

Genome-wide Association Data and Gene-gene Interaction

Jurg Ott

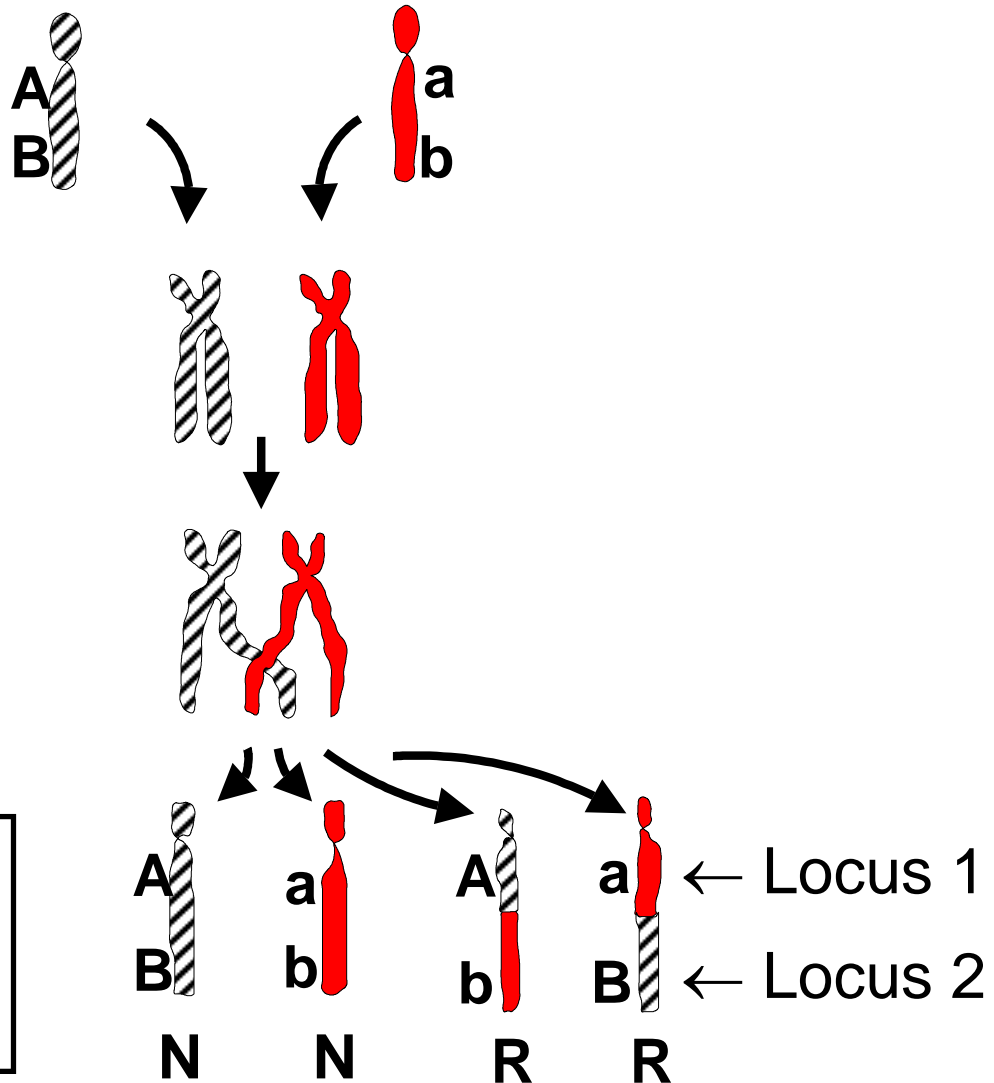
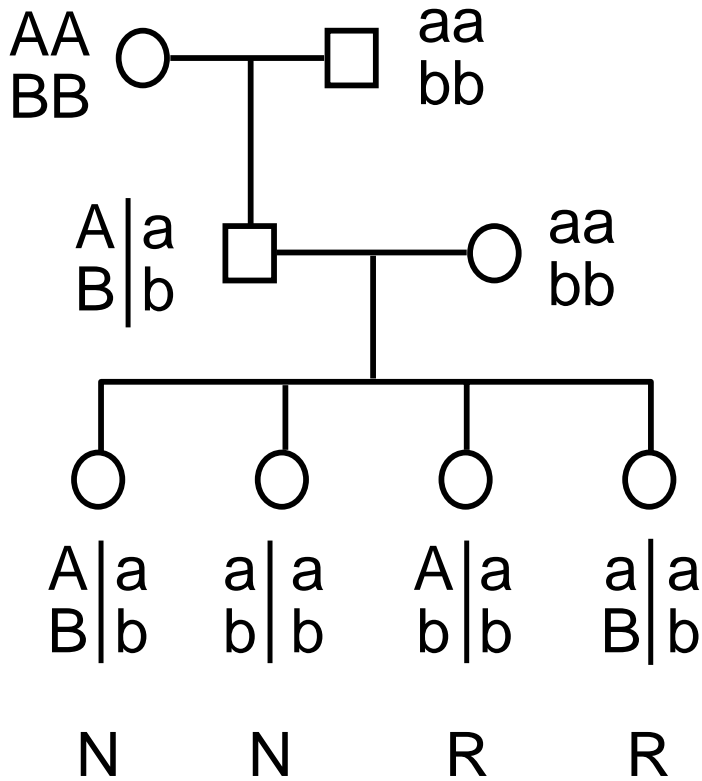
Rockefeller University, New York

ott@rockefeller.edu

Heritable Diseases

- **Rare Diseases** (prevalence $< 1\%$)
 - Mendelian inheritance
 - Examples: Huntington disease, cystic fibrosis
- **Common Diseases**
 - Non-mendelian (“complex”) mode of inheritance. Examples: Diabetes, schizophrenia.
 - Genetically relevant phenotype often unclear
 - Multiple underlying susceptibility genes

Genetic Linkage Analysis



Aim: Determine recombination fraction θ between two loci, where $\theta < 50\%$ \rightarrow linkage.

Linkage Disequilibrium (LD) Genetic Association

<i>Gene</i>	<i>SNP</i>	
A	T	0.30
A	C	0.70



G	T	1
A	T	0.30
A	C	0.70

- Population expands
→ >1 disease allele, **G**
- Crossovers → chromosomes with **G** - **C** alleles
- Motivates case-control studies

	T	C
G	1	0
A	0.30	0.70

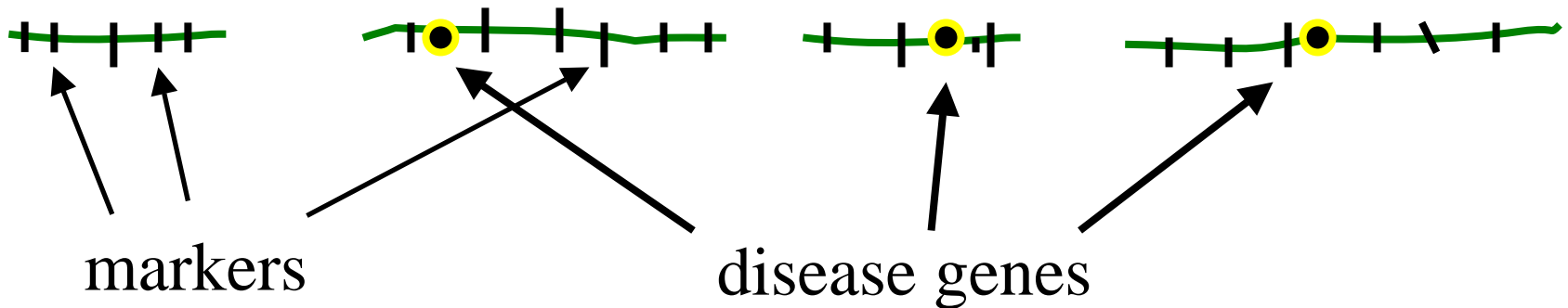
Origin of LD: Idealized Situation!

- Population with small number of founder individuals, rapidly expanding → strong LD.
- Most disease genes show multiple mutations (alleles), having occurred at different times → strength of LD (measured by D') reduced.

- LD is the basis for association studies.

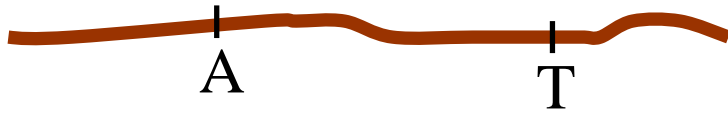
	T	C
Cases	0.95	0.05
controls	0.30	0.70

Genetic Marker Loci on Human Chromosomes



- Candidate genes: Focus on specific regions
- Unknown locations: Genome-wide screening with up to 800 microsatellites, or 100,000s of SNP markers.

Measuring the Extent of LD

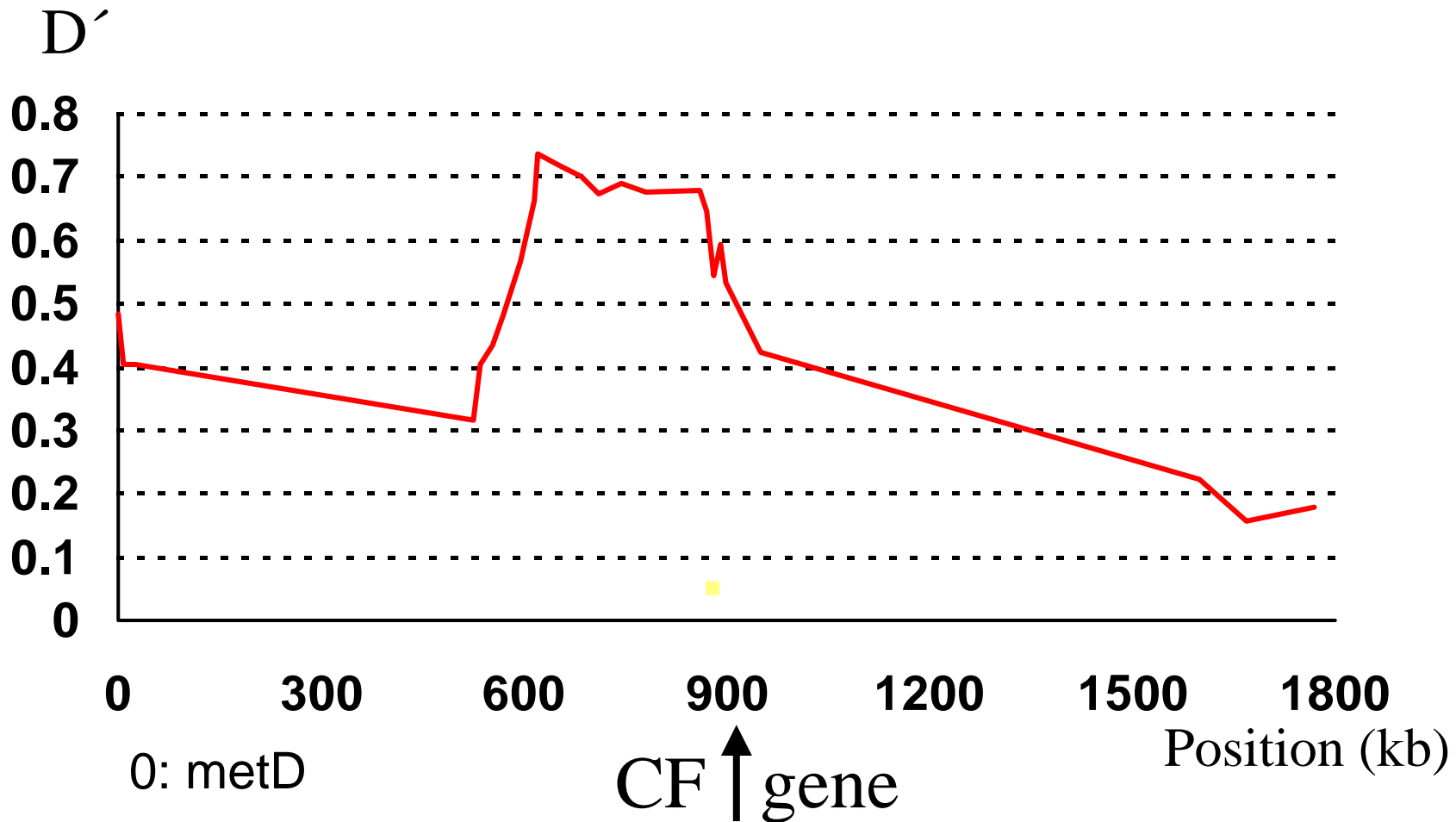


	T	T ^c
A	P(AT)	...
A ^c

- Two alleles, one each at two loci on same chromosome (same haplotype).
- Independence: $P(AT) = P(A) \times P(T)$
- Dependence: $P(AT) = P(A) \times P(T) + D$
- D = disequilibrium parameter, min. and max. values given by allele frequencies.
- $D' = D/D_{\max}$ ranges between 0 and 1.

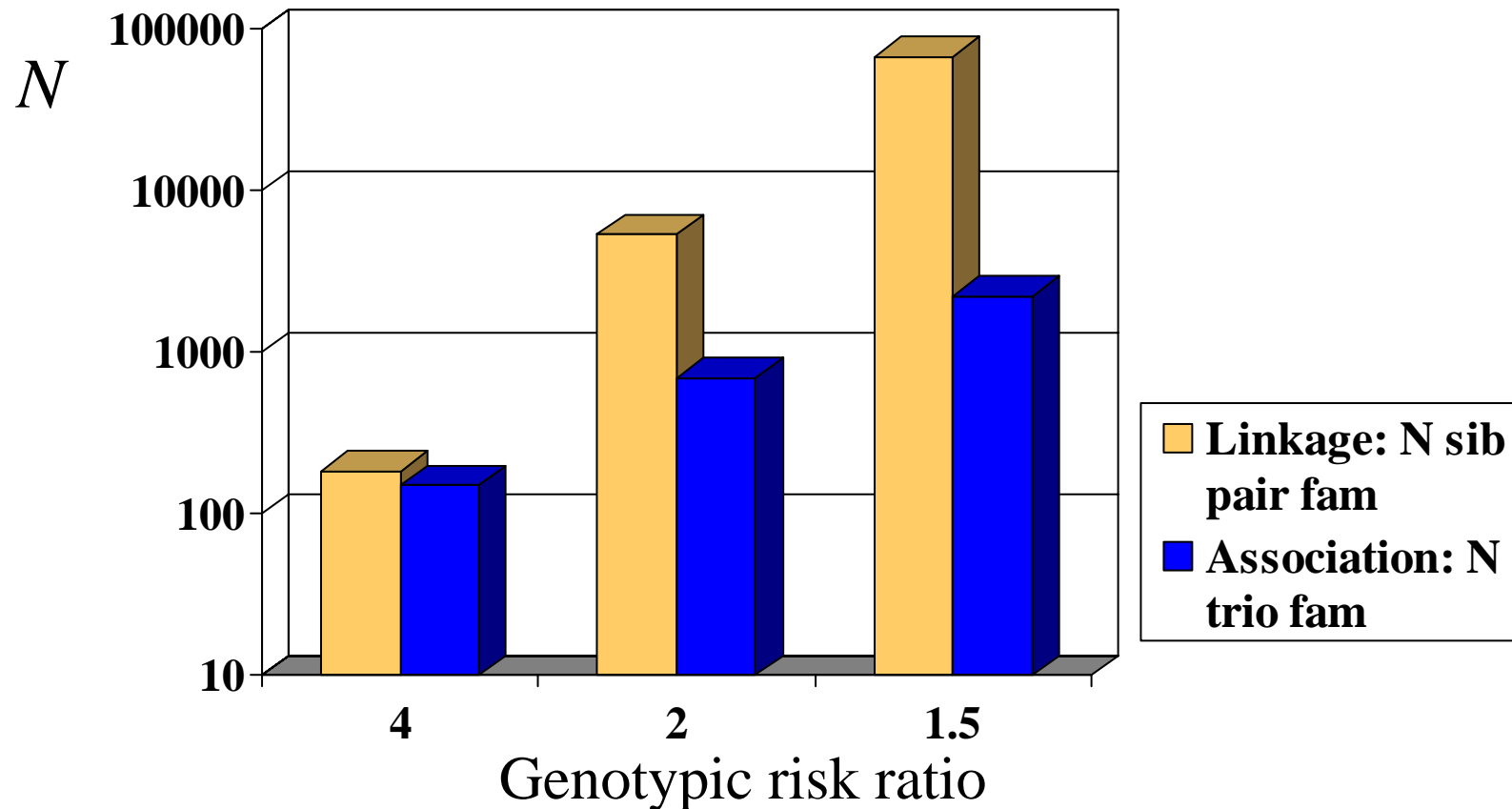
Linkage Disequilibrium at Cystic Fibrosis Locus

Kerem et al (1989) *Science* **245**, 1073



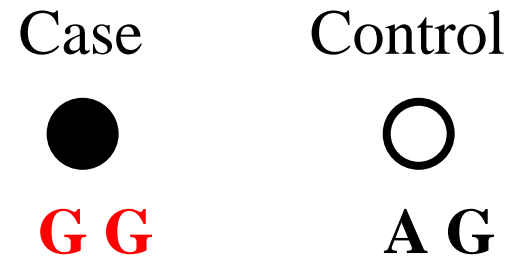
Association Analysis More Powerful Than Linkage Analysis for Weak Loci

Risch & Merikangas (1996) *Science* **273**, 1516

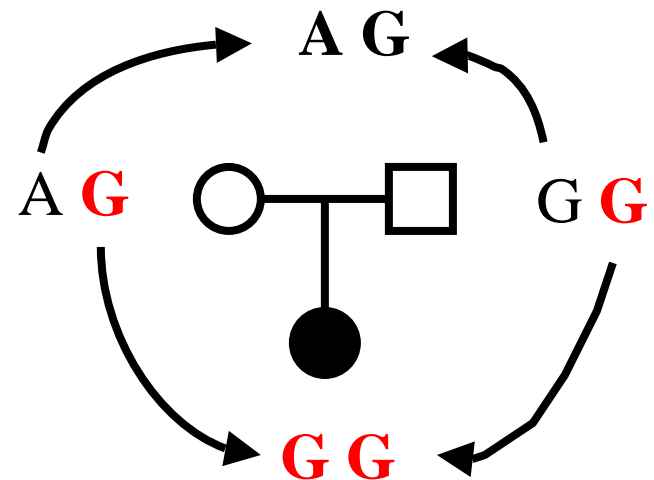


Case-Control Association Studies

Population controls: Easy to collect, efficient yet prone to population stratification (problems can be overcome).



versus



Family based controls:
Compare alleles transmitted to affected child with those not transmitted.

Single-Marker Statistics

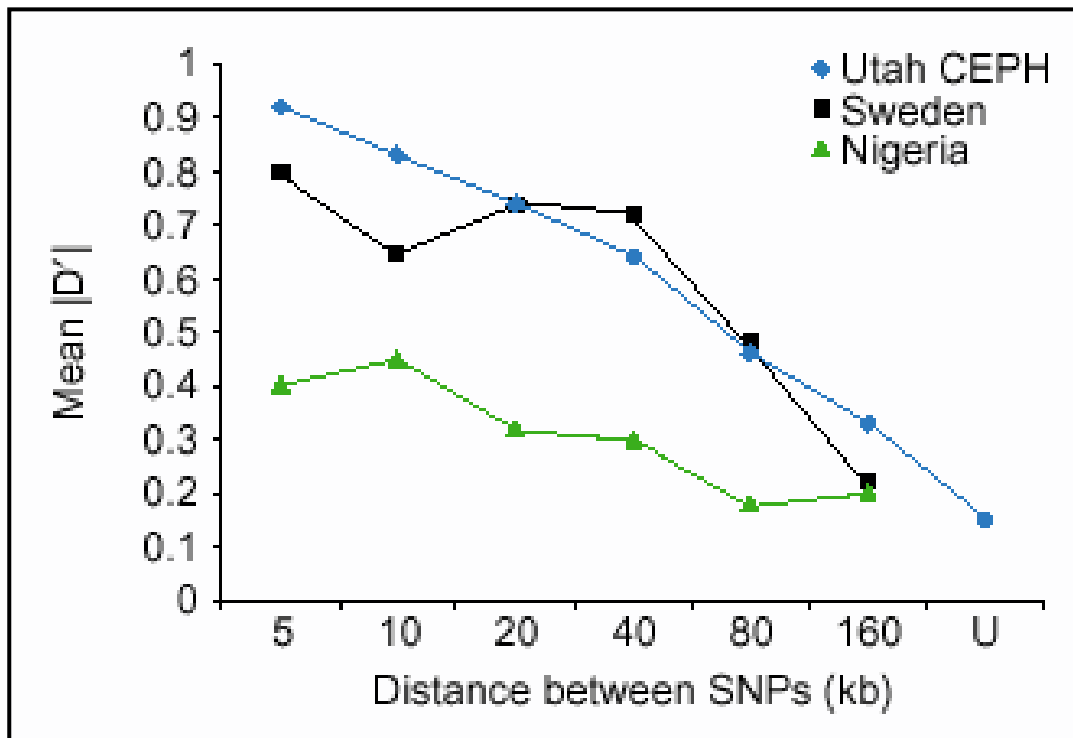
Most genome screens evaluated on a marker-by-marker basis.

	A	G		AA	AG	GG
Cases	or
Controls

Size of χ^2 shows significance of association

LD versus Physical Distance

Weiss & Clark (2002) *Trends in Genetics* **18**, 19

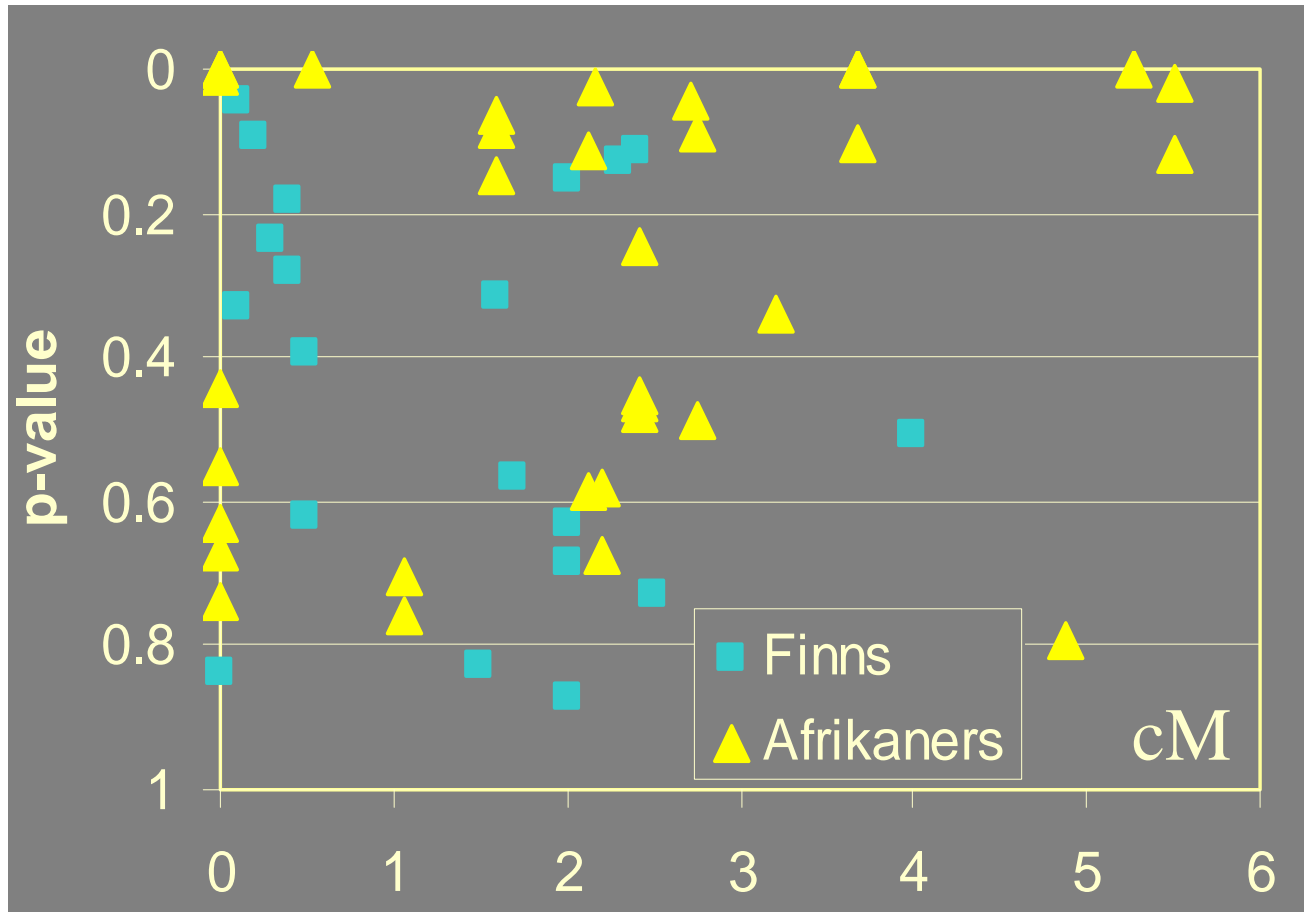


Pairs of SNPs:
Mean LD for 48 individuals from Utah and Sweden and 96 individuals from Nigeria. U = unlinked. Data from Reich et al. (2001) *Nature* **411**, 199

LD in South Africa and Finland

Gordon et al. (2000) *Genomics* **66**:87-92 (Afr. data)

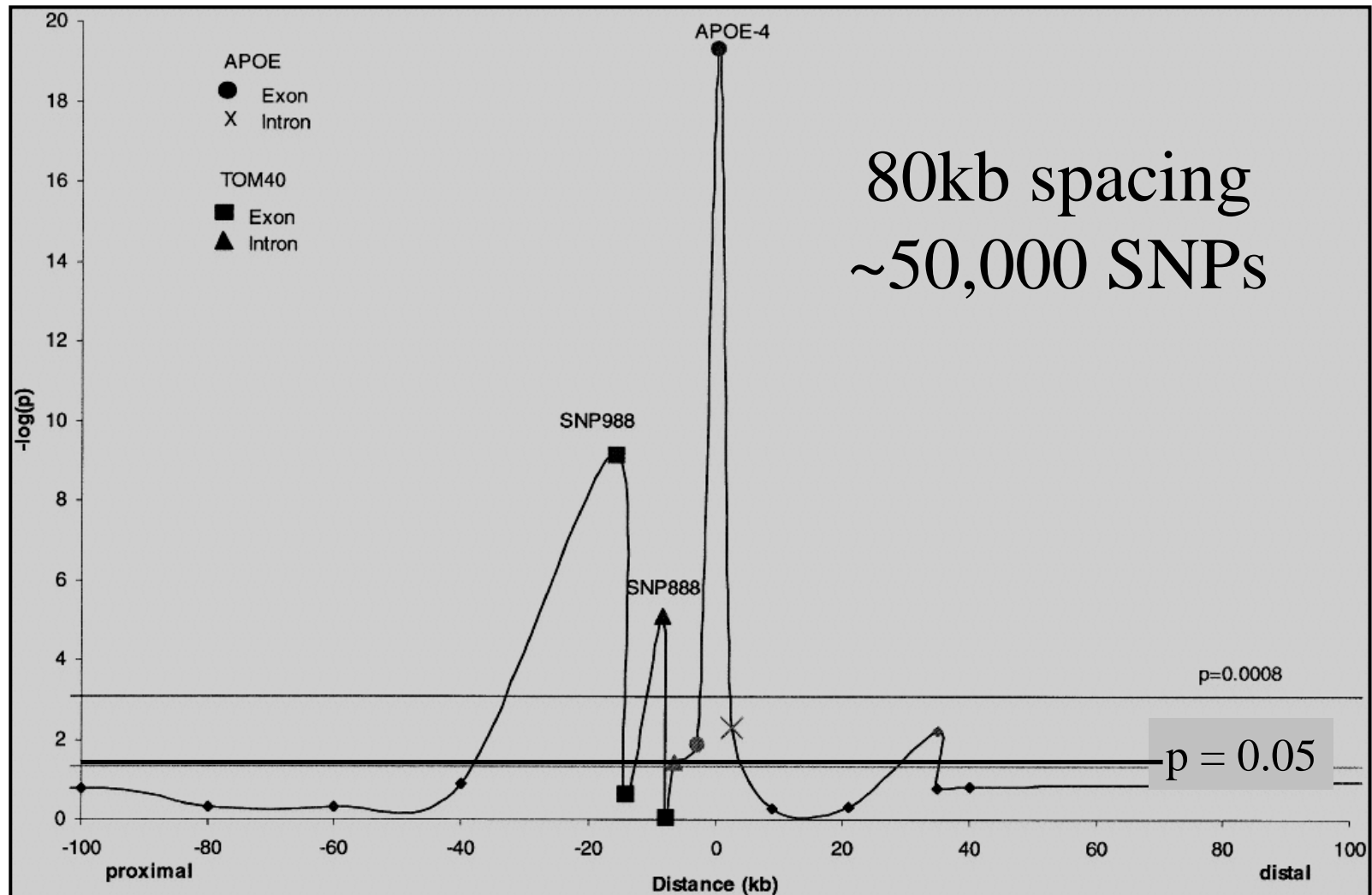
Laan & Pääbo (1997) *Nat Genet* 17:435 (Finnish data)



Strength of LD (p -value) versus map distance between pairs of microsatellite markers. Based on ~90 healthy individuals in each population.

LD Around Alzheimer Disease Gene

Martin et al. (2000) *Am J Hum Genet* 67, 383



Functional SNPs in the lymphotoxin- α gene that are associated with susceptibility to myocardial infarction

Kouichi Ozaki¹, Yoza Ohnishi¹, Aritoshi Iida², Akihiko Sekine², Ryo Yamada³, Tatsuhiko Tsunoda⁴, Hiroshi Sato⁵, Hideyuki Sato⁵, Masatsugu Hori⁵, Yusuke Nakamura^{2,6} & Toshihiro Tanaka¹

By means of a large-scale, case-control association study using 92,788 gene-based single-nucleotide polymorphism (SNP) markers, we identified a candidate locus on chromosome 6p21 associated with susceptibility to myocardial infarction. Subsequent linkage-disequilibrium (LD) mapping and analyses of haplotype structure showed significant associations between myocardial infarction and a single 50 kb haplotype comprised of five SNPs in *LTA* (encoding lymphotoxin- α), *NFKBIL1* (encoding nuclear factor of κ light polypeptide gene enhancer in B cells, inhibitor-like 1) and *BAT1* (encoding HLA-B associated transcript 1). Homozygosity with respect to each of the two SNPs in *LTA* was significantly associated with increased risk for myocardial infarction (odds ratio = 1.78, $\chi^2 = 21.6$, $P = 0.00000033$; 1,133 affected individuals versus 1,006 controls). *In vitro* functional analyses indicated that one SNP in the coding region of *LTA*, which changed an amino-acid residue from threonine to asparagine (Thr26Asn), effected a twofold increase in induction of several cell-adhesion molecules, including VCAM1, in vascular smooth-muscle cells of human coronary artery. Moreover, the SNP in intron 1 of *LTA*, enhanced the transcriptional level of *LTA*. These results indicate that variants in the *LTA* are risk factors for myocardial infarction and implicate *LTA* in the pathogenesis of the disorder.

First genome-wide screen with large number ($n = 92,788$) of SNPs

Complement Factor H Polymorphism in Age-Related Macular Degeneration

Robert J. Klein,¹ Caroline Zeiss,^{2*} Emily Y. Chew,^{3*}
Jen-Yue Tsai,^{4*} Richard S. Sackler,¹ Chad Haynes,¹
Alice K. Henning,⁵ John Paul SanGiovanni,³ Shrikant M. Mane,⁶
Susan T. Mayne,⁷ Michael B. Bracken,⁷ Frederick L. Ferris,³
Jurg Ott,¹ Colin Barnstable,² Josephine Hoh⁷†

Age-related macular degeneration (AMD) is a major cause of blindness in the elderly. We report a genome-wide screen of 96 cases and 50 controls for polymorphisms associated with AMD. Among 116,204 single-nucleotide polymorphisms genotyped, an intronic and common variant in the complement factor H gene (*CFH*) is strongly associated with AMD (nominal *P* value $<10^{-7}$). In individuals homozygous for the risk allele, the likelihood of AMD is increased by a factor of 7.4 (95% confidence interval 2.9 to 19). Resequencing revealed a polymorphism in linkage disequilibrium with the risk allele representing a tyrosine-histidine change at amino acid 402. This polymorphism is in a region of *CFH* that binds heparin and C-reactive protein. The *CFH* gene is located on chromosome 1 in a region repeatedly linked to AMD in family-based studies.

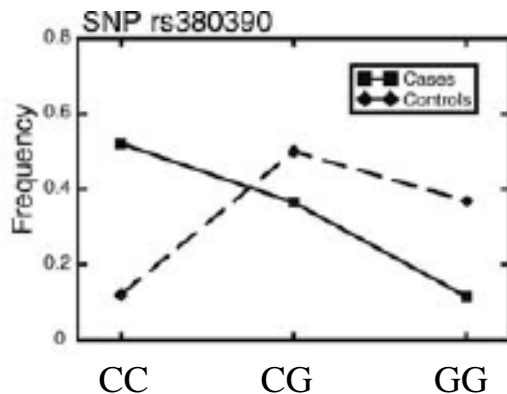


New genome-wide screen, $n = 116,204$ SNPs, strong association of a variant and AMD.



Highlights of AMD Study

- Cases ($n = 96$) and controls ($n = 50$) carefully selected to be extreme
- All of $>100,000$ SNPs used
- Most significant SNP is intronic! Significance depends on mutational “history”, allele frequency, and closeness to functional SNP.



Attribute
 Risk allele
 Allelic association χ^2 nominal P value
 Odds ratio (dominant) (95% CI)
 PAR (95% CI)
 Frequency in HapMap CEU
 Odds ratio (recessive) (95% CI)
 PAR (95% CI)
 Frequency in HapMap CEU

rs380390 (C/G)
 C
 4.1×10^{-8}
 4.6 (2.0–11)
 70% (42–84%)
 0.70
 7.4 (2.9–19)
 46% (31–57%)
 0.23

Epistatic Interaction

- “Complex traits due to multiple interacting genes”
- Most analyses disregard multi-locus nature of complex traits
- Real-life examples of epistatic traits? (next slide)
- Special analysis methods (tomorrow)

Early Example of Interaction of Two Sites

Hum Mutat. 1997;9(5):437-44.



Two mutations in the same low-density lipoprotein receptor allele act in synergy to reduce receptor function in heterozygous familial hypercholesterolemia.

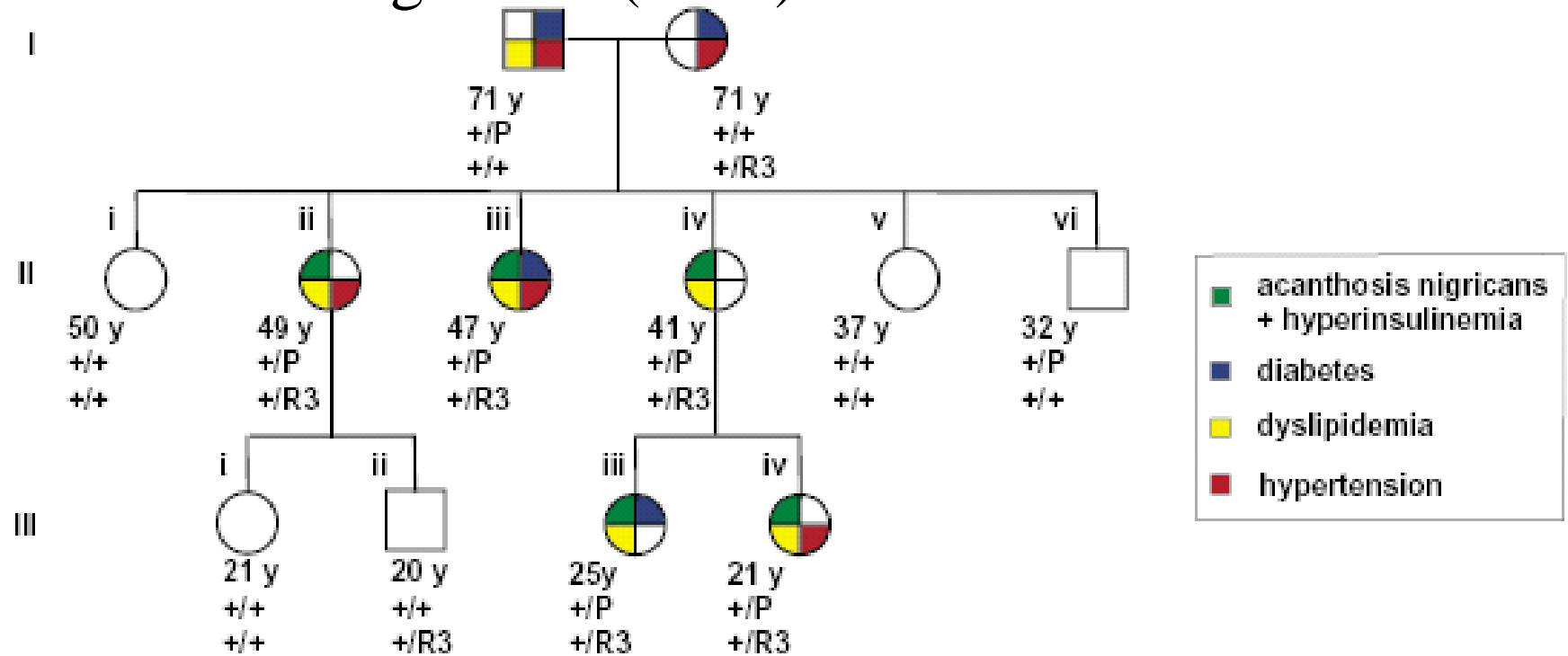
Jensen HK, Jensen TG, Faergeman O, Jensen LG, Andresen BS, Corydon MJ, Andreasen PH, Hansen PS, Heath F, Bolund L, Gregersen N.

Center for Medical Molecular Biology, Skejby Sygehus University Hospital, Aarhus, Denmark.

... two families with familial hypercholesterolemia ... low-density lipoprotein (LDL) receptor gene. His mutation (N543H) in exon 11, and an in-frame 9-bp deletion (2393del9) in exon 17. The two mutations were identified in heterozygous FH index patients in whom no other pathogenic mutations were detected by SSCP analysis of the remaining 16 exons and the promoter region. Both mutations cosegregated with hypercholesterolemia within the families. Each of these mutations had little or no effect on receptor function in transfected COS cells, but when both mutations were present simultaneously, receptor function ... was reduced by 75%. ... synergistic action of these two LDL receptor mutations.

Digenic Inheritance of Severe Insulin Resistance

Savage et al. (2002) *Nat Genet* 31:379



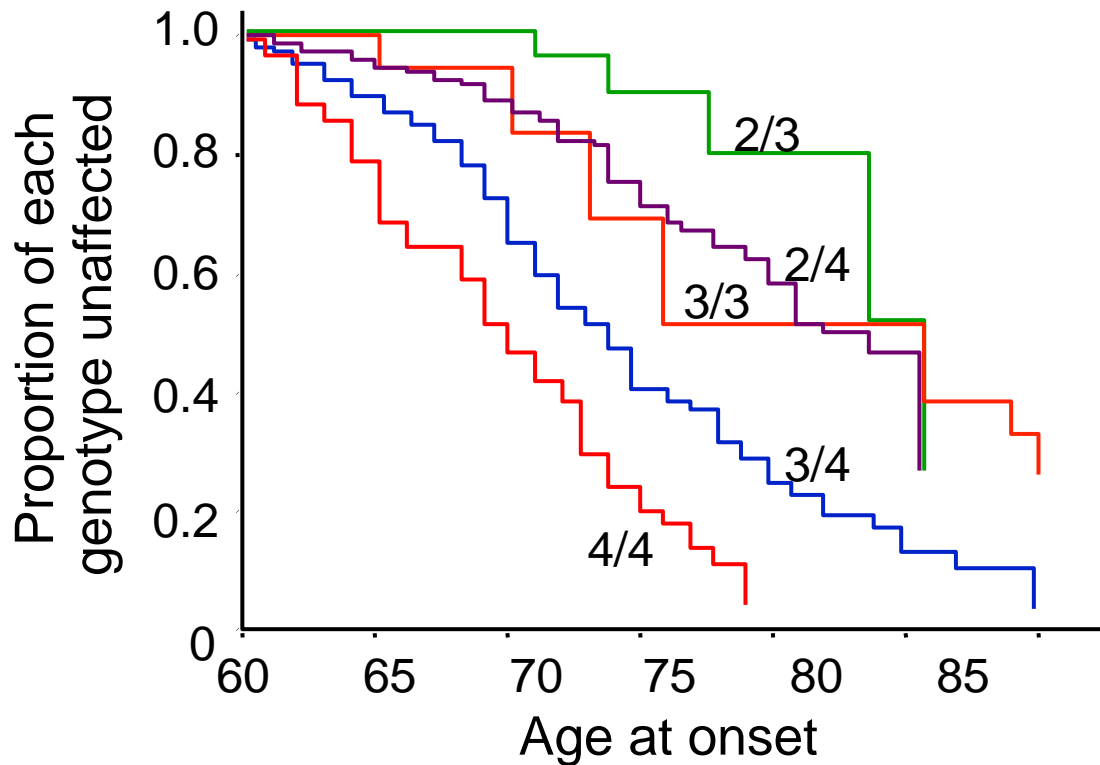
“... all five family members with severe insulin resistance, and no other family members, were compound heterozygous with respect to two frameshift mutations of these two unlinked genes.”

Multiple Hits ... Digenic Diseases

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

EFFECT AND PHENOTYPE	GENE 1		GENE 2	
	Mutation	Phenotype	Mutation	Phenotype
Synergistic:				
RP	<i>ROM1</i> ^{+/G80insG}	Normal	<i>RDS</i> ^{+/L185P}	Normal
RP	<i>ROM1</i> ^{+/L114insG}	Normal	<i>RDS</i> ^{+/L185P}	Normal
Bardet-Biedl	<i>BBS2</i> ^{Y24X/Q59X}	Normal	<i>BBS6</i> ^{+/Q147X}	Normal
Deafness	<i>GJB2</i> ^{+/S5delG}	Normal	<i>GJB6</i> ^{+/-}	Normal
Deafness	<i>GJB2</i> ^{+/L67delT}	Normal	<i>GJB6</i> ^{+/-}	Normal
Hirschsprung	<i>RET</i> ^{+/L647I}	Normal	<i>EDNRB</i> ^{+/S305N}	Normal
Severe insulin resistance	<i>PPARG</i> ^{+/A553delAAAAT}	Normal	<i>PPP1R3A</i> ^{+/C1984delAG}	Normal
Modifier:				
Juvenile-onset glaucoma	<i>MYOC</i> ^{+/G399V}	Adult-onset glaucoma	<i>CYP1B1</i> ^{+/R368H}	Normal
Usher 1	<i>USH3</i> ^{rmut/rmut}	Usher 3	<i>MYO7A</i> ^{+delG (exon 25)}	Normal
Congenital nonlethal JEB	<i>COL17A1</i> ^{R1226X/L855X}	Juvenile JEB	<i>LAMB3</i> ^{+/R635X}	Normal
More severe ADPKD	<i>PKD1</i> ^{+/mut}	Less severe ADPKD	<i>PKD2</i> ^{+/2152delA}	Less severe ADPKD
More severe hearing loss	<i>DFNA1</i>	Mild hearing loss	<i>DFNA2</i>	Mild hearing loss
WS2/OA	<i>MITF</i> ^{+/R44delA}	?WS2	<i>TYR</i> ^{+/R402Q}	Normal
More severe WS2/OA	<i>MITF</i> ^{+/R44delA}	?WS2	<i>TYR</i> ^{R402Q/R402Q}	Normal

Strong Effects of Single Genes



Alzheimer disease: Strong effect of APOE gene.

Survival curves for people with different APOE genotypes

"Fish Consumption and the Risk of Alzheimer Disease"

Friedland (2003) *Arch Neurol* **60**, 923 (editorial)

- Subjects who eat fish at least once a week have a 60% lower risk for developing AD than those who consume fish less frequently (Morris et al. [2003] *Arch Neurol* **60**, 194).
- Previous studies point in same direction.

The *Interheart* Study

- Effects of potentially modifiable risk factors associated with myocardial infarction (coronary heart disease) in 52 countries (Yusuf et al [2004] *Lancet* 364, 937-952). 12,461 cases; 14,637 controls.

Risk factor	Prevalence		Odds ratio (99% CI) adjusted for age, sex, and smoking (OR 1)	PAR (99% CI)
	Controls (%)	Cases (%)		
Current smoking*	26.76	45.17	2.95 (2.72-3.20)	-
Current and former smoking*	48.12	65.19	2.27 (2.11-2.44)	36.4% (33.9-39.0)
Diabetes	7.52	18.45	3.08 (2.77-3.42)	12.3% (11.2-13.5)
Hypertension	21.91	39.02	2.48 (2.30-2.68)	23.4% (21.7-25.1)
Abdominal obesity (2 vs 1)†	33.40	30.21	1.36 (1.24-1.48)	-
Abdominal obesity (3 vs 1)†	33.32	46.31	2.24 (2.06-2.45)	33.7% (30.2-37.4)
All psychosocial‡	-	-	2.51 (2.15-2.93)	28.8% (22.6-35.8)
Vegetables and fruit daily*	42.36	35.79	0.70 (0.64-0.77)	12.9% (10.0-16.6)
Exercise*	19.28	14.27	0.72 (0.65-0.79)	25.5% (20.1-31.8)
Alcohol intake*	24.45	24.01	0.79 (0.73-0.86)	13.9% (9.3-20.2)
ApoB/ApoA1 ratio (2 vs 1)§	19.99	14.26	1.47 (1.28-1.68)	-
ApoB/ApoA1 ratio (3 vs 1)§	20.02	18.05	2.00 (1.74-2.29)	-
ApoB/ApoA1 ratio (4 vs 1)§	19.99	24.22	2.72 (2.38-3.10)	-
ApoB/ApoA1 ratio (5 vs 1)§	20.00	33.49	3.87 (3.39-4.42)	54.1% (49.6-58.6)
All above risk factors combined¶	-	-	129.20 (90.24-184.99)	90.4% (88.1-92.4)



Non-Genetic Risk Factors Create Heterogeneity

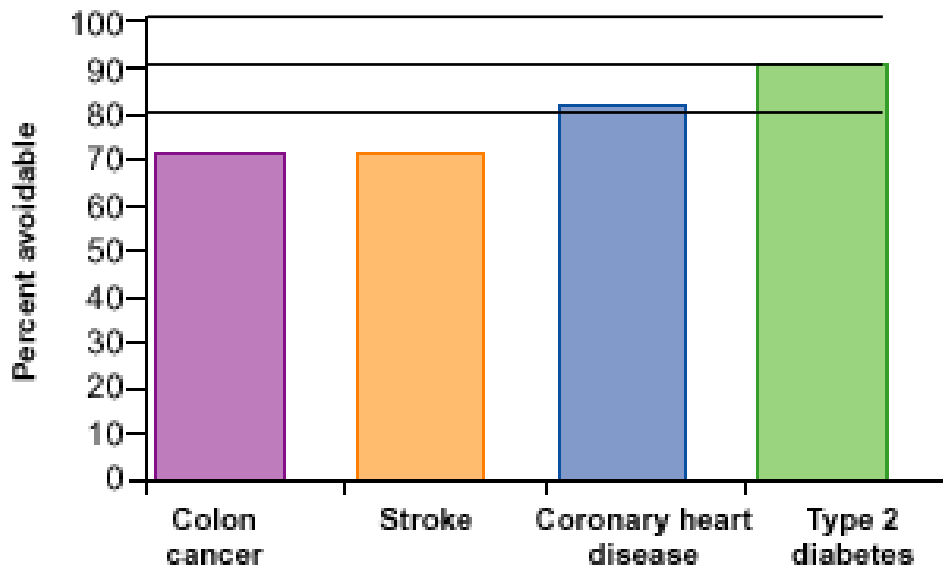
Ott (2004) *Neurology* 63, 955 (editorial)

- Cases with low environmental risk factors are more likely to be genetic
- Group individuals into more homogeneous classes regarding non-genetic risk factors
- Environmental risk factors not generally ascertained by geneticists but they should be.

Percentage of Traits Potentially Preventable by Life-Style Modifications

Willet (2002) *Science* **296**, 695-698

Low risk definitions include (references quoted):



- BMI < 25 kg/m²
- > 30 min. per day of brisk walking
- < 3 alcoholic drinks per day
- Nonsmoking
- < 3 servings of red meat per week

Genomic Priorities and Public Health

Merikangas & Risch (2003) *Science* 302, 599-601

- The translation of genomics to human disease will most likely involve genetic counseling, drug therapy, and gene therapy ... most effective for rare diseases at the level of the individual (e.g., PKU).
- In contrast, public health prevention is most effective when applied to common diseases at the population level.

Merikangas & Risch (2003):

Disease	Phenotype measure	λ	Genes	Specific environmental factors		Prevalence
			Confirmed loci	Known?	Malleable*	
Breast cancer	Biopsy	1.8	BRCA-1 BRCA-2	Parity, 1st child > age 30, physical inactivity	Possible	1.2%
Alzheimer's disease	Clinical, neurocognitive testing, postmortem biopsy	2.8	PS1/PS2 APP APOE	Head injury, low educational level	Possible	5% (>age 65)
Type 1 diabetes	Immunologic markers, glucose metabolism	15.0	HLA INS	Nonspecific	No	0.4%
Multiple sclerosis	Clinical, neuroimaging	20.0	HLA	Nonspecific	No	0.2%
Autism	Clinical	60.0	None	No	No	.02%
Schizophrenia	Clinical	9.0	None	No	No	0.8%
Cervical cancer	Biopsy	1.8	None	Human papilloma virus	Yes	0.16%
Type 2 diabetes	Glucose metabolism	4.3	PPAR γ	Obesity, physical inactivity	Yes	6.1% (>age 20) 15.0% (>age 60)
AIDS	Clinical, antibody, CD4 ⁺ count	NA	CCR5 HLA	HIV	Yes	0.12%
Nicotine dependence	Interview	1.4	None	Nicotine	Yes	24.0%
Alcohol dependence	Interview	7.0	ADH2 ALDH2 (protective)	Alcohol	Yes	4.0%

*) malleable = capable of being altered