

# Glossario

Laboratorio di Bioinformatica - Presentazione NAR

**Chiara Solito e Aurelia Timis**

Corso di Laurea in Bioinformatica  
Università degli studi di Verona  
A.A. 2021/22

# 1 Lessico e nozioni di base

## Nozioni di Statistica

**Single Evidence Score** Tra i livelli di Evidenza, anche definiti come gerarchia dell'evidenza, assegnati agli studi, basandosi sulla qualità metodologica della loro progettazione, validità e applicabilità alla cura del paziente: è il sesto livello di evidenza - Evidenza da un singolo studio descrittivo o qualitativo.

**Proxy** Un proxy è una misura indiretta del risultato desiderato che è esso stesso fortemente correlato a quel risultato. È comunemente usato quando le misure dirette del risultato non sono osservabili e/o non disponibili.

**Credible Sets** *Set Credibili*

sono l'insieme più piccolo di varianti, selezionate in modo tale che la loro stima di copertura rettificata soddisfi la copertura del target. Ad esempio: in letteratura, gli autori in genere riferiscono di aver trovato un insieme credibile al 90% di cui sono fiduciosi almeno al 90% contenga la vera variante causale.

**PP - Posterior Probability** *Probabilità a posteriori*

In statistica bayesiana, la probabilità a posteriori di un evento aleatorio o di una proposizione incerta, è la probabilità condizionata che è assegnata dopo che si è tenuto conto dell'informazione rilevante o degli antefatti relativi a tale evento aleatorio o a tale proposizione incerta. Similmente, la distribuzione di probabilità a posteriori è la distribuzione di una quantità incognita, trattata come una variabile casuale, condizionata sull'informazione posta in evidenza da un esperimento o da un processo di raccolta di informazione rilevanti (es. un'ispezione, un'indagine conoscitiva, ecc.).

**P-Value** La probabilità, per un'ipotesi supposta vera (ipotesi nulla), di ottenere risultati ugualmetnte o meno compatibili, di quelli osservati durante i test, con la suddetta ipotesi.

---

## Nozioni sui GWAS

**GWAS** *Genome Wide Association Study*

Un approccio della ricerca genetica per associare, a specifiche variazioni genetiche, particolari malattie.

**Varianti Causali** Nell'ambito degli studi di associazione, le varianti genetiche responsabili del segnale di associazione in un locus sono indicate nella letteratura come VARIANTI CAUSALI. Esse hanno un effetto biologico sul fenotipo.

**Lead Variant** La variante col miglior p-value per combinazioni gene/fenotipo significative.

**Tag Variants** Varianti rappresentative in una regione del genoma con un alto linkage disequilibrium.

**Trait-Associated Loci** *Loci tratto associati*

Un locus a cui è associato un particolare tratto fenotipico.

**QTL** *Quantitative trait loci*

Un locus dei caratteri quantitativi (ovvero tratti che possono essere studiati e indagati mediante parametri numerici) è un locus che si correla con la variazione di un tratto quantitativo nel fenotipo di una popolazione di organismi.

**Linkage Disequilibrium** Linkage disequilibrium (LD) è l'associazione non casuale di alleli di diversi loci. Non esiste una singola statistica migliore che quantifica l'entità di LD. Sono state proposte diverse statistiche utili per scopi diversi.

**Fine Mapping** Dal momento che i risultati dei GWAS non sempre ci danno una sintesi completa delle statistiche, e avendo a disposizione solo le Variant Lead, dobbiamo ampliare le informazioni con le Variants Tag, in modo da avere un insieme più completo.

Il fine mapping è uno dei metodi con quei viene eseguita questa operazione: è una tecnica dei GWAS per identificare la varianti genetiche che possono influenzare causalmente il tratto esaminato, in particolare cerca di determinare la variante genetica responsabile della malattia o del fenotipo analizzato.

**Linkage Disequilibrium Espansione** Sappiamo che il genoma viene ereditato a blocchi e ogni blocco è definito da un aplotipo, che a sua volta è definito da un insieme di SNPs. OTG prende in considerazione l'LD perchè pazienti che hanno ereditato lo stesso segmento cromosomico, definito dal medesimo aplotipo, possono aver ereditato anche la stessa mutazione. L'LD ha permesso nelle analisi di associazioni, di studiare solo gli SNPs necessari a identificare il blocco di LD in disequilibrium.