD	Controls
mt/mt	1788	*874*
mt/wt	4145	*2165*
wt/wt	2328	*1335*

To perform a chi-squared test, access the URL: http://www.quantpsy.org/chisq/chisq.htm, enter the data of the GAS in the cells, and type "Calculate"

P-value=0.010

There is significant association between *ACE D/I* gene variant and development of CAD

When the association is significant, various genetic models of genotypes are tested by omitting or merging genotypes

These models include:

- **additive model:**
  homozygous for mt vs. homozygous for wt

- **co-dominant model:**
  heterozygous vs. all homozygotes

The significance of the genetic model is assessed using the respective odds ratio (*OR*) and its 95% confidence interval (CI).

Additive model:

The *OR* for the additive model is

$$OR = \frac{\text{" probability" a subject of being diseased when mtmt}}{\text{" probability" a subject of being diseased when wtwt}}$$

For *OR>1*: an *mtmt* subject has greater chance of being diseased than a *wtwt* subject

If the 95% CI does not include 1, then, the *OR* is significant (P<0.05) (i.e. the variant is associated with the disease).

Additive model:

Genotype	Cases with CAD	Controls
mt/mt	1788	*874*
wt/wt	2328	*1335*

$$OR = \frac{\text{" probability" a subject of being with CAD when mtmt}}{\text{" probability" a subject of being with CAD when wtwt}} = \frac{1788/1335}{874/2328} = 1.17$$

$$95\%CI = \left(e^{\ln(OR) - 1.96*\sqrt{\frac{1}{1788}+\frac{1}{2328}+\frac{1}{874}+\frac{1}{1335}}}, e^{\ln(OR) + 1.96*\sqrt{\frac{1}{1788}+\frac{1}{2328}+\frac{1}{874}+\frac{1}{1335}}}\right) = (1.06, 1.30)$$

- ***Since "1" is not included in the 95% CI, we conclude that the OR is significant (P<0.05).***

- ***Since OR>1, we conclude that homozygous for* the *mt* allele have 17% greater risk for CAD than homozygous for the wt allele**

Co-dominant model:

The *OR* for the co-dominant model is

$$OR = \frac{\text{" probability" a subject of being diseased when wtmt}}{\text{" probability" a subject of being diseased when else}}$$

For *OR>1*: an *wtmt* subject has greater chance of being diseased than an homozygous subject

If the 95% CI does not include 1, then, the *OR* is significant (P<0.05) (i.e. the variant is associated with the disease).

Co-dominant model:

Genotype	Cases with CAD	Controls
mt/wt	4145	*2165*
mt/mt+wt/wt	1788+2328=4116	874+1335=2209

$$OR = \frac{\text{" probability" a subject of being with CAD when mtwt}}{\text{" probability" a subject of being with CAD when wtwt + mtmt}} = \frac{4145/2165}{4116/2209} = 1.03$$

$$95\%CI = \left(e^{\ln(OR) - 1.96*\sqrt{\frac{1}{4145}+\frac{1}{4116}+\frac{1}{2165}+\frac{1}{2209}}}, e^{\ln(OR) + 1.96*\sqrt{\frac{1}{4145}+\frac{1}{4116}+\frac{1}{2165}+\frac{1}{2209}}}\right) = (0.96, 1.11)$$

- ***Since "1" is included in the 95% CI, we conclude that the OR is not significant (P≥0.05).***

Pharmacogenetic Studies

In the PG studies, we investigate the association of

i) a gene expression as a binary variable or

ii) a gene expression as a continuous variable or

iii) a gene polymorphism

and clinical outcome in patients after treatment with approved therapy

i) Example: MDR1 gene overexpression is considered to be a major cause of multidrug resistance and it is implicated in the response to chemotherapy in AML patients. In a PG study, the association of MDR1 gene expression and response to chemotherapy in patients with AML has been investigated. The results were as follows:

	mRNA expression levels
	mean±SD
Responders (N=37):	**1.4±2.7**
Non-responders (N=15):	**0.3±0.5**

Is the response to treatment associated with MDR1 mRNA expression levels?

We may use the online t-test at the URL

http://www.quantitativeskills.com/sisa/statistics/t-test.htm

for testing the equality of the two means.

P=0.057, indicating that MDR1 expression levels are different between Responders and Non-responders.

ii) Example: In a PG study, the association of MDR1 gene expression and response to chemotherapy in patients with AML has been investigated. The results were as follows:

	mRNA expression levels	
	+ve	***-ve***
Responders	**27**	**22**
Non-responders	**21**	**5**

Is the response to treatment associated with MDR1 mRNA expression levels?

The association between response to treatment and MDR1 mRNA expression levels is tested using a chi-squared test.

To perform the chi-squared test, access the URL: http://www.quantpsy.org/chisq/chisq.htm, enter the data of the PG study in the cells, and type "Calculate"

P=0.027

There is significant association between response to treatment and MDR1 mRNA expression levels

OR= **Prob. of being Responders when –ve mRNA expression/**
Prob. of being Responders when +ve mRNA expression

OR=(22/5)/(27/21)=3.42

The 95% CI is ($e^{\ln(OR)-1.96*SE}$, $e^{\ln(OR)-1.96*SE}$)

where SE=$\sqrt{}$(1/22 + 1/5 + 1/27 + 1/21)=0.575

Thus, the 95% CI is (1.11, 10.56)

iii) Example: In a PG study, the association of MDR1 C3435T gene polymorphism and response to chemotherapy in patients with AML has been investigated. The results were as follows:

	MDR1 C3435T genotype		
	CC	***CT***	***TT***
Responders	**158**	**65**	**39**
Non-responders	**13**	**18**	**9**

Is the response to treatment associated with the MDR1 C3435T gene polymorphism?

The association between response to treatment and MDR1 C3435T gene polymorphism is tested using a chi-squared test.

To perform the chi-squared test, access the URL: http://www.quantpsy.org/chisq/chisq.htm, enter the data of the PG study in the cells, and type "Calculate"

P=0.004

There is significant association between response to treatment and MDR1 C3435T gene polymorphism