

ANALISI DI LINKAGE

L'analisi di Linkage è un metodo statistico basato sull'analisi di famiglie affette da malattia monogenica mendeliana, con lo scopo di mappare ed identificare la posizione cromosomica di un locus (che si pensa essere il responsabile di una determinata malattia\carattere) rispetto a dei marcatori polimorfici di posizione nota.

Con il termine linkage si indica la tendenza di geni vicini su uno stesso cromosoma ad essere trasmessi insieme durante la meiosi. E' possibile sfruttare questa frequenza di ricombinazione alla meiosi come misura della distanza genetica tra due loci sul cromosoma, con l'analisi di linkage è possibile mappare la distanza tra due loci sul cromosoma andando a determinare la frequenza con la quale questi due loci ricombinano tra di loro alla meiosi.

Essendo a conoscenza della posizione del marcatore, in seguito all'analisi di linkage riesco ad identificare una regione minima ed isolare i geni presenti su di essa, andando poi a sequenziare individui affetti e non affetti nella famiglia per scoprire la mutazione che causa malattia, questo processo è chiamato positional cloning.

I requisiti alla base di questa analisi richiedono di avere una o più famiglie che presentano il carattere mendeliano e dei marcatori genetici. I marcatori genetici per effettuare l'analisi di linkage sono delle varianti polimorfiche (con frequenza di almeno 1% nella popolazione) che possono essere SNP (polimorfismo singolo nt) o microsatelliti.

Il primo passo da effettuare quando si effettua un'analisi di linkage è quello di andare a determinare il modello di ereditarietà del locus malattia per determinare il genotipo, si prosegue, poi, andando ad effettuare la stima di teta, ovvero la frequenza di ricombinazione tra i due loci ($teta = R/R+NR$) :

se i due loci sono lontani sullo stesso cromosoma o sono su cromosomi diversi, segregano in maniera indipendente e la probabilità di essere ereditati insieme è di $1/2$, quindi $teta = 0,50\%$

se i due loci sono vicini fra di loro saranno ereditati insieme più frequentemente, con $teta < 0,50\%$, tanto più sono vicini i due loci, tanto più piccola è la probabilità che questi vengano separati da un evento di crossingover.

Per l'analisi di linkage nell'uomo viene utilizzato il metodo probabilistico del Lod Score (Z), che calcola la probabilità di linkage cercando di estrapolare il massimo dell'informazione dai pedigree a disposizione. Calcola una frazione dove al numeratore c'è la probabilità di osservare linkage per i dati ottenuti, mentre al denominatore la probabilità di osservare questi stessi dati in assenza di linkage.

Con $Z > 0$ = è più probabile l'ipotesi di linkage rispetto a quella di indipendenza

$Z < 0$ = i dati sono in accordo con l'ipotesi di indipendenza

$Z = 3$ l'ipotesi di linkage è 1000 maggiore dell'ipotesi di indipendenza → normalmente si utilizza un limite di significatività di 3, quando $Z > 3$ posso stabilire che esiste linkage tra i due loci.

(Nei caratteri monogenici si può usare l'analisi di linkage per vedere la segregazione del fenotipo di interesse tramite l'analisi del pedigree e cercare di individuare marcatori molecolari che segregano insieme all'ipotetico locus responsabile del fenotipo. Nell'analisi di linkage il punto di partenza è conoscere il modello di ereditarietà del gene malattia che si permette di risalire, dall'osservazione del fenotipo, al genotipo per il locus malattia.

STUDI DI ASSOCIAZIONE

-GWAS-

Quando si studia un carattere multifattoriale in cui il carattere fenotipico è determinato dall'interazione di più geni che individualmente hanno un piccolo effetto sul fenotipo, non è possibile determinare un modello di ereditarietà. Non è possibile, quindi, utilizzare uno studio di linkage per mappare i fattori genetici implicati nei caratteri multifattoriali

Per questo motivo si sono sviluppati metodi di mappatura genetica non basati sull'analisi del pedigree, ma su studi di popolazione. Si analizzano gli individui di una stessa popolazione, suddividendoli in due gruppi:

-casi: individui affetti da una patologia multifattoriale

-controlli: individui non affetti

E' quindi uno studio di associazione casi-controlli di tipo retrospettivo.

Gli studi di associazione permettono di riuscire a mappare i loci di rischio per caratteri multifattoriali tramite un approccio GWAS, cioè studi di associazione che riguardano l'intero genoma. Lo scopo è quello di trovare varianti genetiche associate ad un aumento di rischio di sviluppare la malattia.

In questi studi di associazione si va ad analizzare il genoma dei casi e dei controlli tramite dei marcatori polimorfici (che hanno una frequenza di almeno 1% nella popolazione) SNPs (varianti di un singolo nt), per vedere se c'è un'associazione tra questi ed il fenotipo di interesse. Per associazione si intende la probabilità di osservare il fenotipo insieme ad un determinato allele è significativamente diversa dal prodotto delle singole probabilità.

Da questo confronto si vuole andare a vedere se alcuni di questi SNPs mostrano frequenze alleliche significativamente diversi nei casi e nei controlli. Se uno SNP non ha nulla a che vedere con il fenotipo in considerazione allora la sua frequenza sarà uguale nei casi e nei controlli, se invece c'è differenza significativa fra la distribuzione degli alleli del marcatore nei casi e nei controlli significa che c'è associazione.

Gli studi di associazione si basano sulla presenza nel genoma di linkage disequilibrium (LD), cioè un'associazione non casuale di alleli appartenenti a due loci sullo stesso cromosoma, questi loci sono in LD quando la frequenza di associazione dei differenti alleli è superiore o inferiore a quello che ci si aspetterebbe se i loci fosse indipendente ed associato in modo casuale. Questa specifica associazione costituisce di solito un particolare aplotipo ancestrale, diffuso nella popolazione in cui si osserva perché trasmesso da progenitore come unico segmento attraverso le varie generazioni. Per questo motivo il linkage disequilibrium è maggiore in popolazioni omogenee ed è influenzato da molti fattori, tra cui la selezione, il tasso di ricombinazione genetica, tasso di mutazione, deriva genetica, il sistema di accoppiamento, struttura della popolazione. Bisogna quindi stare attenti che la popolazione sia omogenea da un punto di vista delle frequenze alleliche.

L'associazione tra uno SNP e il fenotipo può avvenire per due diversi motivi:

-associazione diretta SNP e fenotipo: la variante che si studia è la vera variante di rischio

-associazione indiretta tra SNP e fenotipo: data dal fatto che lo SNP in studio è in linkage disequilibrium con il vero allele di rischio.

Negli studi GWAS non si sa in partenza quali sono i geni che possono essere implicati, non si ha quindi un'ipotesi di partenza, ci saranno tante varianti genetiche dove ognuna può leggermente predisporre o proteggere un individuo dal manifestarsi del fenotipo.

In uno studio GWAS si utilizzano i Tag SNPs, cioè una serie di SNPs non ridondanti che rappresentano la quasi totalità della varianza genomica conosciuta, perché grazie alla conoscenza del LD posso risalire a recuperare informazioni anche riguardo altre varianti non direttamente analizzate.

COME VEDERE SE C'E' ASSOCIAZIONE FRA UN ALLELE DEL MARCATORE ED IL FENOTIPO?

Tramite un test di associazione vado a vedere se c'è associazione fra un allele del marcatore ed il fenotipo.

Trasformo le frequenze genotipiche osservate dei casi e dei controlli in frequenze alleliche, calcolo le frequenze attese secondo l'ipotesi H_0 : non c'è associazione. A questo punto si procede con il test del X^2 per vedere se gli scostamenti tra i dati osservati e quelli attesi sono statisticamente significativi o sono dovuti al caso. Da questo test otterrò un valore di X^2 da confrontare con i valori di distribuzione del X^2 con un gdl, e il p-value che corrisponde a questo valore di X^2 lo confronto con il valore di significatività del test, che per lo studio di uno SNP è di 0,005, ma se si vanno a testare, ad esempio, un milione di SNP bisogna correggere con la Correzione di Bonferroni.

Se il p-value che ottengo è inferiore al livello di significatività allora rifiuto H_0 ed accetto l'ipotesi alternativa: c'è associazione.

Oltre al test del X^2 , si possono misurare anche

-odds ratio: calcolo del rapporto tra la probabilità di avere la malattia negli individui con variante di rischio\ chi non ha la variante

-rischio relativo: rapporto tra l'incidenza della malattia negli individui con la variante di rischio\incidenza della malattia negli individui senza variante di rischio.

Negli studi caso-controllo si può calcolare l'odds ratio ma non il rischio relativo!

Un $OR > 1$ mi dice che il fattore di rischio in questione aumenta il rischio di essere un caso rispetto ad un controllo

$OR < 1$ l'allele protegge dalla malattia

$OR=1$ l'allele non ha nessun effetto sul rischio di malattia

La principale utilità degli studi GWAS è quella di fornire indicazione sui meccanismi patologici, in quanto avere identificato delle varianti di rischio ha permesso di identificare le funzioni geniche che possono essere alterate e contribuire ai meccanismi patologici molecolari di una malattia. Questo costituisce una via per sviluppare nuovi farmaci, individuare target farmacologici e capire meglio le basi molecolari di questa patologia.

Uno dei principali motivi per cui sono stati portati avanti gli studi GWAS inizialmente era quello di andare a sviluppare un tipo di medicina personalizzata perchè attraverso la conoscenza delle varianti del genoma di un individuo si pensava di poter fare una predizione di rischio per le malattie complesse e questo permetteva di attuare delle strategie di prevenzione o di cura mirate al genotipo del singolo individuo. Ad oggi, però, questa pratica non trova applicazione in ambito clinico in quanto ogni singola variante di rischio ha un effetto molto piccolo, quindi non ha un valore predittivo. Sono stati, però, sviluppati dei polygenic risk score che mettono insieme tutte le varianti associate ad una determinata patologia e determinano una sorta di punteggio di rischio individuale.

CONFRONTO TRA L'ANALISI DI LINKAGE E GLI STUDI DI ASSOCIAZIONE

Entrambi vengono utilizzati come indagine iniziale per effettuare il mappaggio genetico e la ricerca nel genoma di geni candidati che si presume siano coinvolti nell'insorgenza di diverse malattie. Inoltre, entrambe le metodologie sfruttano gli stessi "strumenti" ovvero utilizzano come marker dei polimorfismi di singolo nt, cioè gli SNP.

Ci sono, però, sostanziali differenze:

ANALISI DI LINKAGE: per i caratteri monogenici mendeliani, determinati da varianti di un singolo gene, in cui c'è una forte correlazione tra la presenza di una determinata variante e la manifestazione del fenotipo, si utilizza un'Analisi di Linkage. Il punto di partenza di questa analisi è conoscere il modello di ereditarietà del gene malattia che ci permette di stabilire un certo genotipo per il locus malattia, partendo dall'osservazione dei fenotipi in un pedigree. Gli studi di linkage vengono effettuati studiando, appunto i membri di famiglie imparentati tra di loro, con lo scopo di vedere la segregazione del fenotipo di interesse e cercare di individuare marcatori molecolari che durante la meiosi segregano insieme all'ipotetico locus responsabile del fenotipo. L'analisi di linkage permette di mappare la distanza di due loci sul cromosoma andando a determinare la frequenza con la quale questi due loci ricombinano tra di loro alla meiosi. Frequenza di ricombinazione (teta) : due loci distanti, segregano indipendentemente, e la probabilità di essere ereditati insieme è di $teta:1/2$ con $teta=50\%$, se due loci sono vicini sullo stesso cromosoma $teta<50\%$, quindi quanto più sono vicini tanto minore sarà la frequenza di ricombinazione. L'analisi di Linkage viene effettuata con un modello probabilistico Lod Score che calcola la probabilità di linkage cercando di estrapolare il massimo delle informazioni dal pedigree in esame. Lod Score: \log_{10} probabilità di osservare linkage per i dati ottenuti \ p. di osservare gli stessi dati in assenza di linkage.

LodScore >0 → è più probabile l'ipotesi di linkage rispetto all'indipendenza

LodScore <0 → dati in accordo con l'ipotesi di indipendenza

LodScore=3 → 1000 volte più probabile l'ipotesi di linkage rispetto all'associazione → L.S >3 → c'è linkage tra i due loci.

STUDI DI ASSOCIAZIONE GWAS: se si hanno, invece, dei caratteri multifattoriali (quantitativi o qualitativi) la quale espressione fenotipica è la risultante del contributo di tanti alleli su geni diversi, non è possibile predire un meccanismo di ereditarietà. Quindi gli studi linkage non sono efficaci in questo caso per mappare fattori genetici implicati in caratteri multifattoriali in quanto non è possibile dire con certezza se sono avvenuti eventi di ricombinazione tra un marcatore ed il gene malattia. Inoltre, gli studi di associazione non riguardano individui imparentati tra loro, ma individui non imparentati appartenenti ad una popolazione. Gli studi di associazione sono studi di popolazione che vanno a vedere se ci sono associazioni significative tra una determinata variante genetica ed un fenotipo ma non sono in grado di osservare da vicino se avvengono eventi di crossing over tra due loci, ci si limita a dire che la variante genetica ed il fenotipo di interesse tendono a presentarsi insieme più frequentemente di quando mi aspetti in base al caso. Questi non hanno un'ipotesi di partenza (invece gli studi di linkage si) servono per avere una scansione del genoma per determinare i fattori di rischio implicati. La strategia è basata sul fatto che esiste un LD che descrive il fatto che varianti genetiche che si trovano sullo stesso cromosoma non si trovano in modo casuale l'uno rispetto all'altro ma tendono ad essere associate.

Per studiare l'associazione fra un allele del marcatore ed il fenotipo si utilizza un test di associazione sugli alleli, dove in base al risultato del test statistico χ^2 si ricava un p-value da confrontare con i livelli di significatività del test, se il test da risultato significativo (p-value < valore significativo) allora indica associazione.