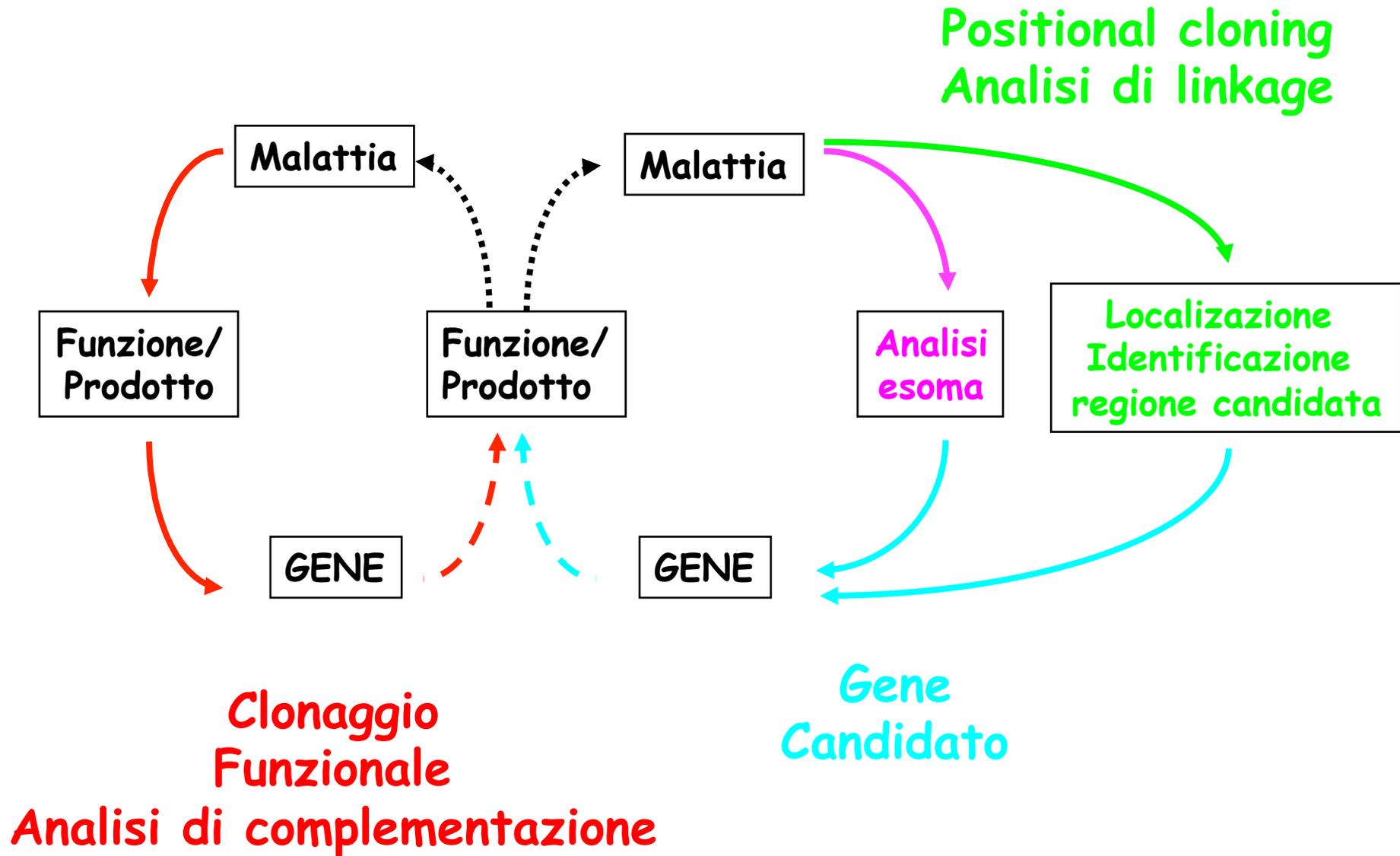
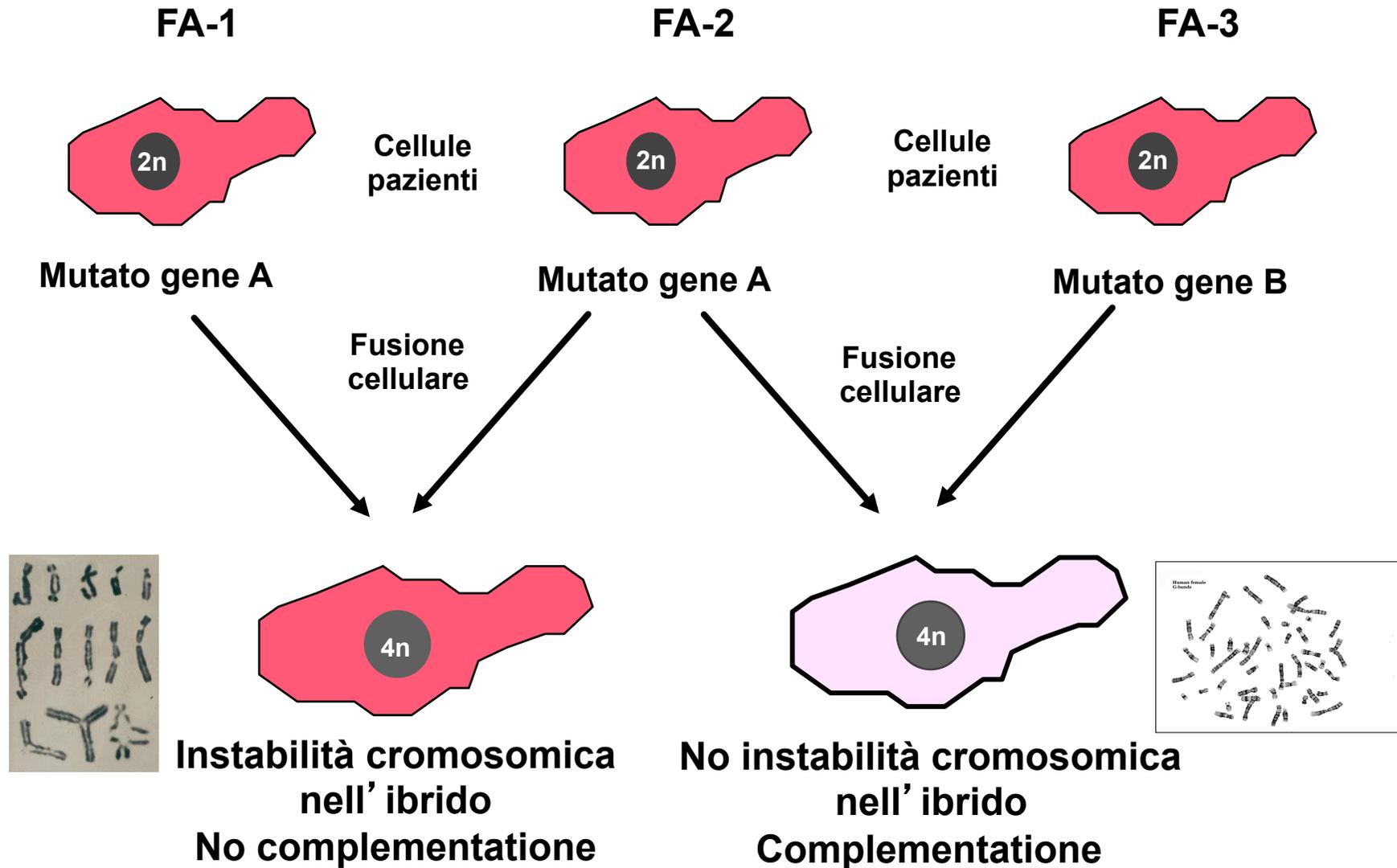


# STRATEGIE di CLONAGGIO



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



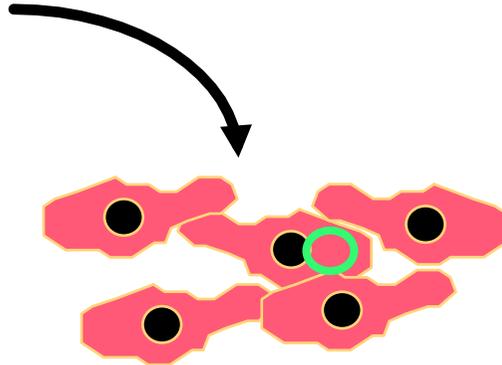
# Functional cloning

Using cDNA libraries



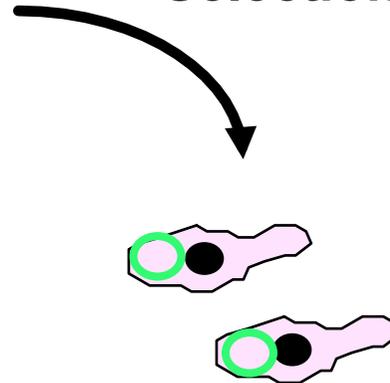
cDNA library

Transfection



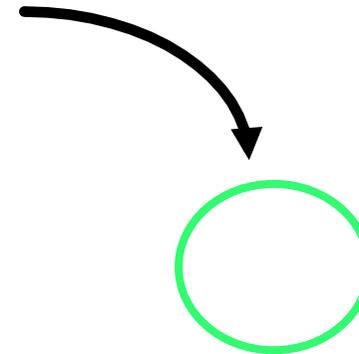
Mutant cells

Selection



Corrected cells

Isolation  
Episomal DNA



Isolation of gene  
leading to correction  
of phenotype

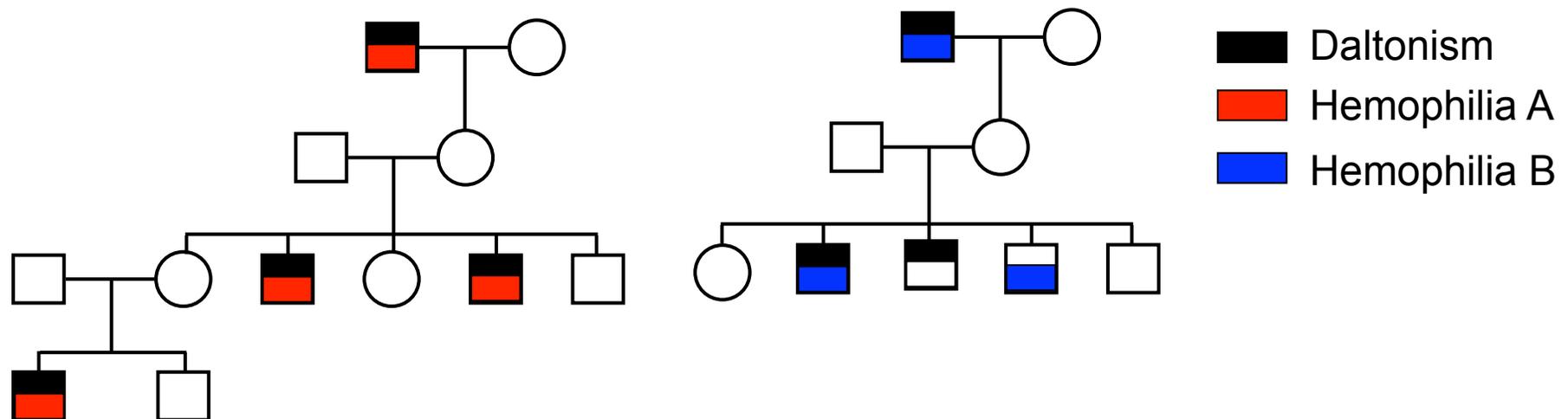
To prove that it is the disease-causing gene:  
1) identification of mutations in patients  
2) segregation analysis in families

# **Positional cloning:**

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

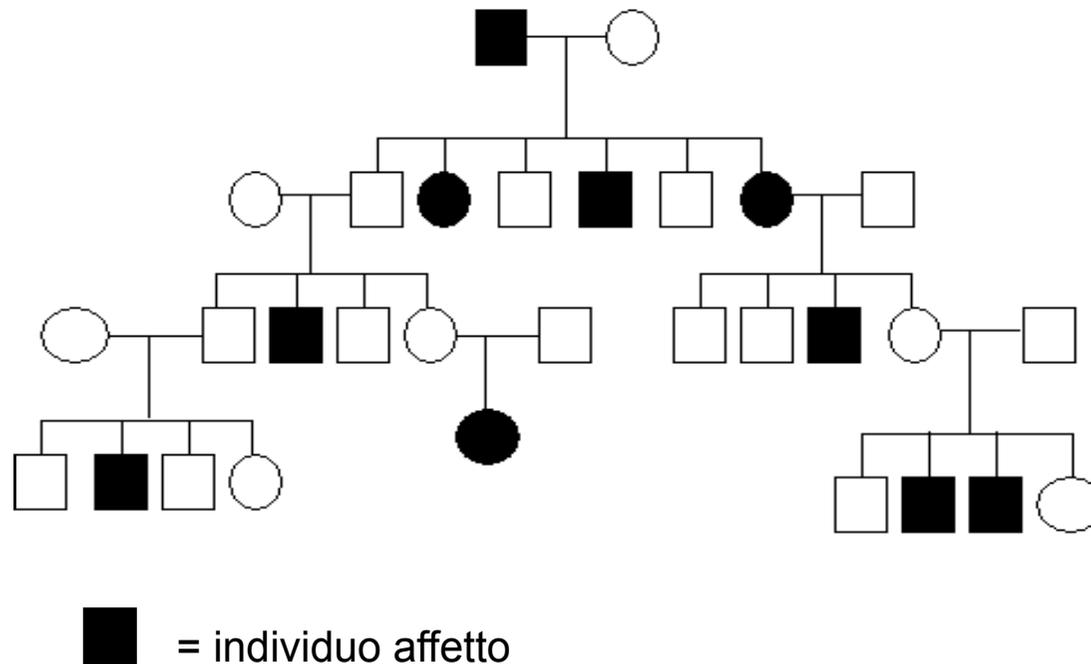
**Analisi di linkage:** permette determinare la posizione di un gene responsabile di una determinata malattia

- Determinare la frequenza di ricombinazione ( $\theta$ ) tra un **locus malattia e marcatori polimorfici a posizione nota**
- Se il locus malattia e il locus polimorfico si trovano su cromosomi diversi, segregano indipendentemente. La probabilità che vengano ereditati insieme è  $1/2$  ( $\theta = 50\%$ )
- Se sono sullo stesso cromosoma saranno ereditati insieme più frequentemente ( $\theta < 50\%$ ). Più sono vicini più il valore di  $\theta$  tenderà a 0

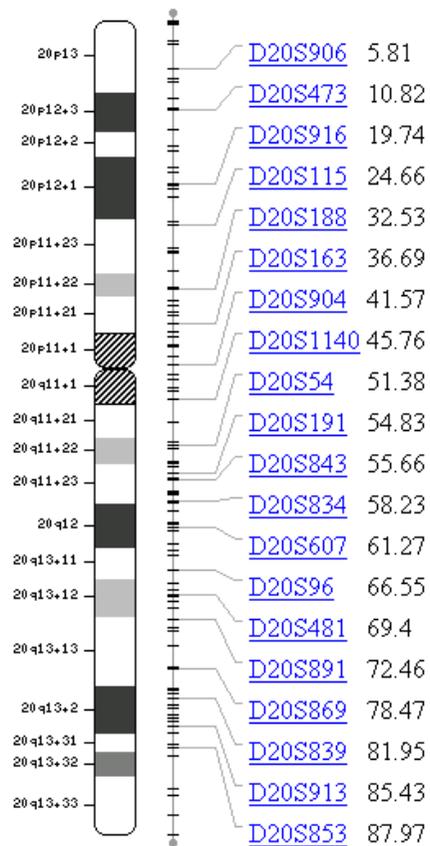


# Analisi di linkage: che cosa serve

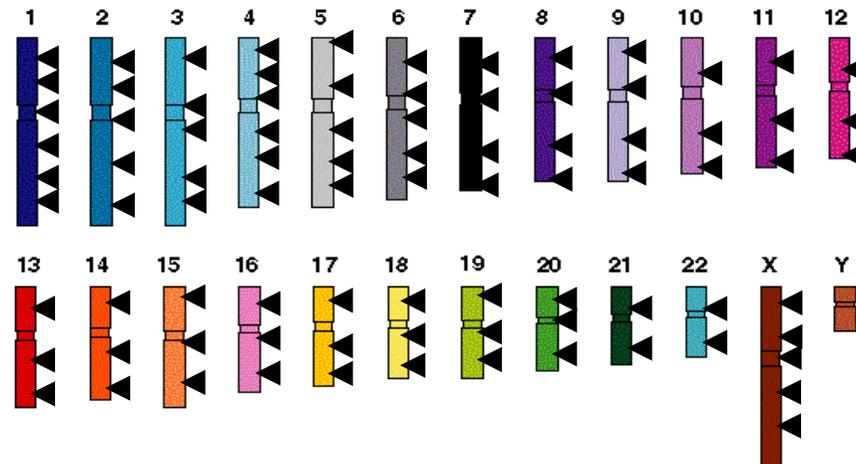
(1) Materiale biologico degli individui appartenenti ad una (ampia) o più (possibilmente omogenee geneticamente) **famiglie** in cui segrega il carattere/malattia



## (2) 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.

### **(3) Test statistico per calcolare la significatività**

**Z (LOD score)**

**Z > 3: linkage significativo**

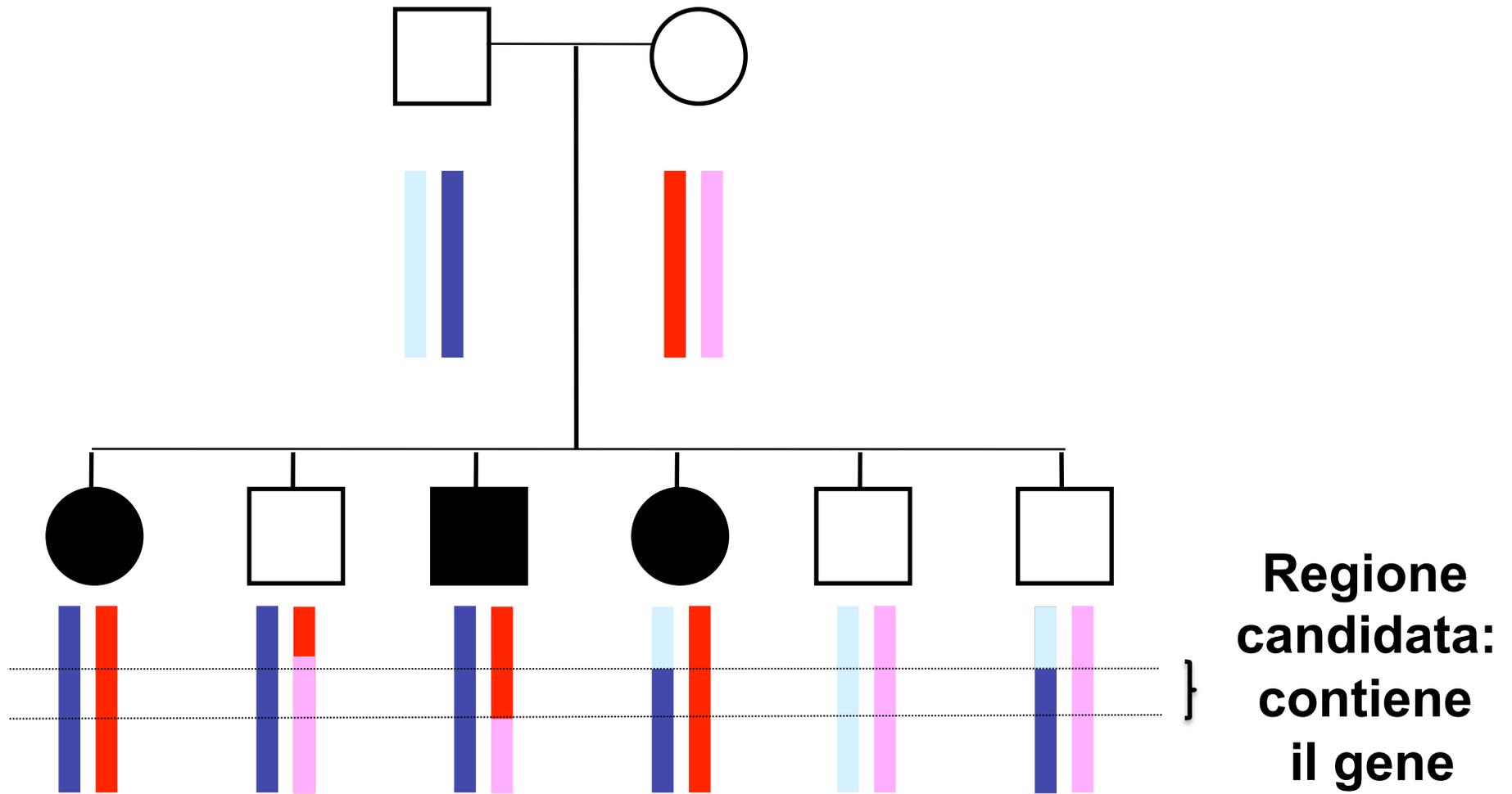
**-2 < Z < 3: non conclusivo**

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

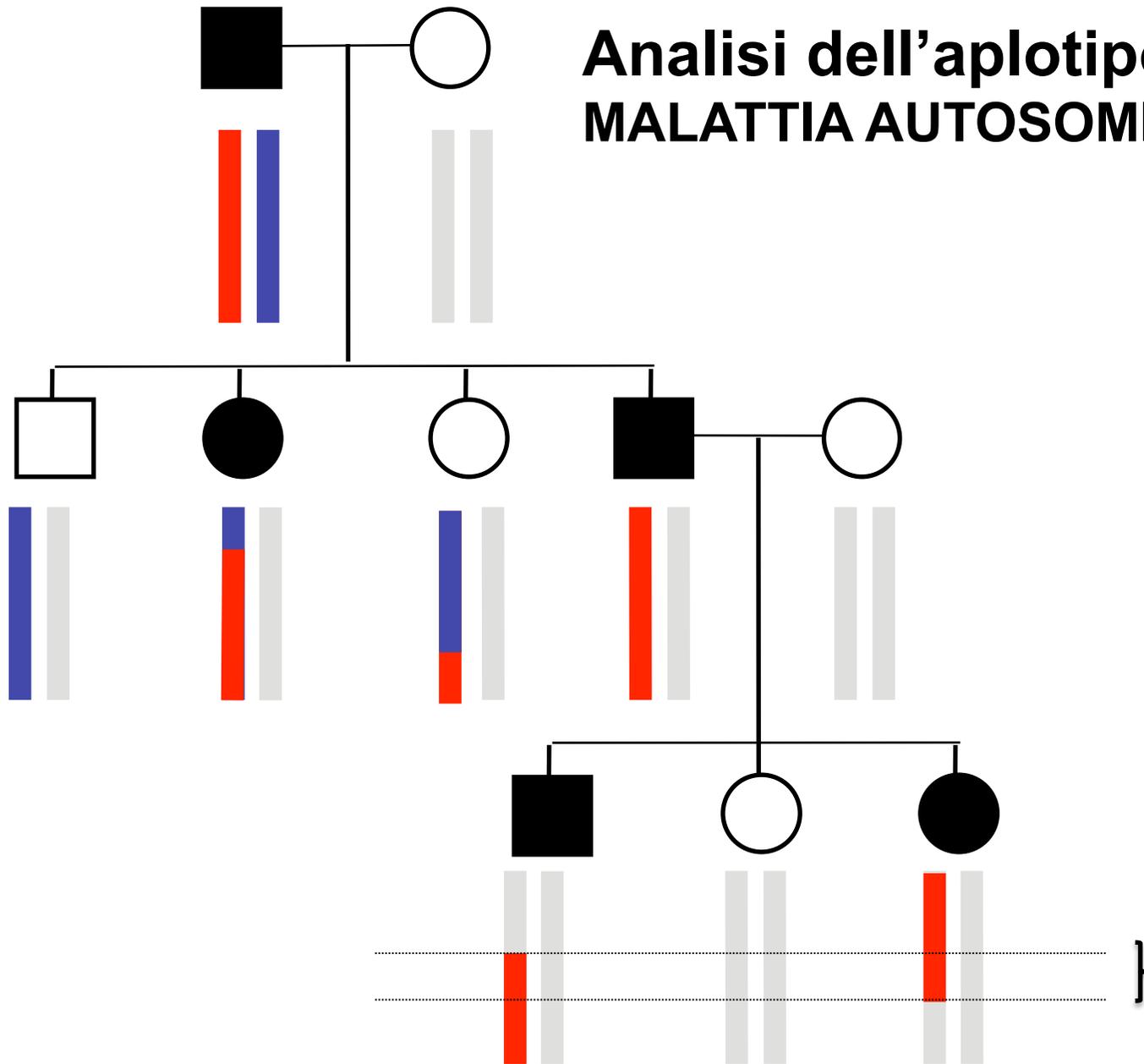
**Linkage tra carattere e marcatore X-linked**

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

# Analisi dell'aplotipo: MALATTIA AUTOSOMICA RECESSIVA

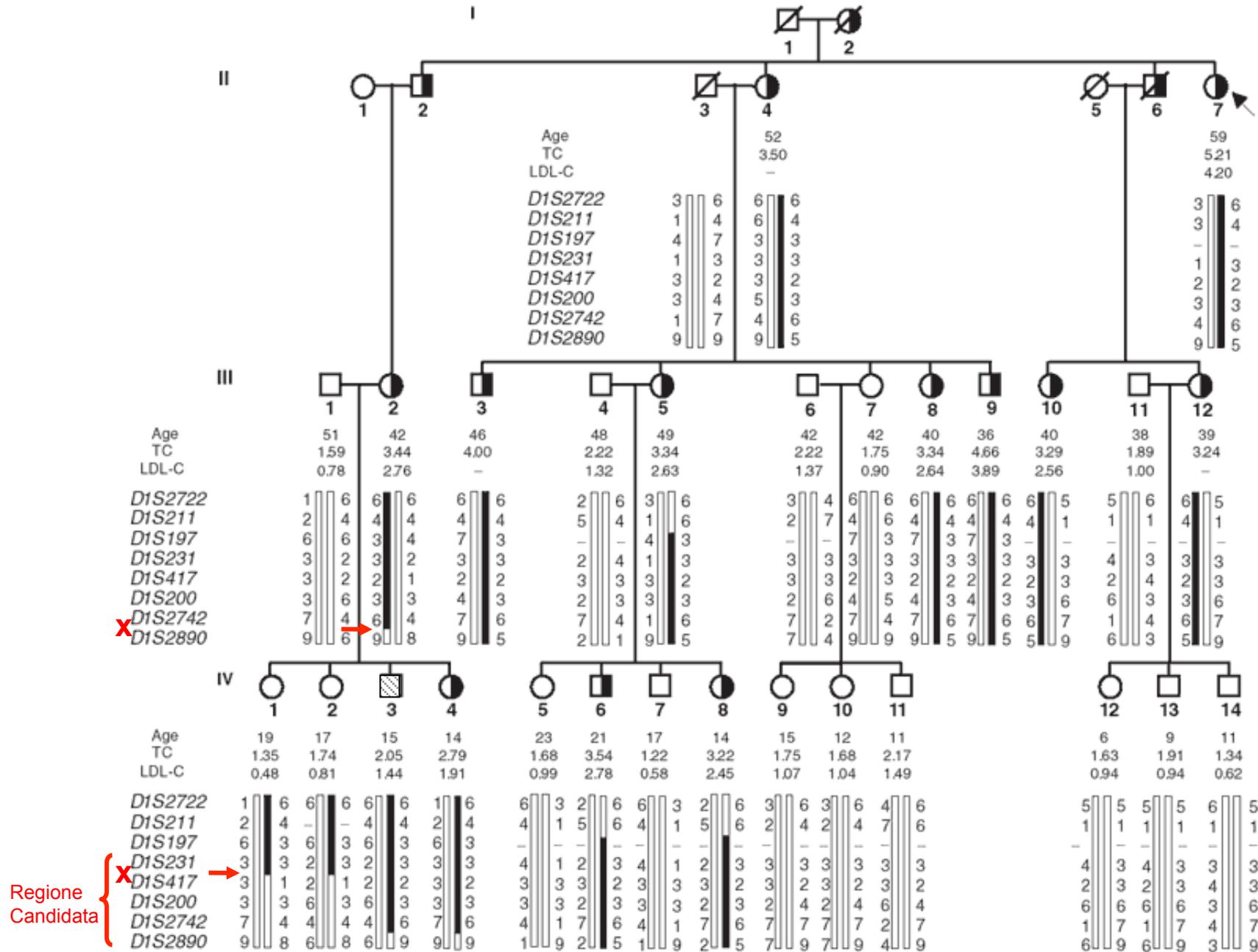


# Analisi dell'aplotipo: MALATTIA AUTOSOMICA DOMINANTE



**Regione  
candidata**

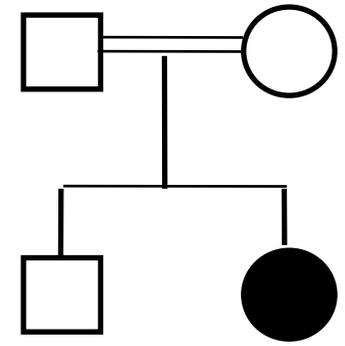
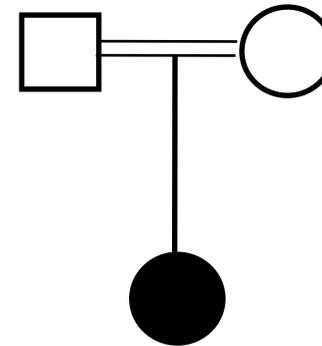
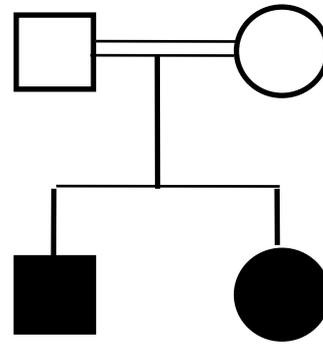
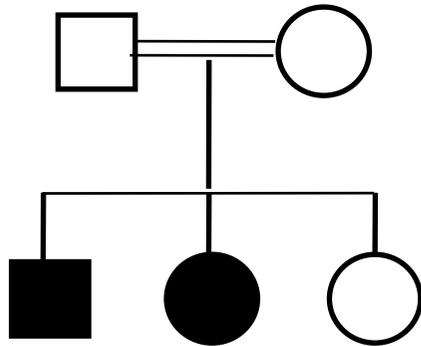
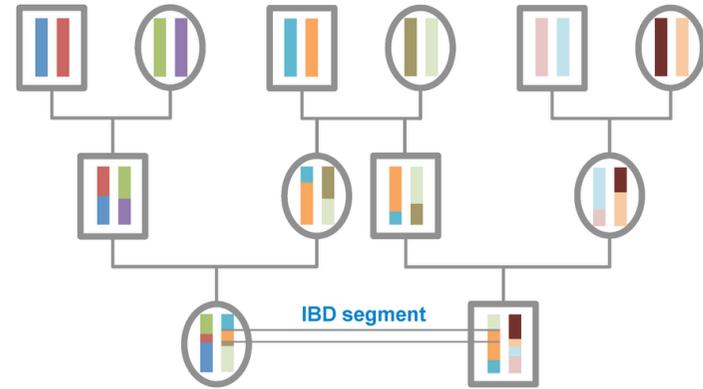
# COSTRUZIONE DI APLOTIPI: definizione della regione critica



# AUTOZIGOSITA'

(IBD: identical by descent)

Marcatori omozigoti perché ereditati da un comune antenato



Regione  
candidata

3	1	3	1	5	1
6	8	6	8	22	8
9	9	9	9	7	9
6	6	6	6	3	6
4	4	4	4	0	0

2	2	2	6
6	6	6	6
3	3	3	3
4	4	4	4
10	10	10	10

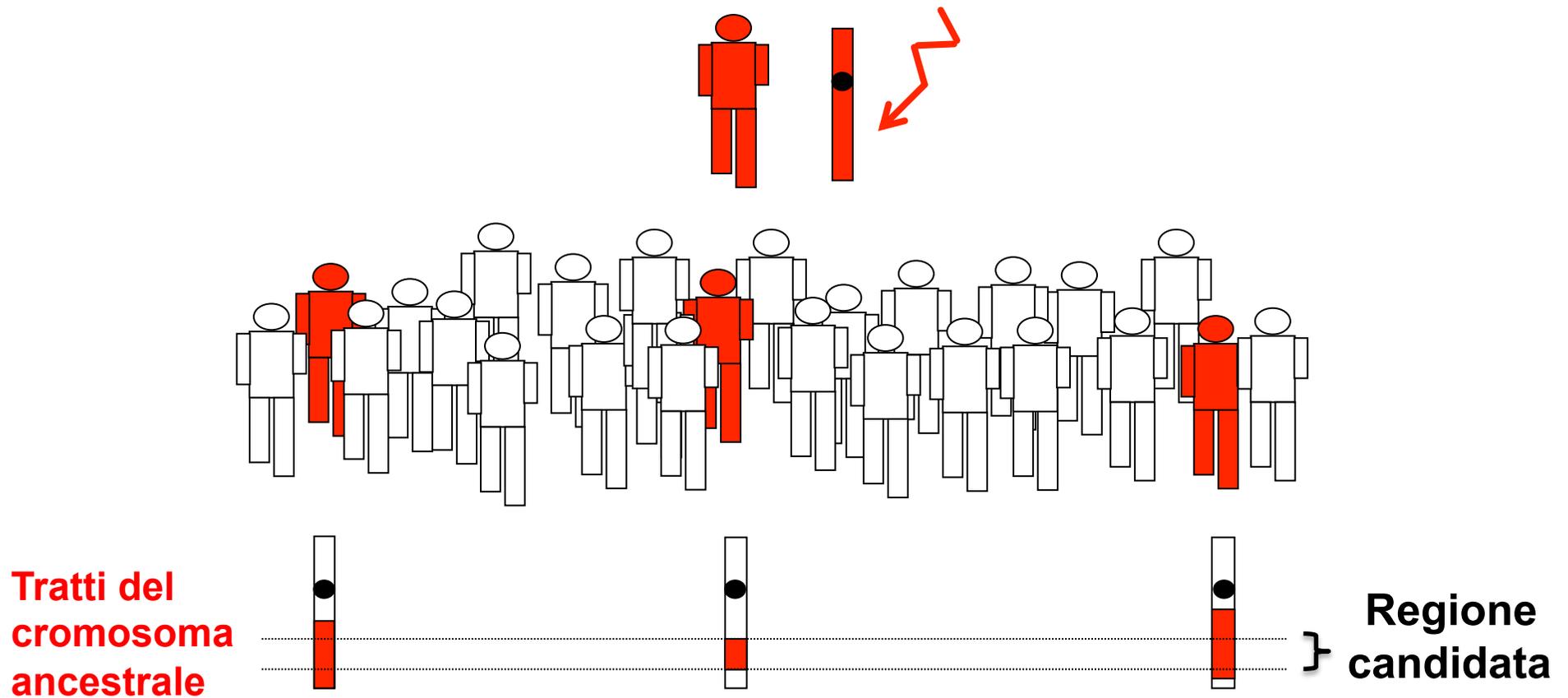
3	3
6	6
5	5
6	6
10	10

5	5
20	8
4	7
9	3
10	10

5	5
8	8
7	7
3	3
10	10

# EFFETTO FONDATORE

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



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

## ***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

# NGS (esoma o genoma): applicazioni nel clonaggio di geni malattia

- 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 che hanno definito una regione candidata

Sulla base del target NGS (genoma, esoma, regione candidata, genicandidati) l'analisi identifica circa una variante ogni kb

**Eliminare varianti poco probabilmente patogenetiche per selezionare quelle causative**

1

**Rarità**

Mutazione è rara  
non presente nei  
database



Eliminare varianti  
con  $MAF < 1$   
(Minor Frequency Allele)

2

**Patogenicità**

Nonsense/frameshit  
Missense: tools bioinformatici  
Sinonime non patogenetiche



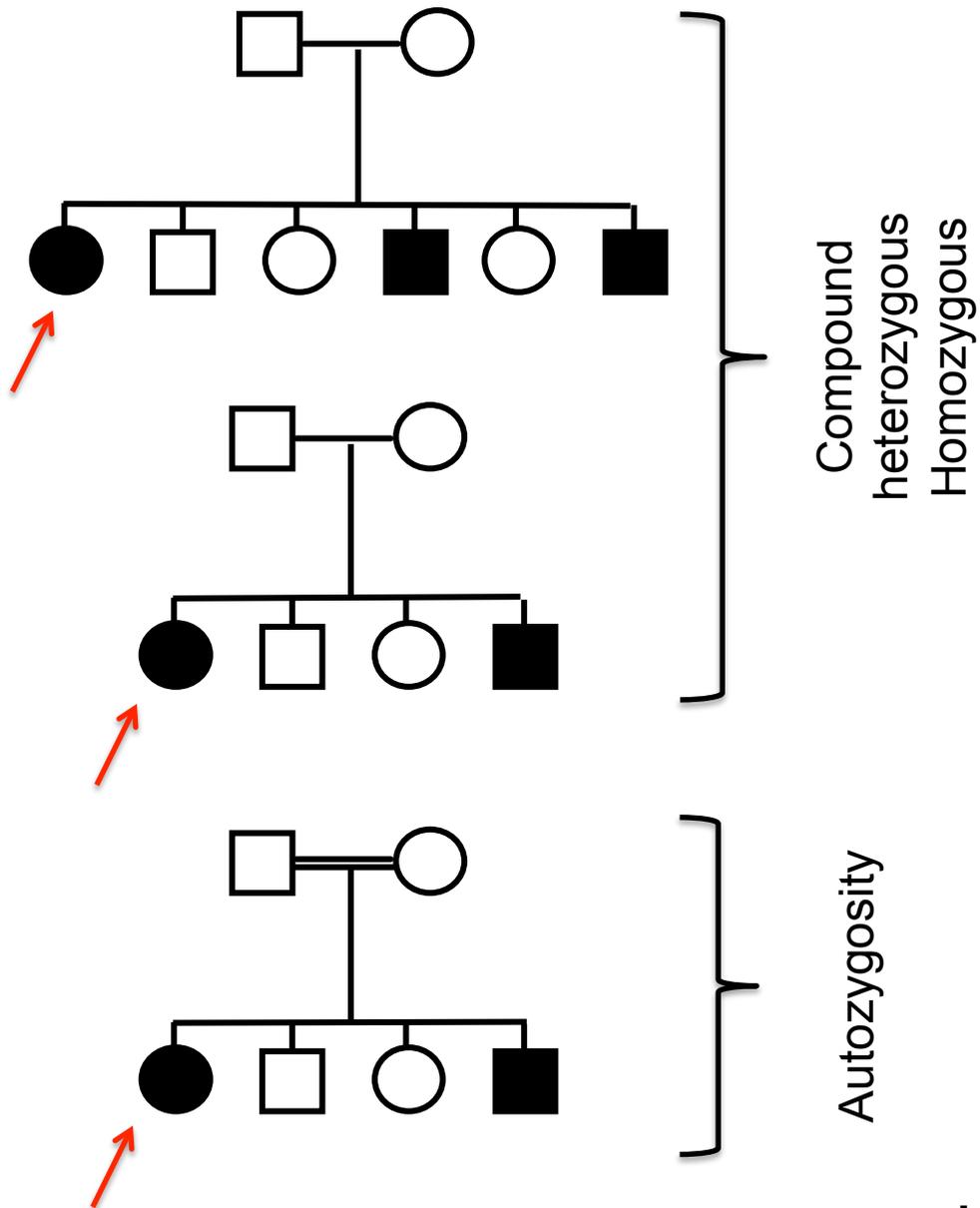
Eliminare varianti non  
patogenetiche

3

**Ereditarietà**

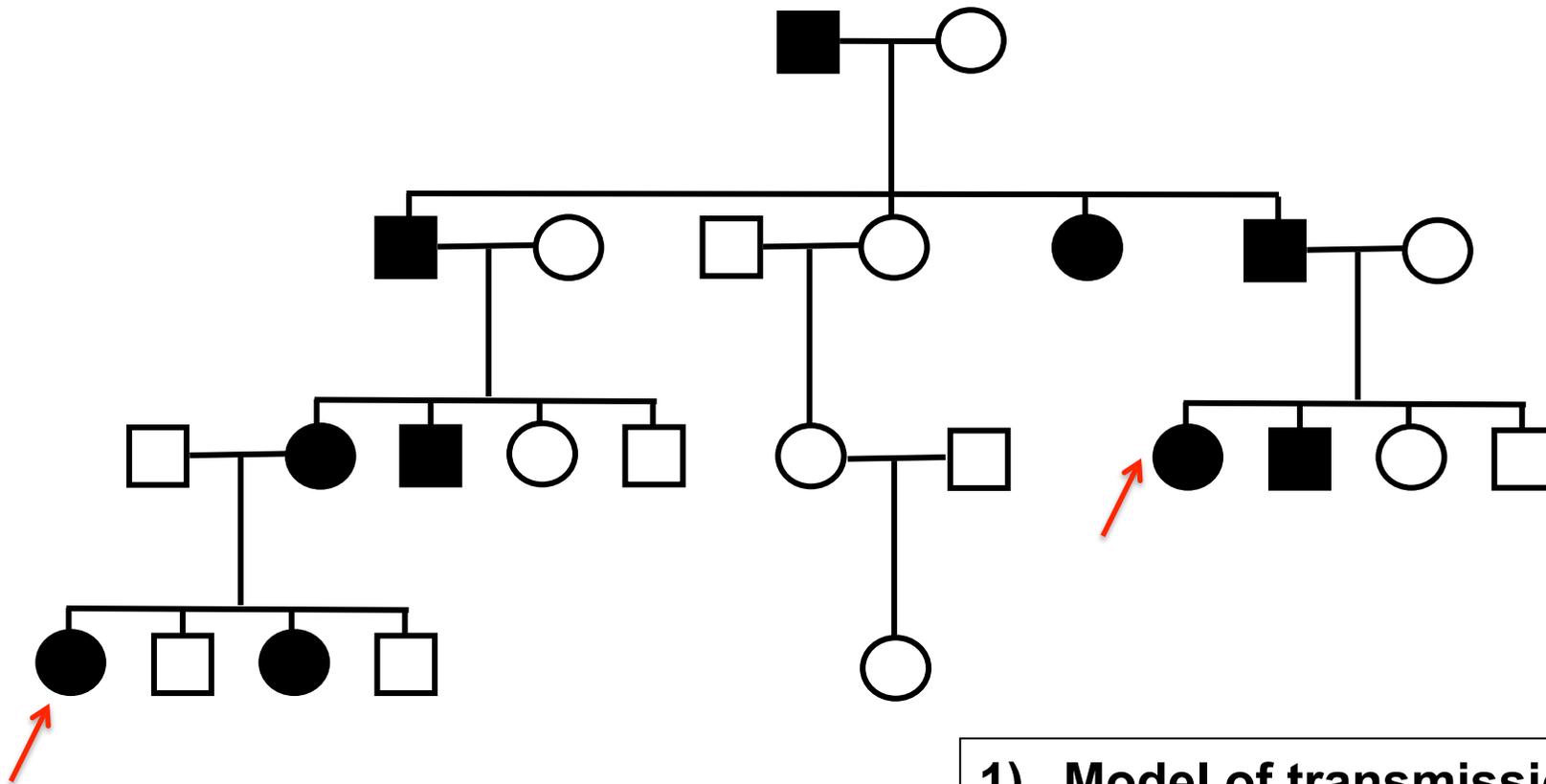


Verificare modello



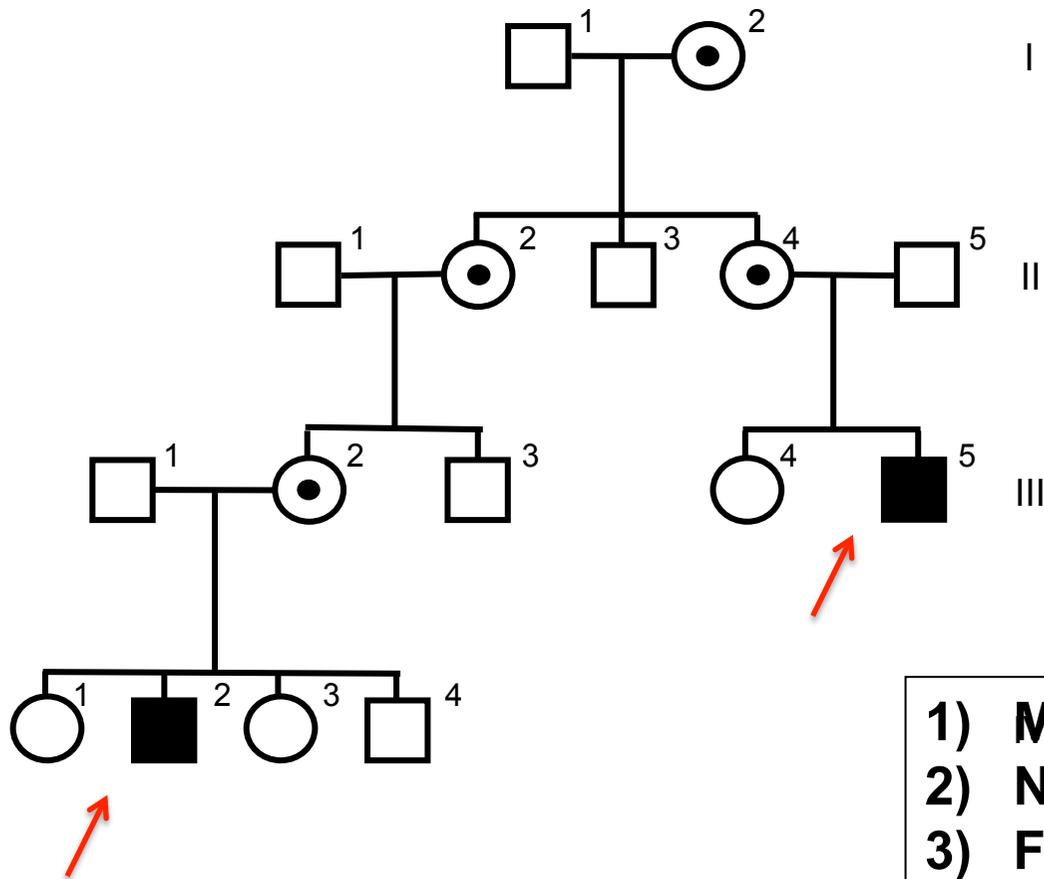
- 1) Model of transmission: AR
- 2) NGS in probands
- 3) Filtering
- 4) List of pathogenic variants
- 5) Variants affecting the same gene(s) in all the families studied
- 6) Both alleles affected (homozygous or compound heterozygous)
- 7) Segregation analysis

... what in case of genetic heterogeneity?

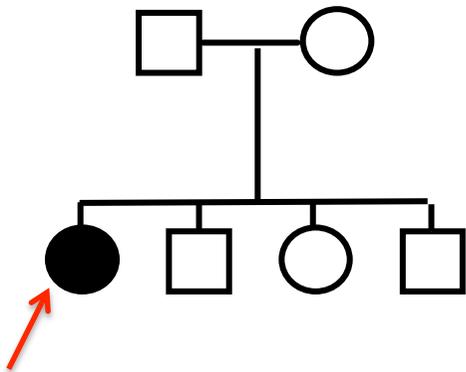
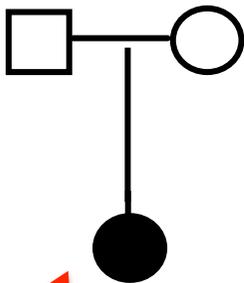
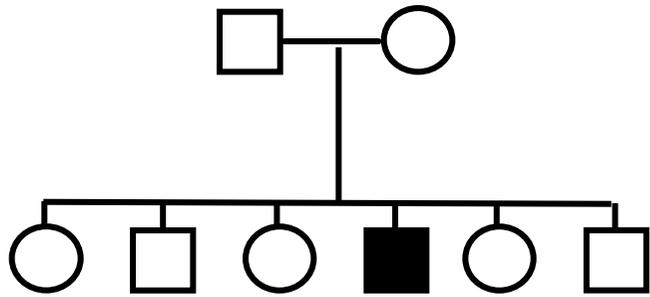


- Large family
- Combination with linkage analysis
- Pathogenic variants only in the candidate region

- 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 db quali RefSeq e CCDS (consensus coding sequence) quindi gli **esoni non ancora annotati non vengono 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

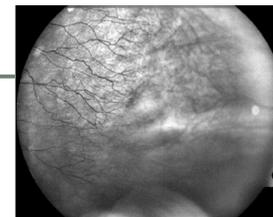
## Verso whole genome sequencing

- 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 )

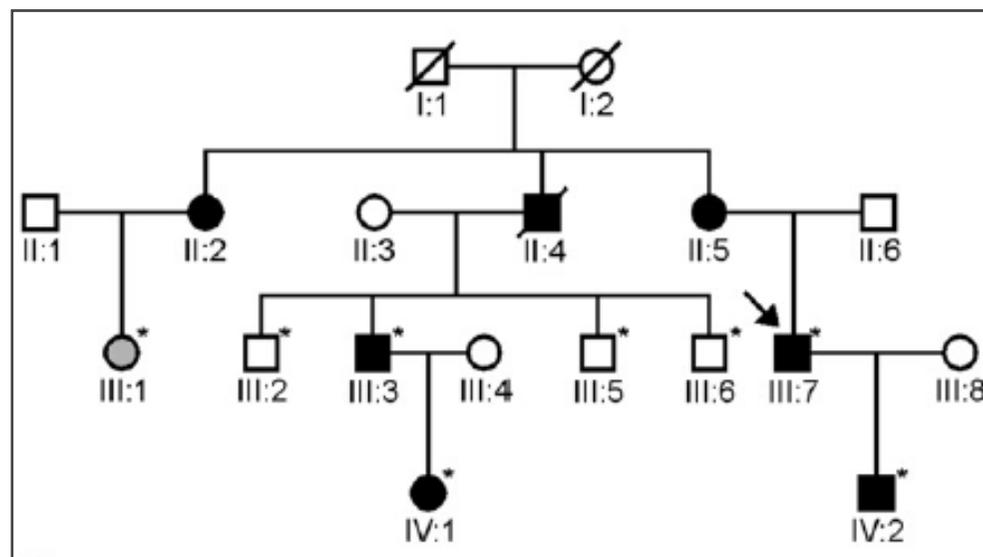


Samples

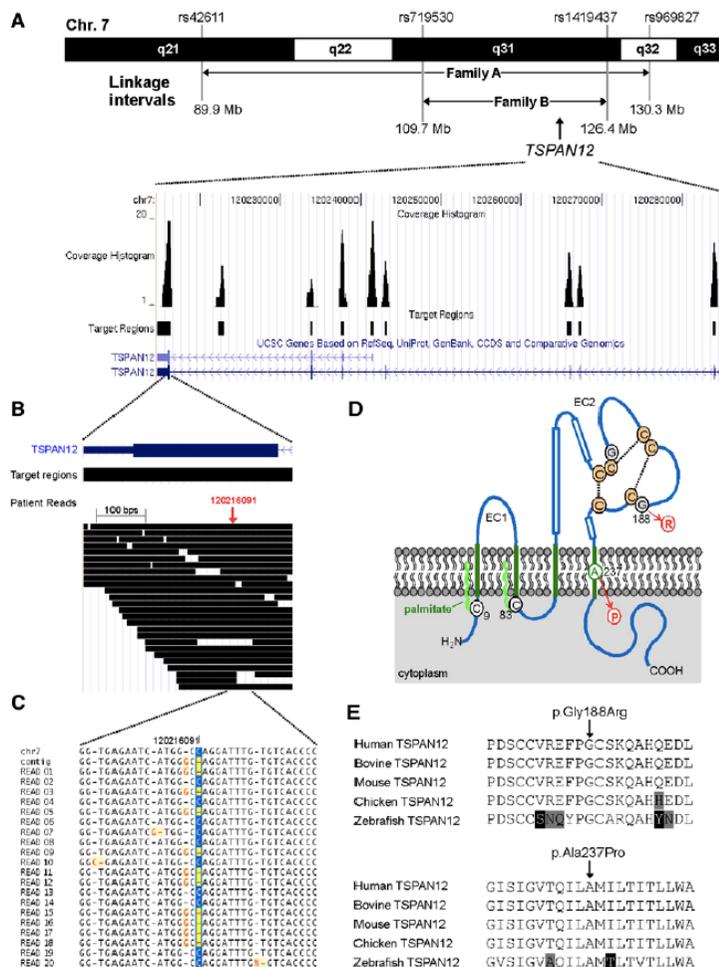
2 famiglie ( Fam A e Fam B) con la forma autosomica dominante con esclusione di mutazioni nei geni noti per la malattia (FZD4, LRP5 e NDP)

Sequencing Analysis

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



# RISULTATI



## 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

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

# 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
  - Cromosomi controllo (assenza mutazione)
  - Database (SNP, 1000 genomes, ExAC)
  - Conservazione amminoacido durante l'evoluzione
  - **Analisi bioinformatica** (Mutation Taster, PolyPhen2, SIFT)
  - **Struttura proteica**
  - **Test funzionali** (localizzazione cellulare - pathway biochimico, tranfezione in modelli cellulari, ecc.)