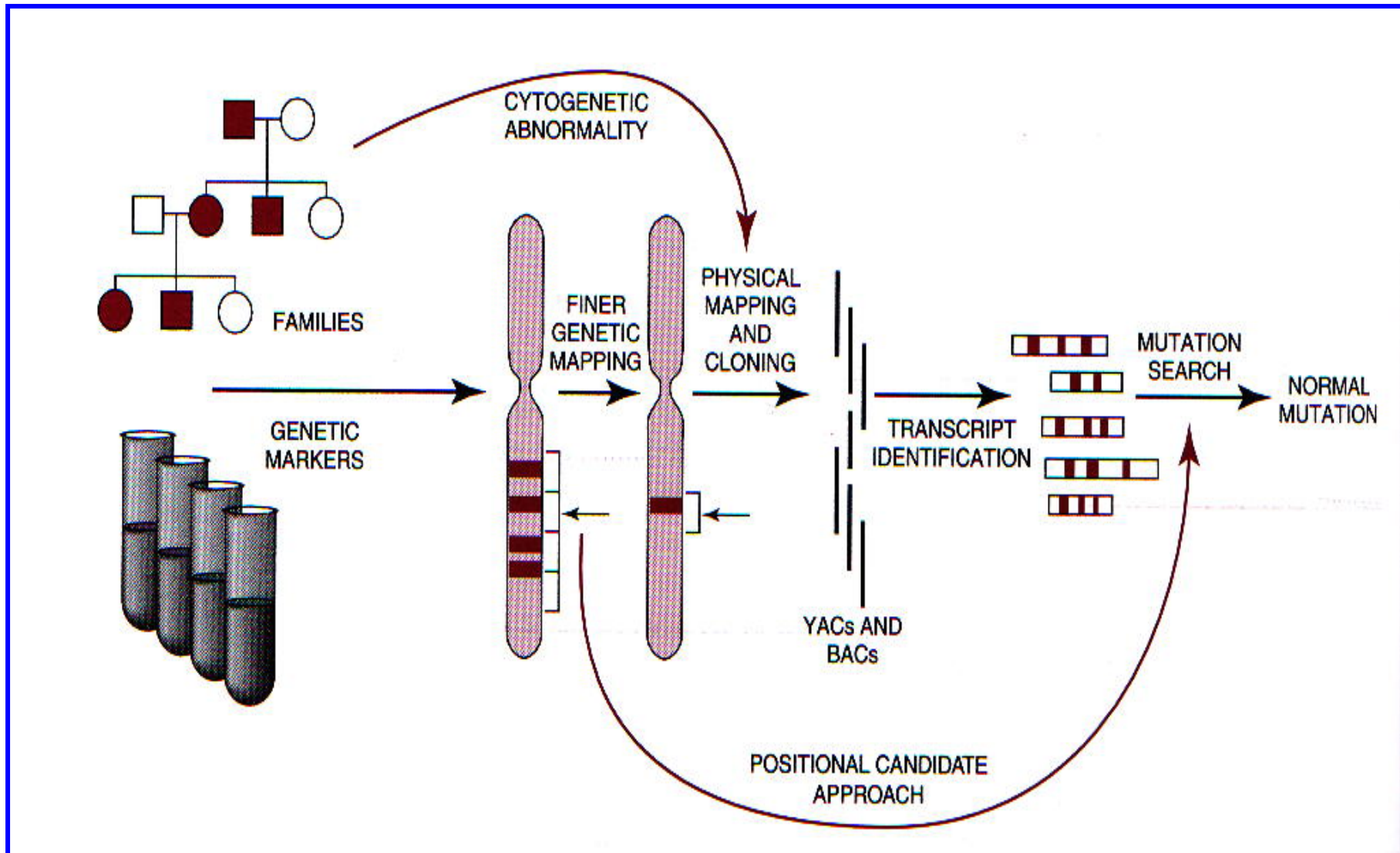


ANALISI DI LINKAGE (CONCATENAZIONE)

- **L'analisi di linkage** permette di determinare la posizione cromosomica di un locus responsabile di una determinata malattia/carattere genetico rispetto a **marcatori polimorfici** la cui localizzazione è nota
- L'analisi di linkage è un approccio molto utile per il mappaggio e l'identificazione di geni responsabili di **malattie genetiche Mendeliane**
(oltre 1,200 geni identificati)
- Più difficoltosa è l'identificazione dei numerosi geni implicati in malattie più comuni, ad ereditarietà **complessa**
(pochissimi geni identificati finora)

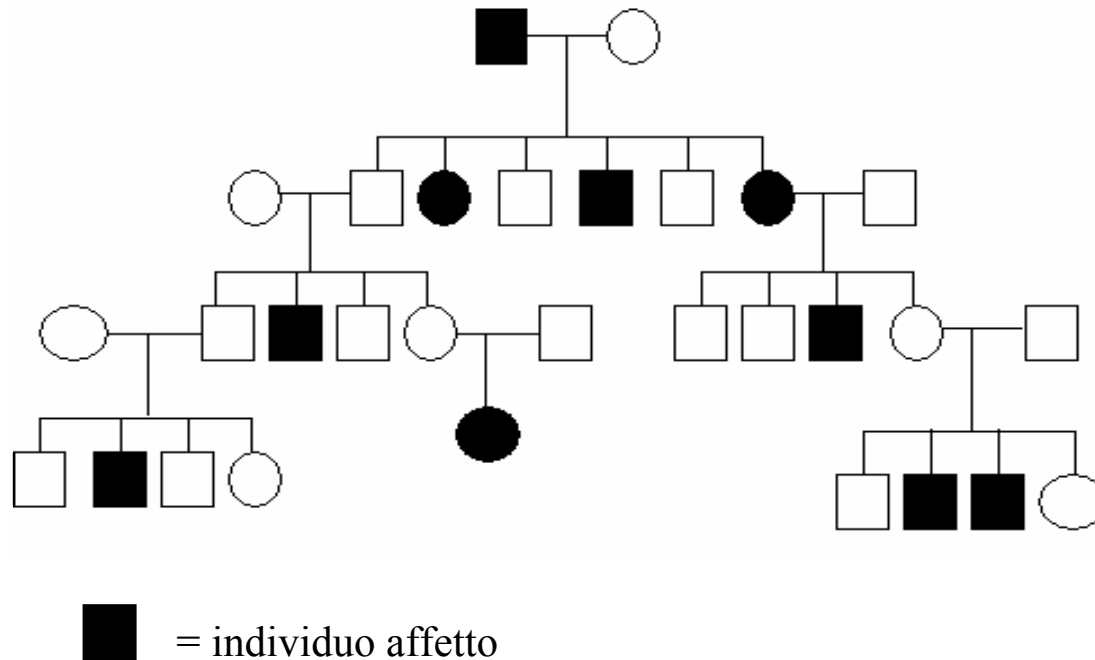
POSITIONAL CLONING



REQUISITI

per l'ANALISI DI LINKAGE DI CARATTERI MENDELIANI

1) Una o più **famiglie** in cui segrega il carattere/malattia in questione

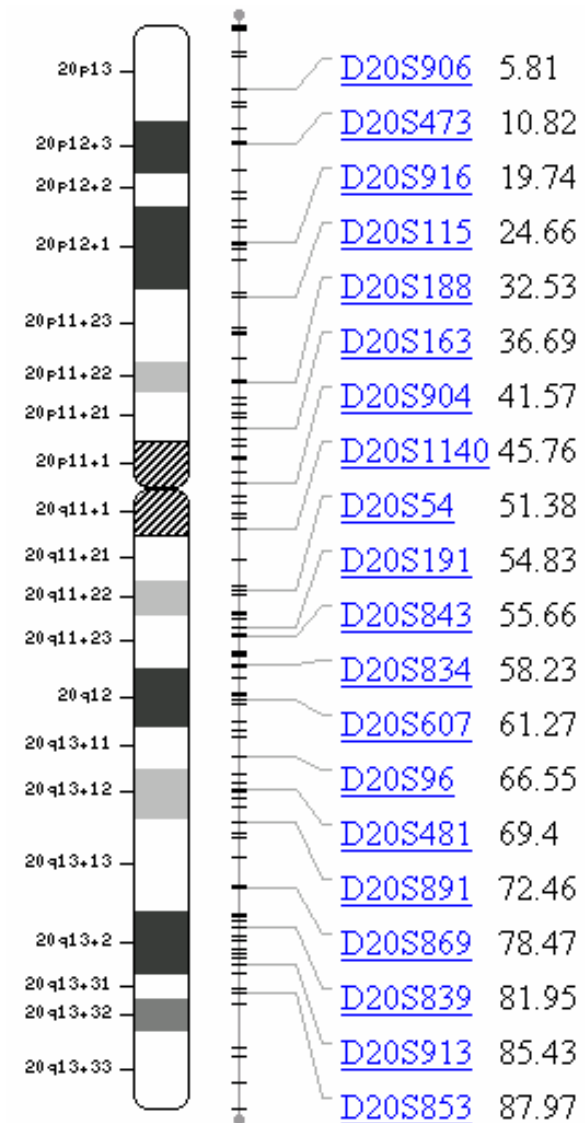


REQUISITI

per l'ANALISI DI LINKAGE DI CARATTERI MENDELIANI

2) MARCATORI GENETICI

- Polimorfici -> presenza di 2 o + alleli alternativi
 - Allele più raro con frequenza di almeno 1% nella popolaz.
 - Facilmente tipizzabili e stabili di generazione in generazione
 - Posizione nota nel genoma
- Gli individui delle famiglie vengono tipizzati per un set di marcatori polimorfici distribuiti su tutto il genoma o in particolari regioni candidate
- Per tutti i cromosomi sono state create **mappe genetiche** di marcatori polimorfici



MARCATORI MOLECOLARI (DEL DNA)

RFLPs (Restriction Fragment Length Polymorphisms)

- Presenza/assenza sito di taglio di enzima di restrizione
- Biallelici
- Southern blot / PCR

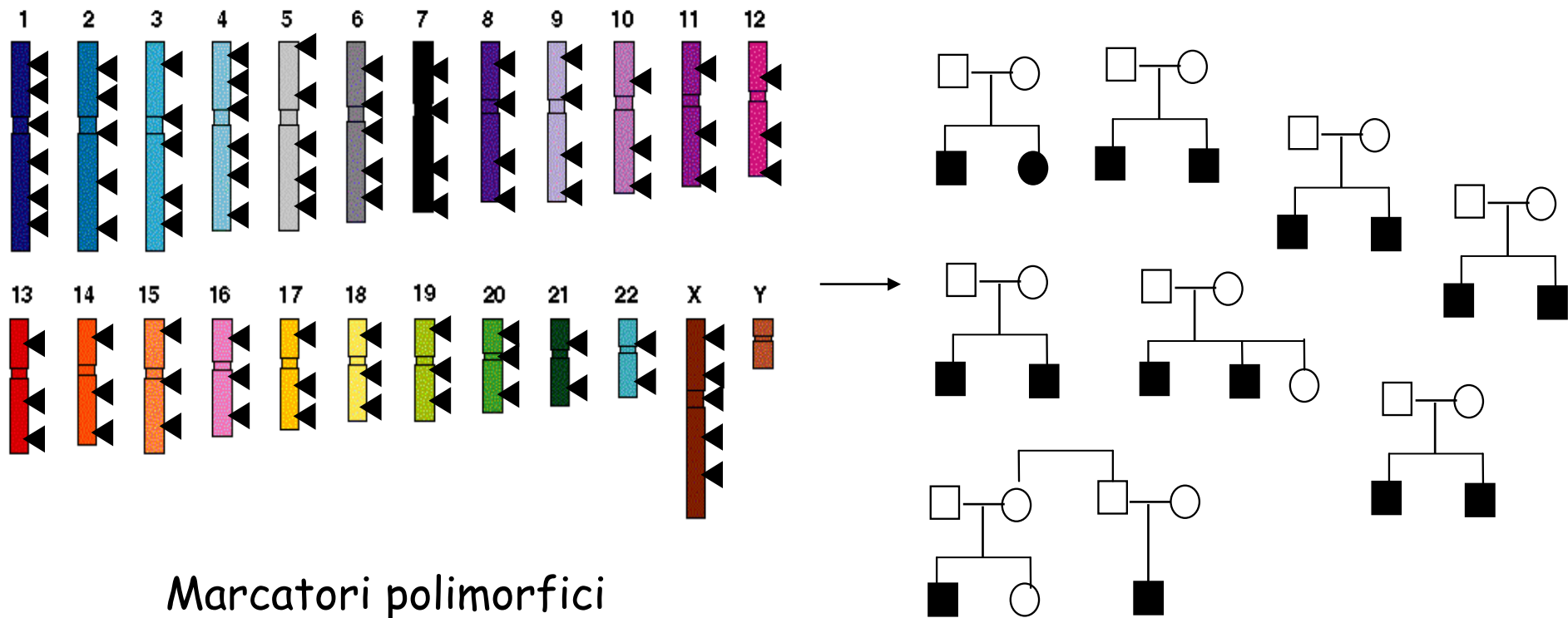
MICROSATELLITI

- ripetiz in tandem di 2-3-4 nucleotidi (CA)_n
- Molti alleli, molto informativi
- Distribuiti in modo uniforme nel genoma. $\approx$ ogni 100 Kb
- PCR / marcatura con fluorescenza

SNPs (Single Nucleotide Polymorphisms)

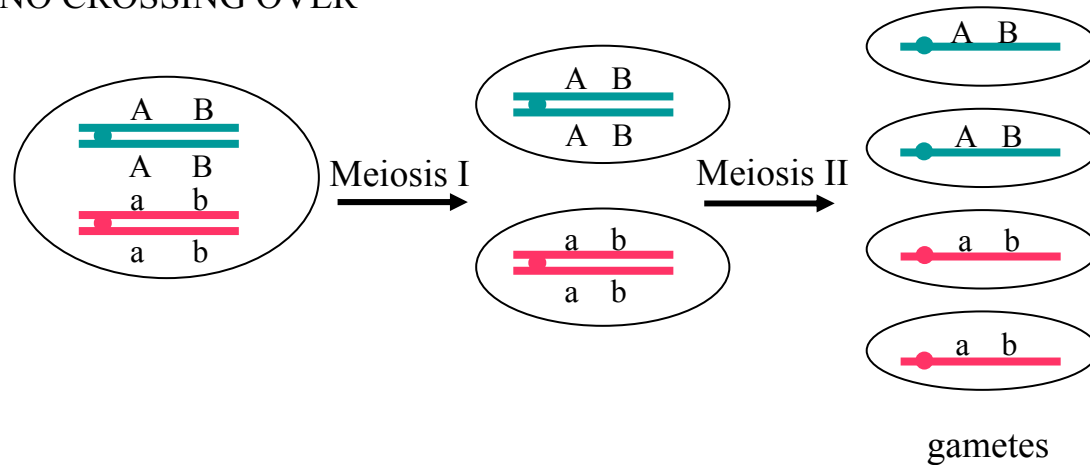
- Differenze di singola base, non necessariamente riconosciuti da enz. di restriz.
- Biallelici
- Molti frequenti,
- Sviluppo di tecniche di genotyping automatizzate e in larga scala

GENOME WIDE SCAN

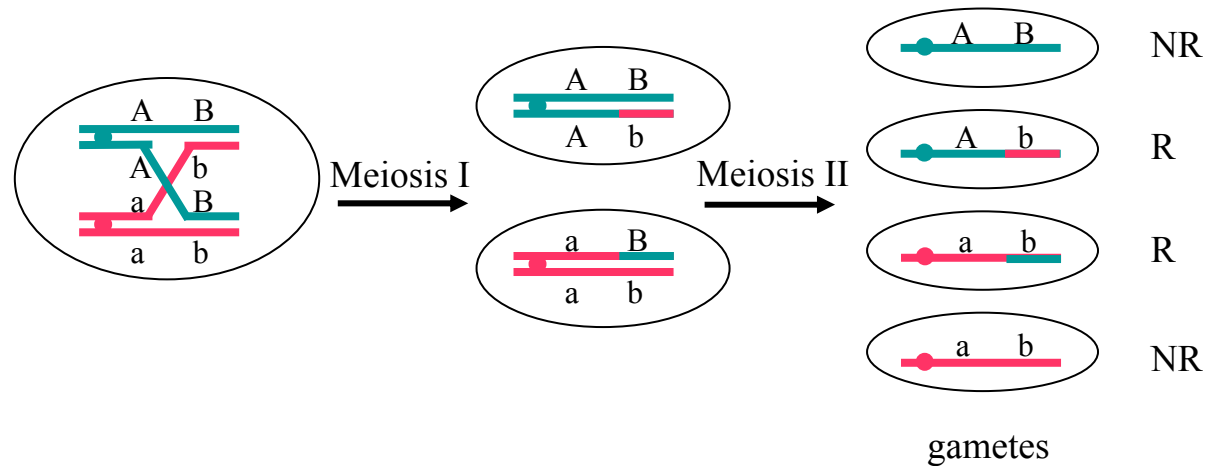


Marcatori polimorfici
distribuiti su tutto il
genoma

NO CROSSING OVER

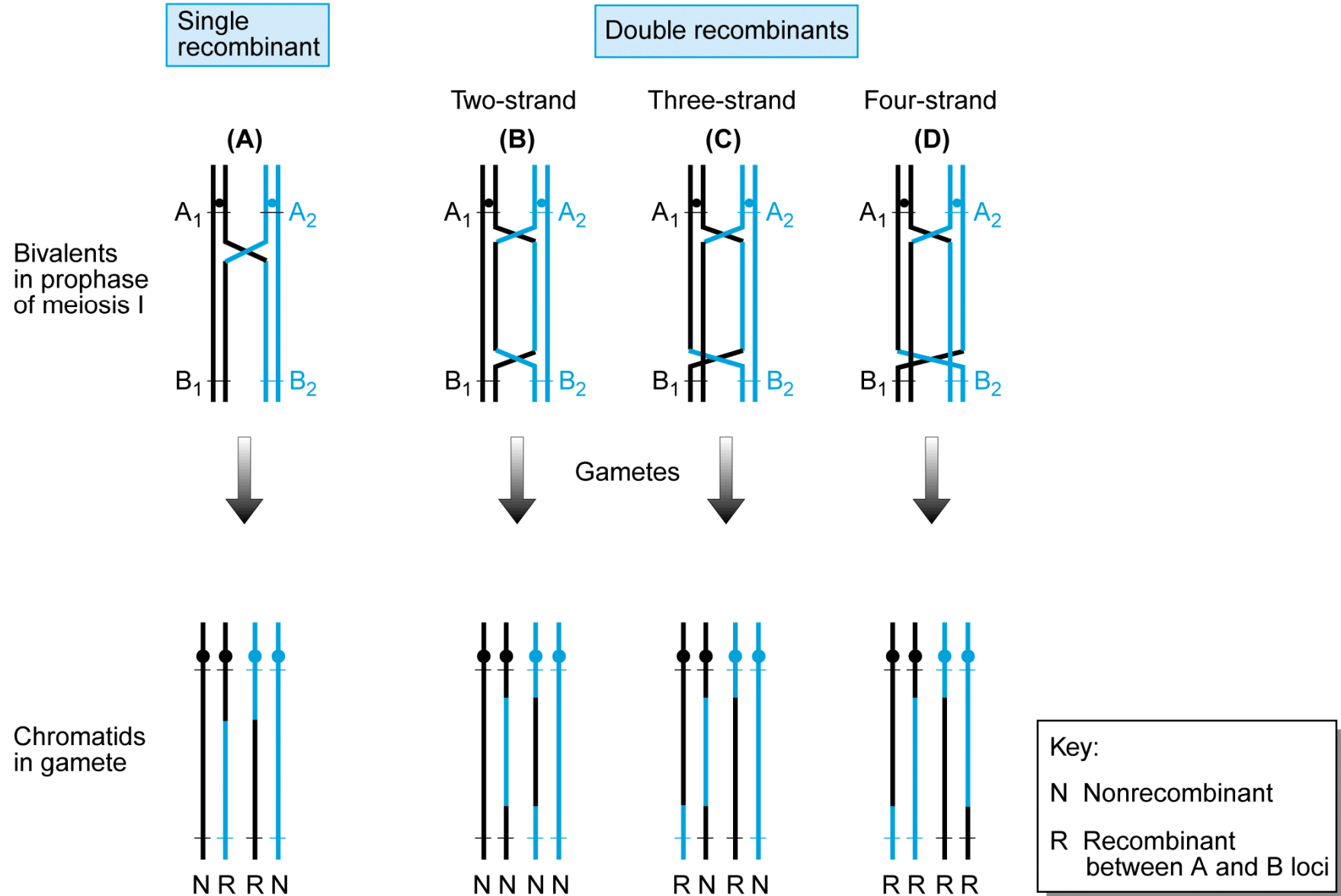


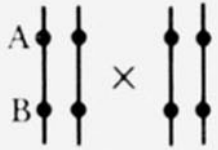
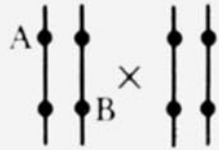
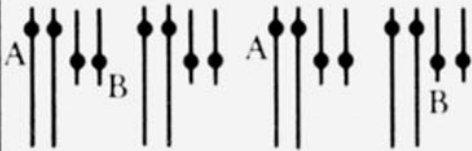
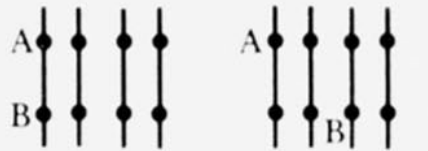
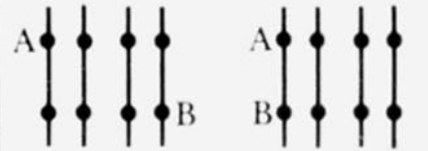
CROSSING OVER

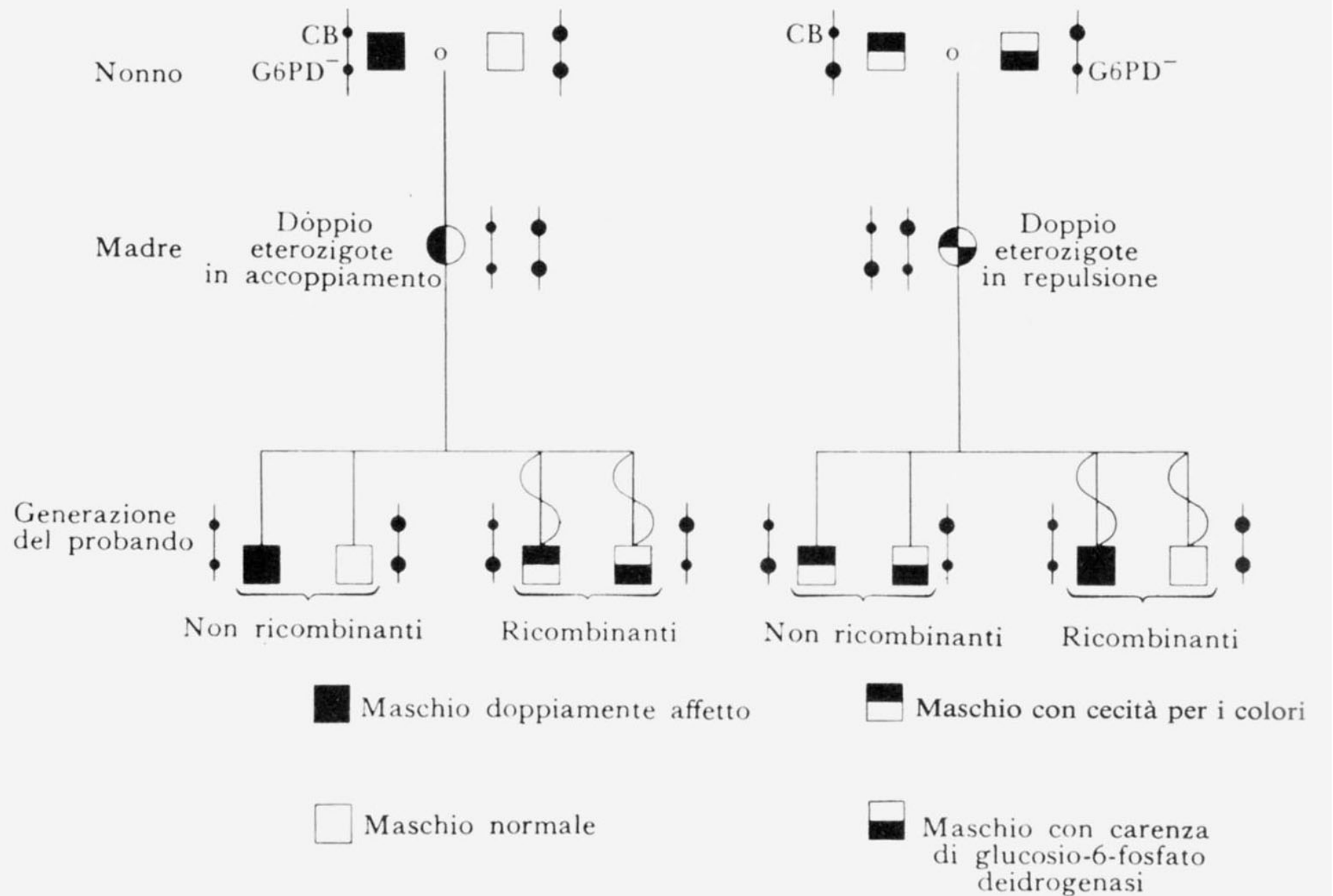


L'analisi di linkage si basa sulla co-segregazione alla meiosi di 2 o più loci più frequentemente di quanto ci si aspetta per caso. Se 2 loci vengono ereditati insieme frequentemente, è probabile che siano localizzati vicini fra di loro sul cromosoma

La frazione di ricombinazione non può mai superare il 50%



Alleli	Non-alleli		
	Su cromosomi differenti oppure lontani sullo stesso cromosoma	Concatenati, ossia vicini su uno stesso cromosoma In accoppiamento (cis)	In repulsione (trans)
Genitori  ↓	 ↓ Ricombinanti	 ↓ Ricombinanti	 ↓ Ricombinanti
Figli  1 : 1 : 1 : 1	 1 : 1 : 1 : 1 Assortimento indipendente	 1 : 1 : <1 : <1	 1 : 1 : <1 : <1
Gli alleli segregano	I non-alleli riassortiscono		



ANALISI DI LINKAGE DI CARATTERI MENDELIANI

- Determinare la frequenza con cui due loci ricombinano fra loro alla meiosi = **TETHA (θ)**
- se due loci sono su cromosomi diversi segregano indipendentemente. La probabilità che vengano ereditati insieme è del $1/2$ -> **$\theta = 50\%$**
 - Se due loci sono vicini fra loro sullo stesso cromosoma saranno ereditati insieme più frequentemente -> **$\theta < 50\%$**
Tanto più sono vicini, tanto più piccola è la probabilità che avvenga un crossing-over

LA FREQUENZA DI RICOMBINAZIONE (θ) E' UNA MISURA DELLA DISTANZA GENETICA

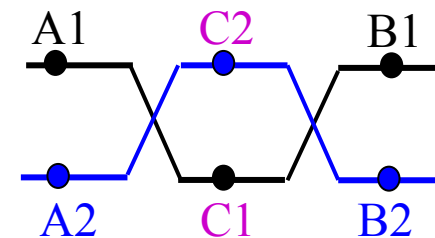
DISTANZA GENETICA =

numero atteso di crossing-overs fra 2 loci per meiosi

- L'unità di misura della distanza genetica è il **centiMorgan (cM)**

1 Morgan = 100 cM

- Le frequenze di ricombinazione fra loci piuttosto distanti non sono additive (a causa dell'effetto dei crossing-over multipli)



- Per $\theta < 10\%$ c'è una corrispondenza $\sim 1:1$ fra θ e cM :
 $1 \text{ cM} \approx \theta = 1\%$
- Per $\theta > 10\%$ si usano delle **funzioni di mappa** per convertire θ in cM e viceversa

Funzioni di mappa -> relazione fra frazione di ricombinazione e distanza genetica

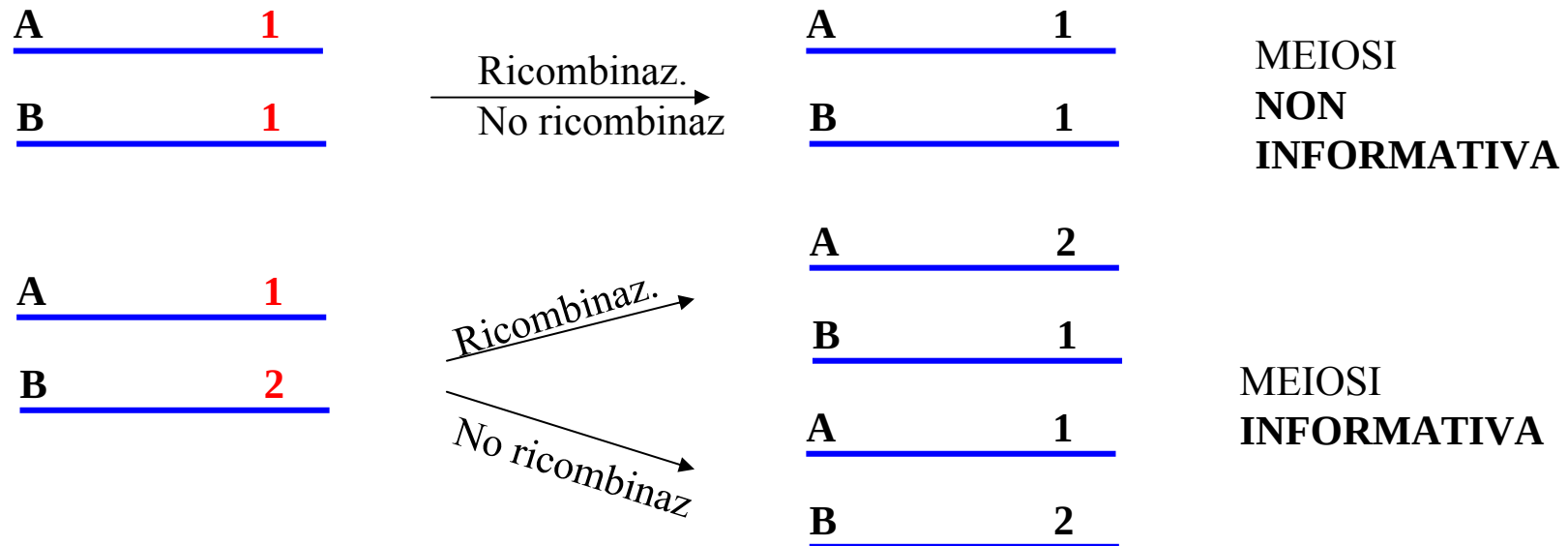
Haldane:

$$\omega = -1/2 \ln(1-2\theta)$$

$$\theta = 1/2 (1 - e^{-2\omega})$$

INFORMATIVITA'

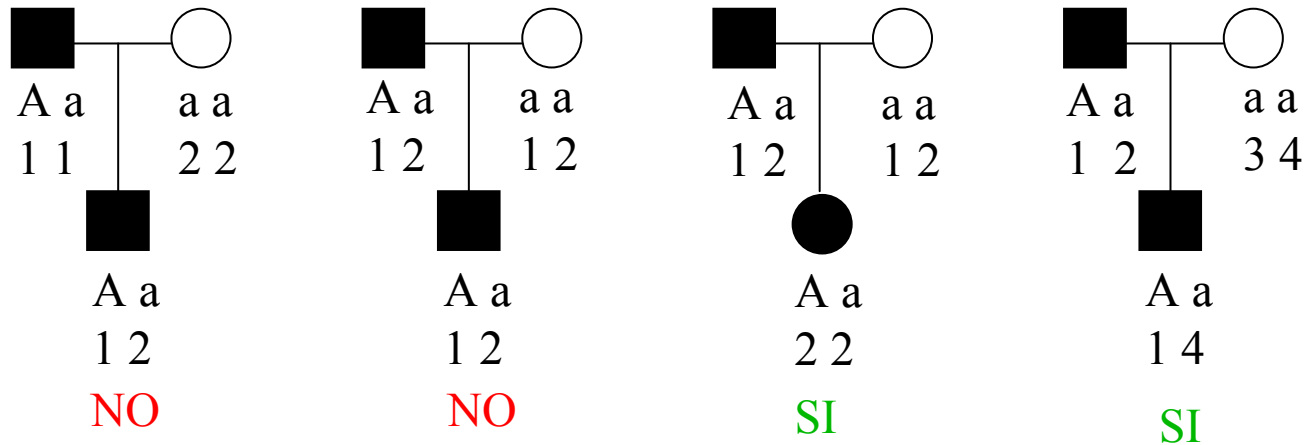
Per misurare la frequenza di ricombinazione fra 2 marcatori e' necessario che gli eventi meiotici siano informativi, cioe' che i marcatori parentali siano **doppi eterozigoti**



FASE = Disposizione degli alleli di due loci adiacenti sullo stesso cromosoma

Una serie di alleli per loci adiacenti che vengono ereditati insieme sullo stesso cromosoma formano un **aplotipo**

MEIOSI INFORMATIVE ?



Eterozigosità (H)

Probabilità che un individuo preso a caso sia eterozigote

$$H = 1 - (p_1^2 + p_2^2 + p_3^2 + p_4^2 + \dots p_i^2)$$

ANALISI DI LINKAGE

MARCATORE - LOCUS MALATTIA

✓ Definizione di un **modello di ereditarietà** del locus-malattia
(es. locus dominante, recessivo, autosomico, cromosoma X...)

Esempio:

Singolo locus **D** = allele “mutato” **d** = allele wild-type

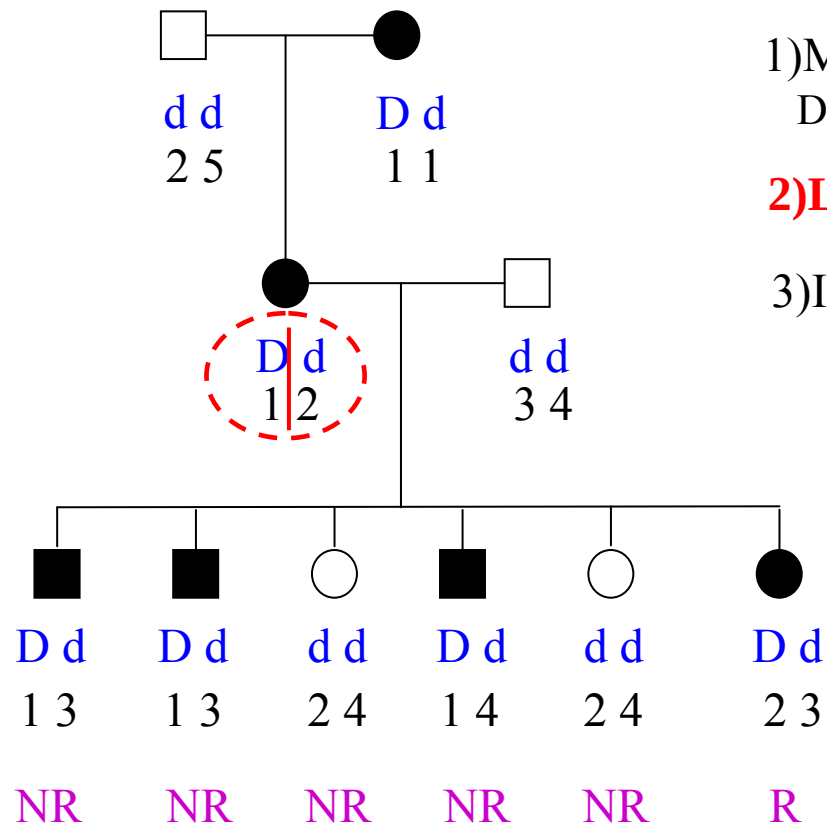
Frequenze alleliche: $P(D) = p$, $P(d) = 1 - p = q$

Penetranze: f_{DD} , f_{Dd} , f_{dd}

✓ Stima di **θ** fra il locus malattia ed il marcatore

a) Conteggio diretto dei ricombinanti $\theta = \frac{\mathbf{R}}{(\mathbf{R} + \mathbf{NR})}$

ESEMPIO CONTEGGIO RICOMBINANTI: Pedigree a fase nota



1)Modello di ereditarietà:
Deduco i genotipi per locus malattia

2)LA FASE E' NOTA

3)Individuo i **ricombinanti**

✓ Stima di θ fra il locus malattia ed il marcatore

b) Metodo del **LOD SCORE**

Likelihood ratio test: θ è significativamente < 0.5 ?

H_0 = i due loci sono indipendenti $\rightarrow \theta = 0.5$

H_1 = i due loci sono in linkage $\rightarrow 0 < \theta < 0.5$

Likelihood [$L(\theta)$] = Probabilità che i dati osservati si verifichino quando la fraz di ricombinaz è θ

LOD = log of the odds **$Z(\theta) = \log_{10} [L(\theta)/L(\theta = 0.5)]$**

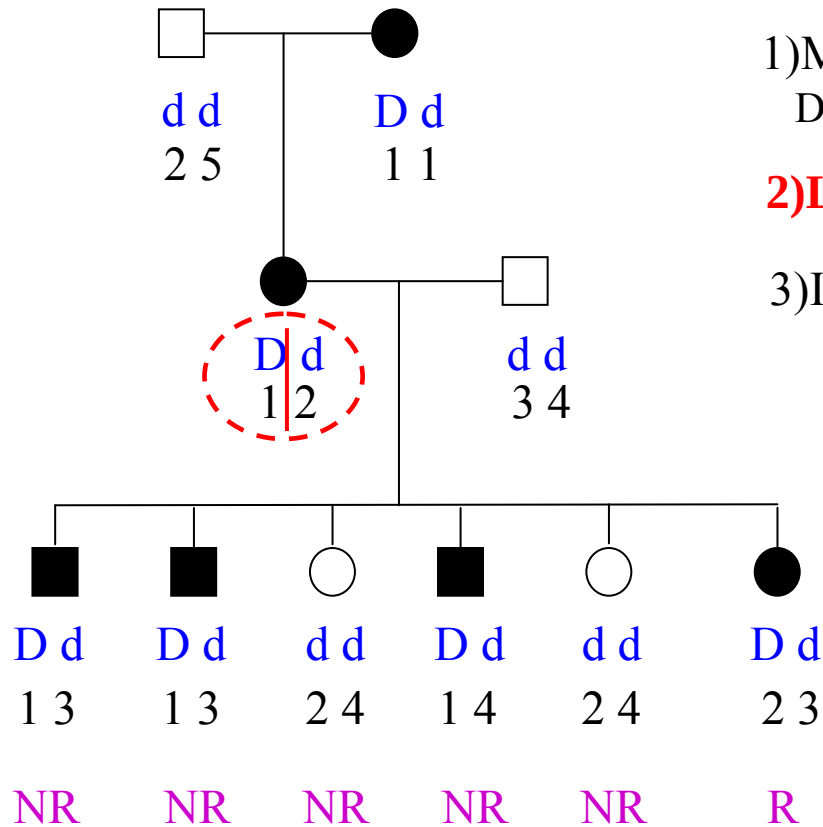
Limiti di significatività

$Z > 3$ evidenza di linkage è significativa

$Z < -2$ si può escludere la presenza di linkage

ESEMPIO CONTEGGIO RICOMBINANTI E CALCOLO DEL LOD SCORE:

1) Pedigree a fase nota



1) Modello di ereditarietà:

Deduco i genotipi per locus malattia

2) LA FASE E' NOTA

3) Individuo i **ricombinanti**

LOD SCORE (LOD)

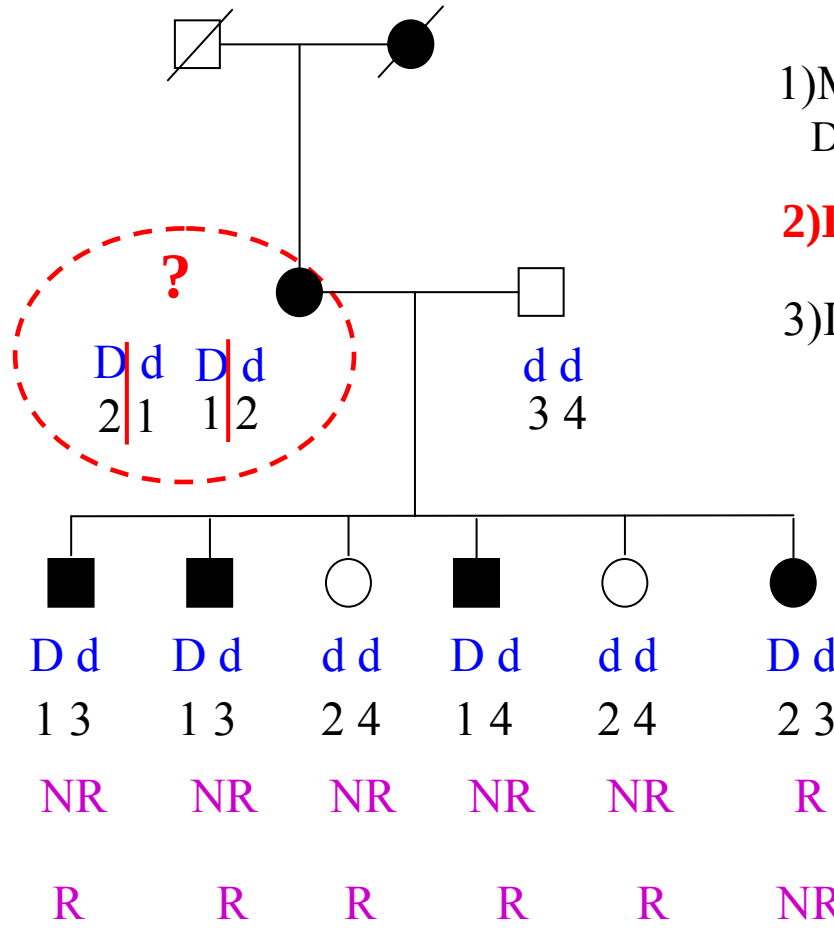
$$Z = \log_{10} \left[\frac{\theta^R (1-\theta)^{NR}}{0.5^{(R+NR)}} \right]$$

Per $\theta = 0$

- No ricomb $Z = \log_{10} \frac{1}{0.5^N}$
- ≥ 1 ricomb. $Z = -\infty$

ESEMPIO CONTEGGIO RICOMBINANTI E CALCOLO DEL LOD SCORE:

2) Pedigree a fase non nota



1) Modello di ereditarietà:

Deduco i genotipi per locus malattia

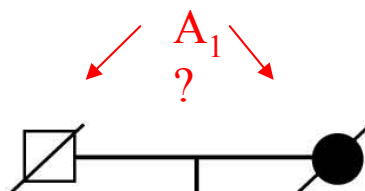
2) LA FASE NON E' NOTA

3) Individuo i **ricombinanti**

LOD SCORE (LOD)

$$Z = \log_{10} \left[\frac{1}{2} \left(\frac{\theta^R (1-\theta)^{NR}}{0.5^{R+NR}} \right) + \frac{1}{2} \left(\frac{\theta^{NR} (1-\theta)^R}{0.5^{R+NR}} \right) \right]$$

I

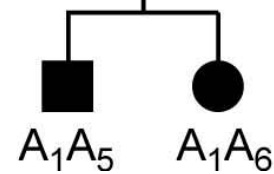
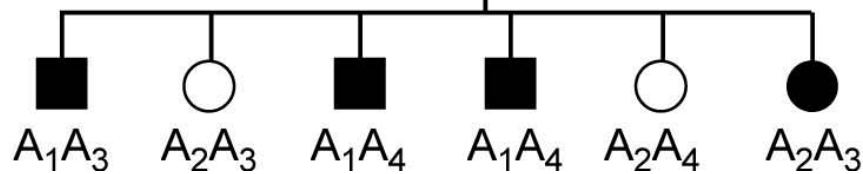


II

Fase non
nota



III



Spesso i dati ricavati da una sola famiglia non sono sufficienti per stabilire presenza /assenza di linkage.

I lod score ottenuti da famiglie indipendenti (per lo stesso valore di θ) si possono [sommare](#) fra loro

Si calcola il lod score per diversi valori di θ ($\theta = 0, 0.1, 0.2, \dots, 0.5$) ->
Tabella

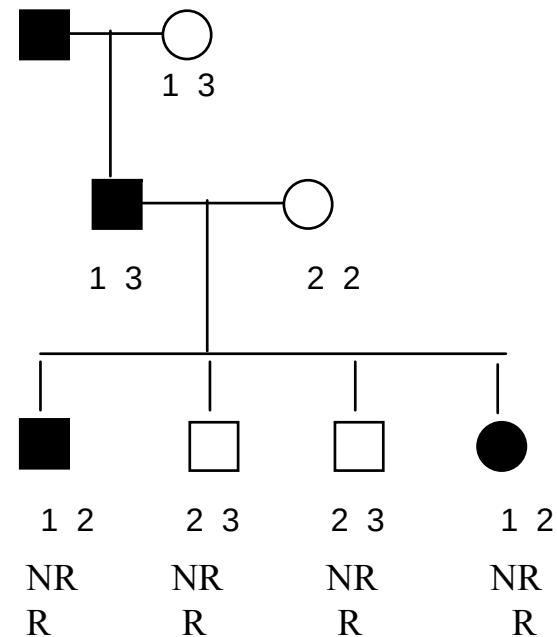
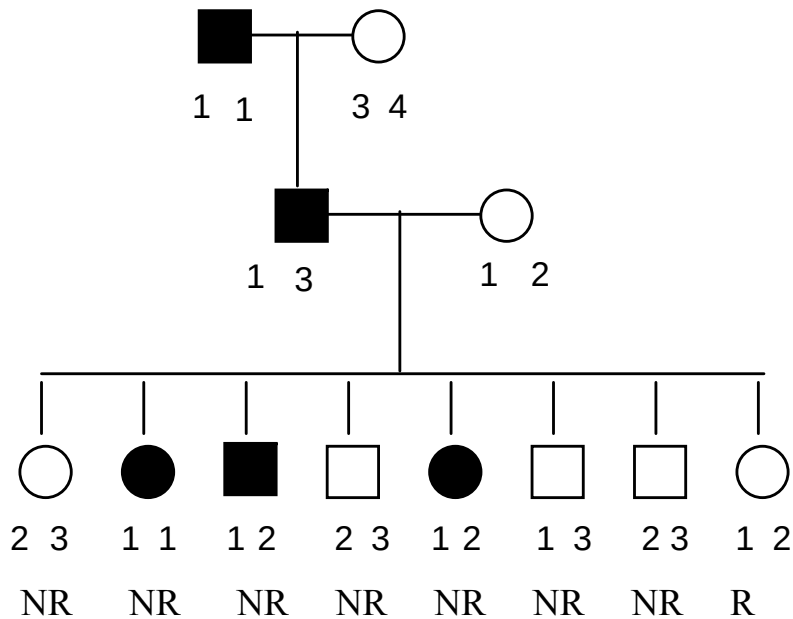
tetha	0.05	0.1	0.2	0.3	0.4	R:NR
Fam 1	0.12	0.32	0.42	0.36	0.22	0:4
Fam2	-0.16	0.07	0.21	0.22	0.14	1:3
Fam3	2.23	2.04	1.63	1.17	0.63	0:8
Totale	2.19	2.43	2.26	1.75	0.99	

Massimo Lod score (MLS)

Valore max di lod score che si ottiene al variare di un parametro (generalmente θ)

Il valore di θ per cui il LOD è massimo è la stima più probabile della frazione di ricombinazione

ESEMPIO DI CALCOLO LOD SCORE



$$Z = \log_{10} \frac{\theta (1 - \theta)^7}{0.5^8}$$

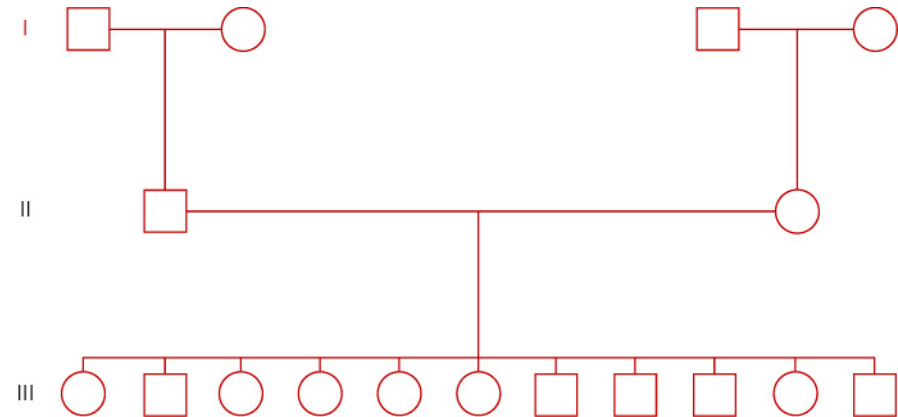
$$Z = \log_{10} \left[\frac{1}{2} \frac{(1 - \theta)^4}{0.5^4} + \frac{1}{2} \frac{\theta^4}{0.5^4} \right]$$

	Frazione di ricombinazione				
Famiglia	0.05	0.1	0.2	0.3	0.4
1	0.95	1.09	1.03	0.80	0.46
2	0.81	0.72	0.52	0.3	0.09
Totale	1.76	1.81	1.55	1.1	0.55

Costruzione di mappe genetiche di riferimento

FAMIGLIE CEPH

Costituiscono un pannello di 40 famiglie di riferimento, selezionate per la loro struttura ideale per l'analisi di linkage (3 generazioni, con i 4 nonni, 2 genitori, e almeno 6 figli).



Mappe genetiche dell'intero genoma:

- Généthon <http://www.genethon.fr>
- Marshfield <http://research.marshfieldclinic.org/genetics/>
- deCODE <http://www.decode.com/>

Passo fondamentale per:

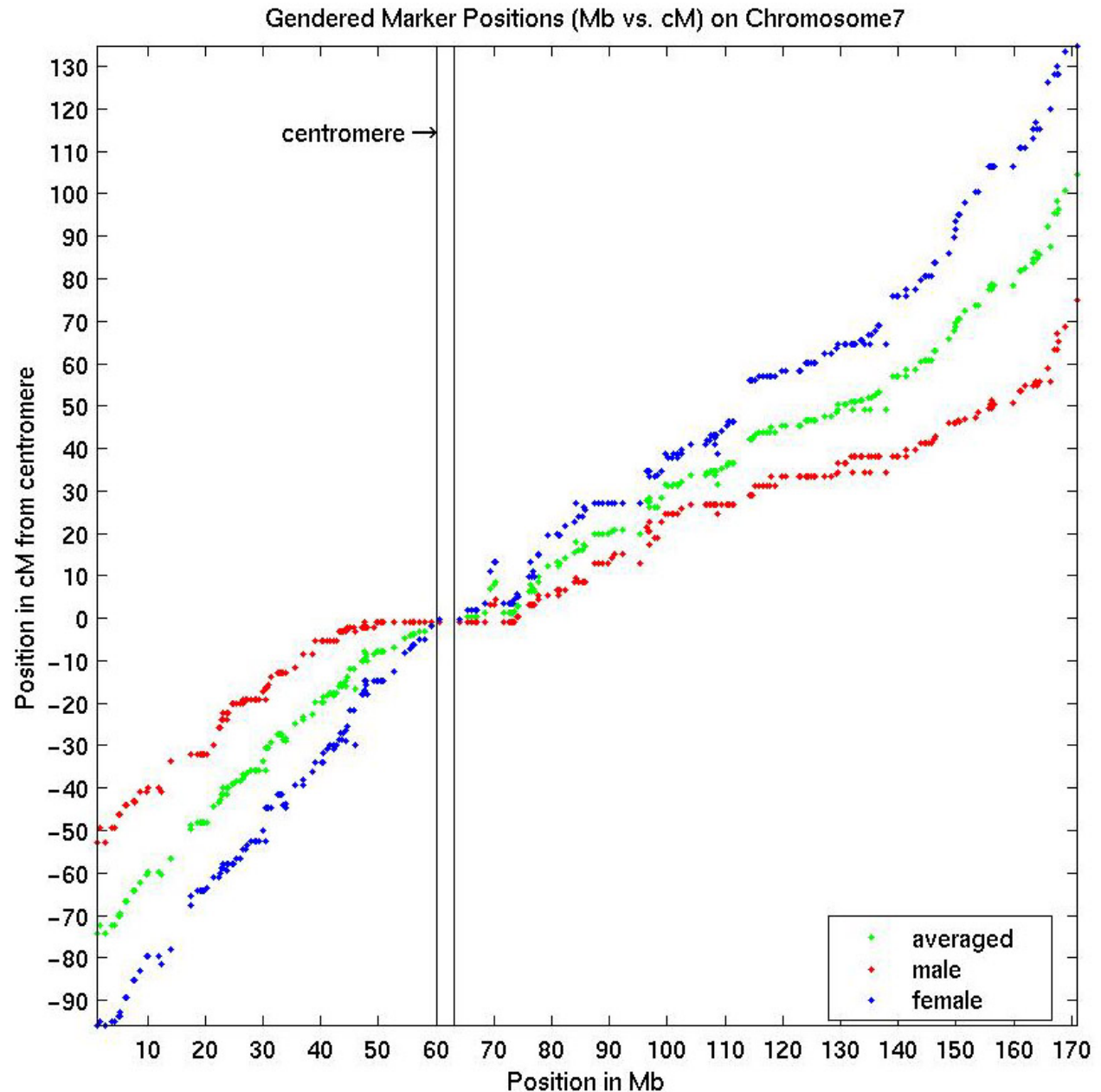
- Mappaggio di loci di malattie
- Hanno costituito una base per la costruzione e la verifica di mappe fisiche

Corrispondenza distanza genetica - fisica

Approssimativamente:
 $1 \text{ cM} \sim 1 \text{ Mb}$ ($1 \times 10^6 \text{ bp}$)

Non è costante

- Varia da regione a regione
($>$ verso i telomeri)
- Mappa femminile è più
lunga



FISIOKINESITERAPIA-NEWS.IT

TABELLA LOD SCORES A 2 PUNTI

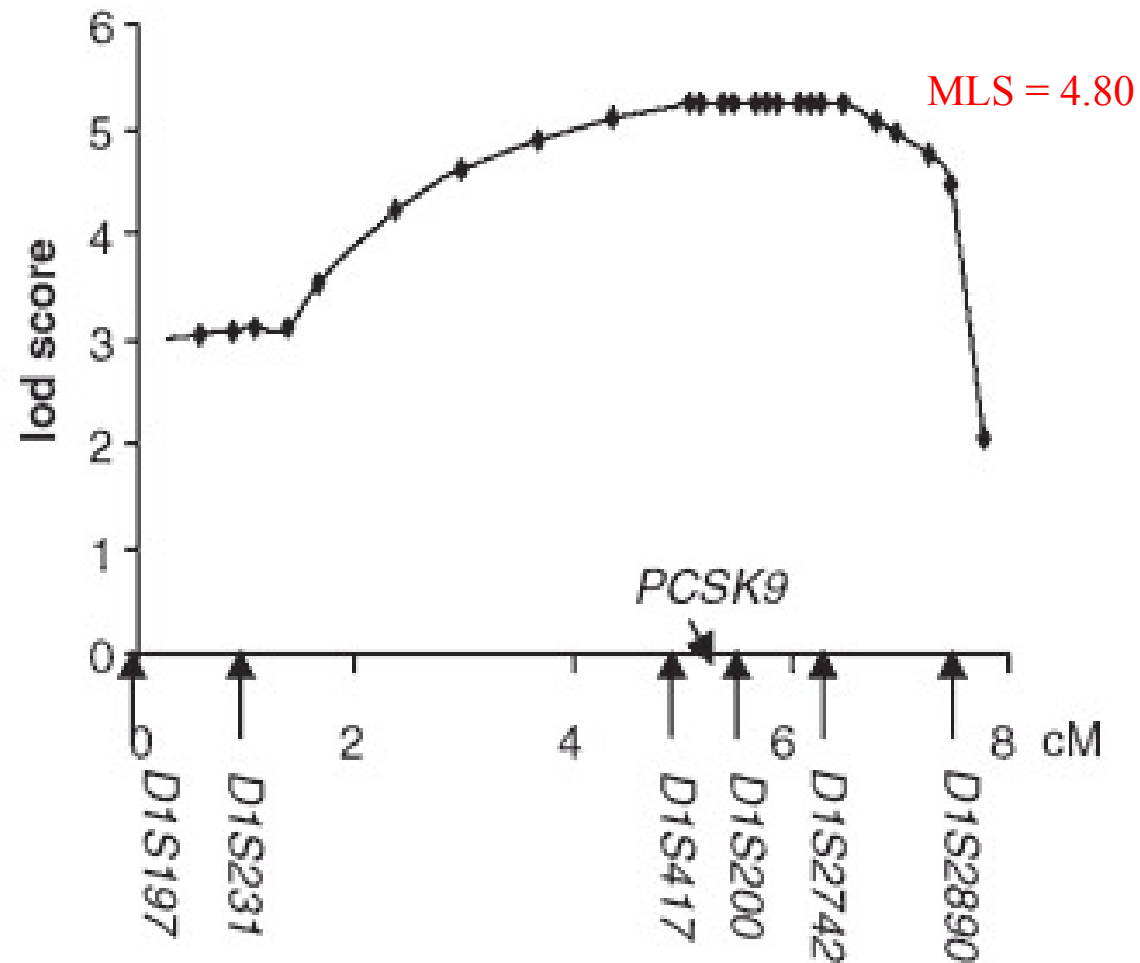
Locus	Distance ¹	LOD score at θ								$Z_{\max}$	$\theta_{\max}$
		0.00	0.001	0.01	0.05	0.10	0.20	0.30	0.40		
<i>DIS2722</i>	...	2.17	2.17	2.13	1.96	1.75	1.31	0.87	0.44	2.17	0.00
<i>DIS211</i>	0.025	-0.77	1.53	2.46	2.86	2.77	2.22	1.48	0.67	2.86	0.05
<i>DIS197</i>	0.063	-1.73	-1.72	-1.62	-1.20	-0.86	-0.48	-0.26	-0.11	-0.11	0.40
<i>DIS231</i>	0.014	1.08	1.08	1.08	1.06	1.00	0.81	0.57	0.30	1.08	0.00
<i>DIS417</i>	0.036	3.12	3.11	3.08	2.89	2.61	1.93	1.16	0.43	3.12	0.00
<i>DIS200</i>	0.007	2.50	2.49	2.45	2.25	1.98	1.44	0.87	0.33	2.50	0.00
<i>DIS2742</i>	0.007	4.26	4.26	4.21	3.97	3.61	2.76	1.79	0.77	4.26	0.00
<i>DIS2890</i>	0.013	0.84	1.29	2.08	2.54	2.52	2.05	1.35	0.58	2.54	0.05

¹ Distance between two adjacent markers in θ

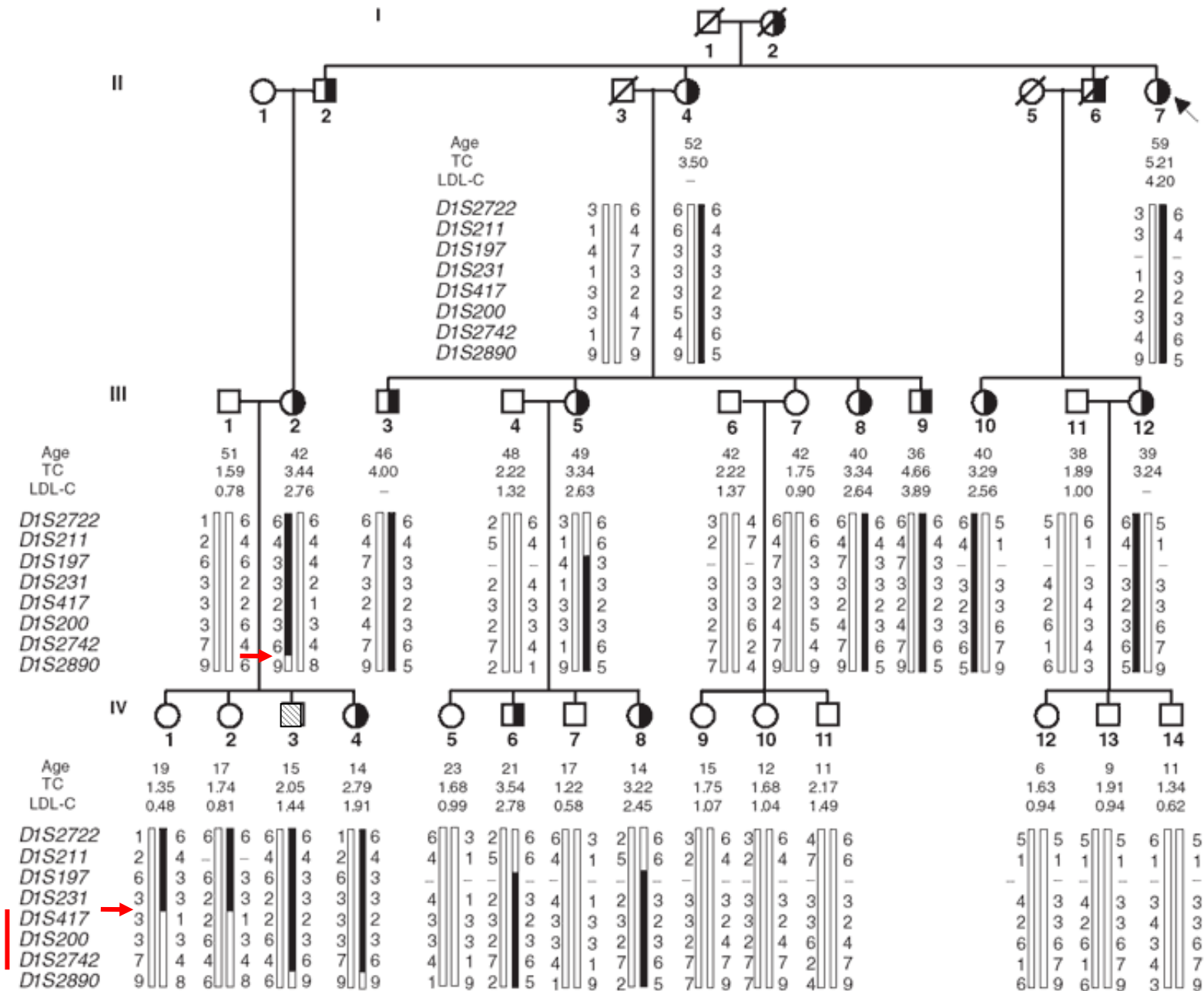
(Abifadel et al. Nat Genet 34:154, 2003)

MULTIPOINT LOD SCORE

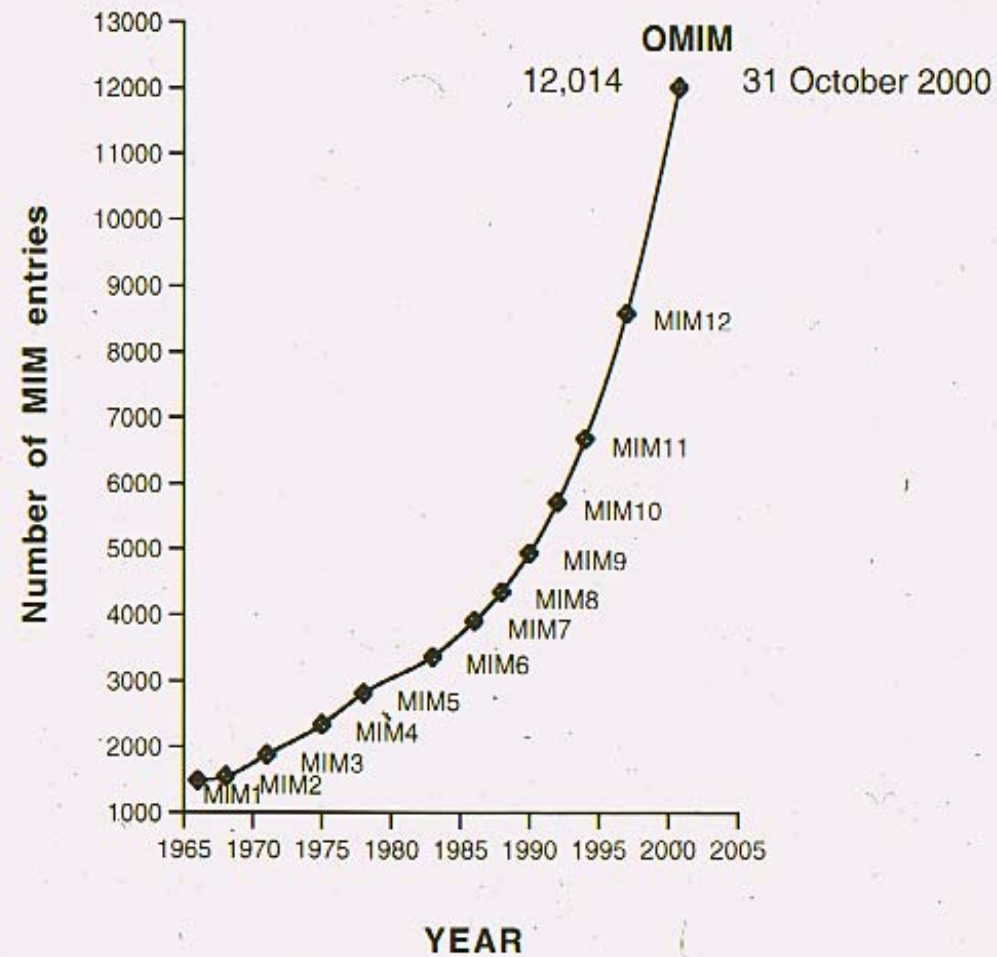
Data una **mappa di markers** con posizione nota, si calcola la likelihood per ogni posizione del locus malattia lungo il cromosoma. Permette di estrarre il massimo dell'informazione data da tutti i markers sul cromosoma.



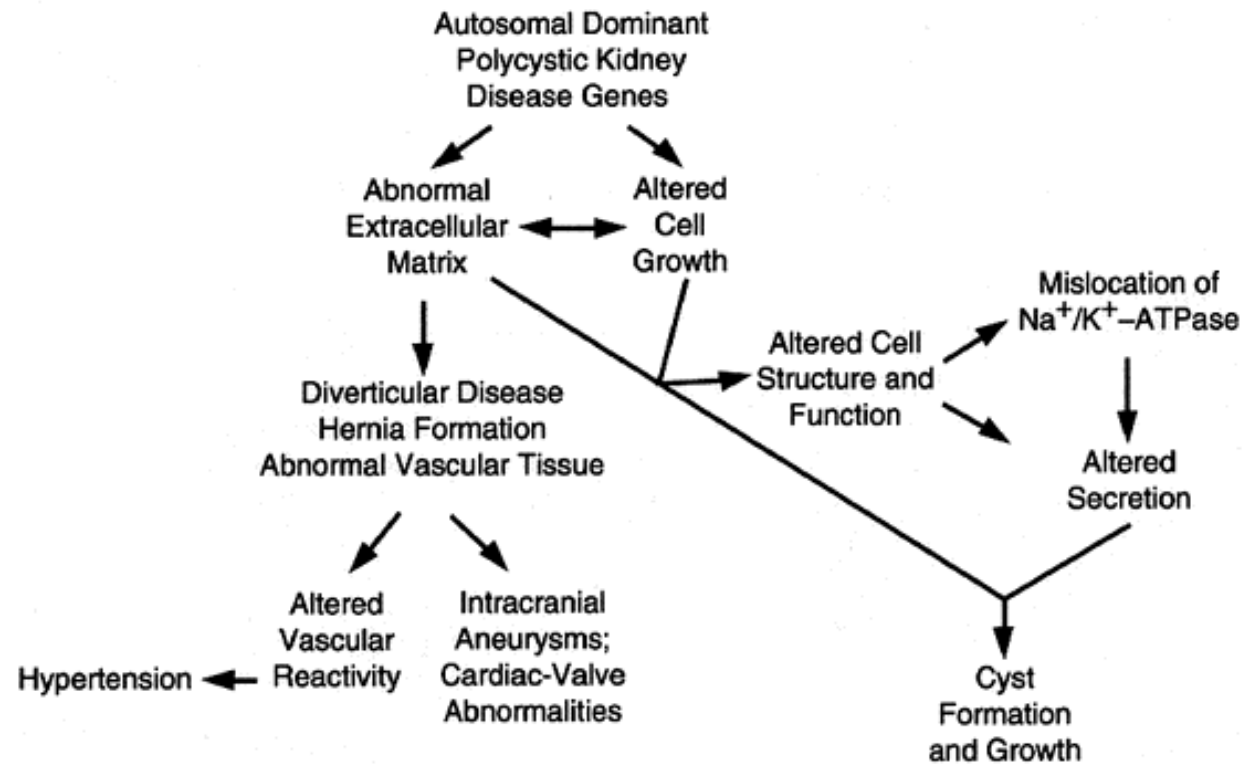
COSTRUZIONE DI APLOTIPI: definizione della regione critica

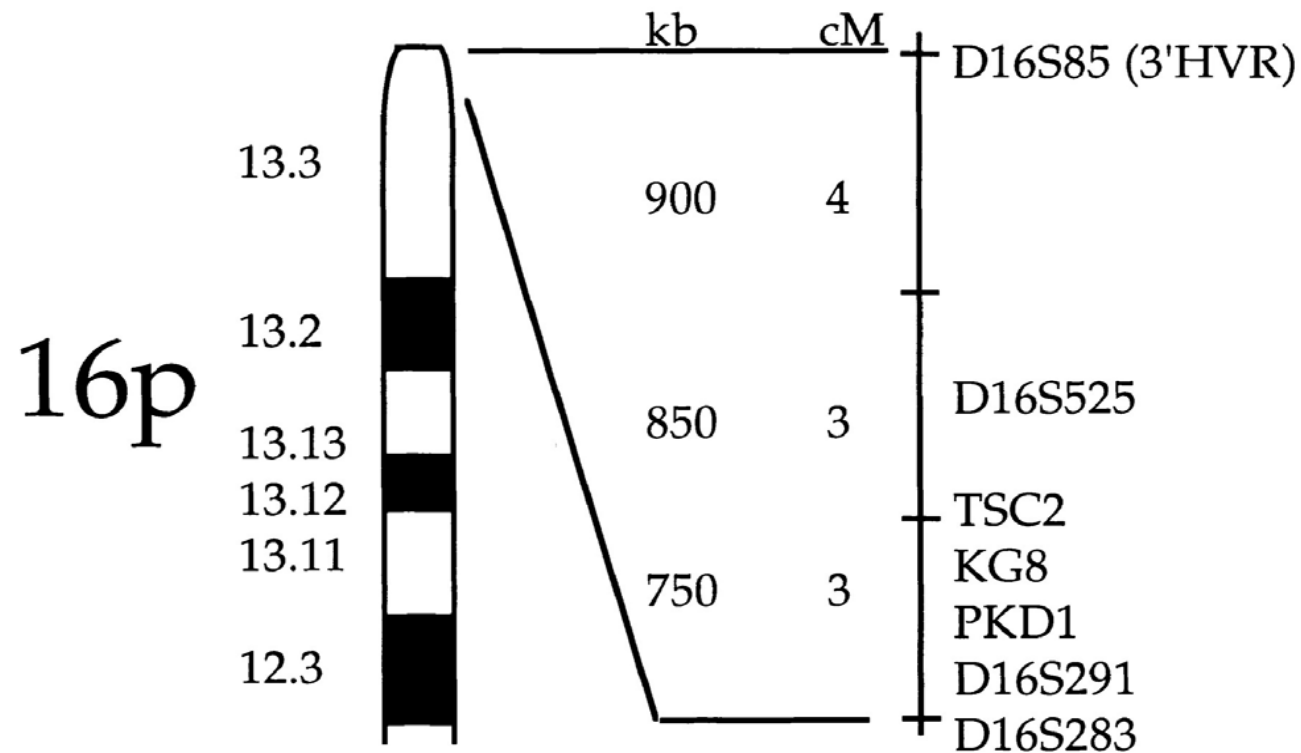


Number of Entries in *Mendelian Inheritance in Man*

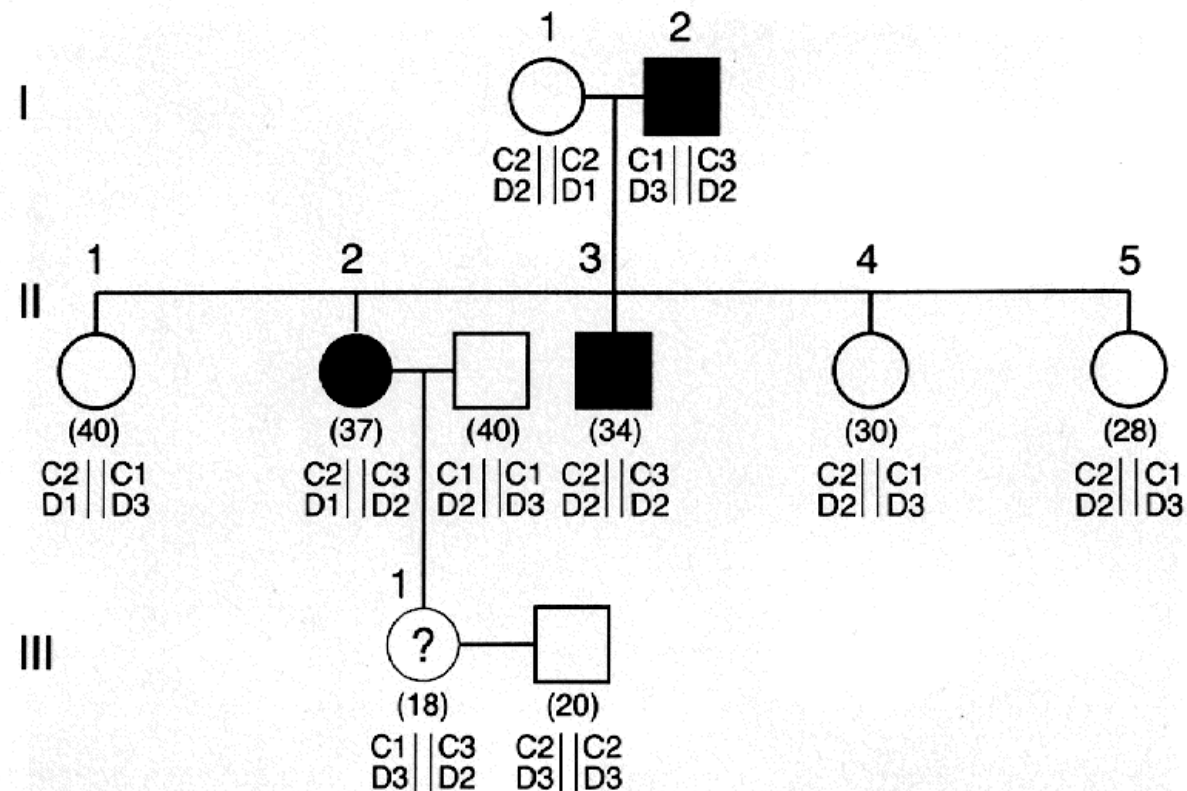


Possible Pathogenetic Mechanism of Autosomal Dominant Polycystic Kidney Disease



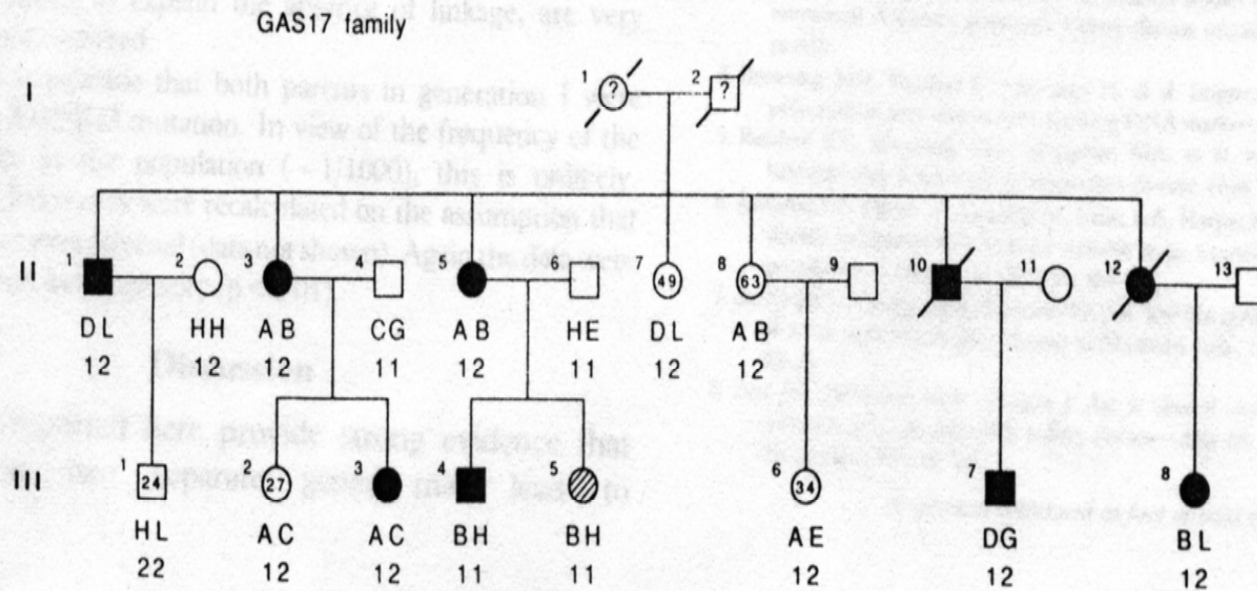


Map of chromosome band 16p13. The relative positions of the markers used in this study and the *PKD1* and *TSC2* genes ([4](#), [12](#)) are shown. Distance in centimorgans (cM) and kilobases (kb) is indicated.



Pedigree of a Kindred with ADPKD1.

Circles denote female family members, squares male family members, and solid symbols affected members. Ages are shown in parentheses. Subject III-1 wished to know her status before her marriage. Ultrasonography was negative. Two types of chromosome 16p markers were used (C and D). The three possible alleles are indicated (1, 2, and 3). Markers C and D are assumed to flank the ADPKD1 gene, each within 5 centimorgans of the ADPKD1 gene locus. These types are depicted below each symbol. In this family the haplotype C3,D2 is transmitted with the ADPKD1 gene. Therefore, Subject III-1 has an approximately 99 percent chance of being a gene carrier.



Family pedigree.

Filled symbols represent affected individuals. The hatched symbol indicates that the diagnosis is equivocal. Numbers within symbols give the age at which the subject last had a normal ultrasound examination. *A* to *L* are 3'HVR alleles. *1* and *2* are 090a alleles.