

# Gene-Gene Interactions



Jurg Ott, PhD  
<http://genemapping.cn>

Rockefeller University  
New York  
ott@rockefeller.edu

Chinese Academy of Sciences  
Beijing Institute of Genomics  
ott@big.ac.cn

## Complex Traits: Multiple disease genes, possibly interacting



□ Unknown locations: Genome-wide screening with 100,000s of SNP markers.

# Epistatic Interactions

Ming & Muenke (2002) *Am J Hum Genet* 71, 1017 (review)

EFFECT AND PHENOTYPE	GENE 1		GENE 2	
	Mutation	Phenotype	Mutation	Phenotype
<b>Synergistic:</b>				
RP	<i>ROM1</i> <sup>+K80insG</sup>	Normal	<i>RDS</i> <sup>+L185P</sup>	Normal
RP	<i>ROM1</i> <sup>+L114insG</sup>	Normal	<i>RDS</i> <sup>+L185P</sup>	Normal
Bardet-Biedl	<i>BBS2</i> <sup>Y245/Q59X</sup>	Normal	<i>BBS6</i> <sup>+Q147X</sup>	Normal
Deafness	<i>GJB2</i> <sup>+I35delG</sup>	Normal	<i>GJB6</i> <sup>+/-</sup>	Normal
Deafness	<i>GJB2</i> <sup>+I67delT</sup>	Normal	<i>GJB6</i> <sup>+/-</sup>	Normal
Hirschsprung	<i>RET</i> <sup>+A647T</sup>	Normal	<i>EDNRB</i> <sup>+I5305N</sup>	Normal
Severe insulin resistance	<i>PPARG</i> <sup>+A553delAAAT</sup>	Normal	<i>PPP1R3A</i> <sup>+K1984delAG</sup>	Normal
<b>Modifier:</b>				
Juvenile-onset glaucoma	<i>MYOC</i> <sup>+G399V</sup>	Adult-onset glaucoma	<i>CYP11B1</i> <sup>+R368H</sup>	Normal
Usher 1	<i>USH3</i> <sup>mut/mt</sup>	Usher 3	<i>MYO7A</i> <sup>+delG (exon 25)</sup>	Normal
Congenital nonlethal JEB	<i>COL17A1</i> <sup>R1226X/L835X</sup>	Juvenile JEB	<i>LAMB3</i> <sup>+R635X</sup>	Normal
More severe ADPKD	<i>PKD1</i> <sup>+/mt</sup>	Less severe ADPKD	<i>PKD2</i> <sup>+/I152delA</sup>	Less severe ADPKD
More severe hearing loss	<i>DFNA1</i>	Mild hearing loss	<i>DFNA2</i>	Mild hearing loss
WS2/OA	<i>MITF</i> <sup>+/944delA</sup>	?WS2	<i>TYR</i> <sup>+R402Q</sup>	Normal
More severe WS2/OA	<i>MITF</i> <sup>+/944delA</sup>	?WS2	<i>TYR</i> <sup>R402Q/R402Q</sup>	Normal

# Dominant × Dominant Model

Disease susceptibility depends on both disease alleles (A, B). Assume model:

Penetrances	Locus 2		
	aa	Aa	AA
bb	0	0	0
Bb	0	f	f
BB	0	f	f

$P(A) = P(B) = 0.15$ .

Penetrance  $f = 0.50 = P(\text{aff}|\text{genotype pattern})$ .

Genotypes in HWE.

2-locus analysis yields much larger chi-square.

2-locus			1-locus		
pattern	$P(\text{pat} \text{aff})$	$P(\text{pat} \text{unaff})$	pattern	$P(\text{pat} \text{aff})$	$P(\text{pat} \text{unaff})$
A-B	1.000	0.039	A	0.278	0.069
other	0.000	0.961	other	0.723	0.931
sum	1	1	sum	1	1
chi-sq	$n = 1$	0.926	chi-sq	$n = 1$	0.076

P(affected genotypes)				
Penetrance	aa	Aa	AA	prob
bb	0	0	1	0.25
Bb	0	0.50	0	0.50
BB	1	0	0	0.25
prob	0.25	0.50	0.25	1

cases P(genotypes affected)				
Penetrance	aa	Aa	AA	prob
bb	0	0	0.2500	0.25
Bb	0	0.5000	0	0.50
BB	0.2500	0	0	0.25
prob	0.25	0.50	0.25	1

controls P(genotypes unaffected)				
Penetrance	aa	Aa	AA	prob
bb	0.083	0.167	0.000	0.25
Bb	0.167	0.167	0.167	0.50
BB	0.000	0.167	0.083	0.25
prob	0.25	0.50	0.25	1

**Pure Epistatic 2Locus Model**

Frankel & Schork (1996) *Nat Genet* **14**, 371

Marginal genotype distributions same for cases and controls and supercontrols.

*Note:* IBD sharing prob. slightly elevated

## Multi-Locus Approaches

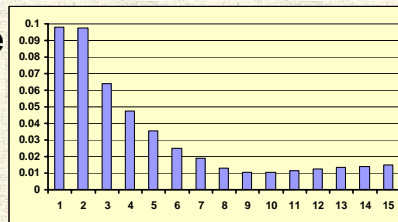
- ❑ Most case-control studies do not yet take into account the multi-locus nature of complex traits.
- ❑ Aim: Analyze multiple SNPs/genes jointly. Two classes of approaches:
  - Combine single-locus statistics over multiple SNPs (wherever they are in genome) as a substitute for multi-locus statistic
  - Look for patterns of genotypes at SNPs in different genomic locations

## Set Association Approach

Hoh *et al.* (2001) *Genome Res* **11**, 2115

<http://www.genemapping.cn/sumstat.html>

- At each SNP, compute association statistic,  $s$
- Build sum over 1, 2, 3, etc. highest  $s$  values
- Evaluate significance of given sum by permutation test
- Sum with smallest  $p$ -value  
→ marker selection
- Smallest  $p$  is single statistic, find sig. level.
- Applicable to many SNPs



## Genotype Patterns

- Marker-by-marker analysis does not exploit all information in data
- Search for genotype patterns: Difficult!

SNP 15    SNP 97    SNP 113517  
A G ----- T T ----- A C

- No easy solutions with 100,000 to 1,000,000 SNPs per individuals

# Genotype Pattern Search Methods

Hoh & Ott (2003) *Nat Rev Genet* 4, 701-709

- ❑ CPM = combinatorial partitioning method (Charlie Sing, U Michigan). Applicable to small number (~50) of SNPs only.
- ❑ MDR = multifactor-dimensionality reduction method (Jason Moore, Vanderbilt U)
- ❑ LAD = logical analysis of data (P. Hammer, Rutgers U)
- ❑ Mining association rules, *Apriori* algorithm (R. Agrawal)
- ❑ Special approaches for microarray data

## Aim: To Find Interaction Effects Associated with Disease

- How to test for epistatic effects above and beyond (independent of) main effects (of single-locus genotype effects)?

	Locus 2		
Locus 1	AA	AB	BB
AA			
AB			
BB			

Main effect locus 1	2 df
Main effect locus 2	2 df
<u>Interactions</u>	<u>4 df</u>
Total	8 df

- “Usual” chi-square for interactions independent of main effects. Isolate individual df’s?
- Difference in interactions between cases and controls? Interactions may indicate pathway



# Partitioning Chi-square

Agresti A (2002) *Categorical Data Analysis*, Wiley



Simple disease model, population frequency  $K = 0.10$   
 $N = 100$  cases, 100 controls.

Predicted numbers of cases and controls in given genotype classes,  
 and resulting odds ratios, OR

Recessive	<table style="width: 100%; border-collapse: collapse;"> <tr><th colspan="3">Penetrance</th></tr> <tr><th>AA</th><th>AB</th><th>BB</th></tr> <tr><td>0.05</td><td>0.05</td><td>0.50</td></tr> </table>	Penetrance			AA	AB	BB	0.05	0.05	0.50	cases	<table style="width: 100%; border-collapse: collapse;"> <tr><th>AA</th><th>BB</th></tr> <tr><td>22</td><td>55</td></tr> </table>	AA	BB	22	55	controls	<table style="width: 100%; border-collapse: collapse;"> <tr><th>AA, BB</th><th>AB</th></tr> <tr><td>78</td><td>22</td></tr> </table>	AA, BB	AB	78	22
	Penetrance																					
	AA	AB	BB																			
0.05	0.05	0.50																				
AA	BB																					
22	55																					
AA, BB	AB																					
78	22																					
<table style="width: 100%; border-collapse: collapse;"> <tr><td>0.05</td><td>0.50</td><td>0.50</td></tr> </table>	0.05	0.50	0.50	<table style="width: 100%; border-collapse: collapse;"> <tr><td>47</td><td>6</td></tr> </table>	47	6	<table style="width: 100%; border-collapse: collapse;"> <tr><td>53</td><td>47</td></tr> </table>	53	47													
0.05	0.50	0.50																				
47	6																					
53	47																					
<table style="width: 100%; border-collapse: collapse;"> <tr><td>0.05</td><td>0.15</td><td>0.45</td></tr> </table>	0.05	0.15	0.45	<table style="width: 100%; border-collapse: collapse;"> <tr><td>OR</td><td>0.05</td><td><b>19</b></td><td><b>1.03</b></td><td>0.97</td></tr> </table>	OR	0.05	<b>19</b>	<b>1.03</b>	0.97													
0.05	0.15	0.45																				
OR	0.05	<b>19</b>	<b>1.03</b>	0.97																		
<table style="width: 100%; border-collapse: collapse;"> <tr><td>0.05</td><td>0.15</td><td>0.45</td></tr> </table>	0.05	0.15	0.45	<table style="width: 100%; border-collapse: collapse;"> <tr><td>OR</td><td>0.05</td><td><b>19</b></td><td>0.05</td><td><b>18.4</b></td></tr> </table>	OR	0.05	<b>19</b>	0.05	<b>18.4</b>													
0.05	0.15	0.45																				
OR	0.05	<b>19</b>	0.05	<b>18.4</b>																		
<table style="width: 100%; border-collapse: collapse;"> <tr><td>0.05</td><td>0.15</td><td>0.45</td></tr> </table>	0.05	0.15	0.45	<table style="width: 100%; border-collapse: collapse;"> <tr><td>OR</td><td>0.06</td><td><b>15.6</b></td><td>0.46</td><td><b>2.16</b></td></tr> </table>	OR	0.06	<b>15.6</b>	0.46	<b>2.16</b>													
0.05	0.15	0.45																				
OR	0.06	<b>15.6</b>	0.46	<b>2.16</b>																		

# Partitioning Chi-square: 2 loci

	Gene 2		
Gene 1	AA	BB	AB
AA			
BB			
AB			

$3 \times 3$  table of genotypes (4 df) may be partitioned into 4 independent components, each with 1 df. Do such partitioning for cases and controls each.

	AA	BB
AA		
BB		

	AA, BB	AB
AA		
BB		

	AA	BB
AA, BB		
AB		

	AA, BB	AB
AA, BB		
AB		

## Testing Equality of ORs

Breslow NE, Day NE (1996) *Statistical Methods in Cancer Research* vol II. IARC Publications

Compare each of the four 2 × 2 sub-tables between cases and controls to see whether their odds ratios are the same.

		Gene 1			OR	Homogeneity
Gene 2		AA	BB	cases	1.84	chi-sq = 1.21
	AA			controls	0.74	<b>p = 0.27</b>
	BB					
		AA,BB	AB	cases	0.84	chi-sq = 1.63
AA,BB			controls	1.26	<b>p = 0.20</b>	
AB						
		AA,BB	AB	cases	1.27	chi-sq = 0.40
AA			controls	0.94	<b>p = 0.53</b>	
BB						
		AA	BB	cases	1.20	chi-sq = 0.89
AA,BB			controls	0.73	<b>p = 0.35</b>	
AB						

## Logistic Regression Model

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots$$

$p$  = probability of being affected;  $x_i$  are input variables like G1a and G1b shown below.

Genotype	G1a	G1b
AA	-1	-1
AB	0	+2
BB	+1	-1

# Logistic Regression Analysis

Parameter	p-value	
CONSTANT	0.0000	
G1a	0.0004	Gene 1 allelic effect
G1b	0.1054	Gene 1 heterozygote effect
G2a	0.0014	Gene 2 allelic effect
G2b	0.0114	Gene 2 heterozygote effect
G1a*G2a	0.2715	
G1a*G2b	0.3692	
G1b*G2a	0.3481	
G1b*G2b	0.4600	

Logistic regression analysis confirms: The two genes act individually, each contributing independently to hypertension.

## 2-Locus Genotype Patterns in Hypertension Data

Gene 1	Gene 2	cases	controls	sum	prop.	OR
AA	AA	25	3	28	0.043	7.35
	AB	41	21	62	0.096	1.70
	BB	32	19	51	0.079	1.43
AB	AA	33	15	48	0.074	1.91
	AB	67	82	149	0.230	0.60
	BB	66	56	122	0.189	0.97
BB	AA	14	9	23	0.036	1.30
	AB	43	46	89	0.138	0.74
	BB	33	42	75	0.116	0.61
	sum	354	293	647	1	

chi-square (8 df) = 35.44,  $p < 0.000001$

Genotype patterns contain information on main effects and interaction effects of varying orders.

## 2-Locus Genotype Patterns in Hypertension Data

Collapse patterns with frequencies < 0.10 into a "rare" class

Gene 1	Gene 2	cases	controls	sum	prop
AB	AB	67	82	149	0.230
AB	BB	66	56	122	0.189
BB	AB	43	46	89	0.138
BB	BB	33	42	75	0.116
AA	AB	41	21	62	0.096
rare (< 0.10)		104	46	150	0.232
total		354	293	647	1.000

chi-square (5 df) = 27.35, p = 0.000049

## Acknowledgments

- Discussions with Prof. Yaning Yang
- Collaborators in New York:
  - Chunsheng He, PhD
  - Haiyan Xu, PhD (and many others)
- Beijing:
  - Qingrun Zhang
  - Quan Long, PhD