



# Association Mapping in Families

---

Gonçalo Abecasis  
University of Oxford



# Linkage Analysis

---

- Sharing between relatives
  - Identifies large regions
    - Include several candidates
- Complex disease
  - Scans on sets of small families popular
  - No strong assumptions about disease alleles
  - Low power
  - Limited resolution

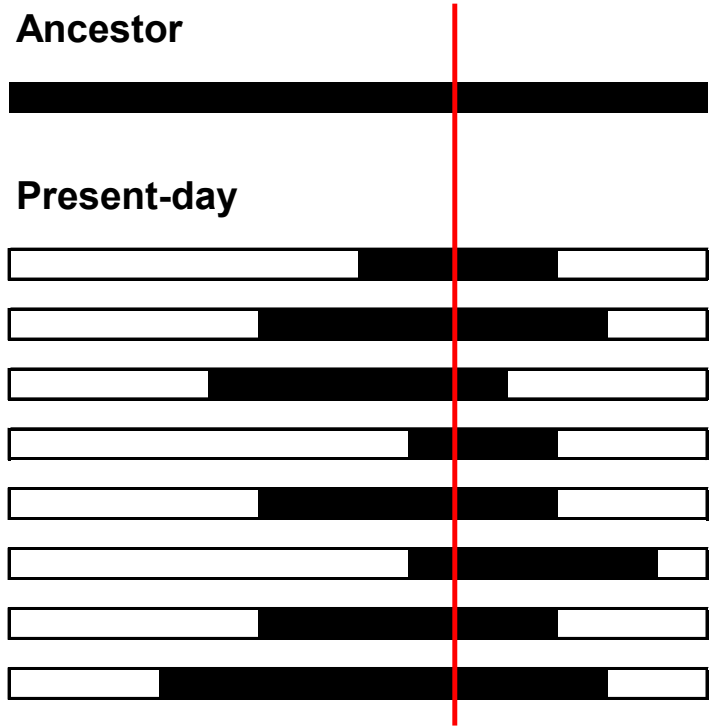
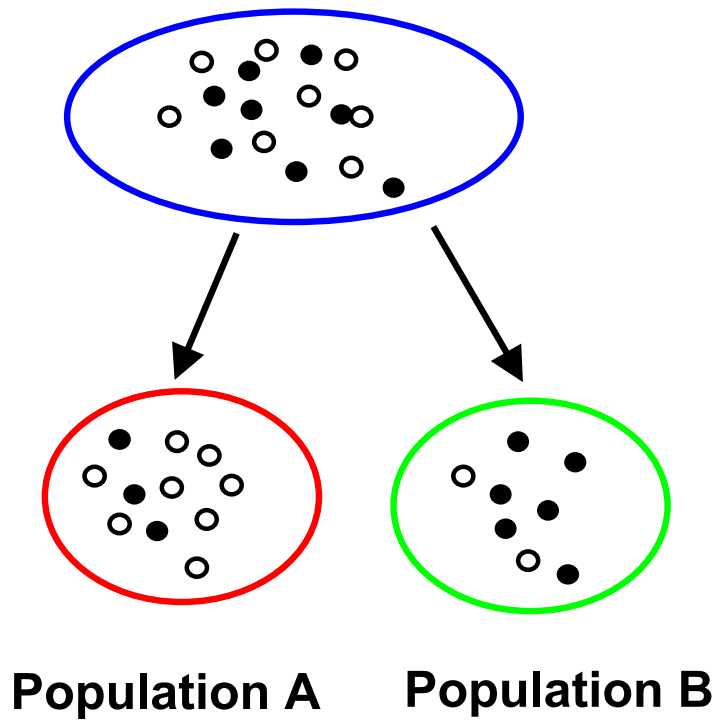


# Association Analysis

---

- Sharing between *unrelated* individuals
  - Trait alleles originate in common ancestor
  - High resolution
    - Recombination since common ancestor
    - Large number of independent tests
- Powerful if assumptions are met
  - Same disease haplotype shared by many patients
- Sensitive to population structure

# Stratification vs Disequilibrium





# Disequilibrium Mapping

---

- Control for possible population structure
  - Distinguish linkage disequilibrium from other types of association
- Family-based association analysis
  - Using families collected for linkage mapping
- Powerful if assumptions are met
  - Same disease haplotype shared by many patients
- High-resolution



# Essential Notation

---

- $i$  families
- $j = 1 .. n_i$  offspring in family  $i$
- $Y_{ij}$  quantitative phenotype
- $g_{ij}$  no. of '1' alleles at marker
- $g_{iF}, g_{iM}$  parental genotypes, optional
- $\pi_{ijk}$  IBD between  $j$  and  $k$  in family  $i$
- $\varphi_{ijk}$  kinship between  $j$  and  $k$  in family  $i$



# Controlling for Stratification

---

- If stratum were known...
  - For each individual genotype ( $g_{ij}$ )
  - Average number of alleles in a strata ( $b_{ij}$ )
  - Adjust for stratum differences ( $w_{ij} = g_{ij} - b_{ij}$ )

$$\hat{y}_{ij} = \mu + \hat{\beta}_b b_{ij} + \hat{\beta}_w w_{ij}$$

- How to define stratum then?
  - Use family data to estimate  $b_{ij}$



# $b_{ij}$ as Family Control

---

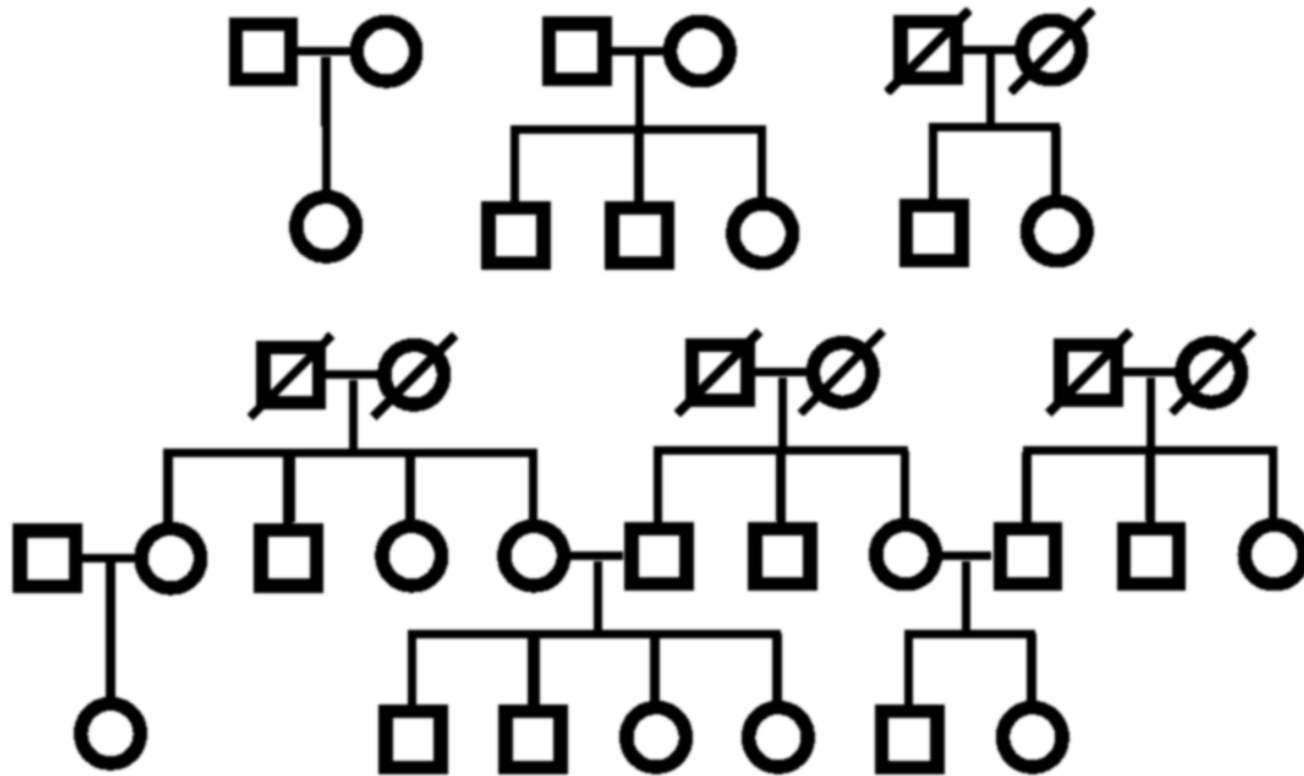
- Expected genotype for each individual
  - Ancestors
  - Siblings
- Informative individuals
  - Genotype may differ from expected
  - Have heterozygous ancestor in pedigree





# Allowable Family Structures

---





# Nuclear Families

---

$$b_{ij} = b_i = \begin{cases} \frac{g_{iF} + g_{iM}}{2} & \text{average of parental genotypes} \\ \sum_k^{sibship} \frac{g_{ik}}{n_{sibs}} & \text{average of sibling genotypes} \end{cases}$$

$$w_{ij} = g_{ij} - b_{ij}$$



# Extended Families

---

$$b_{ij} = \begin{cases} \frac{b_{iF_j} + b_{iM_j}}{2} & \text{average of parental controls} \\ \sum_k^{sibship} \frac{g_{ik}}{n_{sibs}} & \text{average of sibling genotypes} \\ g_{ij} & \text{self - genotype} \\ \text{undefined} & \text{otherwise} \end{cases}$$



# Allowing for Related Data

---

- Similarities between individuals
  - Variance–covariance matrix
- Major gene, polygenic, environment

$$\Omega_{ijk} = \begin{cases} \sigma_a^2 + \sigma_g^2 + \sigma_e^2 & \text{if } j = k \\ \pi_{ijk}\sigma_a^2 + 2\varphi_{ijk}\sigma_g^2 & \text{if } j \neq k \end{cases}$$



# Likelihood function

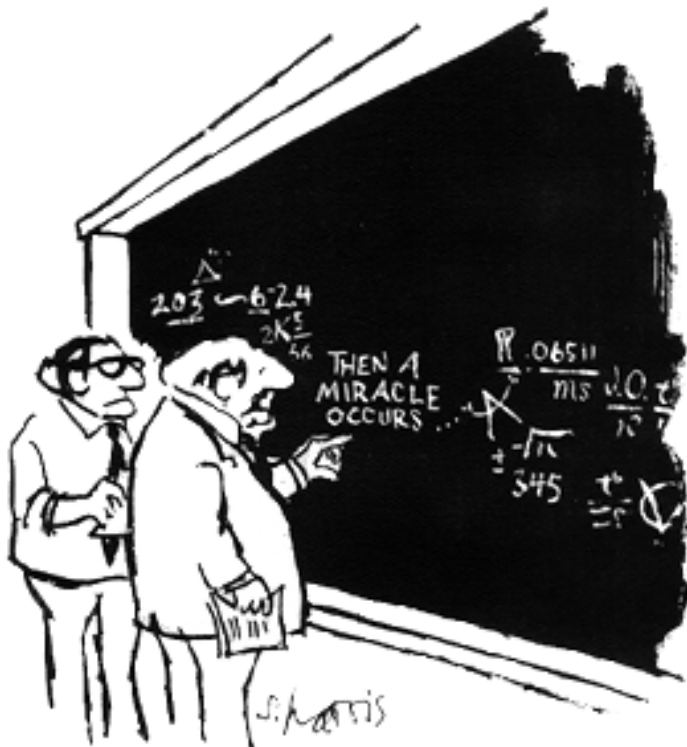
---

- Multivariate Normal Distribution
- Defines asymptotic significance levels

$$L = \prod_i (2\pi)^{-n_i/2} |\hat{\mathbf{\Omega}}_i|^{-1/2} e^{-1/2[(\mathbf{y}_i - \hat{\mathbf{y}}_i)' \hat{\mathbf{\Omega}}_i^{-1} (\mathbf{y}_i - \hat{\mathbf{y}}_i)]}$$

$$\chi^2 = 2 \ln \left( \frac{L_{\text{full model}}}{L_{\text{sub-model}}} \right)$$

# Parameter Derivations



"I think you should be more explicit here in step two."

© 1998 Sidney Harris

$$Model = (\mu, \beta_b, \beta_w, \sigma_e^2, \sigma_g^2, \sigma_a^2)$$

$$\begin{bmatrix} \beta_b \\ \beta_w \end{bmatrix} = \begin{bmatrix} \frac{\sum_i n_i (p_i - q_i) \mu_i}{NV_b} + a \\ a \end{bmatrix}$$

$$a = \frac{D}{pq} a_{QTL}$$

$$\sigma_a^2 = V_{QTL} - 2pqa$$



# Exact Permutation Test

---

- In each family,  $\mathbf{w}_i = [w_{i1}, w_{i2} \dots]$  is the pattern of allelic transmission
  - $\mathbf{w}_i$  and  $-\mathbf{w}_i$  are equally likely ( $H_0$ )
- Null distribution of the data
  - Randomly permute any set families by replacing each  $\mathbf{w}_i$  with itself or  $-\mathbf{w}_i$  with equal probability
  - The permuted data sets define the null distribution of the maximum likelihood statistic
- Empirical significance levels



# Application: Angiotensin-1

---

- British population
- Circulating ACE levels
  - Normalized separately for males / females
- 10 di-allelic polymorphisms
  - 26 kb
  - Common
  - In strong linkage disequilibrium
- Keavney et al, HMG, 1998





# Haplotype Analysis

- 3 clades **A**
  - All common haplotypes
  - >90% of all haplotypes
- "B" = "C"
  - Equal phenotypic effect **B**
  - Functional variant on right
- Keavney et al (1998) **C**

TATATT**A**IA3

TATAT**C**GIA3

TATATTGIA3

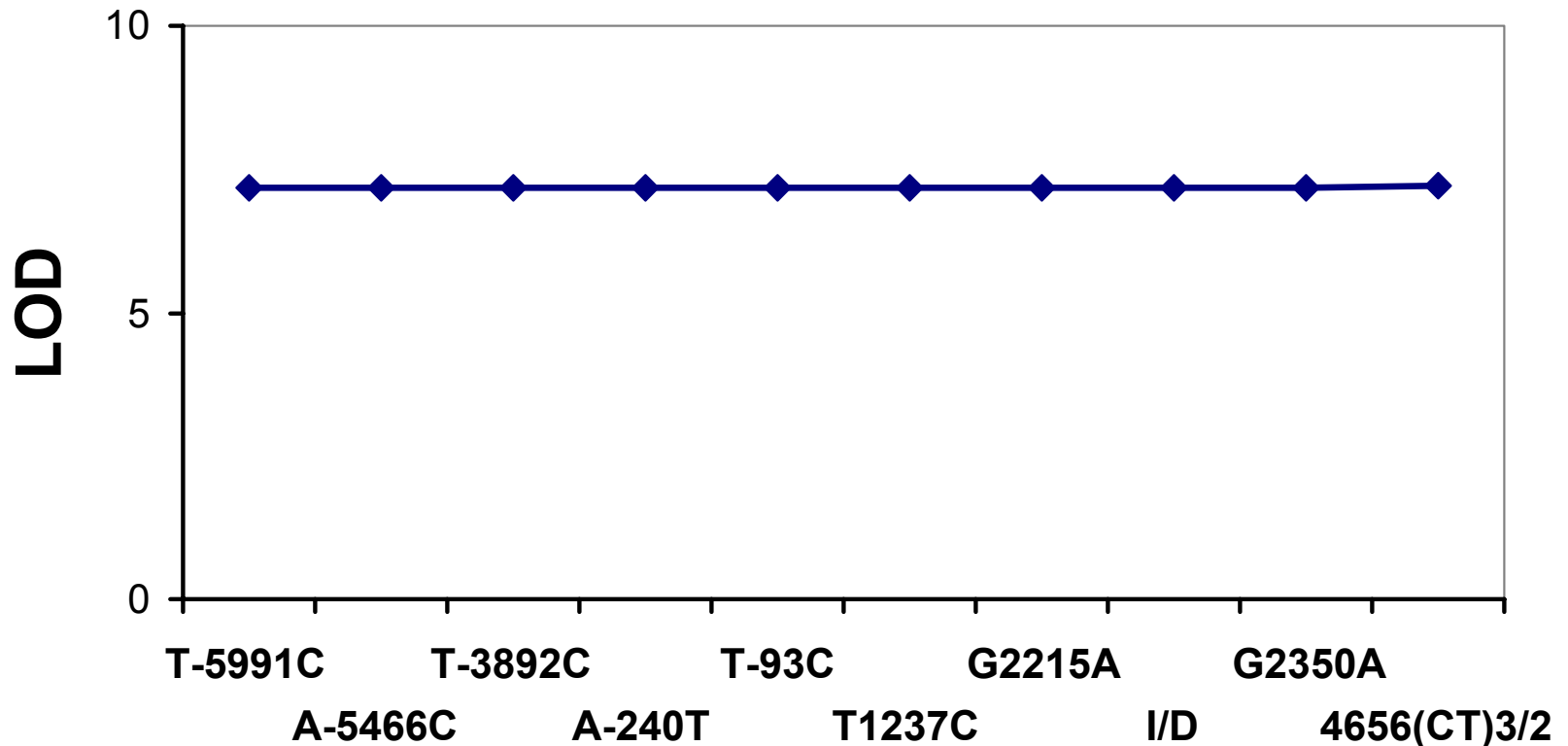
CCCTCC**G**DG2

CCCTCCADG2

TATATCADG2

TAC**A**TCADG2

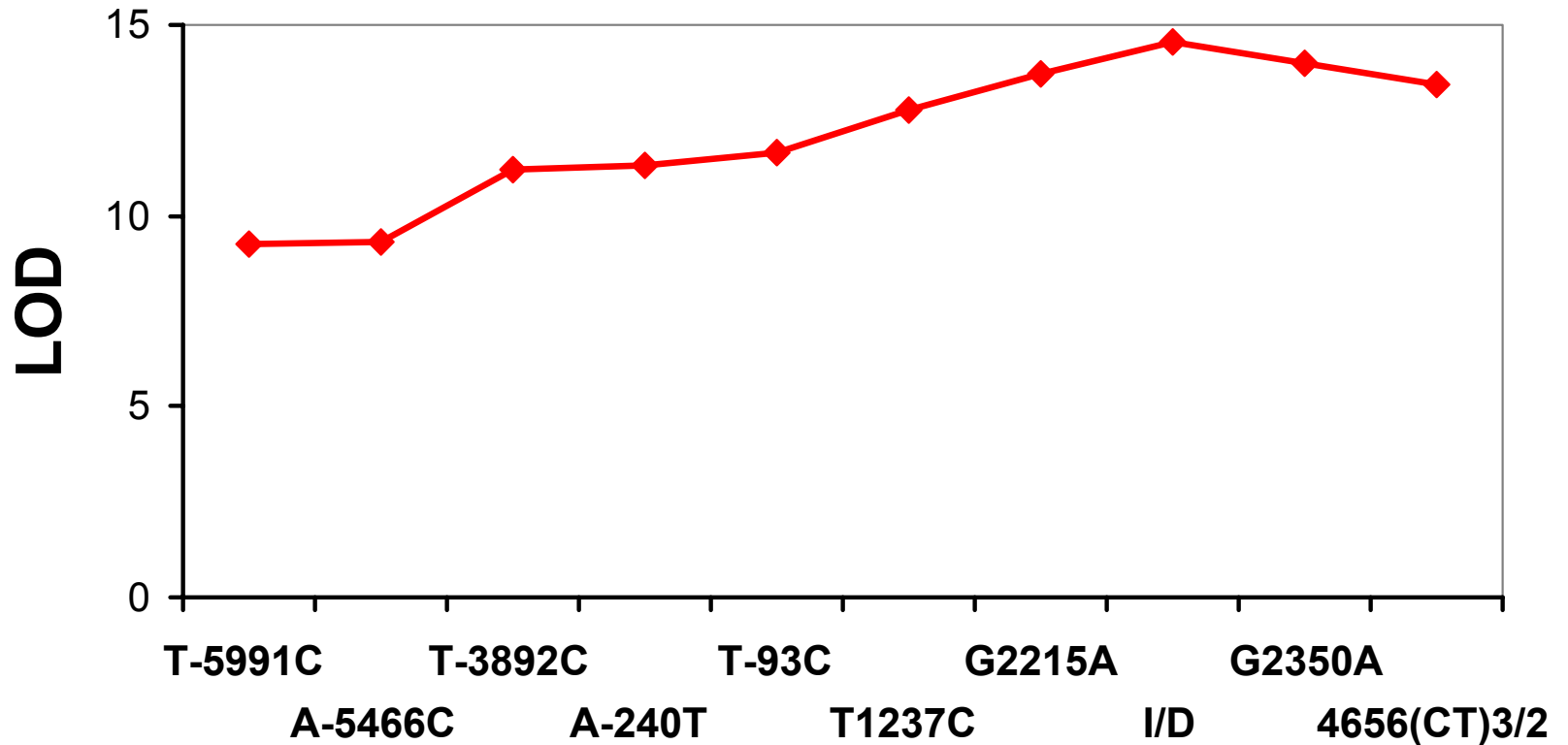
# Evidence for Linkage



$$H_0 : (\mu, \sigma_g^2, \sigma_e^2)$$

$$H_1 : (\mu, \sigma_g^2, \sigma_e^2, \sigma_a^2)$$

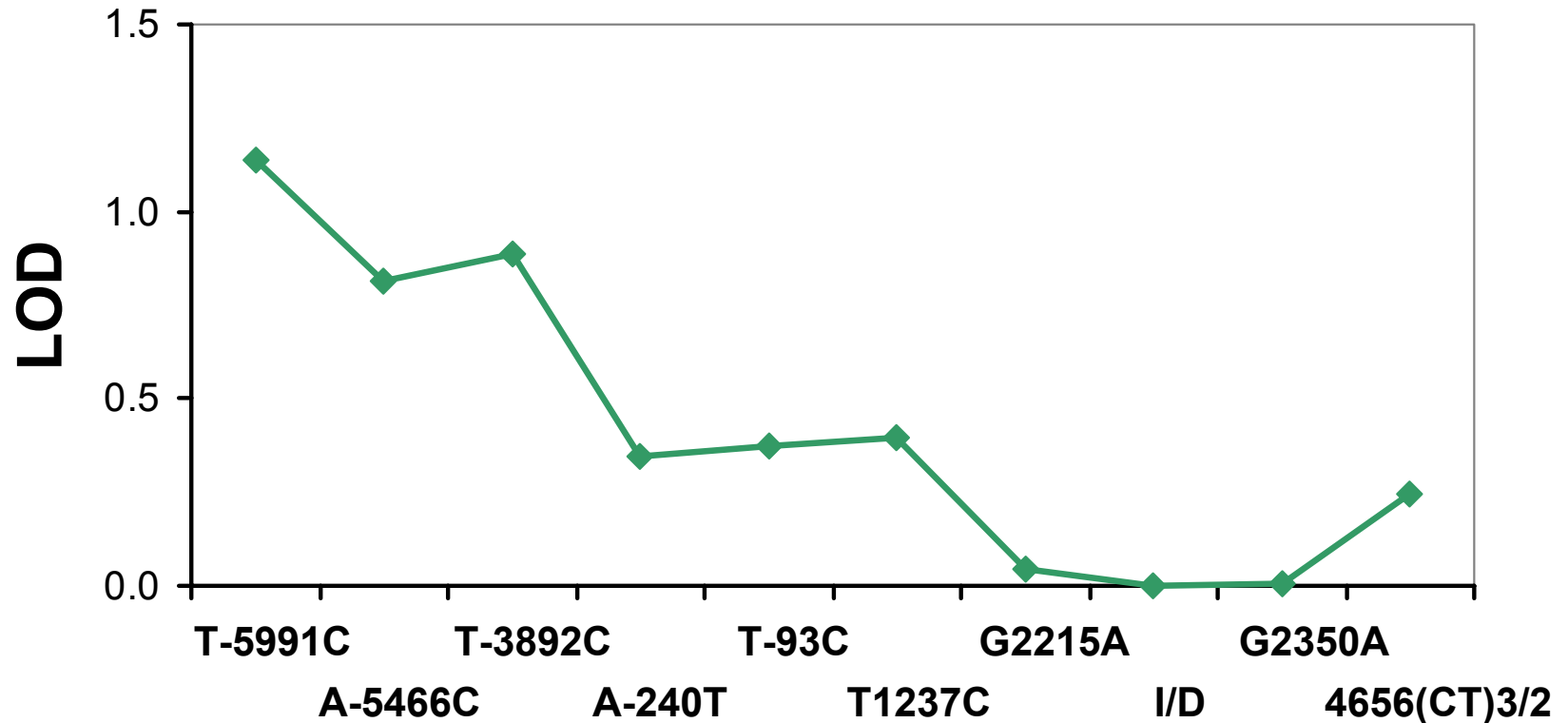
# Evidence for Association



$$H_0 : (\mu, \sigma_g^2, \sigma_a^2, \sigma_e^2, \beta_b)$$

$$H_1 : (\mu, \sigma_g^2, \sigma_a^2, \sigma_e^2, \beta_b, \beta_w)$$

# Evidence Against Complete LD

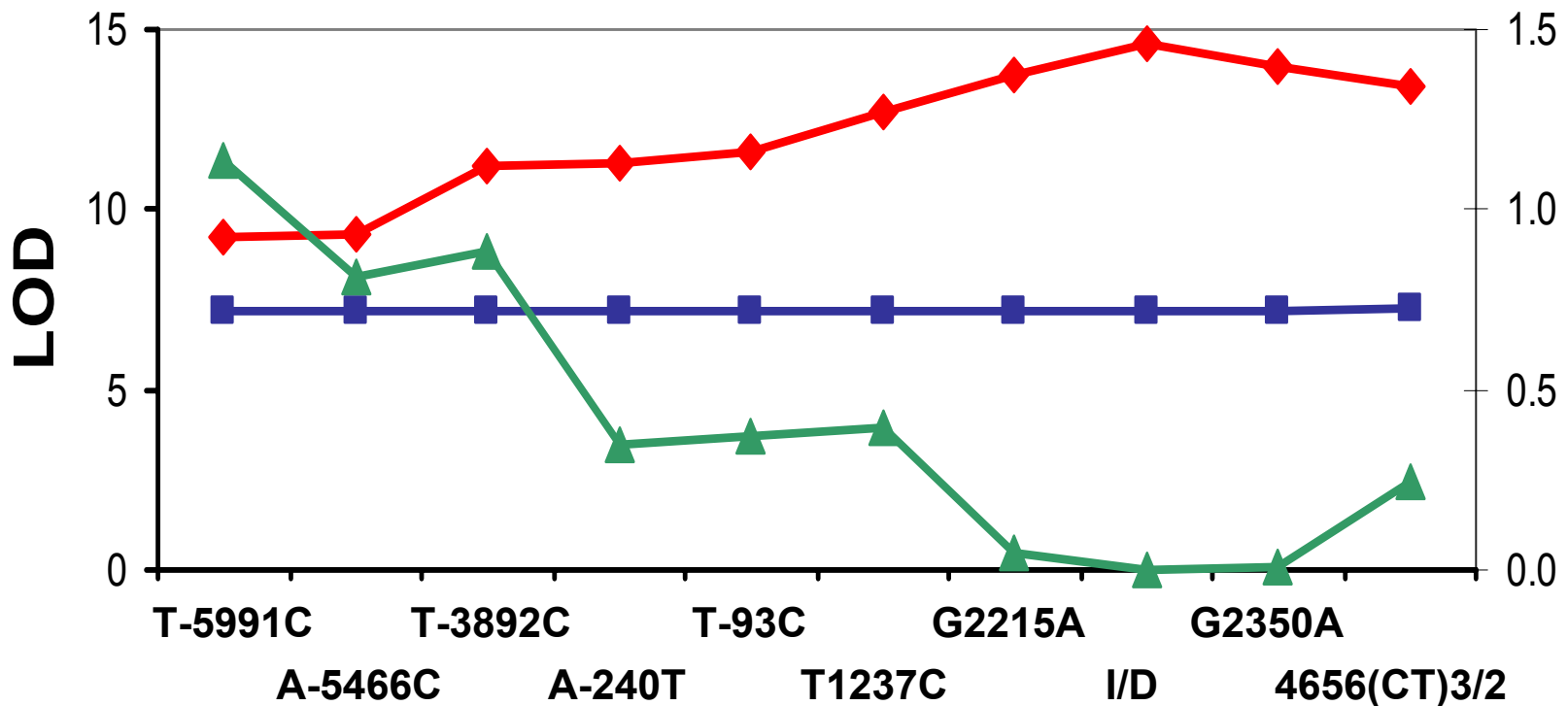


$$H_0 : (\mu, \beta_b, \beta_w, \sigma_g^2, \sigma_e^2)$$

$$H_1 : (\mu, \beta_b, \beta_w, \sigma_g^2, \sigma_e^2, \sigma_a^2)$$

# Drawing Conclusions

■ for Linkage    ◆ for Association    ▲ against Complete LD





# Parameter Estimates

---

- Estimates
  - Total linkage ( $\sigma^2_a$ )
  - Linkage due to LD ( $\sigma^2_{a*LD}$ )
  - Effect size at marker ( $\beta_w$ )
- Depend on
  - QTL allele frequencies ( $p,q$ )
  - QTL effect ( $a$ )
  - Disequilibrium ( $D$ )
  - Marker allele frequencies ( $r,s$ )



# Useful diagnostics

---

- A bit of algebra
- Provide indicator of distance
  - Minimum  $D'$  ( $D'_{\min}$ )
- Select next markers
  - Range for QTL alleles ( $p_{\min}$ ,  $p_{\max}$ )

$$D'_{\min} = \sqrt{\sigma_{a-LD}^2 / \sigma_a^2}$$

$$p_{\min} = \frac{1}{1 + \sigma_a^2 / (2\beta_w^2 r^2)}$$

$$p_{\max} = 1 - \frac{1}{1 + \sigma_a^2 / (2\beta_w^2 s^2)}$$



# Application to T-5991C

---

- LOD scores
  - $\sim 7$  linkage
  - $\sim 9$  association
  - $\sim 1$  linkage minus association
- Trait locus predictions
  - In greater than 78% disequilibrium
  - Minor allele frequency between .15 and .48
- Compare to I/D and neighbors





# ACE: $D'_{\min}$ , $p_{\min}$ and $p_{\max}$

	<b>Expected</b>	<b>Actual</b>		
	<b>T-5991C</b>	<b>G2215A</b>	<b>I/D</b>	<b>G2350A</b>
<b><math>D'</math></b>	> 0.78	0.78	0.82	0.85
<b>Minor allele</b>	.15–.48	.45–.50	.45–.50	.45–.50



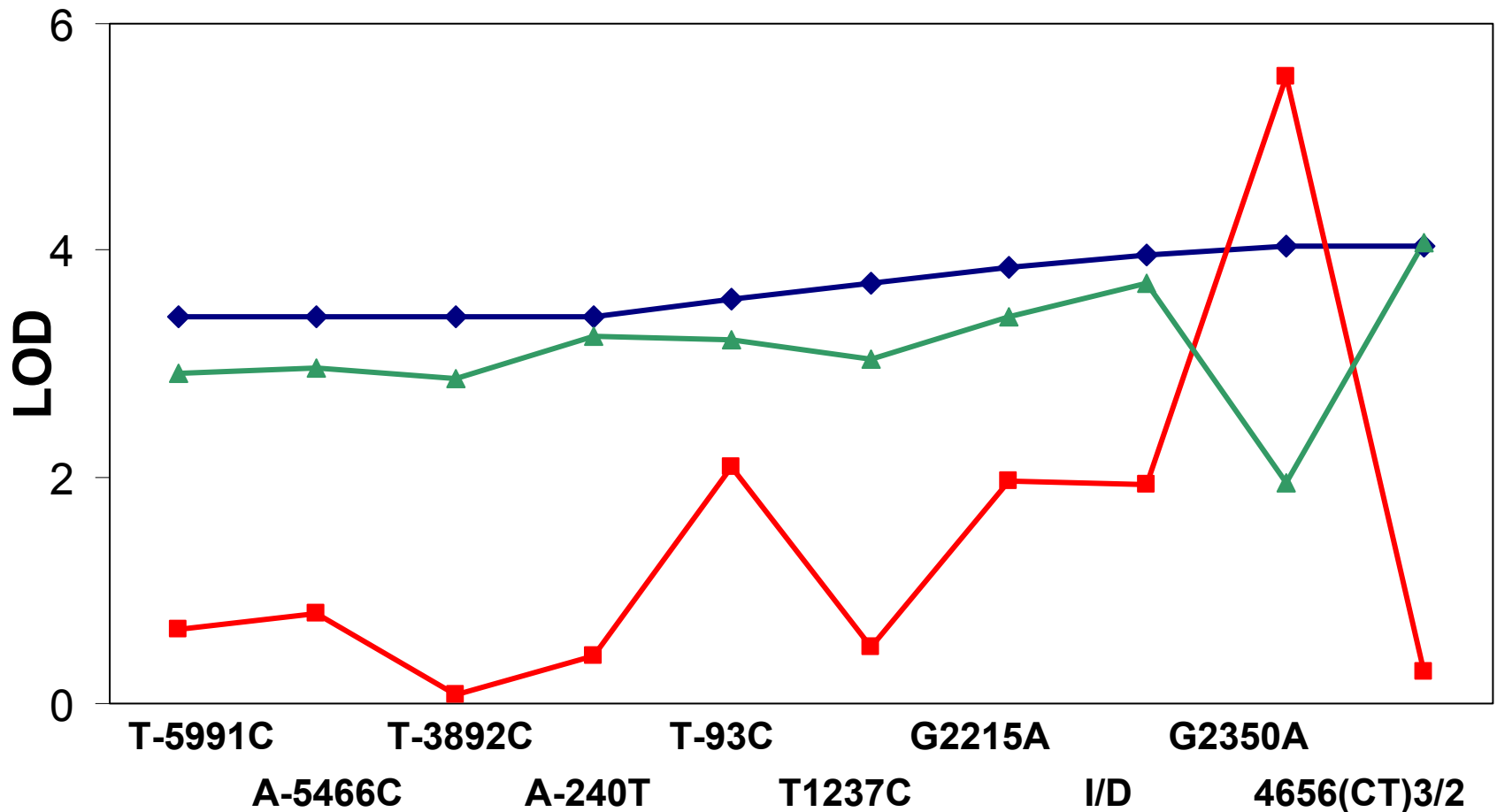
# Finer mapping

---

- 3 mutations (inc. I/D) explain all linkage
  - UK population
- How to identify causal variant?
  - Population with more haplotype diversity
  - Jamaican sample
    - Colin McKenzie, University of West Indies, Jamaica
    - Same di-allelic polymorphisms

# Jamaican Population Summary

◆ for Linkage ■ for Association ▲ against Complete LD





# Example Summary

---

- Agrees with haplotype analysis
- Distinguishes complete and incomplete disequilibrium
  - Measure of distance for incomplete LD
  - Indicator of trait allele frequencies
- Typical or fairy-tale?



# Study design

---

- What markers?
  - Effect of disequilibrium
  - Effect of allele frequencies
- Family sample
  - What size families?
  - Parents or no parents?
- Effect of phenotypic selection



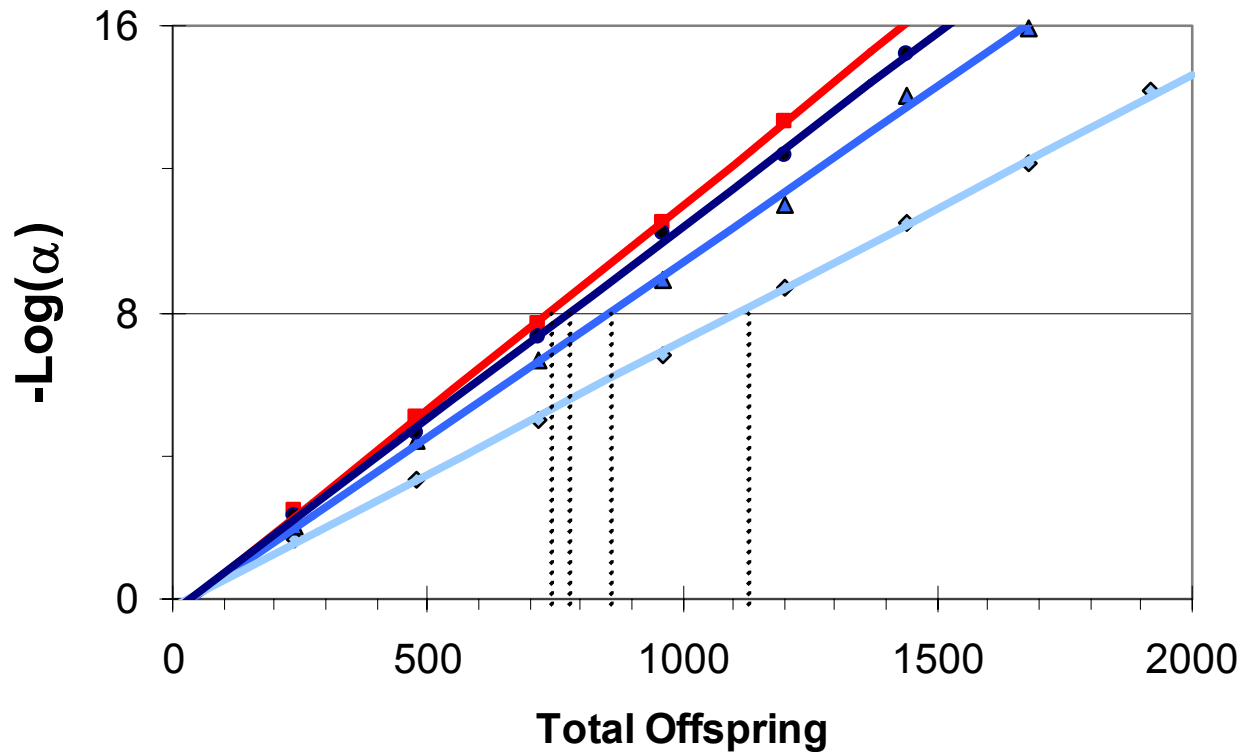
# Sensitivity to Disequilibrium

	Amount of Disequilibrium				
	0%	25%	50%	75%	100%
480 triads	0	2	20	70	97
240 sib-pairs	0	2	23	73	98
120 sib-quads	0	3	27	76	98
Estimate of $a$	0	1.1	2.2	3.4	4.5

Power for  $\alpha=0.001$ ,  $h^2 = .1$ ,  $s^2 = .3$ ,  $\theta = 0$ .

Average additive genetic value estimated at the marker.

# Effect of Family Structure



■ parents   ◇ sib-pair   ▲ sib-triad   ● sib-quad



# Trios For Genome-Wide Scan

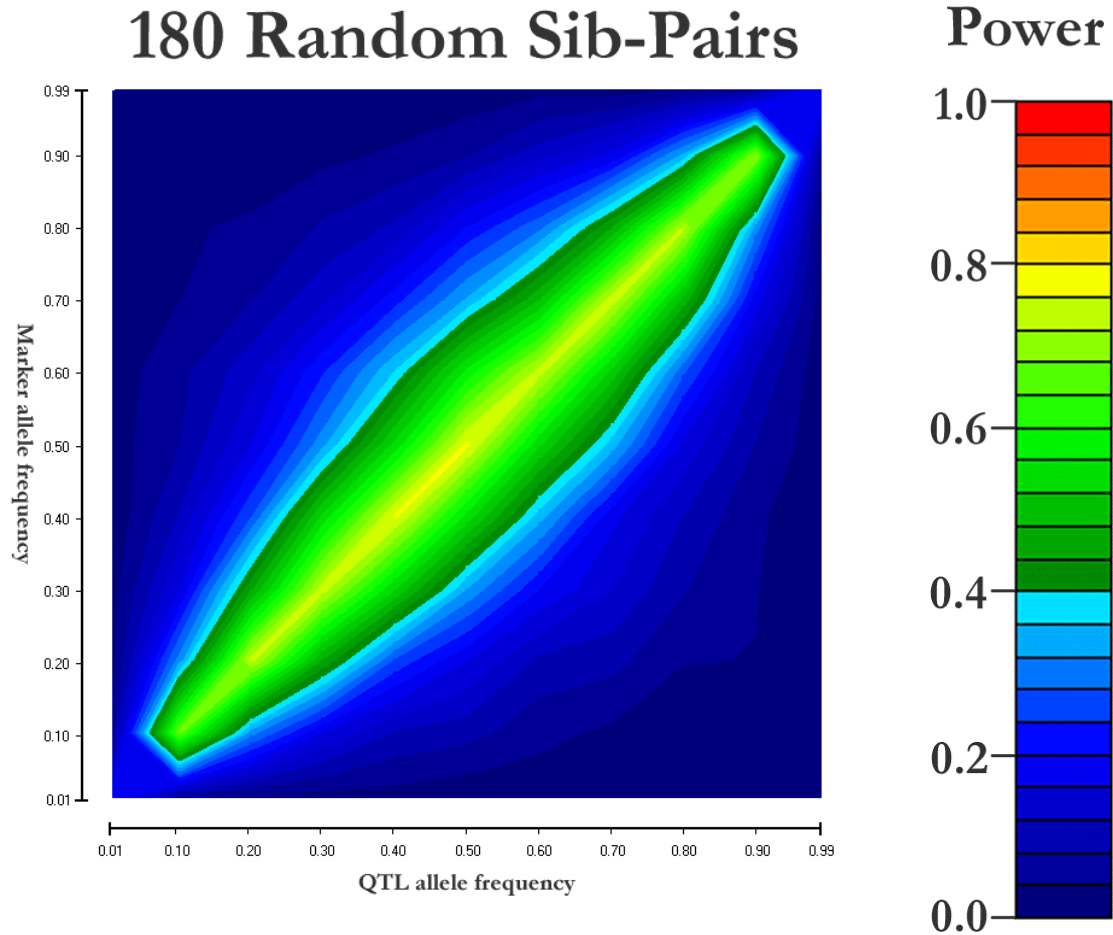
---

Disease Allele Frequency	Marker Allele Frequency				
	0.1	0.3	0.5	0.7	0.9
0.1	<b>248</b>	626	1306	2893	10830
0.3	1018	<b>238</b>	466	996	3651
0.5	2874	702	<b>267</b>	556	2002
0.7	9169	2299	925	<b>337</b>	1187
0.9	73783	18908	7933	3229	<b>616</b>

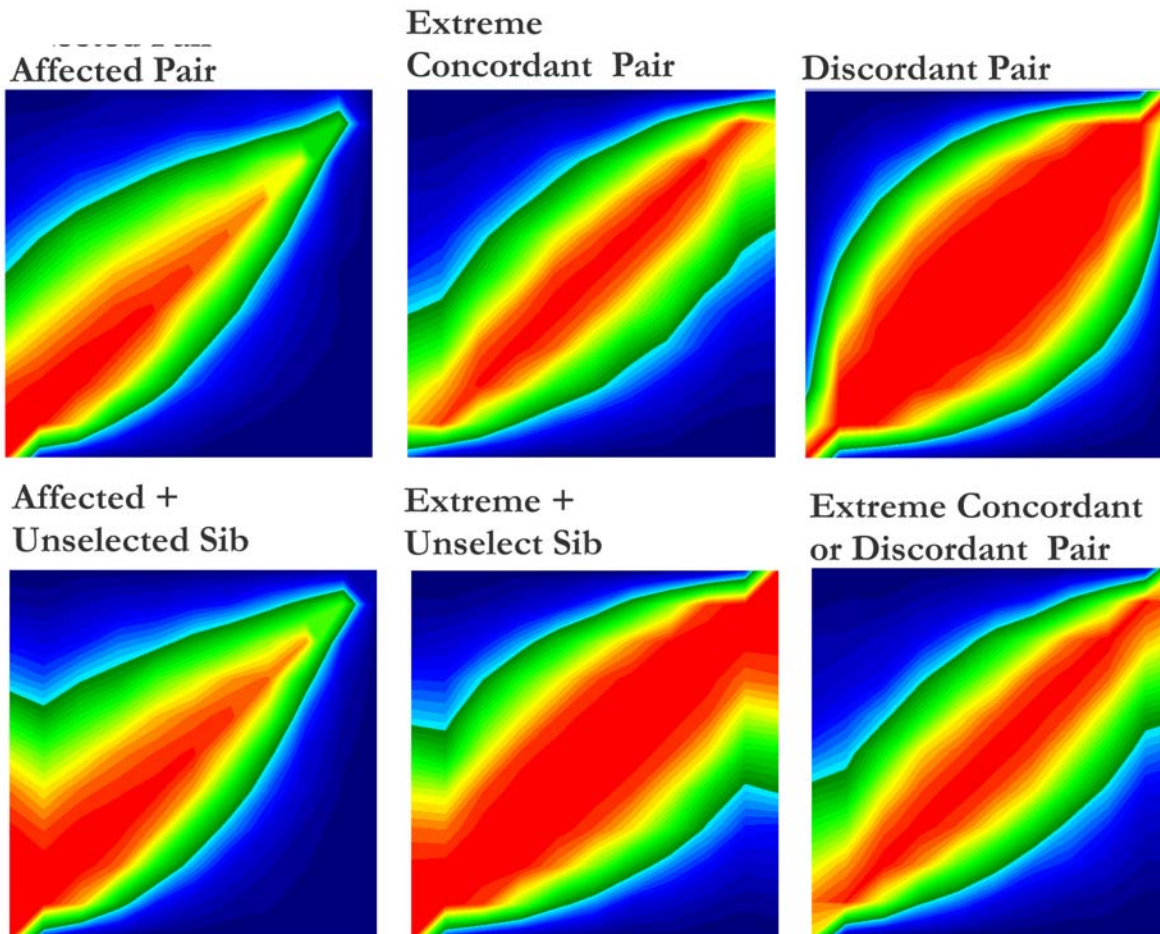
$\lambda_S = 1.5$ ,  $\alpha = 5 \times 10^{-8}$ , Spielman TDT  
(Müller-Myhsok and Abel, 1997)



# Effect of Allele Frequencies



# Effect of Selection





# References

---

Fulker et al (1999) *Am J Hum Genet* **66**:259-267

Abecasis et al (2000) *Am J Hum Genet* **66**:279-292

Abecasis et al (2000) *Eur J Hum Genet* **8**:545-551

Cardon and Abecasis (2000) *Behav Genet* **30**:235-243

McKenzie et al (2001) *Hum Mol Genet* **10**:1077-84

Abecasis et al (2001) *Am J Hum Genet* **68**:1463-74



# www.sph.umich.edu/csg/abecasis



**QTD T Home Page - Netscape**

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop

Bookmarks Netsite: <http://www.well.ox.ac.uk/asthma/QTD T/> What's Related

Instant Message WebMail Contact People Yellow Pages Download Find Sites Channels

## QTD T

### Linkage Disequilibrium Analyses for Quantitative Traits

QTD T provides a convenient one-stop interface for family based tests of linkage disequilibrium. The general model described by [Abecasis \(1999\)](#) applies to families with and without parental data, and includes an optional permutation framework for exact p-values. The tests described by Allison (TDTQ5, 1997), Rabinowitz (1997), Monks (1998) and Fulker (1999) are also supported.

[Background Information](#) Read about family based disequilibrium analysis and QTD T.

[Quick-Start Tutorial](#) Take a tour of the QTD T package. **Recommended.**

[Browse documentation](#) The online documentation includes information on file formats and known bugs.

[Download programs](#) Download the latest version of the programs. [Updated February 2000]

[Registration form](#) If you decide to use QTD T, please take a minute to let us know.

---

[University of Oxford](#) | [Wellcome Trust Centre](#) | [Asthma Genetics](#)

Reload the current page