

# Μαριάνθη Γεωργίτση, PhD

Επίκουρη Καθηγήτρια «Γενετική και Μοριακή Βάση Ασθενειών»

**Εργαστήριο Γονιδιωματικής Ποικιλότητας και Γενετικής Επιδημιολογίας**

E-mail: [mgeorgit@mbg.duth.gr](mailto:mgeorgit@mbg.duth.gr)

---

Εργαστήριο: 1<sup>ος</sup> όροφος – Τμήμα ΜΒΓ

Γραφείο: 1<sup>ος</sup> όροφος – Τμήμα Ιατρικής



**Ανισορροπία  
Σύνδεσης**

**και**

**Απλότυποι**

# Τι θα μάθουμε σήμερα?

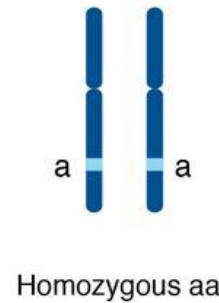
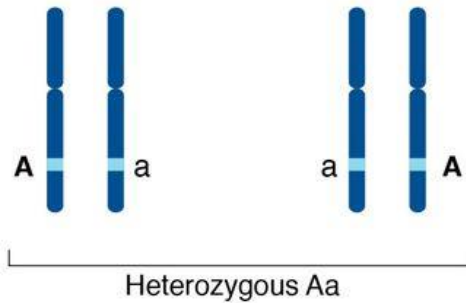
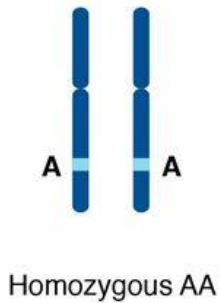
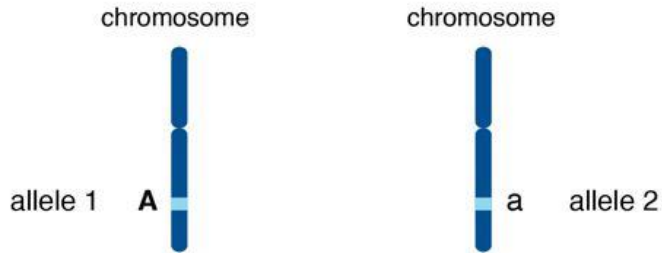
- Τι είναι η ισορροπία σύνδεσης (linkage equilibrium)
- Τι είναι η ανισορροπία σύνδεσης (linkage disequilibrium - LD)
- Τι είναι ο απλότυπος - παραδείγματα
- Ποια είναι τα μέτρα εκτίμησης της ανισορροπίας σύνδεσης
- Πώς αναπαρίσταται η LD
- Από ποιους παράγοντες επηρεάζεται
- Πώς μεταβάλλεται η LD στο χρόνο
- Πού και πώς αξιοποιείται αυτή η γνώση

# Μία γενετική θέση ή Ένας γενετικός τόπος (locus)

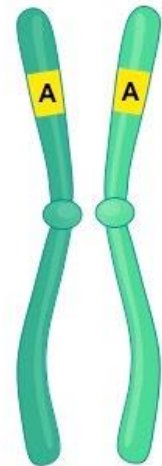
---

# Αλληλόμορφα

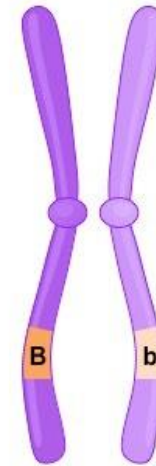
Εναλλακτικές μορφές ενός γονιδίου ή μίας αλληλουχίας (τμήματος) DNA σε ένα συγκεκριμένο γενετικό τόπο



1<sup>ος</sup> Νόμος του Mendel –  
Ο διαχωρισμός των αλληλομόρφων  
στους γαμέτες είναι τυχαίος και  
ανεξάρτητος από άλλα ζεύγη  
αλληλομόρφων



**Homozygous**  
Alleles are same



**Heterozygous**  
Alleles are different



**Hemizygous**  
Only one allele  
(e.g. XY)

Επανάληψη...

$$p^2+2pq+q^2=(p+q)^2=1$$

## Θεώρημα Hardy-Weinberg

Θηλυκοί γαμέτες (συχνότητα)

Αρσενικοί γαμέτες (συχνότητα)

|              | <b>A (p)</b>             | <b>a (q)</b>             |
|--------------|--------------------------|--------------------------|
| <b>A (p)</b> | AA ( $p \cdot p = p^2$ ) | Aa ( $p \cdot q$ )       |
| <b>a (q)</b> | aA ( $q \cdot p$ )       | aa ( $q \cdot q = q^2$ ) |

$$P(AA) = p^2$$

$$P(Aa) = p \cdot q \text{ (μητέρα)} + p \cdot q \text{ (πατέρα)} = 2pq$$

$$P(aa) = q^2$$

# Προϋποθέσεις για την ισχύ του θεωρήματος H-W και . . .

## αίτια απόκλισης από αυτό

- Διπλοειδής οργανισμός
- Σεξουαλική αναπαραγωγή
- Τυχαία σύζευξη των ατόμων (παμμίξια) (**vs** διαστρωμάτωση, ομοιοτυπική σύζευξη, συγγένεια εξ αίματος & ενδογαμία)
- Μη επικαλυπτόμενες γενιές
- Μέγεθος πληθυσμού πολύ μεγάλο (~άπειρο) [**vs** μικρό (γενετική παρέκκλιση)]
- Ίδιες συχνότητες αλληλομόρφων σε αρσενικά και θηλυκά άτομα (**vs** mtDNA, Y-chr, non-PAR1/2 στα αρσενικά)
- Δεν υπάρχει επιλογή γονοτύπου (**vs** θνησιγόνα αλληλόμορφα, πλεονέκτημα ετεροζυγώτη)
- Δεν υπάρχει εισροή αλληλομόρφων στον πληθυσμό και άρα αλλαγή συχνοτήτων των αλληλομόρφων (**vs** μετανάστευση ή μεταλλαγή)

### • • • • Ο Νόμος Hardy-Weinberg

Ο νόμος Hardy-Weinberg ισχύει υπό τις εξής προϋποθέσεις:

- Ο πληθυσμός είναι μεγάλος και οι συζεύξεις είναι τυχαίες, όσον αφορά στον υπό μελέτη γενετικό τόπο.
- Οι συχνότητες των αλληλομόρφων παραμένουν σταθερές με την πάροδο του χρόνου γιατί:

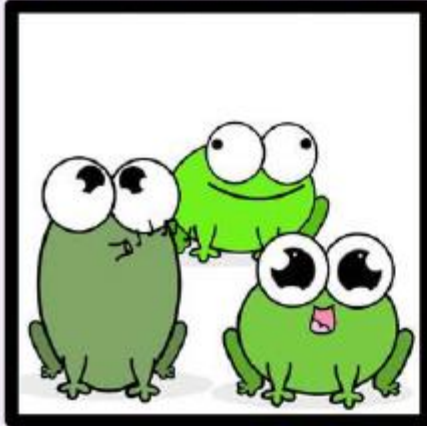
Ο ρυθμός μεταπλαξίγνεσης δεν είναι σημαντικός.

Τα άτομα όλων των γονότυπων είναι εξίσου ικανά να αναπαραχθούν και να μεταβιβάσουν τα γονίδιά τους στους απογόνους τους, δηλαδή δεν υπάρχει επιλογή εις βάρος κάποιου συγκεκριμένου γονοτύπου.

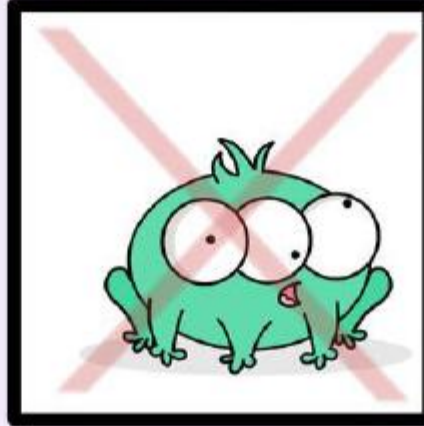
Δεν υπήρχε σημαντική είσοδος ατόμων από έναν άλλο πληθυσμό με συχνότητες αλληλομόρφων πολύ διαφορετικές από αυτές του ενδογενούς πληθυσμού.

# Assumptions of Hardy-Weinberg Equilibrium

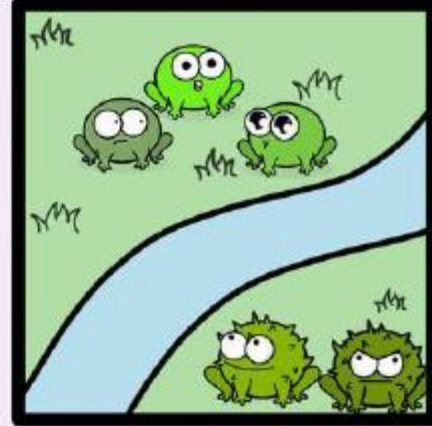
1. No selection



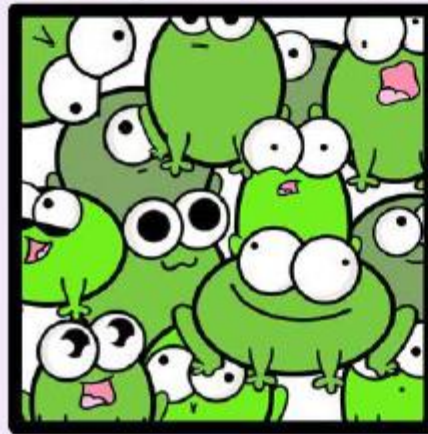
2. No Mutation



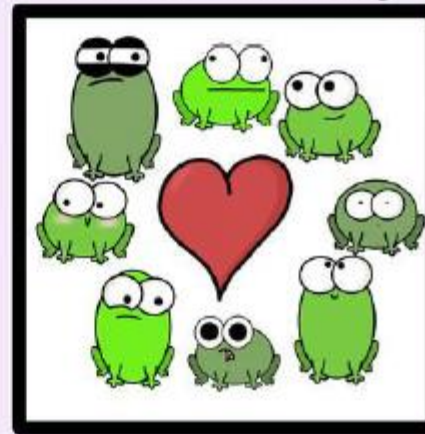
3. No Migration



4. Large Population



5. Random Mating



@AmoebaSisters



## Hardy-Weinberg Equilibrium for a Three-Allele System

$$p + q + r = 1$$

*p, q, and r are the three alleles in the population (system).*

$$(p + q + r)^2 = 1$$

*"Cross" the individual alleles together by squaring the trinomial.*



$$p^2 + q^2 + r^2 + 2pq + 2pr + 2qr = 1$$

@wormthoughts

$(p_1 + p_2 + p_3 + \dots + p_n)^2 = 1$ , εάν σε έναν γενετικό τόπο έχουν περιγραφεί  $n$  διαφορετικά αλληλόμορφα

## Hardy–Weinberg Principle

- X-linked genes are a special case because males have only one X-chromosome
- Genotype frequencies among females:  $HH = p^2$ ;  $Hh = 2pq$ ;  $hh = q^2$
- Genotype frequencies among males are the same as allele frequencies:  
 $H = p$ ,  
 $h = q$

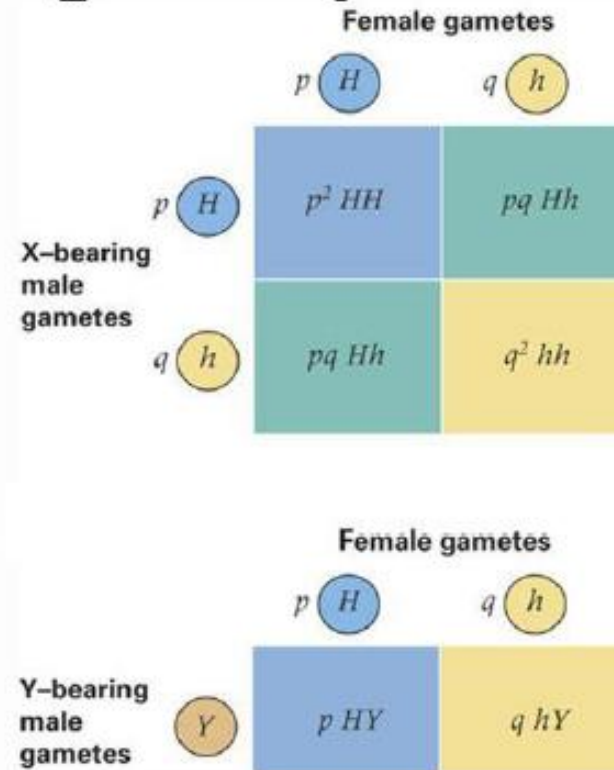


Fig. 14.15

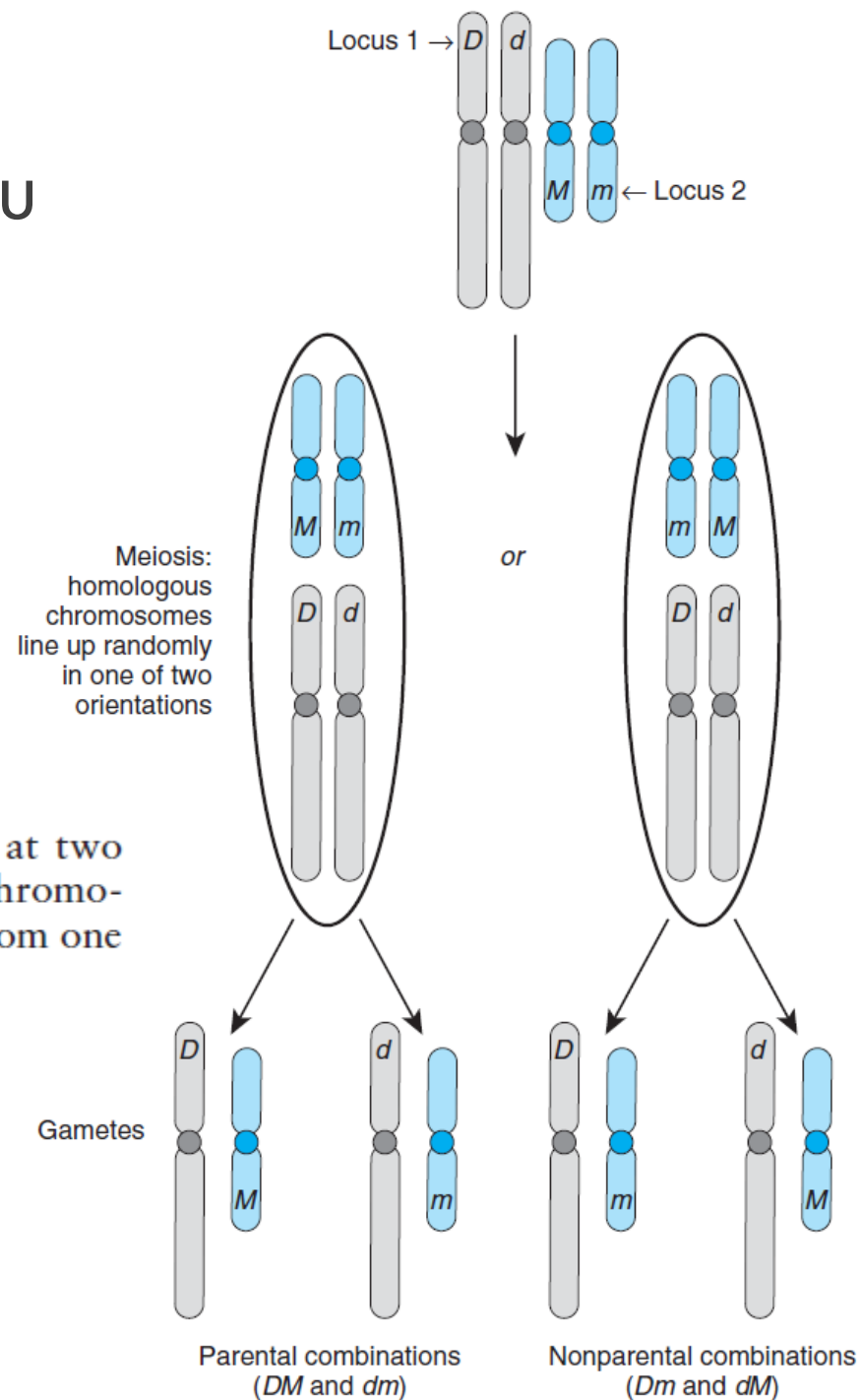
---

# Δύο γενετικές θέσεις ή Δύο γενετικοί τόποι (loci)

# 2<sup>ος</sup> Νόμος του ανεξάρτητου συνδυασμού του Mendel

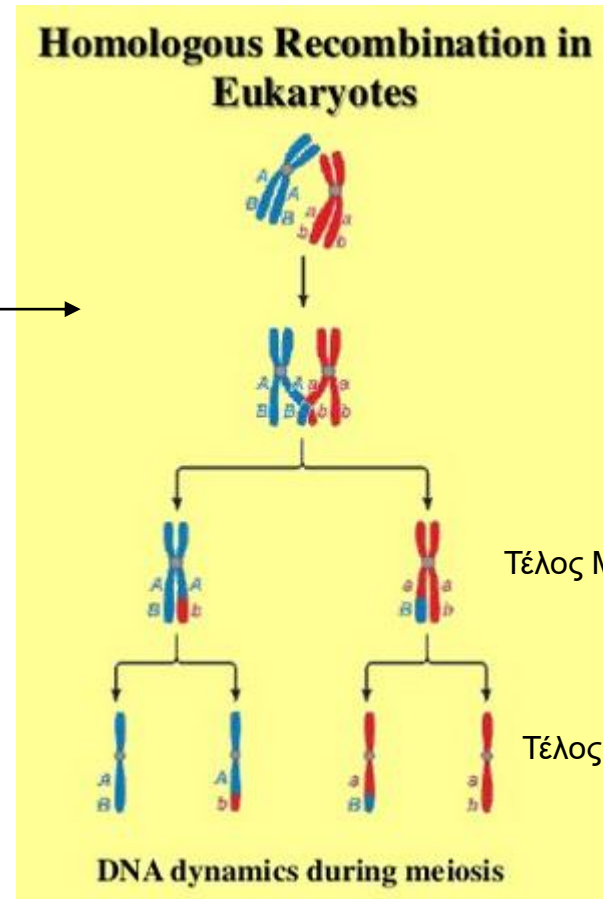
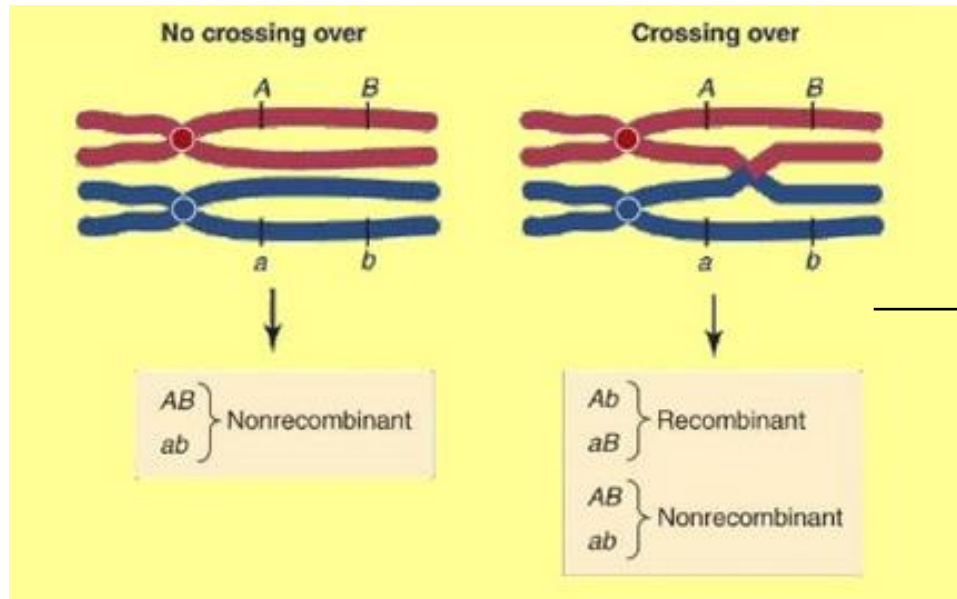
Τα αλληλόμορφα  $D/d$  και  $M/m$  χαρτογραφούνται σε **διαφορετικά** χρωμοσώματα

**Figure 10-2** ■ Independent assortment of alleles at two loci, 1 and 2, when they are located on different chromosomes. Assume that alleles  $D$  and  $M$  were inherited from one parent,  $d$  and  $m$  from the other.



Τα αλληλόμορφα A/a και B/b χαρτογραφούνται στο **ίδιο** χρωμόσωμα

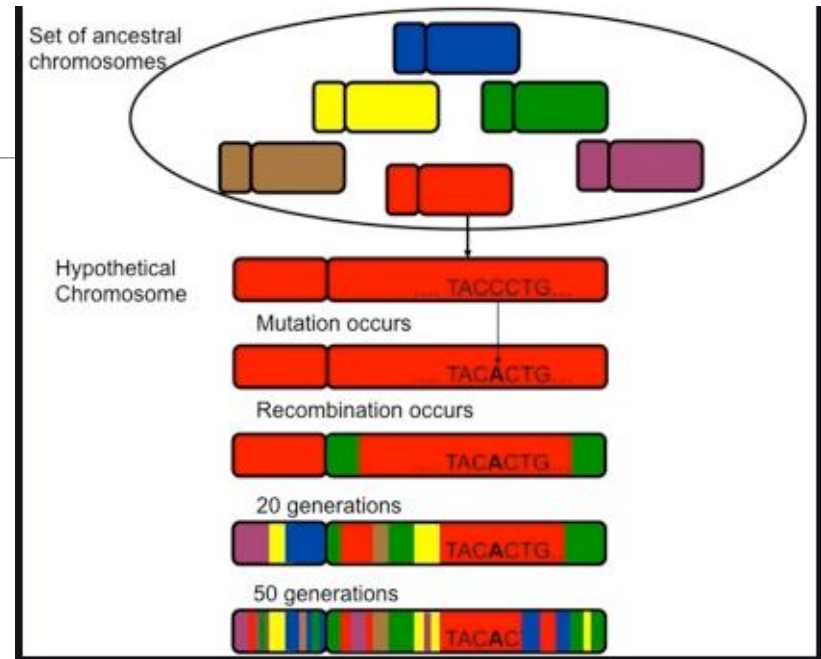
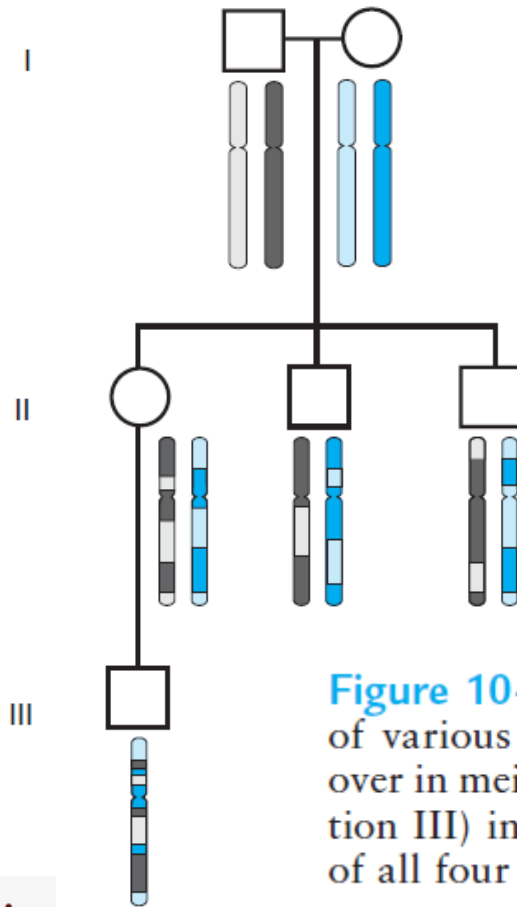
## Ομόλογος ανασυνδυασμός



Οι γενετικές θέσεις των αλληλόμορφων A/a και B/b που χαρτογραφούνται στο **ίδιο** χρωμόσωμα μπορεί να απέχουν πολύ ή λίγο μεταξύ τους

# Ομόλογος ανασυνδυασμός

Σε επίπεδο οικογένειας



Σε επίπεδο πληθυσμού

**Figure 10-1** ■ The effect of recombination on the origin of various portions of a chromosome. Because of crossing over in meiosis, the copy of the chromosome the boy (generation III) inherited from his mother is a mosaic of segments of all four of his grandparents' copies of that chromosome.

# Ανισορροπία σύνδεσης *Linkage disequilibrium (LD)*

Η μη τυχαία συσχέτιση αλληλομόρφων σε διαφορετικούς γενετικούς τόπους επάνω στο **ίδιο χρωμόσωμα!!!**

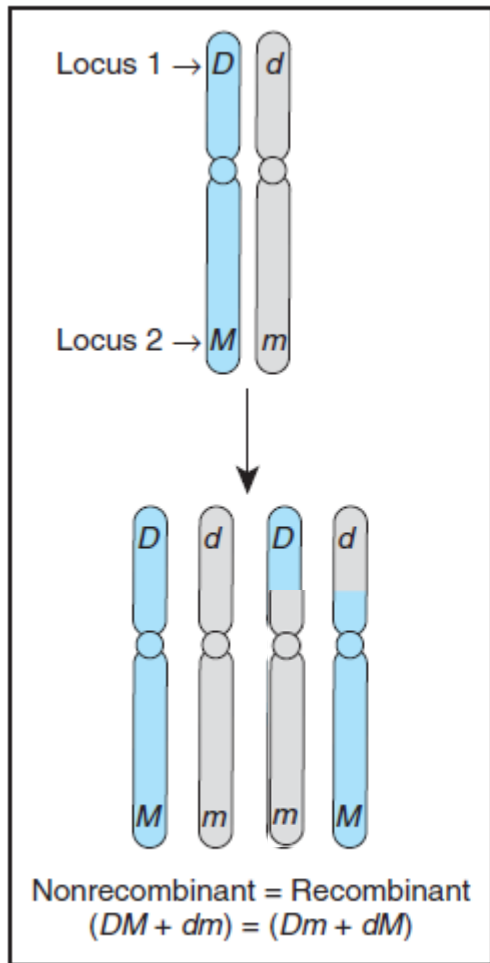
**Απόκλιση** του νόμου του ανεξάρτητου συνδυασμού των αλληλομόρφων του Mendel

**ΓΙΑΤΙ ?**

$\theta = 0.5$

(ανεξάρτητος διαχωρισμός)

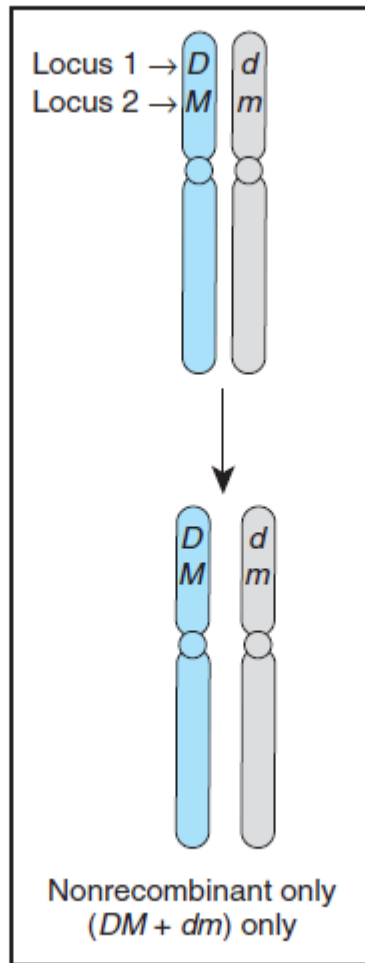
(ισορροπία σύνδεσης)



A

$\theta \sim 0$

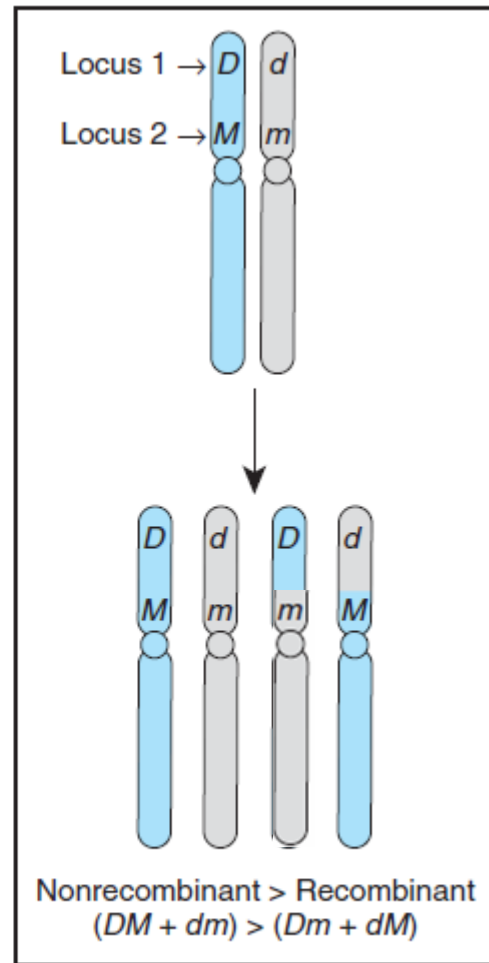
(τέλεια ανισορροπία σύνδεσης)



B

$0.01 \leq \theta < 0.5$

(μερική/ατελής ανισορροπία σύνδεσης)



C

Τα αλληλόμορφα  $D/d$  και  $M/m$  χαρτογραφούνται στο **ίδιο** χρωμόσωμα

Figure 10-4 ■ Assortment of alleles at two loci, 1 and 2, when they are located on the same chromosome. A, The loci are far apart and at least one crossover between them is likely to occur in every meiosis. B, The loci are so close together that crossing over between them is very unlikely. C, The loci are close together on the same chromosome but far enough apart that crossing over occurs in the interval between the two loci only in some meioses and not in others.

Όσο πιο μικρό είναι το ποσοστό ανασυνδυασμού ( $\theta$ ), τόσο πιο κοντά βρίσκονται οι δύο γενετικοί τόποι (εδώ τα  $D/d, M/m$ )  $\rightarrow$  όπου  $0 < \theta < 0.5$

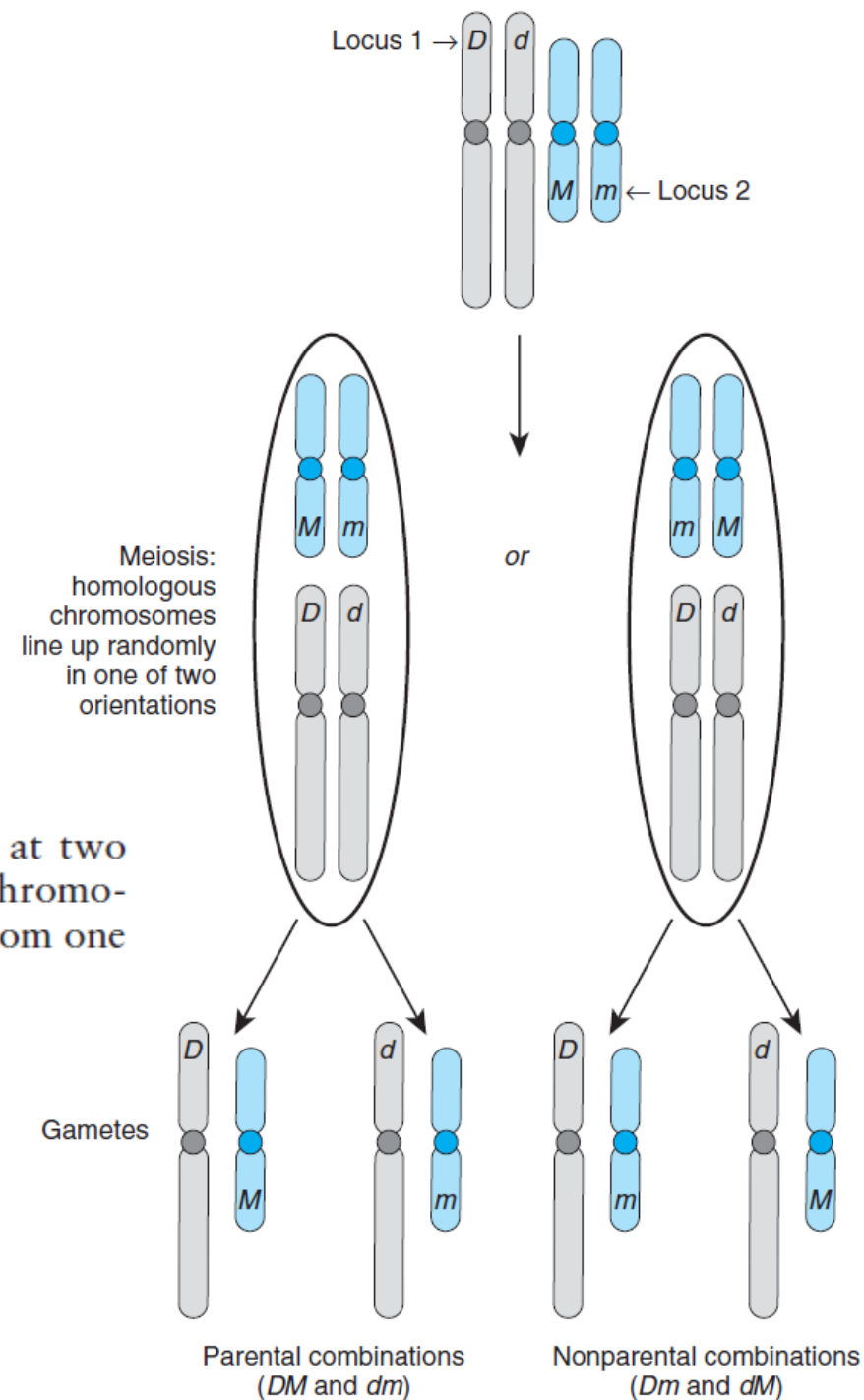


## Υπενθύμιση

# Νόμος του ανεξάρτητου συνδυασμού του Mendel

Τα αλληλόμορφα  $D/d$  και  $M/m$  χαρτογραφούνται σε **διαφορετικά** χρωμοσώματα

**Figure 10-2** ■ Independent assortment of alleles at two loci, 1 and 2, when they are located on different chromosomes. Assume that alleles  $D$  and  $M$  were inherited from one parent,  $d$  and  $m$  from the other.



# Ανασυνδυασμός

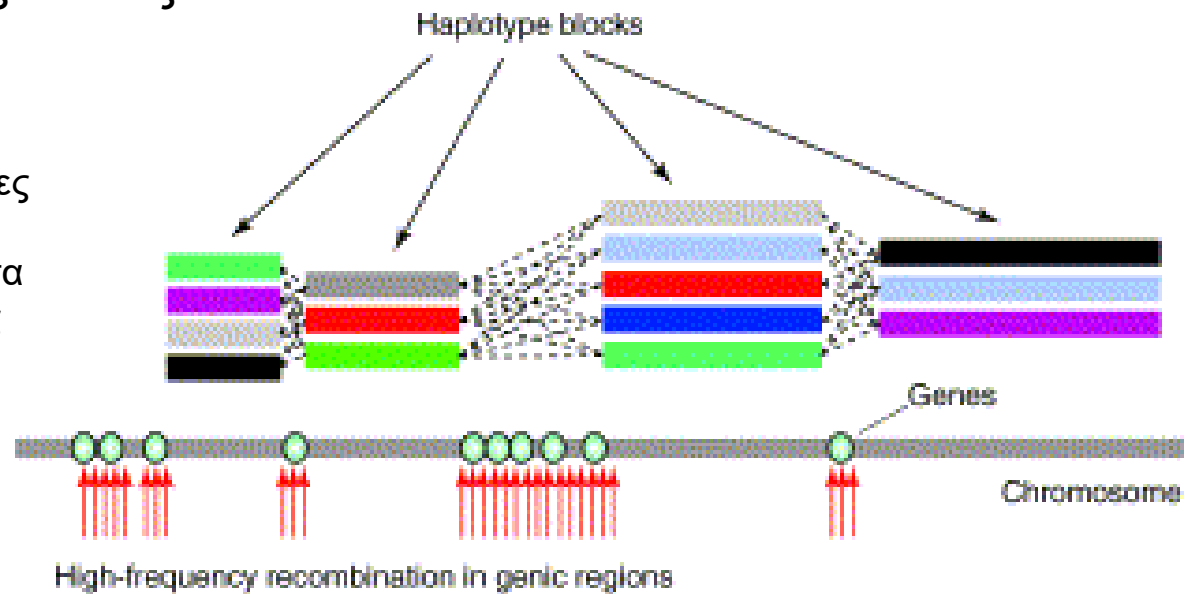
- $\theta$  (ή  $c$  ή  $r$ ) = κλάσμα ανασυνδυασμού (recombination fraction)
- $\theta$  = το ποσοστό ανασυνδυασμένων γαμετών που παράγονται από έναν διπλό ετεροζυγώτη
- Τιμές  $\theta$  από 0 ως 0.5
  - $\theta=0$  πλήρης σύνδεση
  - $\theta=0.5$  ανεξάρτητος διαχωρισμός
  - $\theta=0.01$  δηλ. 1/100 γαμέτες είναι ανασυνδυασμένοι (απόσταση 1cM ανάμεσα στους δύο γενετικούς τόπους)
- Τόποι που βρίσκονται σε απόσταση  $>50\text{cM}$  διαχωρίζονται ανεξάρτητα

# Επίδραση του ανασυνδυασμού στη μη τυχαία συσχέτιση των αλληλομόρφων

Ο ανασυνδυασμός ανάμεσα σε δύο γενετικούς τόπους τείνει να τυχαιοποιεί τους συνδυασμούς των αλληλομόρφων από γενιά σε γενιά.

Ο ρυθμός αυτής της διαδικασίας εξαρτάται από τον ρυθμό ανασυνδυασμού ανάμεσα σε αυτούς τους τόπους.

Δεν γνωρίζουμε ακριβώς τη βιολογική βάση γιατί ορισμένες περιοχές του γονιδιώματος εμφανίζουν μεγάλη συχνότητα ανασυνδυασμού ενώ άλλες εξαιρετικά χαμηλή.



TRENDS in Genetics

# ΑΠΛΌΤΥΠΟΙ

Γονότυπος: AaBb

Δυνατοί απλότυποι:

Απλότυπος 1

A |  
B |

Απλότυπος 2

| a  
| b

A |  
b |

Απλότυπος 3

| a  
| B

Απλότυπος 4

# Γονότυποι και Απλότυποι

## Ένας γενετικός τόπος:

|                        |    |
|------------------------|----|
| Homozygous wild type   | AA |
| Homozygous alternative | aa |
| Heterozygous           | Aa |

## Δύο γενετικοί τόποι:

| <u>Locus 1</u> | <u>Locus 2</u> | <u>Απλότυποι:</u> |
|----------------|----------------|-------------------|
| AA             | BB             | AB                |
| aa             | bb             | Ab                |
| Aa             | Bb             | aB                |
|                |                | ab                |

## Πολλαπλοί γενετικοί τόποι:

| <u>Locus 1</u> | <u>Locus 2</u> | <u>Locus 3</u> | <u>Απλότυποι:</u> |
|----------------|----------------|----------------|-------------------|
| AA             | BB             | CC             | ABC               |
| aa             | bb             | cc             | ABc               |
| Aa             | Bb             | Cc             | AbC               |
|                |                |                | Abc               |
|                |                |                | aBC               |
|                |                |                | aBc               |
|                |                |                | abC               |
|                |                |                | abc               |

# Απλότυπος (haplotype)

Ένα σύνολο αλληλομόρφων γονιδίων/γενετικών τόπων σε έναν οργανισμό που κληρονομούνται μαζί από ένα γονέα

- Π.χ. ισχύει για συνδεδεμένα γονίδια/γενετικούς τόπους που βρίσκονται κοντά, πάνω στο ίδιο χρωμόσωμα
- Π.χ. οι πολυμορφισμοί μιτοχονδριακού και χλωροπλαστικού DNA, καθώς και του DNA του Y χρωμοσώματος (μη ψευδοαυτοσωμικές περιοχές PAR1 & PAR2) που κληρονομούνται από ένα γονέα χωρίς ανασυνδυασμό

Παραδείγματα άλλων χρήσεων του όρου (βλέπε επόμενη διαφάνεια):

- Ένα σύνολο από παραλλαγές (=πολυμορφισμοί) (π.χ. SNPs) σε ένα χρωμόσωμα που τείνουν να εμφανίζονται μαζί (είναι στατιστικά σχετιζόμενες)
  - π.χ. έρευνα ασθενειών στον άνθρωπο μέσα από το πρόγραμμα HarMap
- Ένα σύνολο συγκεκριμένων παραλλαγών μέσα σε ένα τμήμα DNA σε ένα άτομο

# Απλότυποι πολυμορφικών θέσεων

- Συνδυασμοί αλληλομόρφων διαφορετικών παραλλαγών (=πολυμορφισμών) σε ένα χρωμόσωμα
- Οι παραλλαγές μπορεί να είναι τύπου RFLPs, μικροδορυφόροι, SNPs, κτλ
- 3 διαλληλικοί (biallelic) πολυμορφισμοί  $2^3=8$  δυνατοί απλότυποι
- 7 διαλληλικοί πολυμορφισμοί  $2^7=128$  δυνατοί απλότυποι
- Αν υπάρχουν λιγότεροι απλότυποι από ότι αναμένεται σύμφωνα με το νόμο του ανεξάρτητου συνδυασμού, τότε οι παραλλαγές που μελετάμε είναι σε **ανισορροπία σύνδεσης**
- Linkage Disequilibrium ή **LD**

# Παράδειγμα με δείκτες RFLPs

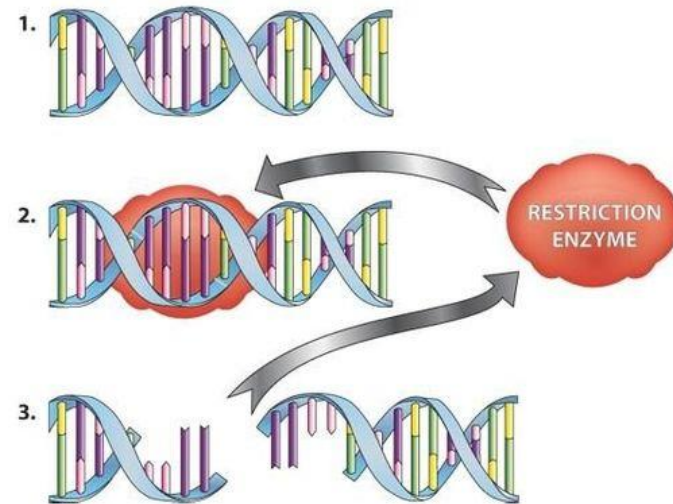
Για 3 δείκτες RFLPs υπάρχουν  $2^3 = 8$   
δυνατοί απλότυποι - **ΘΕΩΡΗΤΙΚΑ**

Πόσοι απλότυποι θα παρατηρηθούν  
σε ένα δείγμα; - **ΠΡΑΚΤΙΚΑ**

Εξαρτάται από:

- την ανισορροπία σύνδεσης ανάμεσα στις θέσεις (RFLPs, SNPs, άλλα loci)
- τον πληθυσμό υπό μελέτη
- το μέγεθος του πληθυσμιακού δείγματος
- τις συχνότητες των αλληλομόρφων στον δεδομένο πληθυσμό

1. +++
2. ++-
3. +-+
4. -++
5. +- -
6. -+ -
7. --+
8. ---





**Ας μελετήσουμε δύο γενετικούς τόπους A και B με αλληλόμορφα**

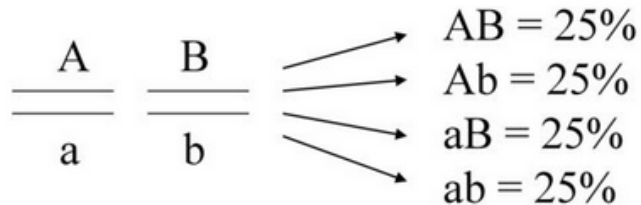
### **Αλληλόμορφα Συχνότητα**

|     |       |       |                 |
|-----|-------|-------|-----------------|
| A = | $A_1$ | $p_1$ | $p_1 + p_2 = 1$ |
| a = | $A_2$ | $p_2$ |                 |
| B = | $B_1$ | $q_1$ | $q_1 + q_2 = 1$ |
| b = | $B_2$ | $q_2$ |                 |

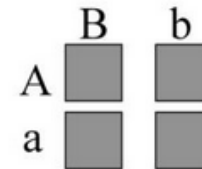
## Υπάρχουν 4 πιθανοί γαμέτες

|      | Γαμέτες   | Συχνότητα απλοτύπου |                             |
|------|-----------|---------------------|-----------------------------|
| AB = | $A_1 B_1$ | $x_1$               | $x_1 + x_2 + x_3 + x_4 = 1$ |
| Ab = | $A_1 B_2$ | $x_2$               |                             |
| aB = | $A_2 B_1$ | $x_3$               |                             |
| ab = | $A_2 B_2$ | $x_4$               |                             |

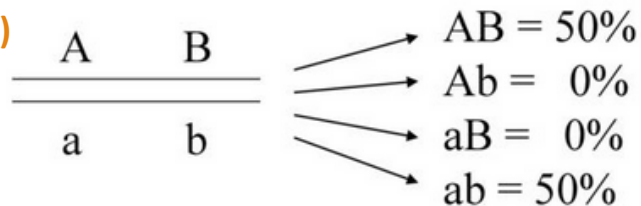
Ισορροπία Σύνδεσης  
(linkage equilibrium)



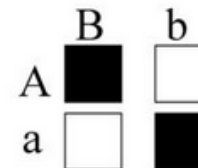
$$f(Ab) = f(A) \times f(b)$$



Ανισορροπία Σύνδεσης  
(linkage disequilibrium - LD)  
(πλήρης)



$$f(Ab) \neq f(A) \times f(b)$$



# Linkage Equilibrium

Τυχαία συσχέτιση αλληλομόρφων στους γαμέτες

Τόπος Β

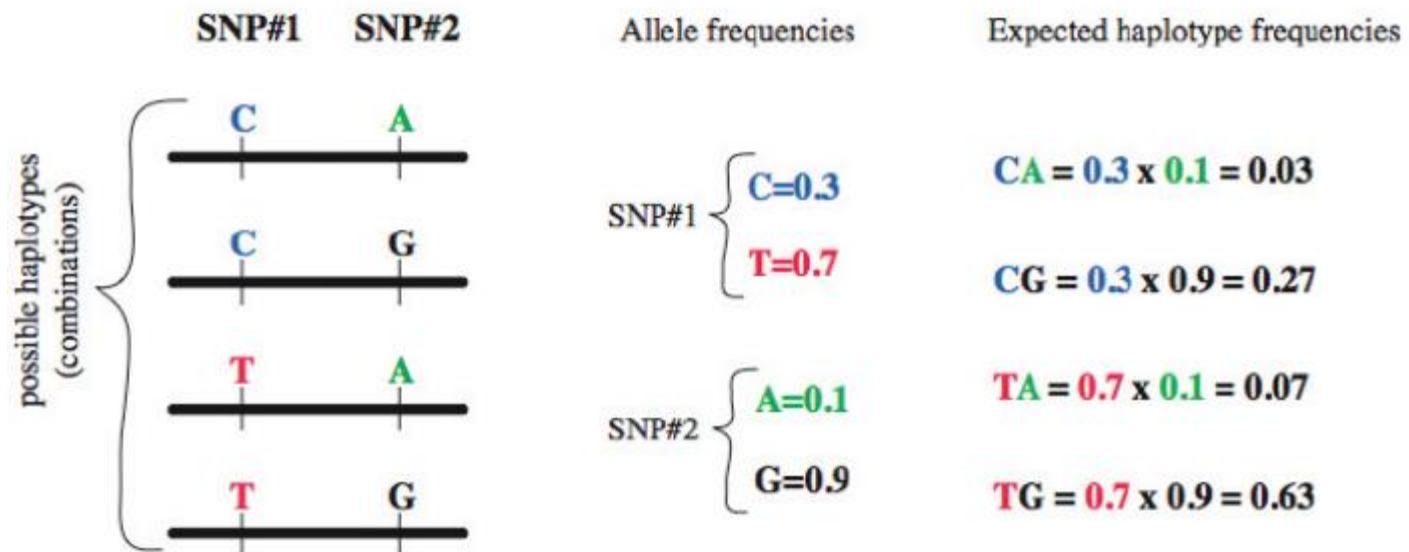
$B_1$        $B_2$   
 $q_1$        $q_2$

|                |                             |                             |
|----------------|-----------------------------|-----------------------------|
| $A_1$<br>$p_1$ | $A_1B_1$<br>$x_1 = p_1 q_1$ | $A_1B_2$<br>$x_2 = p_1 q_2$ |
| $A_2$<br>$p_2$ | $A_2B_1$<br>$x_3 = p_2 q_1$ | $A_2B_2$<br>$x_4 = p_2 q_2$ |

Τόπος Α

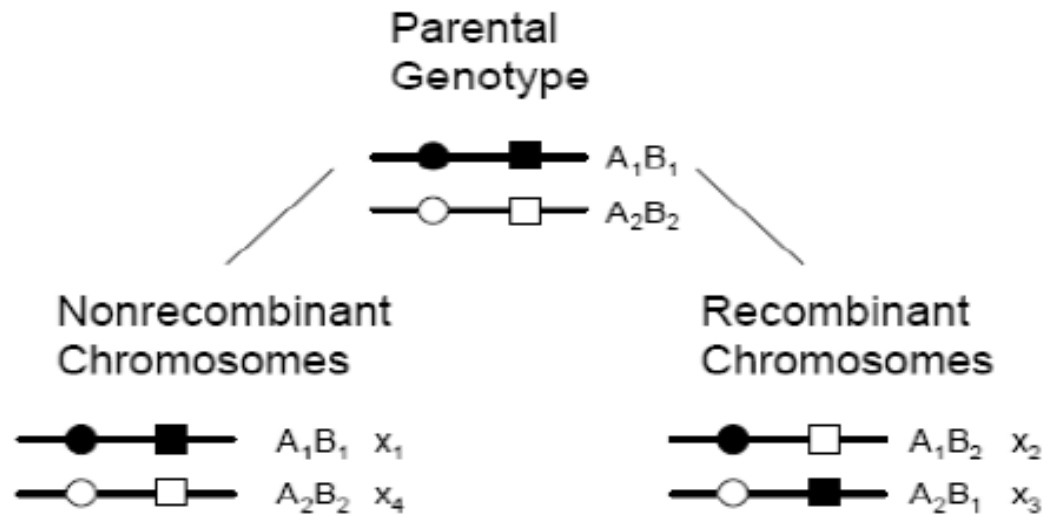
$$x_1 + x_2 + x_3 + x_4 = 1$$

# Ισορροπία Σύνδεσης



Στο δείγμα μας παρατηρούμε όλους τους δυνατούς απλότυπους, με ποικίλες συχνότητες

# Linkage Equilibrium



| Gamete   | Frequency |            | Allele | Frequency         |
|----------|-----------|------------|--------|-------------------|
| $A_1B_1$ | $x_1$     | $= p_1q_1$ | $A_1$  | $p_1 = x_1 + x_2$ |
| $A_1B_2$ | $x_2$     | $= p_1q_2$ | $A_2$  | $p_2 = x_3 + x_4$ |
| $A_2B_1$ | $x_3$     | $= p_2q_1$ | $B_1$  | $q_1 = x_1 + x_3$ |
| $A_2B_2$ | $x_4$     | $= p_2q_2$ | $B_2$  | $q_2 = x_2 + x_4$ |

# Πλήρης Ανισορροπία Σύνδεσης

Παράδειγμα για 2 SNPs:

SNP1: A 50% C 50%

SNP2: A 50% G 50%

| SNP1 | SNP2 | αναμενόμενη συχνότητα   |
|------|------|-------------------------|
| A    | A    | $0,5 \times 0,5 = 25\%$ |
| A    | G    | $0,5 \times 0,5 = 25\%$ |
| C    | A    | $0,5 \times 0,5 = 25\%$ |
| C    | G    | $0,5 \times 0,5 = 25\%$ |

Στο δείγμα μας μπορεί να **παρατηρούμε** μόνο δύο απλότυπους

A A C C **A** T A T C ... C **G** A T T...  $\rightarrow p(AG)=0.5$  ή 50%

A A C C **C** T A T C ... C **A** A T T...  $\rightarrow p(CA)=0.5$  ή 50%

# Παράμετρος ανισορροπίας σύνδεσης, D

Lewontin and Kojima (1960) Evolution 14:450-472.

| Γαμέτης  | Συχνότητα |                |
|----------|-----------|----------------|
| $A_1B_1$ | $x_1$     | $= p_1q_1 + D$ |
| $A_1B_2$ | $x_2$     | $= p_1q_2 - D$ |
| $A_2B_1$ | $x_3$     | $= p_2q_1 - D$ |
| $A_2B_2$ | $x_4$     | $= p_2q_2 + D$ |

$$D = x_1 - p_1q_1$$

= αυτό που παρατηρούμε

= αυτό που περιμένουμε σε  
ισορροπία σύνδεσης

Η τιμή D είναι μια παράμετρος της απόκλισης της συχνότητας του κάθε απλότυπου από το σενάριο της ισορροπίας σύνδεσης

# Παράμετρος ανισοροπίας σύνδεσης, D

## Αντικαθιστώντας:

$$D = x_1 - p_1 q_1$$

|

$$D = x_1 x_4 - x_2 x_3$$

$$p_1 = x_1 + x_2$$

$$p_2 = x_3 + x_4$$

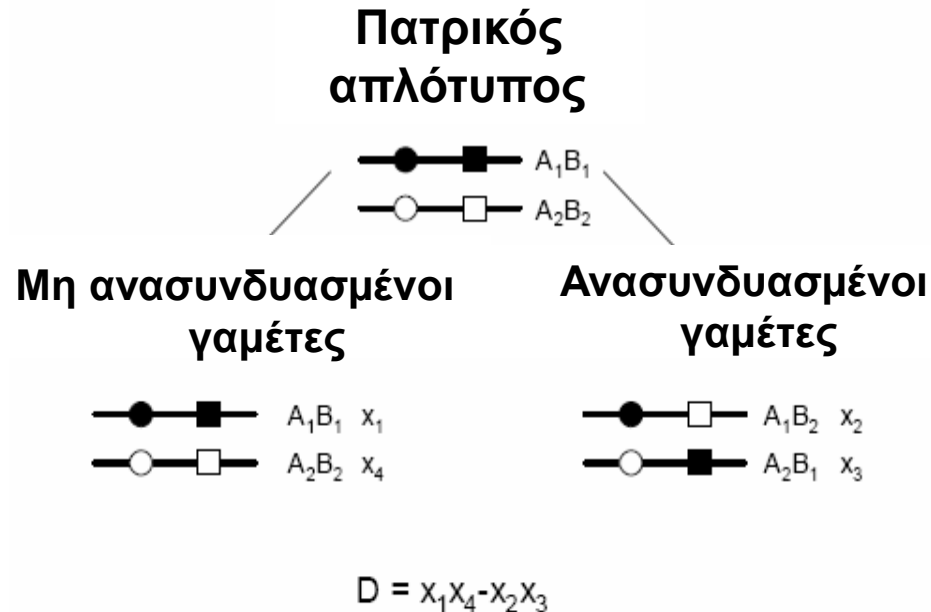
$$q_1 = x_1 + x_3$$

$$q_2 = x_2 + x_4$$

$$x_4 = 1 - x_1 - x_2 - x_3$$



# Παράμετρος ανισοροπίας σύνδεσης, D



**D = το γινόμενο των συχνοτήτων των μη ανασυνδυασμένων γαμετών μείον το γινόμενο των συχνοτήτων των ανασυνδυασμένων γαμετών**

## Estimate of D in case of Linkage Equilibrium

If allele frequencies of p1 and q1 are both 0.5 and equilibrium occurs

$$x_1 = p_1q_1 = 0.5 \times 0.5 = 0.25$$

$$x_4 = p_2q_2 = 0.5 \times 0.5 = 0.25$$

$$x_2 = p_1q_2 = 0.5 \times 0.5 = 0.25$$

$$x_3 = p_2q_1 = 0.5 \times 0.5 = 0.25$$

$$D = x_1 * x_4 - x_2 * x_3$$

$$D = (0.25)(0.25) - (0.25)(0.25) = 0$$

## Estimate of D in case of Linkage Disequilibrium

If allele frequency of  $p_1$  and  $q_1$  are both 0.5 and there is complete non-random association (only  $A_1B_1$  and  $A_2B_2$  exist in the population) with equal allele frequencies at all loci

$$x_1 = p_1q_1 + D = 0.25 + D = 0.5$$

$$x_4 = p_2q_2 + D = 0.25 + D = 0.5$$

$$x_2 = p_1q_2 - D = 0.25 - D = 0$$

$$x_3 = p_2q_1 - D = 0.25 - D = 0$$

$$D = x_1 * x_4 - x_2 * x_3$$

$$D = (0.5)(0.5) - (0)(0) = 0.25$$

## Estimate of D in case of Linkage Disequilibrium

If allele frequency of  $p_1$  and  $q_1$  are both 0.5 and there is complete non-random association (only  $A_1B_2$  and  $A_2B_1$  exist in the population) with equal allele frequencies at all loci

$$x_1 = p_1q_1 + D = 0.25 + D = 0$$

$$x_4 = p_2q_2 + D = 0.25 + D = 0$$

$$x_2 = p_1q_2 - D = 0.25 - D = 0.5$$

$$x_3 = p_2q_1 - D = 0.25 - D = 0.5$$

$$D = x_1 * x_4 - x_2 * x_3$$

$$D = (0) (0) - (0.5) (0.5) = -0.25$$

# Ιδιότητες του D

---

Δύο τόποι ανεξάρτητοι έχουν  $D=0$

- Όταν υπάρχει ανισορροπία σύνδεσης έχουμε  $D \neq 0$

Το πρόσημο είναι αυθαίρετο

- Λαμβάνει τιμές από  $-0,25$  ως  $0,25$

Δύσκολη η σύγκριση ανάμεσα σε περισσότερα ζεύγη δεικτών

Η τιμή του εξαρτάται από τις συχνότητες των αλληλομόρφων

- Απαιτείται στάθμιση
- Μπορούμε να διαιρέσουμε το D με τη μέγιστη τιμή που μπορεί να λάβει για τις συγκεκριμένες συχνότητες αλληλομόρφων

$D_{\max} =$  η μικρότερη τιμή των  $p_1q_2$  και  $p_2q_1$  όταν  $D > 0$

ή

$D_{\max} =$  η μικρότερη τιμή των  $p_1q_1$  και  $p_2q_2$  όταν  $D < 0$

# Σταθμισμένος δείκτης $D'$

---

$$D' = \frac{D}{D_{\max}}$$

$|D'| = 1$ , τότε έχουμε απόλυτη ανισορροπία σύνδεσης

- τουλάχιστον ένας απλότυπος δεν παρατηρείται
- $|D'| < 1$ , σημαίνει πως η απόλυτη προγονική ανισορροπία σύνδεσης έχει διαταραχθεί

Υπερεκτίμηση της ανισορροπίας σύνδεσης

- Ιδιαίτερα όταν το μέγεθος του δείγματος είναι μικρό
- Όταν ένα τουλάχιστον από τα αλληλόμορφα είναι σπάνιο
  - Μεγάλη πιθανότητα να μην δούμε ένα αλληλόμορφο στο δείγμα από τύχη

# Παράμετρος ανισορροπίας σύνδεσης, $r^2$

Hill and Robertson (1968) Genet Res 8:269-294.

Square of the correlation coefficient.

$$r^2 = \frac{D^2}{p_1 p_2 q_1 q_2}$$

$$0 < r^2 < 1$$

$$-1 < r < 1$$

Συντελεστής Συσχέτισης  
κατά Pearson

# Στατιστική συσχέτιση

$$r^2 = (x_1 \cdot x_4 - x_2 \cdot x_3)^2 / (p_1 \cdot p_2 \cdot q_1 \cdot q_2)$$

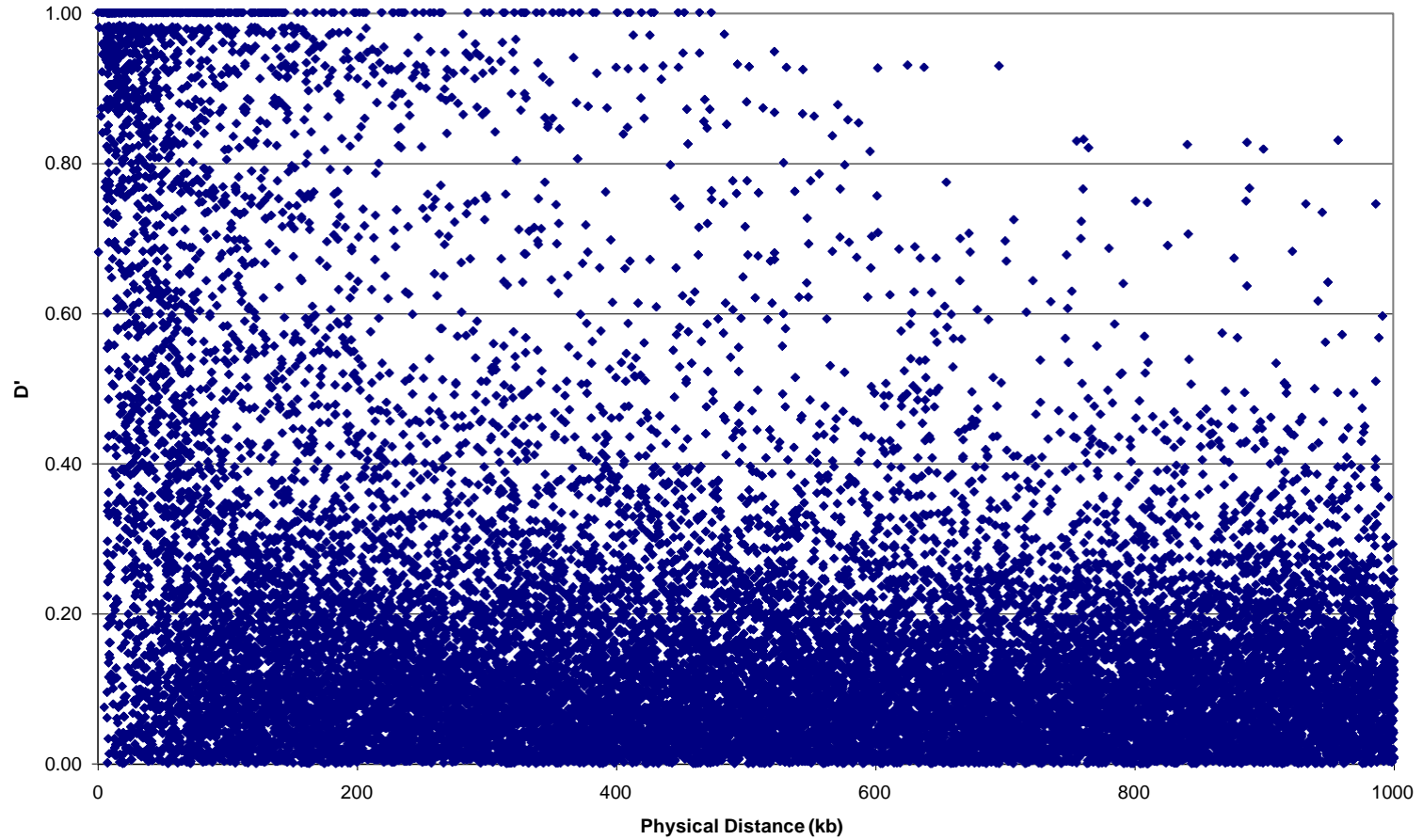
- Εκτιμά στατιστική συσχέτιση
- Για χαμηλές συχνότητες αλληλομόρφων, το  $r^2$  είναι πιο αξιόπιστο από το  $|D'|$ .
- $r^2 = 1$  τέλεια LD. Οι παρατηρήσεις σε ένα δείκτη προσφέρουν πλήρεις πληροφορίες για τον άλλο – δεν είναι απαραίτητη η τυποποίηση και των δύο
- $r^2 = 1$  αν και μόνο αν οι δείκτες δεν έχουν διαχωριστεί ποτέ από ανασυνδυασμό και έχουν ακριβώς τις ίδιες συχνότητες αλληλομόρφων. Τότε, μόνο παρατηρούνται δύο από τους τέσσερις πιθανούς απλότυπους



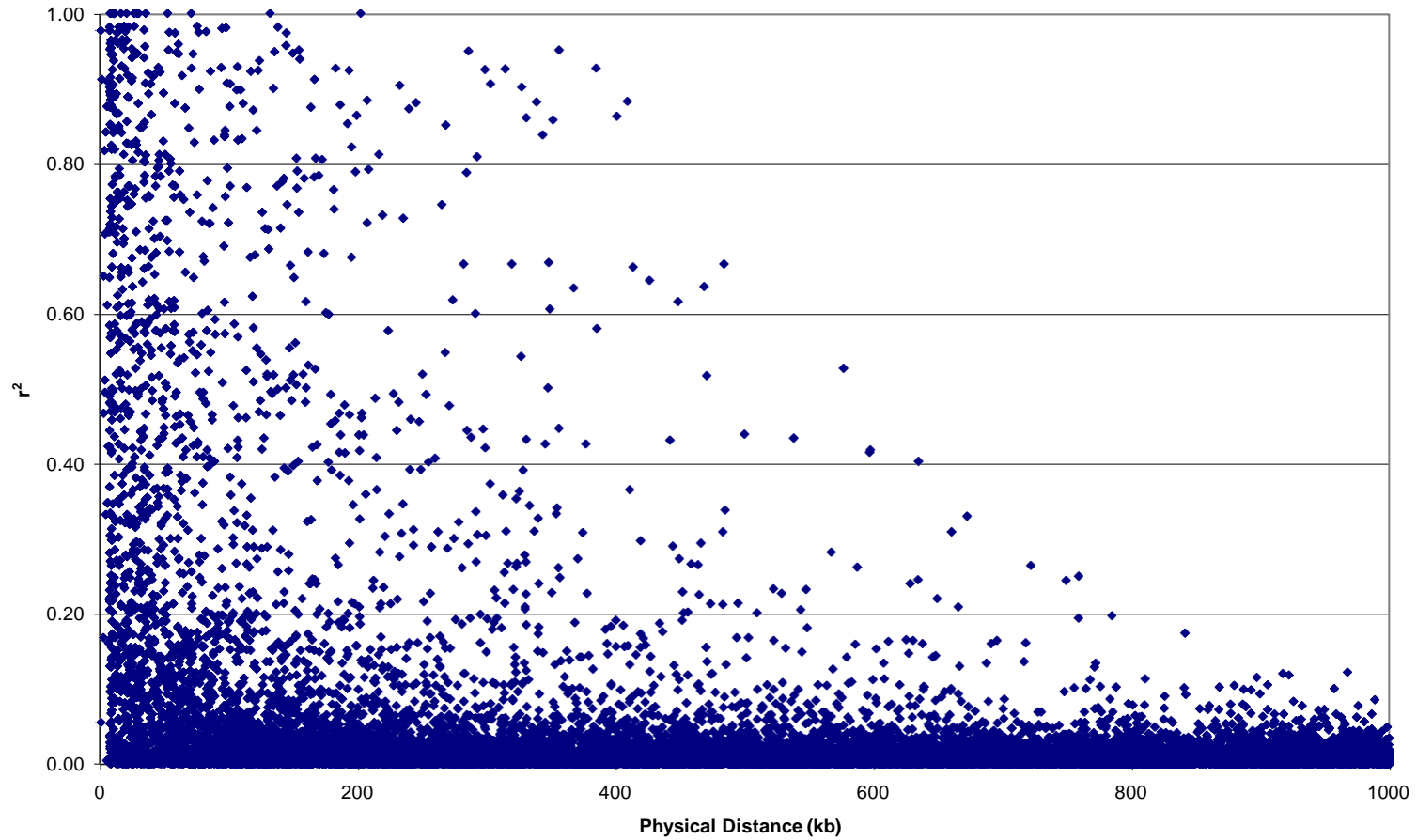
# Σύγκριση

- $r^2$  εξαρτάται από τις συχνότητες των αλληλομόρφων και είναι δύσκολη η ερμηνεία του για σύγκριση ανάμεσα σε πολλαπλούς δείκτες (Hedrick, 1987)
- $|D'|$  σχετίζεται άμεσα με το κλάσμα ανασυνδυασμού και είναι το μόνο μέτρο LD που δεν επηρεάζεται από αλληλικές συχνότητες
- $|D'|=1$  αν εμφανίζονται 2 ή 3 απλότυποι
- $r^2=1$  εμφανίζονται μόνο 2 απλότυποι – τέλεια LD
- Ενδιάμεσες τιμές του  $|D'|$  είναι δύσκολο να ερμηνευτούν

# Raw $|D'|$ data from Chr22



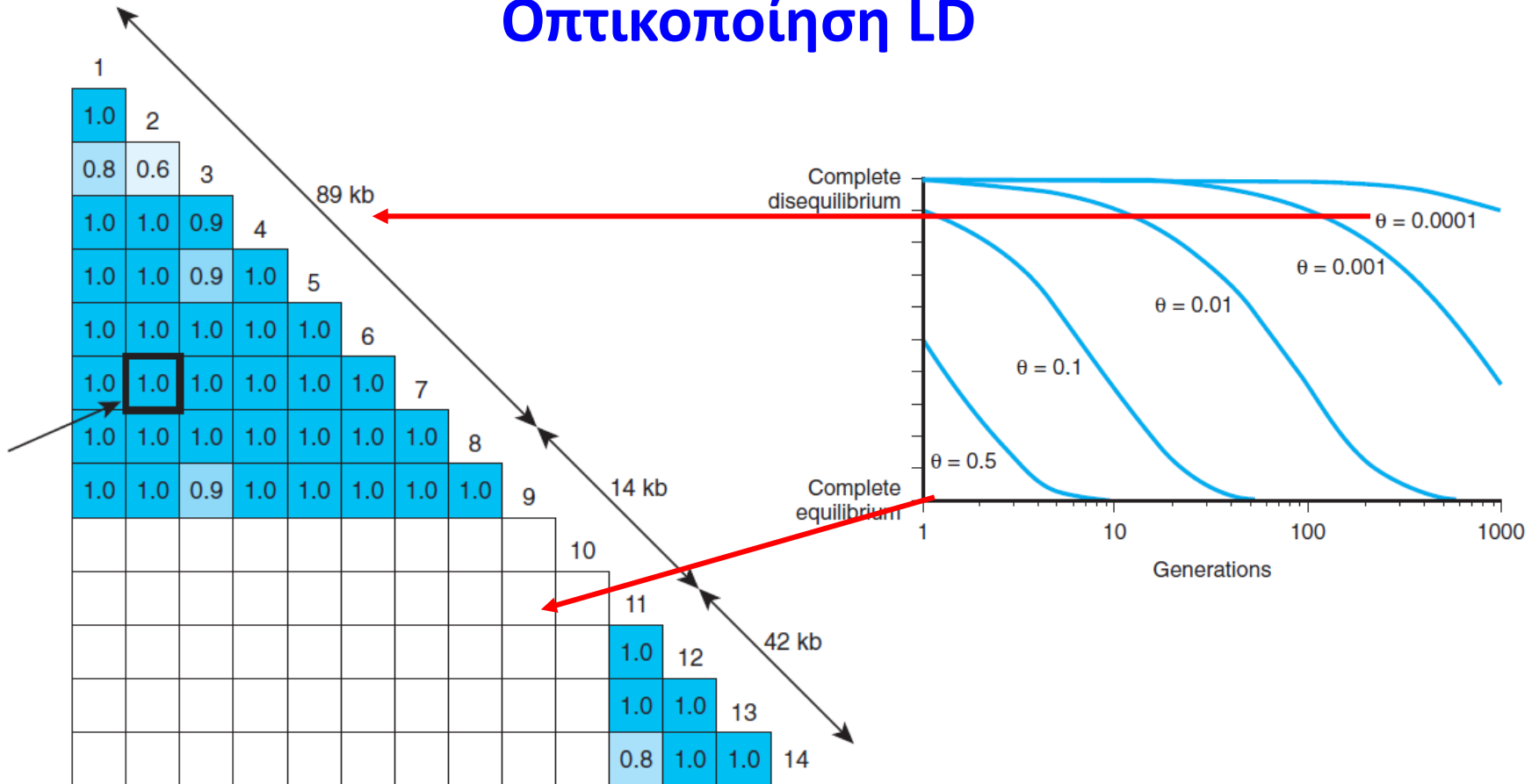
# Raw $r^2$ data from Chr22

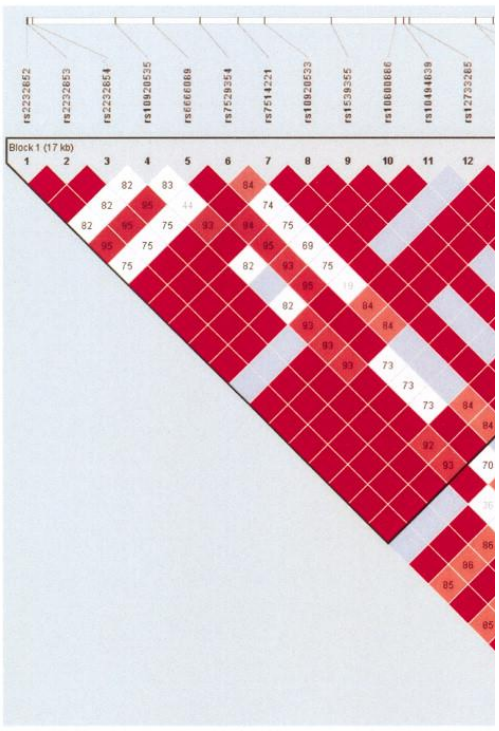




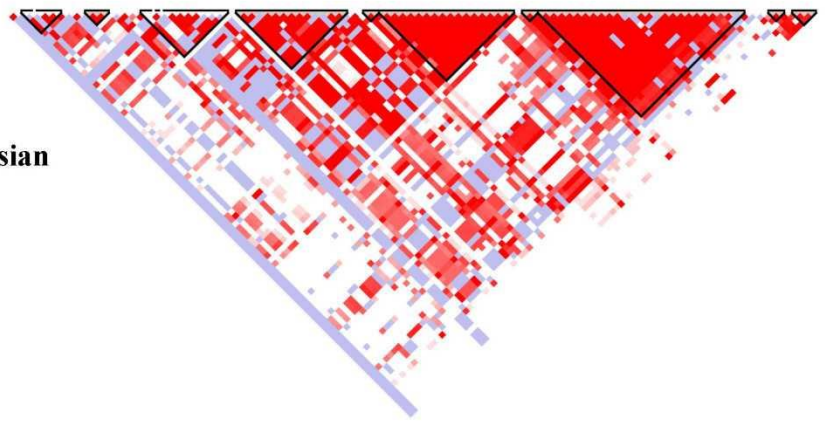
B

# Οπτικοποίηση LD

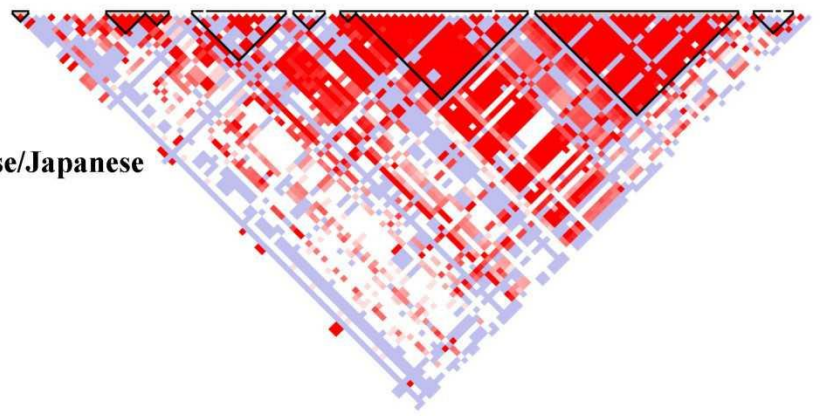




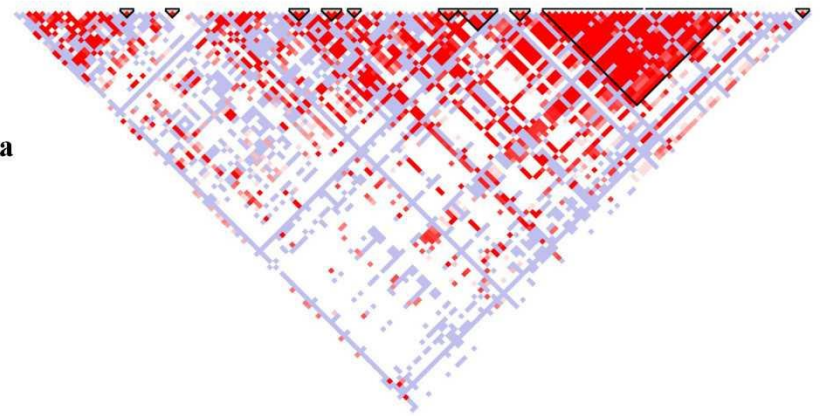
Caucasian



Chinese/Japanese



Yoruba



**B**

|           |           |           |           |           |           |           |            |           |            |            |            |           |           |            |
|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|-----------|------------|------------|------------|-----------|-----------|------------|
| rs2232852 | rs2232853 | rs2232854 | hCV454720 | rs6660089 | rs7529354 | rs7514221 | rs10920533 | rs1539355 | rs10800886 | rs10494839 | rs12733285 | rs2275738 | rs2275737 | rs16850799 |
| T         | T         | T         | C         | A         | C         | T         | A          | G         | T          | C          | C          | C         | C         | G          |
| G         | C         | C         | T         | G         | C         | C         | G          | A         | C          | T          | T          | T         | A         | G          |
| G         | C         | C         | C         | G         | C         | T         | G          | A         | C          | T          | C          | C         | C         | A          |
| G         | C         | C         | C         | G         | T         | C         | G          | A         | C          | T          | C          | T         | A         | G          |



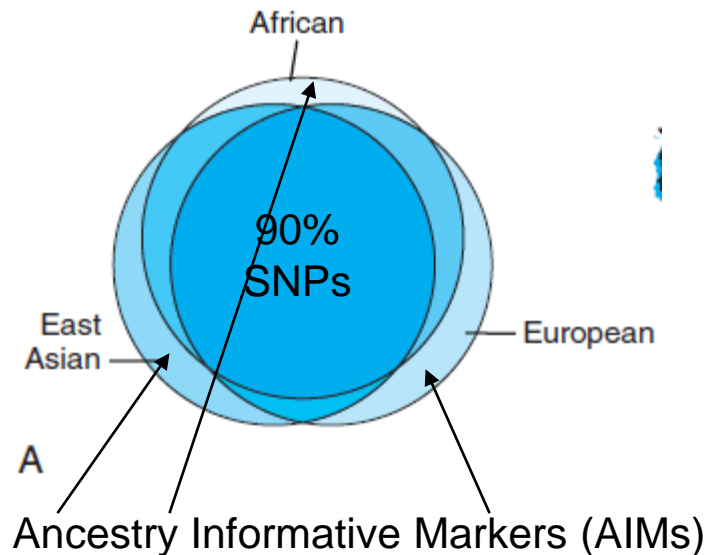
# Γεωγραφική προέλευση και LD

Καθώς η LD εξαρτάται από τη γενετική απόσταση όσο και από τη χρονική περίοδο που πρωτοεμφανίστηκε μία παραλλαγή (πολυμορφισμός), πληθυσμοί με διαφορετική ιστορία (πχ Αφρικανοί vs Καυκάσιοι vs Ασιάτες) μπορεί να δείχνουν διαφορετικές τιμές  $D'$  μεταξύ των ίδιων δεικτών στο γονιδίωμα:

Africans: Ave LD block = 7.3kb

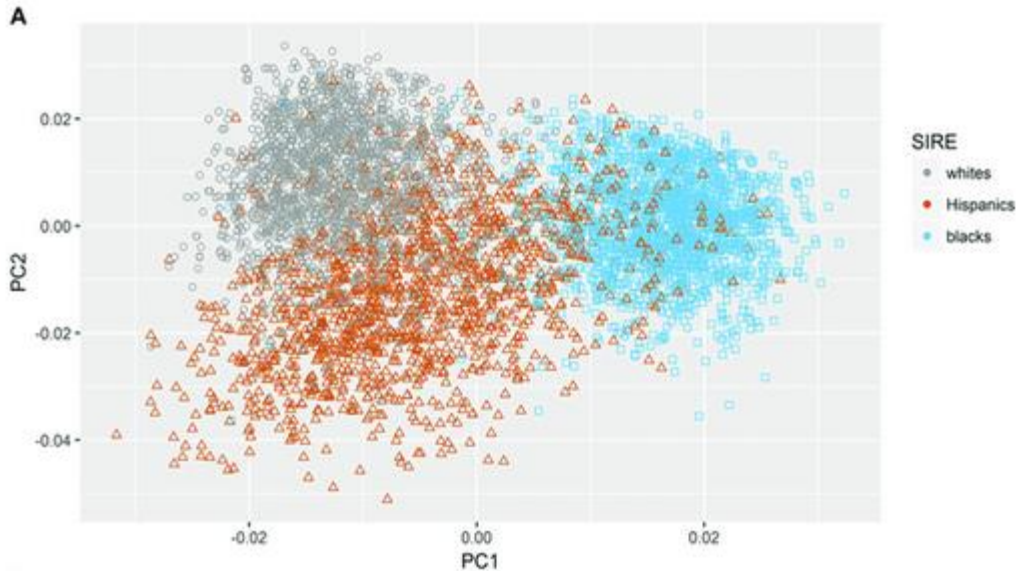
Asians: Ave LD block = 13.2kb

Caucasians: Ave LD block = 16.3kb



# Ancestry Informative Markers – AIMs

## Πληροφοριακοί δείκτες της καταγωγής



Principal Component Analysis (PCA)

Ανάλυση κύριων (ή βασικών) συνιστωσών

- Σημαντικές διαφορές στη συχνότητα αλληλομόρφων ορισμένων παραλλαγών τύπου SNPs παρατηρήθηκαν μεταξύ διαφορετικών πληθυσμών.
- Αποτέλεσμα γενετικής παρέκκλισης ή του φαινομένου του ιδρυτή ή διαφορές επιλογής σε συγκεκριμένες γεωγραφικές περιοχές/ συνθήκες?
- Χρησιμοποιούνται σε μελέτες καταγωγής του ανθρώπινου είδους, της μετανάστευσης και στην ιατροδικαστική (πληθυσμιακή καταγωγή του δράστη) και τις μελέτες συσχέτισης σε επίπεδο ολικού γονιδιώματος (GWAS) για την ελάττωση των ψευδώς θετικών ευρημάτων



Testing for statistical *association* between two *known* SNPs. In this example  $D=0.21-0.0 = 0.21$  and  $D'=1$   
*SNP1* and *SNP2* are in LD

$D_{max}$  = η μικρότερη τιμή των  $p_1q_2$  και  $p_2q_1$  όταν  $D>0$

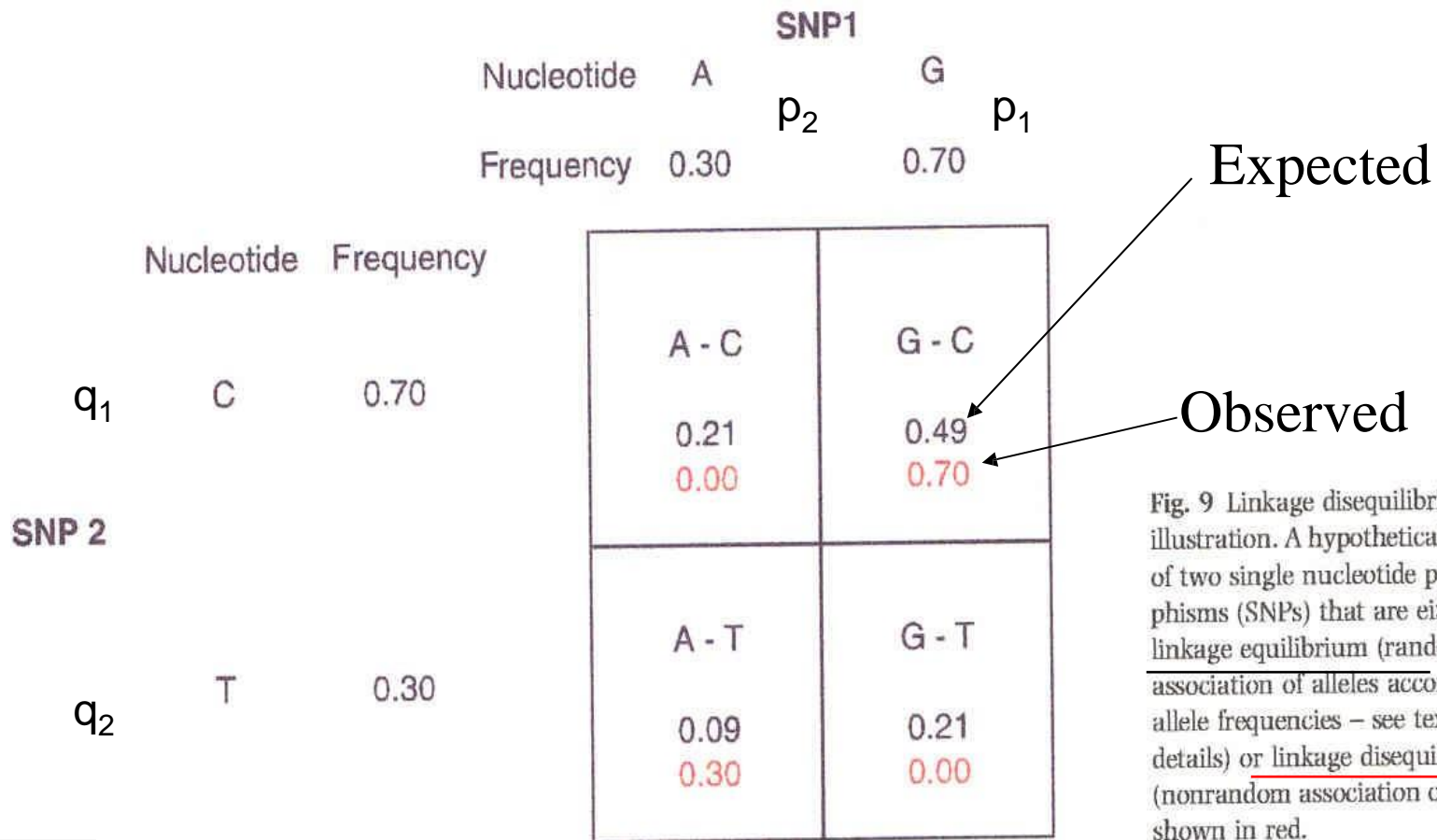


Fig. 9 Linkage disequilibrium illustration. A hypothetical example of two single nucleotide polymorphisms (SNPs) that are either in linkage equilibrium (random association of alleles according to allele frequencies – see text for details) or linkage disequilibrium (nonrandom association of alleles)– shown in red.

From Kaessman & Paabo 2002

(The statistical test is a chi-square test)

# Προσδιορισμός της στατιστικής σημαντικότητας

Determining Significance

$$X^2 = \sum_{i=1}^k \frac{(O - E)^2}{E}$$

$$df = 1$$

<https://www.ncbi.nlm.nih.gov/books/NBK22084/>

# Παράδειγμα – Ομάδες Αίματος

N=2000 άτομα  
(=γονότυποι)

|   | S   | s   |
|---|-----|-----|
| M | 474 | 611 |
| N | 142 | 773 |

|   | S              | s               |
|---|----------------|-----------------|
| M | $x_1$<br>0.237 | $x_2$<br>0.3055 |
| N | $x_3$<br>0.071 | $x_4$<br>0.3865 |

$$MS=474/2000=0.237$$

$$Ms=611/2000=0.3055$$

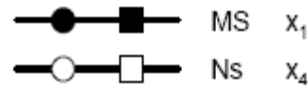
$$NS=142/2000=0.071$$

$$Ns=773/2000=0.3865$$

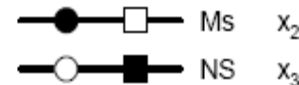
**Πατρικός  
απλότυπος**



**Μη ανασυνδυασμένα  
χρωμοσώματα**



**Αناسυνδυασμένα  
χρωμοσώματα**



[https://pbgworks.org/sites/pbgworks.org/files/measuresoflinkagedisequilibrium-11119214123-phpapp01\\_0.pdf](https://pbgworks.org/sites/pbgworks.org/files/measuresoflinkagedisequilibrium-11119214123-phpapp01_0.pdf)

# Παράδειγμα – Ομάδες Αίματος

|   |                |                 |
|---|----------------|-----------------|
|   | S              | s               |
| M | $x_1$<br>0.237 | $x_2$<br>0.3055 |
| N | $x_3$<br>0.071 | $x_4$<br>0.3865 |

| Gamete | Frequency      | Allele | Frequency                                   |
|--------|----------------|--------|---|
| MS     | $x_1 = 0.2370$ | M      | $p_1 = x_1 + x_2 = 0.237 + 0.3055 = 0.5425$ |
| Ms     | $x_2 = 0.3055$ | N      | $p_2 = x_3 + x_4 = 0.071 + 0.3865 = 0.4575$ |
| NS     | $x_3 = 0.0710$ | S      | $q_1 = x_1 + x_3 = 0.237 + 0.071 = 0.308$   |
| Ns     | $x_4 = 0.3865$ | s      | $q_2 = x_2 + x_4 = 0.3055 + 0.3865 = 0.692$ |

Υπόθεση  $H_0$ : Εφόσον παρατηρούμε όλους τους απλότυπους, οι τόποι βρίσκονται σε ισορροπία σύνδεσης (ασύνδετοι)

# Παράδειγμα – Ομάδες Αίματος

| Gamete | Frequency      | Allele | Frequency                      |
|--------|----------------|--------|--------------------------------|
| MS     | $X_1 = 0.237$  | M      | $p_1 = 0.5425$                 |
| Ms     | $X_2 = 0.3055$ | N      | $p_2 = 0.4575$                 |
| NS     | $X_3 = 0.071$  | S      | $q_1 = 0.308$                  |
| Ns     | $X_4 = 0.3865$ | s      | $q_2 = 0.695$ <del>0.692</del> |

$$D = X_1X_4 - X_2X_3$$

$$D = (0.237)(0.3865) - (0.3055)(0.071) = 0.07$$

$$D' = \frac{D}{D_{\max}}$$

$$D' = \frac{0.07}{0.141} = 0.5 \quad p_2q_1 = 0.141 < p_1q_2 = 0.377$$

$$r^2 = \frac{D^2}{p_1q_2q_1q_2}$$

~~$p_2$~~

$$r^2 = \frac{0.07^2}{(0.5425)(0.695)(0.308)(0.695)} = 0.06$$

~~0.692~~      ~~0.092~~

0.4575      0.692      0.092

$$r = \sqrt{\frac{D^2}{p_1q_2q_1q_2}}$$

~~$p_2$~~

$$r = \sqrt{0.06} = 0.25$$

0.092      0.304

# Παράδειγμα – Ομάδες Αίματος

$$X^2 = \sum_{i=1}^k \frac{(O - E)^2}{E} \quad df = 1$$

degrees of freedom,  $df = (r - 1) * (c - 1)$   
 where r = number of rows  
 c = number of columns

Observed

|   | S   | s   |
|---|-----|-----|
| M | 474 | 611 |
| N | 142 | 773 |

Expected

|   | S      | s      |
|---|--------|--------|
| M | 334.18 | 750.82 |
| N | 281.82 | 633.18 |

$E_i$  is the expected frequency

$$E_i = \frac{r_{i\text{total}} * c_{i\text{total}}}{\text{grand}_{\text{total}}}$$

$$X^2 = \frac{(474 - 334.18)^2}{334.18} + \frac{(611 - 750.82)^2}{750.82} + \frac{(142 - 281.82)^2}{281.82} + \frac{(773 - 633.18)^2}{633.18} \approx 184$$

Καταλήγουμε πως η υπόθεση  $H_0$  καταρρίπτεται: Υπάρχει σημαντικό LD μεταξύ των δύο γενετικών τόπων ( $D' = 0.5$ ) - Η παρατηρούμενη απόκλιση ανάμεσα στο πείραμα (observed) και τις θεωρητικές (expected) τιμές είναι μάλλον απίθανο (όχι όμως αδύνατο) ότι οφείλεται μόνο στην τύχη.

$$p < 0.001$$

# Determining Significance

$$X^2 = \sum_{i=1}^k \frac{(O - E)^2}{E} \quad df = 1$$

|   | S  | s  |
|---|----|----|
| M | 1  | 0  |
| N | 50 | 49 |

|   | S     | s     |
|---|-------|-------|
| M | 0.51  | 0.49  |
| N | 50.49 | 48.51 |

$$D'=1$$

$$X^2 = 0.97$$

$$p > 0.10$$

Υπόθεση  $H_0$ : Εφόσον δεν παρατηρούμε όλους τους απλότυπους, οι τόποι βρίσκονται σε ανισορροπία σύνδεσης (συνδεδεμένοι)

Καταλήγουμε πως η υπόθεση  $H_0$  ΔΕΝ καταρρίπτεται: Υπάρχει σημαντικό LD μεταξύ των δύο γενετικών τόπων ( $D'=1$ ) – Η παρατηρούμενη απόκλιση ανάμεσα στις παρατηρούμενες (observed, αριστερά) και τις θεωρητικές (expected, δεξιά) τιμές οφείλεται στην τύχη - Το πείραμα συμβαδίζει με την υπόθεση που κάναμε και δεν πρέπει να απορρίψουμε την υπόθεσή μας.

Προσοχή στη χρήση του *Chi square* test για τιμές κελιών <5

πληθυσμός

## Παράγοντες που επηρεάζουν την ανισορροπία σύνδεσης

γονιδίωμα

Η ανισορροπία σύνδεσης διαφέρει ανάμεσα σε διαφορετικούς πληθυσμούς και ανάμεσα σε διαφορετικές γενετικές περιοχές

- Φυσική απόσταση ανάμεσα σε δύο γενετικούς τόπους
  - Ρυθμός ανασυνδυασμού του γονιδιώματος
  - Μεταλλάξεις (γονιδιακή ροή)
  - Φυσική επιλογή
- 
- Τυχαία γεγονότα (γενετική παρέκκλιση)
    - Φαινόμενο του ιδρυτή (founder effect)
    - Γενετική στενωπός (bottleneck)
  - Πληθυσμιακές προσμίξεις (admixture) ή ομομιξία (απομονωμένοι πληθυσμοί)



Το  $D$  σχετίζεται με το ρυθμό  
ανασυνδυασμού

$$D_t = (1 - \theta)^t D_0$$

Όπου  $\theta$  το κλάσμα ανασυνδυασμού

# Χωρίς ανασυνδυασμό

|     |                    |                    |       |
|-----|--------------------|--------------------|-------|
|     | $B$                | $b$                |       |
| $A$ | $p_A p_B + D_{AB}$ | $p_A p_b - D_{AB}$ | $p_A$ |
| $a$ | $p_a p_B - D_{AB}$ | $p_a p_b + D_{AB}$ | $p_a$ |
|     | $p_B$              | $p_b$              |       |

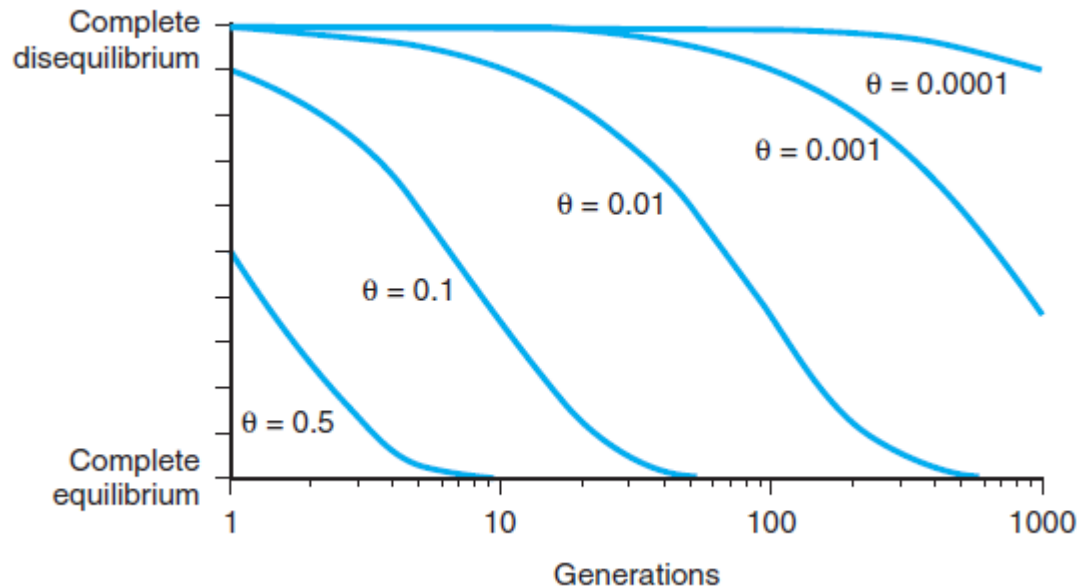
Οι συχνότητες των απλοτύπων παραμένουν σταθερές στην διάρκεια του χρόνου

# Με ανασυνδυασμό

|     |                                 |  |                                 |
|-----|---------------------------------|--|---------------------------------|
|     | $B$                             |  | $b$                             |
| $A$ | $p_A p_B + (1 - \theta) D_{AB}$ |  | $p_A p_b - (1 - \theta) D_{AB}$ |
| $a$ | $p_A p_b - (1 - \theta) D_{AB}$ |  | $p_a p_b + (1 - \theta) D_{AB}$ |
|     | $p_B$                           |  | $p_b$                           |

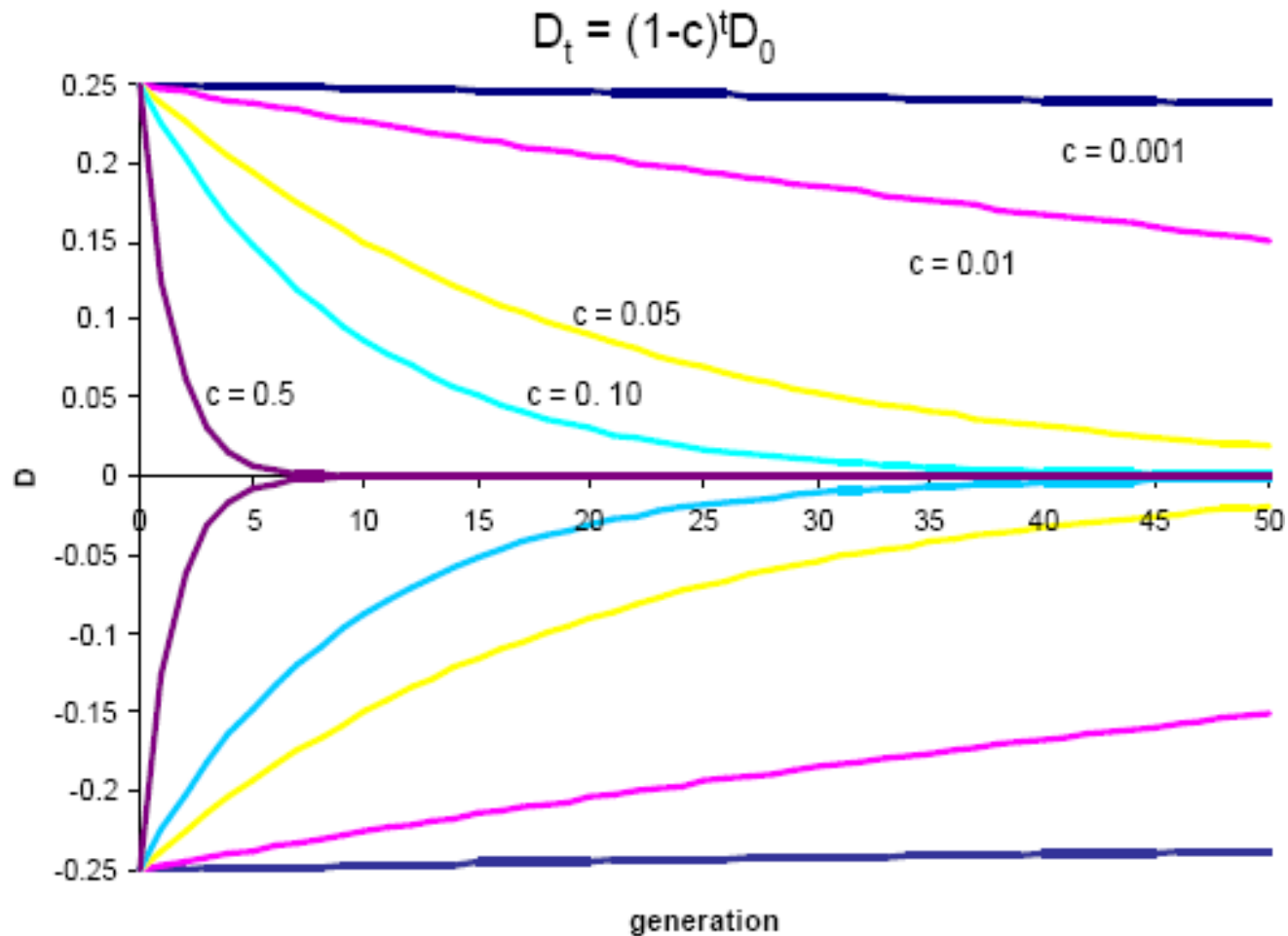
Η ανισορροπία σύνδεσης μειώνεται...

# Θεωρητικός ρυθμός μείωσης της LD μεταξύ δύο γενετικών τόπων ως συνάρτηση του χρόνου (γενιές)



**Figure 10-11** ■ Theoretical rate at which initial linkage disequilibrium between alleles at two loci decays and the alleles approach linkage equilibrium as a function of time and various values of recombination frequency,  $\theta$ , between the markers. (Adapted from original figure by G. Abecasis, University of Michigan. <http://www.sph.umich.edu/csg/labecasis/class/666.03.pdf>.)

# Change in D over time



Each generation will decay until  $D = 0$ .

# Linkage disequilibrium – understanding the evolutionary past and mapping the medical future

Montgomery Slatkin

*Nature Reviews Genetics* 9, 477–485(2008) | [Cite this article](#)

Linkage disequilibrium (LD) is one of those unfortunate terms that does not reveal its meaning. As every instructor of population genetics knows, the term is a barrier not an aid to understanding. LD means simply a nonrandom association of alleles at two or more loci, and detecting LD does not ensure either linkage or a lack of equilibrium. The term was first used in 1960 by Lewontin and Kojima and it persists because LD was initially the concern of population geneticists who were

not picky about definition was c which to study L biology and huma of population ge rapidly in the 1980 mapping became evid closely linked loci was too well establ

- Linkage disequilibrium (LD) is the nonrandom association of alleles of different loci. There is no single best statistic that quantifies the extent of LD. Several statistics have been proposed that are useful for different purposes.
- In humans and other model organisms, LD between marker alleles and traits of interest allow fine-scale gene mapping. Many recent genome-wide association studies have successfully mapped SNPs associated with complex inherited diseases in humans.

# Η ανισορροπία σύνδεσης είναι χρήσιμη...

- Για τη διερεύνηση της ιστορίας των πληθυσμών
- Για τη χαρτογράφηση γονιδίων  
(γνωρίσματα/χαρακτήρες ή νοσήματα/διαταραχές)

# Χαρτογράφηση γονιδίων και ανισορροπία σύνδεσης



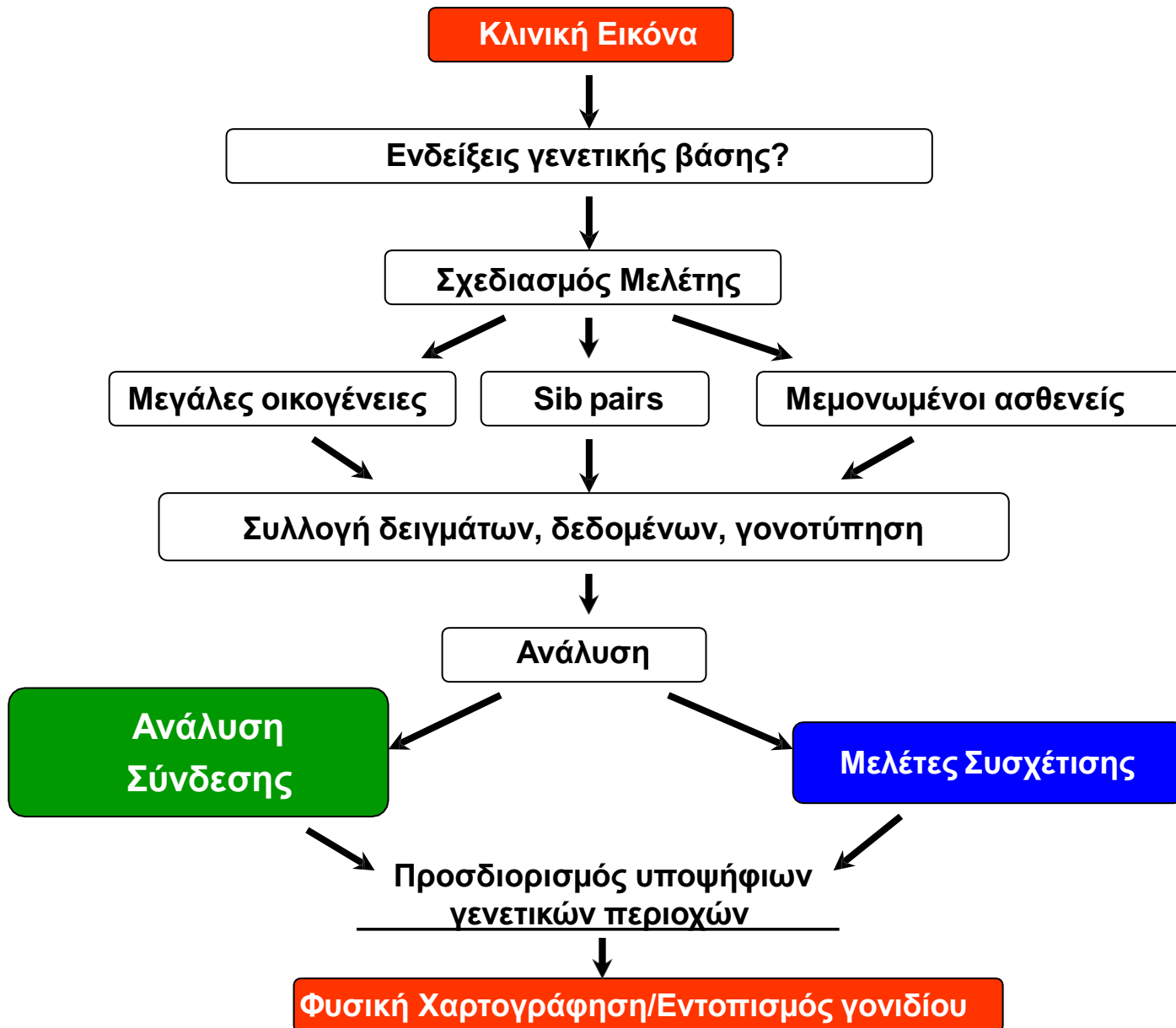
## Μονογονιδιακά νοσήματα

- Μικρή επίπτωση για τη δημόσια υγεία
- Ένα γονίδιο – μία νόσος
- Μενδελιανή κληρονομικότητα
- Σπάνιες μεταλλάξεις
- Κλασσική γενετική
- Παραδείγματα:
  - Νόσος του Huntington
  - Κυστική Ίνωση

## Πολύπλοκα νοσήματα

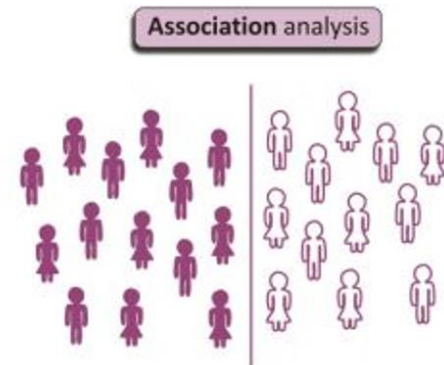
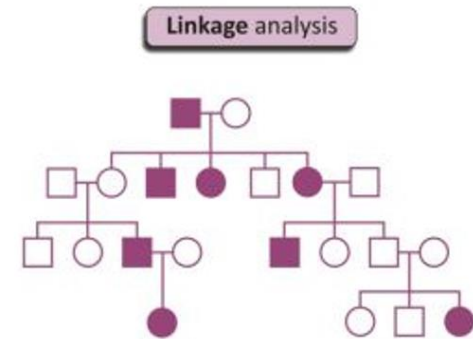
- Σοβαρή επίπτωση για τη δημόσια υγεία
- Πολλά γονίδια – μία νόσος
- Πολύπλοκο πρότυπο κληρονομικότητας
- Συνήθεις γενετικοί πολυμορφισμοί
- Αδυναμία κλασσικής γενετικής
- Παραδείγματα:
  - Έμφραγμα
  - Διαβήτης
  - Σχιζοφρένεια

# Γενικό σχήμα χαρτογράφησης γονιδίων



# Σύνδεση και Συσχέτιση (Linkage and Association)

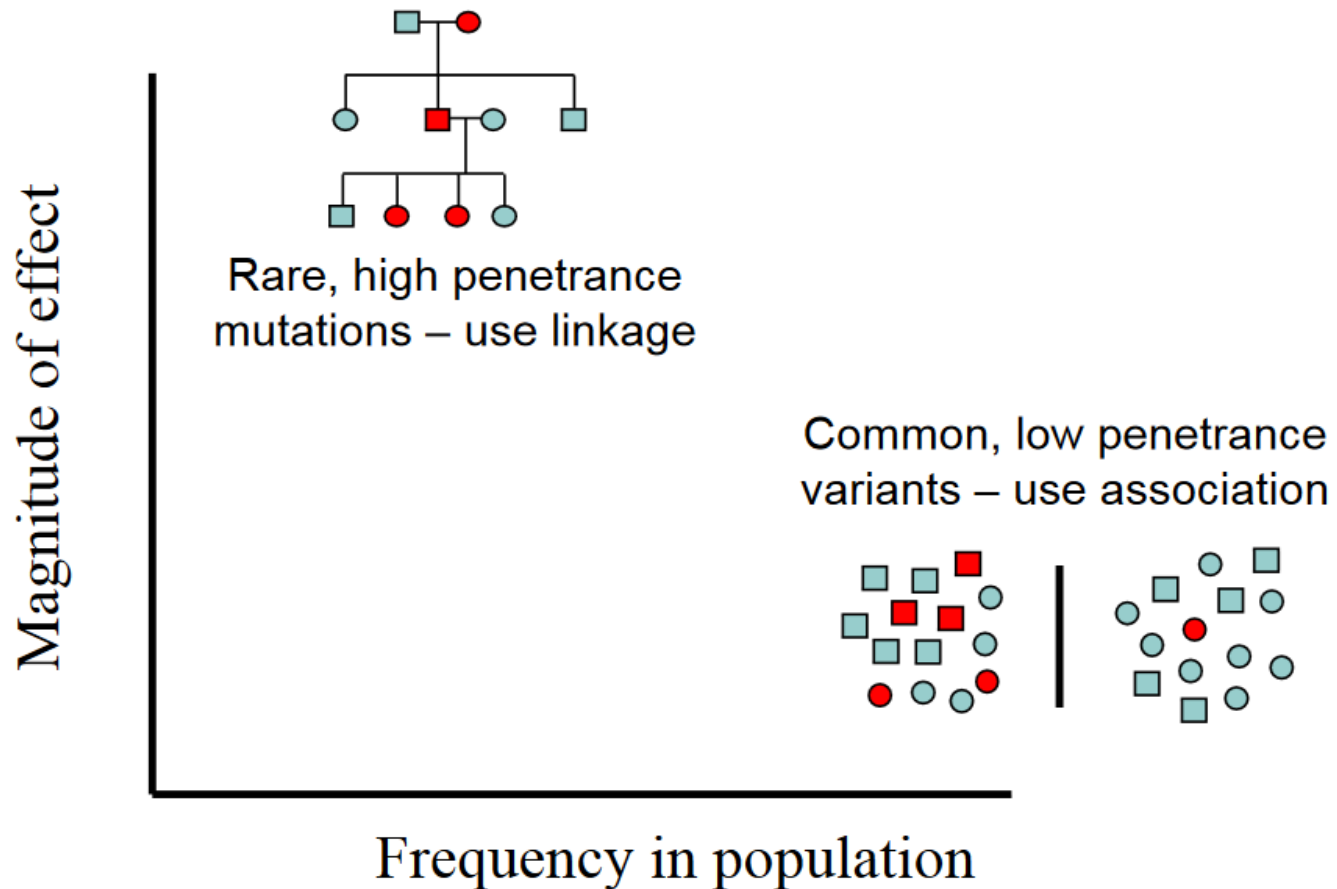
- Οι **μελέτες ανάλυσης σύνδεσης** χρησιμοποιούν μεμονωμένες οικογένειες με προσβεβλημένα μέλη και προσπαθούν να δείξουν σύνδεση ανάμεσα στον φαινότυπο και γενετικούς δείκτες (συσχετίζονται γενετικοί τόποι και όχι αλληλόμορφα)
- Οι **μελέτες συσχέτισης** βασίζονται σε πληθυσμούς και προσπαθούν να δείξουν συσχέτιση ανάμεσα σε ένα συγκεκριμένο αλληλόμορφο ενός γενετικού δείκτη και την προδιάθεση π.χ. για μια νόσο (στατιστικός έλεγχος της ταυτόχρονης παρουσίας ενός συγκεκριμένου αλληλομόρφου και ενός φαινότυπου)

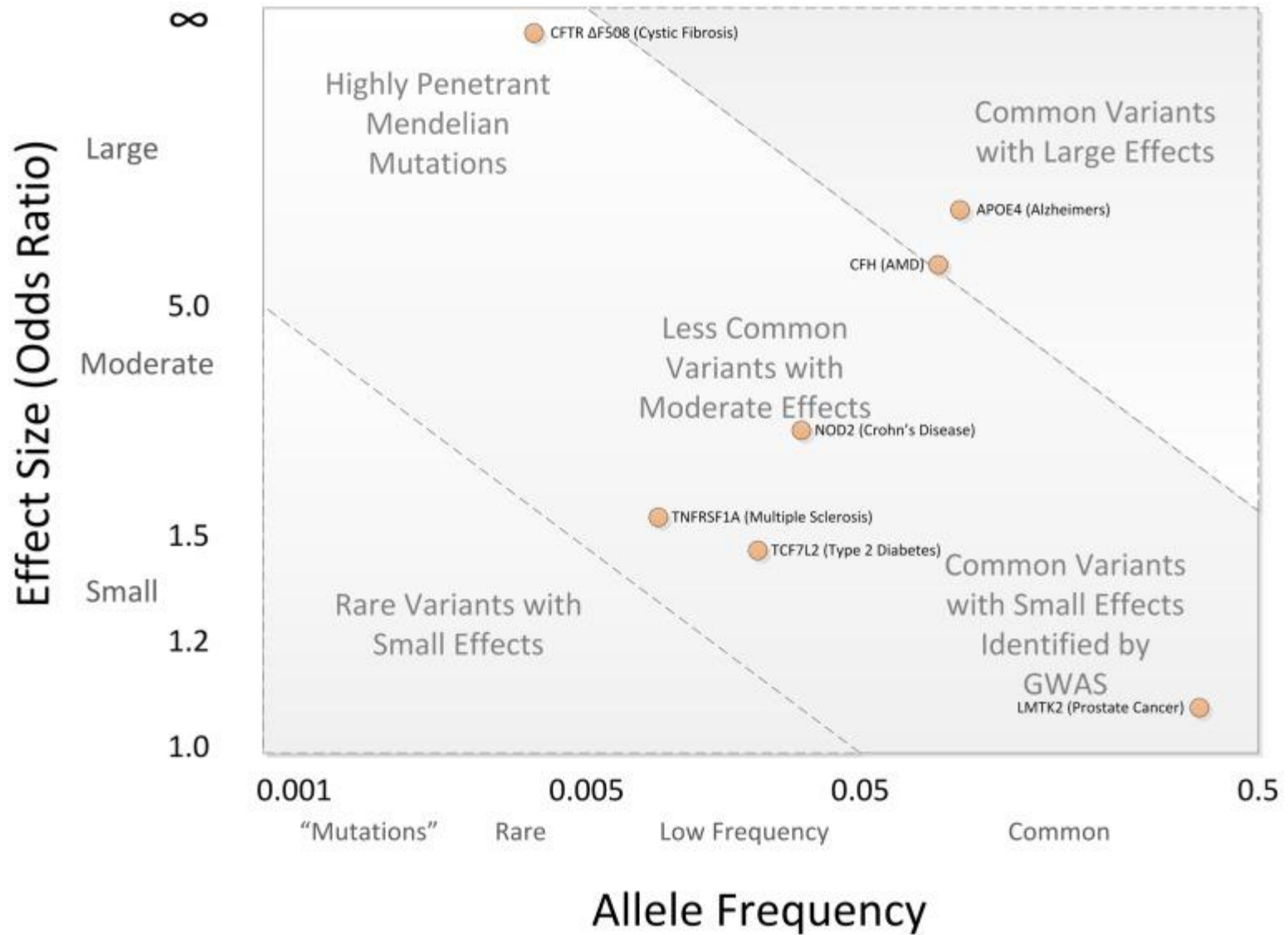


# The Allelic Architecture of Disease

What is it and how do we discover it?

---



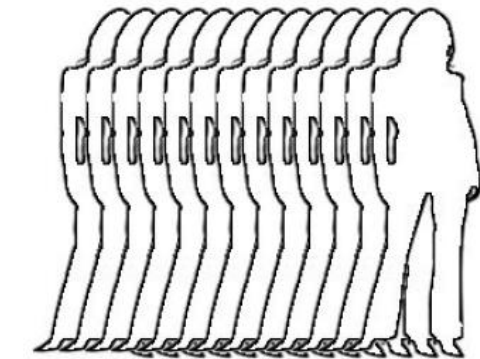


# Ποικιλότητα του ανθρώπινου DNA

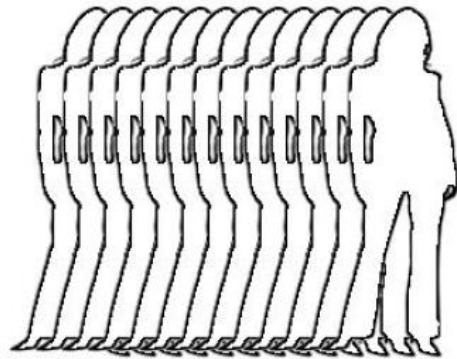
- Δύο άτομα είναι όμοια κατά 99.8%
- Το 90% της διαφοράς οφείλεται σε SNPs
- Υπολογίζονται 10.000.000 συνήθη SNPs (MAF>10%)
- dbSNP: >9.000.000 καταχωρήσεις

Ποιούς δείκτες πρέπει να επιλέξουμε;

# Μελέτες συσχέτισης



GC CC GG GC CC GC GC  
GG CC GC GG GC GG



GC CC GC GC GG CC CC  
CC GC GC GG GC GG

SNP1

**Cases**

Count of G:  
2104 of 4000

Frequency of G:  
52.6%

**Controls**

Count of G:  
2676 of 6000

Frequency of G:  
44.6%

**P-value:**

$5.0 \cdot 10^{-15}$

SNP2

**Cases**

Count of G:  
1648 of 4000

Frequency of G:  
41.2%

**Controls**

Count of G:  
2532 of 6000

Frequency of G:  
42.2%

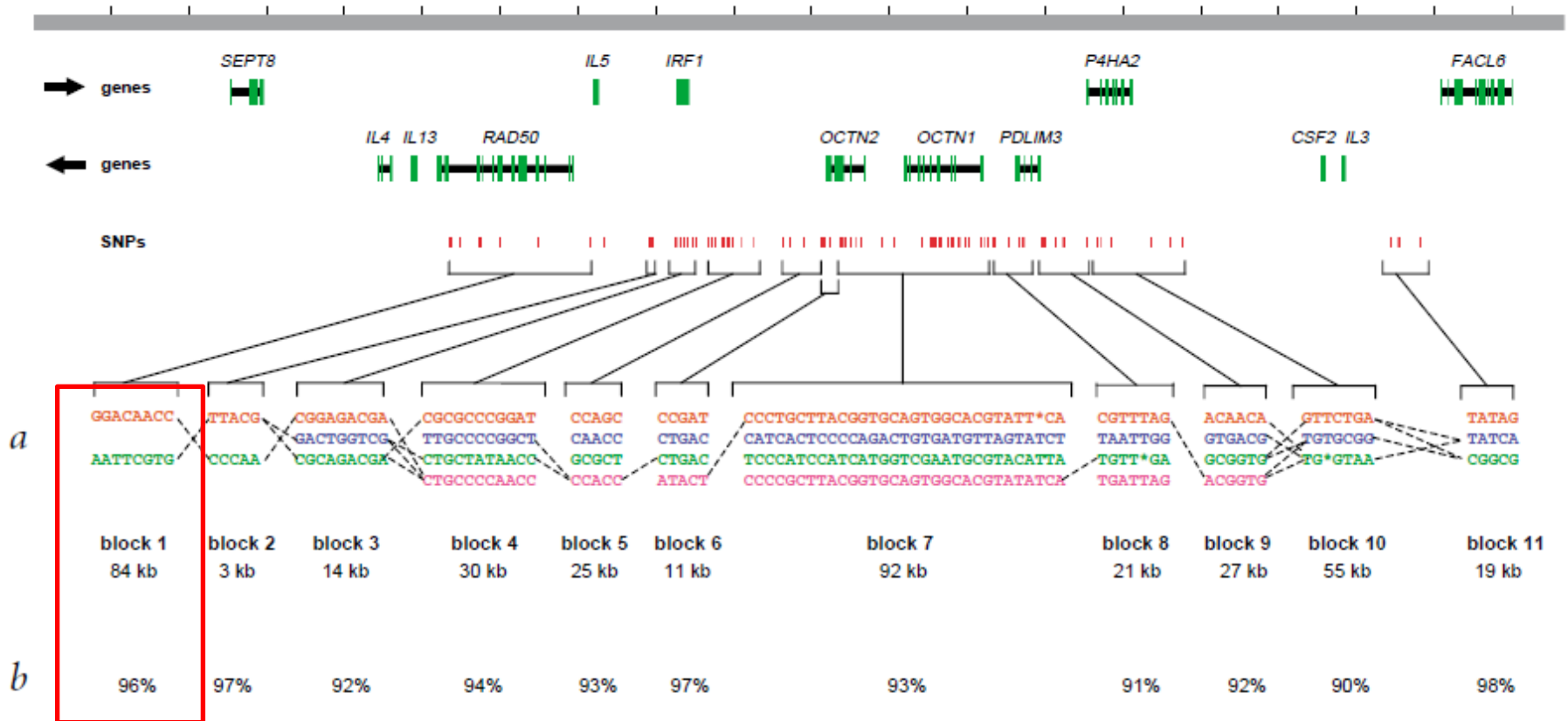
**P-value:**

0.33

SNP ...

*Repeat for all  
SNPs*

# A block-like structure of the human genome...



**Table 1 • Haplotypes of SNPs in block 1 (8 SNPs/84 kb)\***

| Haplotype       | Observations            |
|-----------------|-------------------------|
| G G A C A A C C | 283 (83.2%) haplotype A |
| A A T T C G G G | 40 (11.8%) haplotype B  |
| G A T T A G C C | 2 (0.6%)                |
| G G T C A G C C | 2 (0.6%)                |

\*Another 13 chromosomes (3.8%) were observed that matched haplotype A or B at all alleles except one, and might represent gene conversion or an undetected genotyping error.

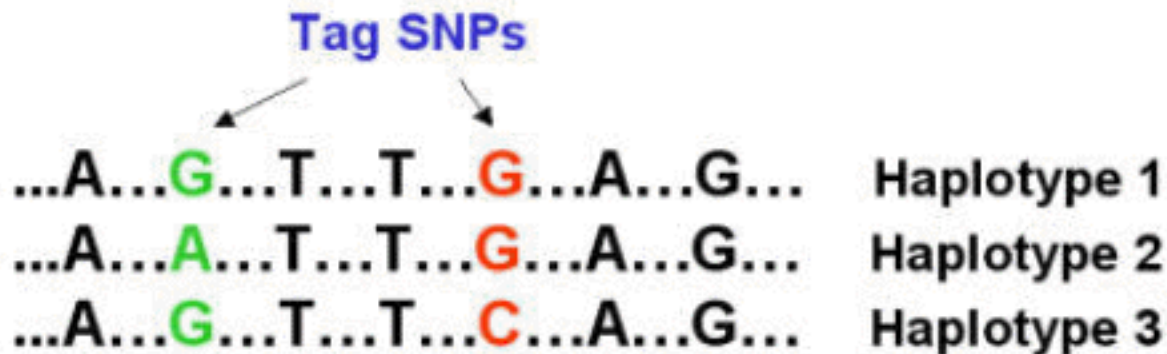


# Block απλοτύπων και μελέτες συσχέτισης

- Ιδανικά θα θέλαμε να μπορούμε να συγκρίνουμε όλα τα υπάρχοντα SNPs ανάμεσα σε μια μεγάλη ομάδα ασθενών και μια ομάδα controls
- Απαγορευτικό κόστος
- Χάρη στη δομή του ανθρώπινου γονιδιώματος δεν είναι απαραίτητο να γονοτυπήσουμε όλα τα SNPs
- Οι μελέτες συσχέτισης μπορούν να επικεντρωθούν μόνο στα αντιπροσωπευτικά **tagging SNPs (tagSNPs ή tSNPs)**

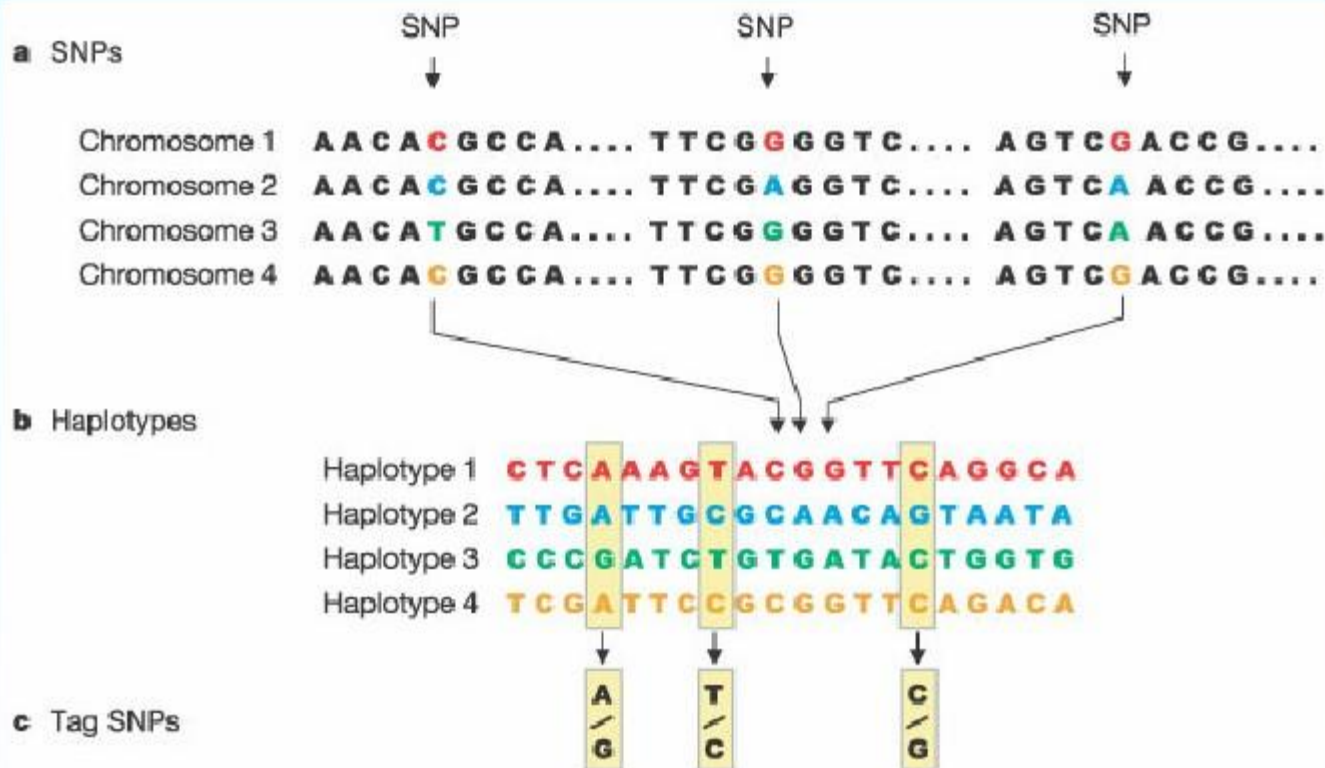
# htSNPs-haplotype tagging SNPs

- Μια υπο-ομάδα SNPs σε έναν απλότυπο που είναι αρκετά για να καθορίσουν όλα τα υπόλοιπα



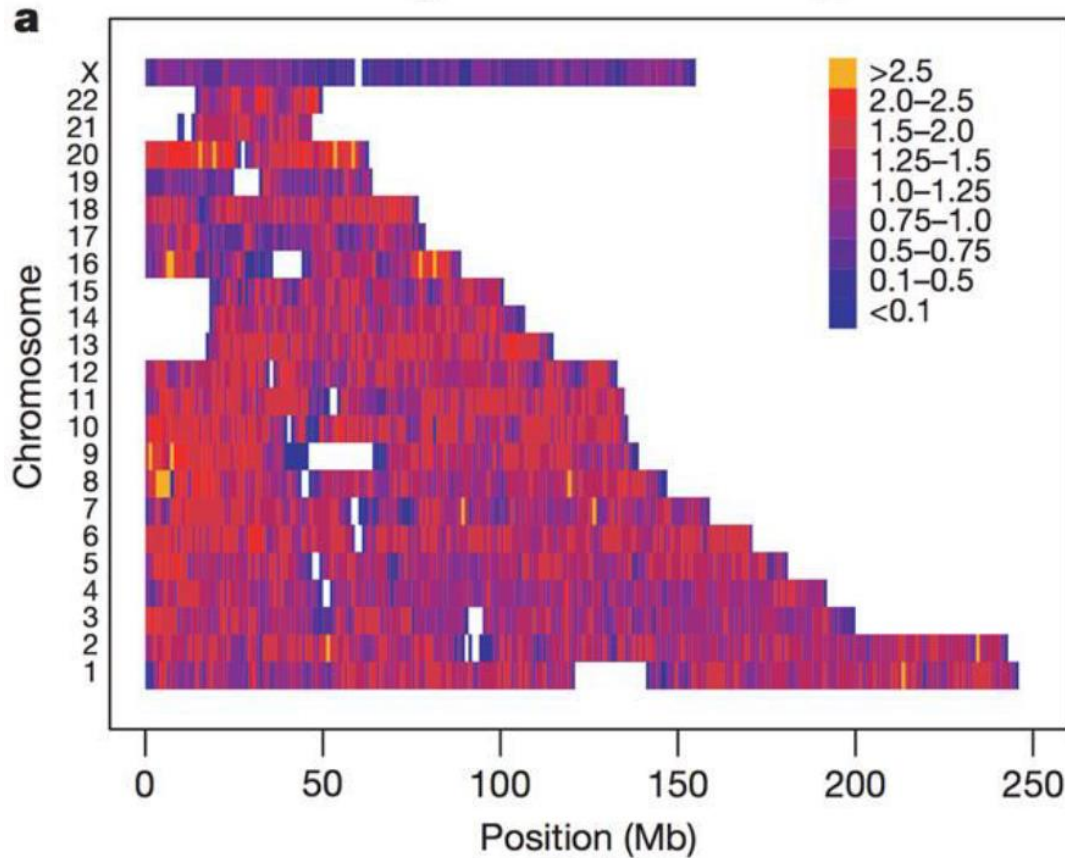
if a chromosome has alleles G and C at the two tag SNPs position, then it has the **Haplotype 3**.

# Human Sequence Variation



The International HapMap Project, Nature 2003

# SNP density across genome



The color scale shows SNP density per 1000 bases.

# The HapMap Project

## Στόχοι:

- Καθορισμός της δομής του ανθρώπινου γονιδιώματος
- Προσδιορισμός tSNPs
- Δημοσιοποίηση της πληροφορίας
- Ισχυρό εργαλείο για πιο αποτελεσματικές μελέτες γενετικής συσχέτισης

Φάση I: 1.000.000 SNPs (1 SNP/5kb) για 270 ανθρώπους

*The International HapMap Consortium (2005), Nature*



# HapMap Population Samples

Project launched in 2002 to provide a public resource for accelerating medical genetic research

## 270 Individuals from 4 Geographically Diverse Populations

**YRI:** 90 Yorubans from Ibadan, Nigeria

30 parent-offspring trios

**CEU:** 90 northern and western European-descent living in Utah, USA from the Centre d'Etude du Polymorphisme Humain (CEPH) collection

30 parent-offspring trios

**CHB:** 45 unrelated Han Chinese from Beijing, China

**JPT:** 45 unrelated Japanese from Tokyo, Japan

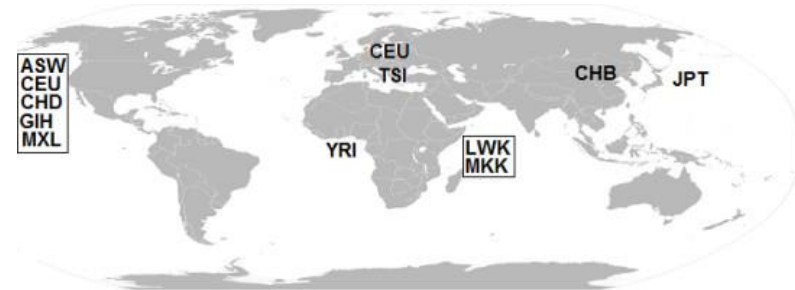
<http://www.hapmap.org/>

<http://www.genome.gov/page.cfm?pageID=10001688>

**HapMap**  
**NHGRI**



# HapMap Project



|                      | Phase 1                         | Phase 2                | Phase 3                        |
|----------------------|---------------------------------|------------------------|--------------------------------|
| Samples & POP panels | 269 samples (4 panels)          | 270 samples (4 panels) | 1,115 samples (11 panels)      |
| Genotyping centers   | HapMap International Consortium | Perlegen               | Broad & Sanger                 |
| Unique QC+ SNPs      | 1.1 M                           | 3.8 M (phase I+II)     | 1.6 M (Affy 6.0 & Illumina 1M) |
| Reference            | Nature (2005) 437:p1299         | Nature (2007) 449:p851 | Draft Rel. 1 (May 2008)        |

**Table 1** HapMap details

|           | No. of SNPs genotyped | Targeted SNPs  | Populations studied                                   |
|-----------|-----------------------|--|---|
| Phase I   | 1 million             | Prioritized coding SNPs to attain 1 SNP for each 5 kb region | CEU, YRI, CHB, JPT                                    |
| Phase II  | 3 million             | Prioritized non-synonymous SNPs in coding regions            | CEU, YRI, CHB, JPT                                    |
| Phase III | 1.4 million           | Prioritized rare variants                                    | CEU, YRI, CHB, JPT, ASW, CHB, GIH, LWK, MXL, MKK, TSI |

SNP, single nucleotide polymorphism.

# Συμπεράσματα...

- Πλεονασμός (redundancy) της πληροφορίας των SNPs
- Δυνατή η επιλογή tagging SNPs (tagSNPs, tSNPs)
- Δυνατή η ανάκτηση (πρόβλεψη) ολόκληρης της πληροφορίας με δεδομένο μόνο ένα μικρό αριθμό tSNPs
- Μείωση κόστους και χρόνου
- Έλευση τεχνολογιών next generation sequencing