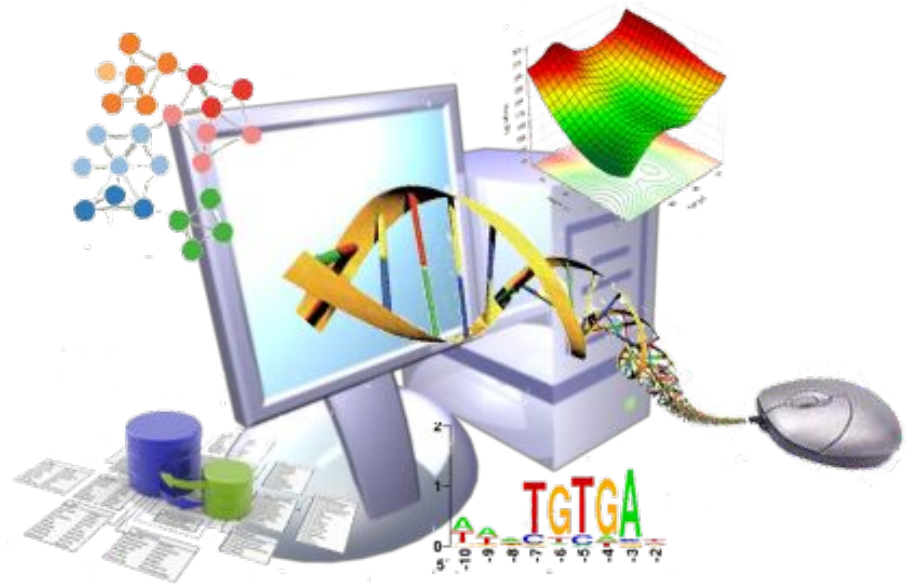


Campi di applicazione delle NGS

- GENOMICA
- TRASCRITTOMICA
- METAGENOMICA
- EPIGENOMICA
- POPULATION GENOMICS



GENOMICA



Cos'è la genomica?

E' una branca della biologia molecolare che si occupa dello studio completo del genoma di un organismo. Coinvolge il sequencing, la mappatura e l'analisi dei genomi. Tutto questo è reso possibile grazie ai progressi tecnologici di sequencing del DNA, che hanno reso più efficiente ed economico il processo di determinazione dell'ordine delle basi nucleotidiche.

Cosa possiamo determinare attraverso l'analisi genomica?

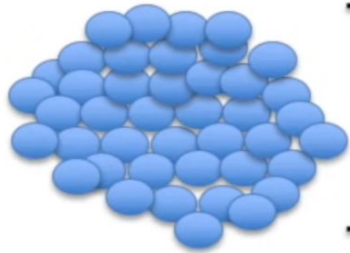
- **Identificazione dei geni:** ci consente di identificare e mappare i geni all'interno del genoma. Questo è essenziale per comprendere quali geni sono presenti in un organismo e dove si trovano sulla sua struttura genetica.
- **Annotazione funzionale dei geni:** determinare le funzioni dei geni. Questo è fondamentale per capire come i geni influenzano i processi biologici e fisiologici.
- **Variazioni genetiche:** può rivelare variazioni genetiche tra individui o popolazioni. Queste informazioni sono importanti per la medicina genomica e per la comprensione delle malattie genetiche.
- **Evoluzione e filogenesi:** utilizzata per studiare l'evoluzione delle specie e le relazioni filogenetiche tra di esse. Confrontando i genomi di diverse specie, è possibile determinare quanto sono imparentate e come si sono evolute nel corso del tempo.

- **Medicina genomica:** consente di effettuare una diagnosi più precisa delle malattie genetiche e una personalizzazione dei trattamenti in base al profilo genetico di un individuo. Ciò può migliorare l'efficacia dei trattamenti e ridurre gli effetti collaterali.
- **Biotecnologia agricola:** utilizzata per migliorare le colture e la produzione alimentare. Questo comprende la selezione di piante resistenti alle malattie, la migliorata qualità nutrizionale dei prodotti alimentari e la creazione di colture più adattabili a condizioni ambientali sfavorevoli.
- **Conservazione della biodiversità:** permette di studiare la diversità genetica nelle popolazioni di specie minacciate e di sviluppare strategie per la loro conservazione.
- **Farmacogenomica:** per comprendere come la variabilità genetica influenzi la risposta dei pazienti ai farmaci e quindi aiutare a personalizzare i trattamenti farmacologici per massimizzare l'efficacia e ridurre gli effetti collaterali.

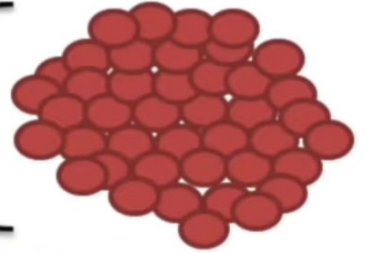
TRASCRIPTOMICA

● = a normal neural cell

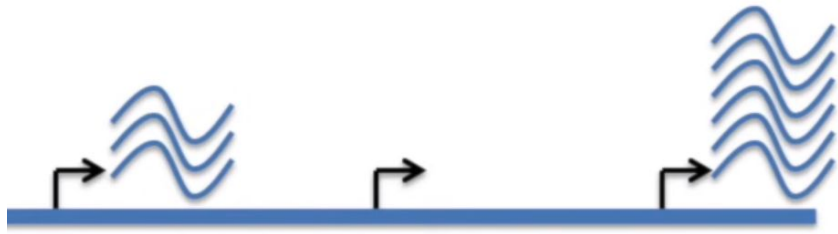
● = a mutated neural cell



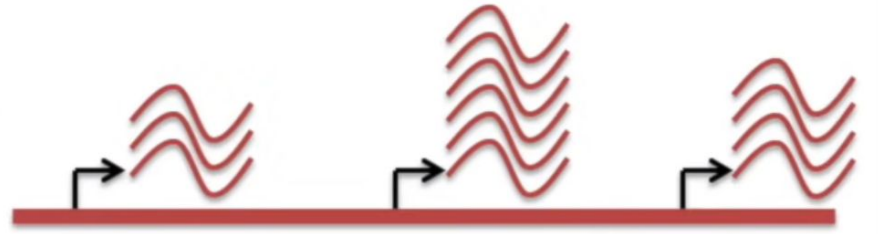
A bunch of normal neural cells.



A bunch of mutated neural cells.



We can use RNA-seq to measure gene expression in normal cells...



... then use it to measure gene expression in mutated cells...

Assegnamento delle read ai geni e quantificazione dell'espressione genica

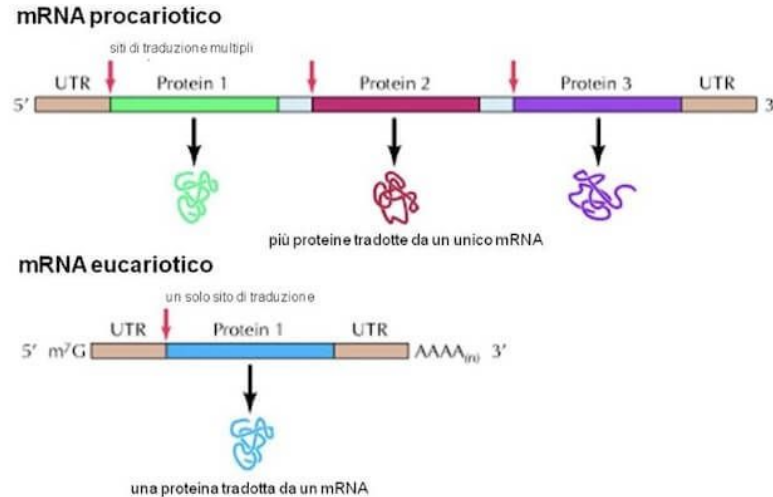
- **Il numero di read che mappano su un gene è proporzionale al livello di espressione**
- I valori di espressione ottenuti dall'RNA-Seq deriva dalla conta diretta delle read che mappano su un gene: **misura digitale**
- Non richiede la conoscenza a priori delle posizioni dei geni
- Intervallo dinamico più ampio comparato a microarray

Cos'è la trascrittomica?

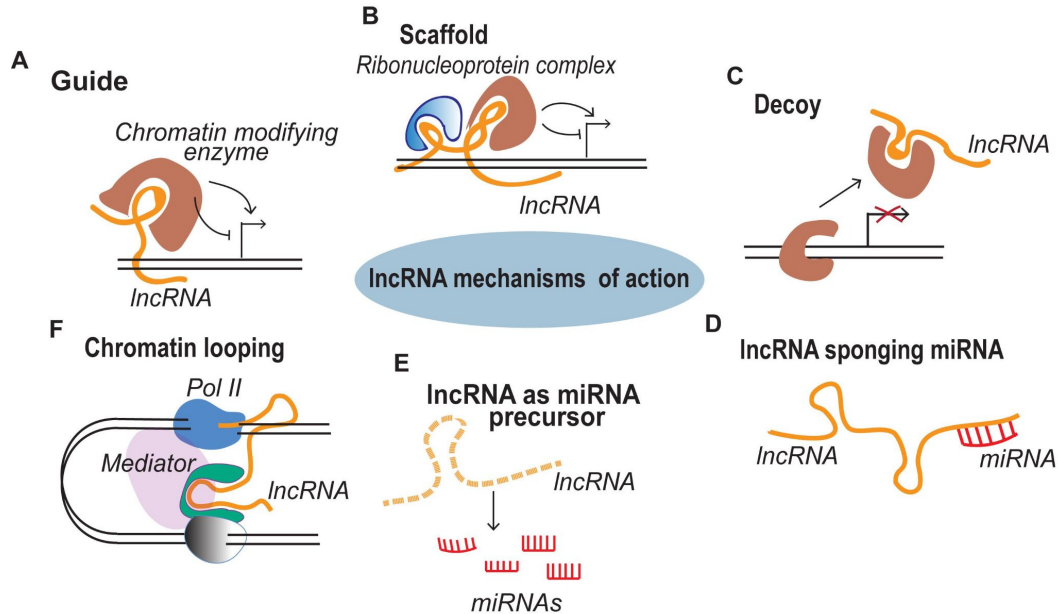
La trascrittomica è un campo della biologia molecolare che si occupa dello studio dei trascritti cellulari, ovvero delle molecole di RNA messaggero (mRNA) prodotte da un gene durante il processo di trascrizione. Questi trascritti svolgono un ruolo cruciale nella regolazione dell'espressione genica e nella sintesi delle proteine all'interno delle cellule.

Cosa va ad amplificare la trascrittomicca

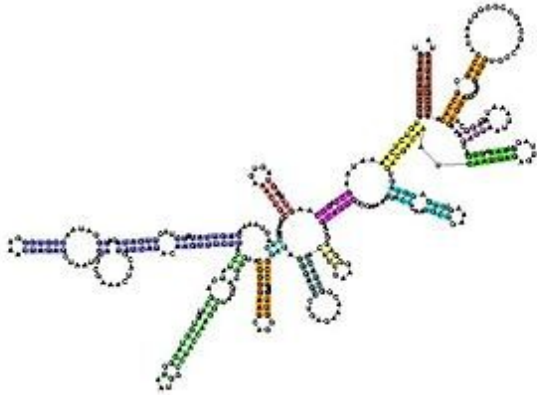
- **Trascritti di RNA messaggero (mRNA):** si concentra principalmente sull'amplificazione degli mRNA, che rappresentano il prodotto finale del processo di trascrizione genica. Questi trascritti contengono le informazioni per la sintesi delle proteine e riflettono l'espressione dei geni.



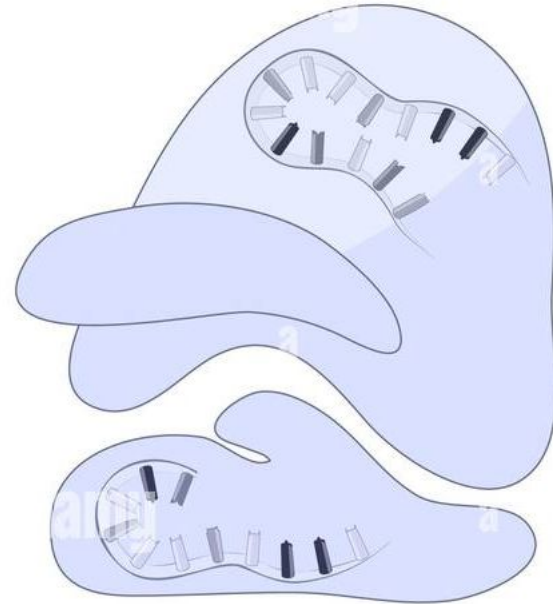
- **RNA non codificanti (lncRNA)**: sono trascritti di RNA che non codificano per proteine ma possono avere ruoli nel regolare l'espressione genica e altri processi cellulari.



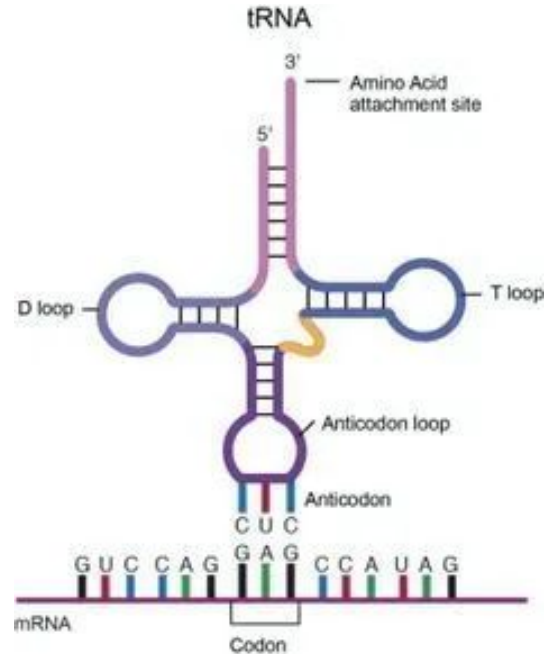
- **RNA ribosomali (rRNA)**: costituiscono una parte significativa delle molecole di RNA all'interno delle cellule e sono coinvolti nella sintesi delle proteine.



rRNA
(ribosomal RNA)



- **RNA transfer (tRNA)**: svolgono un ruolo chiave nel processo di traduzione delle proteine, possono essere inclusi nelle analisi trascrittomiche.



Che ruolo ha la bioinformatica in questo campo?

- **Allineamento e assemblaggio di sequenze:** I dati di RNA-seq possono contenere milioni di sequenze di RNA che devono essere allineate al genoma di riferimento o assemblate in trascritti completi.
- **Quantificazione dell'espressione genica:** algoritmi per quantificare l'espressione genica dai dati di RNA-seq o microarray, che possono essere utilizzati per confrontare l'espressione tra campioni diversi.
- **Analisi di splicing alternativo:** rivelare informazioni sull'uso di diversi esoni e giunzioni splicing alternativi.

- **Identificazione di nuovi trascritti e isoforme:** gli algoritmi di identificazione delle nuove isoforme aiutano a scoprire nuovi trascritti e varianti di RNA.
- **Analisi della funzione biologica:** interpretazione biologica dei dati trascrittomici, ad esempio, identificando le vie metaboliche coinvolte o le funzioni cellulari influenzate dai cambiamenti nell'espressione genica.
- **Integrazione con altre dati omici:** integrare dati trascrittomici con dati genomici, proteomici e metabolomici per ottenere una visione più completa dei processi biologici.
- **Predizione di marcatori biologici:** identificare geni o trascritti che fungono da potenziali marcatori biologici per malattie o condizioni specifiche.

METAGENOMICA



Cos'è la metagenomica?

La metagenomica è un campo della biologia molecolare che si occupa dello studio del DNA estratto da campioni ambientali complessi, come suoli, acque, campioni di microbioma umano e altri habitat naturali. A differenza della genomica tradizionale, che si concentra sullo studio del DNA di un organismo specifico, la metagenomica mira a analizzare l'insieme del materiale genetico presente in una comunità microbica o in un ambiente specifico.

Che ruolo ha la bioinformatica in questo campo?

Identificazione e classificazione degli organismi:

- **Tassonomia:** i geni o le sequenze di DNA vengono confrontati con database di riferimento per identificare e classificare gli organismi presenti nei campioni.
- **Metagenomica comparativa:** si valuta la composizione tassonomica della comunità microbica attraverso l'analisi delle sequenze e la loro correlazione con organismi noti.

Analisi della diversità:

- **Diversità alfa e beta:** si calcola la diversità all'interno dei campioni (diversità alfa) e tra i campioni (diversità beta) per valutare la variabilità nella composizione delle comunità microbiche.
- **Analisi di clustering:** si utilizzano algoritmi di clustering per raggruppare campioni o sequenze simili.

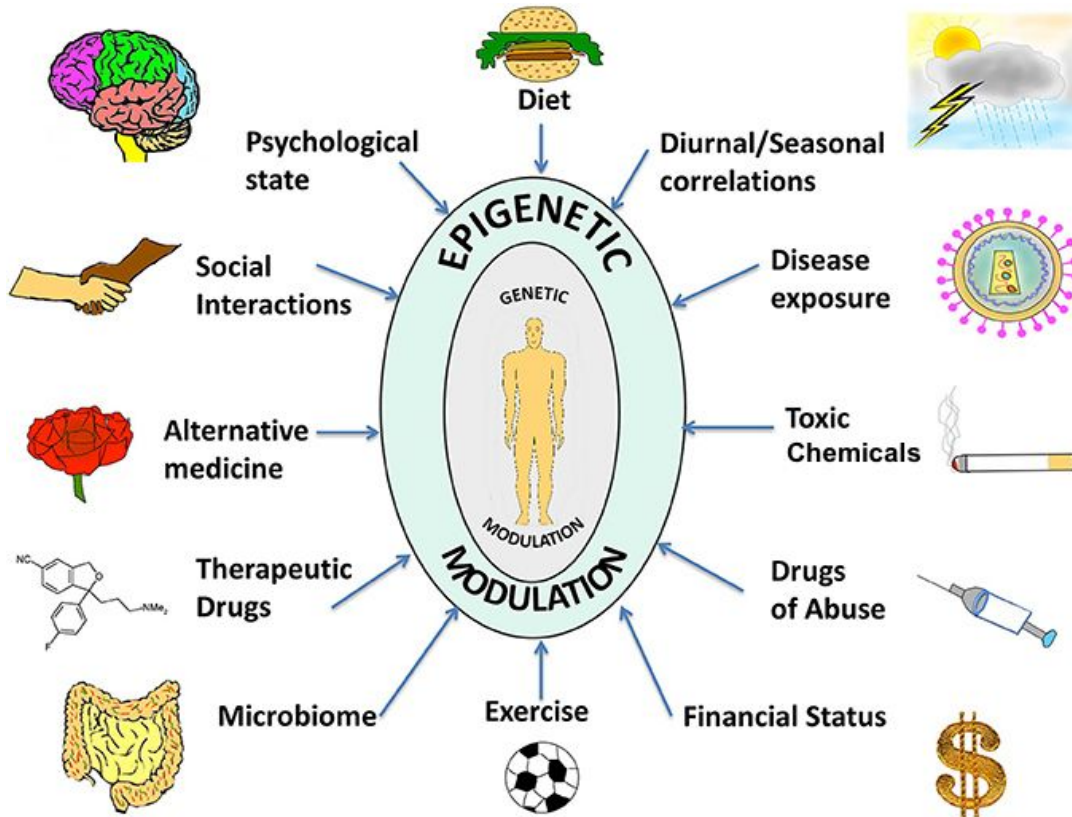
Predizione delle funzioni e delle interazioni: si possono utilizzare strumenti per prevedere le funzioni metaboliche e le interazioni tra membri della comunità microbica.

Principali applicazioni e risultati ottenuti

- **Caratterizzazione della diversità microbica:** Uno degli obiettivi principali della metagenomica è identificare e caratterizzare la diversità dei microorganismi presenti in un ambiente specifico. Questo può aiutare a comprendere meglio la composizione e l'ecologia delle comunità microbiche in diversi contesti ambientali.
- **Identificazione di nuovi organismi e geni:** permette di scoprire nuovi organismi microbici e geni che potrebbero avere importanza biologica o applicazioni biotecnologiche.
- **Analisi delle funzioni genetiche:** identificare le funzioni genetiche presenti nelle comunità microbiche. Questo include la ricerca di geni coinvolti in processi biologici chiave, come la decomposizione della materia organica, la fissazione dell'azoto, la produzione di metaboliti utili, ecc.
- **Studi ecologici:** informazioni sull'ecologia microbica, inclusi i ruoli dei microorganismi all'interno di un ecosistema, le loro interazioni reciproche e con l'ambiente, e le risposte alle variazioni ambientali.

- **Risposta a disturbi ambientali:** L utilizzata per studiare come le comunità microbiche rispondono a disturbi ambientali, come l'inquinamento, i cambiamenti climatici o le perturbazioni antropiche.
- **Applicazioni biotecnologiche e mediche:** applicazioni significative nella scoperta di nuove molecole biologiche, enzimi utili, antibiotici e altre risorse bioattive che possono essere utilizzate in medicina, agricoltura e industria.
- **Progettazione di strategie di gestione ambientale:** ad esempio strategie per il ripristino di ecosistemi o il controllo di contaminazioni microbiche.
- **Comprendere il microbioma umano:** cioè l'insieme dei microbi presenti nel corpo umano, e il suo impatto sulla salute umana.

EPIGENOMICA



Cos'è l'epigenomica?

L'epigenomica è un campo scientifico che studia le modificazioni chimiche e strutturali che avvengono nel DNA e nelle proteine associate ad esso. Queste modificazioni influenzano l'espressione dei geni, senza alterare la sequenza del DNA in sé.

Inoltre è un campo di grande importanza nella ricerca biologica/medica, poiché fornisce una comprensione approfondita di come fattori esterni come la dieta, l'ambiente e lo stile di vita possano influenzare la salute e la suscettibilità alle malattie (compresi i tumori e le malattie genetiche).

Che ruolo ha la bioinformatica in questo campo?

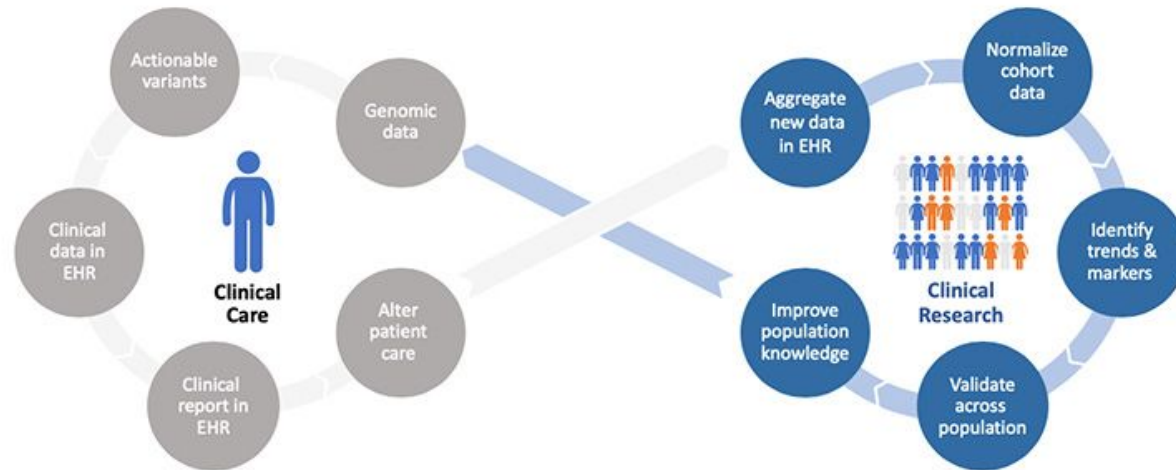
- **Analisi dei dati epigenomici:** la generazione di dati epigenomici, come la mappatura delle modifiche chimiche del DNA o delle proteine istoniche, produce enormi quantità di informazioni. La bioinformatica sviluppa algoritmi e strumenti per elaborare, analizzare e interpretare questi dati, consentendo l'identificazione di modelli, regioni di interesse e relazioni tra modificazioni epigenetiche e regolazione genica.
- **Predizione delle regioni regolatorie:** utilizzata per identificare potenziali regioni regolatorie nel genoma in base alle modifiche epigenetiche presenti.
- **Ricerca di marcatori epigenetici:** che possono essere utilizzati per scopi di diagnosi e prognosi in medicina. Ad esempio, l'identificazione di specifiche modificazioni epigenetiche associate a malattie può essere utile per la diagnosi precoce e la personalizzazione dei trattamenti.

Principali applicazioni e risultati ottenuti

- **Mappatura delle modifiche epigenetiche:** permettono di mappare le modifiche epigenetiche, come la metilazione del DNA o le modifiche delle proteine istoniche, in tutto il genoma. Questo fornisce una comprensione dettagliata di quali regioni del DNA sono modificate e in che modo.
- **Identificazione di marcatori epigenetici:** possono rivelare modifiche specifiche associate a condizioni patologiche, come il cancro o le malattie neurodegenerative. Questi marcatori possono essere utilizzati per la diagnosi precoce, la prognosi e lo sviluppo di terapie mirate.
- **Regolazione dell'espressione genica:** influenzano direttamente l'espressione dei geni. Le analisi epigenomiche consentono di identificare quali geni sono attivati o disattivati in risposta a determinate modifiche epigenetiche, fornendo una comprensione dei meccanismi di regolazione genica.

- **Studio della differenziazione cellulare:** Le modifiche epigenetiche sono fondamentali per la differenziazione cellulare, ovvero il processo attraverso il quale le cellule si specializzano in tipi specifici di cellule (ad esempio, cellule del sangue o neuroni). Le analisi epigenomiche possono rivelare come queste modifiche guidino la differenziazione cellulare.
- **Risposta all'ambiente:** Le modifiche epigenetiche possono cambiare in risposta a fattori ambientali come la dieta, l'esposizione a sostanze chimiche o lo stress. Queste analisi possono fornire informazioni su come l'ambiente influisce sulla biologia.
- **Personalizzazione dei trattamenti:** Nel campo della medicina personalizzata, le analisi epigenomiche possono aiutare a determinare quali terapie potrebbero funzionare meglio per un individuo, in base al suo profilo epigenetico.
- **Ricerca di farmaci:** L'identificazione di modifiche epigenetiche specifiche può suggerire bersagli farmacologici per lo sviluppo di nuovi farmaci mirati.

POPULATION GENOMICS



Cos'è la population genomics?

La population genomics, o genomica delle popolazioni, è un campo di studio che si concentra sulla genetica delle popolazioni umane o di altre specie. Si tratta di una disciplina multidisciplinare che combina la genetica, la biologia evolutiva, la biologia delle popolazioni e la genomica per analizzare le variazioni genetiche all'interno di gruppi di individui all'interno di una specie.

Che ruolo ha la bioinformatica in questo campo?

- **Analisi delle sequenze genomiche:** aiuta a elaborare e analizzare le sequenze genomiche di individui appartenenti a diverse popolazioni. Ciò include l'allineamento delle sequenze, la rilevazione delle varianti genetiche (come SNP e indel) e l'identificazione delle regioni geneticamente conservate o variabili.
- **Filtraggio e interpretazione dei dati:** consente di filtrare i dati genomici per identificare varianti genetiche rilevanti o regioni di interesse. Ciò può aiutare a individuare associazioni tra geni e malattie o a identificare segni di adattamento evolutivo.
- **Analisi delle popolazioni e delle strutture genetiche:** utilizza metodi statistici e algoritmi per analizzare la struttura genetica delle popolazioni, identificare cluster o gruppi di individui geneticamente simili e ricostruire alberi filogenetici che rappresentano le relazioni tra le popolazioni.
- **Studio delle traiettorie evolutive:** permette di esaminare le traiettorie evolutive delle popolazioni, inclusi i cambiamenti genetici nel tempo e nello spazio. Questo è utile per comprendere come le popolazioni si sono adattate a diversi ambienti e condizioni.

Principali applicazioni e risultati ottenuti

- **Tracciamento delle migrazioni umane:** ricostruire le migrazioni umane storiche, identificando le rotte migratorie e le popolazioni coinvolte. Ad esempio, è stato possibile tracciare le migrazioni dell'uomo moderno fuori dall'Africa e il popolamento delle diverse regioni del mondo.
- **Identificazione delle varianti genetiche associate alle malattie:** come il diabete, l'Alzheimer, il cancro e le malattie cardiache. Questi risultati hanno aperto la strada allo sviluppo di terapie mirate e strategie di prevenzione.
- **Adattamento evolutivo:** gli studi hanno rivelato come le popolazioni umane si siano adattate all'ambiente circostante attraverso cambiamenti genetici. Ad esempio, sono stati identificati geni coinvolti nell'adattamento all'altitudine, alla dieta e all'esposizione alle malattie.
- **Ibridazione e introgressione genetica:** cioè il passaggio di materiale genetico tra popolazioni. Questi processi possono influenzare la diversità genetica e l'evoluzione delle specie.

- **Diversità genetica nelle specie non umane:** utilizzata per studiare la diversità genetica nelle specie non umane, aiutando nella conservazione della biodiversità e nella gestione delle popolazioni in pericolo di estinzione.
- **Storia evolutiva delle popolazioni:** alberi filogenetici basati sulla genomica delle popolazioni hanno permesso di comprendere meglio la storia evolutiva delle popolazioni umane e delle specie non umane, inclusi i tempi di divergenza e i legami filogenetici.
- **Conservazione della biodiversità:** nelle specie non umane, la genomica delle popolazioni fornisce informazioni critiche per la conservazione delle specie a rischio, aiutando a identificare popolazioni geneticamente distintive e a pianificare strategie di conservazione.

Software e strumenti specializzati di maggior utilizzo

Strumenti per il sequenziamento e l'assemblaggio del DNA:

- **Bowtie2:** Un programma per allineare letture di sequenziamento del DNA contro un genoma di riferimento.
- **BWA (Burrows-Wheeler Aligner):** Utilizzato per l'allineamento di letture di sequenziamento contro un genoma di riferimento.
- **SPAdes:** Un programma di assemblaggio genomico per creare sequenze complete di genomi da dati di sequenziamento.

Annotazione genomica:

- **Augustus:** Un software per la previsione di geni all'interno del genoma.
- **GeneMark:** Utilizzato per l'identificazione di geni in sequenze genomiche.
- **AUGUSTUS:** Un programma per la predizione dei geni nelle sequenze genomiche.

Analisi di variazioni genetiche:

- **GATK (Genome Analysis Toolkit):** Utilizzato per identificare SNP e indel da dati di sequenziamento.
- **Samtools:** Una suite di programmi per la manipolazione di file SAM/BAM, comunemente utilizzati per l'analisi di dati di sequenziamento.
- **VarScan:** Utilizzato per rilevare varianti genetiche in dati di sequenziamento.

Software e strumenti utilizzati per la trascrittomica:

- **TopHat e HISAT:** programmi utilizzati per l'allineamento dei dati di sequenziamento RNA-Seq sul genoma di riferimento.
- **Cufflinks e StringTie:** utilizzati per l'assemblaggio delle trascrizioni da dati di sequenziamento RNA-Seq e per la quantificazione dell'espressione genica. Forniscono informazioni sulle isoforme trascrizionali e sui livelli di espressione genica.
- **STAR (Spliced Transcripts Alignment to a Reference):** è uno strumento di allineamento RNA-Seq ad alte prestazioni che consente di identificare giunzioni di splicing e mappare i trascritti in modo efficiente.
- **Salmon:** è un programma per la quantificazione dell'espressione genica che utilizza un approccio basato su k-mer invece dell'allineamento diretto dei dati RNA-Seq. È noto per la sua velocità.
- **Trinity:** è un assembler de novo per dati RNA-Seq, che permette di assemblare trascritti in assenza di un genoma di riferimento.

Bioinformatica strutturale:

- **Modeller:** Utilizzato per la modellazione di proteine e la previsione di strutture tridimensionali.
- **PyMOL:** Un visualizzatore di strutture tridimensionali di proteine.
- **Rosalind:** Una piattaforma per l'apprendimento della bioinformatica strutturale.

Software e strumenti utilizzati per la metagenomica:

- **MG-RAST (Metagenomics Rapid Annotations using Subsystems Technology):** è un servizio online per l'analisi e l'annotazione di dati metagenomici. Fornisce informazioni sulla composizione della comunità microbica e sulle funzioni genetiche.
- **MetaPhlAn:** è uno strumento per l'identificazione dei microorganismi presenti nei campioni metagenomici basandosi su sequenze di marker specifici. È utile per determinare la composizione tassonomica delle comunità microbiche.
- **Kraken:** è un programma per l'assegnazione tassonomica rapida di sequenze metagenomiche. Utilizza un database di riferimento per attribuire le sequenze a organismi specifici.
- **MEGAN (MetaGenome Analyzer):** è un software per l'analisi della composizione tassonomica e funzionale delle comunità microbiche. Può essere utilizzato per visualizzare e interpretare i dati metagenomici.

Software utilizzati per l'epigenomica:

- **Bismark:** è uno strumento utilizzato per il mappaggio e l'analisi dei dati di sequenziamento del bisolfuro di metile (BS-Seq), che permette di rilevare modificazioni del DNA come la metilazione del citosina.
- **MACS (Model-based Analysis of ChIP-Seq):** è uno strumento ampiamente utilizzato per l'analisi dei dati di sequenziamento di immunoprecipitazione della cromatina (ChIP-Seq), che consente di identificare i picchi di legame delle proteine con il DNA, come le proteine di legame al DNA o le modificazioni epigenetiche.
- **HOMER (Hypergeometric Optimization of Motif EnRichment):** è un software per l'analisi dei dati ChIP-Seq che permette di identificare picchi di legame delle proteine, scoprire motivi di legame e condurre analisi di annotazione genica.
- **DifferentialPeakCaller:** Questo software è specifico per l'identificazione di regioni epigenomiche differenzialmente metilate o modificate tra diverse condizioni sperimentali.

Software e strumenti utilizzati per la population genomics:

- **PLINK:** è uno dei software più ampiamente utilizzati per l'analisi delle associazioni genomica in studi di genomica delle popolazioni. È utilizzato per condurre analisi di associazione con SNP (polimorfismi a singolo nucleotide) e per calcolare statistiche come l' F_{ST} per la diversità genetica tra popolazioni.
- **ADMIXTURE:** è uno strumento utilizzato per l'analisi della struttura genetica delle popolazioni. Consente di stimare la proporzione di ascendenza ancestrale in individui o gruppi di popolazione basandosi su dati SNP.
- **EIGENSOFT:** è una suite di programmi utilizzata per l'analisi della struttura genomica delle popolazioni, la riduzione della dimensionalità dei dati genomici e l'identificazione di popolazioni outlier.

Analisi filogenetica:

- **PhyML e RAxML:** Programmi per la costruzione di alberi filogenetici basati su sequenze genomiche.
- **MEGA (Molecular Evolutionary Genetics Analysis):** Un software completo per l'analisi filogenetica.
- **TreeMix:** è utilizzato per costruire alberi filogenetici delle popolazioni basati su dati genomici e per stimare eventi di migrazione tra popolazioni.
- **BEAST (Bayesian Evolutionary Analysis by Sampling Trees):** è uno strumento utilizzato per l'analisi filogenetica bayesiana e l'analisi dell'evoluzione delle popolazioni basata su dati genetici.