I Have the Power in QTL linkage: single and multilocus analysis

Benjamin Neale[1], Sir Shaun Purcell[2] & Pak Sham[13]

[1]SGDP, IoP, London, UK

[2]Harvard School of Public Health, Cambridge, MA, USA

[3]Department of Psychiatry, Hong Kong University, HK

Overview

1) Brief power primer

Practical 1 : Using GPC for elementary power calculations

2) Calculating power for QTL linkage analysis

Practical 2 : Using GPC for linkage power calculations

3) Structure of Mx power script

What will be discussed

What is power? (refresher)

Why and when to do power?

What affects power in linkage analysis?

How do we calculate power for QTL linkage analysis

Practical 1 : Using GPC for linkage power calculations

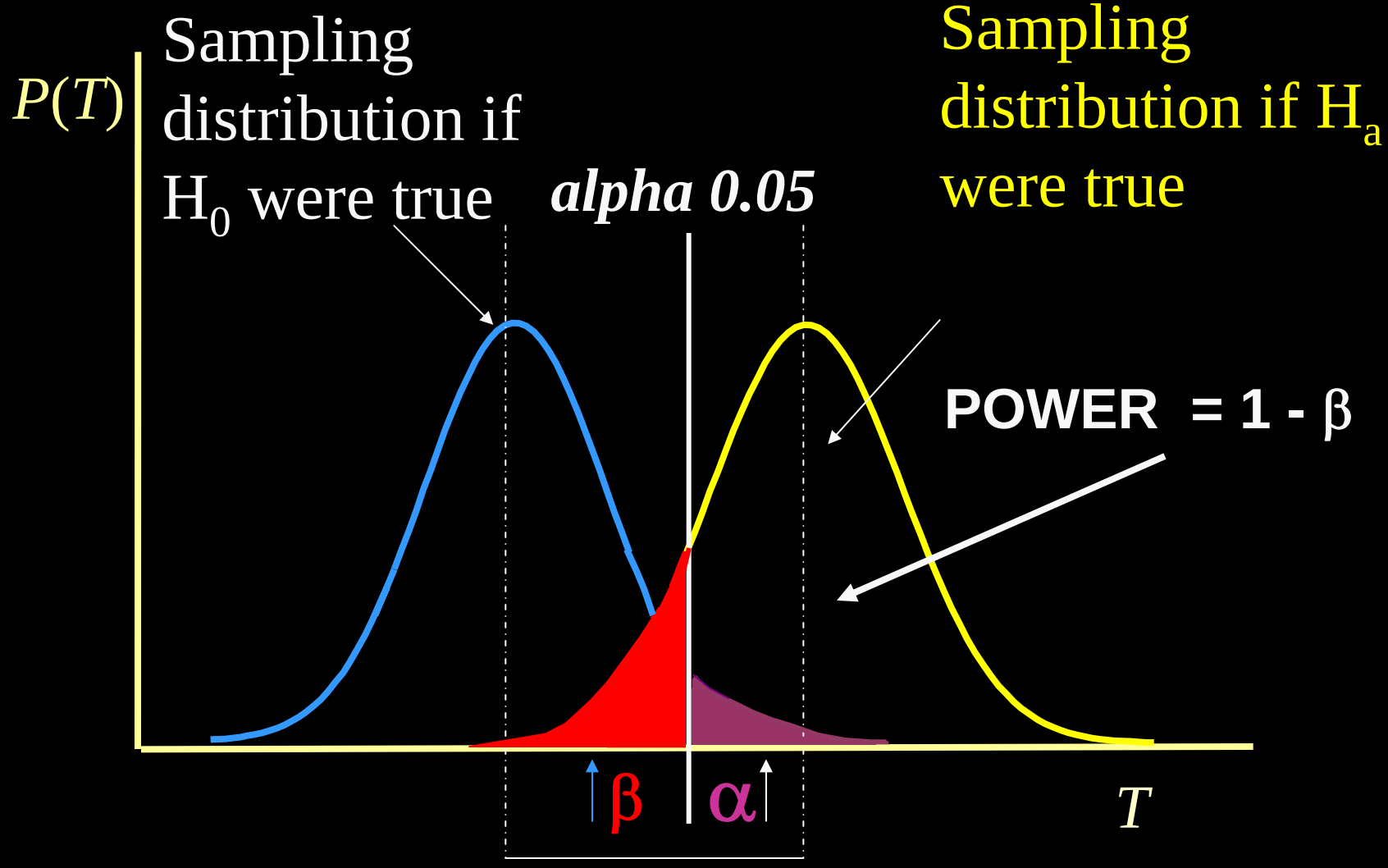
The adequacy of additive single locus analysis

Practical 2 : Using Mx for linkage power calculations

Needed for power calculations

- Test statistic
- Distribution of test statistic under H_0
  - to set significance threshold
- Distribution of test statistic under H_a
  - to calculate probability of exceeding significance threshold

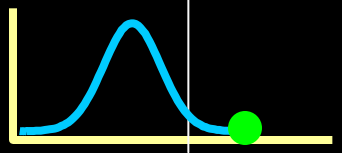
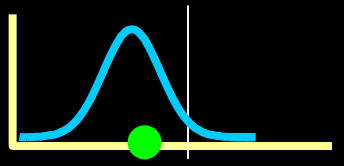
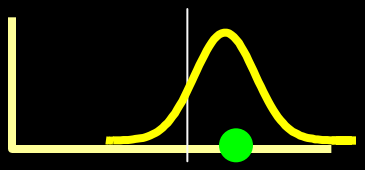
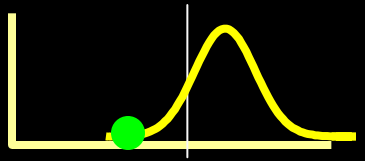
Standard Case



Type-I & Type-II error probabilities

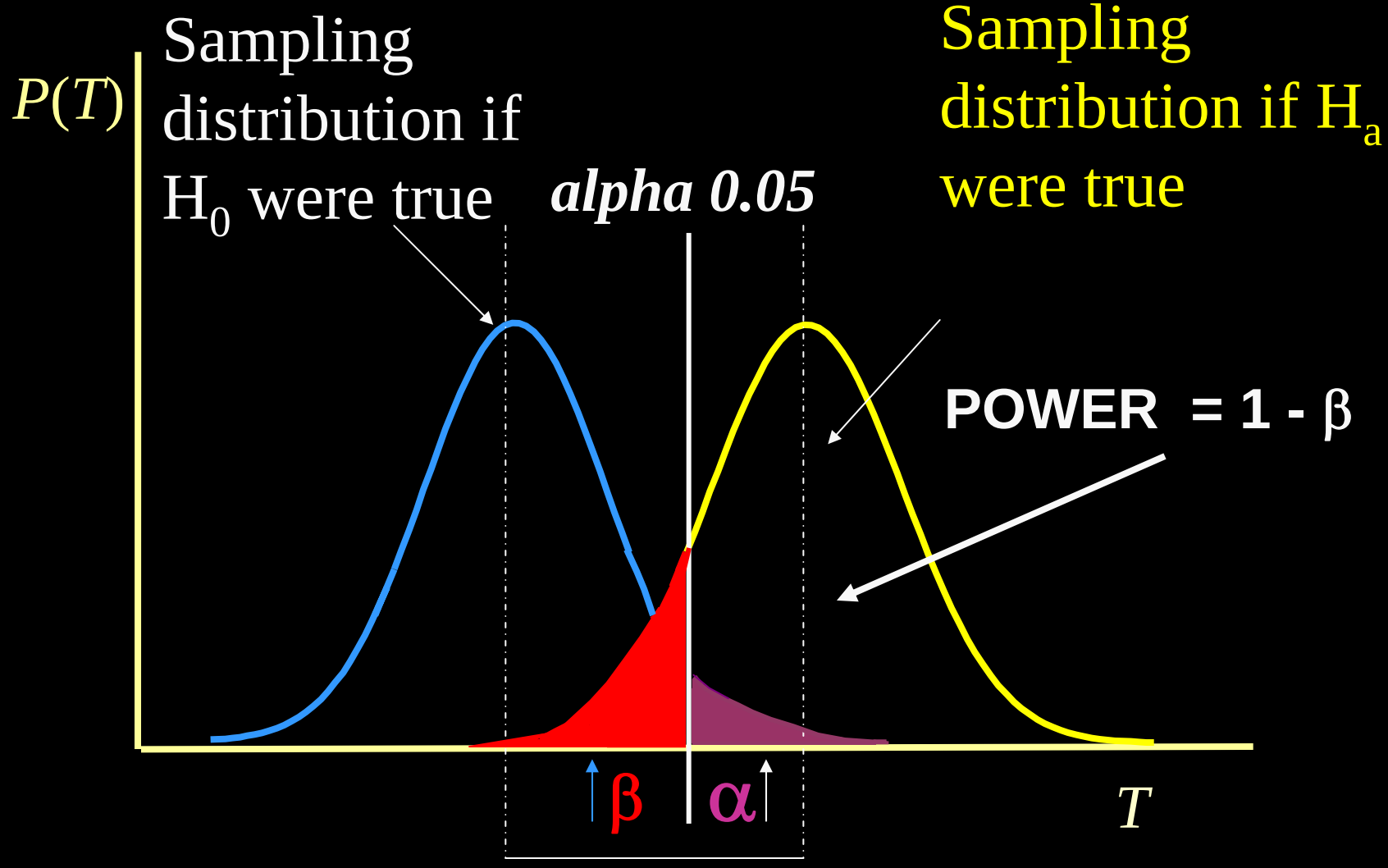
	Null hypothesis True	Null hypothesis False
Accept H_0	$1-\alpha$	β (type-II error) (false negative)
Reject H_0	α (type-I error) (false positive)	$1-\beta$ (power)

STATISTICS

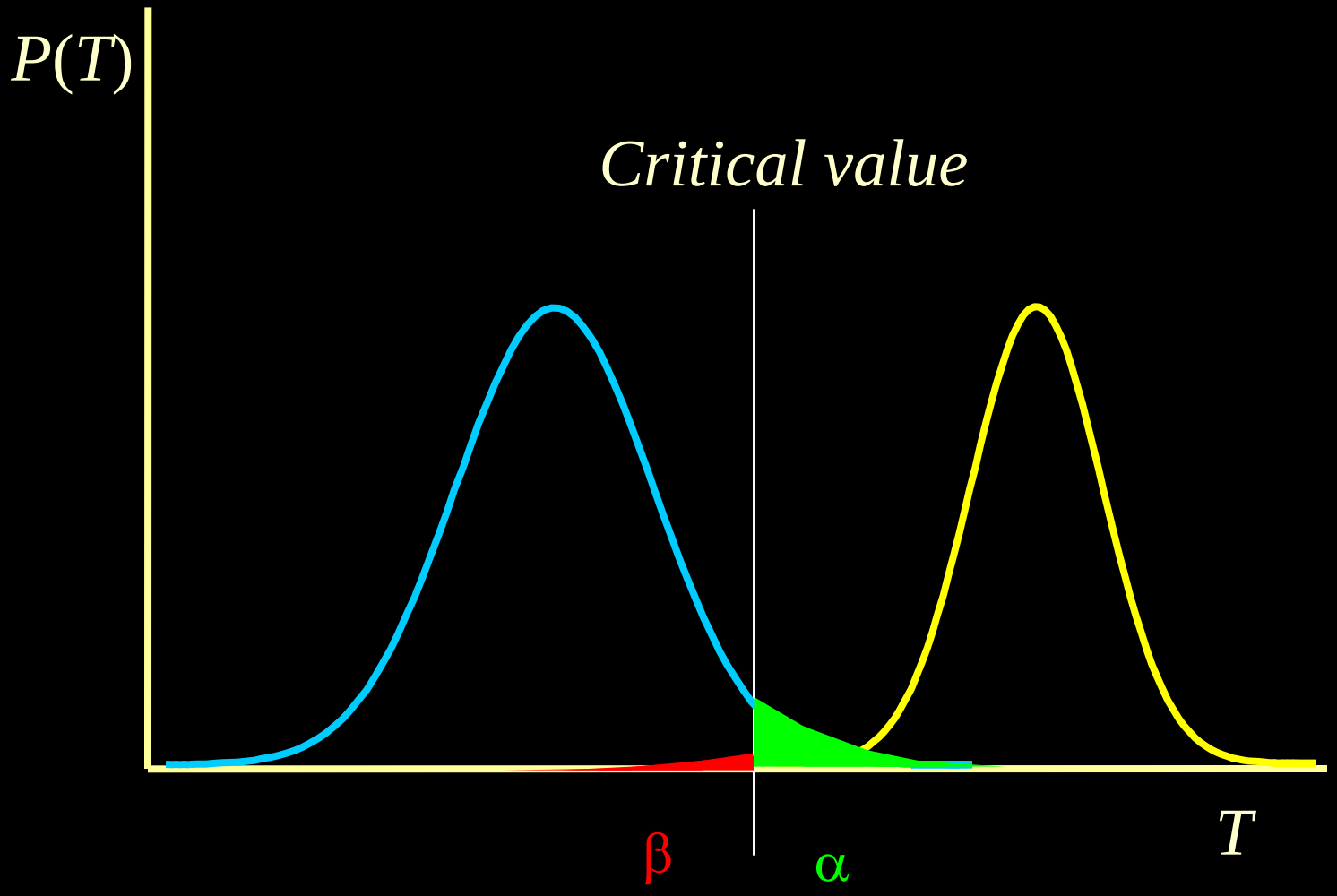
REALITY	Rejection of H_0	Nonrejection of H_0
H_0 true	Type I error at rate α 	Nonsignificant result 
H_A true	Significant result 	Type II error at rate β 

POWER $=(1-\beta)$

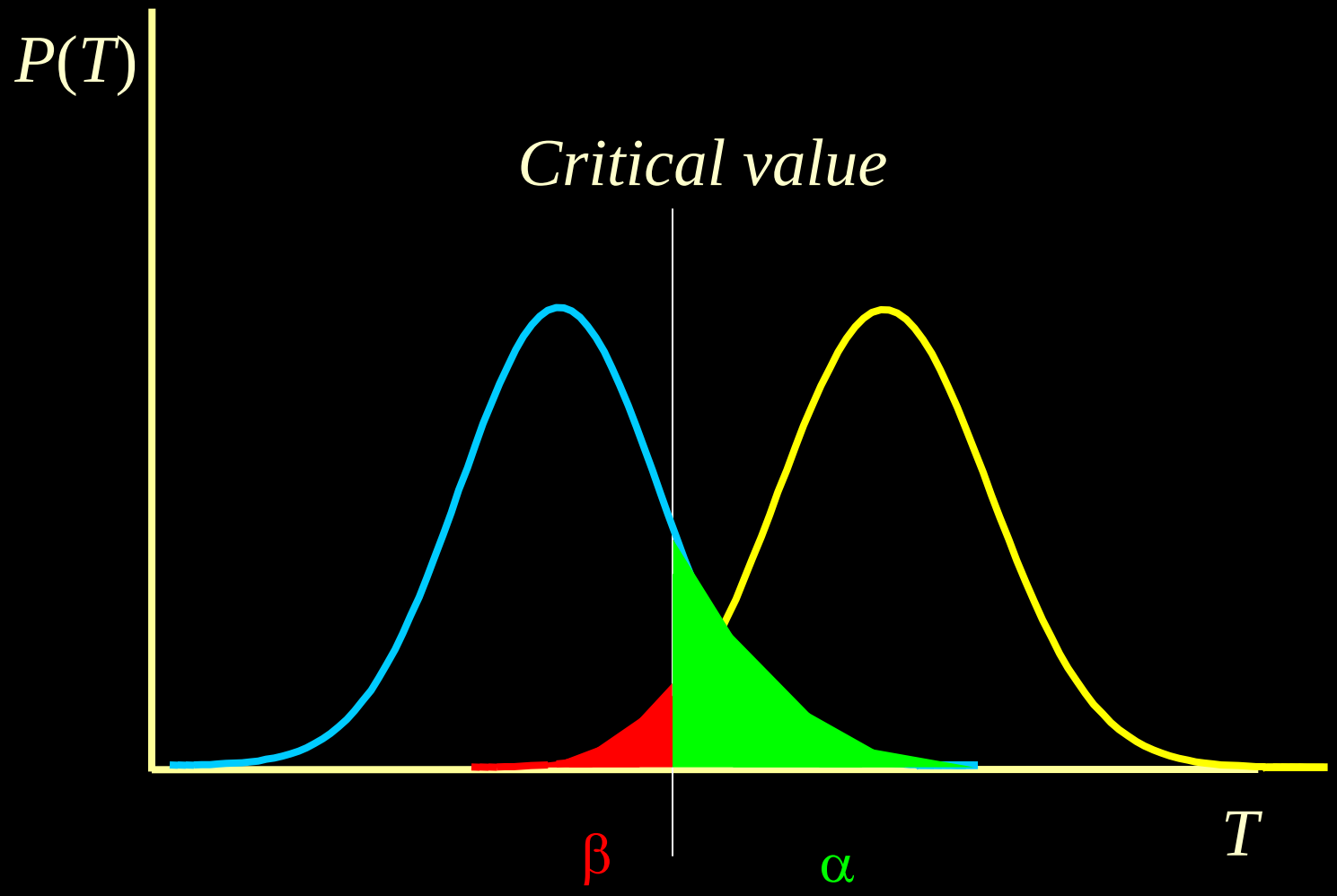
Standard Case



Impact of ↑ effect size, N

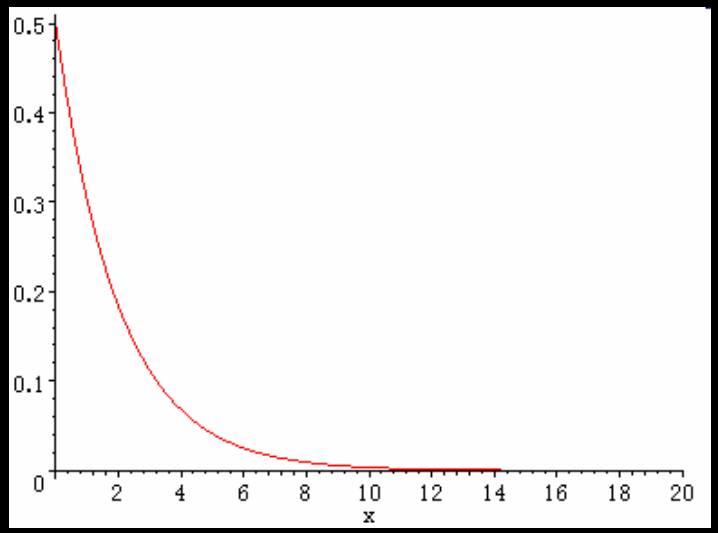


Impact of ↑ α

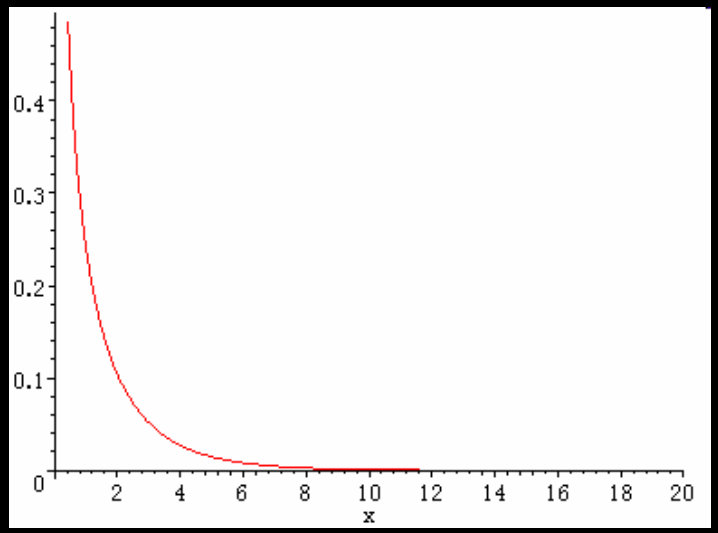


χ^2 distributions

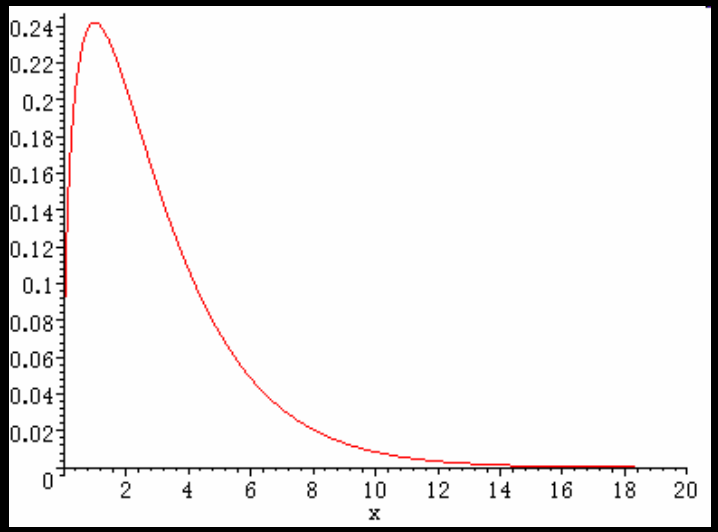
1 df



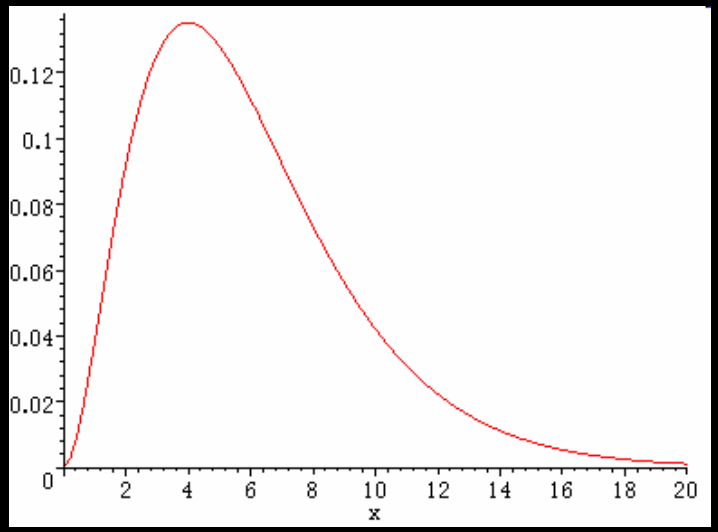
2 df



3 df



6 df



http://www2.ipcku.kansai-u.ac.jp/~aki/pdf/chi21.htm

Noncentral χ^2

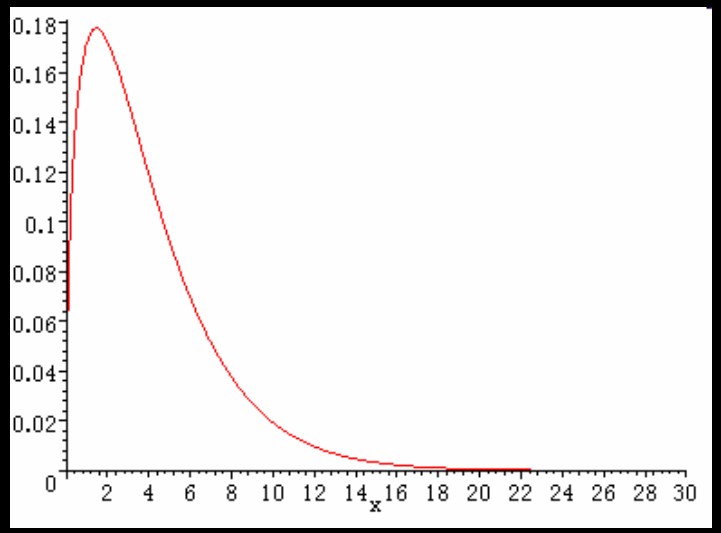
Null χ^2 has $\mu=df$ and $\sigma^2=2df$

Noncentral χ^2 has $\mu=df + \lambda$ and $\sigma^2=2df + 4\lambda$

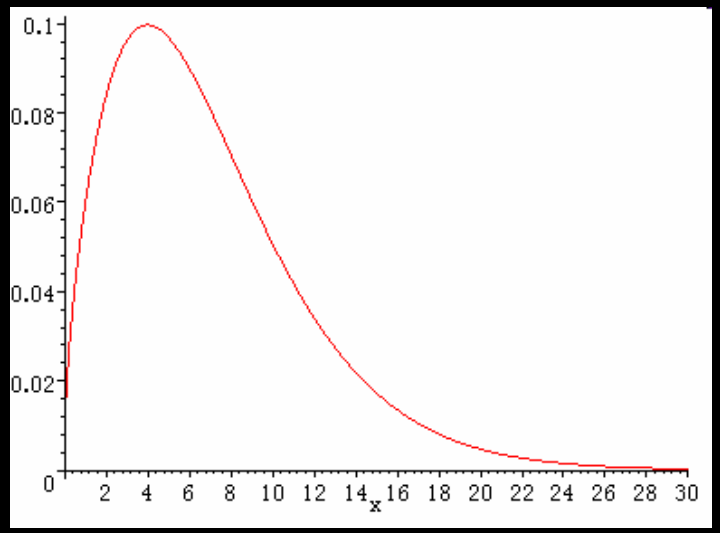
Where df are degrees of freedom and λ is the noncentrality parameter

Noncentral χ^2 3 degrees of freedom

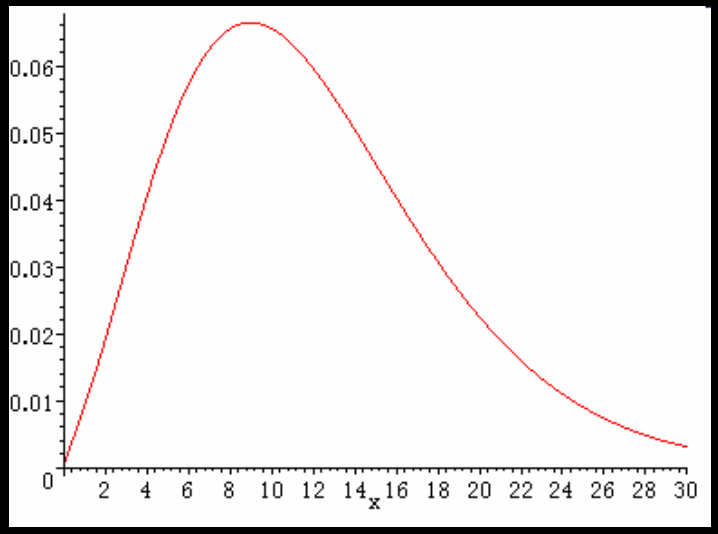
λ=1



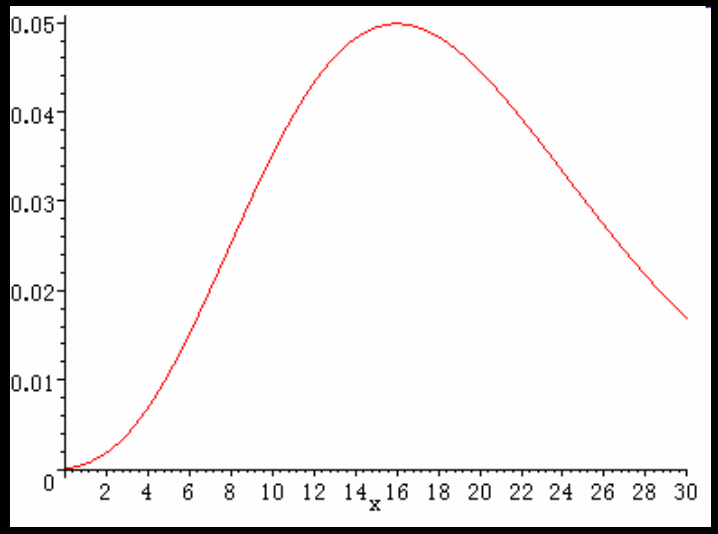
λ=4



λ=9



λ=16



http://www2.ipcku.kansai-u.ac.jp/~aki/pdf/chi21.htm

Short practical on GPC

Genetic Power Calculator is an online resource for carrying out basic power calculations

For our 1st example we will use the probability function calculator to play with power

http://ibgwww.colorado.edu/~pshaun/gpc/

Parameters in probability function calculator

Click on the link to probability function calculator

4 main terms:

X: critical value of the chi-square

P(X>x): Power

df: degrees of freedom

NCP: non-centrality parameter

Exercises

1) Find the power when NCP=5, degrees of freedom=1, and the critical X is 3.84
2) Find the NCP for power of .8, degrees of freedom=1 and critical X is 13.8

Answers

1) Power=0.608922, when NCP=5, degrees of freedom=1, and the critical X is 3.84
2) NCP=20.7613 when power of .8, degrees of freedom=1 and critical X is 13.8

2) Power for QTL linkage

For chi-squared tests on large samples, power is determined by non-centrality parameter (λ) and degrees of freedom (df)

$$\lambda = E(2\ln L_A - 2\ln L_0)$$

$$= E(2\ln L_A) - E(2\ln L_0)$$

where expectations are taken at asymptotic values of maximum likelihood estimates (MLE) under an assumed true model

Linkage test

$$2\ln L = -\ln\left|\Sigma\right| - x'\Sigma^{-1}x$$

H_A

$$\left[\Sigma_L\right]_{ij} = \begin{cases} V_A + V_D + V_S + V_N & \text{for i=j} \\ \hat{\pi}V_A + \hat{z}V_D + V_S & \text{for i≠j} \end{cases}$$

H_0

$$\left[\Sigma_N\right]_{ij} = \begin{cases} V_A + V_D + V_S + V_N & \text{for i=j} \\ \frac{V_A}{2} + \frac{V_D}{4} + V_S & \text{for i≠j} \end{cases}$$

Linkage test

Expected NCP

$$\lambda = \ln\left|\Sigma_0\right| - \sum_{i=1}^{m} P_i \ln\left|\Sigma_i\right|$$

For sib-pairs under complete marker information

$$\lambda = \ln\left|\Sigma_0\right| - \left[\tfrac{1}{4}\ln\left|\Sigma_{\pi=0}\right| + \tfrac{1}{2}\ln\left|\Sigma_{\pi=1}\right| + \tfrac{1}{4}\ln\left|\Sigma_{\pi=2}\right|\right]$$

Determinant of 2-by-2 standardised covariance matrix = 1 - r^2

$$\lambda_L = -\frac{1}{4}\ln(1-r_0^2) - \frac{1}{2}\ln(1-r_1^2) - \frac{1}{4}\ln(1-r_2^2) + \ln(1-r_S^2)$$

Note: standardised trait

See Sham et al (2000) AJHG, 66. for further details

Concrete example

200 sibling pairs; sibling correlation 0.5.

To calculate NCP if QTL explained 10% variance:

$$\lambda_L = -\frac{1}{4}\ln(1-r_0^2) - \frac{1}{2}\ln(1-r_1^2) - \frac{1}{4}\ln(1-r_2^2) + \ln(1-r_S^2)$$

$$= -\frac{1}{4}\ln(1-0.45^2) - \frac{1}{2}\ln(1-0.5^2) - \frac{1}{4}\ln(1-0.55^2) + \ln(1-0.5^2)$$

$$= 0.0565 + 0.1438 + 0.0900 - 0.2877$$

$$= 0.002791$$

200 × 0.002791 = **0.5581**

Approximation of NCP

$$NCP \approx \frac{s(s-1)}{2}\frac{(1+r^2)}{(1-r^2)^2}Var(r_\pi)$$

$$\approx \frac{s(s-1)}{2}\frac{(1+r^2)}{(1-r^2)^2}\left[V_A^2 Var(\pi) + V_D^2 Var(z) + V_A V_D Cov(\pi, z)\right]$$

NCP per sibship is proportional to

- the # of pairs in the sibship
  (large sibships are powerful)

- the square of the additive QTL variance
  (decreases rapidly for QTL of v. small effect)

- the sibling correlation
  (structure of residual variance is important)

Using GPC

Comparison to Haseman-Elston regression linkage

Amos & Elston (1989) H-E regression

- 90% power (at significant level 0.05)
- QTL variance 0.5
- marker & major gene completely linked ($\theta = 0$)

$\rightarrow$ 320 sib pairs

- if $\theta = 0.1$

$\rightarrow$ 778 sib pairs

GPC input parameters

Proportions of variance

- additive QTL variance
- dominance QTL variance
- residual variance (shared / nonshared)

Recombination fraction (0 - 0.5)

Sample size & Sibship size (2 - 8)

Type I error rate

Type II error rate

GPC output parameters

Expected sibling correlations

- by IBD status at the QTL
- by IBD status at the marker

Expected NCP per sibship

Power

- at different levels of alpha given sample size

Sample size

- for specified power at different levels of alpha given power

GPC

http://ibgwww.colorado.edu/~pshaun/gpc/

Practical 2

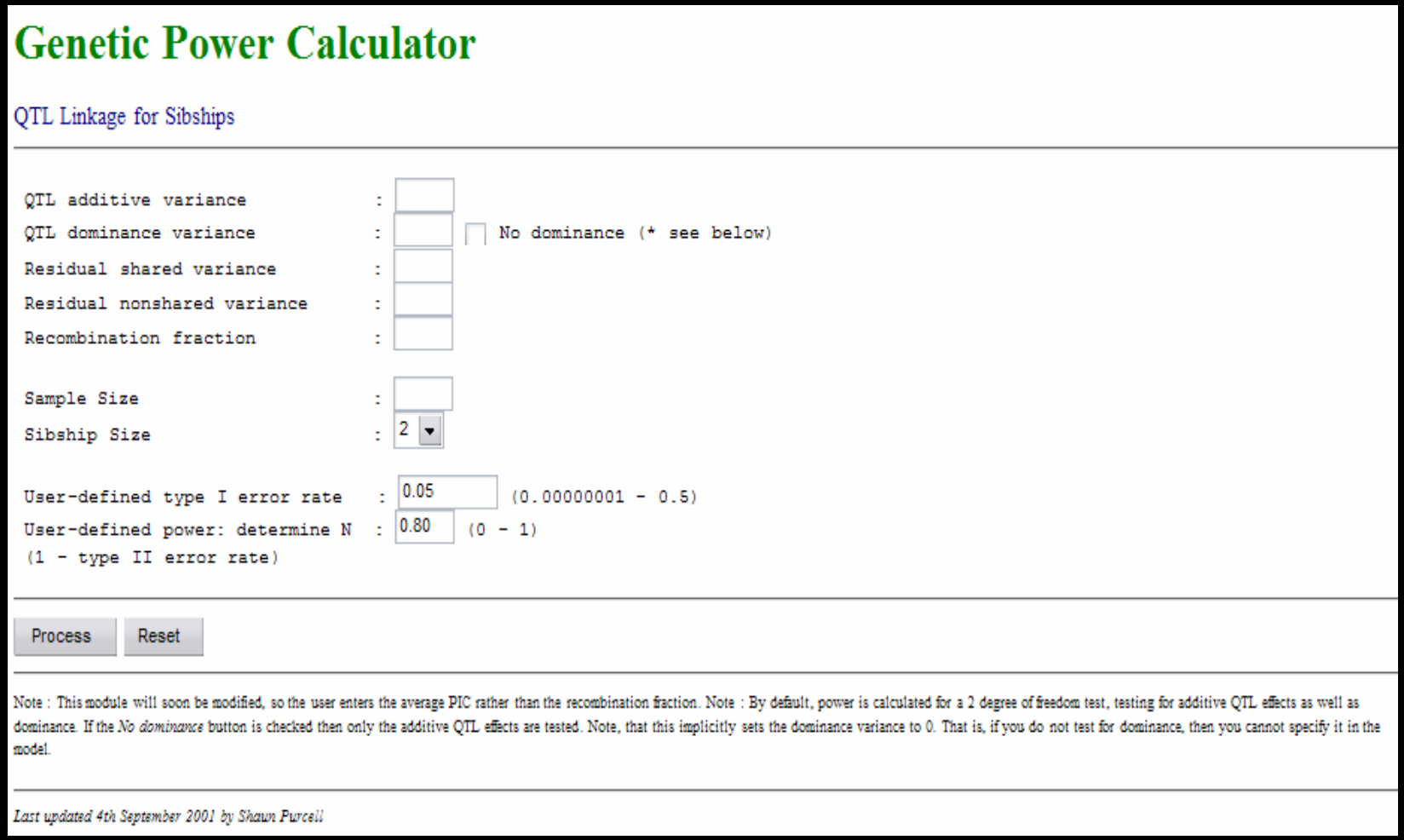
Using GPC, what is the effect on power to detect linkage of :

1. QTL variance?

2. residual sibling correlation?

3. marker QTL recombination fraction?

GPC Input



Genetic Power Calculator

QTL Linkage for Sibships

QTL additive variance :

QTL dominance variance : No dominance (* see below)

Residual shared variance :

Residual nonshared variance :

Recombination fraction :

Sample Size :

Sibship Size : 2

User-defined type I error rate : 0.05 (0.00000001 - 0.5)

User-defined power: determine N : 0.80 (0 - 1)
(1 - type II error rate)

Process Reset

Note : This module will soon be modified, so the user enters the average PIC rather than the recombination fraction. Note : By default, power is calculated for a 2 degree of freedom test, testing for additive QTL effects as well as dominance. If the *No dominance* button is checked then only the additive QTL effects are tested. Note, that this implicitly sets the dominance variance to 0. That is, if you do not test for dominance, then you cannot specify it in the model.

Last updated 4th September 2001 by Shawn Purcell

GPC output

Genetic Power Calculator

QTL Linkage : Sibships

Proportions of variance at QTL

Additive QTL variance	0.1818
Dominance QTL variance	0.1818
Shared residual variance	0.2727
Nonshared residual variance	0.3636

Sibling correlations by IBD status at QTL

IBD 0	0.2727
IBD 1	0.3636
IBD 2	0.6364

Sibling correlations by IBD status at marker

IBD 0	0.3113
IBD 1	0.3905
IBD 2	0.5441

Misc. statistics

Sibship Size	2
Sample Size	2000
Recombination fraction	0.1

Test Statisitics : Power Analysis

QTL Linkage NCP = 25.61

Alpha	Power	Sample for 80% power
0.1	0.9997	443.6
0.05	0.9989	602.1
0.01	0.9917	943.4
0.001	0.951	1402
0.05	0.9989	602.1

All tests are for additive and dominance effects (2 df)

Practical 2

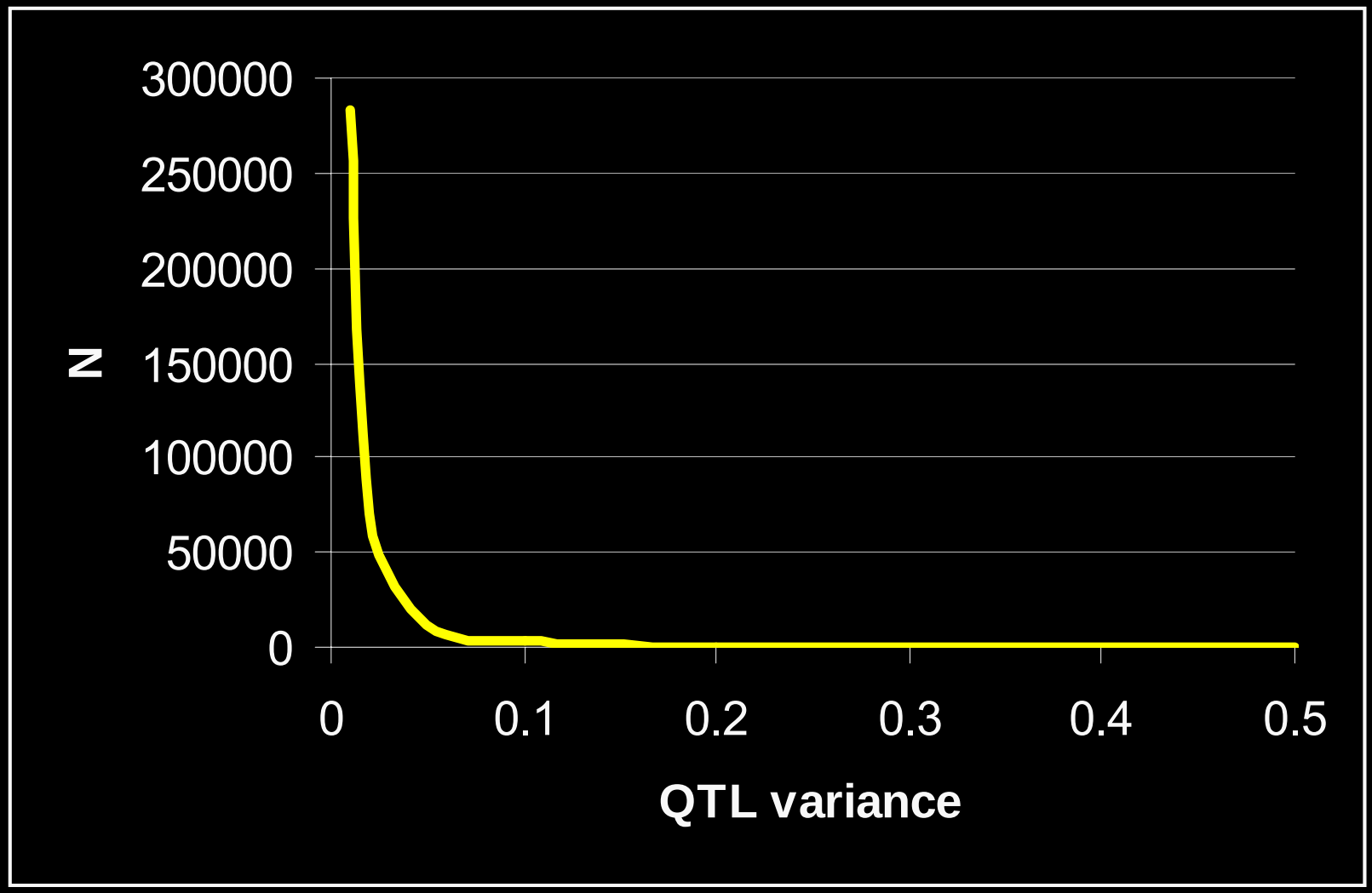
1) One good way of understanding power is to start with a basic case and then change relevant factors in both directions one at a time
2) Let's begin with a basic case of:
    1) Additive QTL .15
    2) No dominance (check the box)
    3) Residual shared variance .35
    4) Residual nonshared environment .5
    5) Recombination fraction .1
    6) Sample size 200
    7) Sibship size 2
    8) User-defined Type I error rate .0001
    9) User-defined power .8

GPC

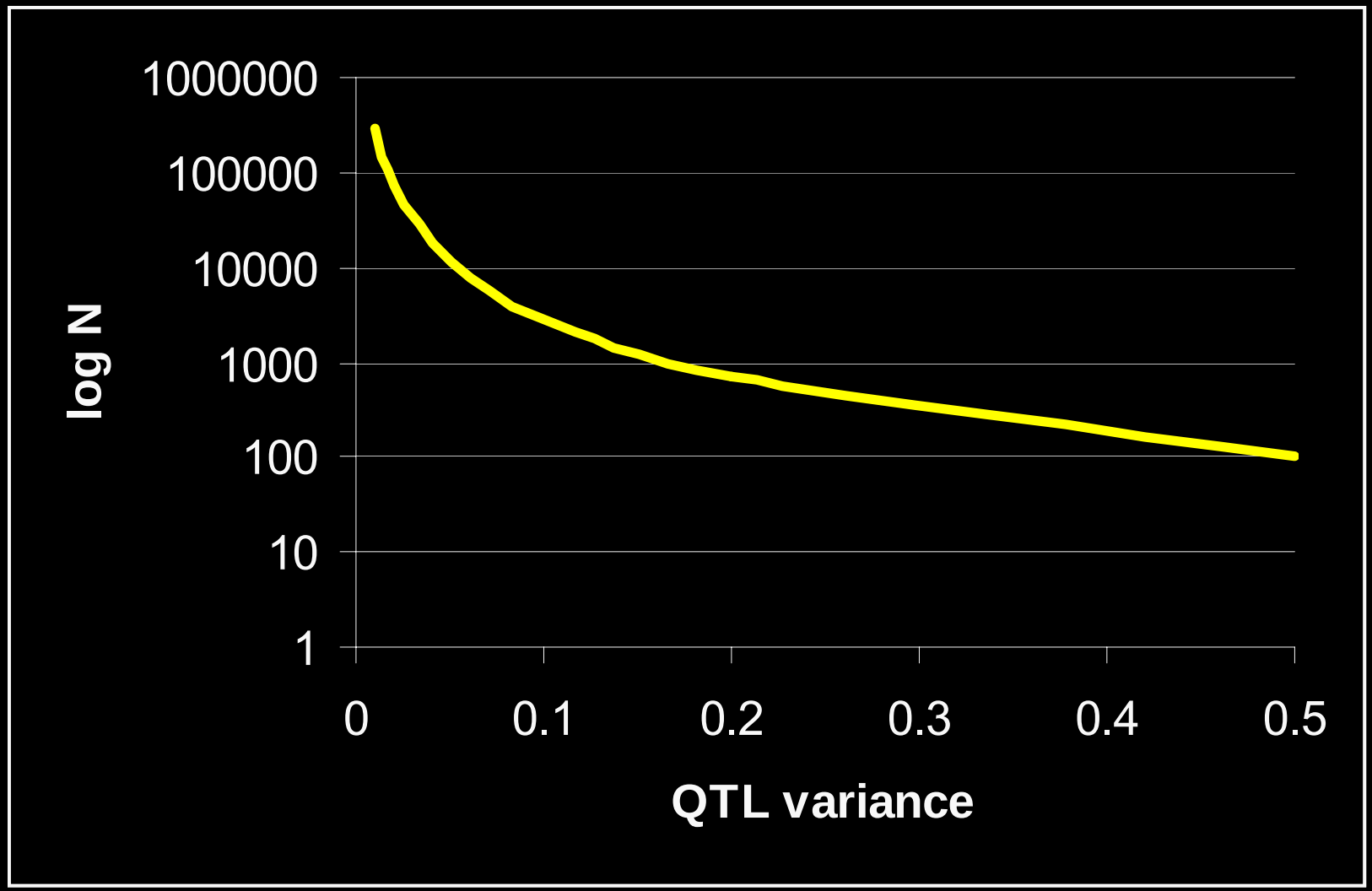
What happens when you vary:

1. QTL variance
2. Dominance vs. additive QTL variance
3. Residual sibling shared variance
4. Recombination fraction
5. Sibship sizes

Pairs required
(θ=0, p=0.05, power=0.8)



Pairs required
(θ=0, p=0.05, power=0.8)



Effect of residual correlation

QTL additive effects account for 10% trait variance

Sample size required for 80% power (α=0.05)

No dominance

θ = 0.1

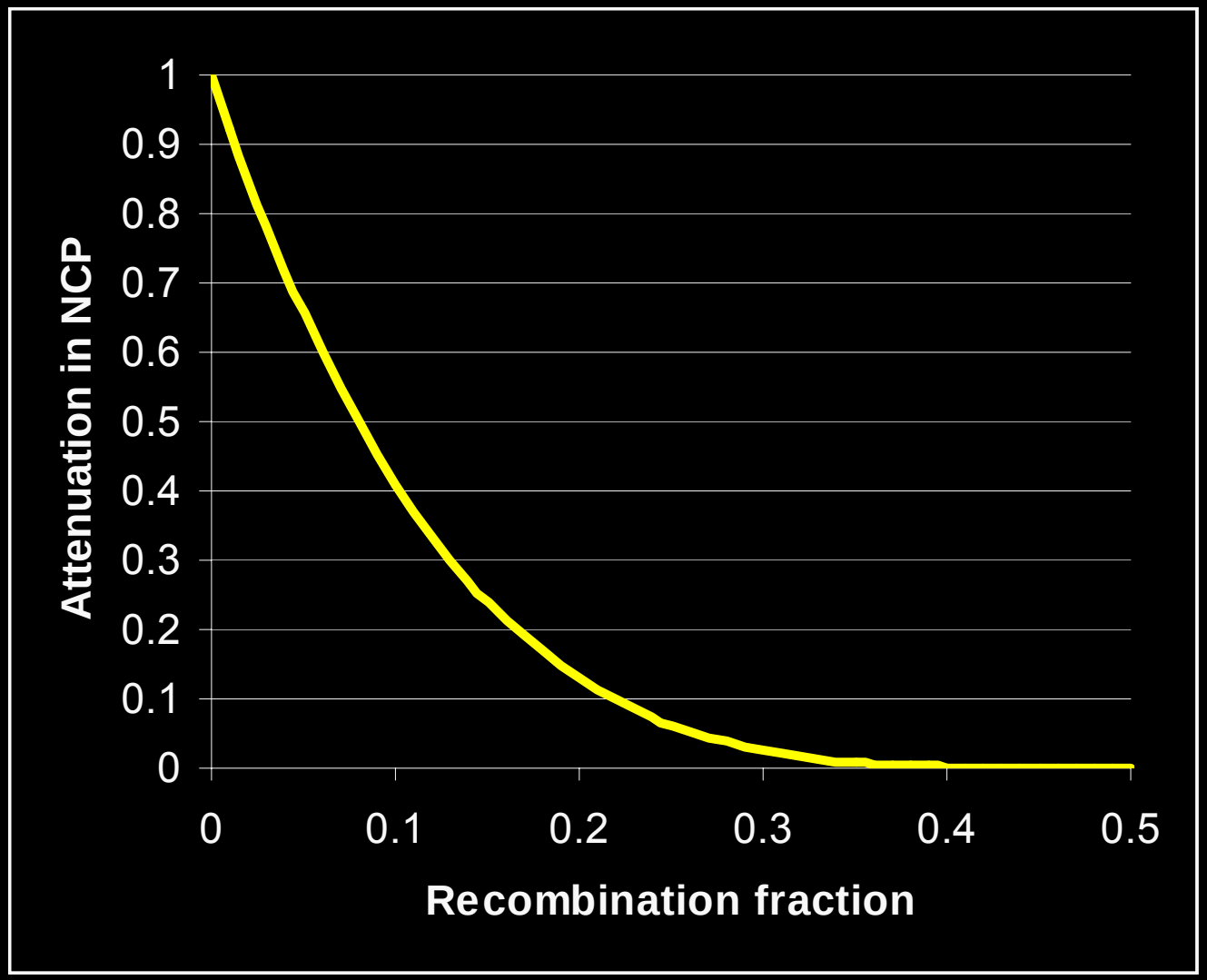
A residual correlation 0.35

B residual correlation 0.50

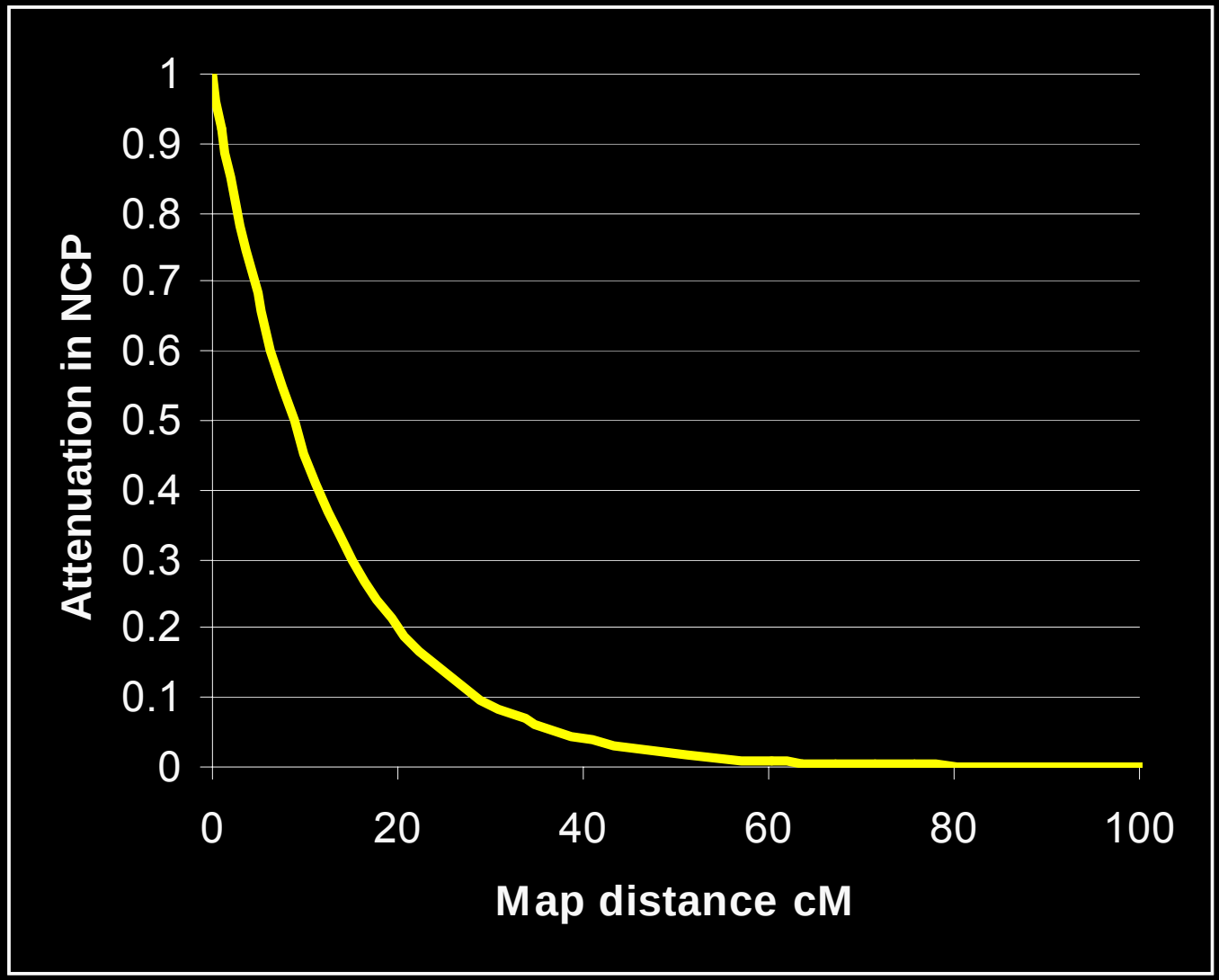
C residual correlation 0.65

Individuals required

Effect of incomplete linkage



Effect of incomplete linkage



Some factors influencing power

1. QTL variance
2. Sib correlation
3. Sibship size

4. Marker informativeness & density
5. Phenotypic selection

Marker informativeness:

Markers should be highly polymorphic

- alleles inherited from different sources are likely to be distinguishable

Heterozygosity (H)

Polymorphism Information Content (PIC)

- measure number and frequency of alleles at a locus

Polymorphism Information Content

IF a parent is heterozygous,

their gametes will usually be informative.

BUT if both parents & child are heterozygous for the same genotype,

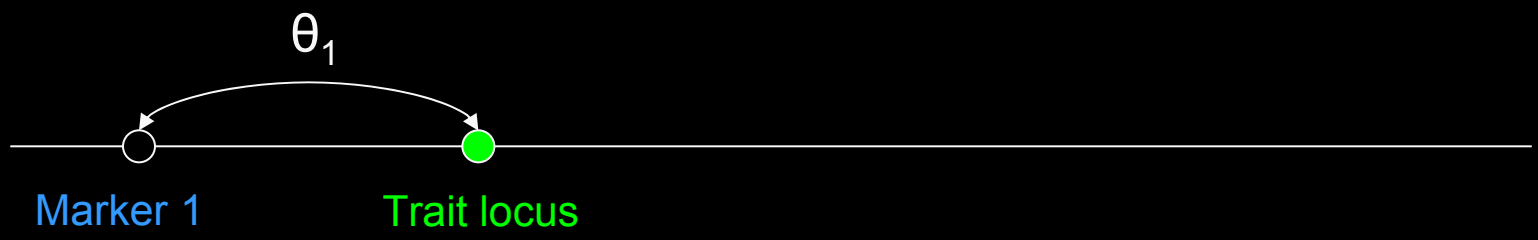
origins of child's alleles are ambiguous

IF C = the probability of this occurring,

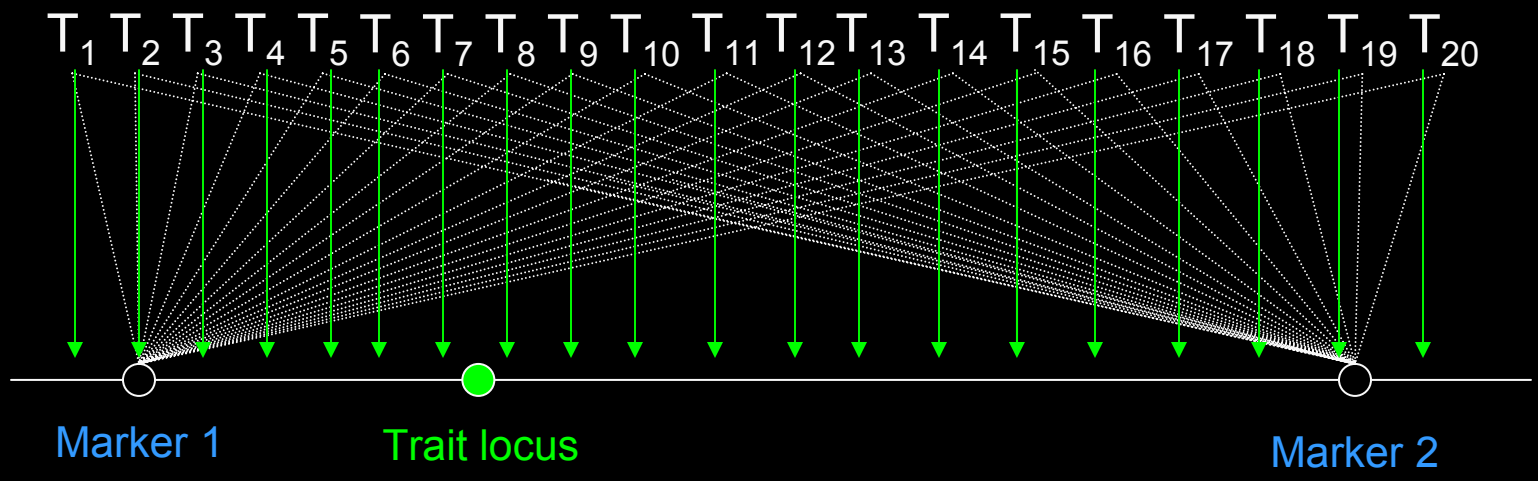
$$PIC = H - C$$

$$= 1 - \sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n} \sum_{j=i+1}^{n} 2 p_i^2 p_j^2$$

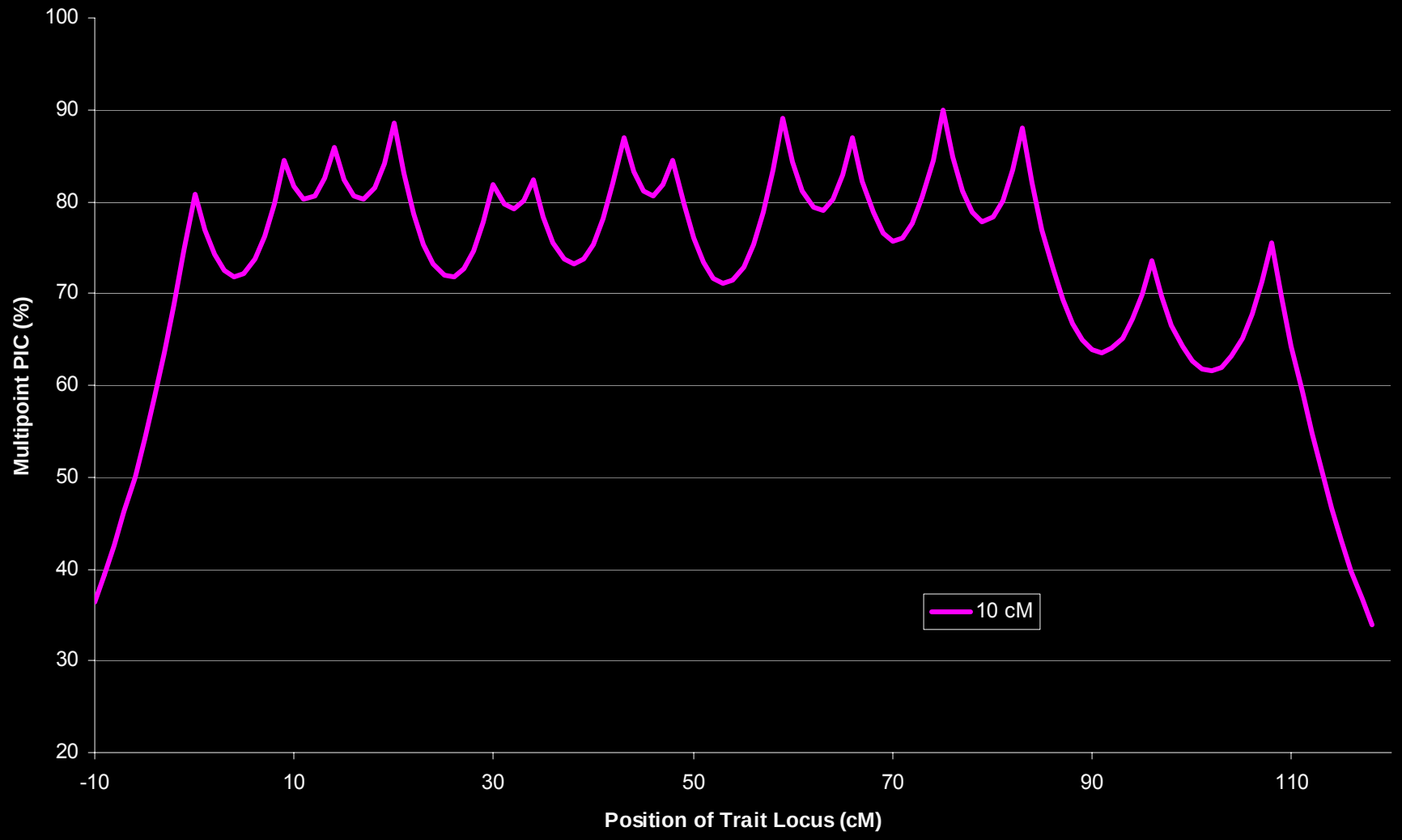
Singlepoint



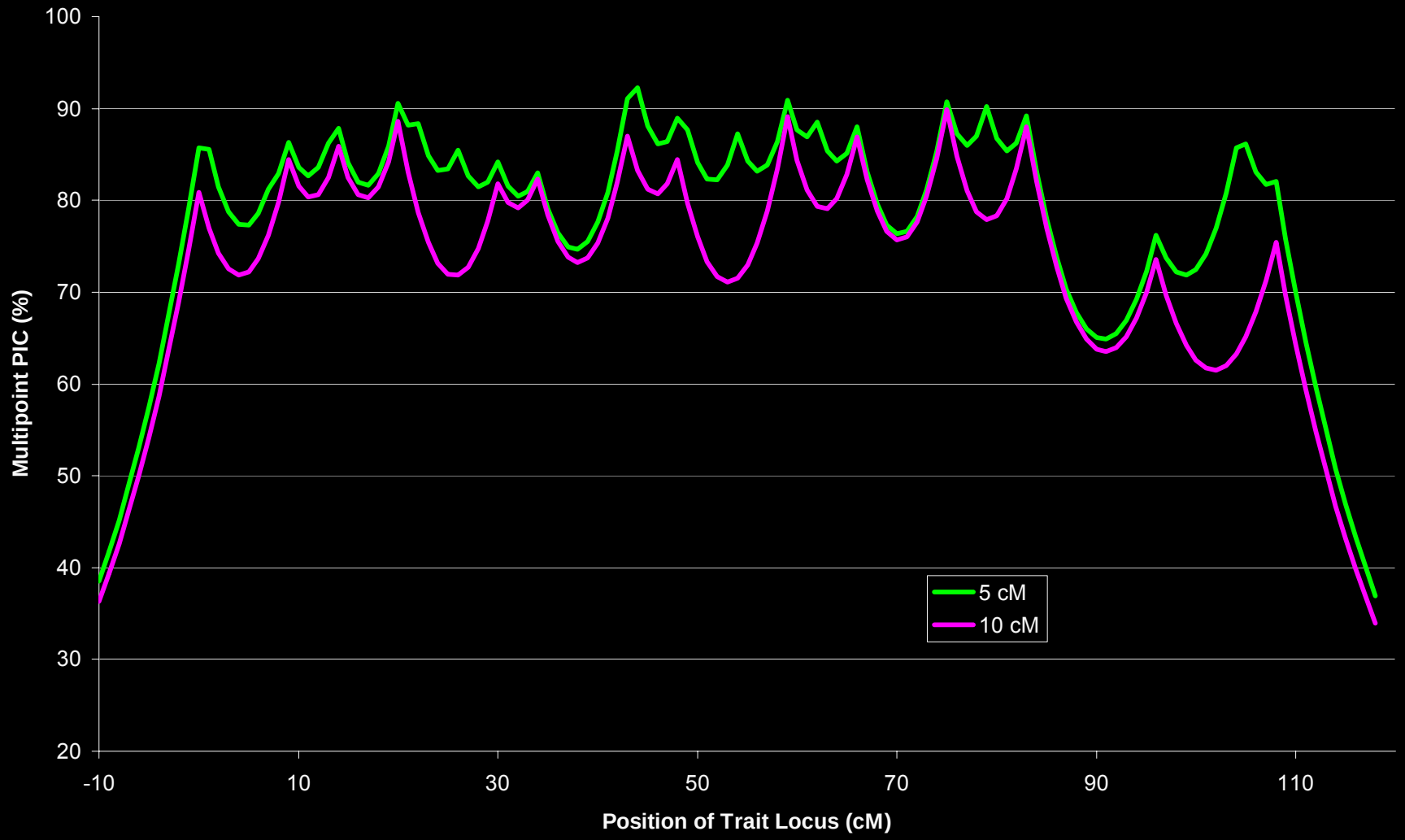
Multipoint



Multipoint PIC: 10 cM map



Multipoint PIC: 5 cM map



Statistical Genetics Group - Mozilla

File Edit View Go Bookmarks Tools Window Help

Back Forward Reload Stop http://statgen.iop.kcl.ac.uk/gpc/mpic.html

Home Bookmarks The Mozilla Organization Latest Builds

Genetic Power Calculator

MPIC: Multipoint Polymorphism Information Content

Implements the Rijsdijk & Sham MPIC method.

IMPORTANT: Please ensure there are no blank lines in either text box (including the last line). Also, please whitespace at the end of each line.

See notes for guidance on input parameters.

Locus Specification

```
3
2 0.5 0.5
2 0.5 0.5
2 0.5 0.5
```

Map Distance

```
10
10
```

Process Reset

The Singlepoint Information Content of the markers:

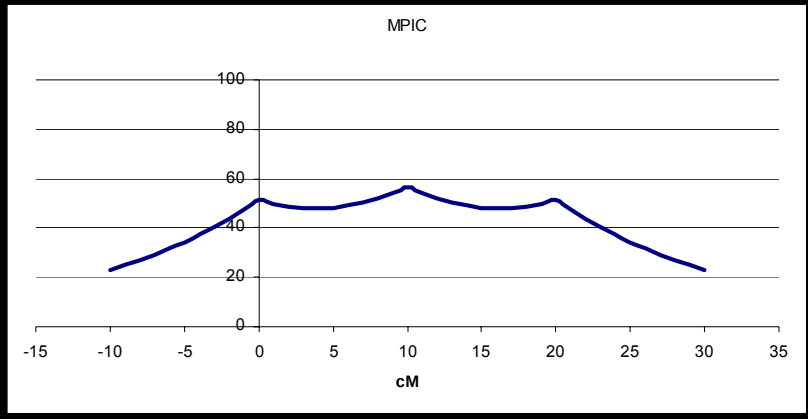
Locus 1 PIC = 0.375
Locus 2 PIC = 0.375
Locus 3 PIC = 0.375

The Multipoint Information Content of the markers:

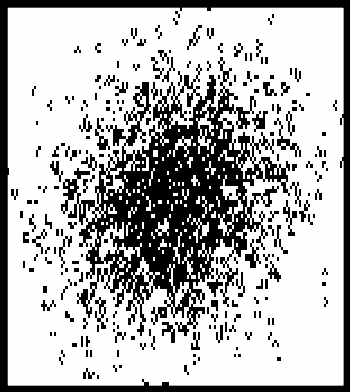
Pos	MPIC
-10	22.9946
-9	24.9097
-8	26.9843
-7	29.2319
-6	31.6665
-5	34.304
-4	37.1609
-3	40.256
-2	43.6087
-1	47.2408
0	51.1754
1	49.6898

…

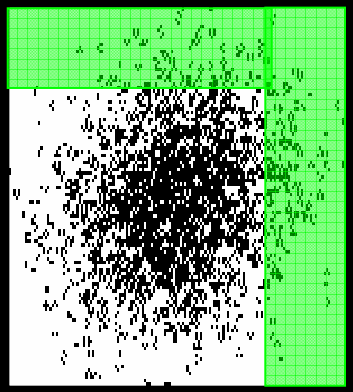
meaninf 50.2027



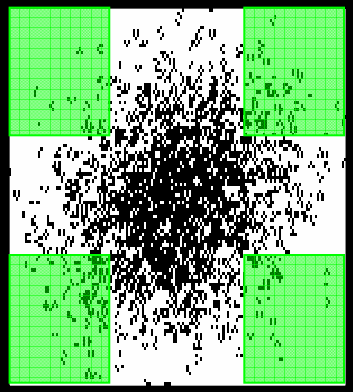
Selective genotyping



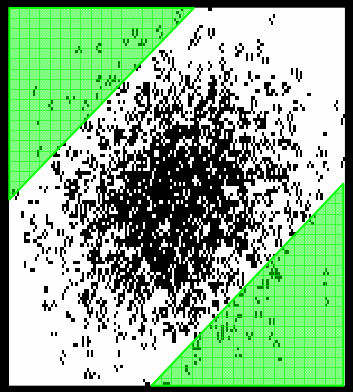
Unselected



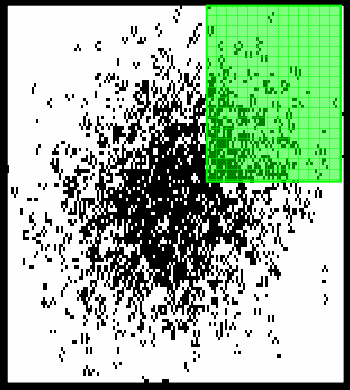
Proband Selection



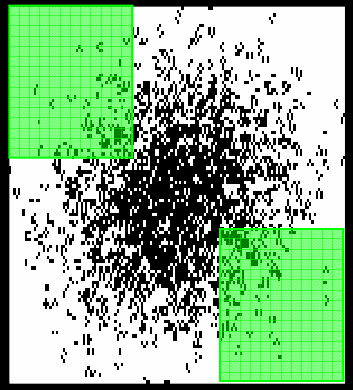
EDAC



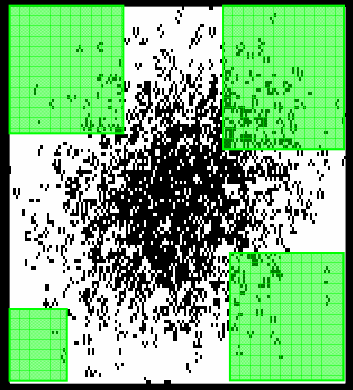
Maximally Dissimilar



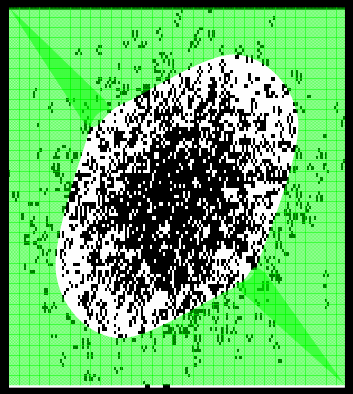
ASP



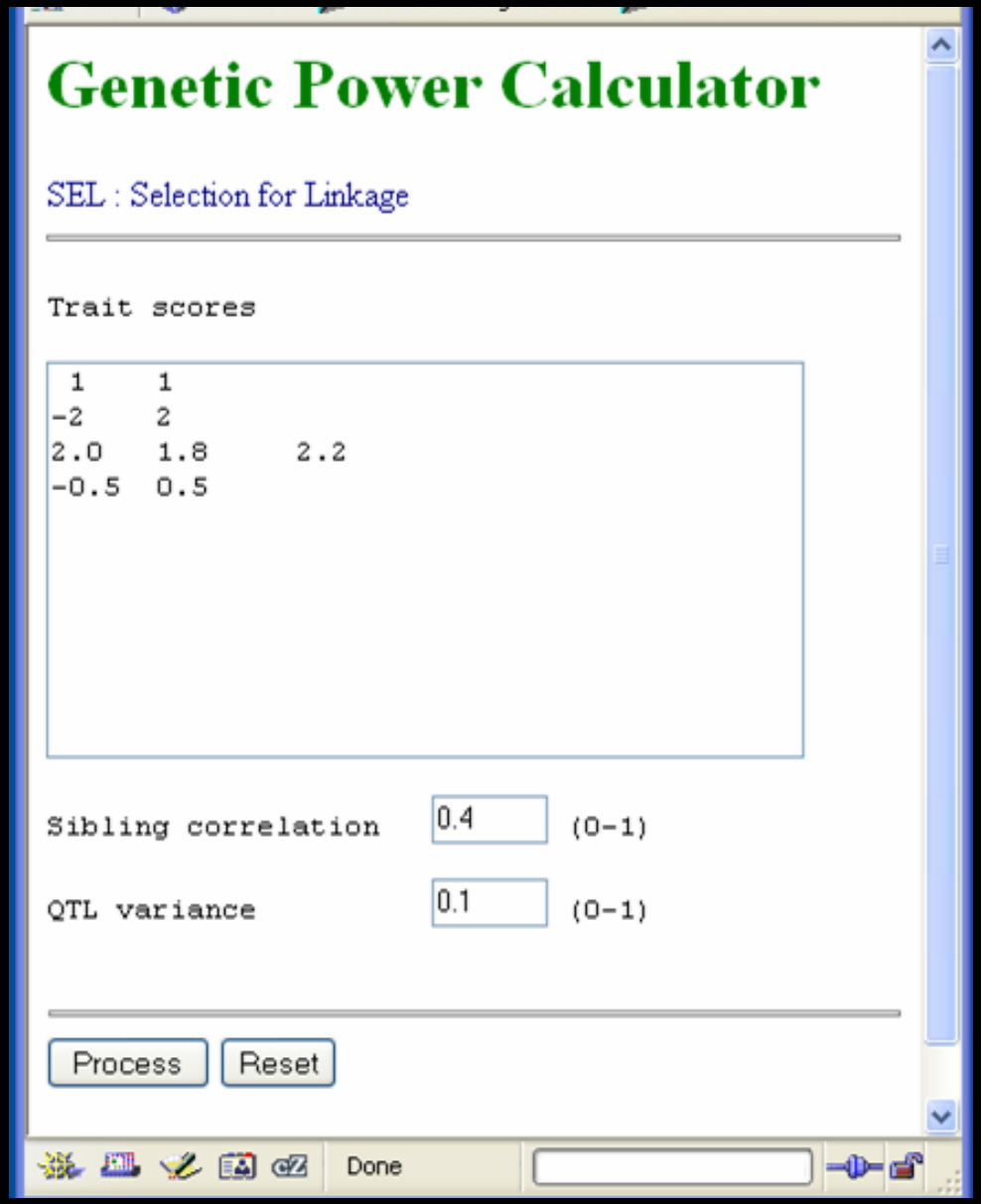
Extreme Discordant



EDAC

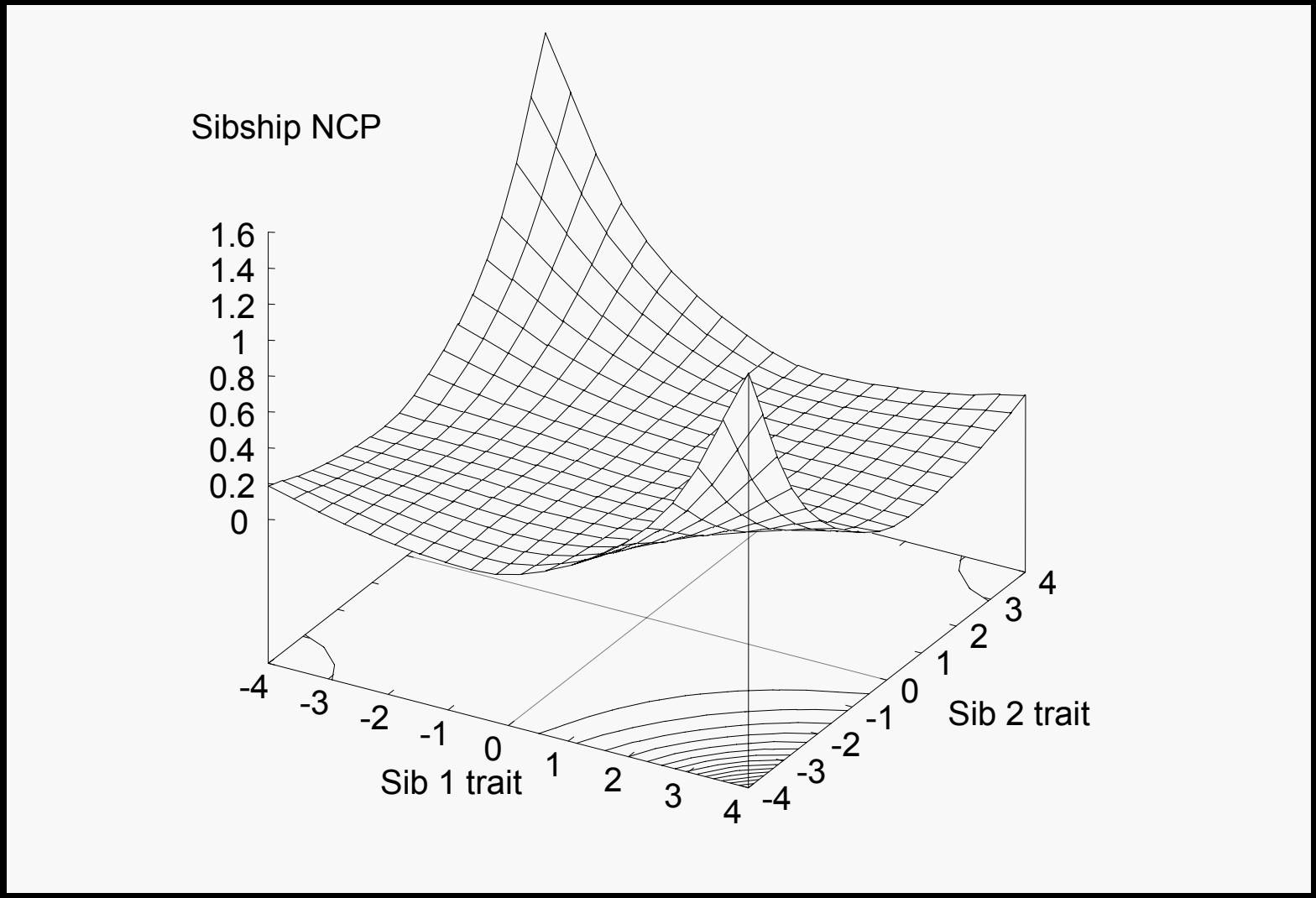


Mahalanobis Distance

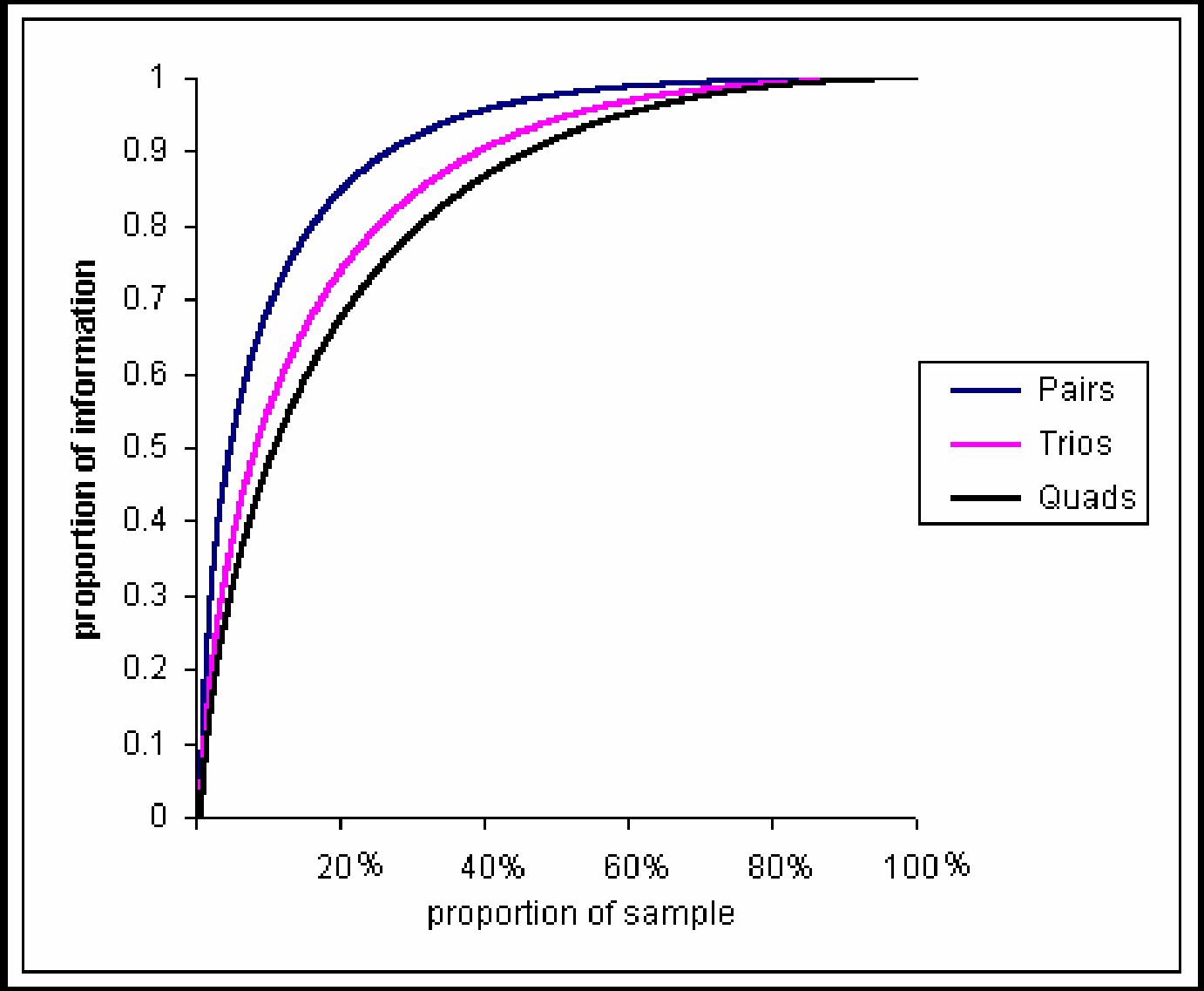


E(-2LL)	Sib 1	Sib 2	Sib 3
0.00121621	1.00	1.00	
0.14137692	-2.00	2.00	
0.00957190	2.00	1.80	2.20
0.00005954	-0.50	0.50	

Sibship informativeness : sib pairs



Impact of selection



QTL power using Mx

- Power can be calculated theoretically or empirically
- We have shown the theoretical power calculations from Sham et al. 2000
- Empirical calculations can be computed in Mx or from simulated data
  - Most of us are too busy (short IQ pts.) to figure out the theoretical power calculation so empirical is useful

Mx power script

1) Download the script powerFEQ.mx
2) I'll open it and walk through precisely what Mx is doing
3) Briefly, Mx requires that you set up the model under the 'true model', using algebra generating the variance covariance matrices
4) Refit the model from the variance covariance models fixing the parameter you wish to test to 0.
5) At end of script include the option power= α, df

Same again with raw data

Mx can now estimate the power distribution from raw data. The change in likelihood is taken to be the NCP and this governs the power.

Download realFEQpower.mx and we will use the lipidall.dat data from Danielle's session.

I've highlighted position 79—the maximum.

Summary

The power of linkage analysis is related to:

1. QTL variance
2. Sib correlation
3. Sibship size
4. Marker informativeness & density
5. Phenotypic selection

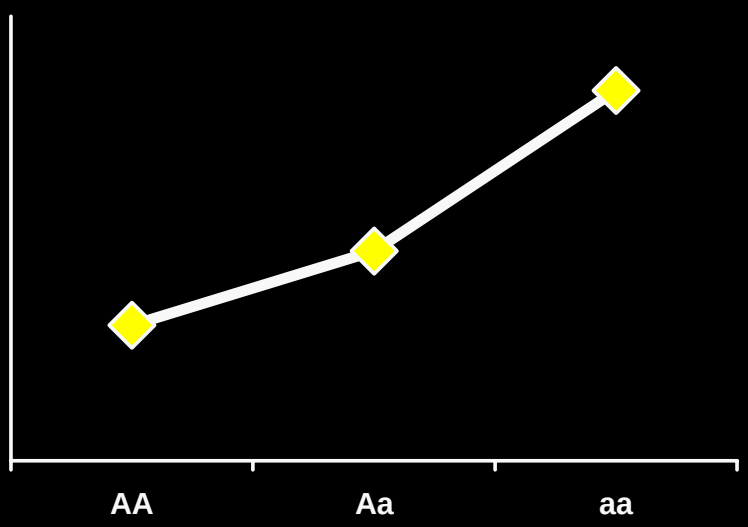
If we have time slide

We'll move on to 2 locus models

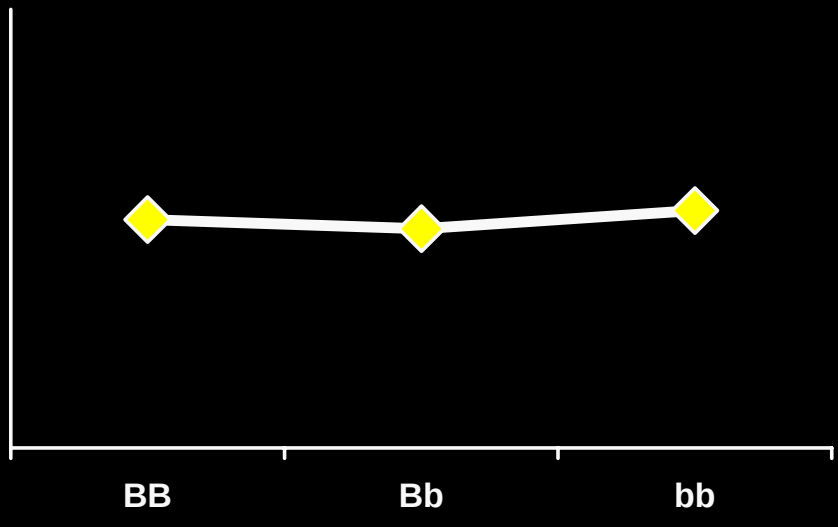
3) Single additive locus model

locus A shows an association with the trait

locus B appears unrelated



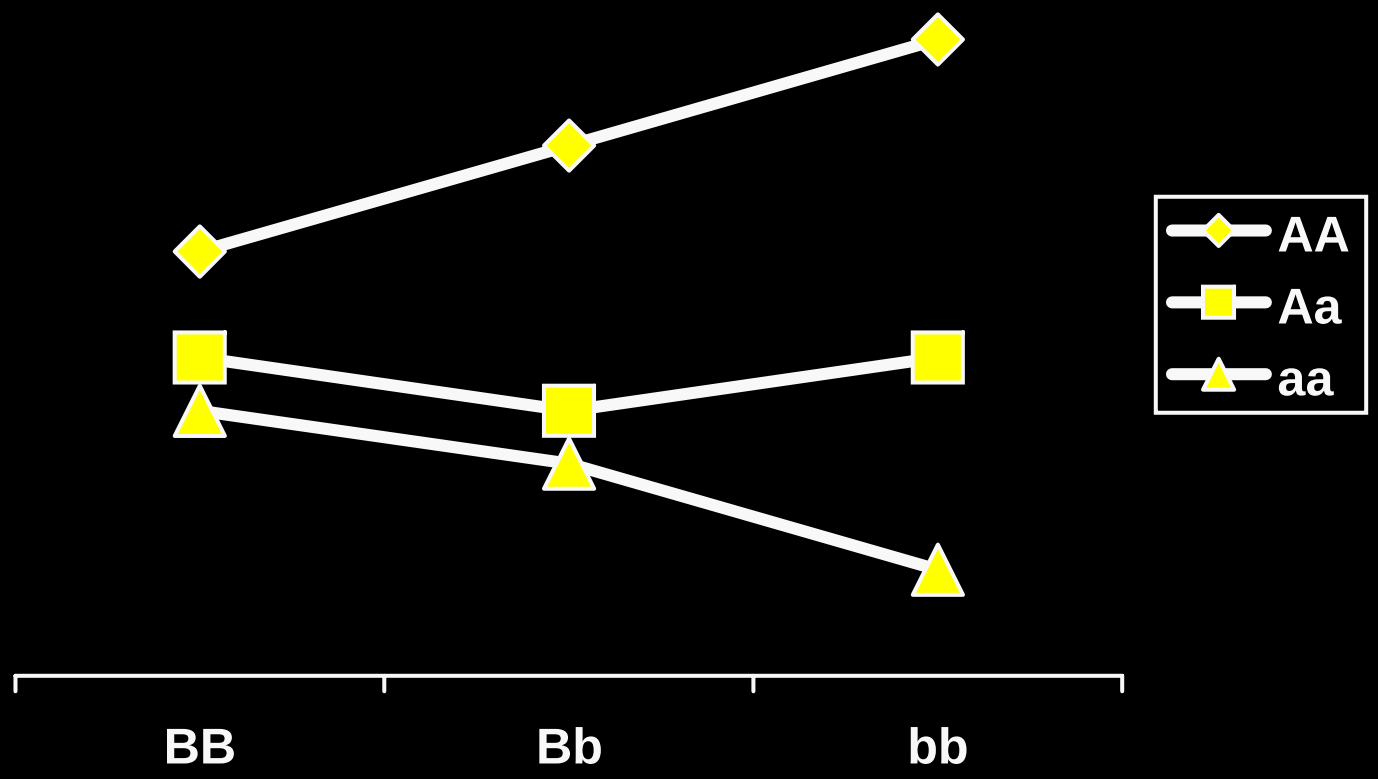
Locus A



Locus B

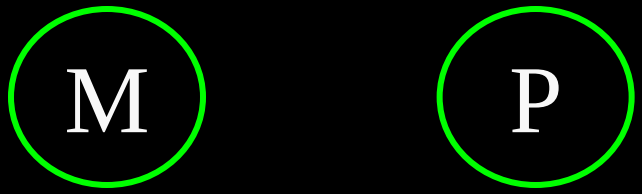
Joint analysis

locus B modifies the effects of locus A: epistasis

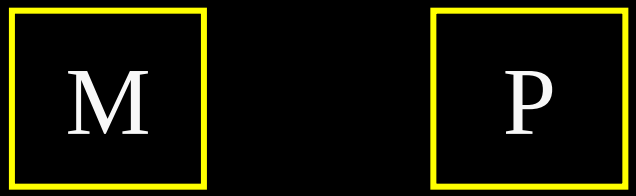


Partitioning of effects

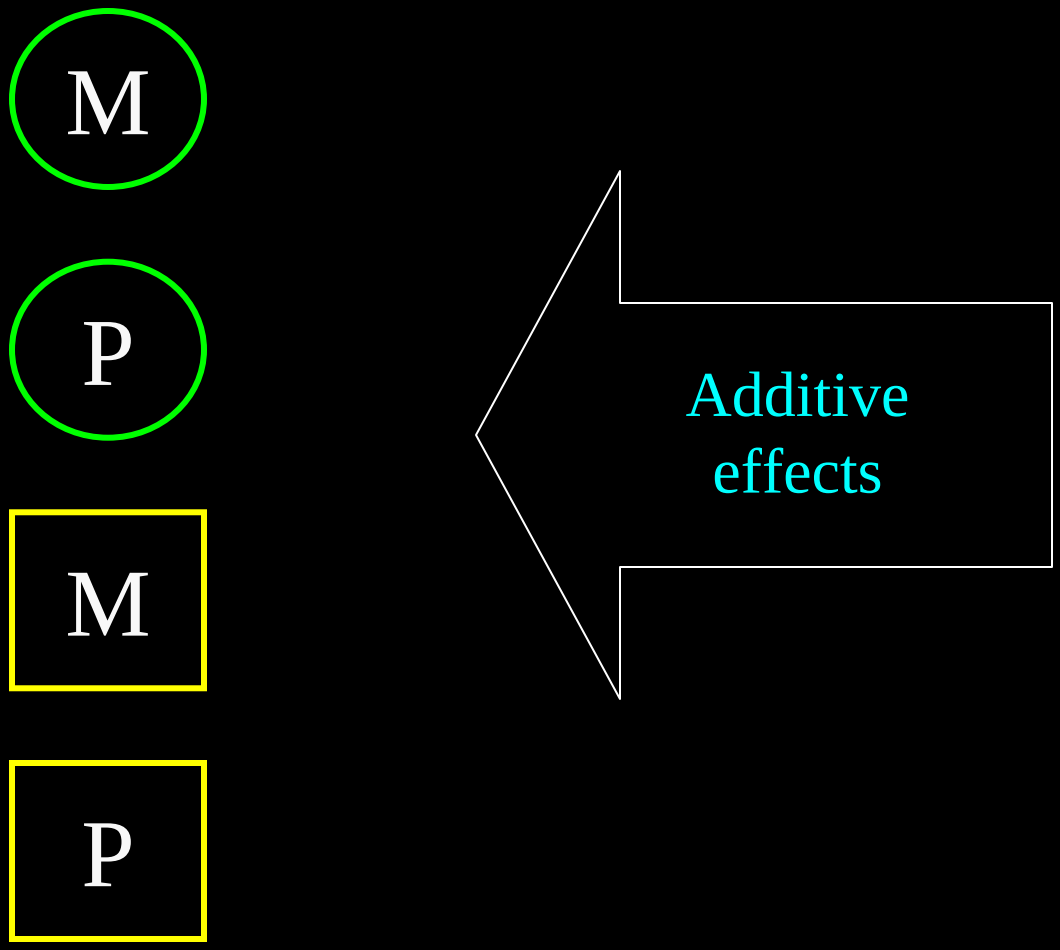
Locus A



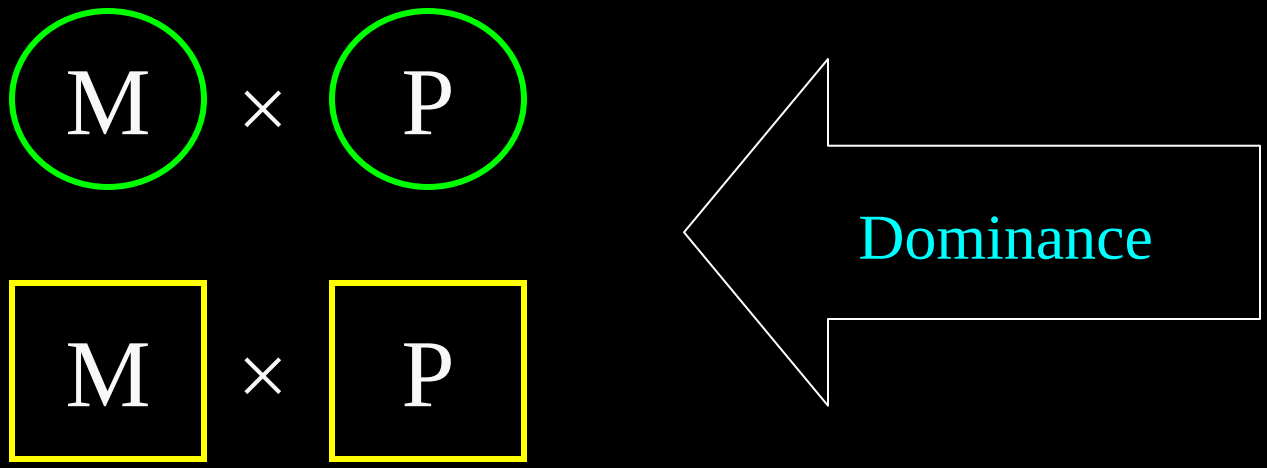
Locus B



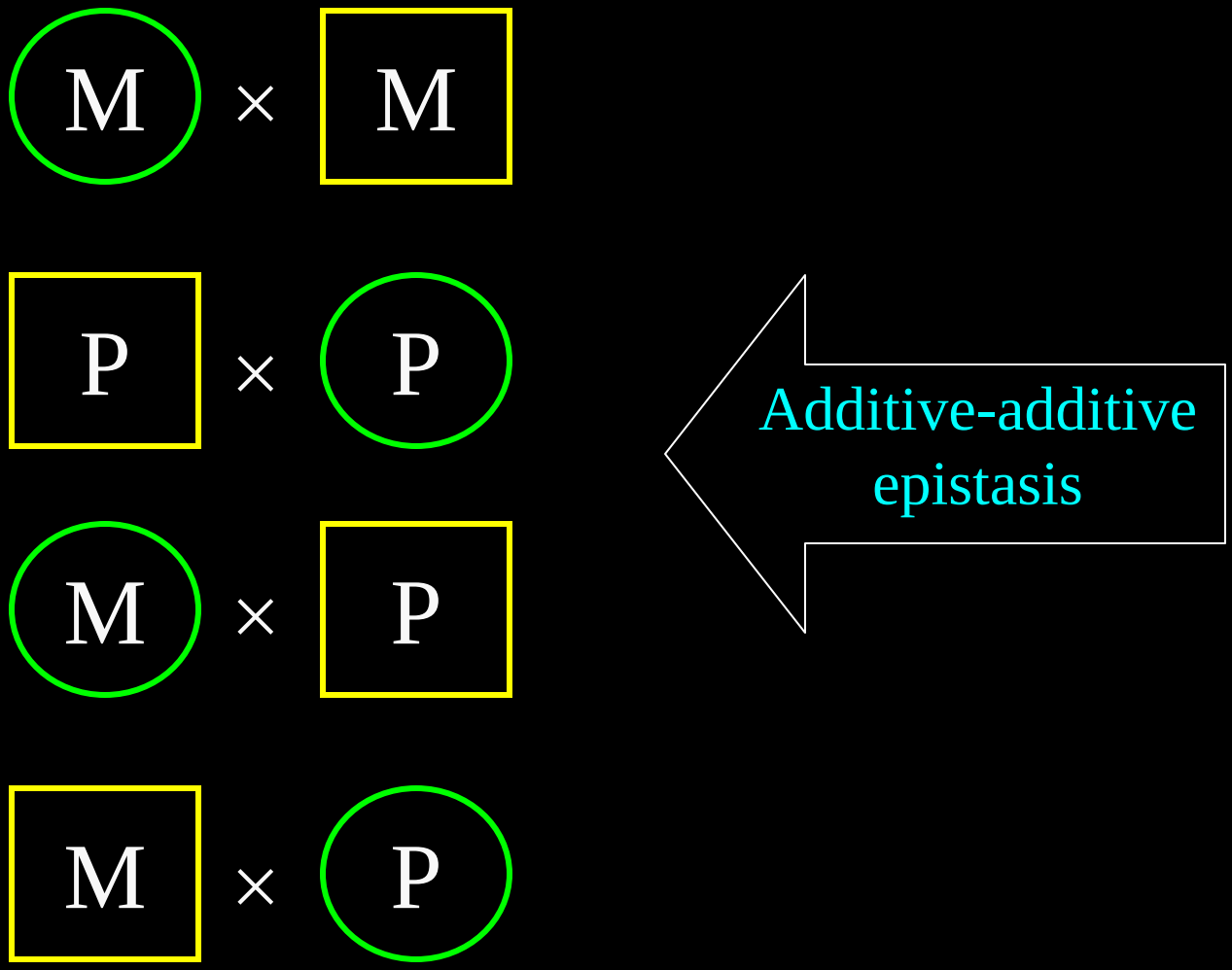
4 main effects



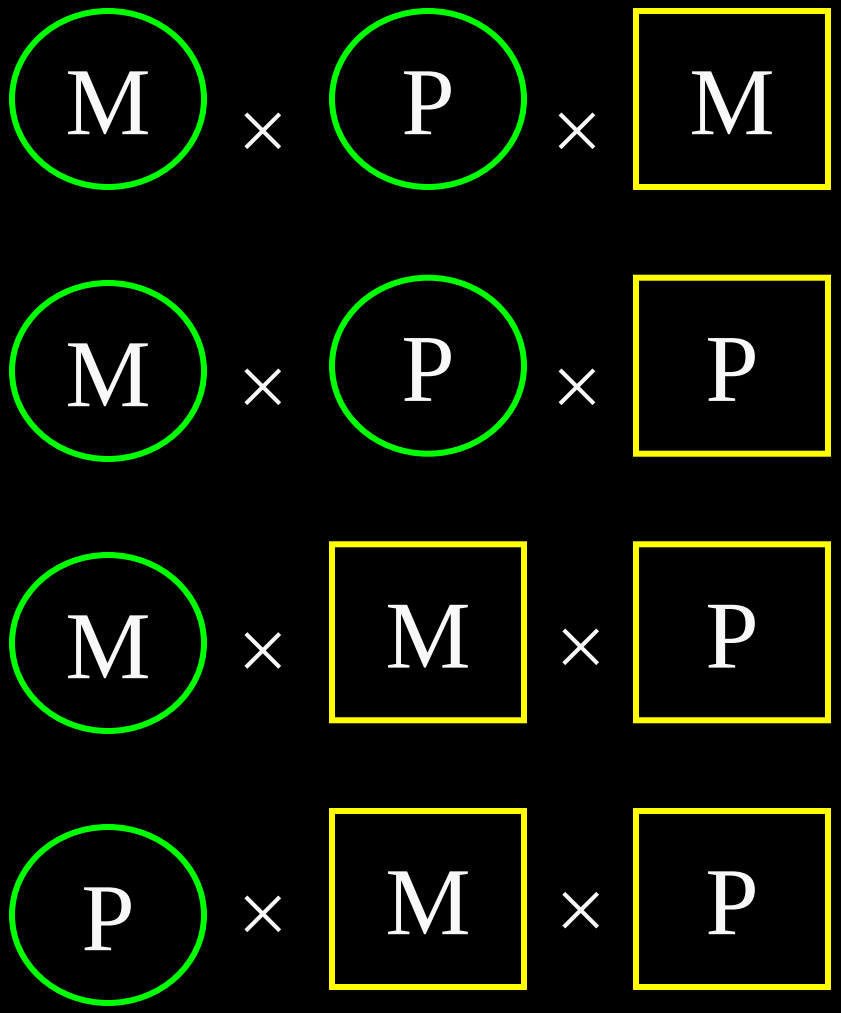
6 twoway interactions



6 twoway interactions

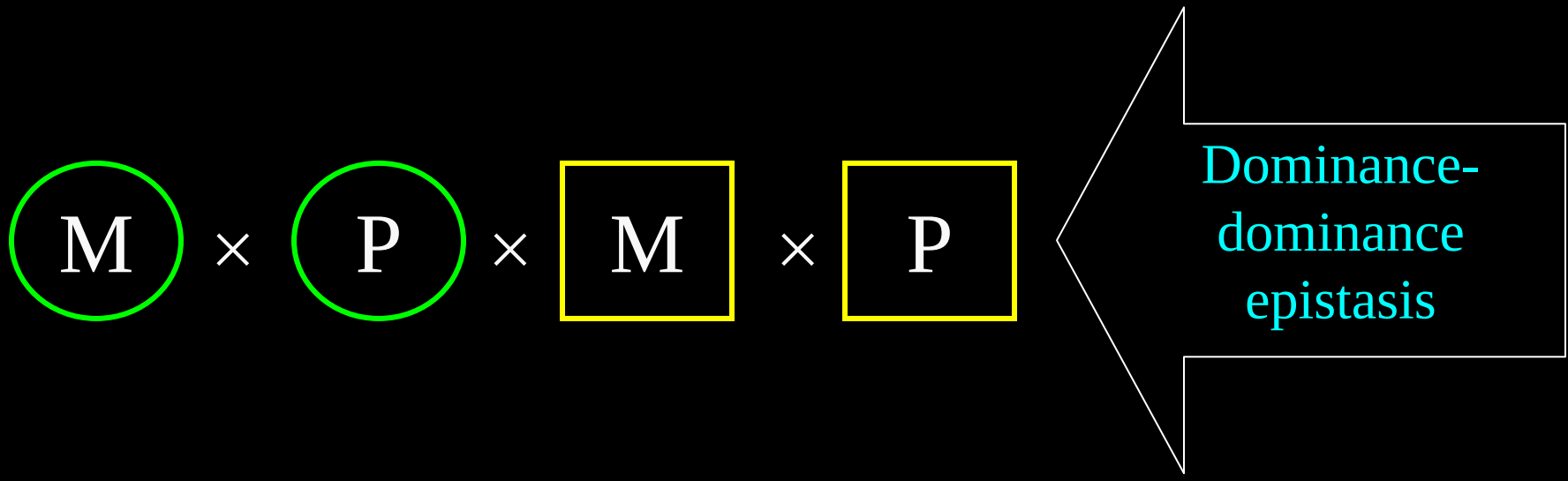


4 threeway interactions



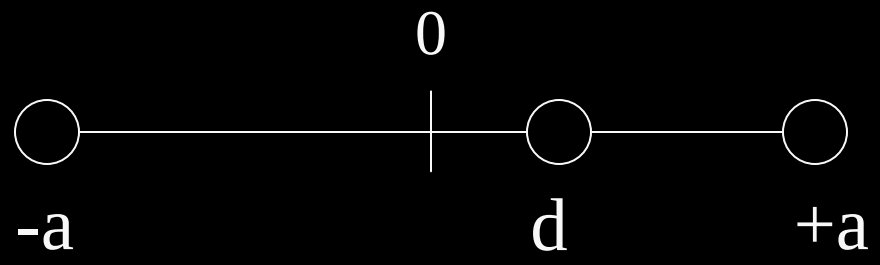
Additive-dominance epistasis

1 fourway interaction



One locus

	Genotypic means
AA	m + a
Aa	m + d
aa	m - a



Two loci

	AA	Aa	aa
BB	$m + a_A + a_B + aa$	$m + d_A + a_B + da$	$m - a_A + a_B - aa$
Bb	$m + a_A + d_B + ad$	$m + d_A + d_B + dd$	$m - a_A + d_B - ad$
bb	$m + a_A - a_B - aa$	$m + d_A - a_B - da$	$m - a_A - a_B + aa$

IBD locus 1	2	Expected Sib Correlation
0	0	σ^2_S
0	1	$\sigma^2_A/2 + \sigma^2_S$
0	2	$\sigma^2_A + \sigma^2_D + \sigma^2_S$
1	0	$\sigma^2_A/2 + \sigma^2_S$
1	1	$\sigma^2_A/2 + \sigma^2_A/2 + \sigma^2_{AA}/4 + \sigma^2_S$
1	2	$\sigma^2_A/2 + \sigma^2_A + \sigma^2_D + \sigma^2_{AA}/2 + \sigma^2_{AD}/2 + \sigma^2_S$
2	0	$\sigma^2_A + \sigma^2_D + \sigma^2_S$
2	1	$\sigma^2_A + \sigma^2_D + \sigma^2_A/2 + \sigma^2_{AA}/2 + \sigma^2_{DA}/2 + \sigma^2_S$
2	2	$\sigma^2_A + \sigma^2_D + \sigma^2_A + \sigma^2_D + \sigma^2_{AA} + \sigma^2_{AD} + \sigma^2_{DA} + \sigma^2_{DD} + \sigma^2_S$

Estimating power for QTL models

Using Mx to calculate power

i. Calculate expected covariance matrices under the full model

ii. Fit model to data with value of interest fixed to null value

	i.True model	ii. Submodel
	Q	0
	S	S
	N	N
-2LL	0.000	=NCP

Model misspecification

Using the `domqtl.mx` script

	i.True	ii. Full	iii. Null
	Q_A	Q_A	0
	Q_D	0	0
	S	S	S
	N	N	N
-2LL	0.000	T_1	T_2

Test:	dominance only	T_1
	additive & dominance	T_2
	additive only	**T_2-T_1**

Results

Using the `domqtl.mx` script

	i.True	ii. Full	iii. Null
Q_A	0.1	0.217	0
Q_D	0.1	0	0
S	0.4	0.367	0.475
N	0.4	0.417	0.525
-2LL	0.000	1.269	12.549

Test:	dominance only (1df)	1.269
	additive & dominance (2df)	12.549
	additive only (1df)	12.549 - 1.269 = 11.28

Expected variances, covariances

	i.True	ii. Full	iii. Null
Var	1.00	1.0005	1.0000
Cov(IBD=0)	0.40	0.3667	0.4750
Cov(IBD=1)	0.45	0.4753	0.4750
Cov(IBD=2)	0.60	0.5839	0.4750

Potential importance of epistasis

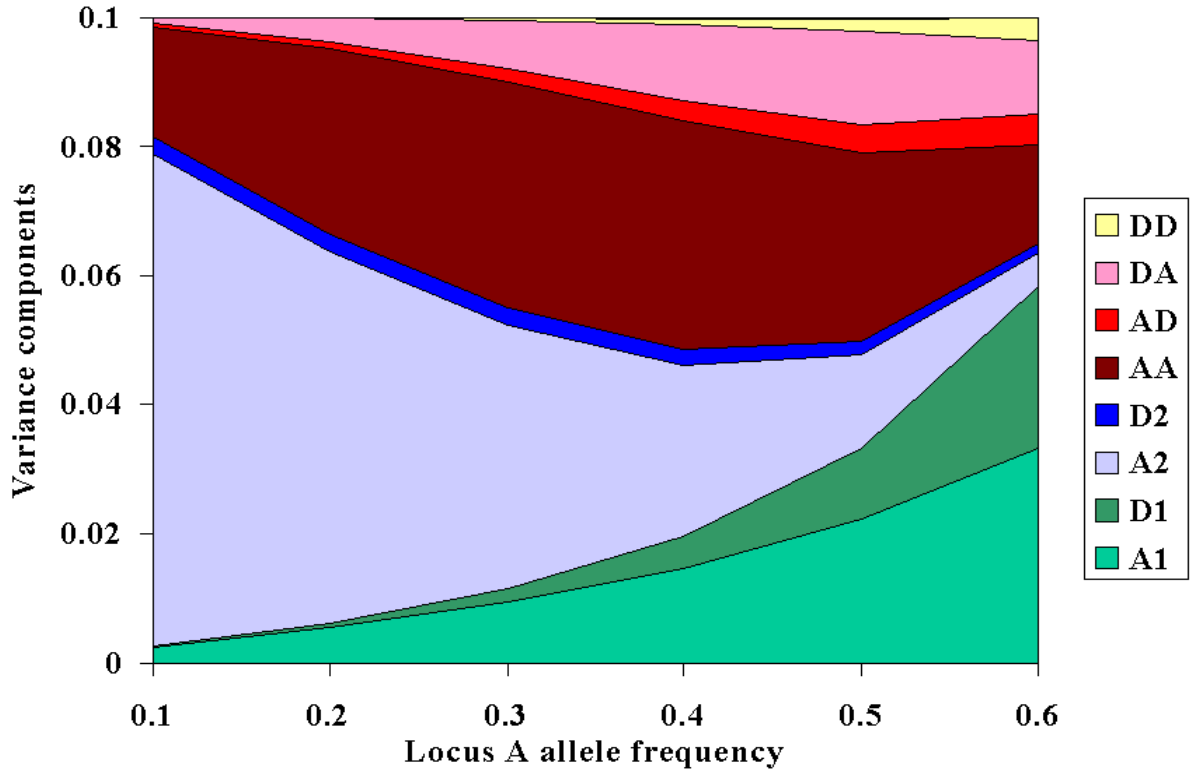
"… a gene's effect might only be detected within a framework that accommodates epistasis…"

			Locus A			
			A_1A_1	A_1A_2	A_2A_2	Marginal
		Freq.	0.25	0.50	0.25	
Locus B	B_1B_1	0.25	**0**	**0**	**1**	0.25
	B_1B_2	0.50	**0**	**0.5**	**0**	0.25
	B_2B_2	0.25	**1**	**0**	**0**	0.25
	Marginal		0.25	0.25	0.25	

Full	V_{A1}	V_{D1}	V_{A2}	V_{D2}	V_{AA}	V_{AD}	V_{DA}	V_{DD}
- DD	V^*_{A1}	V^*_{D1}	V^*_{A2}	V^*_{D2}	V^*_{AA}	V^*_{AD}	V^*_{DA}	-
- AD	V^*_{A1}	V^*_{D1}	V^*_{A2}	V^*_{D2}	V^*_{AA}	-	-	-
- AA	V^*_{A1}	V^*_{D1}	V^*_{A2}	V^*_{D2}	-	-	-	-
- D	V^*_{A1}	-	V^*_{A2}	-	-	-	-	-
- A	V^*_{A1}	-	-	-	-	-	-	-
H_0	-	-	-	-	-	-	-	-

V_S and V_N estimated in all models

True model VC



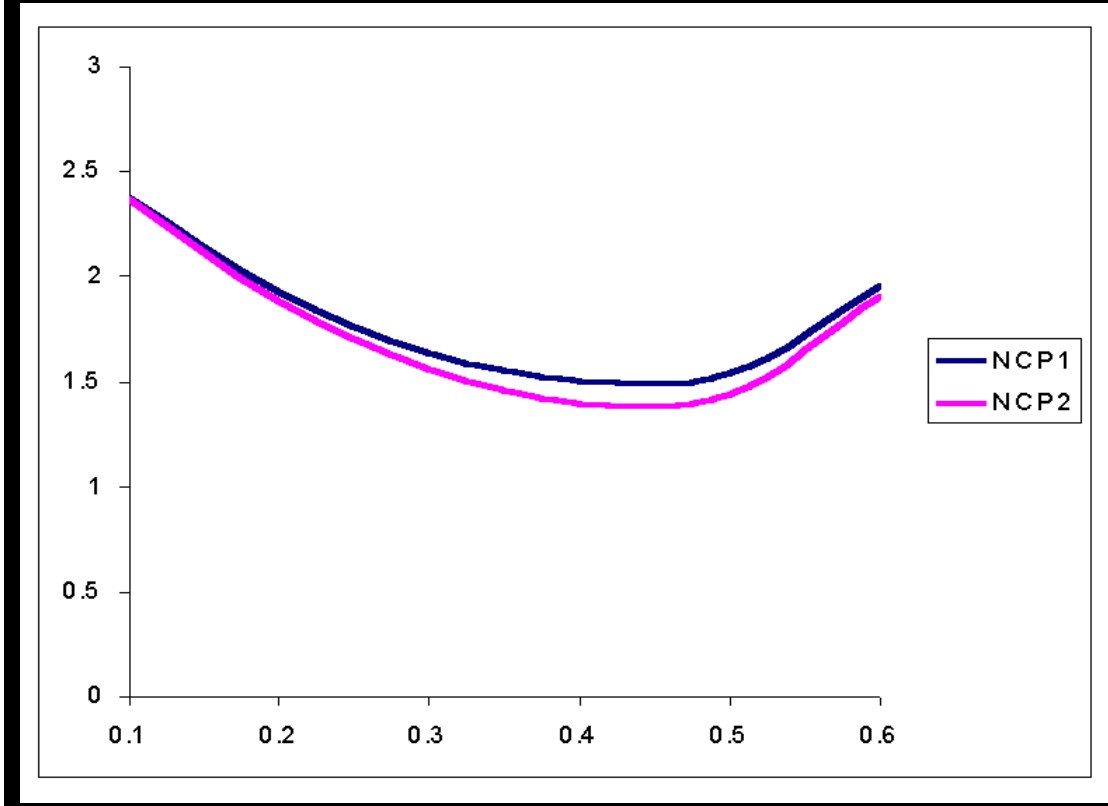
Means matrix

0 0 0

0 0 0

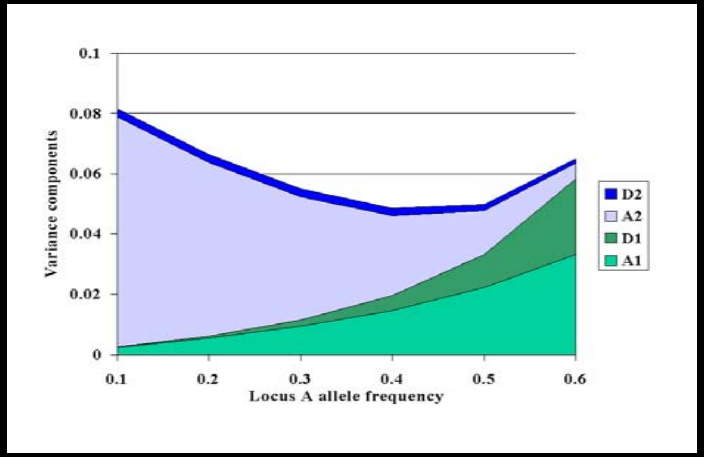
0 1 1

NCP for test of linkage

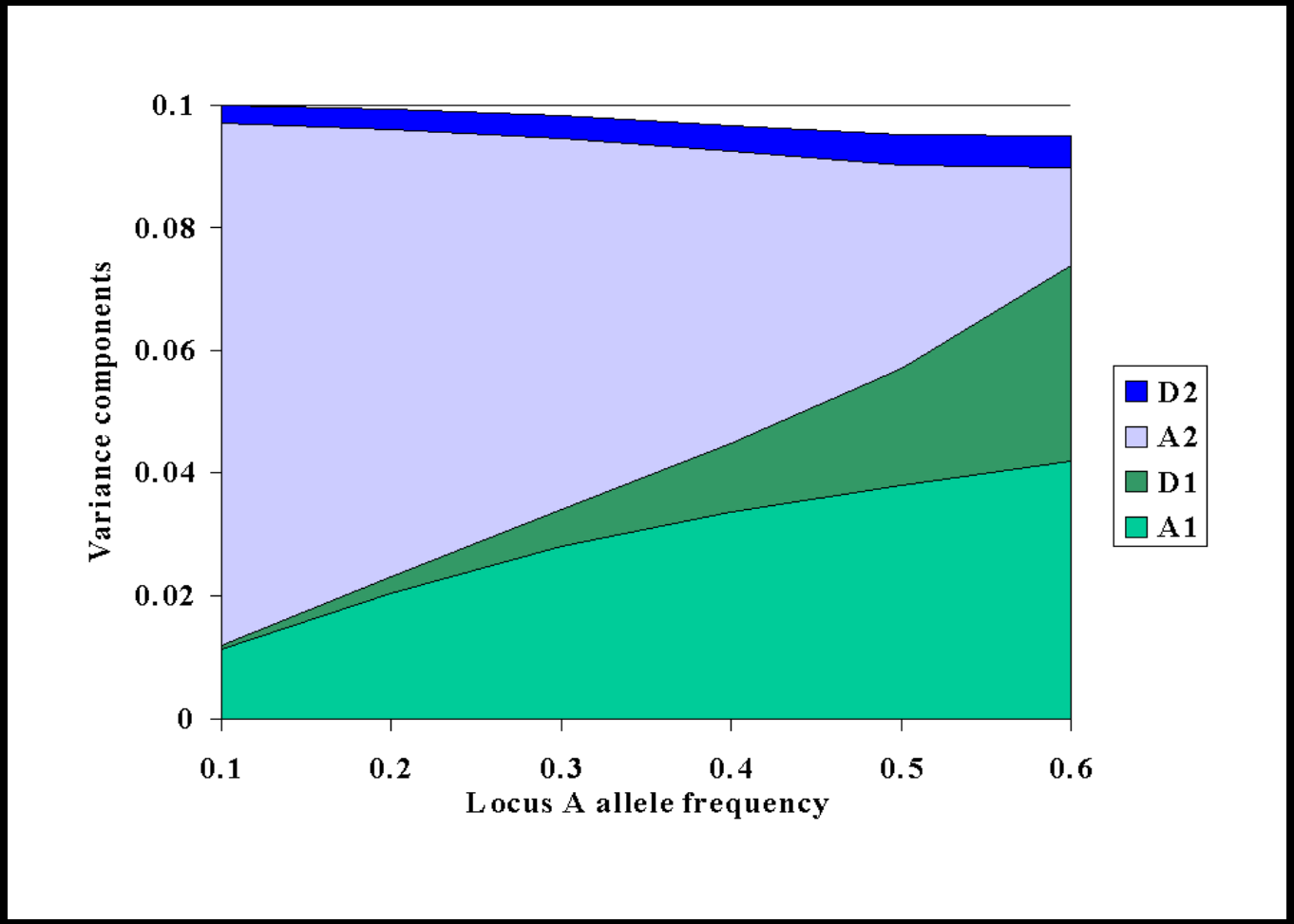


NCP1 Full model

NCP2 Non-epistatic model



Apparent VC under non-epistatic model



Means matrix

0 0 0

0 0 0

0 1 1

Summary

Linkage has low power to detect QTL of small effect

Using selected and/or larger sibships increases power

Single locus additive analysis is usually acceptable

GPC: two-locus linkage

Using the module, for unlinked loci A and B with

Means :

0	0	1
0	0.5	0
1	0	0

Frequencies:

$p_A = p_B = 0.5$

Power of the full model to detect linkage?

Power to detect epistasis?

Power of the single additive locus model?

(1000 pairs, 20% joint QTL effect, VS=VN)