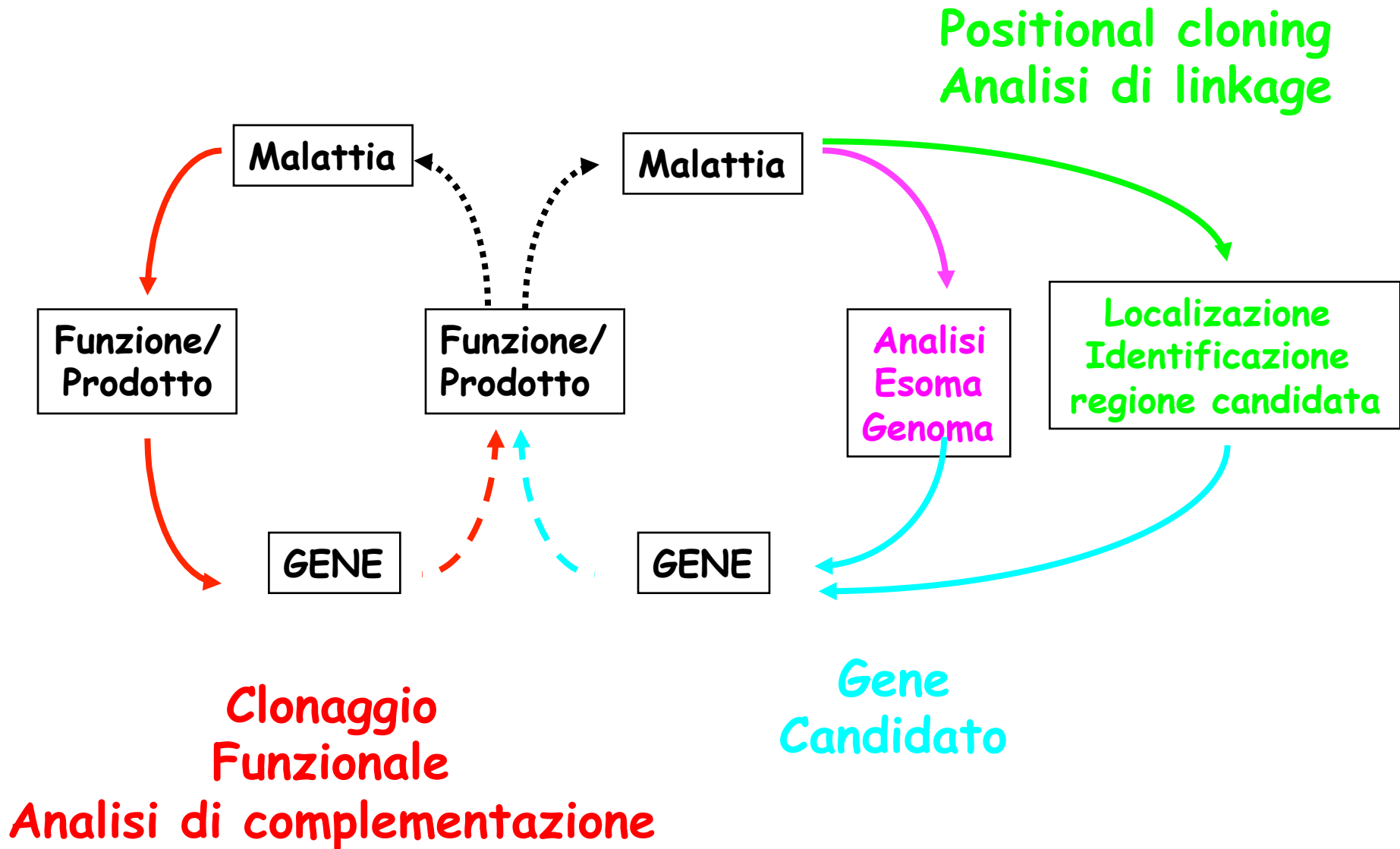
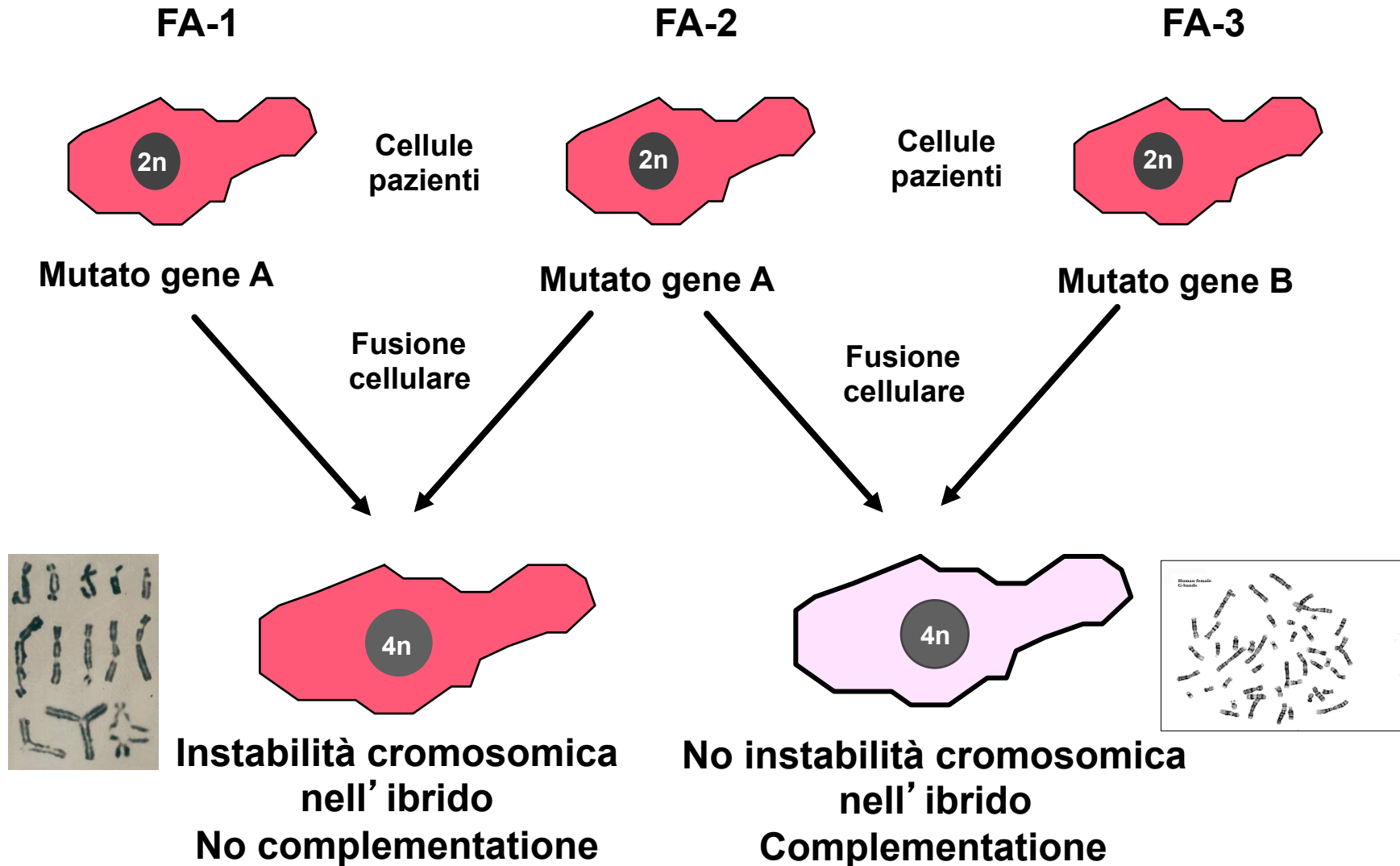


STRATEGIE di CLONAGGIO



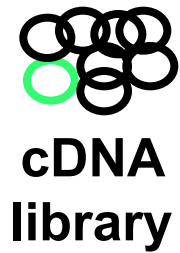
Analisi di complementazione:

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

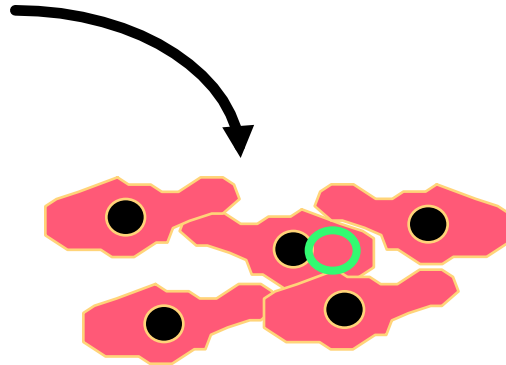


Functional cloning

Using cDNA libraries

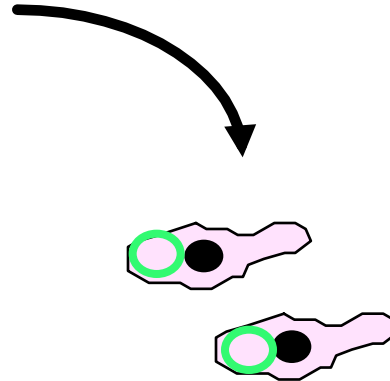


Transfection



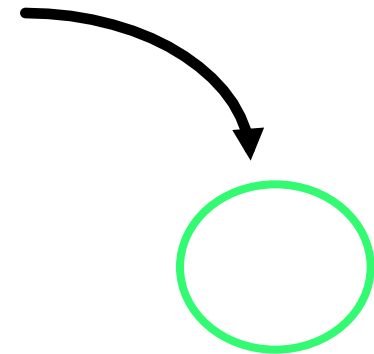
Mutant cells

Selection



Corrected cells

Isolation
Episomal DNA



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

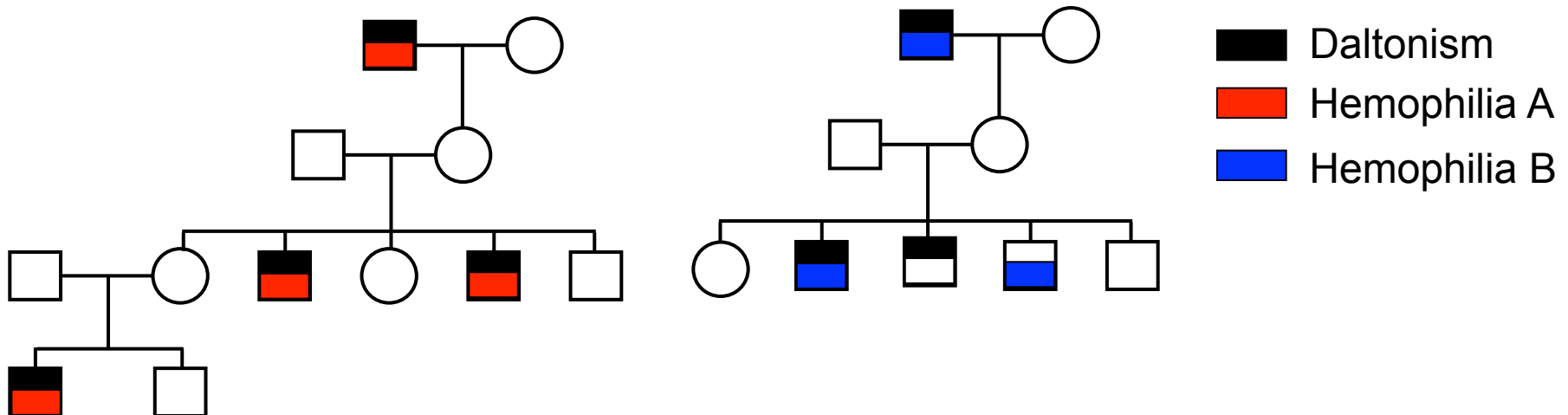
3) Screening for mutations

NGS – WES of candidate region

Sanger sequencing of candidate genes

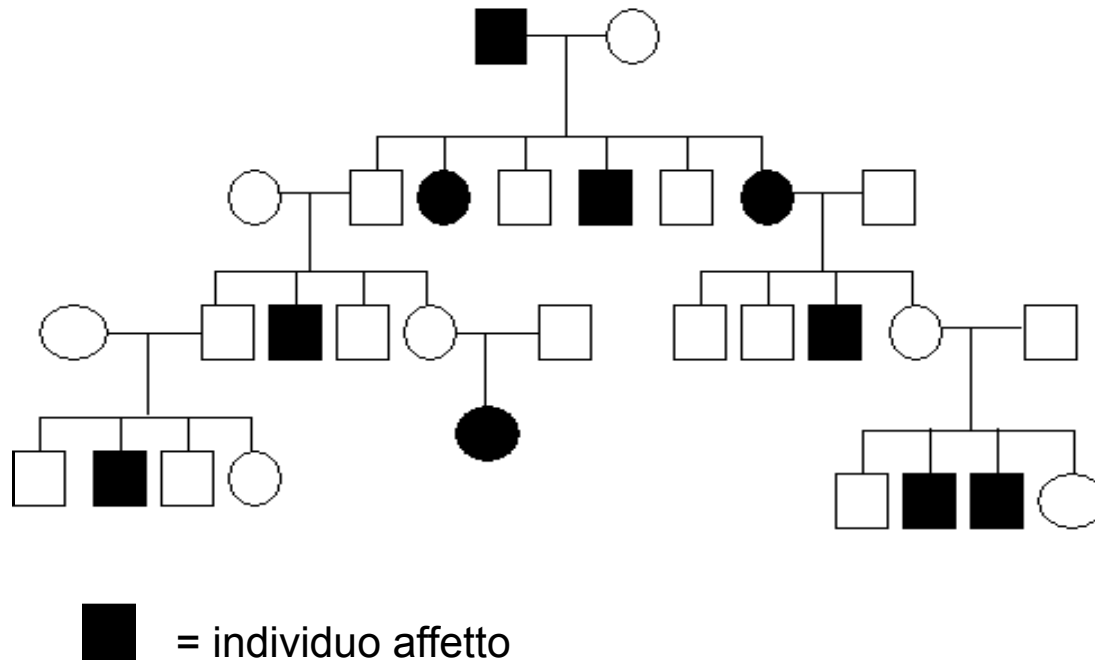
Analisi di linkage: permette determinare la posizione di un gene responsabile di una determinata malattia

- Determinare la frequenza di ricombinazione (θ) 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 θ tenderà a 0

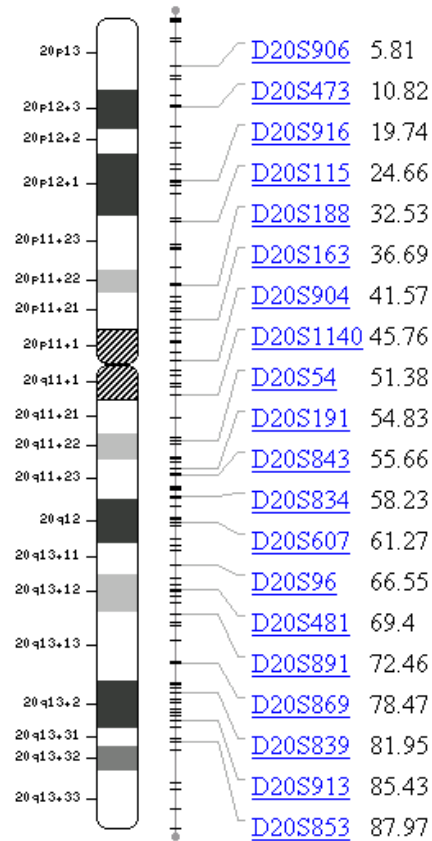


Analisi di linkage: che cosa serve

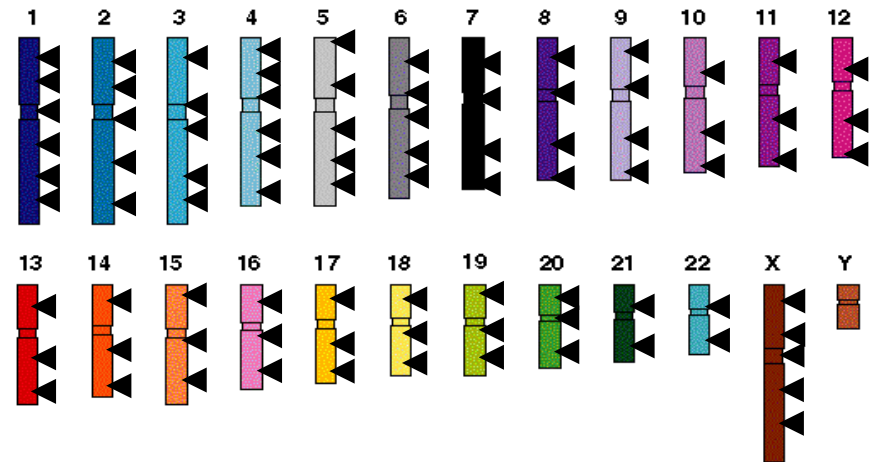
(1) Materiale biologico degli individui appartenenti ad una (ampia) o più (possibilmente omogenee geneticamente) **famiglie** in cui segrega il carattere/malattia



(2) Marcatori adeguati (microsatelliti o SNP) distribuiti uniformemente lungo tutti i cromosomi



Mappa genetica del cromosoma 2



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

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

(3) Test statistico per calcolare la significatività

LOD SCORE - $Z(\theta)$

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

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

**Frazione
di
ricombinazione**

$Z > 3$: linkage significativo

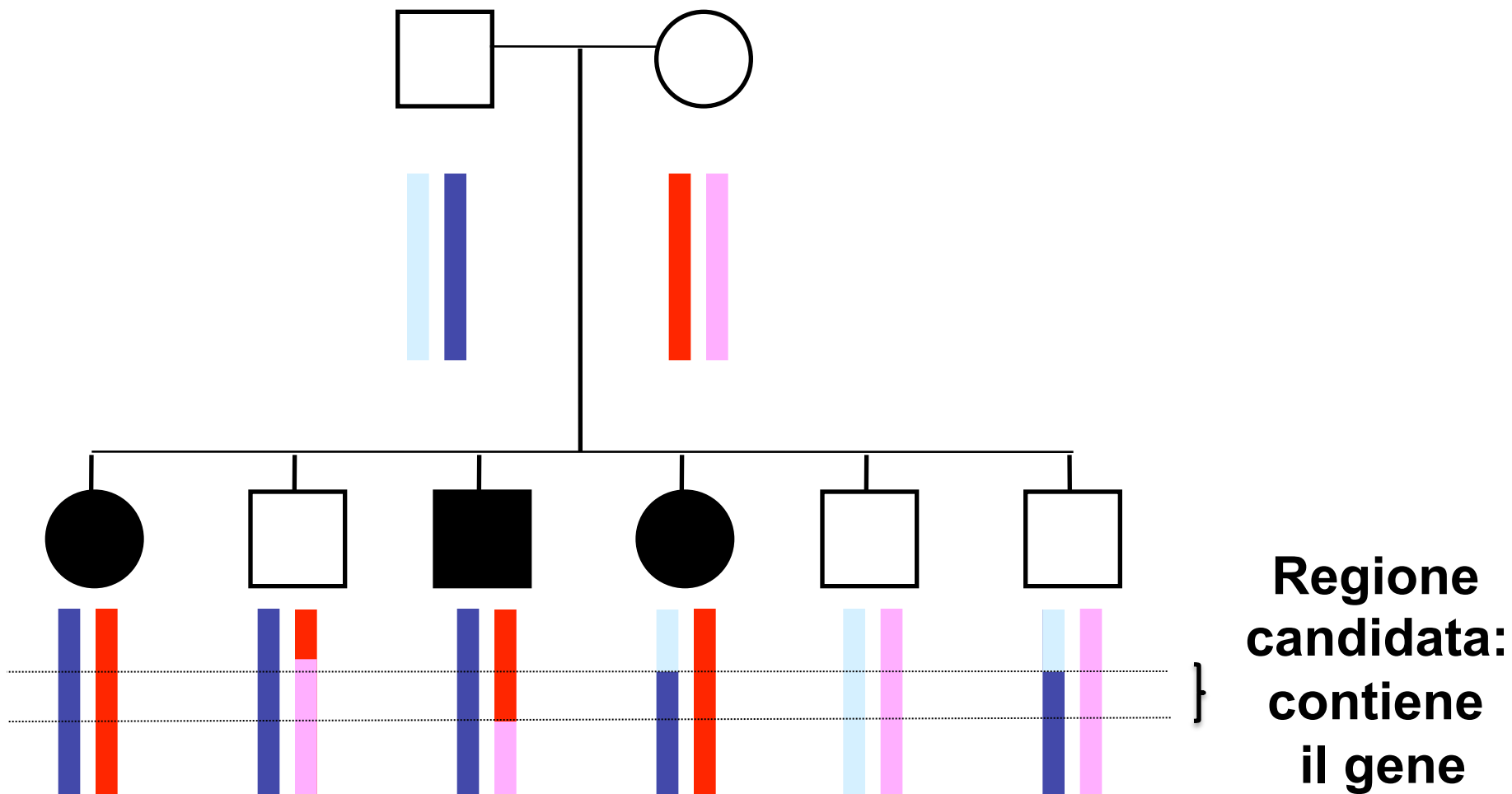
$-2 < Z < 3$: non conclusivo

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

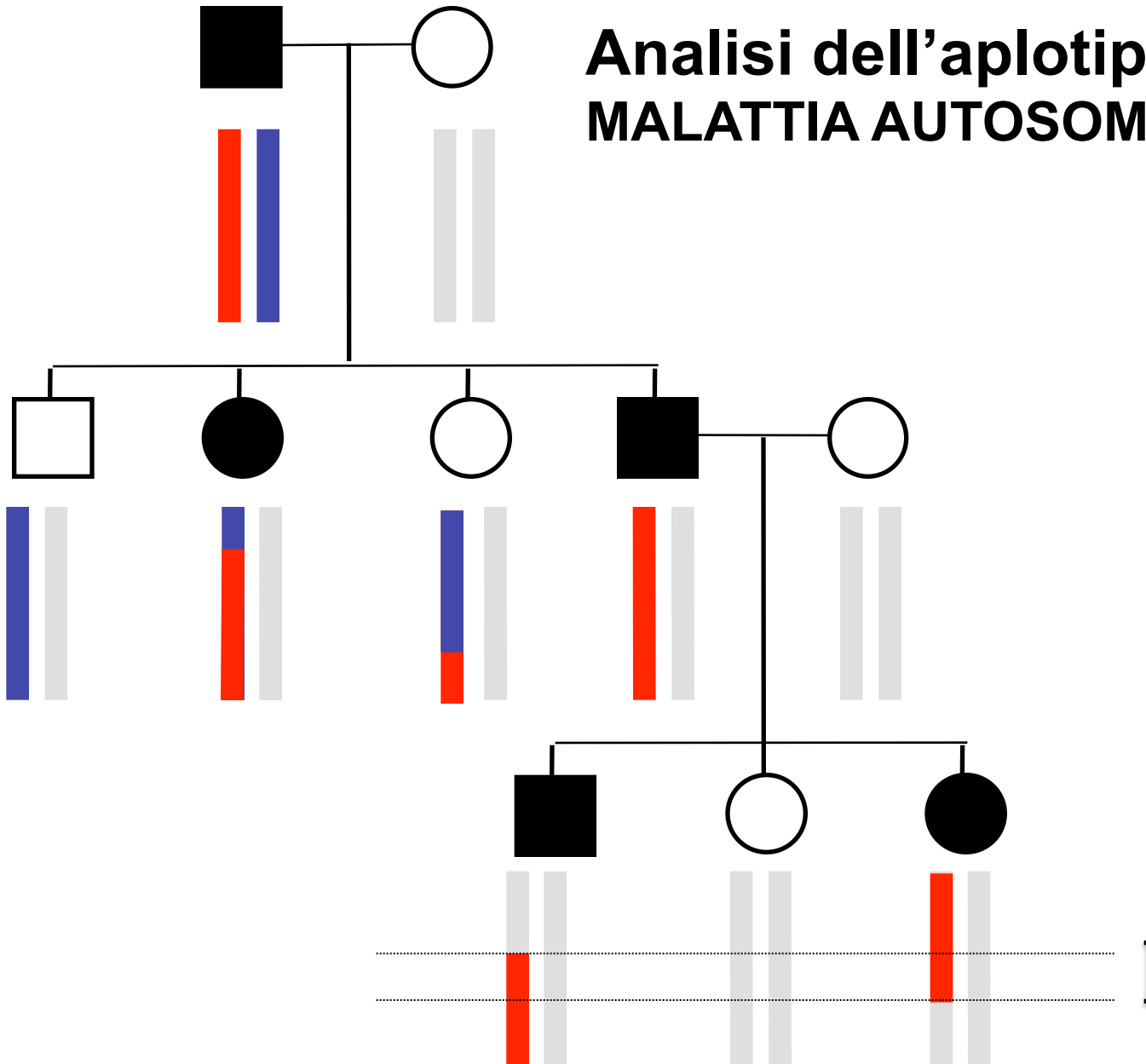
Linkage tra carattere e marcatore X-linked

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

Analisi dell'aplotipo: MALATTIA AUTOSOMICA RECESSIVA

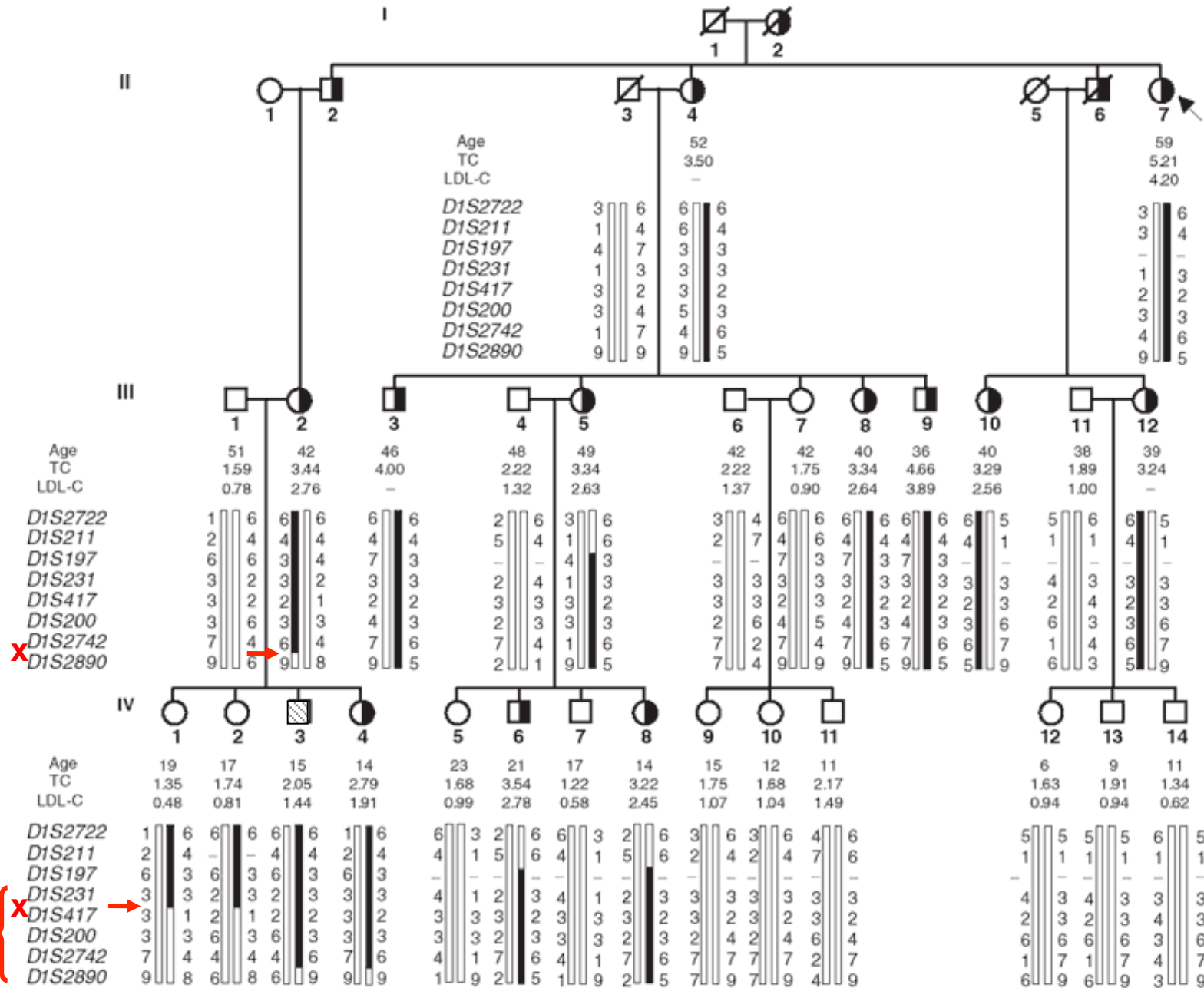


Analisi dell'aplotipo: MALATTIA AUTOSOMICA DOMINANTE



**Regione
candidata**

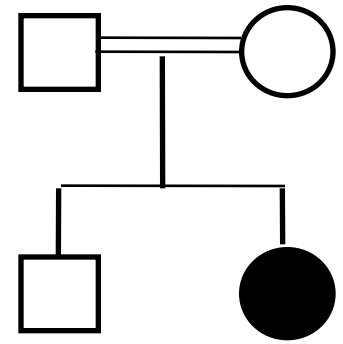
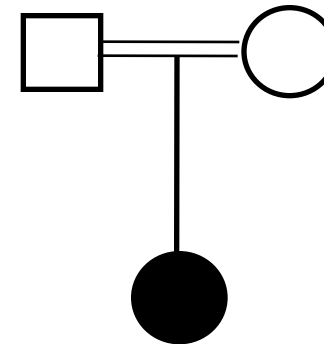
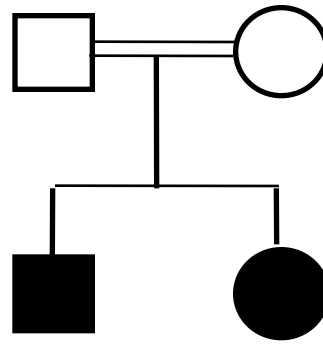
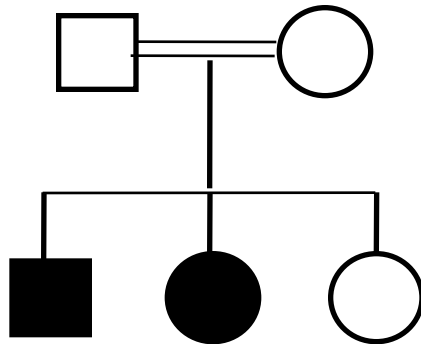
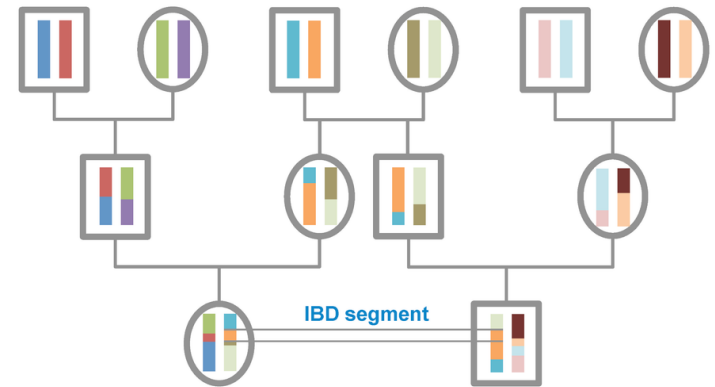
COSTRUZIONE DI APLOTIPI: definizione della regione critica



AUTOZIGOSITA'

(IBD: **identical by descent**)

Marcatori omozigoti perché ereditati da un comune antenato



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

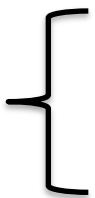
2	2	2	6
6	6	6	6
3	3	3	3
4	4	4	4
10	10	10	10

3	3
6	6
5	5
6	6
10	10

5	5
20	8
4	7
9	3
10	10

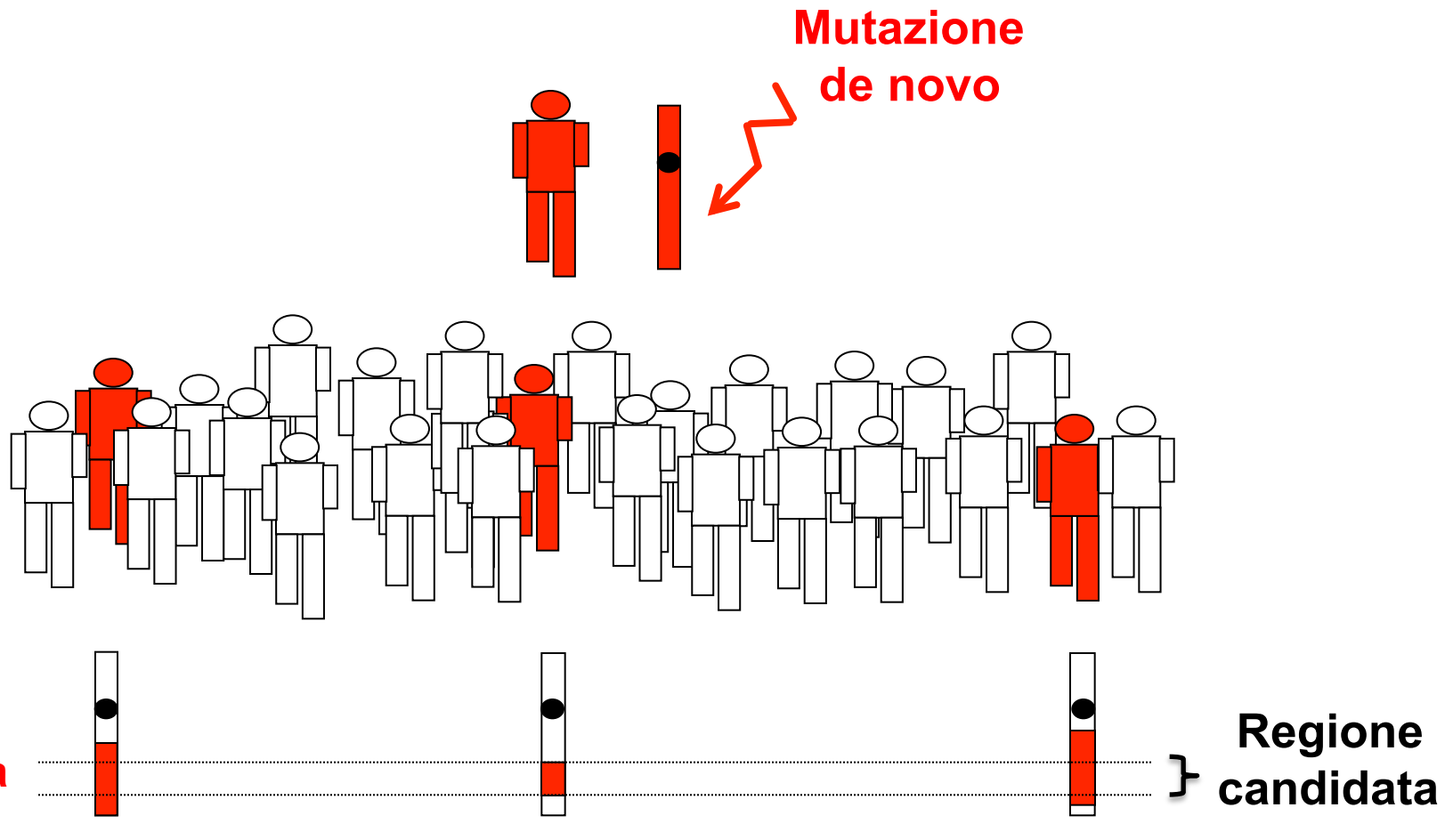
5	5
8	8
7	7
3	3
10	10

Regione
candidata



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, geni candidati) **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



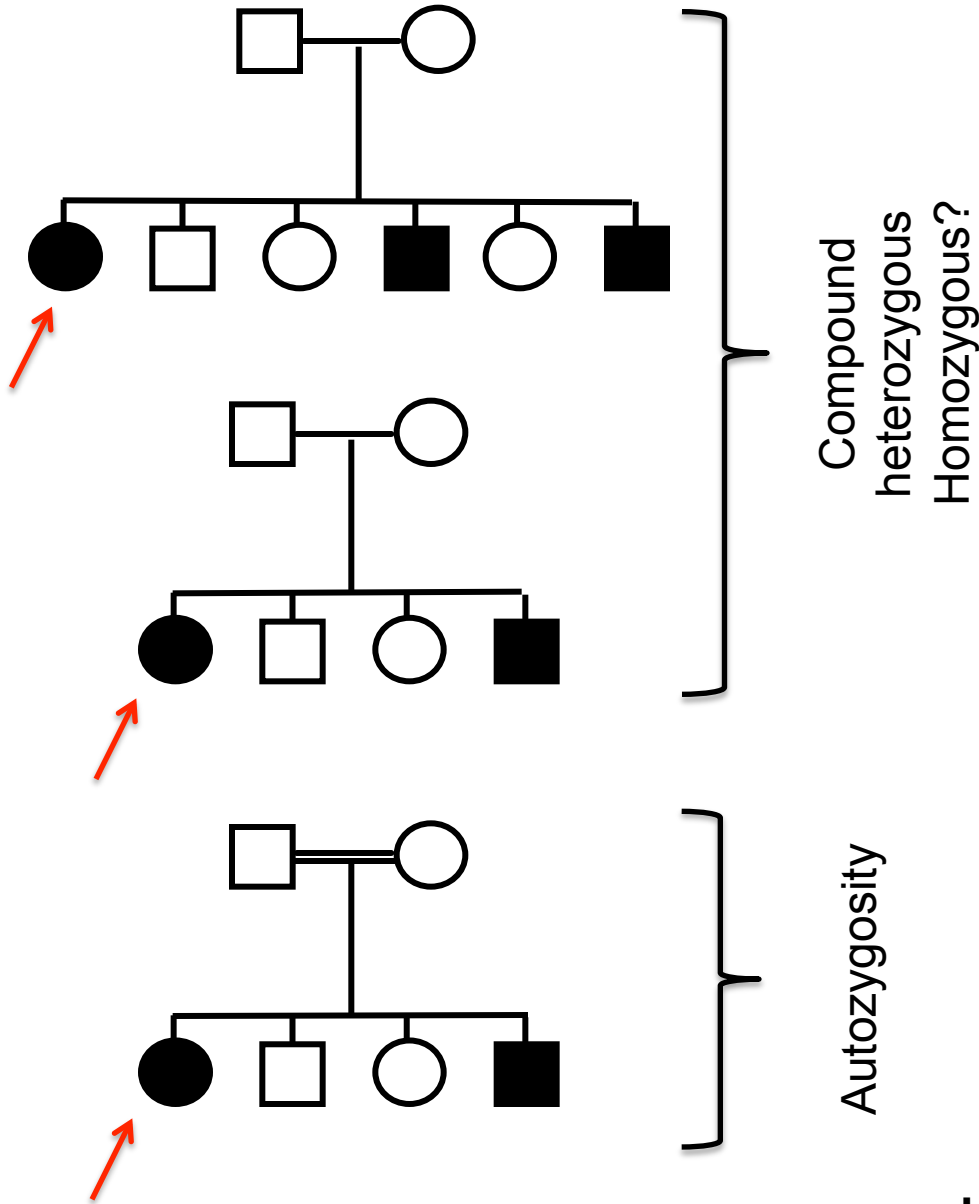
Eliminare varianti non
patogenetiche

3

Ereditarietà

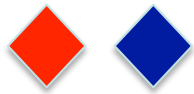


Verificare modello
mediante analisi di
segregazione

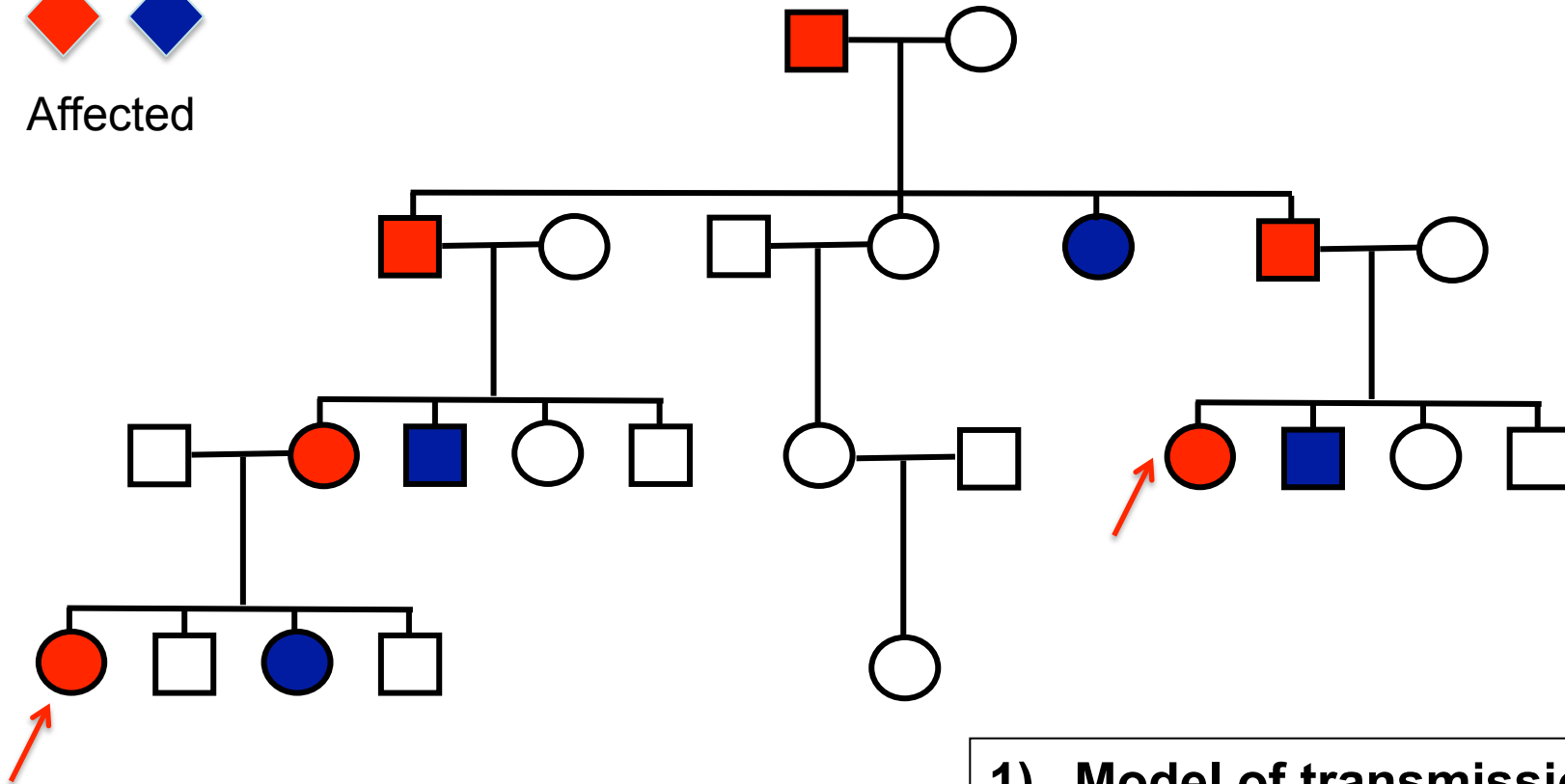


- 1) Model of transmission: AR
- 2) NGS in probands (and other members)
- 3) Filtering
- 4) List of potential pathogenic variants
- 5) Variants affecting the same gene(s) in all the families studied
- 6) Both alleles affected (homozygous or compound heterozygous)
- 7) Segregation analysis

... what in case of genetic heterogeneity?

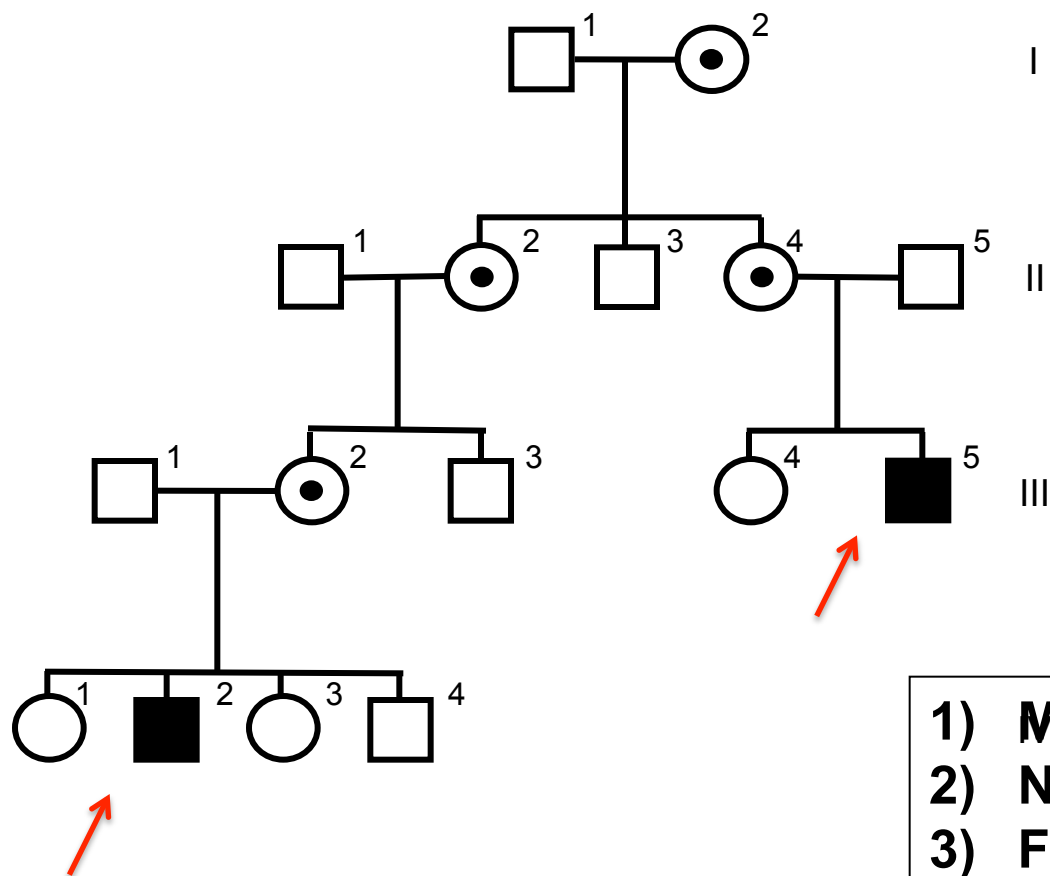


Affected

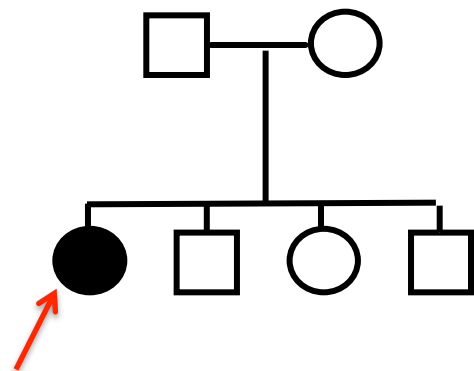
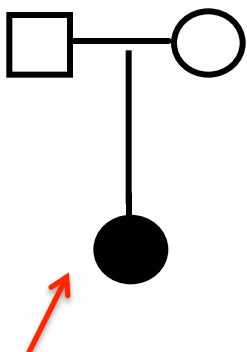
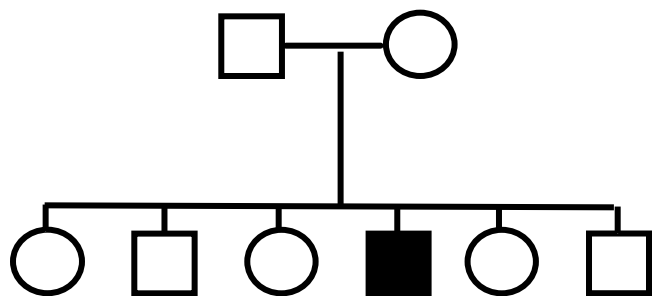


- Large family
- Combination with linkage analysis
- Pathogenic variants only in the candidate region

- 1) Model of transmission: AD
- 2) NGS in the most distant (meiosis) relatives
- 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 parents**
- 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 **non si riesce a sequenziare tutto l'esoma**
- **Coperture (numero di volte che uno specifico nucleotide o regione) varia da una regione all'altra**
- Le sonde si basano su database, quali RefSeq e CCDS (consensus coding sequence) con **esoni non ancora annotati e quindi 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

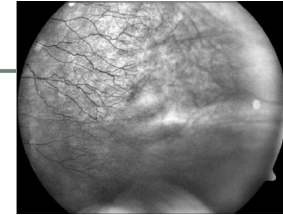
Verso whole genome sequencing

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

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 in famiglie negative per mutazioni nei geni noti:
regione candidata in 40.5 Mb nel chr. 7

Identificare nuovo(i) gene(i) responsabile(i) della malattia in famiglie nelle quali non sono state identificate mutazioni dei geni noti.

Biological
Background

Aim

MATERIALI E METODI

Approach

Sequenza degli esoni localizzati nella regione candidata (338 geni, 3048 esoni, 2.366.514 bp)

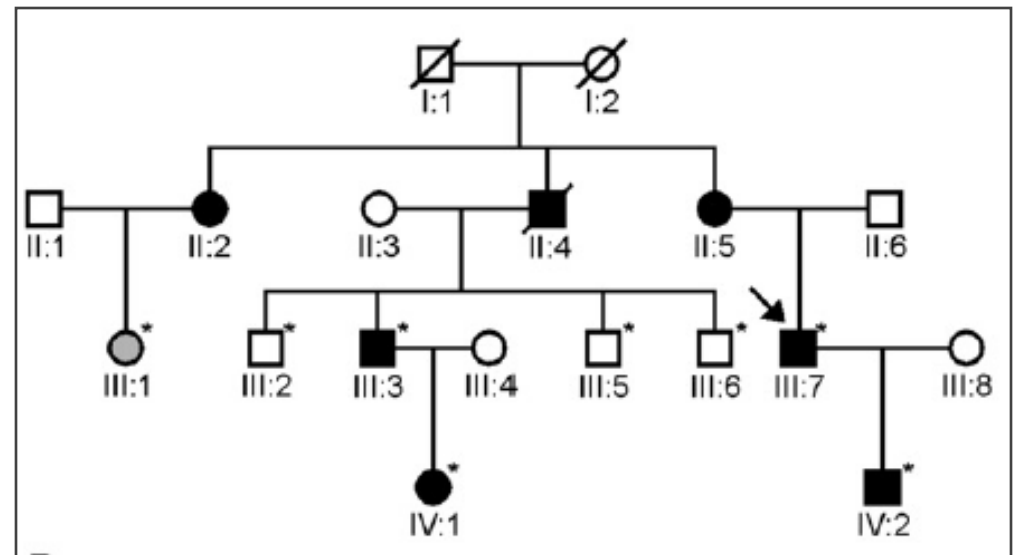


Samples

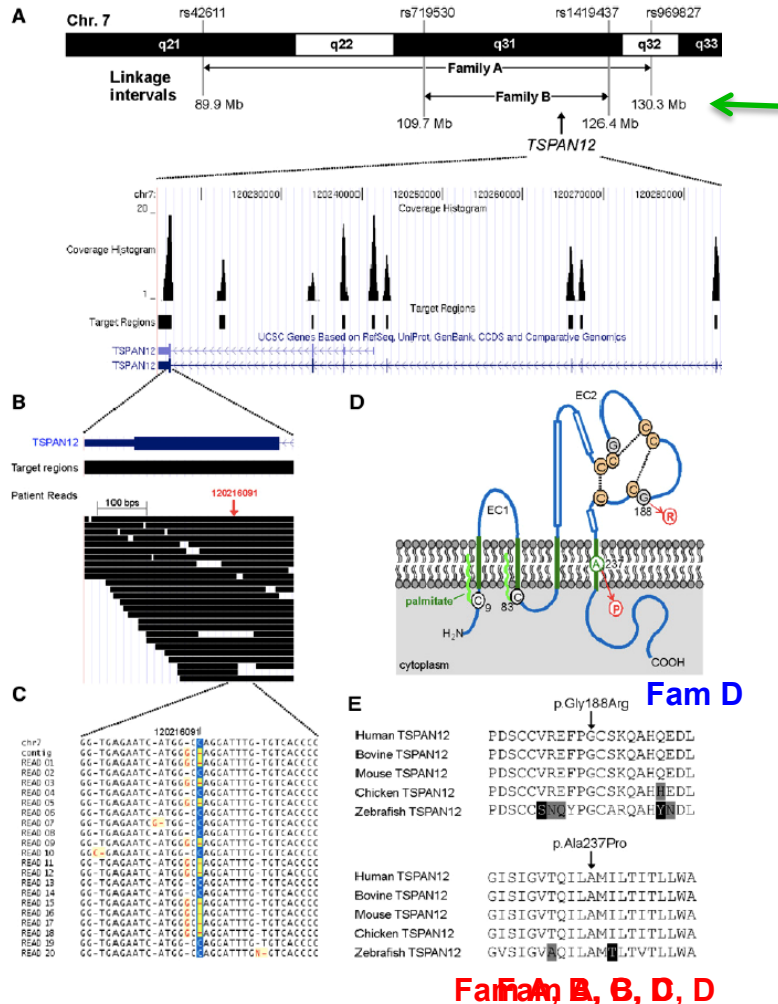
2 famiglie (Fam A e Fam B) con la forma autosomica dominante con esclusione di mutazioni nei geni noti per la malattia (FZD4, LRP5 e NDP)

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 identificate

1749

escluse perché non rare

14

varianti esoniche, non sinonime e
presenti in eterozigosi

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

TSPAN12 regola lo sviluppo vascolare della retina.

- Identificata variante **p.Ala237Pro** nella Fam A.
- La stessa mutazione (p.Ala237Pro) è presente negli affetti delle Fam B (analisi di linkage localizza il gene nella stessa regione, restringendo la porzione candidata)
- Mutazione anche in altre due Fam C e Fam D (non sottoposte ad analisi di linkage e esoma).
- Una seconda mutazione (**p.Gly188Arg**) di TSPAN12 è stata identificata in Fam D

Conclusioni: TSPAN12 è il quarto gene 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
 - 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.)