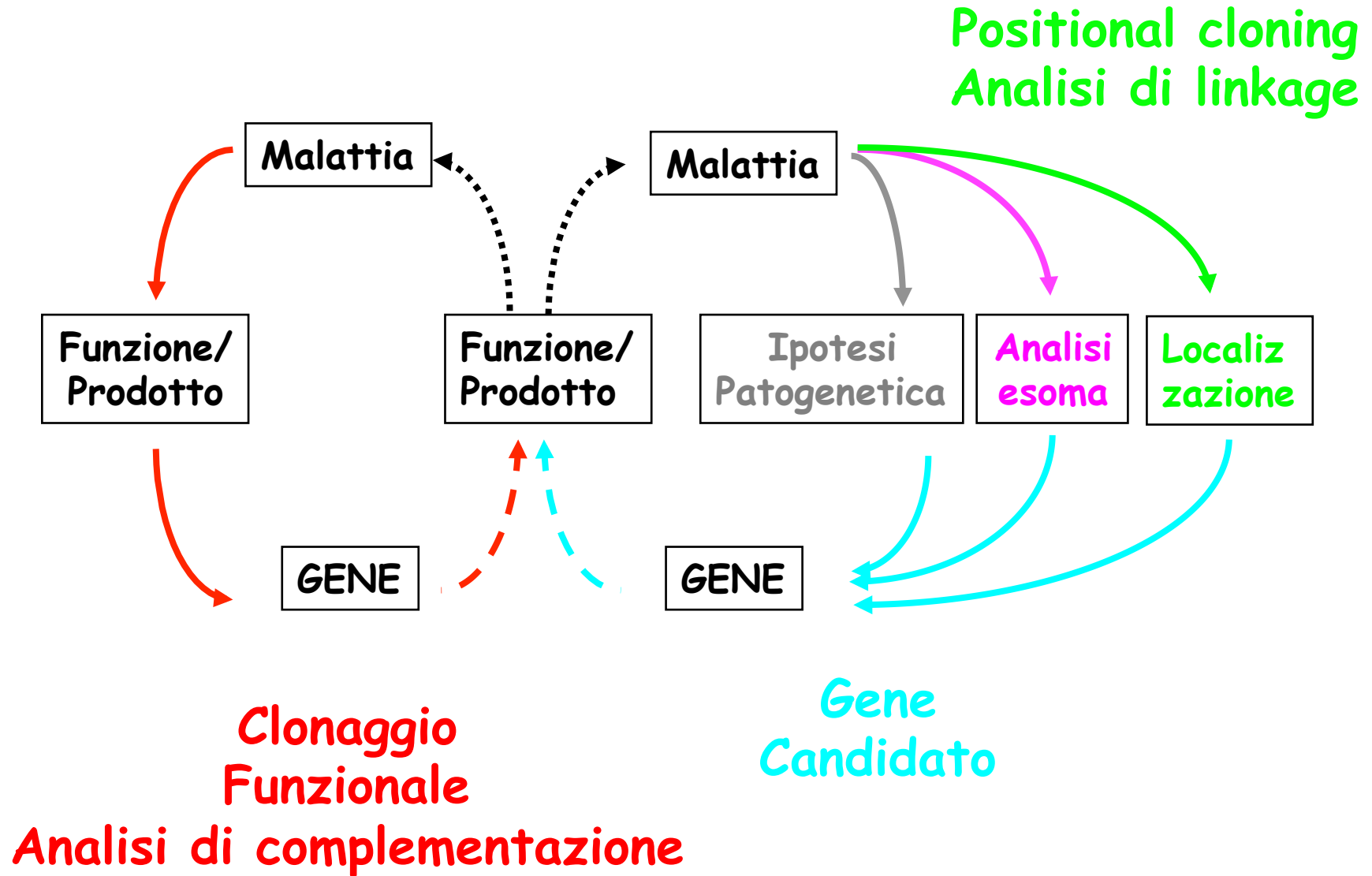
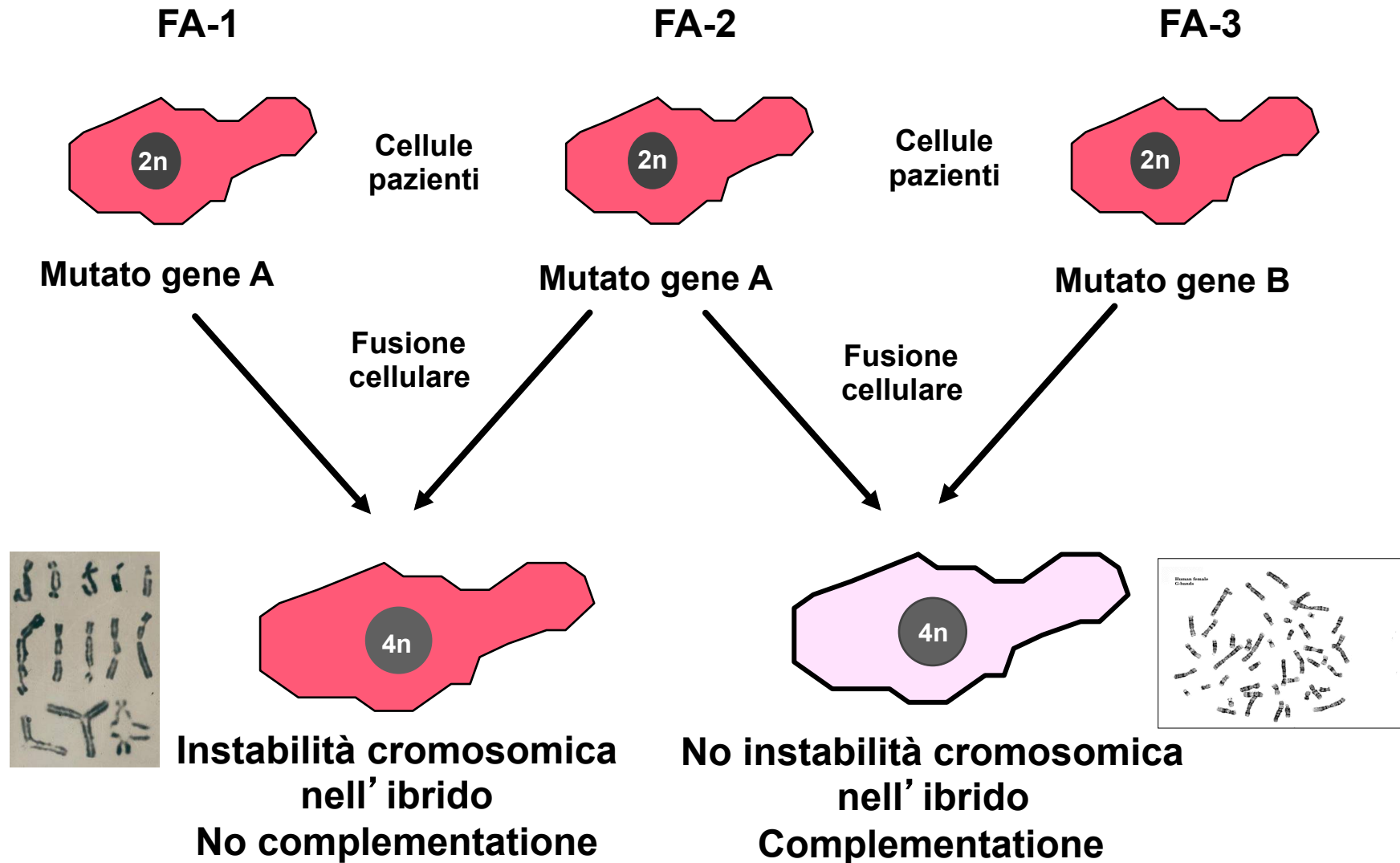


# STRATEGIE di CLONAGGIO

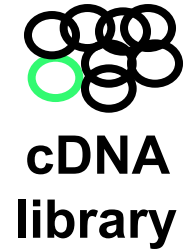


Analisi di complementazione:  
determina la presenza di ereditarietà genetica  
esempio anemia di Fanconi - instabilità cromosomica

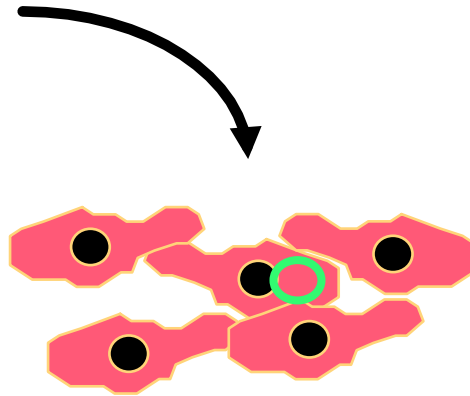


# Functional cloning

Using cDNA libraries

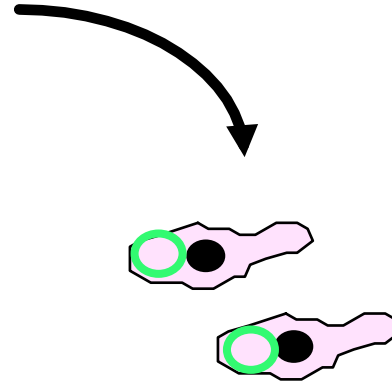


Transfection



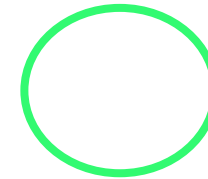
Mutant cells

Selection

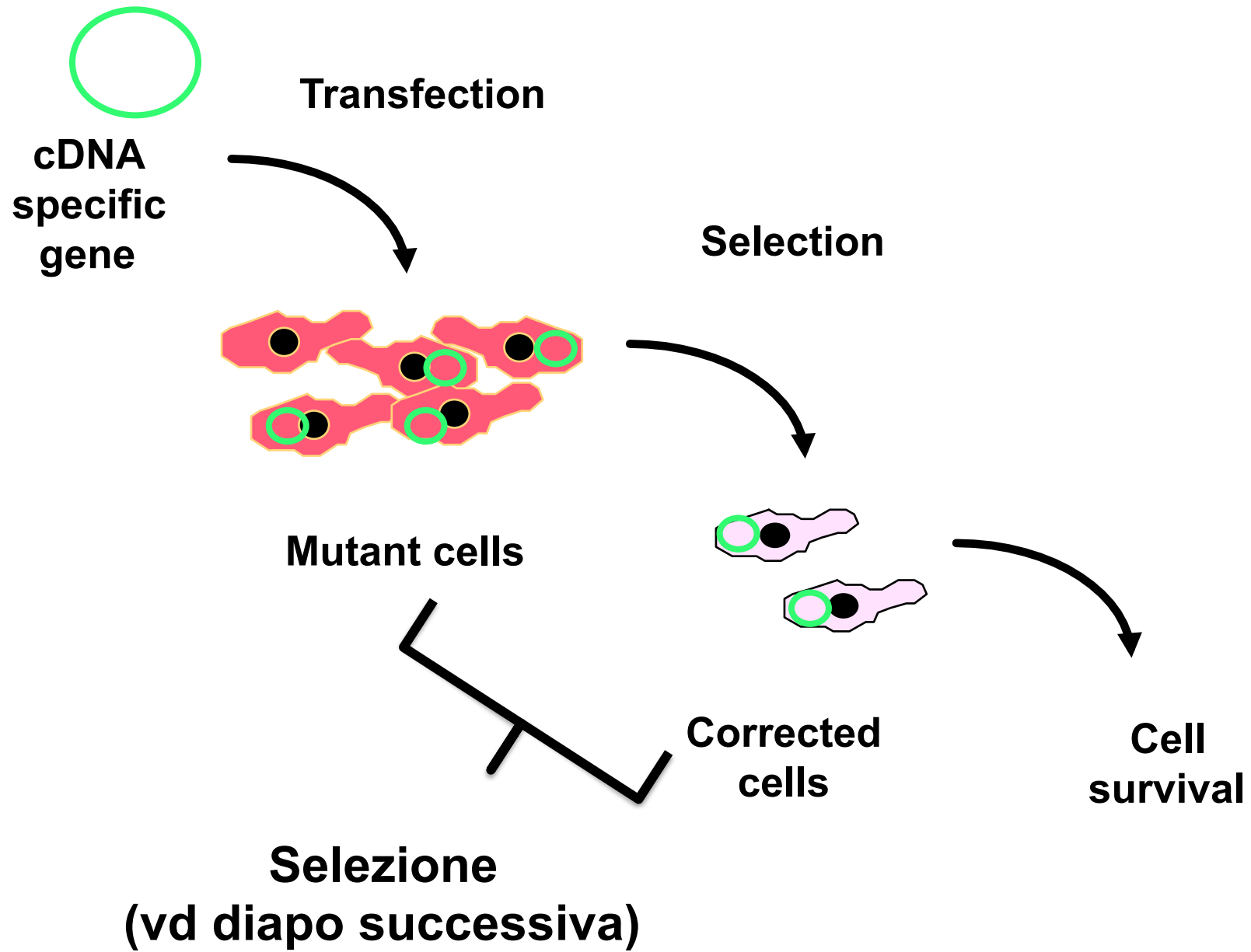


Corrected cells

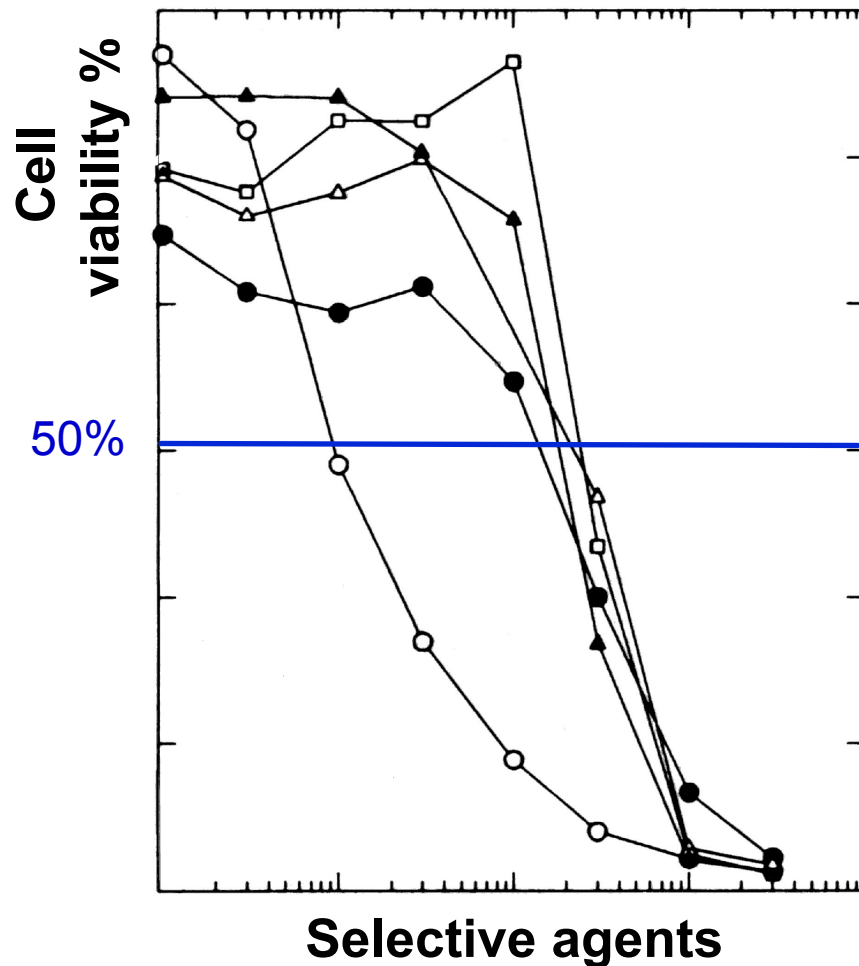
Isolation  
Episomal DNA



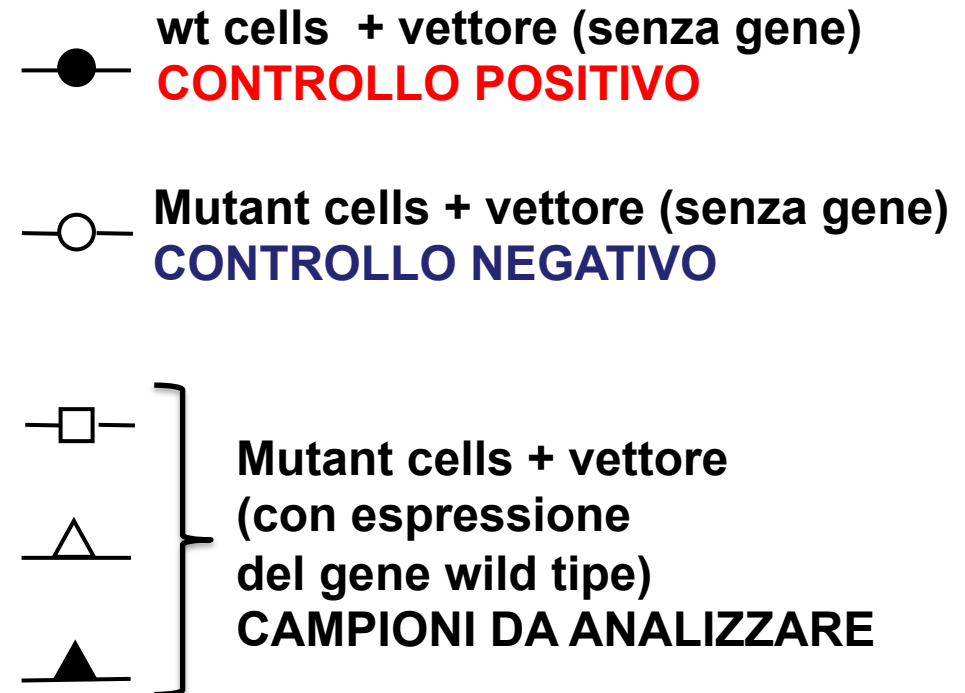
Gene isolation



# Complementation analysis of the isolated cDNA (phenotypic marker: cell survival)



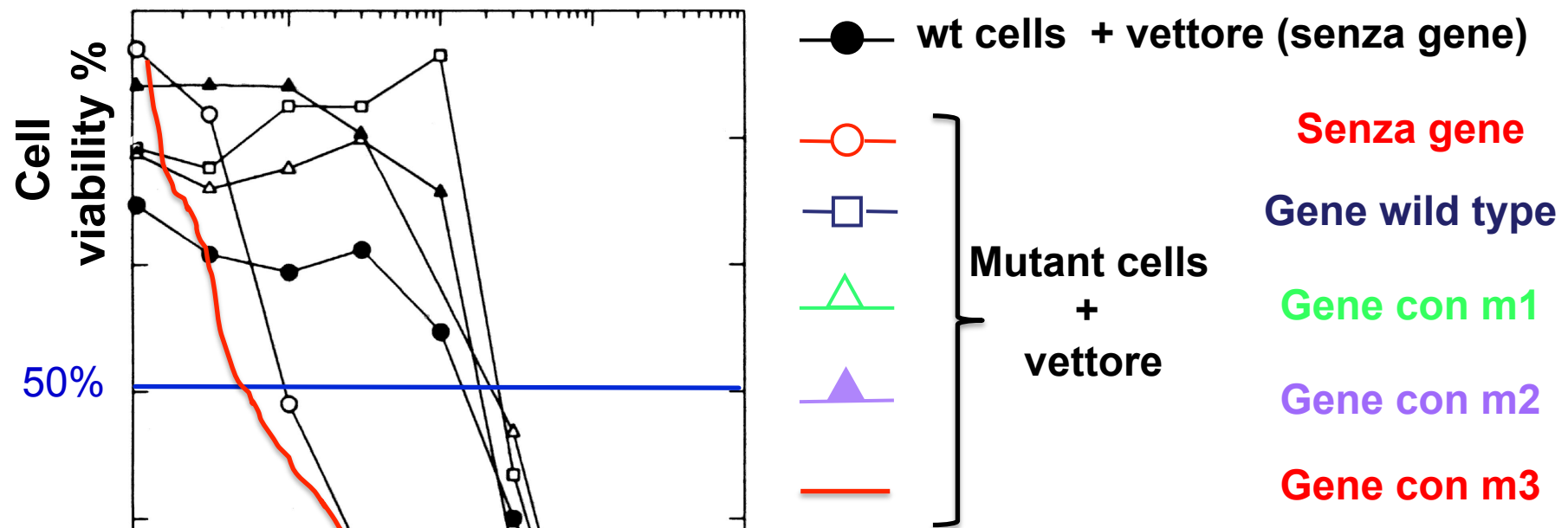
(sostanze che provonano danni al DNA nel modello, per esempio, dell'anemia di Fanconi)



**In tutti i tre I casi le cellule che esprimono il gene wild type (producono una proteina funzionante) si comportano come il controllo positivo. Il gene introdotto complementa le cellule mutanti.**

# Complementation studies

- 1) Genetic heterogeneity determination
- 2) Gene cloning
- 3) Pathogenic effect of missense mutations (functional studies)



Dei geni con le tre diverse varianti, quelli con m1 e m2 si comportano come geni wild type (m1 e m2 sono varianti neutre). Il gene con m3 non complementa: m3 è una variante patogenetica e quindi una mutazione

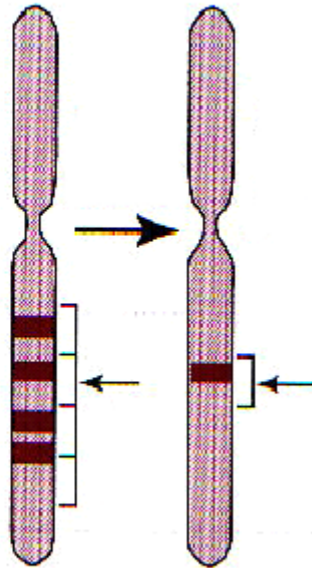
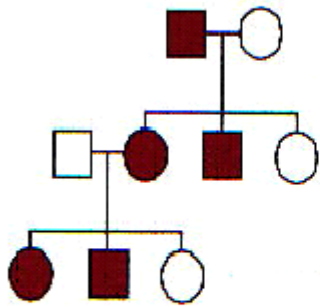
# CLONAGGIO PER POSIZIONE



## Localizzazione gene

Analisi di linkage

Identificazione regione minima



## Analisi regione

Lista geni candidati  
Espressione  
Funzione  
Omologie  
Modelli animali



## Screening mutazioni

Patogenicità  
Segregazione  
Studi funzionali

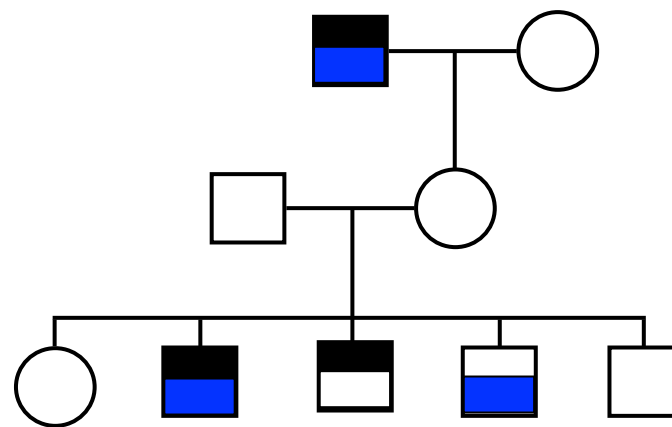
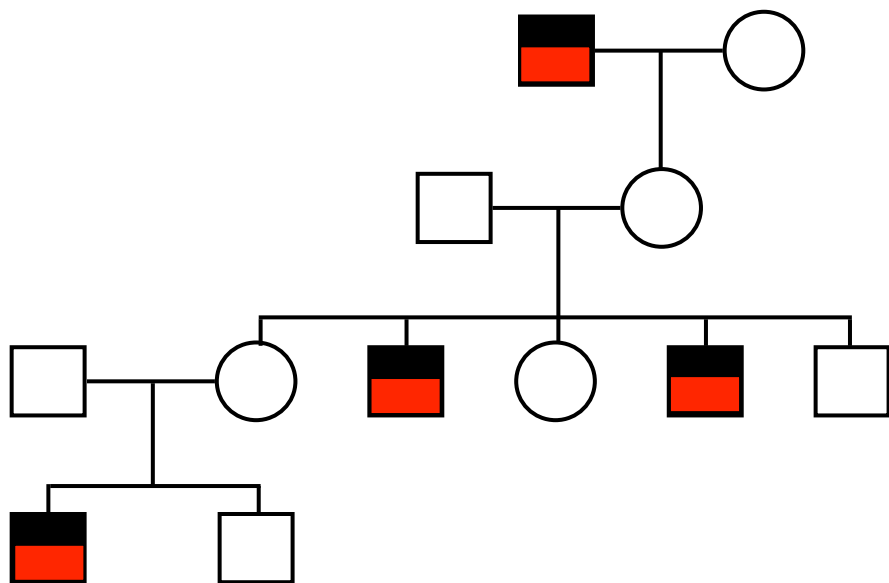
**GENE**

# **Positional cloning:**

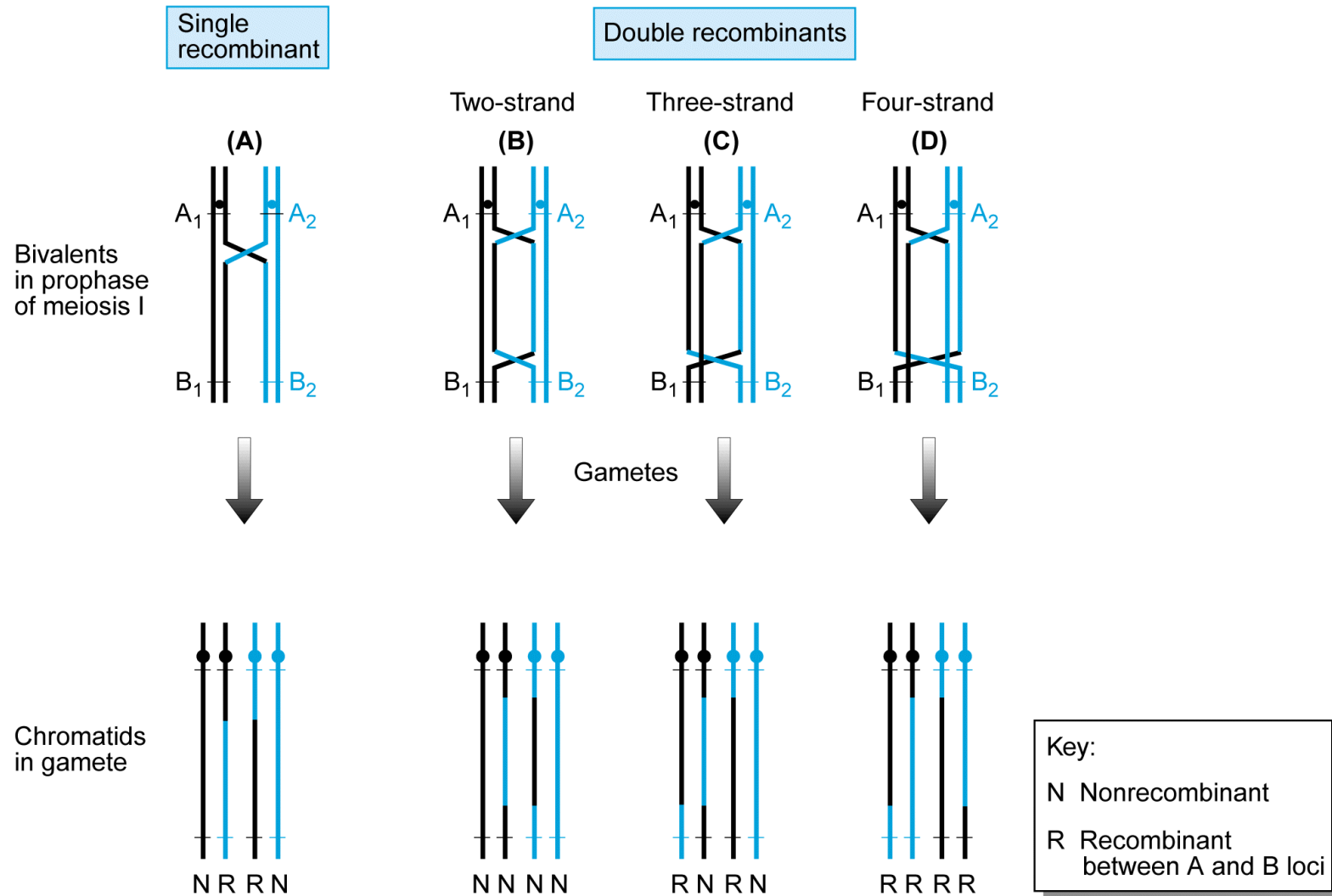
- 1) Localization of gene on chromosome**  
**Linkage analysis**
  
- 2) Definition of candidate gene**  
**Haplotype analysis**  
**Autozigosidad (consanguineous families)**  
**Ancestral haplotype – Founder effect**
  
- 3) Screening for mutations**  
**NGS – WES of candidate region**  
**Sanger sequencing of candidate genes**



■ Daltonismo  
■ Emofilia A  
■ Emofilia B



# La frazione di ricombinazione non supera mai il 50%



# Relazione fra Frequenza di ricombinazione e distanza genetica

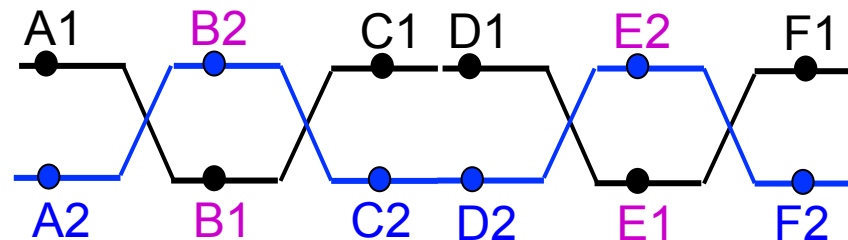
Unità di misura

Frequenza di ricombinazione ( $\theta$ ): %

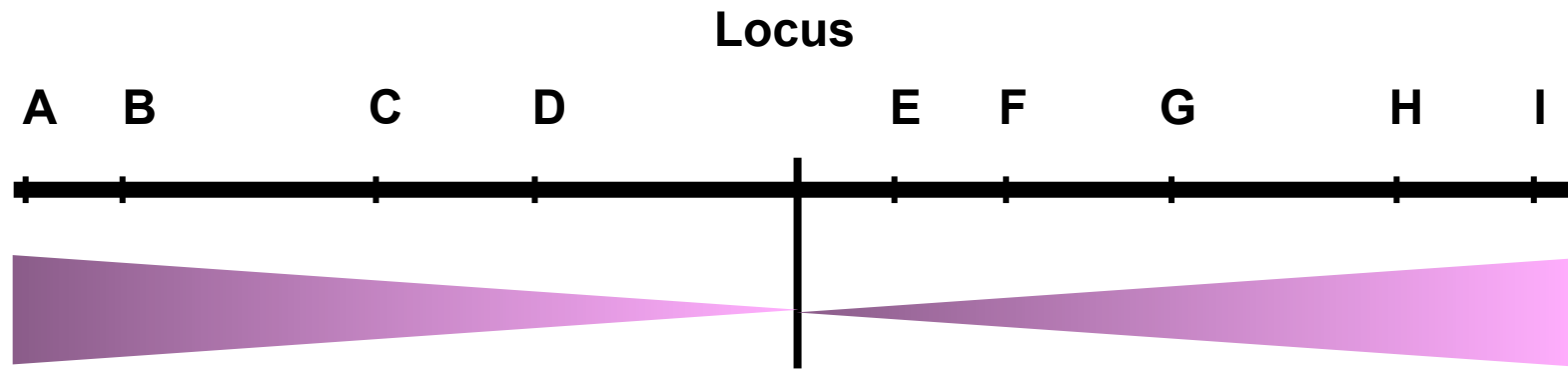
Distanza genetica: cM (centiMorgan)

**1 cM: 1% probabilità di ricombinazione ( $\theta = 1\%$ )**

Le frequenze di ricombinazione non sono additive: effetto dei crossing-over multipli, crossing-over inibisce un crossing-over vicino (interferenza)

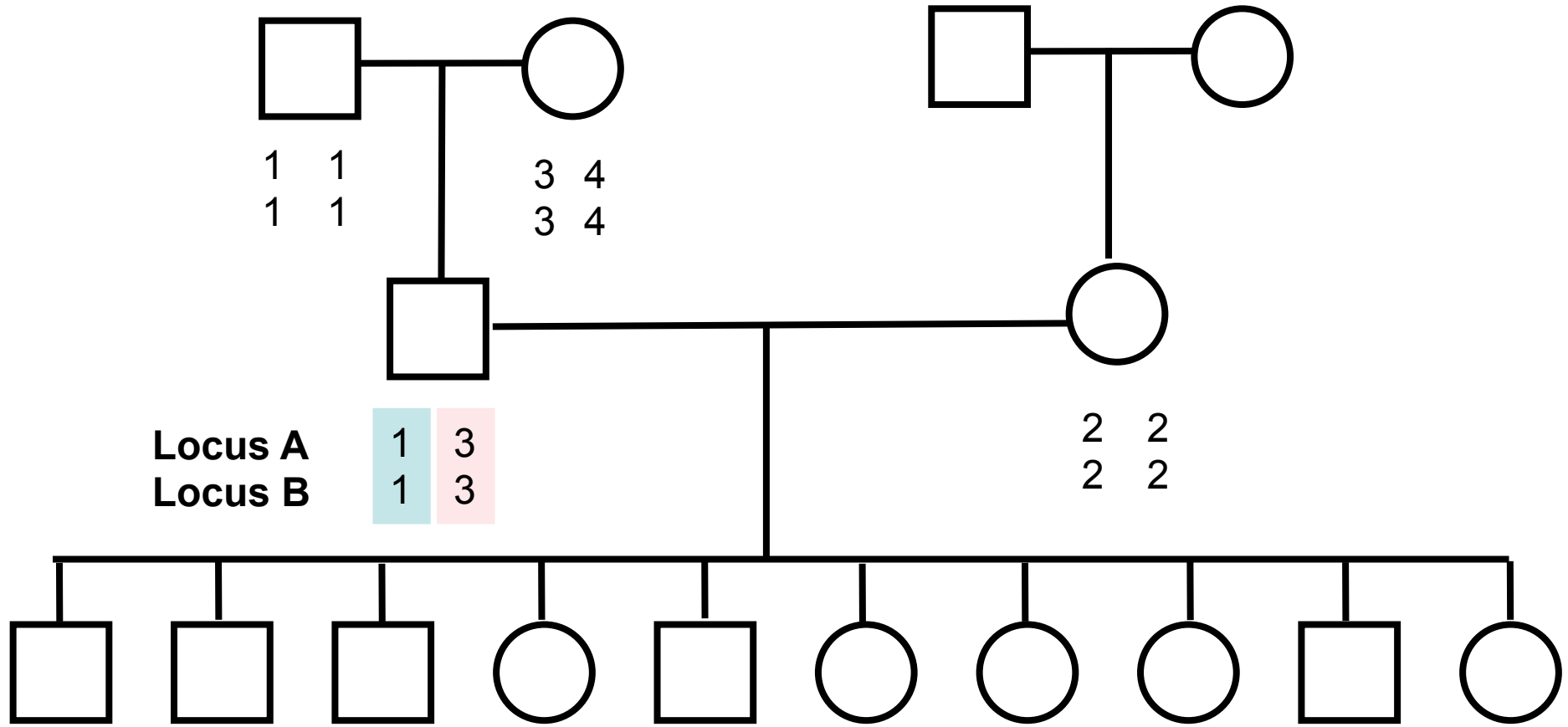


[Funzioni di mappa](#) per convertire  $\theta$  in cM



Frequenza di ricombinazione:

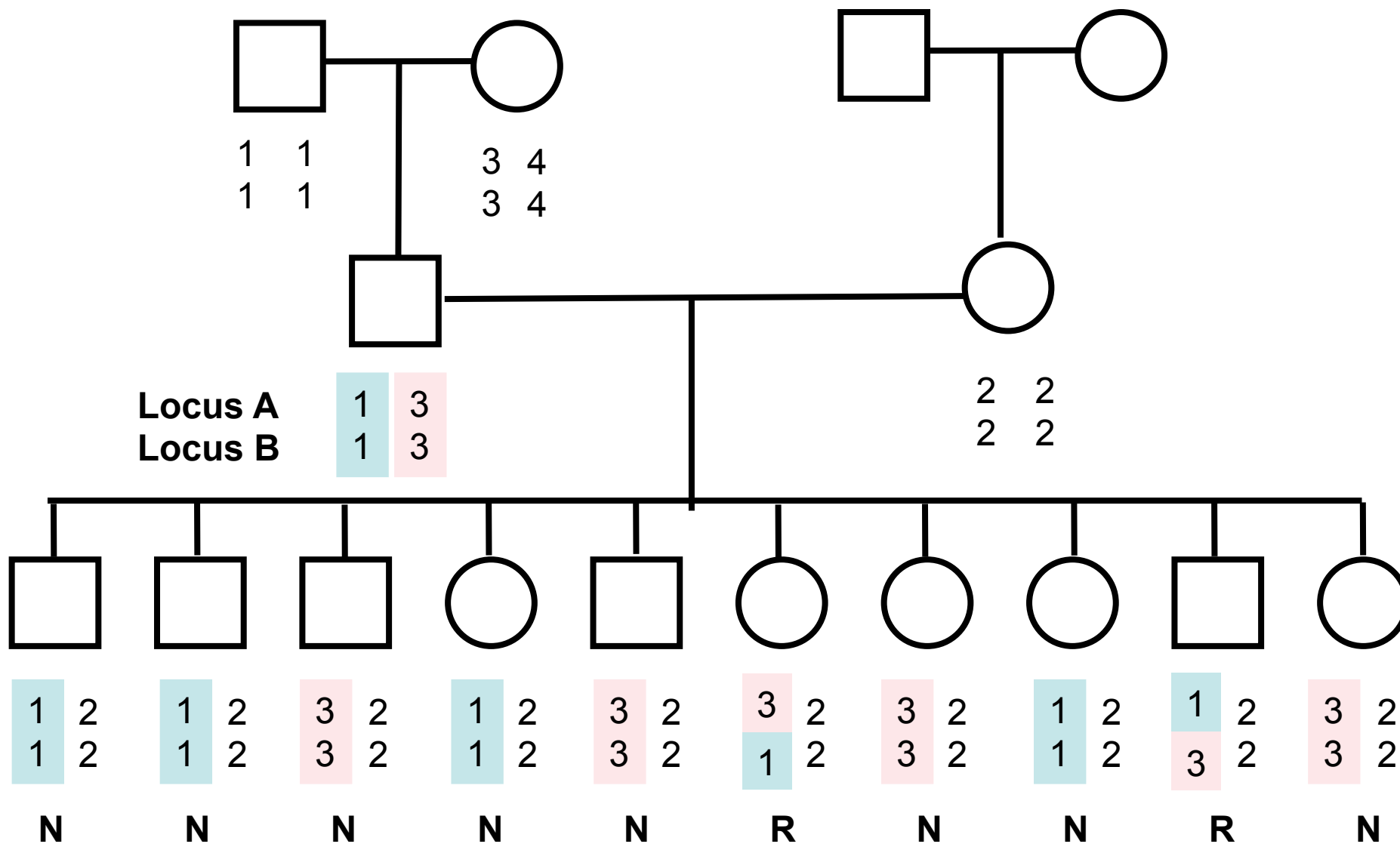
- **Due loci, vicini tra di loro sullo stesso cromosoma, tendono ad essere trasmessi insieme.**
- **Se 2 loci vengono frequentemente ereditati insieme, è probabile che siano localizzati vicini tra di loro sul cromosoma**



**Analisi genotipica della progenie**

3	2	1	2	3	2	3	2	1	2	3	2	3	2	1	2	1	2	3	2
1	2	1	2	3	2	1	2	3	2	3	2	1	2	3	2	1	2	3	2
<b>R</b>	<b>N</b>	<b>N</b>	<b>R</b>	<b>R</b>	<b>R</b>	<b>N</b>	<b>N</b>	<b>R</b>	<b>N</b>										

# Loci sintenici (sullo stesso cromosoma)



## frazione di ricombinazione, $\theta$

Probabilità che avvenga un evento di ricombinazione tra due loci,

$$\theta = \frac{R}{(R + NR)}$$

N meiosi ricombinanti

N totale meiosi

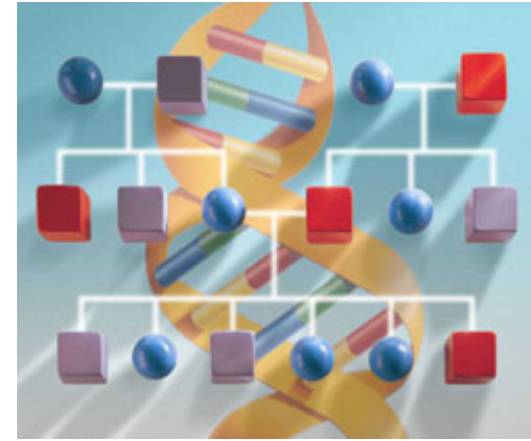
$$\theta = \frac{5}{10}$$

= 0.5 I famiglia  
Distanza A – B = 50% = 50cM  
Assortimento indipendente

$$\theta = \frac{2}{10}$$

= 0.2 II famiglia  
Distanza A – B = 20% = 20cM  
Concatenazione/Linkage

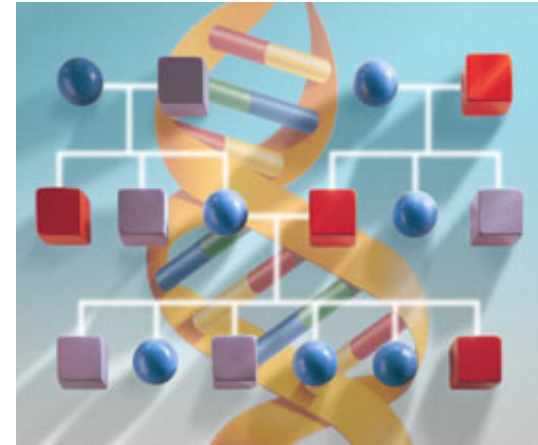
# ANALISI DI LINKAGE



- L'analisi di linkage permette 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

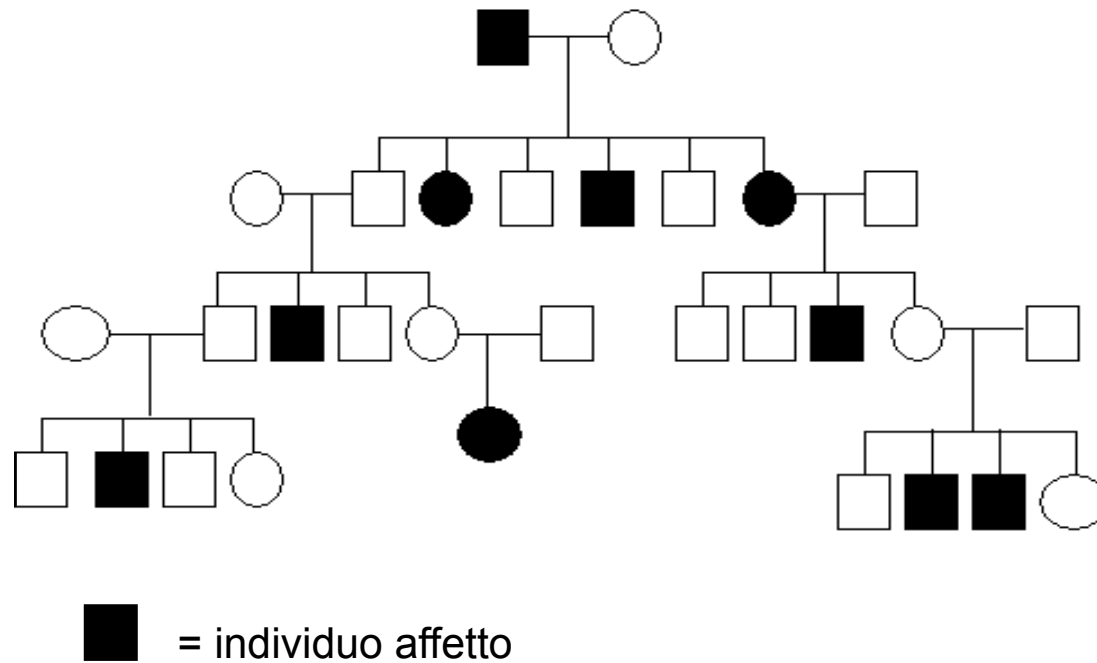


# ANALISI DI LINKAGE



- Determinare la frequenza di ricombinazione ( $\theta$ ) con cui due loci, **locus malattia e un marcatore**, ricombinano fra loro
- Se due loci sono su cromosomi diversi segregano indipendentemente. La probabilità che vengano ereditati insieme è  $1/2$  ( $\theta = 50\%$ )
- Se due loci sono vicini fra loro sullo stesso cromosoma saranno ereditati insieme più frequentemente ( $\theta < 50\%$ )
- **La frequenza di ricombinazione ( $\theta$ ) misura la distanza genetica (non fisica) tra due loci.  $\theta = 1\% = 1$  cM (centimorgan)**

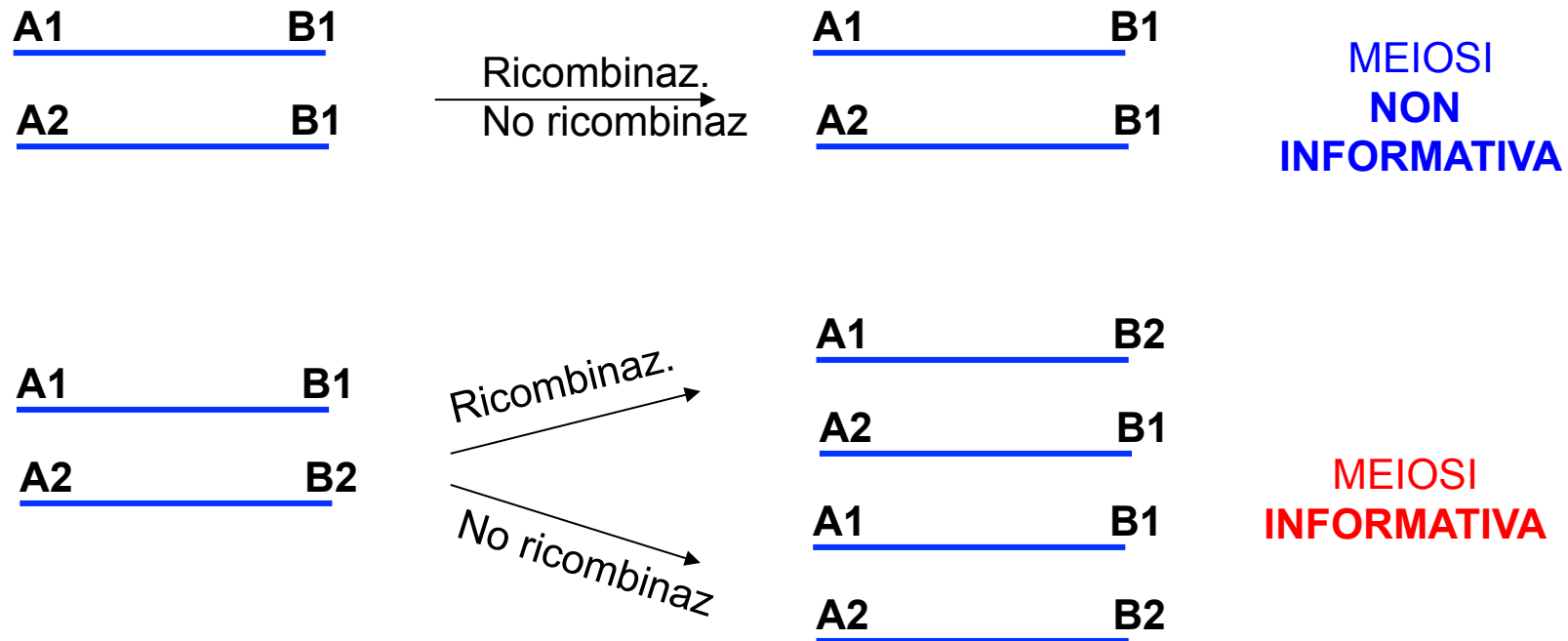
(1) Materiale biologico degli individui appartenenti ad una (ampia) o più (possibilmente omogenee geneticamente) **famiglie** in cui segrega il carattere/malattia



# INFORMATIVITA'

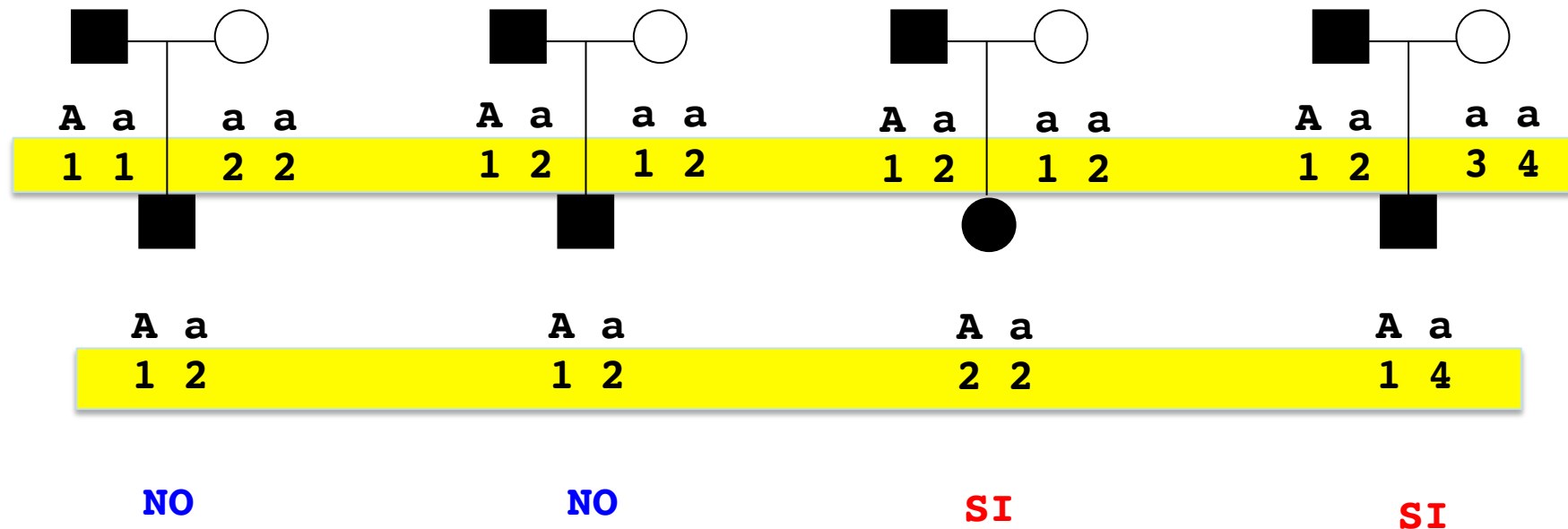
Analisi di linkage:  
misurare la frequenza di ricombinazione fra 2 marcatori.

Necessità di eventi meiotici informativi:  
marcatori parentali **doppi eterozigoti**



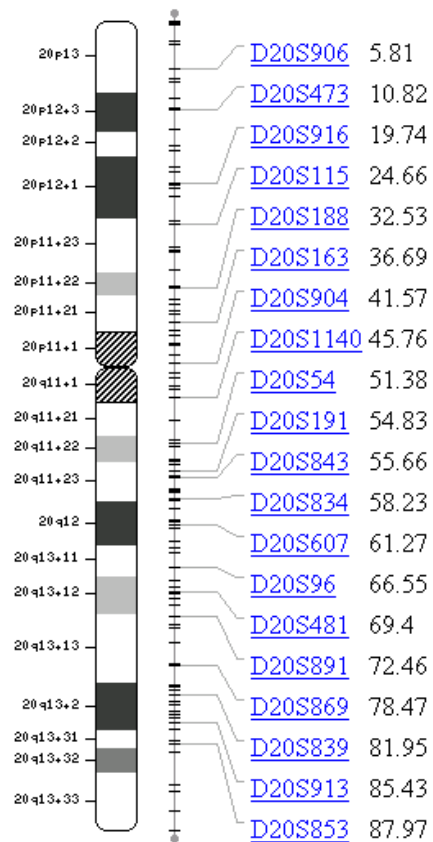
## (2) Meiosi informative: consentono

di individuare quale dei due alleli al locus in esame (giallo) è ereditato dai singoli genitori per i determinare le meiosi ricombinanti

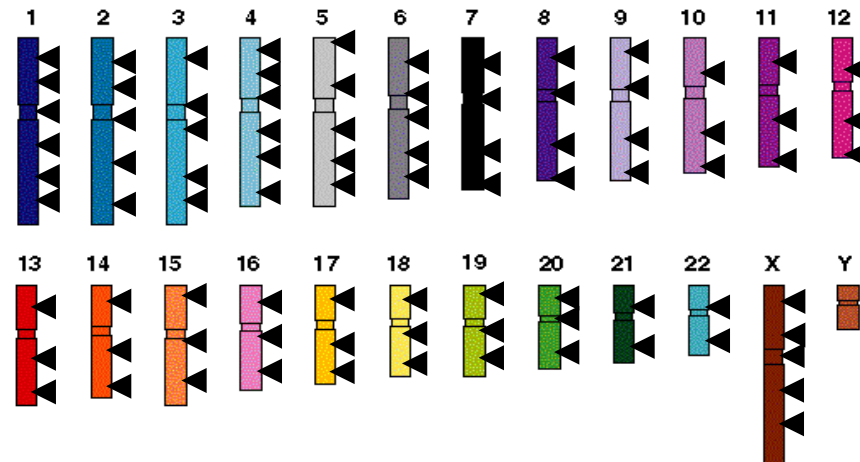


Eterozigotità ad entrambi i loci (malattia e marcatore) non è sempre una condizione sufficiente per determinare l' informatività di una meiosi

### (3) Marcatori adeguati (microsatelliti o e/o SNP) distribuiti uniformemente lungo tutti i cromosomi



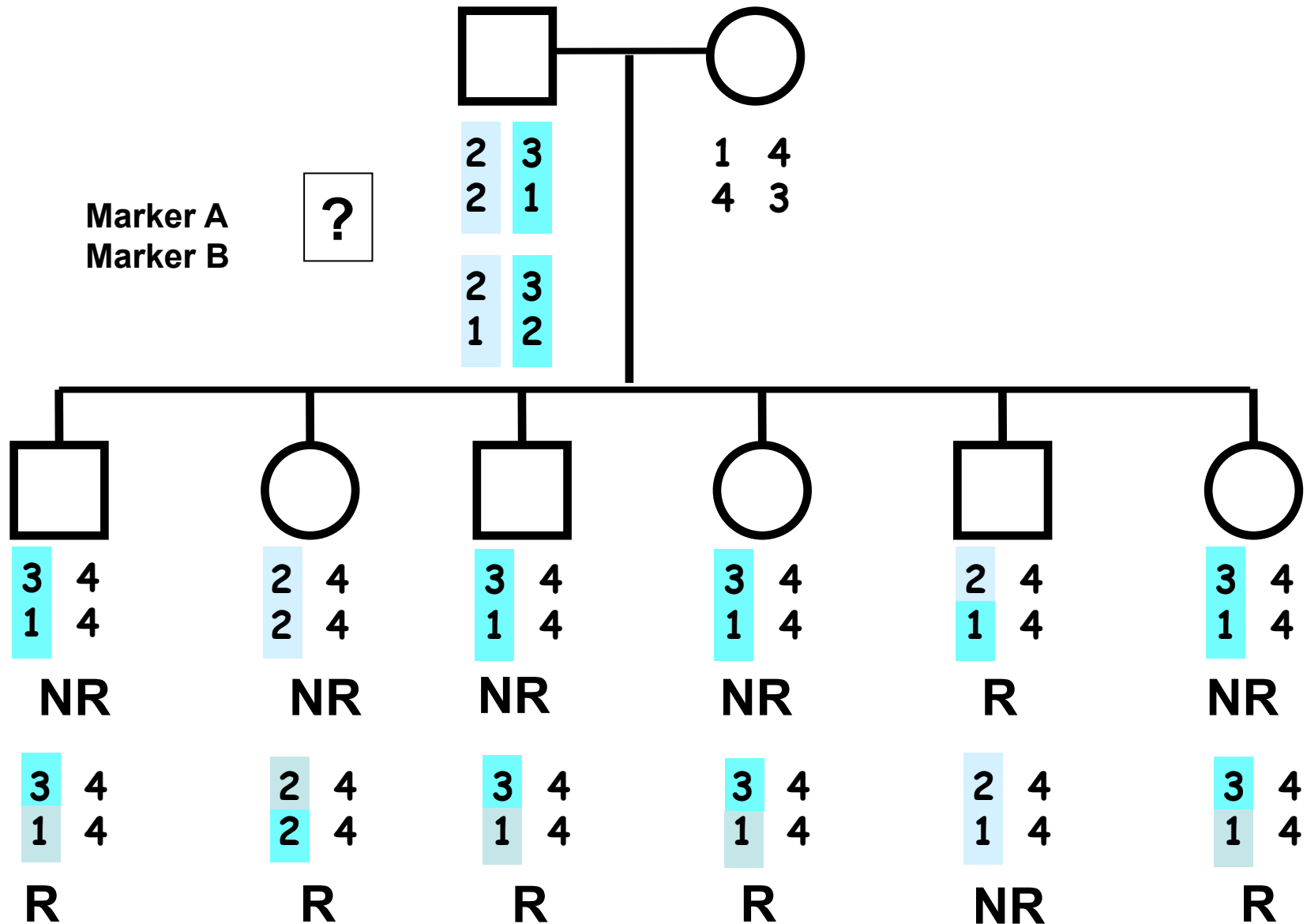
Mapa genetica del cromosoma 2



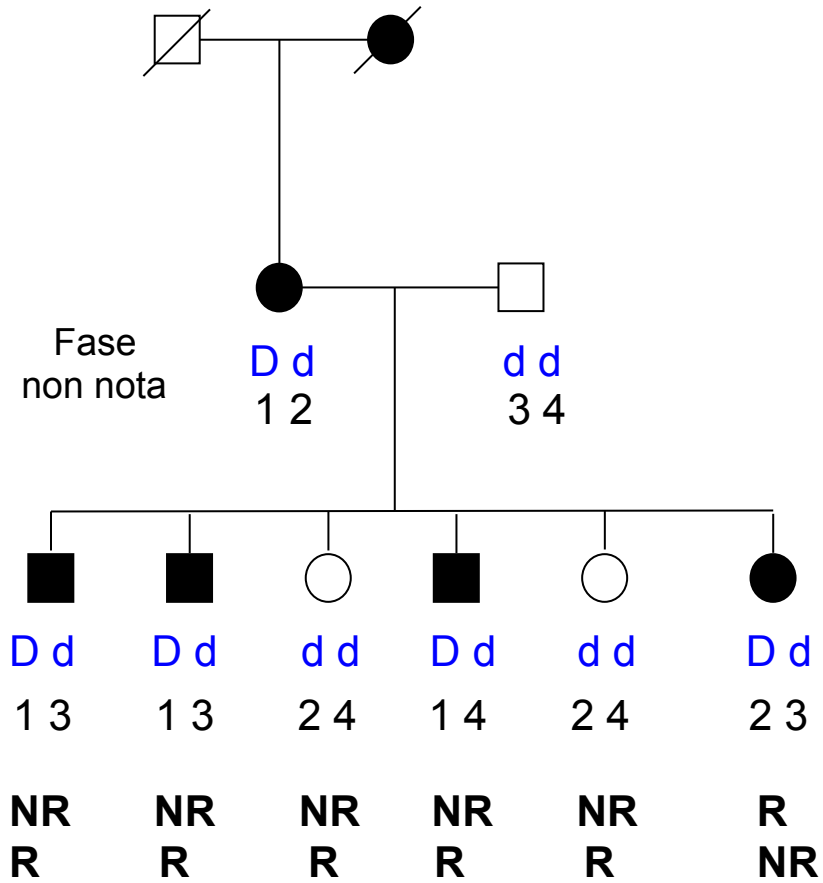
1) Set di microsatelliti (circa 400) distanti 10 cM l'uno dall'altro. Molto polimorfici > meiosi informative

2) Set di SNP (500.000 o più SNP su microarray). Poco polimorfici > meiosi poco informative. Molto densi.

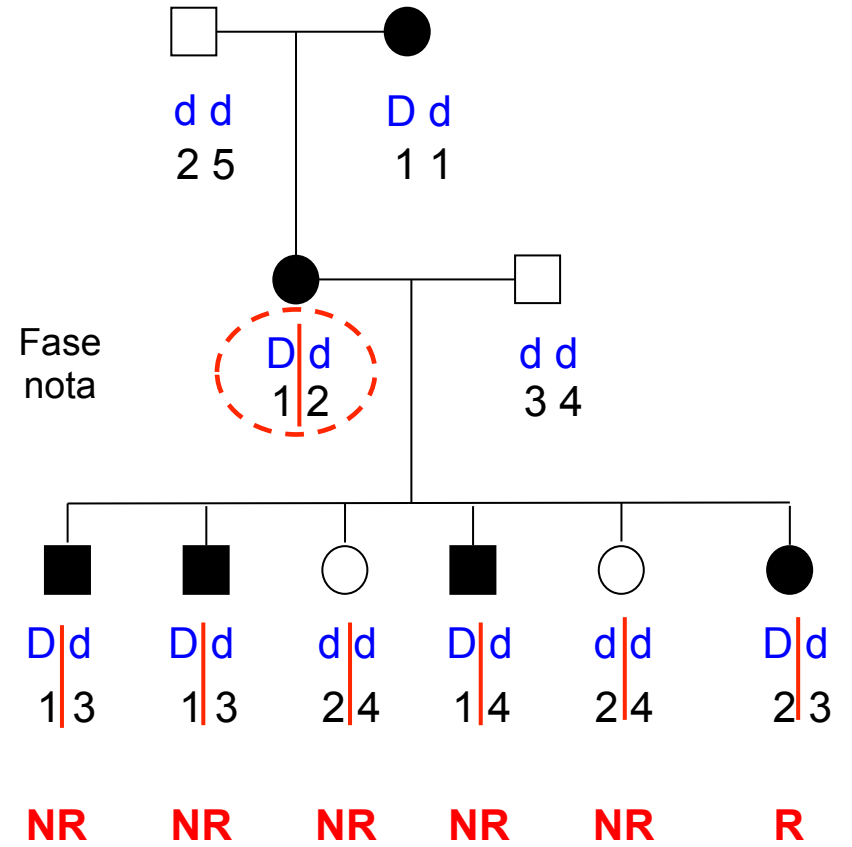
**(4) Stima della frequenza di ricombinazione  $\theta$   
tra il locus malattia e il marcatore**



## Determinazione della fase quando possibile



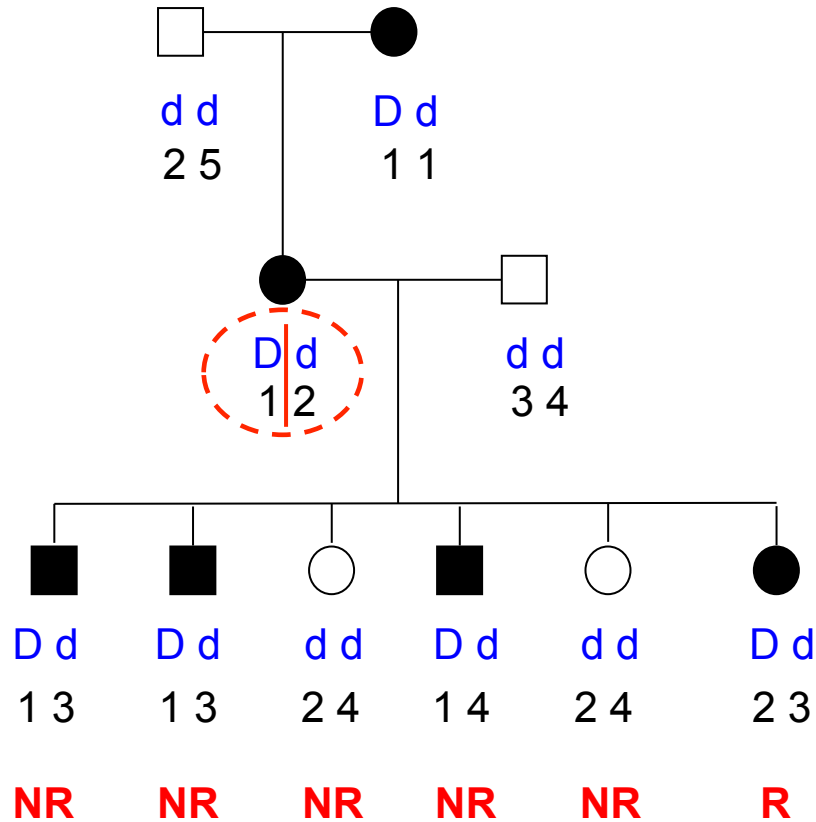
Fase non nota



Fase nota

# Frazione di ricombinazione, $\theta$

Probabilità che avvenga un evento di ricombinazione tra due loci,



Conteggio diretto dei ricombinanti

$$\theta = \frac{R}{(R + NR)}$$

**FR**

- 1) Modello di ereditarietà: genotipi per locus malattia
- 2) **Fase nota** (marcatori informativi, aplotipo)
- 3) Identificazione **ricombinanti e non-ricombinanti**



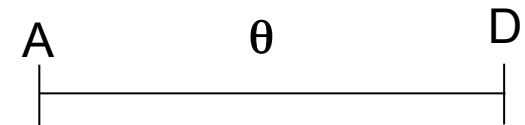
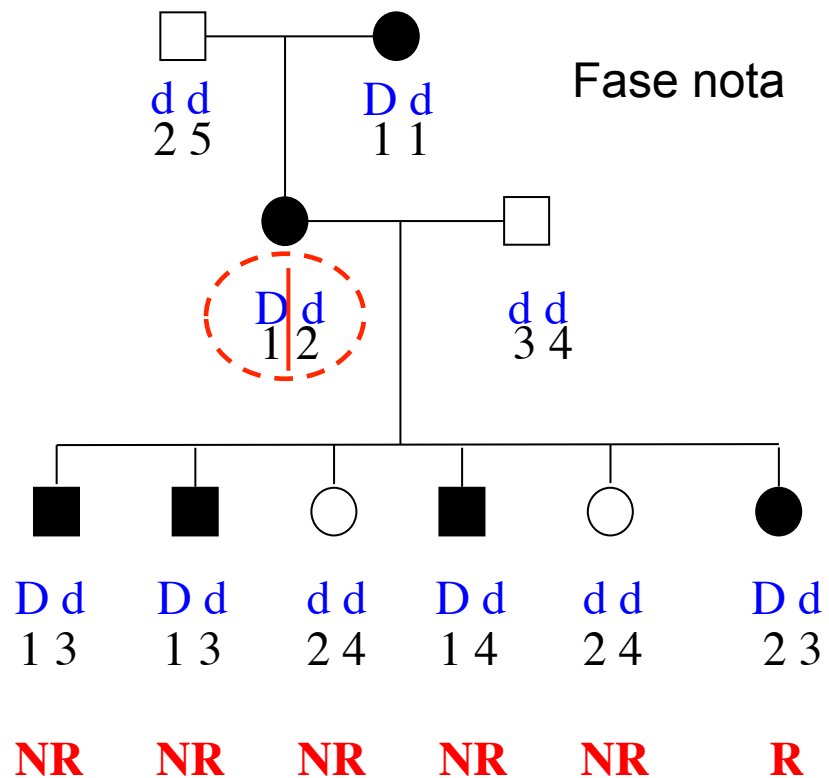
**(5) Test statistico :  $\theta$  è significativamente  $< 0,5$ ?**

Due ipotesi alternative

- 1) Linkage (concatenazione)  $0 < \theta < 0.5$
- 2) Indipendenza  $\theta = 0.5$

$$Z(\theta) = \log_{10} [L(\theta)/L(\theta = 0.5)]$$

**LOD SCORE (Z)** (log of the odds, logaritmo delle probabilità)



**LOD SCORE -  $Z(\theta)$**

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

Essendo  $Z$  una funzione di  $\theta$ , si calcola per diversi valori di  $\theta$

$\theta$	0.0	0.1	0.2	0.3	0.4	0.5
$Z$	$-\infty$	0.577	<b>0.623</b>	0.509	0.299	0

$\theta$	0.0	0.1	0.2	0.3	0.4	0.5
Z	$-\infty$	0.577	0.623	0.509	0.299	0

### Massimo Lod score (MLS)

Valore max di lod score al variare di  $\theta$

Valore di  $\theta$  con massimo LOD:  
frazione di ricombinazione più probabile

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

$$\theta=0 \begin{cases} \text{No ricomb} & Z = \log_{10} \frac{1}{(0.5)^{NR}} \\ \text{Ricomb} \geq 1 & Z = -\infty \end{cases}$$

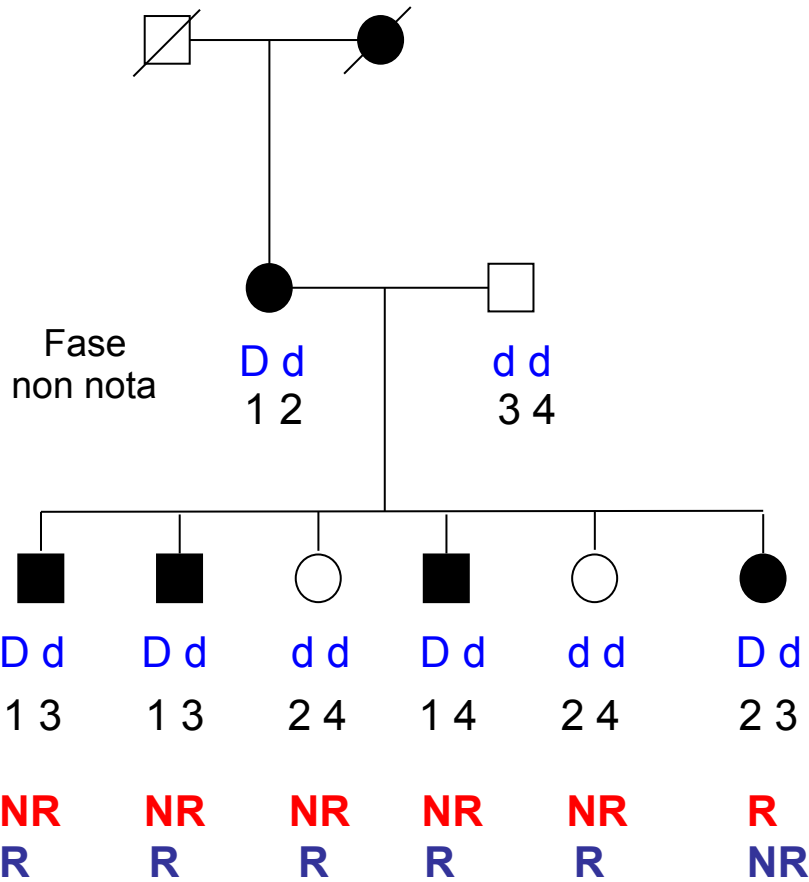
**MLS**

$$Z = \log_{10} \frac{1}{(0.5)^{NR}}$$

Fase non nota

**LOD SCORE - Z(θ)**

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



se II-1 eredita l'allele 1 con la malattia  
se II-1 eredita l'allele 2 con la malattia

θ	0.0	0.1	0.2	0.3	0.4	0.5
Z	- ∞	0.276	<b>0.323</b>	0.222	0.076	0

**MLS ↓**

## **(6) Significatività del test**

**$Z > 3$ : linkage significativo (1000 probabilità contro 1 che il linkage non sia dovuto al caso)**

**$-2 < Z < 3$ : non conclusivo**

**$Z < -2$ : esclusione di linkage (100 a 1 contro il linkage)**

**Linkage tra carattere e marcatore X-linked**

**$Z > 2.3$ : linkage significativo ( $p < 0.05$ )**

SOMMA dei lod score ottenuti da famiglie indipendenti  
(quando singole famiglie non sono sufficientemente informative)

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

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	1:4
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
<b>Totale</b>	2.19	<b>2.43</b>	2.26	1.75	0.99	

### Massimo Lod score (MLS)

Valore max di lod score che si ottiene al variare di  $\theta$

Il valore di  $\theta$  con massimo LOD è la stima più probabile della frazione di ricombinazione

## TABELLA LOD SCORES A 2 PUNTI

Locus	Distance <sup>1</sup>	LOD score at $\theta$								$Z_{\max}$	$\theta_{\max}$
		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

<sup>1</sup>Distance between two adjacent markers in  $\theta$

Localizzazione di un locus implicato nell' ipercolesterolemia autosomica dominante (mutazioni nel gene PCSK9 -proprotein convertase subtilisin/kexin type 9)

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

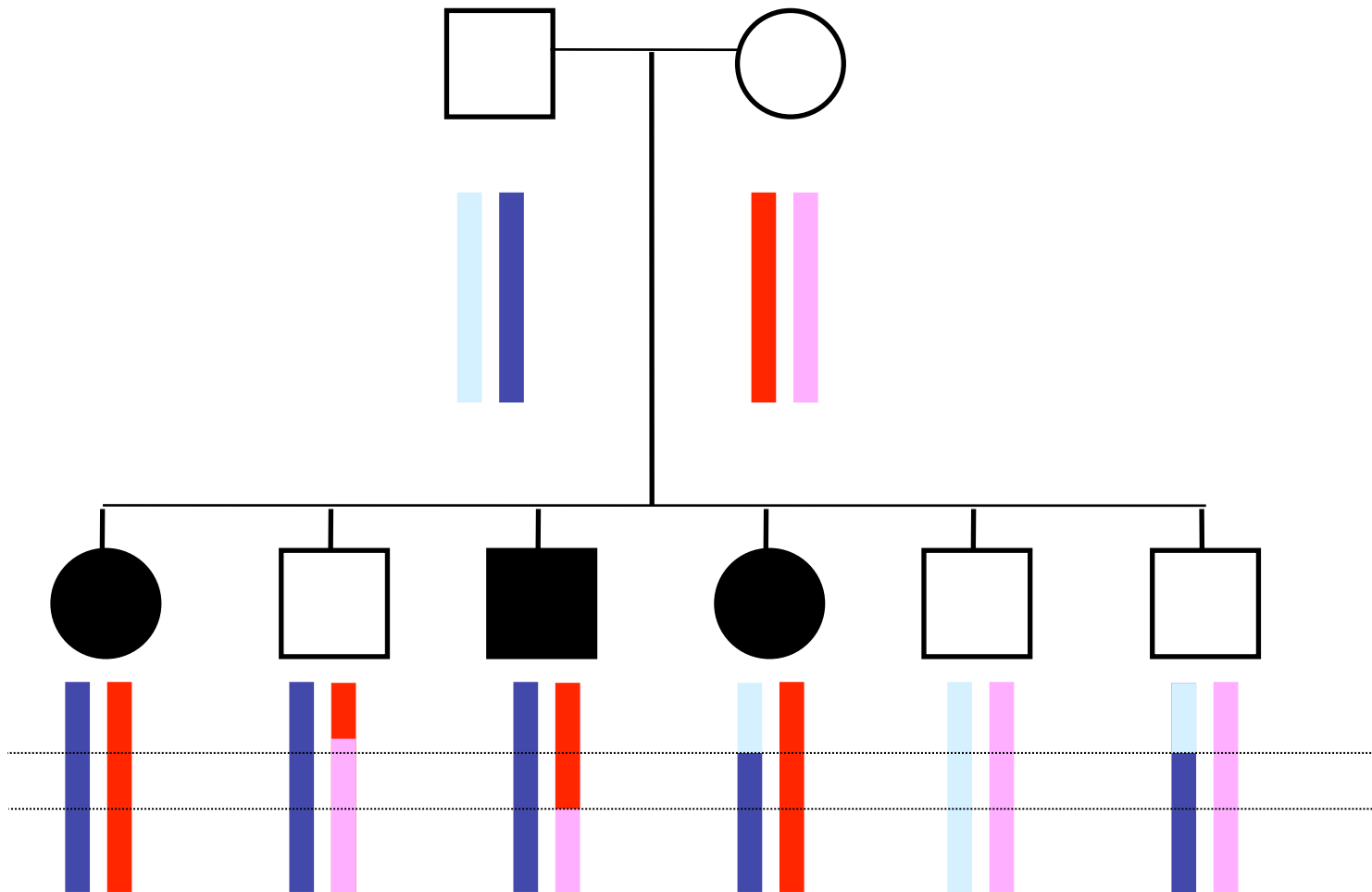
## Mappatura fine del locus malattia

- Analisi dell' APLOTIPO con segregazione sulla base del modello di ereditarietà
- AUTOZIGOSI: omozigosi per alleli identici per discendenza
- Presenza di APLOTIPO ANCESTRALE: linkage disequilibrium
- Anomalie cromosomiche

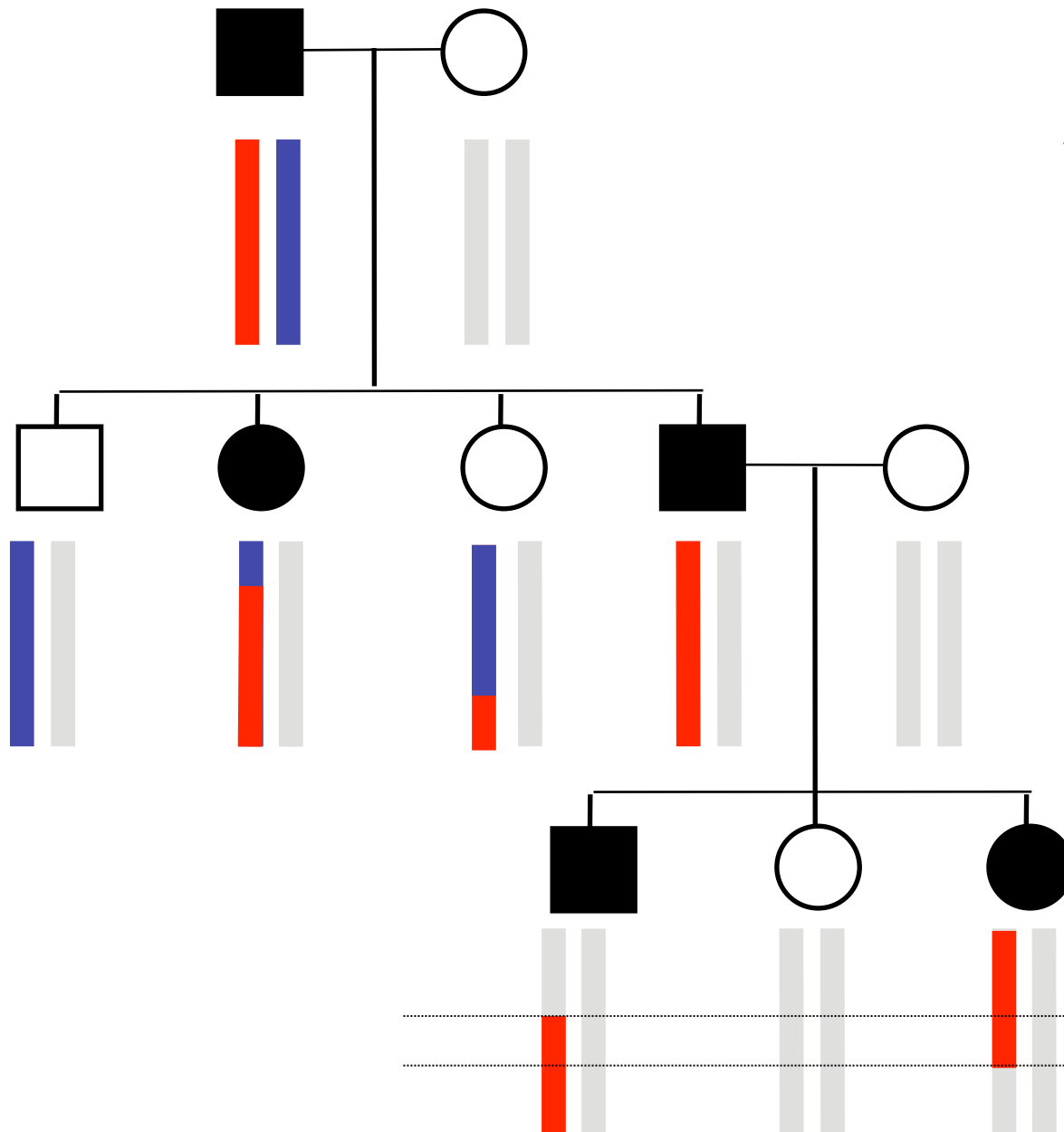


# MALATTIA AUTOSOMICA RECESSIVA

- 1) Genitori sono generalmente asintomatici
- 2) Aumenta incidenza in famiglie consanguinee
- 3) Entrambi i sessi sono affetti
- 4) Rischio 25%



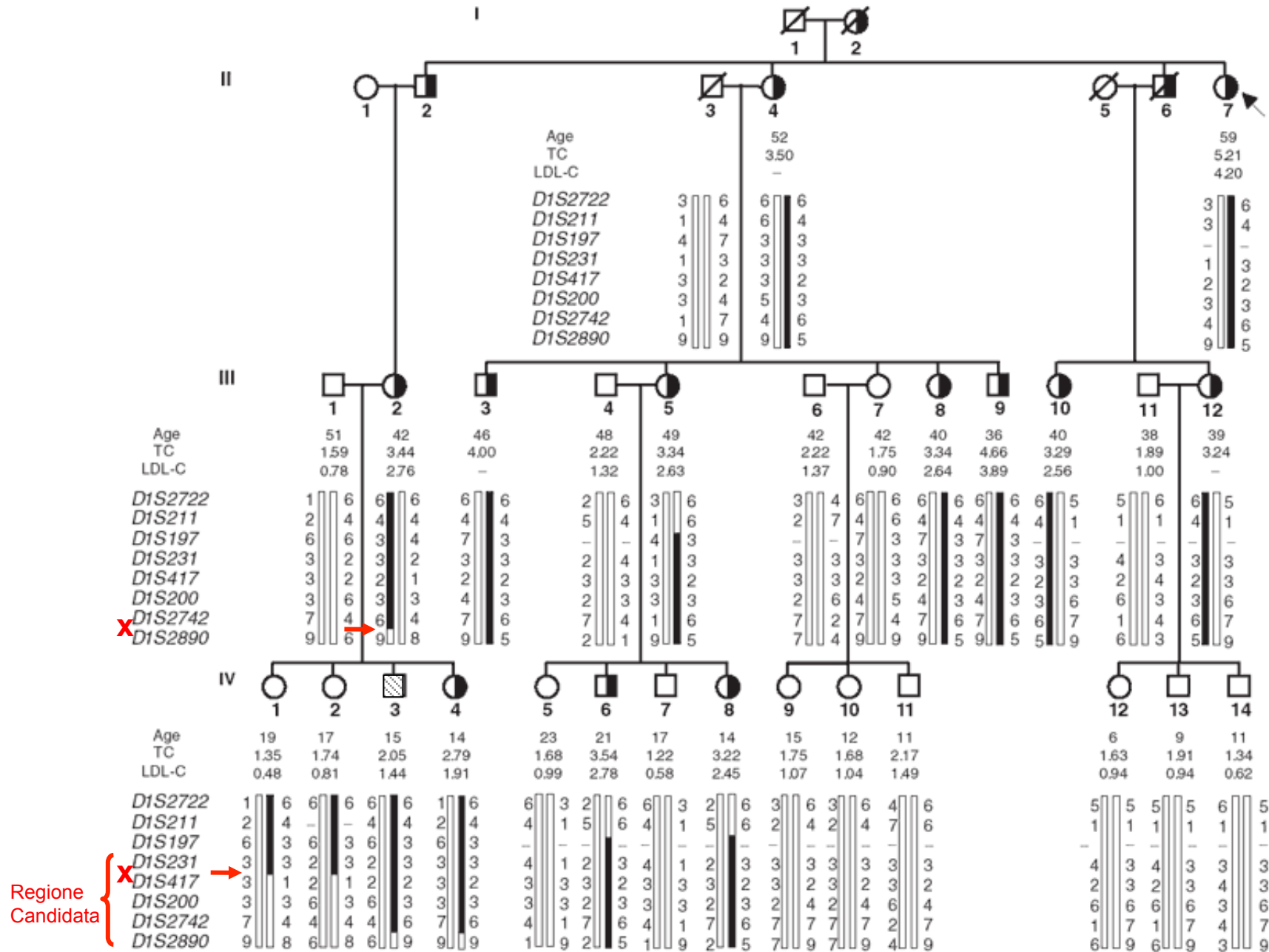
# MALATTIA AUTOSOMICA DOMINANTE



Un genitore affetto  
Affetti entrambi i sessi  
Rischio 50%

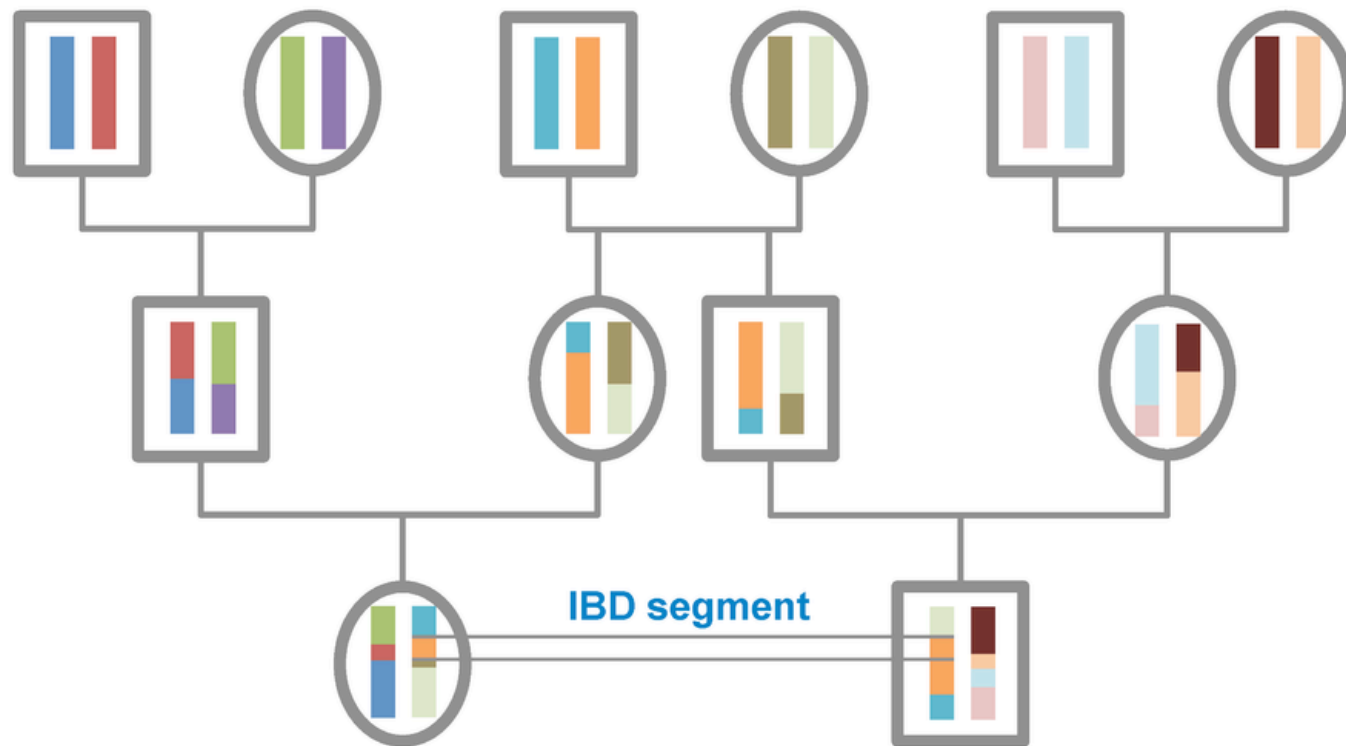
La regione (rossa)  
del cromosoma che  
contiene il gene  
malattia è condivisa  
da tutti gli affetti

# COSTRUZIONE DI APLOTIPI: definizione della regione critica



**IBS:** NA segment **identical by state** in two or more individuals

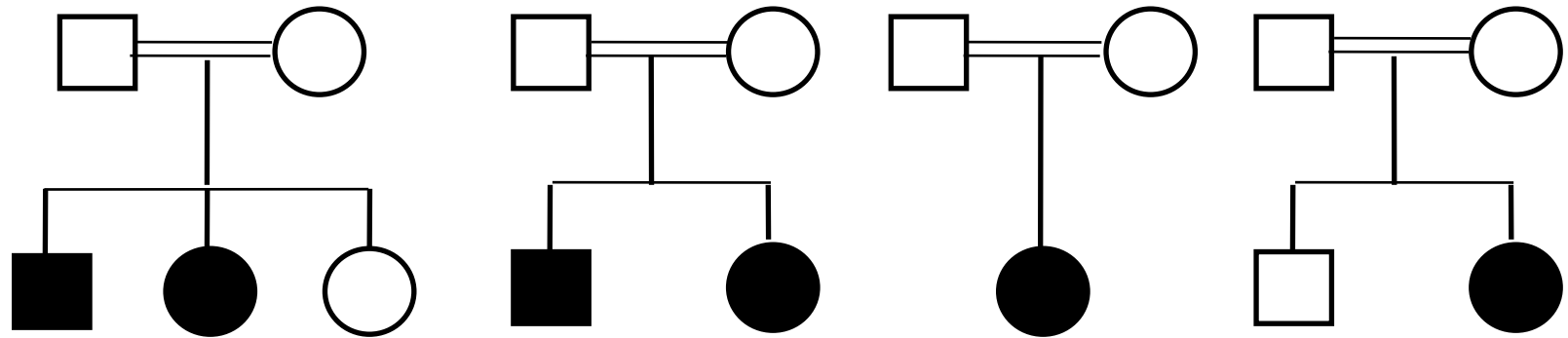
**IBD:** DNA segment **identical by descent** in two or more individuals if they have inherited it from a common ancestor without recombination



# AUTOZIGOSITA' :

Marcatori omozigoti perché ereditati da un comune antenato

Localizzazione di geni: malattie autosomiche recessive



D16S520	3 1	3 1	5 1	2 2	2 6	3 3	5 5	5 5
D16S498	6 8	6 8	22 8	6 6	6 6	6 6	20 8	8 8
D16S305	9 9	9 9	7 9	3 3	3 3	5 5	4 7	7 7
D16S413	6 6	6 6	3 6	4 4	4 4	6 6	9 3	3 3
D16S303	4 4	4 4	0 0	10 10	10 10	10 10	10 10	10 10

**16q**

21.0

22.1

22.2

22.3

23.1

23.2

23.3

24.1

24.2

24.3



cM

3

13

1

2

1

7

**D16S505**

**D16S402**

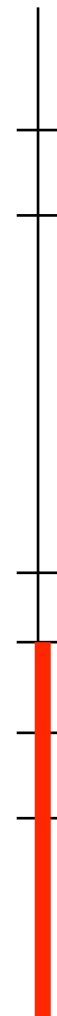
**D16S520**

**D16S498**

**D16S305**

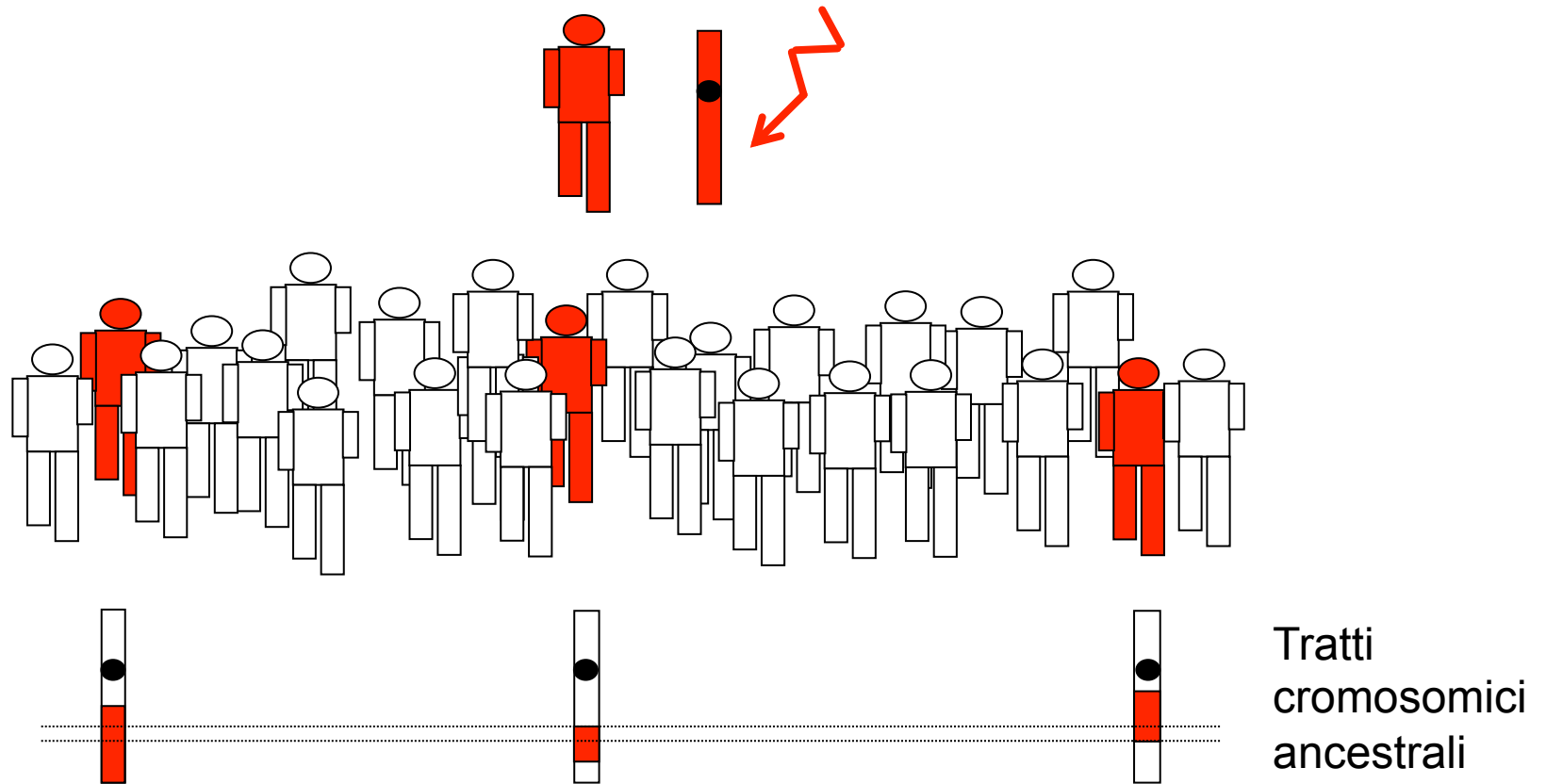
**D16S413**

**D16S303**



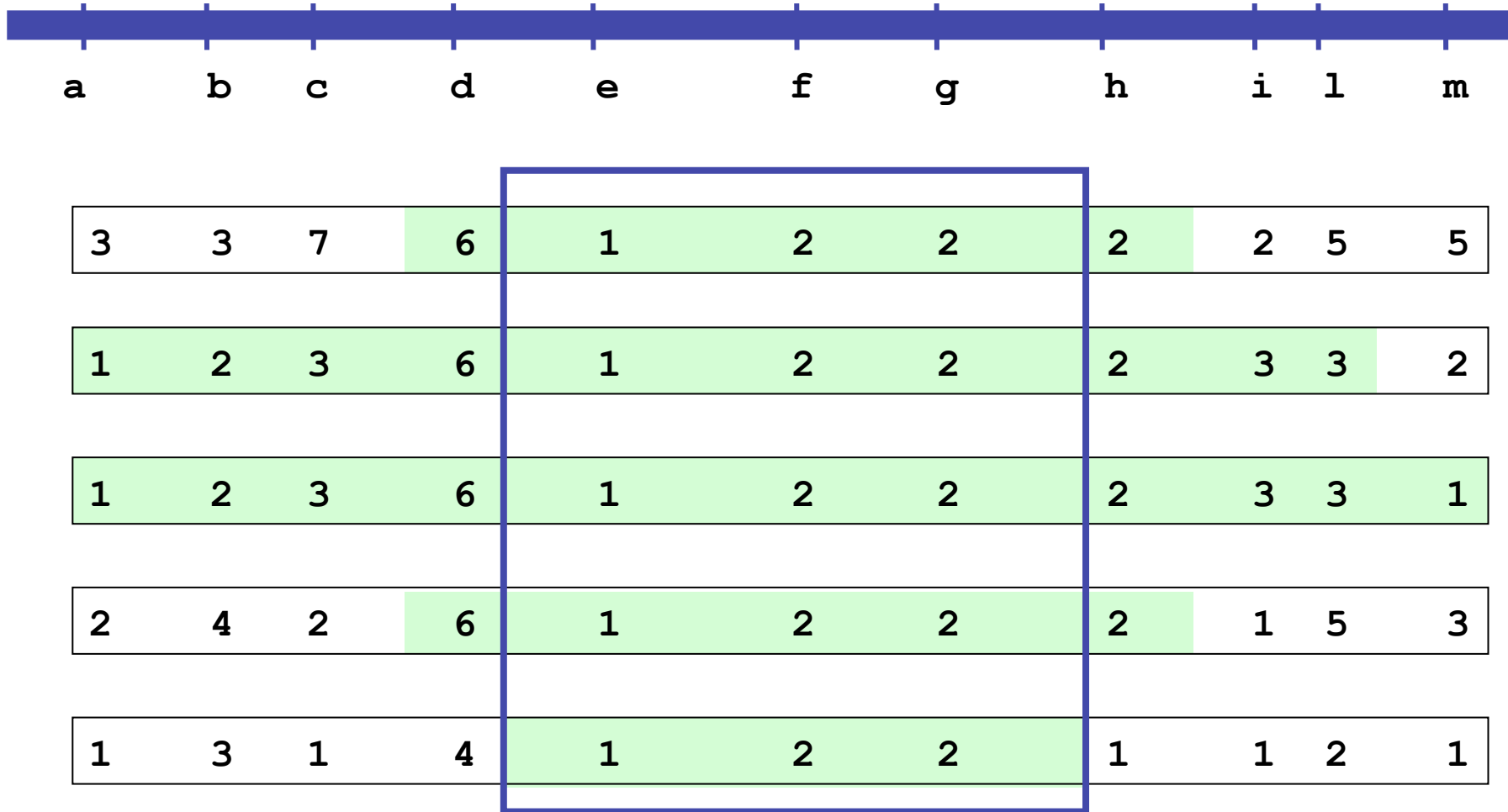
# EFFETTO FONDATORE

Cromosomi con una mutazione che causa malattia derivano da un comune cromosoma ancestrale



Individui affetti condividono l' **aplotipo** nella regione comune

# REGIONE CANDIDATA: analisi di linkage



**REGIONE CANDIDATA**



# ANOMALIE CROMOSOMICHE

- Ricerca sistematica di microdelezioni e duplicazioni (CGH e SNP array)
- Pazienti con anomalie cromosomiche bilanciate (traslocazioni e inversioni) e un fenotipo anomalo

**Una delle due rotture può essere la causa del fenotipo**



**Localizzazione e caratterizzazione delle rotture**

## ***Identificata la regione candidata ....***

**Ricerca delle mutazioni in:**

- 1) NGS (esoma o genoma) della regione candidata**
- 2) Sequenziamento Sanger dei geni candidati**

**Criteri per definire i geni candidati:**

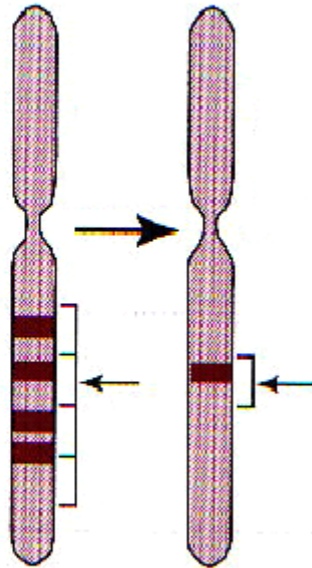
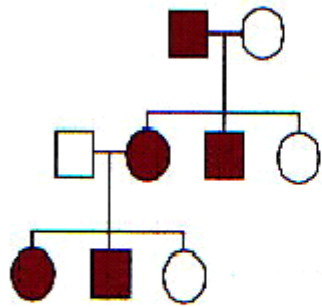
- Pattern di **espressione**
- **Funzione** (membro di una famiglia, interattori, pathway molecolare, ...)
- **Omologia** con gene implicato in una malattia simile
- **Modelli animali:** mutazioni nel gene omologo causa un fenotipo simile alla malattia in esame

# CLONAGGIO PER POSIZIONE



## Localizzazione gene

Analisi di linkage  
Identificazione regione minima



## Analisi regione

Lista geni candidati  
Espressione  
Funzione  
Omologie  
Modelli animali



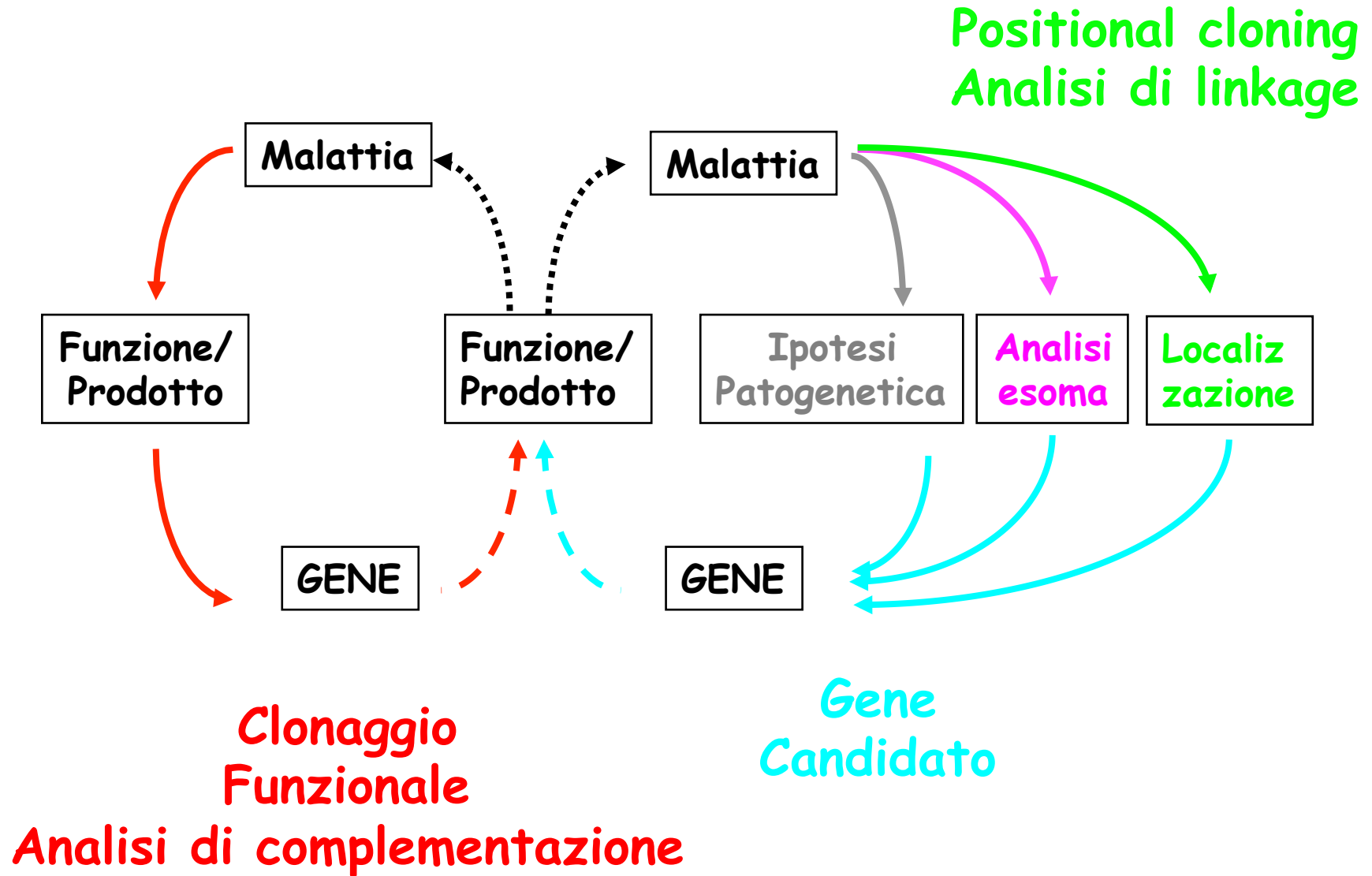
## Screening mutazioni

Patogenicità  
Segregazione

**GENE**



# STRATEGIE di CLONAGGIO



# Sanger sequencing vs Next Generation Sequencing (NGS)

500-1000 bp > entire genome



## *NGS platforms*

- Illumina (MiSeq, HiSeq, Genome Analyzer, etc.)
- Ion Torrent (PGM, Proton)
- Roche/454 (FLX+, Junior)
- SOLiD
- Third Generation/Single Molecule Sequencing

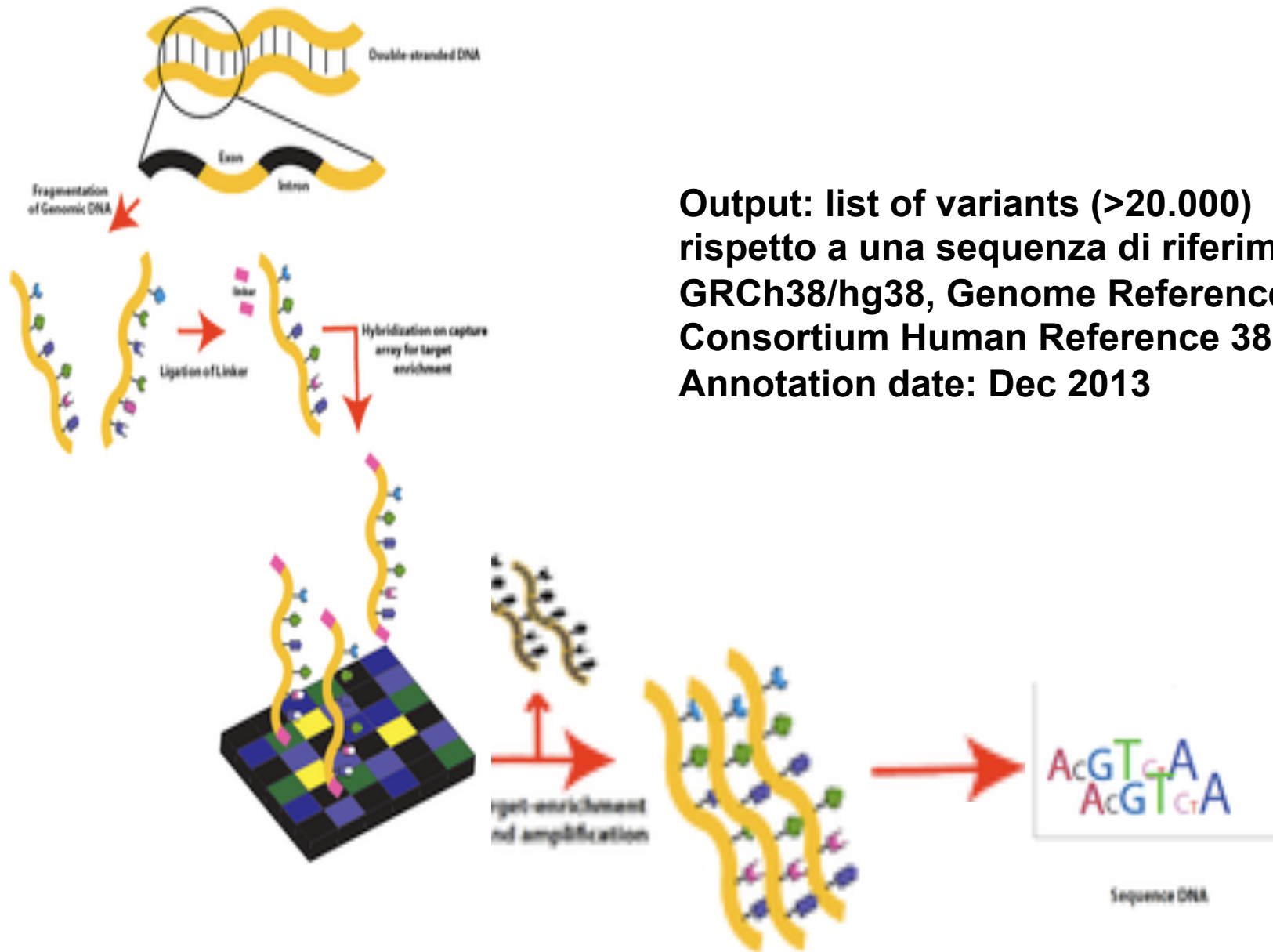
**Sanger sequencing:** consente di sequenziare un frammento alla volta

**NGS** (anche high-throughput sequencing, sequenziamento ad alta resa): consente di sequenziare moltissimi frammenti in parallelo con riduzione tempi e i costi della ricerca di mutazioni.

## ***NGS applications***

- **Whole Genome Sequencing (WGS):** entire genome (3000 Mb)
- **Whole Exome Sequencing (WES):** all exons (180,000-200,000 exons; 1% of genome; 30 Mb)
- **Transcriptome analysis (RNA-seq):** the quantification of transcript levels and the sequence information
- **Epigenomics:** study of changes in the regulation of gene activity and expression that are not dependent on the DNA sequence
- **Metagenomics:** access of the genetic content of entire communities of organisms

# Exome sequencing



**Output: list of variants (>20.000)  
rispetto a una sequenza di riferimento:  
GRCh38/hg38, Genome Reference  
Consortium Human Reference 38  
Annotation date: Dec 2013**

## Filtering: identificazione delle potenziali varianti causative

1

Rarità

Mutazione è rara  
non presente nei  
database



Eliminare varianti con MAF>1 nelle  
banche dati; dbSNP database  
1000 genome, GnomAD  
Exome aggregation Consortium

**20.000**



2

Patogenicità

Nonsense/frameshit  
Missense: tools bioinformatici  
Sinonime non patogenetiche



Eliminare varianti non  
patogenetiche

**300**



3

Ereditarietà



Verificare modello

**1-5** varianti



**GnomAD:** <http://gnomad.broadinstitute.org/>

- **Genome Aggregation Database (gnomAD) is a resource developed with the goal of aggregating both exome and genome sequencing data from a wide variety of large-scale sequencing projects.**
- **It contains data set from**
  - **123,136 exome sequences**
  - **15,496 whole-genome sequences****from unrelated individuals sequenced as part of various disease-specific and population genetic studies**
- **Used to known variants identified, their MAF:  
Connection to other useful links**

# gnomAD browser beta | genome Aggregation Database

BRCA1

Example - Gene: [PCSK9](#), Variant: [1-55516888-G-GA](#)

## Gene: BRCA1

**BRCA1** breast cancer 1, early onset  
**Number of variants** 2936 (Including filtered: 3296)  
**UCSC Browser** [17:41196312-41277500](#)  
**GeneCards** [BRCA1](#)  
**OMIM** [113705](#)  
**Other** [External References](#)

Transcripts ▾

### Gene summary

(Coverage shown for [canonical transcript](#): ENST00000471181)

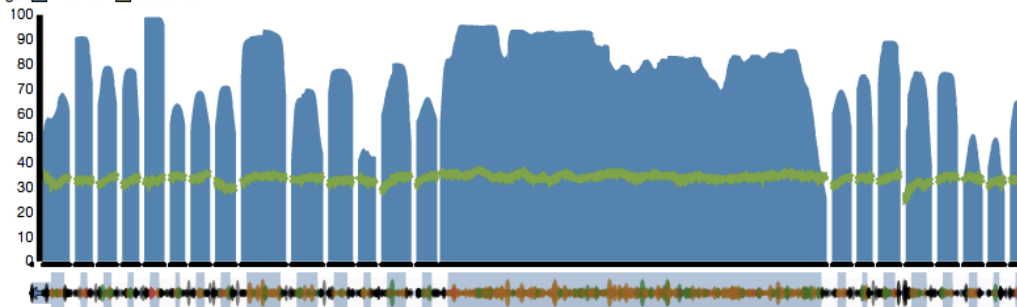
Mean coverage 75.32

Display: **Overview** **Detail**  Include UTRs in plot

Coverage metric: **Average** **Individuals over X**

Metric:

Coverage:  Exomes  Genomes



Save coverage plot

Save gene image

# BRCA1: very common variant

Variant: 17:41196408 G / A

	Exomes	Genomes	Total
Filter	Pass	Pass	
Allele Count	45170	9682	54852
Allele Number	131564	30898	162462
Allele Frequency	0.3433	0.3134	0.3376
dbSNP	rs12516		
UCSC	17-41196408-G-A <a href="#">↗</a>		
ClinVar	<a href="#">Click to search for variant in Clinvar</a> <a href="#">↗</a>		

Genotype Quality Metrics

Site Quality Metrics

Report this variant

## Annotations

This variant falls on 14 transcripts in 1 genes:

### 3' UTR

- BRCA1

Transcripts ▾

**Note:** This list may not include additional transcripts in the same gene that the variant does not overlap.

## Population Frequencies

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
South Asian	11276	22518	2895	0.5008
European (Finnish)	3442	8790	684	0.3916
East Asian	4385	11516	851	0.3808
Ashkenazi Jewish*	3107	8346	595	0.3723
Other	1564	4596	278	0.3403
European (Non-Finnish)	22345	67246	3709	0.3323
African	3445	14922	413	0.2309
Latino	5288	24528	615	0.2156
Total	54852	162462	10040	0.3376

Include:  Exomes  Genomes

\* For detailed analysis of Ashkenazi Jewish frequency see the [IBD Exomes Browser](#).

# BRCA1: variant with MAF 0.018

Variant: 17:41222975 C / T

	Exomes	Genomes	Total
Filter	Pass	Pass	
Allele Count	4483	607	5090
Allele Number	245672	30984	276656
Allele Frequency	0.01825	0.01959	0.01840
dbSNP	<a href="#">rs1799967</a>		
UCSC	<a href="#">17-41222975-C-T</a>		
ClinVar	<a href="#">Click to search for variant in Clinvar</a>		

Genotype Quality Metrics

Site Quality Metrics

Report this variant

## Annotations

This variant falls on 18 transcripts in 1 genes:

### missense

- BRCA1

Transcripts ▾

### 3' UTR

- BRCA1 - ENST00000461221

### non coding transcript exon

- BRCA1 - ENST00000472490

### intron

- BRCA1

Transcripts ▾

**Note:** This list may not include additional transcripts in the same gene that the variant does not overlap.

## Population Frequencies

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
European (Finnish)	1323	25788	41	0.05130
South Asian	1163	30782	37	0.03778
Ashkenazi Jewish*	235	10150	2	0.02315
Other	121	6458	2	0.01874
European (Non-Finnish)	2069	126152	26	0.01640
Latino	137	34420	0	0.003980
African	39	24036	0	0.001623
East Asian	3	18870	0	0.0001590
<b>Total</b>	<b>5090</b>	<b>276656</b>	<b>108</b>	<b>0.01840</b>

# BRCA1: very rare variant

Variant: 17:41215963 C / A

	Exomes	Genomes	Total
Filter	Pass	No variant	
Allele Count	1		1
Allele Number	245982		245982
Allele Frequency	4.065e-6		4.065e-6
dbSNP	rs80356896		
UCSC	17-41215963-C-A		
ClinVar	Click to search for variant in Clinvar		

Genotype Quality Metrics

Site Quality Metrics

Report this variant

## Annotations

This variant falls on 18 transcripts in 1 genes:

### stop gained

- BRCA1 Transcripts

### intron

- BRCA1 Transcripts

### 3' UTR

- BRCA1 - ENST00000461221 ()  
LoF: Low-confidence (Non-protein-coding gene)

**Note:** This list may not include additional transcripts in the same gene that the variant does not overlap.

## Population Frequencies

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
European (Non-Finnish)	1	111542	0	0.000008965
African	0	15272	0	0.000
Ashkenazi Jewish	0	9830	0	0.000
East Asian	0	17230	0	0.000
European (Finnish)	0	22290	0	0.000
Latino	0	33574	0	0.000
Other	0	5474	0	0.000
South Asian	0	30770	0	0.000
<b>Total</b>	<b>1</b>	<b>245982</b>	<b>0</b>	<b>0.000004065</b>

Include:  Exomes  Genomes (no variant)

# dbSNP database: <https://www.ncbi.nlm.nih.gov/projects/SNP/>

Reference SNP (refSNP) Cluster Report: rs80356896 **\*\* With Pathogenic allele \*\***

RefSNP	Allele	HGVS Names
<b>Organism:</b> human ( <i>Homo sapiens</i> )	<b>Variation Class:</b> SNV: single nucleotide variation	AAA73985.1:p.Glu1694Ter CM000679.2:g.43063946C>A NC_000017.10:g.41215963C>A NC_000017.11:g.43063946C>A NG_005905.2:g.154038G>T NM_007294.3:c.5080G>T NM_007297.3:c.4939G>T NM_007298.3:c.1768G>T NM_007299.3:c.1768G>T NM_007300.3:c.5143G>T NP_009225.1:p.Glu1694Ter NP_009228.2:p.Glu1647Ter NP_009229.2:p.Glu590Ter NP_009230.2:p.Glu590Ter NP_009231.2:p.Glu1715Ter NR_027676.1:n.5216G>T U14680.1:c.5080G>T
<b>Molecule Type:</b> Genomic	<b>RefSNP Alleles:</b> G/T (REV)	
<b>Created/Updated in build:</b> 132/151	<b>Allele Origin:</b> G:germline T:germline	
<b>Map to Genome Build:</b> <a href="#">108/Weight 1</a>	<b>Ancestral Allele:</b> G	
<b>Validation Status:</b>	<b>Variation Viewer:</b>	
<b>Citation:</b> <a href="#">PubMed</a> <a href="#">LitVar</a> <sup>NEW</sup>	<b>Clinical Significance:</b> <b>With Pathogenic allele</b> <a href="#">[ClinVar]</a>	
	NA	

[...less](#)

# ClinVar: <https://www.ncbi.nlm.nih.gov/clinvar/>

ACTGATGGTATGGGGCCAAGAGATATATCT  
CAGGTACGGCTGTCATCACTTAGACCTCAC  
CAGGGCTGGGCATAAAAGTCAGGGCAGAGC  
CCATGGTGCATCTGACTCCTGAGGAGAAGT  
GCAGGTTGGTATCAAGGTTACAAGACAGGT  
GGCCTGACTCTCTGCTTGGTCTAT

**ClinVar**  
ClinVar aggregates information about genomic variation and its relationship to human health.

**NEW** [Click here](#) to see the new Variation Report design!

## NM\_007294.3(BRCA1):c.5080G>T (p.Glu1694Ter)

Variation ID: [?](#) 55387  
Review status: [?](#) ★ ★ ★ ☆ reviewed by expert panel

### Interpretation [?](#)

Go to: [⌵](#) [⌶](#)

Clinical significance: [Pathogenic](#)  
Last evaluated: Sep 8, 2016  
Number of submission(s): 10  
Condition(s):

- Breast-ovarian cancer, familial 1 [\[MedGen - OMIM\]](#)
- Hereditary breast and ovarian cancer syndrome [\[MedGen - Orphanet - OMIM\]](#)
- Hereditary cancer-predisposing syndrome [\[MedGen\]](#)

[See supporting ClinVar records](#) [↗](#)

### Allele(s) [?](#)

Go to: [⌵](#) [⌶](#)

#### NM\_007294.3(BRCA1):c.5080G>T (p.Glu1694Ter)

Allele ID: 70054  
Variant type: single nucleotide variant  
Cytogenetic location: 17q21.3  
Genomic location:

- Chr17: 43063946 (on Assembly GRCh38)
- Chr17: 41215963 (on Assembly GRCh37)

Protein change: E1694\*  
HGVS:

- NG\_005905.2:g.154038G>T
- NM\_007294.3:c.5080G>T
- NP\_009225.1:p.Glu1694Ter

[...more](#)

# Exome sequencing: applicazioni

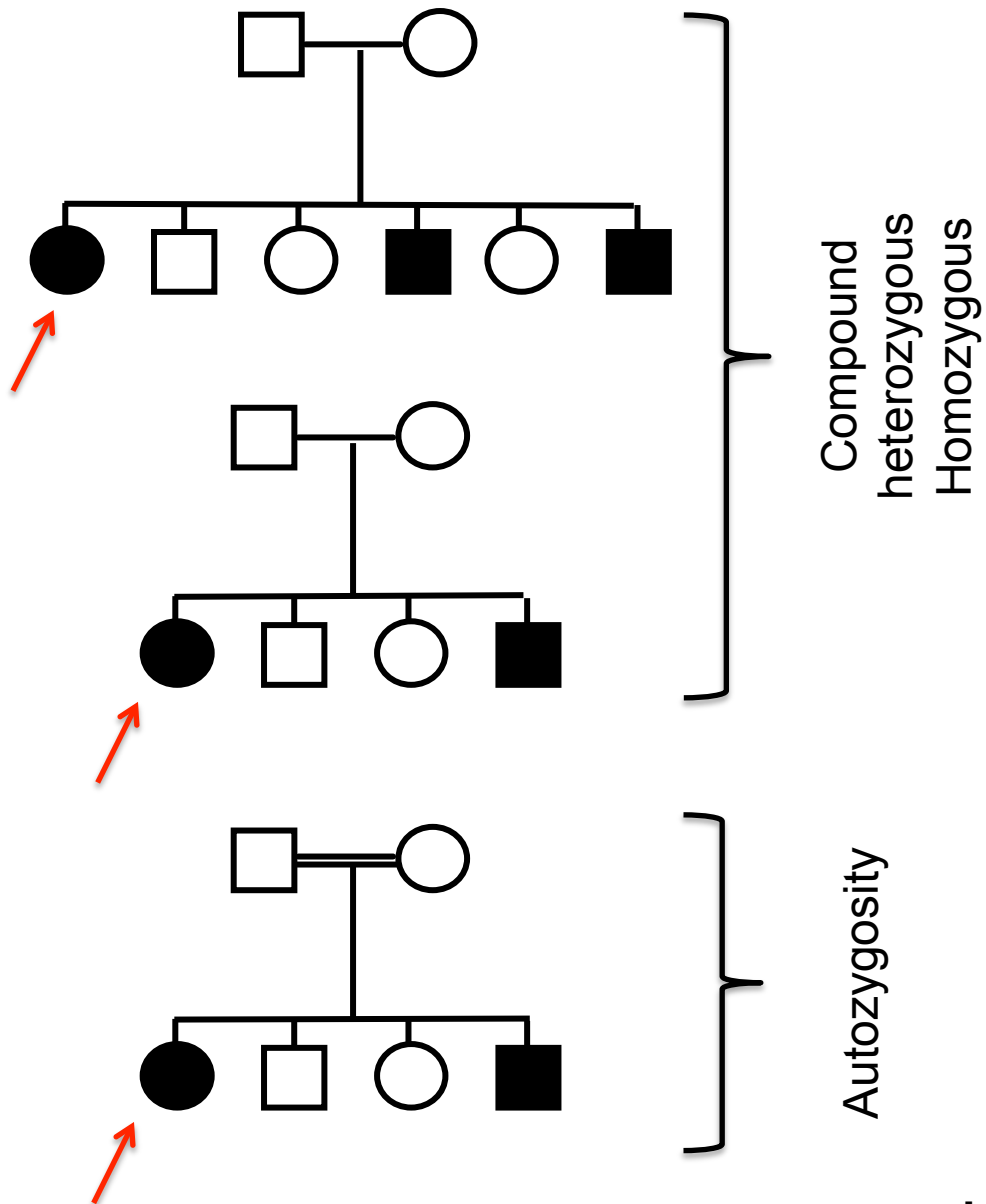
- Malattie mendeliane estremamente **rare** (non applicabili gli studi di linkage) con **pochi casi** derivanti da famiglie non imparentate
- Molti affetti nella stessa famiglia
- Casi sporadici dovuti a **mutazioni de novo**
- Malattie geneticamente e fenotipicamente **eterogenee**
- Come integrazione a

Studi di linkage (in famiglie numerose o con molti affetti)

Homozygosity map (malattie recessive in consanguinei)

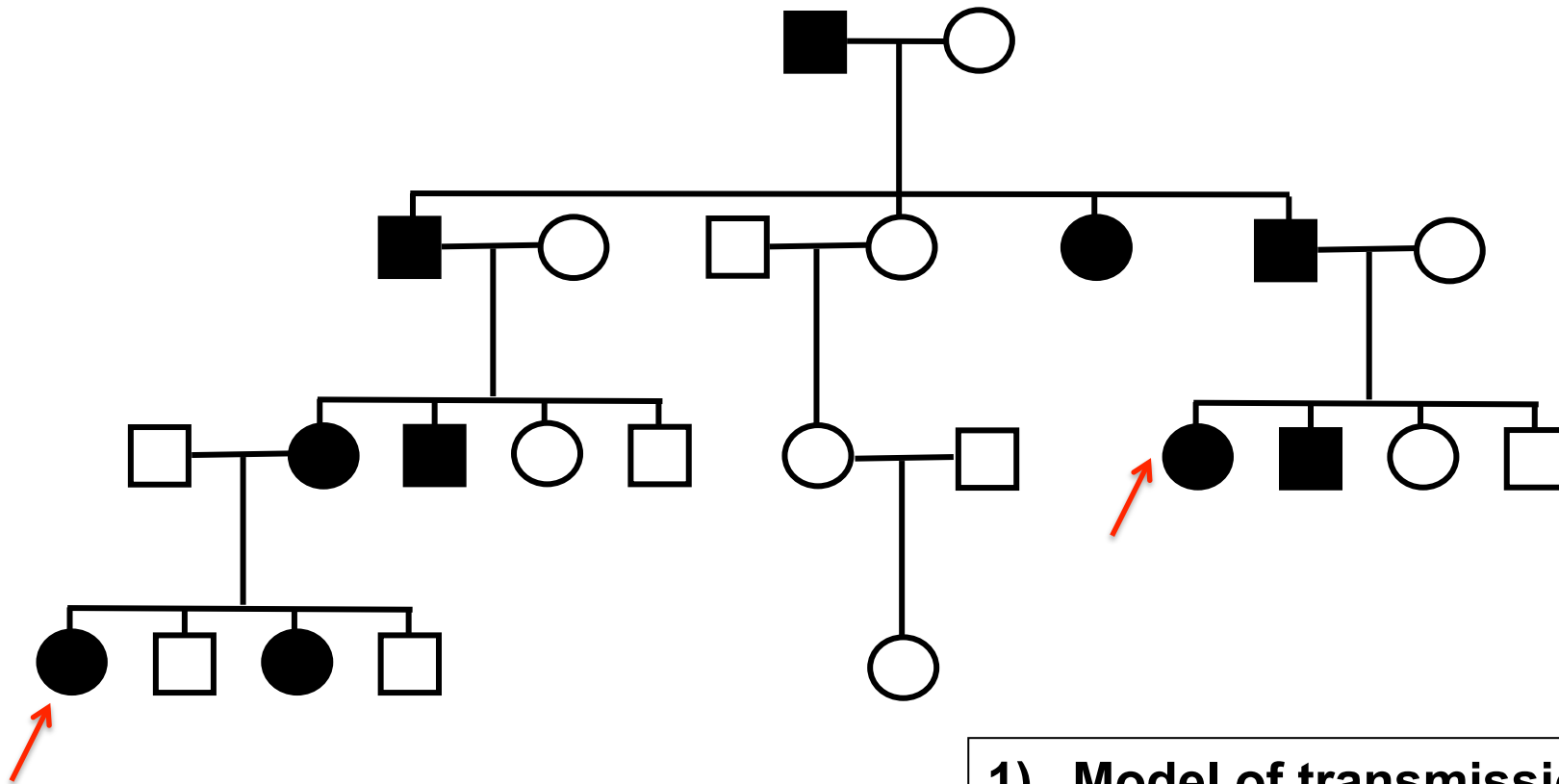
Come si utilizza l'ereditarietà per identificare il gene





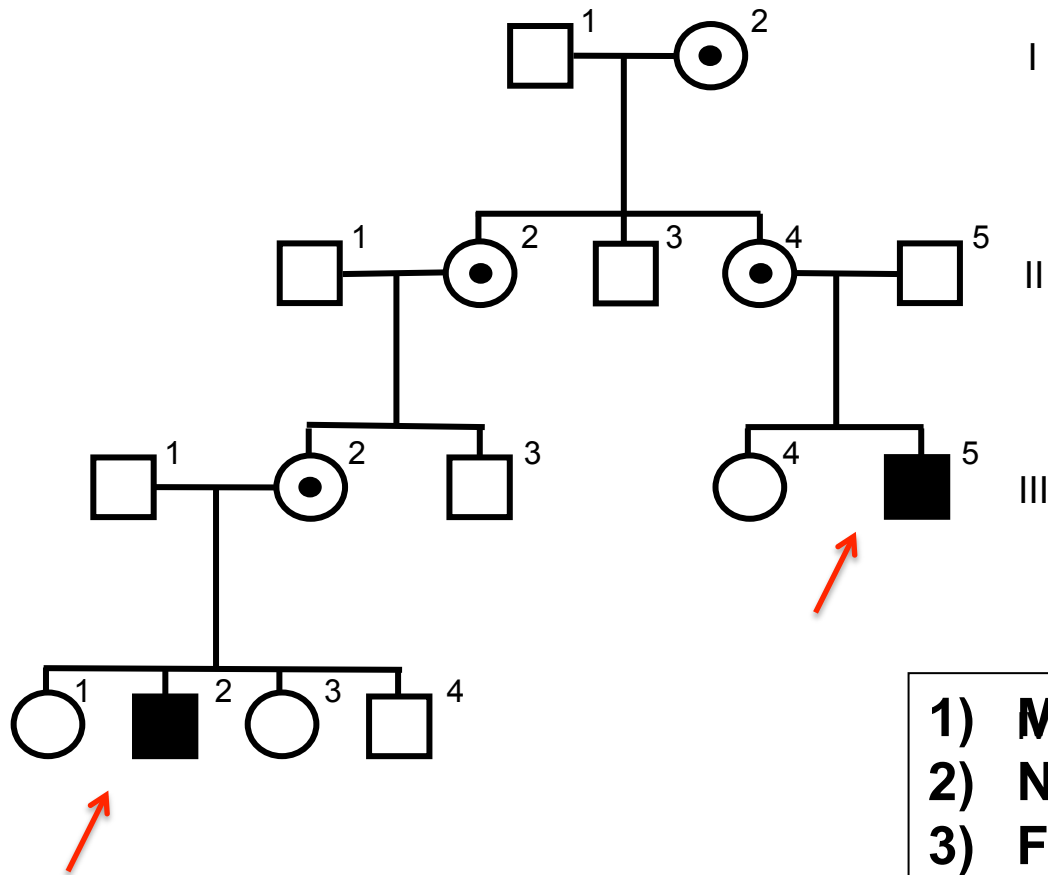
- 1) Model of transmission: AR
- 2) NGS in probands
- 3) Filtering
- 4) List of pathogenic variants
- 5) Variants affecting the same gene(s) in all the families studied
- 6) Both alleles affected (homozygous or compound heterozygous)
- 7) Segregation analysis

... what in case of genetic heterogeneity?

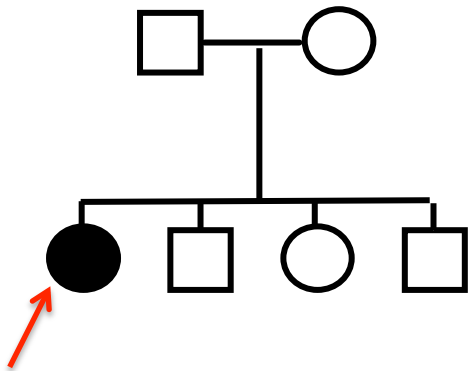
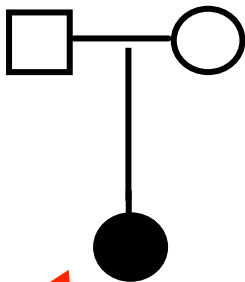
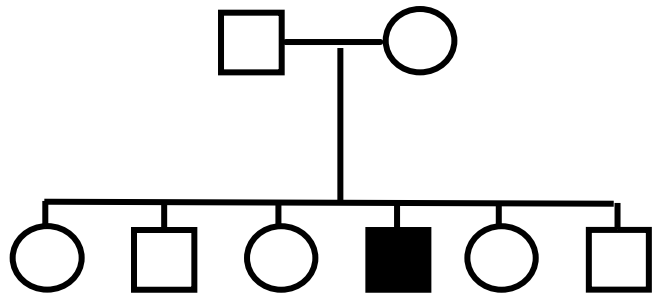


- Large family
- Combination with linkage analysis
- Pathogenic variants only in the candidate region

- 1) Model of transmission: AD
- 2) NGS in the most distant relative
- 3) Filtering
- 4) List of pathogenic variants
- 5) **Common heterozygous variants**
- 6) Segregation analysis



- 1) Model of transmission: X-LR
- 2) NGS in the probands
- 3) Filtering
- 4) List of pathogenic variants
- 5) **Common hemizygous variants**
- 6) Segregation analysis



- 1) Model of transmission: sporadic (de novo mutations)
- 2) NGS in probands and their relatives
- 3) Filtering and list of pathogenic variants
- 4) Variants not transmitted by parents
- 5) Variants affecting the same gene(s) in the probands

**Trio analysis**

## Exome sequencing: problemi

- Sperimentalmente, anche con un alto coverage, **non si riesce a sequenziare tutto l'esoma**
- Le sonde si basano su RefSeq e CCDS (consensus coding sequence): ci sono quindi **esoni non ancora annotati che non sono inclusi**
- Sono identificate solo variazioni nelle regioni codificanti e nei siti di splicing **quindi si escludono dall'analisi:**
  - promotori/enhancers
  - miRNA
  - elementi regolatori
  - sequenze non codificanti ma evolutivamente conservate

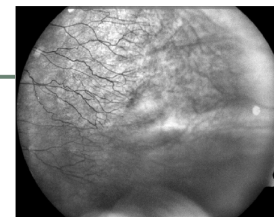
## Verso whole genome sequencing

- L'NGS fa più **errori di sequenziamento** rispetto al metodo Sanger
  - Falsi +
  - Falsi -

# ESEMPIO: ANALISI DI UNA MALATTIA

## REPORT

### Next-Generation Sequencing of a 40 Mb Linkage Interval Reveals *TSPAN12* Mutations in Patients with Familial Exudative Vitreoretinopathy



Konstantinos Nikopoulos,<sup>1,2,10</sup> Christian Gilissen,<sup>1,2,10</sup> Alexander Hoischen,<sup>1,2</sup> C. Erik van Nouhuys,<sup>4</sup> F. Nienke Boonstra,<sup>5</sup> Ellen A.W. Blokland,<sup>1</sup> Peer Arts,<sup>1,2</sup> Nienke Wieskamp,<sup>1,2</sup> Tim M. Strom,<sup>6</sup> Carmen Ayuso,<sup>7,8</sup> Mauk A.D. Tilanus,<sup>3</sup> Sanne Bouwhuis,<sup>1</sup> Arijit Mukhopadhyay,<sup>1,2,9</sup> Hans Scheffer,<sup>1,2</sup> Lies H. Hoefsloot,<sup>1</sup> Joris A. Veltman,<sup>1,2</sup> Frans P.M. Cremers,<sup>1,2,\*</sup> and Rob W.J. Collin<sup>1,2,3</sup>

## Malattia OCULARE EREDITARIA

**FENOTIPO:** sviluppo anormale della retina

**EREDITARIETA'** : autosomica dominante, autosomica recessiva, X-linked

3 geni noti coinvolti nella patologia:

*FZD4* on 11q14.2

*LRP5* on 11q13.2

*NDP* on Xp11.2

Analisi di linkage: **regione candidata in 40.5 Mb nel chr. 7**

Analisi dell' esoma compreso nei 40,5 Mb nel chr.7 allo scopo di identificare la variante associata alla malattia

Biological  
Background

Aim

# MATERIALI E METODI

Approach

**Sequenza degli esoni** all' interno della regione candidata (338 geni, 3048 esoni, 2.366.514 bp )

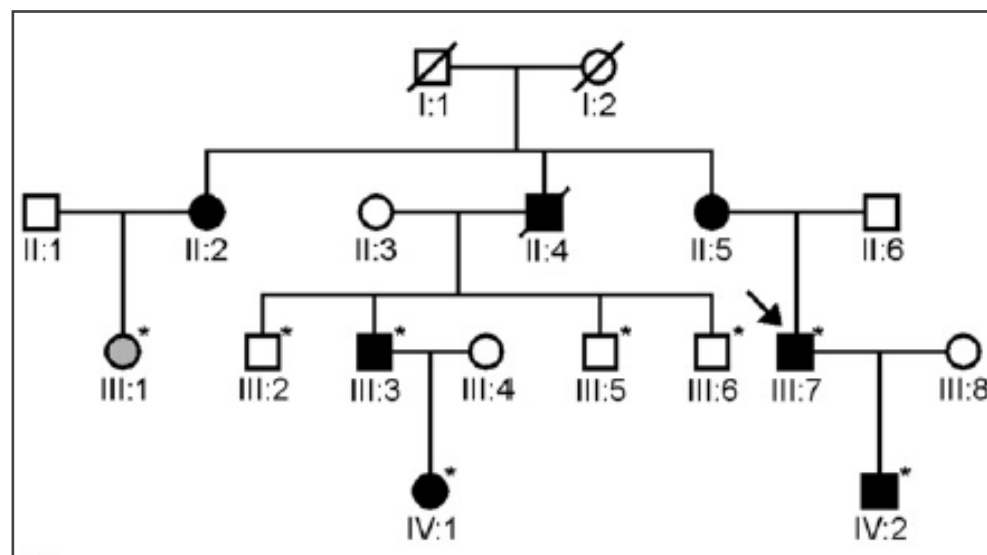


Samples

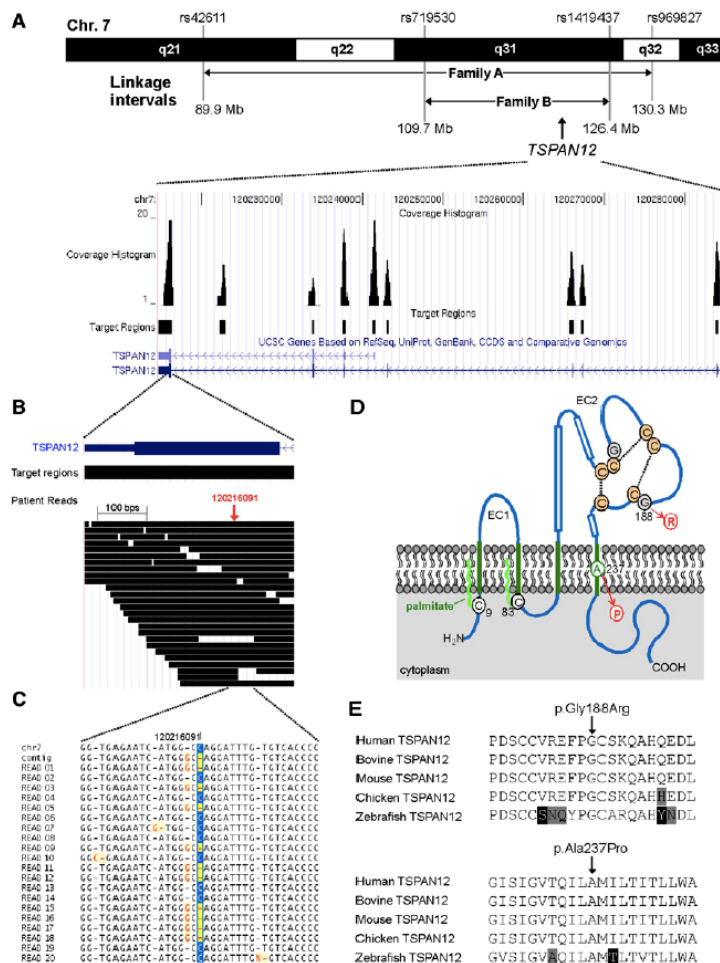
2 famiglie ( Fam A e Fam B) con la forma autosomica dominante con esclusione di mutazioni nei geni noti per la malattia (FZD4, LRP5 e NDP)

Sequencing Analysis

Fam A: la freccia indica l'individuo sottoposto a NGS



# RISULTATI



Per dettagli consultare l'articolo di riferimento  
Nikopoulos et al. An J Hum Genet 86:240, 2010  
(file pdf caricato come materiale didattico)

Exome in III-7 della Fam A:

1915 varianti totali identificate

1749 escluse perchè presenti in dbSNP

14 varianti esoniche, non sinonime e presenti in eterozigosi

3 geni candidati: *PTCD1*, *ZAN* e *TSPAN12*

**TSPAN12** (con variante p.Ala237Pro): gene localizzato nella regione candidata comune a Fam A e Fam B. TSPAN12 regola lo sviluppo vascolare della retina. Perciò, TSPAN12 è considerato il gene che causa la malattia in Fam A.

La stessa mutazione (p.Ala237Pro) è presente negli affetti della Fam B e in altre due famiglie (Fam C e Fam D).

Una seconda mutazione (p.Gly188Arg) di TSPAN12 è stata identificata in Fam D

**Conclusioni:** TSPAN12 è il quarto gene identificato responsabile della malattia



# Screening di mutazioni: alterazioni patogenetiche?

- **Frame-shift e nonsense** (generalmente si tratta di una mutazione patogenetica)
- **Missense (mutazione, polimorfismo, variante rara)**  
**(VUS: Variant of Uncertain Significance)**
  - Segregazione nella famiglia
  - Presenza in altre famiglie
  - Database (SNP, 1000 genomes, ExAC, GnomAD)
  - Conservazione amminoacido durante l'evoluzione
  - **Analisi bioinformatica** (Mutation Taster, PolyPhen2, SIFT)
  - **Struttura proteica**
  - **Test funzionali** (localizzazione cellulare - pathway biochimico, tranfezione in modelli cellulari, ecc.)

# CLONAGGIO PER CANDIDATO

Indipendente dalla conoscenza della localizzazione

Gli stessi criteri si applicano alla selezione dei geni nella regione  
candidata dopo analisi di linkage

## Gene Candidato?

Pattern di **espressione**

**Funzione** (membro di una famiglia, interattori, pathway molecolare, ...)

**Omologia** con gene implicato in una malattia simile

Esperimenti di **expression profiling** (controlli vs pazienti)

Gene (gene omologo) nei **modelli animali** o in vitro (**iRNA**) causa un  
fenotipo simile alla malattia in esame

Identificazione di **tripleste ripetute** se si sospetta un'espansione di  
trinucleotidi

**Subunità di complessi** implicati in pathway specifici

