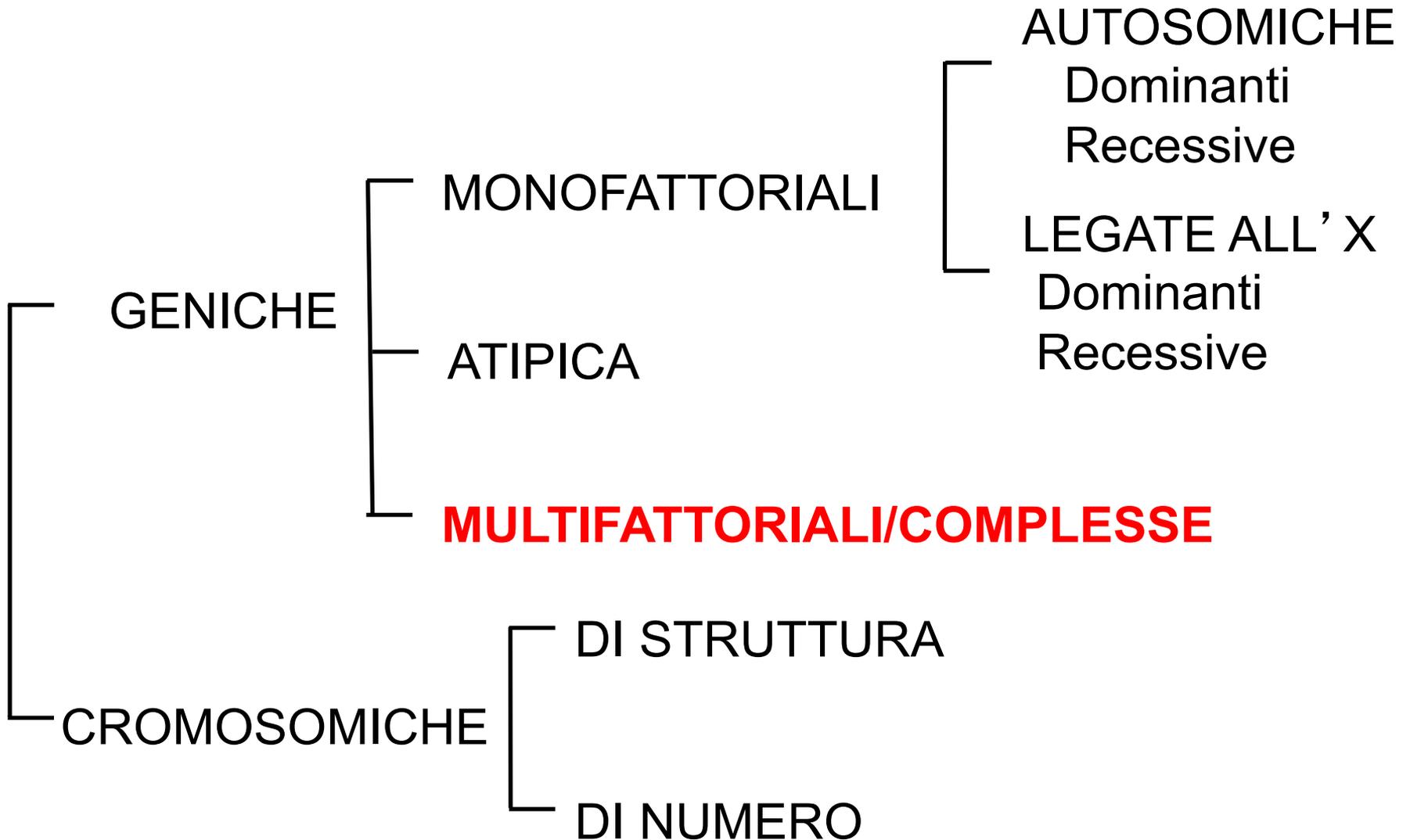


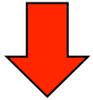
MALATTIE GENETICHE



Mendelian vs complex diseases

Penetrance

Disease-causing
gene



Modifier
Gene 1



Modifier
Gene 2



Modifier
Gene 3



Mendelian
disease



Environment

Modulation of
expressivity

RF 1
AD



RF 2
AR



RF3
XL



RF4
ADpp



Complex
disease



RFn



Environment

Do not confuse with
genetic heterogeneity

Malattie mendeliane



Malattie multifattoriali

Ereditarietà

Trasmissione mendeliana

Autosomica recessiva

Autosomica dominante

X-linked

Familiarità

Malattia è più comune tra i parentidi un individuo affetto
che nella popolazione in generale

Trasmissione non-mendeliana

Rare

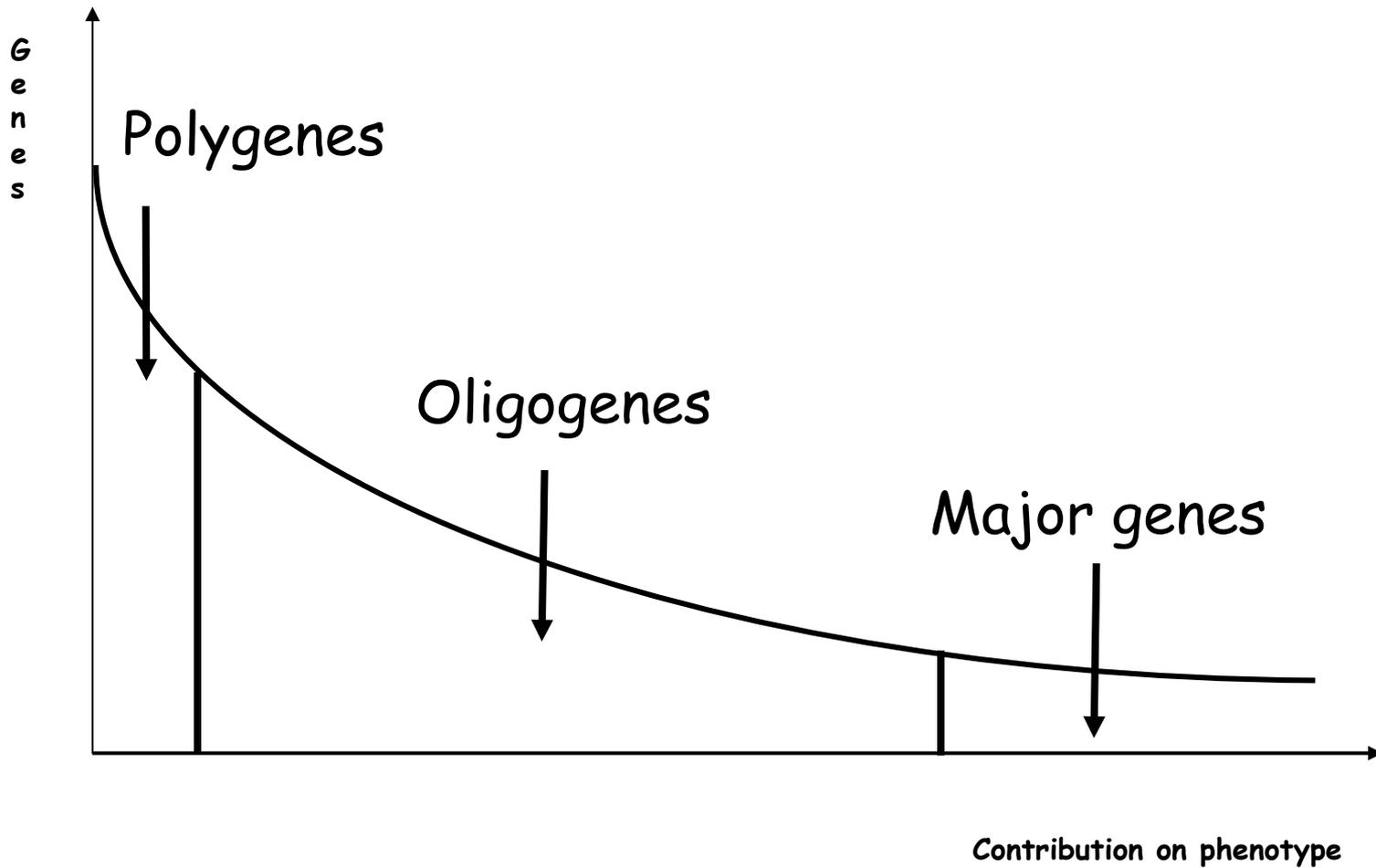
- fibrosi cistica
- sordità
- distrofia muscolare di Duchenne

Comuni

- ipertensione
- asma
- diabete mellito
- malattia trombo-embolica
- infarto miocardio
- depressione
- cefalea
- osteoporosi
- patologia tiroidea

Carattere poligenico: fenotipo determinato dall'azione combinata di numerosi loci, ciascuno con un piccolo effetto. **Ereditarietà poligenica**

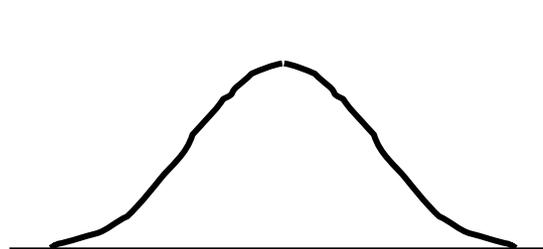
Carattere multifattoriale: fenotipo determinato dalla combinazione di fattori ambientali e genetici. **Carattere complesso**



Caratteri

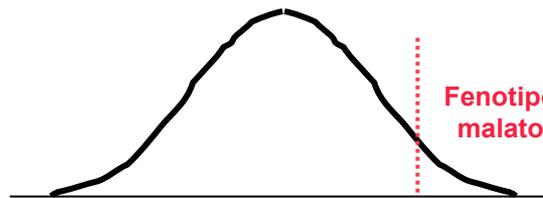
Ereditarietà

Quantitativi/continui



Esempio altezza

*A “soglia”
(discontinui)*

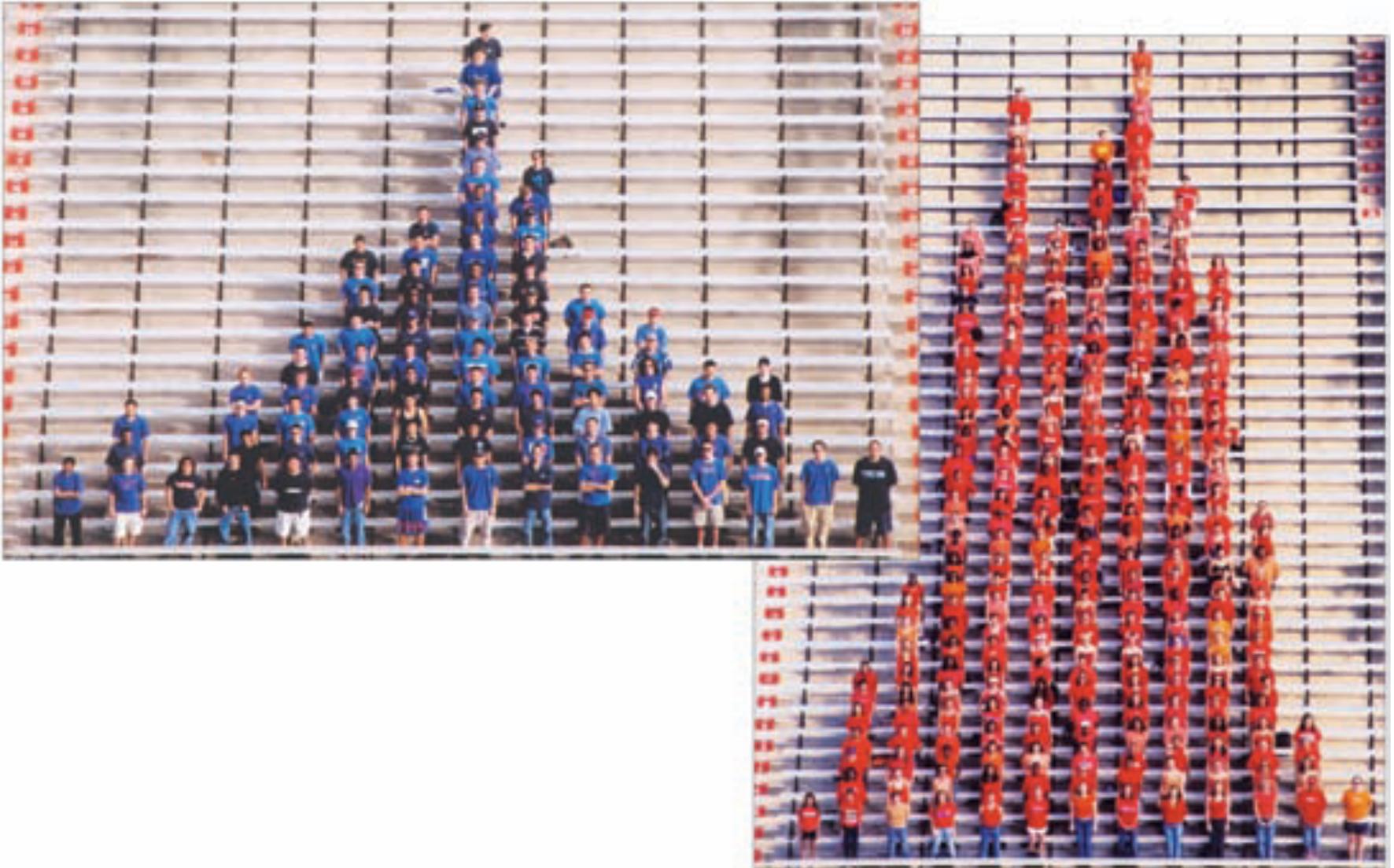


Esempio predisposizione

Teoria poligenica

Penetranza ed espressività variano con il numero di loci mutati e in relazione all'effetto delle diverse mutazioni, alle modalità di ereditarietà (recessiva, dominante, codominante) e alla mescolanza di componenti poligeniche e non.

Variazione continua: distribuzione delle altezze

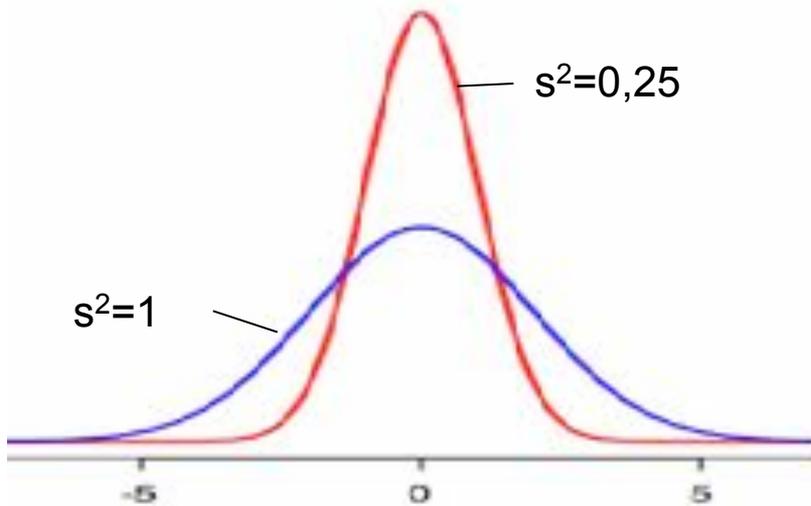


Distribuzione normale (simmetrica o a campana) delle frequenze dei fenotipi

- Si osserva quando numerosi fattori indipendenti influenzano il fenotipo
- Molti caratteri continui sono influenzati da molti fattori (genetici e ambientali)

Media

$$\bar{X} = \frac{\sum_{i=1}^N X_i}{N}$$



Due distribuzioni con la stessa media
ma con diverse varianze
(o deviazioni standard)

Varianza

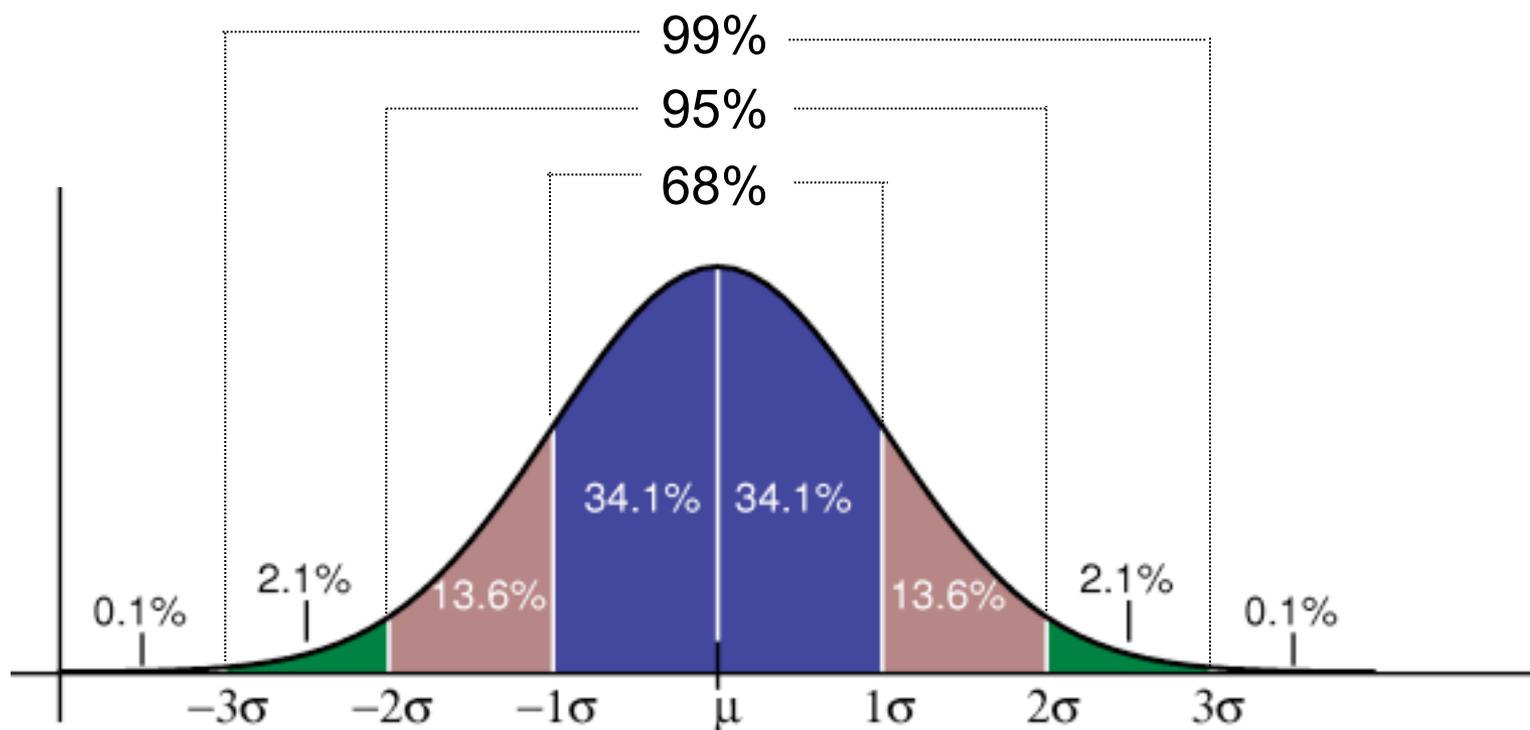
$$s^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N}$$

Deviazione standard

$$s = \sqrt{s^2}$$

Curva relativa ad una distribuzione normale

Distribuzione rispetto a multipli della deviazione standard



Teoria poligenica: Caratteri quantitativi

QTL (quantitative trait loci)

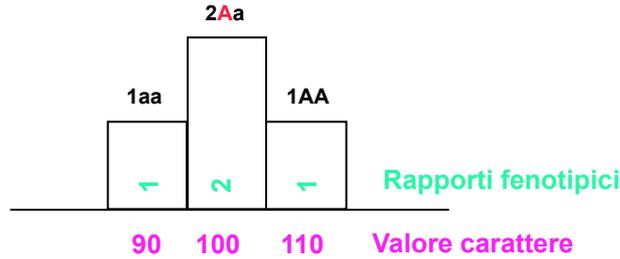
Effetto additivo degli alleli

Ipotetico carattere
con valore medio = 100

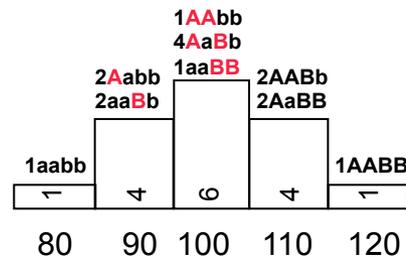
Effetto = 10
Per ogni allele dominante

Locus biallelico
 $p = q = 0,5$

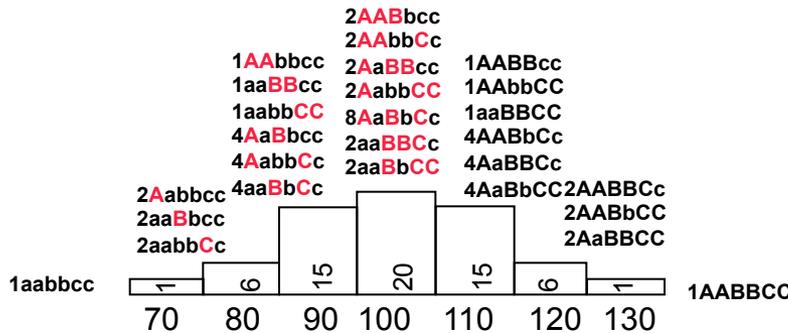
Effetto co-dominante
alleli



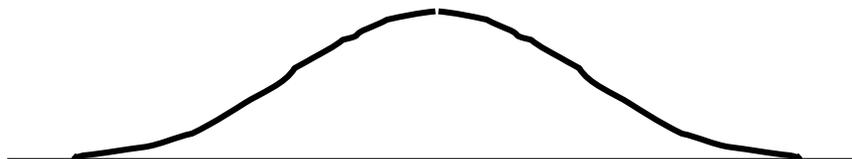
Unico locus



Due loci



Tre loci



N loci: poligeni

Solo cambiando un parametro del box rosso ...

Ipotetico carattere
con valore medio = 100

Effetto = 10
Per ogni allele dominante

Locus biallelico
p = 0,8
q = 0,2

Effetto co-dominante
alleli



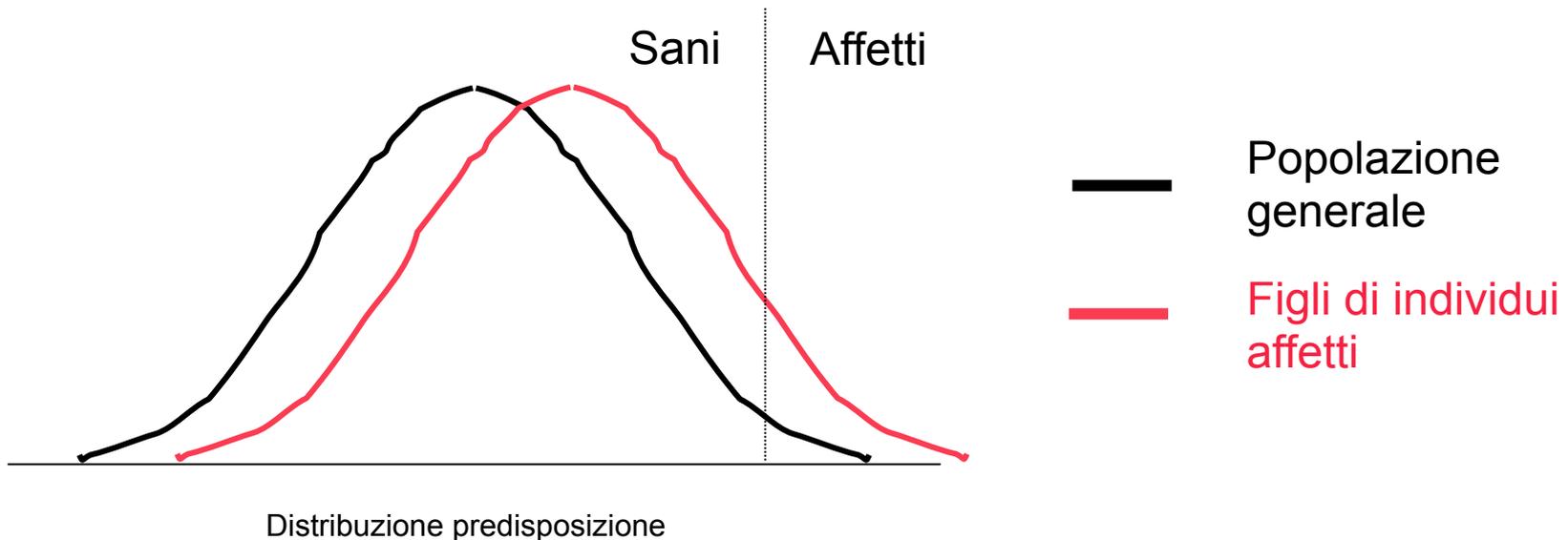
... per ogni locus i parametri del box rosso cambiano

Numero di genotipi e fenotipi attesi
al variare del numero delle coppie alleliche
che segregano indipendentemente e che hanno effetti additivi e uguali

Numero loci (2 alleli)	Alleli segreganti	Numero di genotipi F ₂	Numero fenotipi F ₂
1	2	$3^1 = 3$	3
2	4	$3^2 = 9$	5
3	6	$3^3 = 27$	7
4	8	$3^4 = 81$	9
n	2n	3^n	2n+1

Teoria poligenica: caratteri discontinui

- Tutti gli individui hanno un qualche grado di predisposizione a manifestare un carattere
- I fattori di suscettibilità (genetici e ambientali) sono molteplici
- Il numero dei fattori è distribuito secondo la curva gaussiana
- La comparsa del fenotipo dipende da una soglia di suscettibilità
- I consanguinei avranno un valore medio di fattori di predisposizione maggiore



Caratteri discontinui: rischio di ricorrenza λ_R
 $\lambda_R = 1$, rischio uguale a quello della popolazione generale

Stima del rischio di ricorrenza si basa su osservazioni empiriche

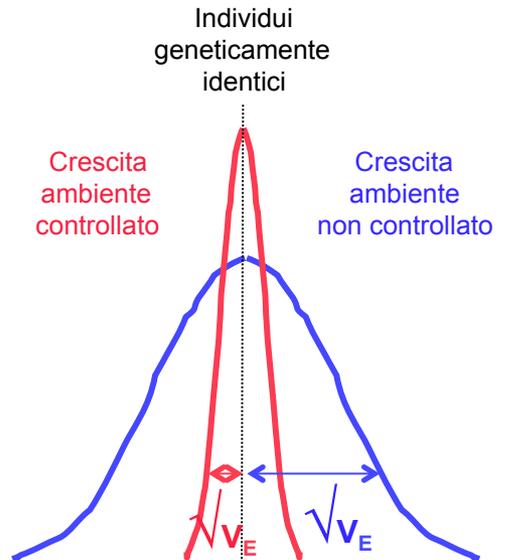
- Numero dei malati in una famiglia (maggiore è il numero di individui malati nella famiglia maggiore è λ_R)
- Gravità della malattia nel consanguineo affetto (individui colpiti in modo grave hanno in media un numero maggiore di fattori di predisposizione)
- Incidenza della malattia nella popolazione (tanto maggiore è la malattia nella popolazione tanto minore è il peso dei fattori genetici)
- Incidenza diversa nei sessi

Esempio: stenosi del piloro

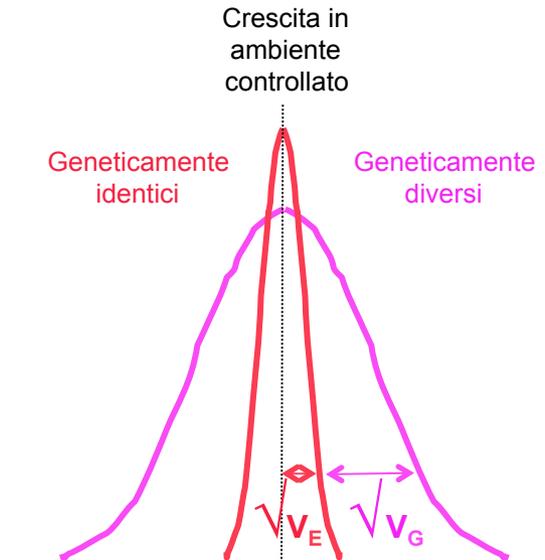
	Figli	Figlie	Fratelli	Sorelle
Maschio affetto	6,5	2,5	2,2	2,0
Femmina affetta	23	11,5	11	9

NB. Condivisione dello stesso ambiente può essere un'alternativa alla familiarità

Genetica o ambiente?

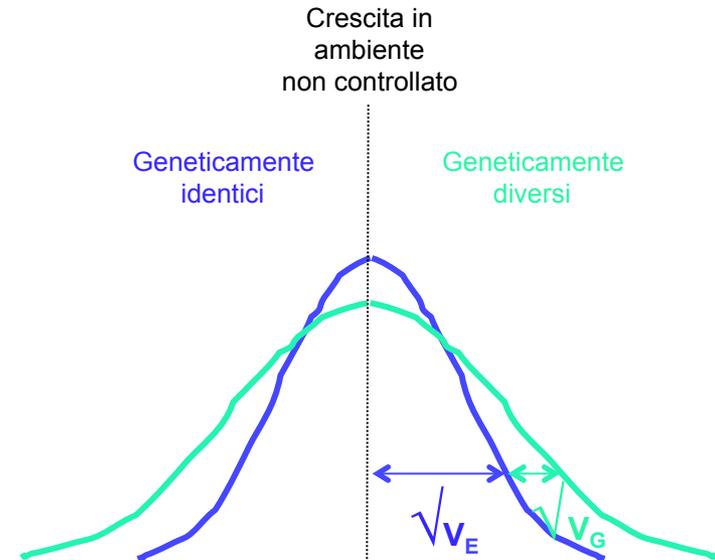


$$V_E \rightarrow 0$$



V_P : varianza fenotipica
 V_E : varianza ambientale
 V_G : varianza genetica

h^2 : ereditabilità



$$V_P = V_E + V_G$$

$$h^2 = V_G / V_P$$

Studi per definire la componente genetica (in blu le criticità degli studi)

- **Ricorrenza nelle famiglie:** rischio di un parente di sviluppare la malattia rispetto alla popolazione generale (condivisione dell'ambiente potrebbe essere un'alternativa alla causa genetica)
- **Studi su gemelli "identici" monozigoti (MZ)** per valutare la concordanza (% di coppie di gemelli che condividono un carattere). Se il carattere è controllato geneticamente, è maggiore nei MZ rispetto a quella dei gemelli dizigotici (DZ) (condivisione ambiente, genere diverso tra i DZ)
- **Studi su gemelli MZ affidati in adozione:** condividono il genoma ma non l'ambiente (la separazione però non è totale)
- **Studi su figli adottati:** valutare se la malattia che colpisce i figli adottati è presente o meno nelle famiglie adottive o naturali; viceversa valutare la malattia in figli dati in adozione quando la malattia è presente nei loro genitori.

Concordance: percentage of pairs of relatives (mono- or di-zygotic twins) in which both members have:

- 1) Certain qualitative trait
- 2) Quantitative trait similar in magnitude

Malattia	Concordanza (%)	
	MZ	DZ
Epilessia	70	6
Sclerosi multipla	18	2
Diabete tipo 1	40	5
Schizofrenia	53	15
Artrite reumatoide	12	3,5
Psoriasi	72	15
Palatoschisi	30	5

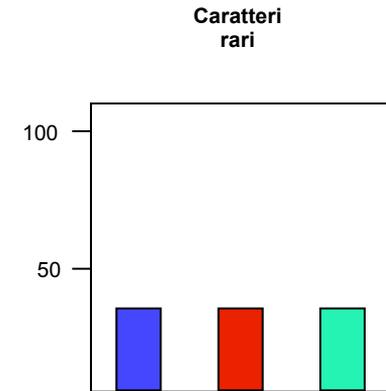
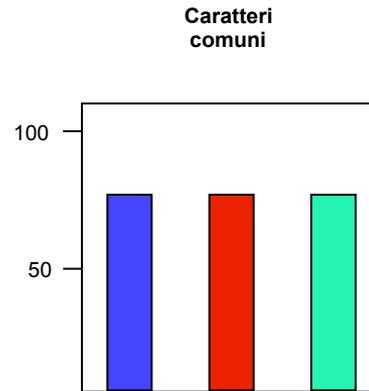
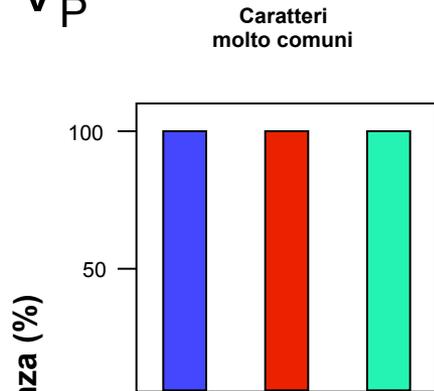
Più i valori si differenziano maggiore è l'influenza genetica

In caso di epilessia, nel 70% delle coppie MZ entrambi i gemelli sono affetti; in solo il 6% delle coppie DZ entrambi i gemelli hanno la malattia

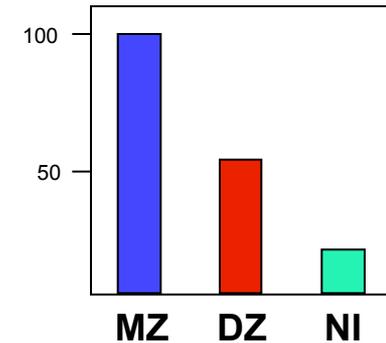
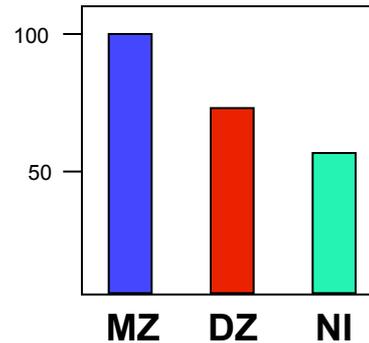
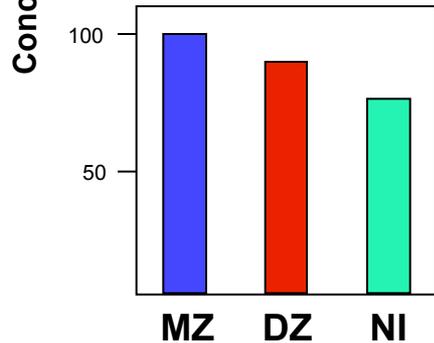
Effetto dell'ereditabilità sulla concordanza

$$h^2 = V_G / V_P$$

$$h^2 = 0$$



$$h^2 = 1$$



MZ: gemelli omozigoti
DZ: gemelli dizigoti
NI: adottivi non imparentati

Fattori di rischio genetici: caratteristiche

- Varianti comuni nella popolazione (es. SNP)
- Azione non “neutra” sulla funzione di un prodotto genetico (effetti/variazioni lievi)
- Esempi
 - SNP nella regione codificante di un gene che influisce sulla funzione della proteina (es. riduzione dell’attività del 5%)
 - SNP in regioni regolatrici che interferiscono con i normali livelli di espressione della proteina (ridotti/aumentati di una piccola percentuale)

Strategie per l'identificazione dei fattori genetici implicati nelle malattie multifattoriali

1) Studi di associazione

Studio casi-controlli. Consiste nel confronto della frequenza di un determinato allele ad un locus in due gruppi: presenza (casi) o assenza (controlli) del carattere. Se un allele mostra una frequenza significativamente maggiore nei “casi” rispetto ai “controlli” si dice che esso è associato al fenotipo in esame.

L'allele A1 al locus A è associato alla malattia D se gli individui affetti da D hanno A1 ad una frequenza maggiore di quella attesa sulla base della frequenza di D e di A1 nella popolazione.

Esempio: allele HLA-DR4 ha una frequenza del 36% nella popolazione (UK) generale ma del 78% in individui con l'artrite reumatoide.

Studio casi-controlli

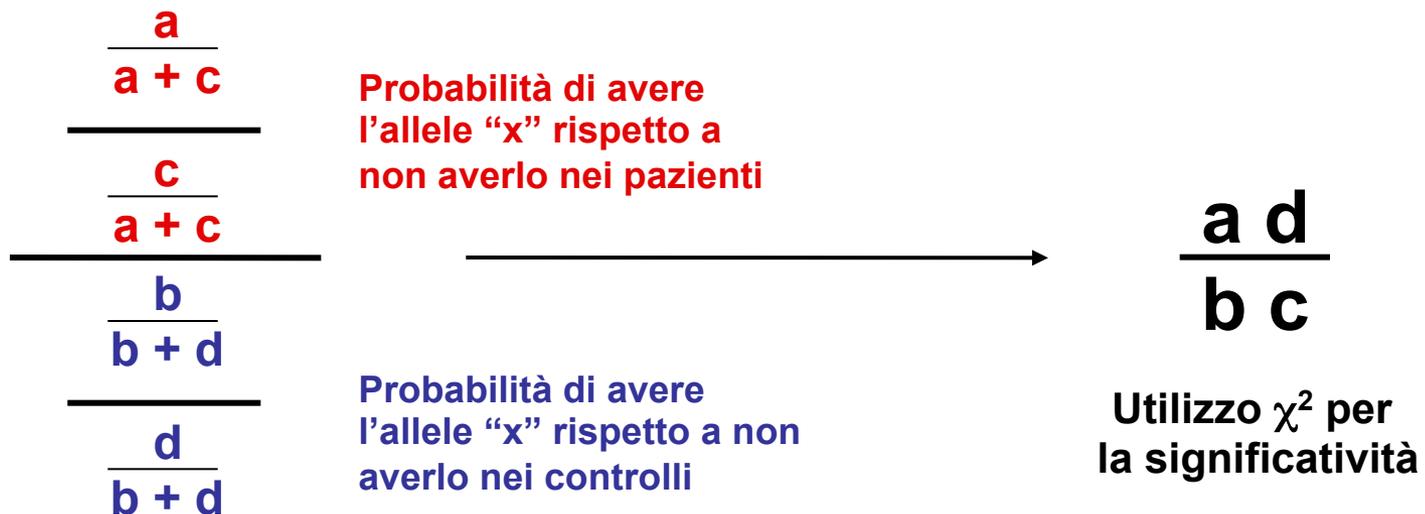
Vantaggi: facilità di reclutamento della casistica.
Approccio utilizzato per valutare il coinvolgimento di un gene candidato nella predisposizione di un carattere.

Aspetti limitanti: i) campione “casi” omogeneo, ii) numerosità del campione (tanto più frequente è un polimorfismo tanto minore sarà l’effetto), iii) selezione adeguata del campione “controlli”: **sexo, età, origine geografica per limitare eventuali differenze delle frequenze alleliche tra le popolazioni.**

Studi di associazione caso-controllo

Calcolo dell'**Odds Ratio** (OR, rapporto di probabilità)
per valutare la forza di associazione

Allele "x"	Pazienti (N=100)	Controlli (N=120)
Presente	90 (a)	40 (b)
Assente	10 (c)	80 (d)



Studio di associazione

b) Studio di **associazione** con controlli interni:
transmission disequilibrium test (TDT) prevede l'analisi del probando e di entrambi i genitori. Può essere utilizzato solo quando i genitori sono disponibili

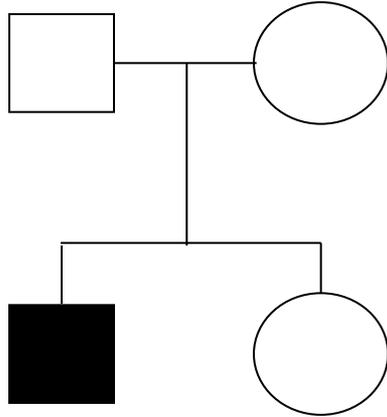
Vantaggi: si evita il problema della selezione del campione controllo

Aspetti limitanti: genitori non sempre disponibile. Malattie sono spesso ad esordio tardivo

TDT (transmission disequilibrium test)

Uso di **CONTROLLI INTERNI**, gli stessi usati per ricavare i dati

A1A2



- Determinare genotipo dei probandi e dei genitori
- Selezionare genitori eterozigoti (possono essere o non essere ammalati) ai loci di interesse
- “a” è il numero di volte che un genitore eterozigote trasmette l’allele A1 ai figli ammalati, e “b” è il numero di volte che trasmette l’altro allele.
- Se c’è una differenza significativa tra “a” e “b”, c’è associazione. **La trasmissione di un allele si discosta significativamente dal valore atteso del 50%**

$$(a - b)^2 / (a + b)$$

Distribuzione χ^2
con 1 grado di libertà

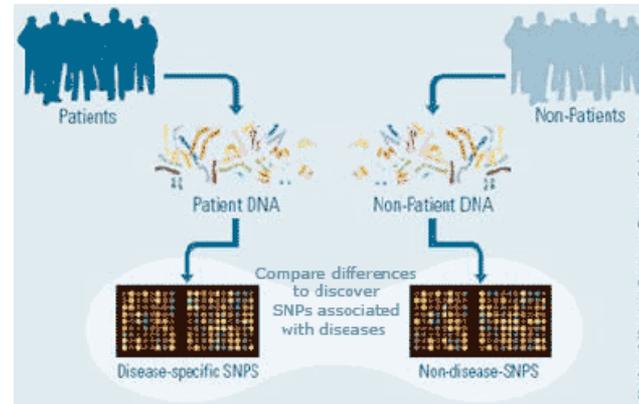
Allele	A1	A2	
Trasmisione	a (49)	b (51)	Atteso
	a (12)	b (88)	Assoc

Studio di associazione

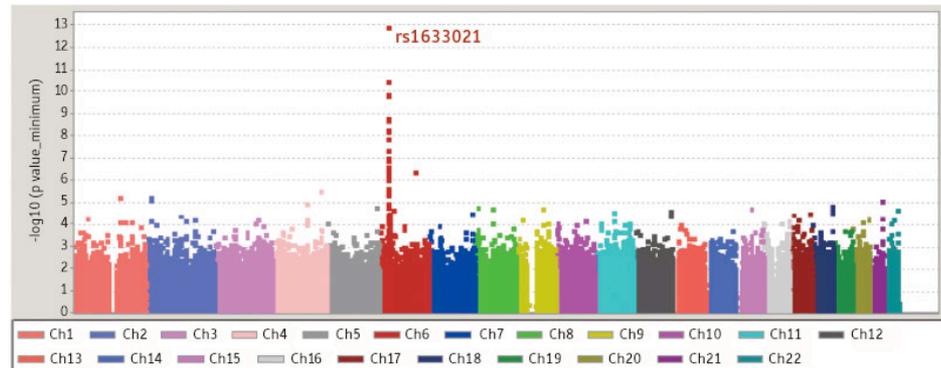
- Gene candidato: analisi di SNP ad esso associati
- Genome wide association study: analisi SNP su tutto il genoma

Genome Wide Association Study (GWAS)

**Piattaforme da
300.000/
5.000.000 SNPs**



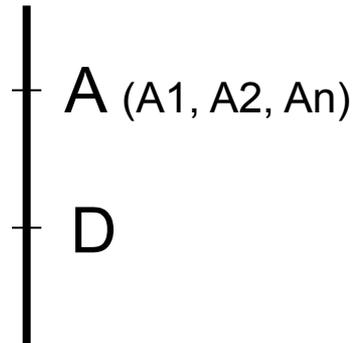
**Rappresentazione
“Manhattan”**



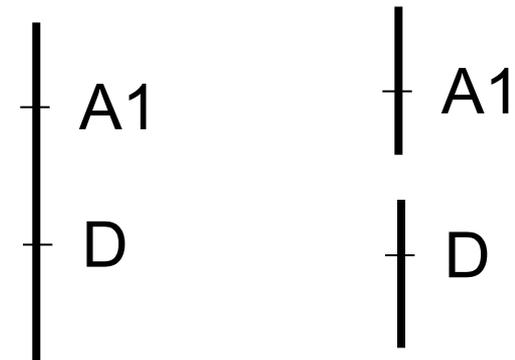
- Identificazione di varianti con una frequenza maggiore (statisticamente significativa) nel gruppo di pazienti rispetto ai controlli
- Rispetto agli studi del gene candidato in cui si analizzano alcuni SNP nel gene stesso, nel GWAS si analizzano SNPs distribuiti su tutto il genoma.

Linkage \neq Associazione

Linkage



Associazione



Relazione tra loci (genetica)

- distribuzione degli alleli A in un gruppo di pazienti è uguale a quella della popolazione generale;
- all'interno di una famiglia affetta che segrega con il locus A i malati condividono lo stesso allele;
- linkage crea associazione all'interno delle singole famiglie ma non tra individui non imparentati.

Relazione tra alleli (Osservazione statistica)

Perché si osserva associazione?

- **Causa diretta:** presenza dell' allele A aumenta la predisposizione. Effetto patogenetico dell' allele.
- **Selezione naturale:** gli individui che hanno la malattia D hanno una maggiore probabilità di avere figli se hanno anche l' allele A. Allele non implicato nella malattia
- **Stratificazione della popolazione:** popolazione può essere costituita da sottogruppi nei quali la malattia D e l' allele A1 possono essere relativamente frequenti. Allele non implicato nella malattia
- **Linkage disequilibrium:** alleli sono associati l' uno all' altro ad una frequenza significativamente più elevata rispetto a quella attesa calcolata sulla base delle frequenze alleliche e l' associazione random. Allele è in stretta vicinanza alla variante patogenetica.

Linkage disequilibrium (LD)

Teoricamente, la distribuzione delle combinazioni alleliche nella popolazione dovrebbe seguire le leggi della segregazione causale e della probabilità.

Dati 2 loci, A e B, ciascuno con due alleli
le cui frequenze sono:

$$F(A) = p(A)$$

$$F(a) = q(a)$$

$$F(B) = p(B)$$

$$F(b) = q(b)$$

le frequenze attese delle combinazioni (aplotipi) dipendono dalle frequenze alleliche, come indicato in tabella.

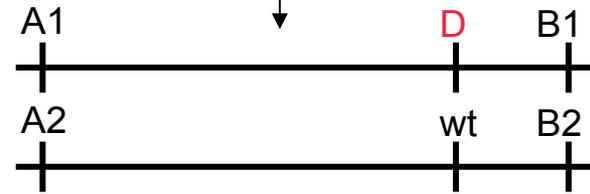
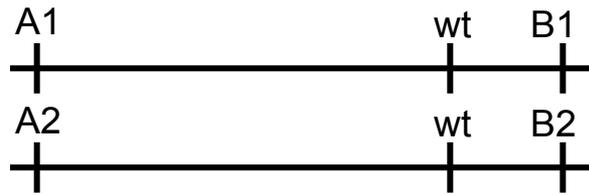
	p(A)	q(a)
p(B)	AB $p(A)p(B)$	aB $q(a)p(B)$
q(b)	Ab $p(A)q(b)$	ab $q(a)q(b)$

Linkage disequilibrium: deviazione delle frequenze aplotiche osservate rispetto a quelle attese. Ciò accade quando i loci sono strettamente concatenati. Per effetto del crossing-over, prima o poi l'equilibrio dovrebbe essere raggiunto con una velocità che dipende dalla distanza tra i loci.

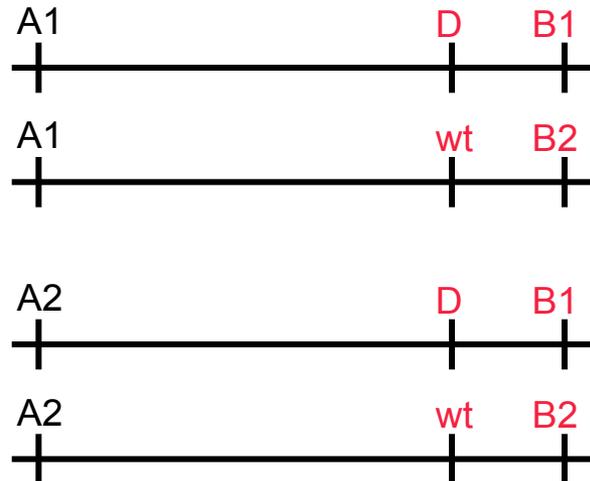
LD = Frequenza aplotipo osservata - Frequenza aplotipo attesa

$$LD = F(AB) - [p(A)p(B)]$$

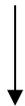
$$LD = 0, \text{ frequenze alleliche in equilibrio}$$



Crossing over



Crossing over
più frequente
tra locus A e D



Equilibrium
tra gli alleli
al locus A e
al locus D

Crossing over
meno frequente
tra locus D e B



**Linkage
disequilibrium**

(non equilibrio causato dal linkage)

Allele B1 tenderà
ad essere trasmesso
insieme all'allele D

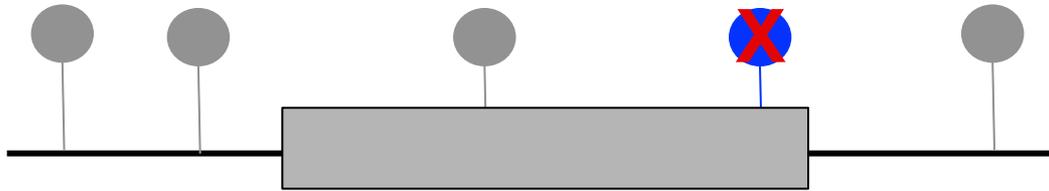


SNP analizzati

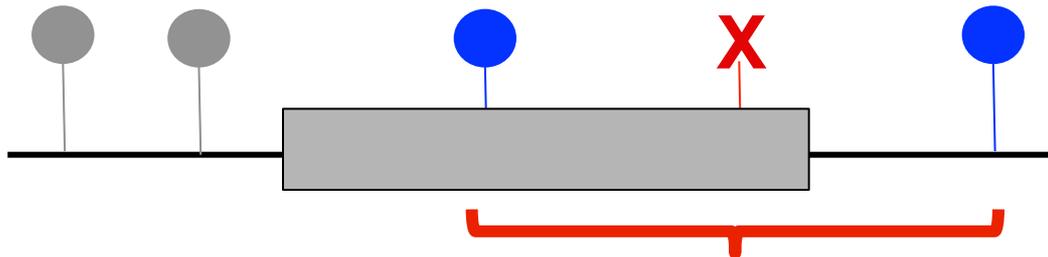


SNP in associazione

X Fattore di rischio



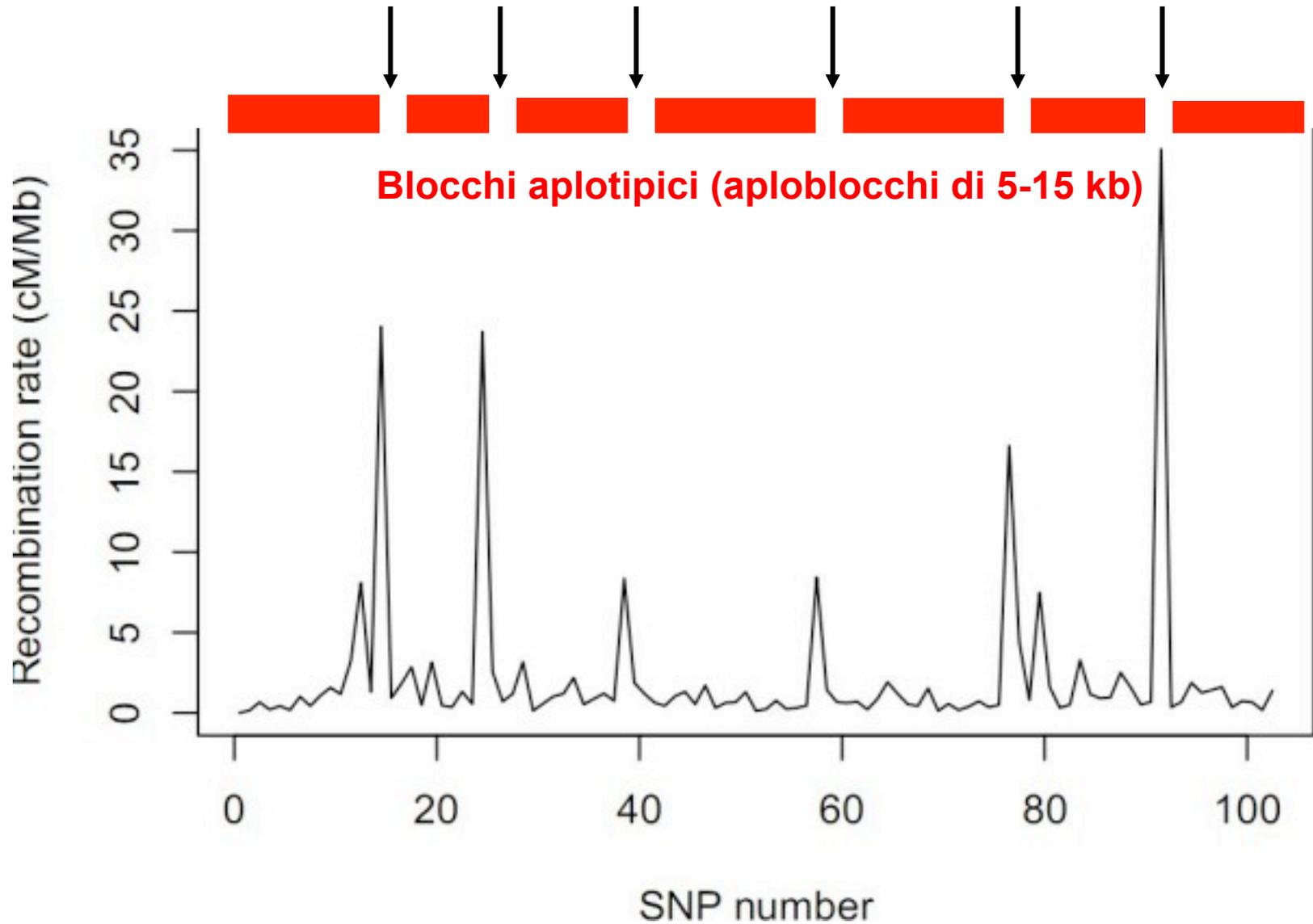
Causa diretta



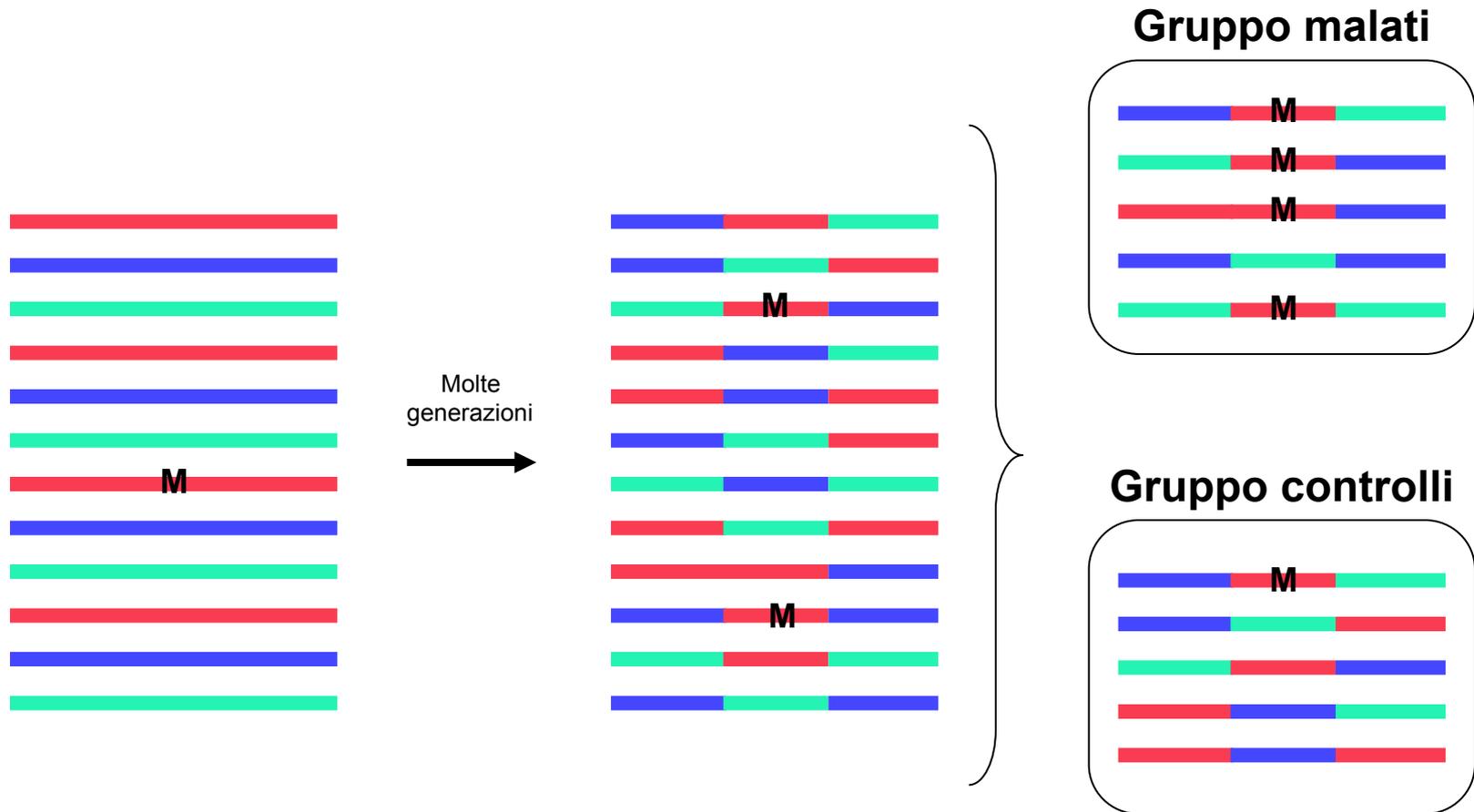
Causa indiretta

Linkage disequilibrium

Hotspot di ricombinazione



Associazione aplotipica



Alleli dell' aplotipo rosso presenteranno una frequenza più elevata nel gruppo di individui malati rispetto a quella del gruppo controlli

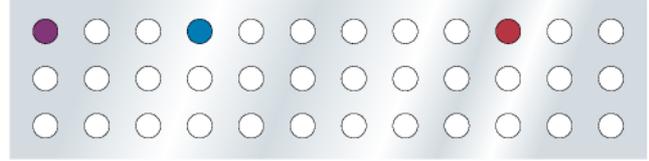
HapMap

selezione degli SNP per "etichettare" gli aplotipi



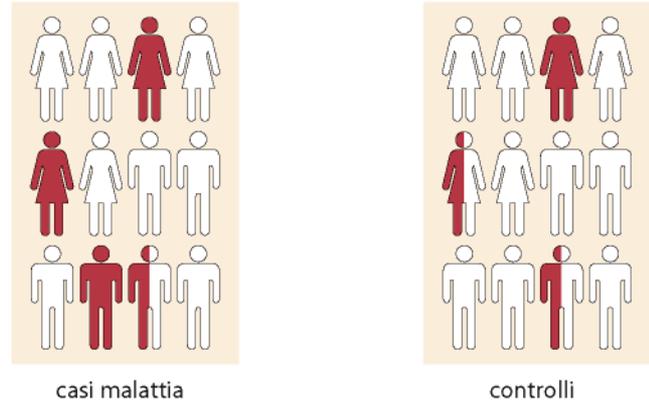
genotipizzazione

300 000–500 000 SNP caratterizzati tramite *array* ad alta densità



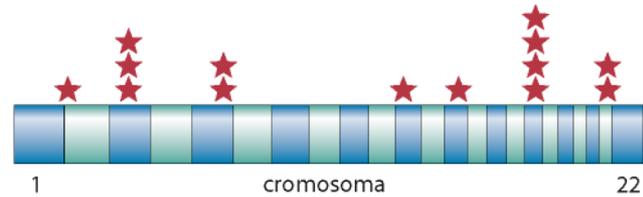
studi caso-controllo

confronto delle frequenze alleliche nei casi malattia e nei controlli



risultati della scansione del genoma

differenze significative nelle frequenze alleliche degli SNP indicano la possibile presenza di nuovi loci e geni malattia



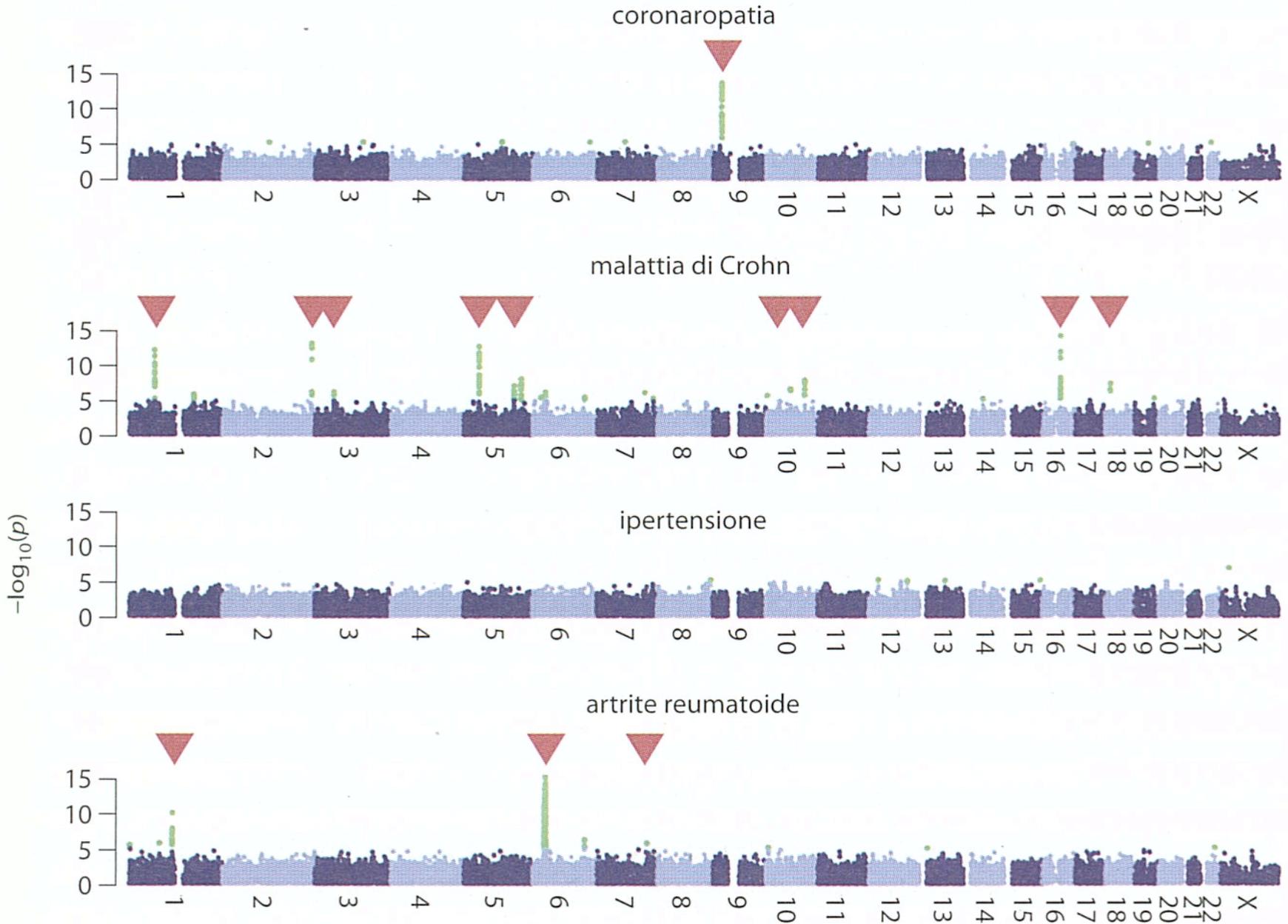
test di replica

conferma dei dati della scansione

SNP associati a un genotipo in un campione indipendente di casi e controlli

Rappresentazione "Manhattan"

Esempi di studi di associazione



Considerazioni:

- Migliaia di SNP sono associati a diverse malattie complesse (<http://www.genome.gov/GWASStudies>)
- Si conoscono le varianti patogenetiche solo in rari casi (es uno SNP del gene di IRF6 che predispone alla schisi del labbro)
- Molto difficile identificare le varianti causative con bassa penetranza, contrariamente a quanto si verifica nelle malattie mendeliane che sono determinate da varianti (mutazioni) a completa (o quasi completa) penetranza