Corso Bioinformatica 1 - 2024

Programma del corso di Bioinformatica 1 - 2024:

Introduzione alle banche dati biologiche primarie e secondarie.

Algoritmi di allineamento a coppie di sequenze.

Algoritmi euristici per l'allineamento di sequenze in banca dati.

Algoritmi per l'allineamento multiplo.

Metodi di costruzione degli alberi filogenetici.

Metodi bioinformatici per la predizione della struttura dell'RNA.

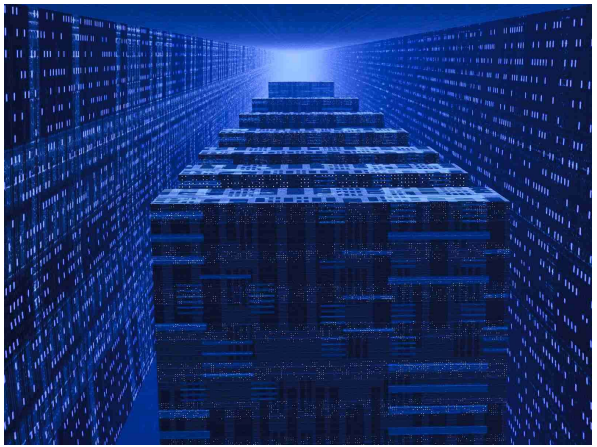
Metodi bioinformatici per l'analisi di sequenze proteiche.

Algoritmi di predizione della struttura secondaria e del folding delle proteine.

Analisi delle interazioni proteiche.

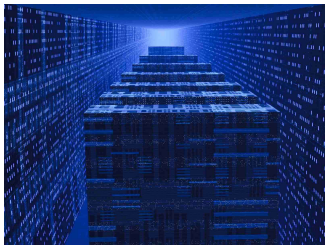
Il docking molecolare.

LE BANCHE DATI



- Introduzione alla bioinformatica: banche dati biologiche
- Importanza delle banche dati biologiche
- Caratteristiche di una banca dati biologica
- Tipi di banche dati biologiche: primarie e secondarie
- Banche dati di sequenza
- Banche dati di struttura
- Banche dati di funzione
- Banche dati di espressione
- Utilizzo pratico delle banche dati biologiche
- Riepilogo dell'importanza delle banche dati biologiche nella ricerca biologica

LE BANCHE DATI

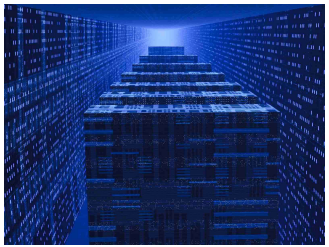


- Cos'è la bioinformatica?

La bioinformatica è una disciplina interdisciplinare che utilizza **l'informatica** e le tecniche di analisi dei dati per studiare i **processi biologici**.

Attraverso l'uso di algoritmi e strumenti computazionali, la bioinformatica consente l'analisi di grandi quantità di dati biologici, come ad esempio sequenze genomiche, dati di espressione genica e informazioni di struttura proteica.

LE BANCHE DATI

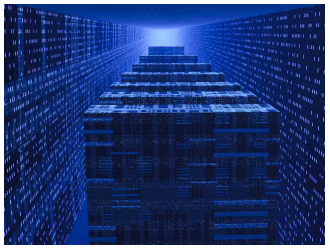


- Ruolo della bioinformatica nella ricerca biologica

La bioinformatica ha un ruolo fondamentale nella ricerca biologica, in quanto consente ai ricercatori di **gestire**, **analizzare** e **interpretare** i dati biologici in modo efficiente e accurato.

Ad esempio, la bioinformatica viene utilizzata per identificare nuovi geni e proteine, per analizzare l'evoluzione di sequenze genomiche e per studiare la struttura e la funzione delle proteine.

LE BANCHE DATI

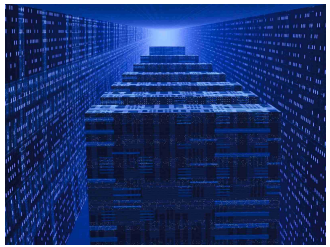


- Tecniche e strumenti di bioinformatica

La bioinformatica utilizza una vasta gamma di **tecniche** e **strumenti**, tra cui algoritmi per l'allineamento di sequenze, algoritmi per la modellizzazione di strutture proteiche, la predizione di funzioni proteiche, algoritmi per l'analisi di reti di interazione proteina-proteina e l'analisi di dati di espressione genica.

Ad esempio, l'allineamento di sequenze viene utilizzato per confrontare sequenze di DNA o proteine e identificare regioni conservate, mentre la modellizzazione di strutture proteiche viene utilizzata per prevedere la struttura tridimensionale di una proteina

LE BANCHE DATI



Uno dei principali strumenti della bioinformatica per l'analisi e l'interpretazione dei dati biologici sono le banche dati biologiche.

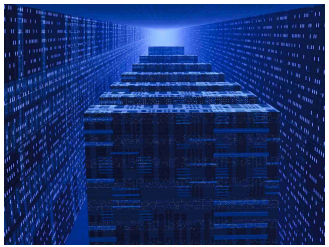
Cos'è una banca dati ?

Una banca dati (o **database**) è un insieme organizzato di dati, memorizzati in un computer, che possono essere facilmente recuperati, aggiornati e interrogati. In altre parole, una banca dati è un sistema di archiviazione e gestione dei dati che consente di organizzare e consultare grandi quantità di informazioni.

Cos'è una banca dati biologica?

Una banca dati biologica è una banca dati che contiene informazioni biologiche, come sequenze genomiche, strutture proteiche, funzioni di proteine, espressione genica e molto altro. Le banche dati biologiche sono state create per gestire e archiviare grandi quantità di dati biologici, consentendo ai ricercatori di accedere a queste informazioni in modo efficiente e di utilizzarle per i loro studi scientifici.

LE BANCHE DATI

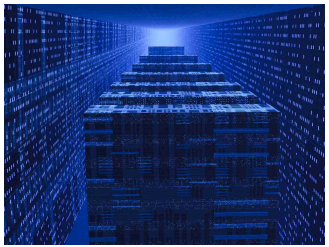


- Importanza delle banche dati biologiche

La ricerca biologica moderna genera **enormi quantità di dati,** dalle sequenze genomiche ai dati di espressione genica, che possono essere utilizzati per rispondere a importanti domande biologiche.

Tuttavia, per fare ciò in modo efficace, è necessario archiviare e gestire questi dati in modo accurato e organizzato

LE BANCHE DATI

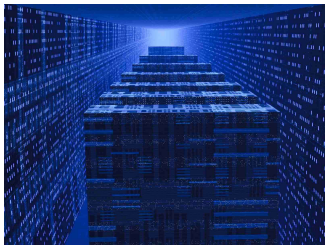


- Quantità di dati

Per avere un'idea della quantità di dati biologici generati dalla ricerca moderna, ad esempio, il Genoma Umano, il primo genoma umano sequenziato nel 2001, contiene circa **3 miliardi di basi** di DNA, e l'intero insieme di genomi umani conosciuti (compresi i genomi mitocondriali) contiene oltre 100 miliardi di basi di DNA.

Inoltre, una singola sequenza proteica può contenere migliaia di amminoacidi, ciascuno dei quali può influenzare la struttura e la funzione della proteina. Pertanto, la comprensione della struttura e della funzione delle proteine richiede una comprensione dettagliata della sequenza di amminoacidi e dei loro effetti sulla proteina stessa. Questo richiede la raccolta di **grandi quantità di dati** sulle proteine e sulla loro funzione, nonché l'uso di sofisticati metodi di analisi dei dati per comprendere le relazioni tra le sequenze di amminoacidi e le proprietà strutturali e funzionali delle proteine.

LE BANCHE DATI



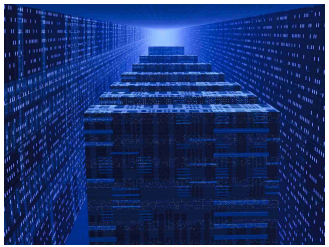
- Complessità dei dati

I dati biologici sono spesso molto **complessi** e richiedono l'uso di tecniche avanzate di analisi dei dati per essere compresi e utilizzati efficacemente.

Ad esempio, le sequenze genomiche possono essere lunghe e complesse, con molte regioni non codificanti e regioni codificanti che producono proteine diverse.

Inoltre, i dati biologici possono provenire da **diverse fonti**, come la genomica, la trascrittomica, la proteomica e la metabolomica, ognuna delle quali fornisce informazioni su diversi aspetti del sistema biologico.

LE BANCHE DATI

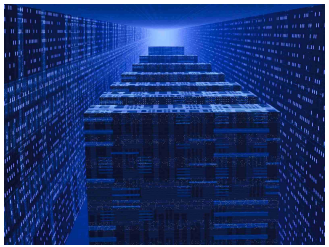


- Necessità di archiviare e gestire i dati:

La **gestione** e la **condivisione** dei dati biologici sono essenziali per evitare la duplicazione degli sforzi di ricerca e aumentare la produttività scientifica complessiva.

Inoltre, l'archiviazione dei dati biologici consente ai ricercatori di accedere ai dati stessi e di utilizzarli per rispondere alle loro domande di ricerca.

LE BANCHE DATI



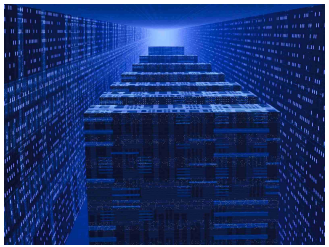
- Ruolo delle banche dati:

Quindi riassumendo possiamo affermare come …

Le banche dati biologiche svolgono un ruolo fondamentale nella **gestione** e nell'**archiviazione** dei dati biologici. Consentono di **conservare** in modo organizzato le informazioni biologiche, rendendole facilmente **accessibili** ai ricercatori di tutto il mondo.

In questo modo, le banche dati biologiche sono uno strumento fondamentale per la ricerca biologica moderna

LE BANCHE DATI



- Caratteristiche di una banca dati biologica:

Alcune delle principali caratteristiche di un database biologico includono:

1. **Accessibilità:** un database biologico deve essere facilmente accessibile a tutti gli utenti interessati, indipendentemente dal loro livello di competenza informatica.
2. **Aggiornamento regolare:** i dati contenuti nel database biologico devono essere aggiornati regolarmente per includere nuove informazioni e correzioni.
3. **Organizzazione:** un database biologico deve essere organizzato in modo da consentire una facile ricerca e recupero delle informazioni. Ciò può includere l'organizzazione per specie, tipo di molecola biologica o funzione biologica.

LE BANCHE DATI

- Caratteristiche di una banca dati biologica:

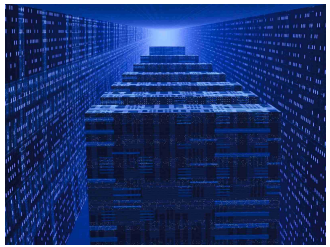
4. **Standardizzazione:** i dati contenuti nel database biologico devono essere standardizzati in modo da garantire che possano essere utilizzati in modo coerente da tutti gli utenti

5. **Compatibilità:** un database biologico deve essere compatibile con altri database biologici e software di analisi, in modo che le informazioni possano essere facilmente scambiate e utilizzate in diversi contesti.

6. **Sicurezza:** un database biologico deve essere protetto da accessi non autorizzati e attacchi informatici per garantire la sicurezza delle informazioni sensibili contenute al suo interno

7. **Documentazione:** un database biologico deve essere accompagnato da documentazione chiara e completa che spiega come accedere ai dati e come sono stati raccolti e curati.

LE BANCHE DATI



- Tipi di banche dati biologiche

Le banche dati biologiche possono essere suddivise in due categorie principali: banche dati **primarie** e banche dati **secondarie**.

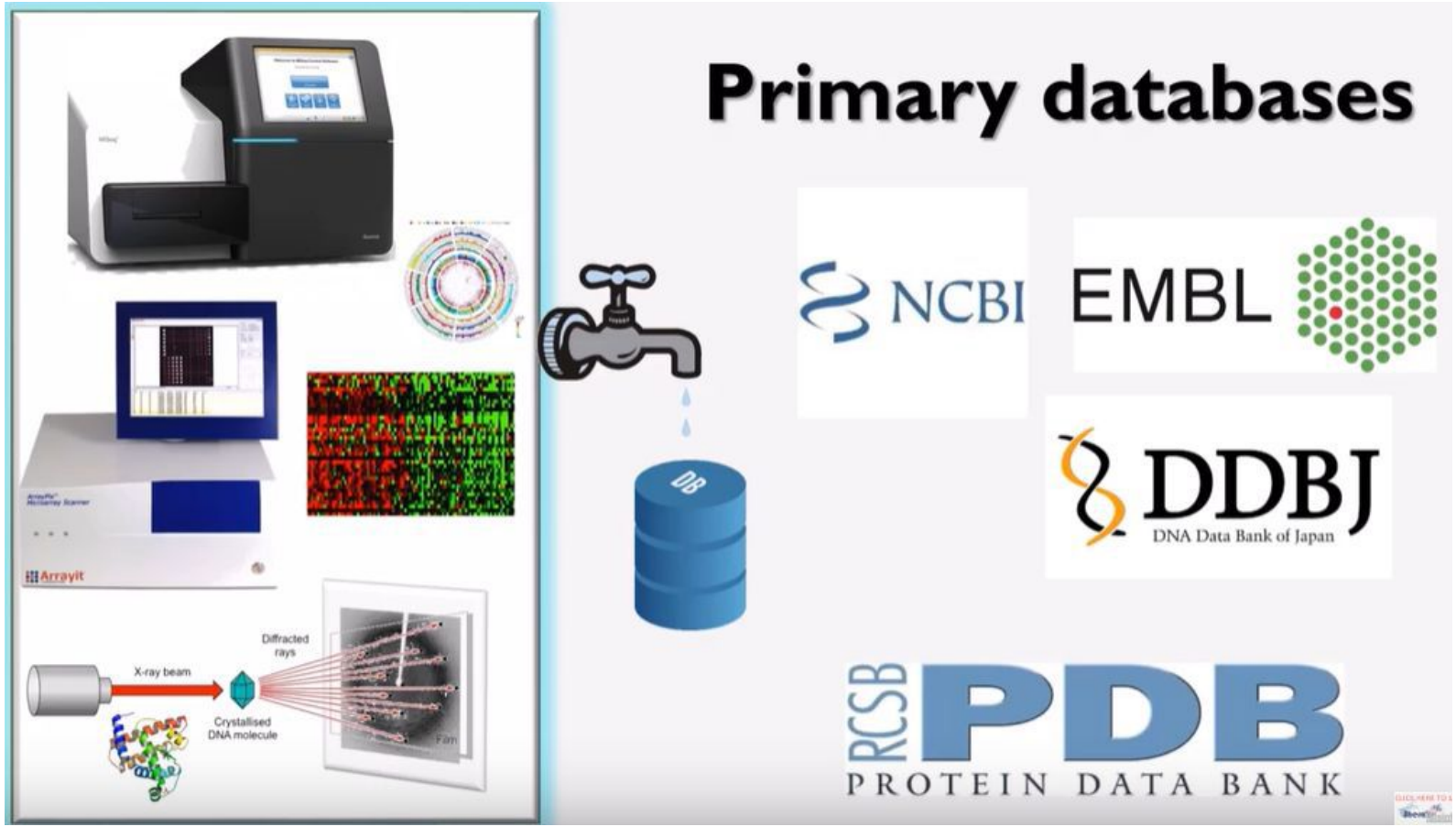
- Le banche dati primarie (dette anche archivi o collettori primari), sono quelle che contengono i dati originali raccolti dai **ricercatori**, come le sequenze di DNA, le strutture proteiche e le informazioni sulla funzione biologica delle molecole.

- In linea di principio, chiunque può sottomettere dati ad una banca dati primaria, rispettando alcune regole.

- Le banche dati primarie sono di solito gestite dalle comunità scientifiche che raccolgono i dati, come il National Center for Biotechnology Information (NCBI) negli Stati Uniti o l'European Bioinformatics Institute (EBI) in Europa.

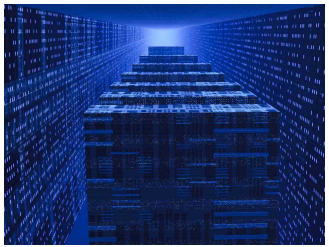
LE BANCHE DATI

- Tipi di banche dati biologiche



LE BANCHE DATI

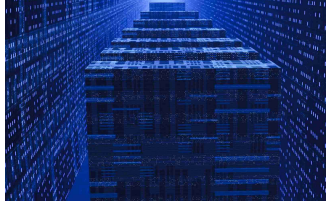
- Tipi di banche dati biologiche



I database **primari** in biologia sono raccolte di dati sperimentali che servono come risorse di informazioni fondamentali per la ricerca scientifica. Questi database contengono sequenze di DNA, RNA o proteine, strutture molecolari, dati genetici e altre misurazioni biologiche ottenute direttamente da esperimenti di laboratorio.

Il contenuto dei database primari è solitamente soggetto a un processo di revisione per assicurare l'accuratezza e l'affidabilità dei dati.

LE BANCHE DATI



Ecco alcuni esempi di database primari in biologia:

Nucleotide (GenBank): È il database nazionale di sequenze degli Stati Uniti e contiene informazioni sulle sequenze di nucleotidi e loro annotazioni. Le sequenze vengono sottomesse dai ricercatori e automaticamente integrate nel database.

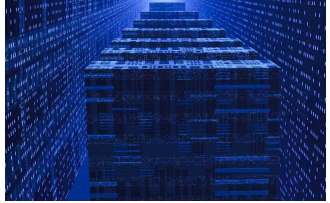
EMBL Bank: Simile a GenBank, ma gestito dall'European Molecular Biology Laboratory. Raccoglie dati sulle sequenze di DNA e RNA.

DDBJ (DNA Data Bank of Japan): Anch'esso un database di sequenze nucleotidiche, che costituisce parte del sistema internazionale di database di sequenze, insieme a GenBank e EMBL.

Protein Data Bank (PDB): Database per la struttura tridimensionale delle grandi molecole biologiche, come le proteine e gli acidi nucleici. Le strutture vengono ottenute principalmente tramite cristallografia a raggi X e risonanza magnetica nucleare (NMR).

UniProt (Universal Protein Resource): Fornisce un'ampia gamma di informazioni sulle proteine, come le

LE BANCHE DATI



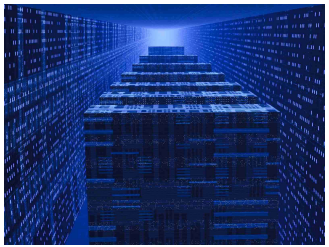
Ecco alcuni esempi di database primari in biologia:

Protein Data Bank (PDB): Database per la struttura tridimensionale delle grandi molecole biologiche, come le proteine e gli acidi nucleici. Le strutture vengono ottenute principalmente tramite cristallografia a raggi X e risonanza magnetica nucleare (NMR).

UniProt (Universal Protein Resource): Fornisce un'ampia gamma di informazioni sulle proteine, come le sequenze e le annotazioni funzionali. È una risorsa composita che combina informazioni provenienti da diverse altre fonti.

LE BANCHE DATI

- Tipi di banche dati biologiche

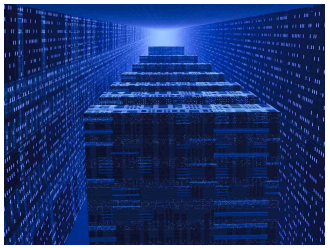


I database secondari in biologia, noti anche come database derivati, sono raccolte di informazioni che sono state elaborate o interpretate a partire dai dati originali (primari).

Questi database non contengono dati sperimentali grezzi, ma piuttosto informazioni che sono state analizzate, riassunte, predette o annotate da esperti. Sono utilizzati per facilitare la comprensione e l'analisi delle informazioni biologiche e per assistere nella ricerca e nello sviluppo.

LE BANCHE DATI

- Tipi di banche dati biologiche



- Le banche dati **secondarie**, quindi, sono derivate dalle primarie e possono raccogliere un sottoinsieme delle informazioni in esse contenute.

- Sono di solito create da gruppi di ricerca specifici, che utilizzano i dati primari per generare nuove informazioni e migliorare la comprensione della biologia.

- La differenza principale è che i dati in queste banche dati sono "**curati**" manualmente da persone esperte, che applicano filtri e procedure di controllo per assicurarsi che ogni entry della banca dati non sia ridondante, non contenga errori e inesattezze e sia il più possibile affidabile.

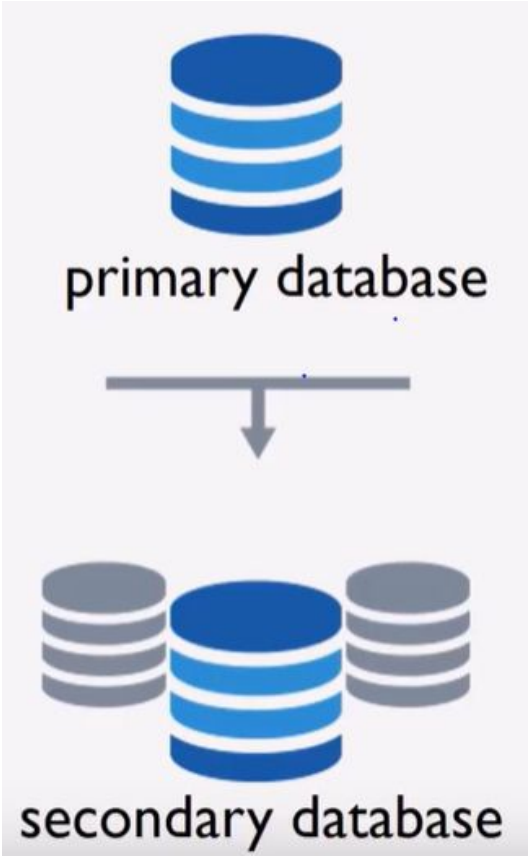
LE BANCHE DATI

- Tipi di banche dati biologiche

  - Spesso infatti è preferibile avere un "set" di dati più piccolo ma più affidabile rispetto ad uno più grande, ridondante e con troppi errori.

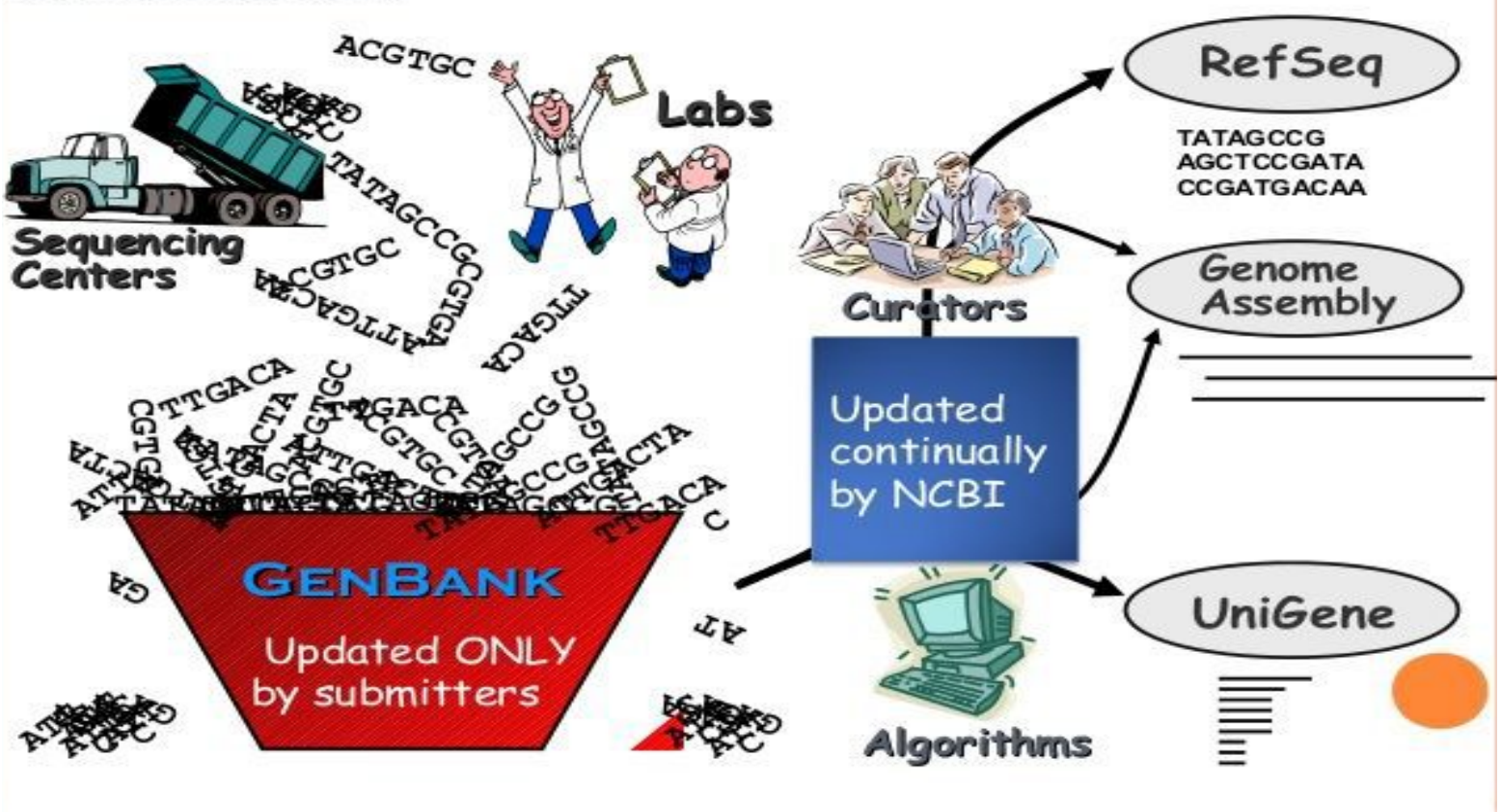
  - I database secondari sono database ottenuti tramite l'elaborazione dei dati contenuti nei database primari.

  - Una banca dati primaria può alimentare diverse banche dati secondarie



LE BANCHE DATI

Primary vs.secondary sequence database



LE BANCHE DATI

Alcuni esempi di database secondari includono:

Pfam: Un database di famiglie di proteine e domini proteici. Ogni famiglia è rappresentata da uno o più modelli di allineamento delle sequenze, che sono usati per identificare le proteine appartenenti a quella famiglia.

InterPro: Fornisce informazioni integrate riguardo alle famiglie, domini e siti funzionali delle proteine attraverso la combinazione di diverse risorse di database proteici come Pfam, PRINTS, ProDom, e SMART. Aiuta a classificare le sequenze proteiche in famiglie e a prevedere la presenza di domini e siti importanti.

Swiss-Prot: Parte di UniProt, è un database di sequenze proteiche annotato manualmente che fornisce una descrizione di alta qualità delle sequenze delle proteine, oltre a informazioni sulla funzione della proteina, sull'interazione, sulla localizzazione cellulare e molto altro.

LE BANCHE DATI

Alcuni esempi di database secondari includono:

CATH e **SCOP**: Sono database di strutture proteiche che classificano le proteine in base alla loro struttura e alla loro evoluzione. Forniscono un modo per studiare le relazioni tra le diverse strutture proteiche e le loro funzioni.

KEGG (Kyoto Encyclopedia of Genes and Genomes):

È un database che integra dati genomici, chimici e sistemici. Fornisce informazioni sulle reti di interazione tra biomolecole e sui percorsi metabolici, assistendo nella comprensione delle funzioni biologiche e delle interazioni all'interno delle celle.

LE BANCHE DATI

- Tipi di banche dati biologiche

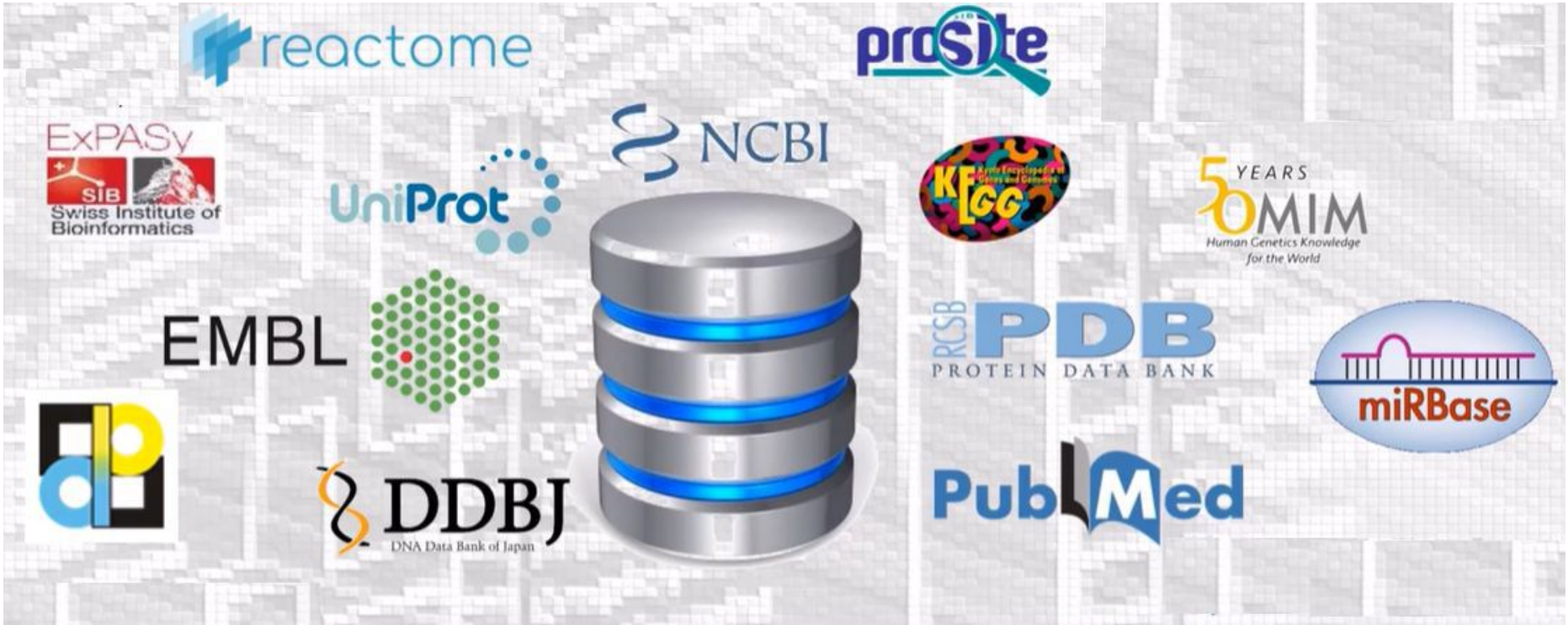
Le banche dati biologiche possono essere ulteriormente suddivise in base al **tipo di dati** che contengono.

Ad esempio, le banche dati di **sequenza** contengono le sequenze di DNA o RNA, mentre le banche dati di **struttura** contengono le informazioni sulla struttura tridimensionale delle proteine.

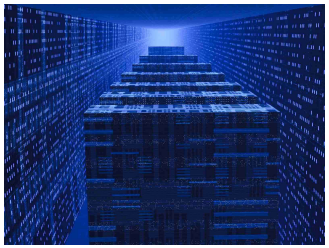
Le banche dati di **funzione** forniscono informazioni sulla funzione biologica delle molecole, come le attività enzimatiche o le vie metaboliche coinvolte nella sintesi di un metabolita.

Le banche dati di **espressione** contengono informazioni sulla quantità di mRNA o proteine prodotte da un gene in diverse condizioni o tessuti.

LE BANCHE DATI



LE BANCHE DATI

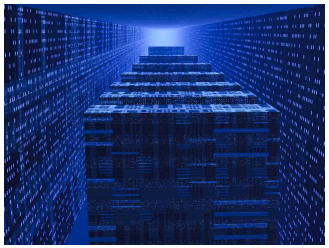


- Banche dati di sequenza

Una banca dati di sequenza è un **archivio** digitale che raccoglie e organizza le **sequenze** di DNA, RNA e proteine provenienti da varie fonti e organismi. Queste banche dati consentono di accedere alle sequenze, di analizzarle e di utilizzarle per scopi di ricerca e di sviluppo di applicazioni. Tra le più importanti a livello mondiale:

- NCBI GenBank
- European Nucleotide Archive (ENA)
- DNA Data Bank of Japan (DDBJ)

LE BANCHE DATI



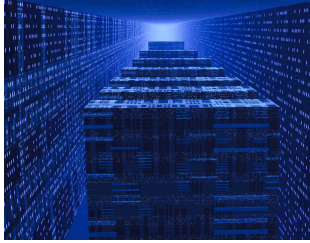
- Banche dati di sequenza

NCBI GenBank

è un database di sequenze di DNA, RNA e proteine gestito dal National Center for Biotechnology Information (NCBI) degli Stati Uniti. GenBank contiene oltre 300 miliardi di basi di dati genetiche, tra cui sequenze di organismi eucarioti, procarioti e virali. GenBank è stato fondato nel 1982 ed è uno dei più antichi e grandi database di sequenze a livello mondiale.

L'interfaccia di ricerca di Genbank (attualmente chiamata Nucleotide)

LE BANCHE DATI



- Banche dati di sequenza

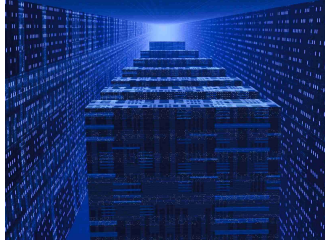
European Nucleotide Archive (ENA)

è un database di sequenze di DNA, RNA e proteine gestito dal European Bioinformatics Institute (EMBL-EBI). ENA è uno dei più grandi database di sequenze a livello mondiale. È stato fondato nel 2002 e raccoglie, conserva e rende liberamente accessibili i dati di sequenza provenienti da tutte le forme di vita.

L'ENA è progettato per essere una risorsa globale che supporta non solo i ricercatori nel campo della genomica, ma anche quelli in settori come la metagenomica, la filogenetica, la biotecnologia e l'epidemiologia. Il database include sequenze di DNA e RNA di vari tipi, tra cui sequenze complete di genomi, sequenze di cromosomi, sequenze di esoni e introni, e molto altro.

ENA
European Nucleotide Archive

LE BANCHE DATI



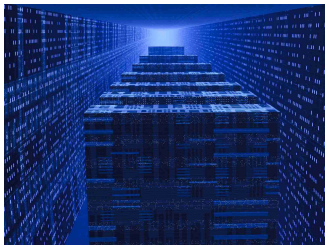
- Banche dati di sequenza

DNA Data Bank of Japan (DDBJ)

DDBJ
DNA Data Bank of Japan

è un database di sequenze di DNA, RNA e proteine gestito dall'Università di Chiba, in Giappone. DDBJ è uno dei tre membri fondatori del International Nucleotide Sequence Database Collaboration (INSDC) insieme a GenBank e ENA. Il database contiene sequenze di organismi eucarioti, procarioti e virali e collabora con GenBank ed ENA per mantenere un unico database di sequenze a livello mondiale.

LE BANCHE DATI



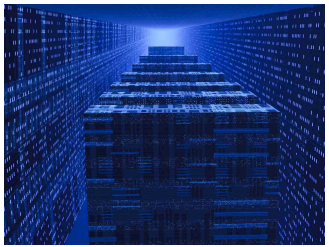
- Banche dati di strutture

I database di struttura sono banche dati che contengono informazioni sulla struttura tridimensionale delle macromolecole biologiche, come le proteine e l'RNA. Queste banche dati permettono di accedere alle informazioni sulla struttura delle molecole e di utilizzarle per scopi di ricerca e di sviluppo di applicazioni.

Tra i più noti abbiamo:

- Il Protein Data Bank (PDB)

LE BANCHE DATI



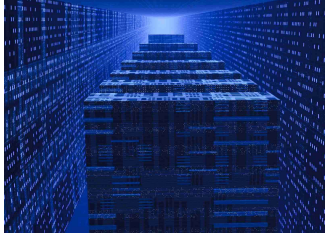
- Banche dati di strutture

Il Protein Data Bank (PDB)

è uno dei database di struttura più noti e utilizzati. Contiene informazioni sulla struttura tridimensionale delle proteine, ottenute attraverso tecniche sperimentali come la cristallografia ai raggi X e la risonanza magnetica nucleare (NMR). Le informazioni sulle strutture tridimensionali delle proteine presenti nel PDB includono le coordinate atomiche e le informazioni sulla sequenza di amminoacidi. Queste informazioni possono essere utilizzate per comprendere le proprietà e le funzioni delle proteine, per identificare potenziali bersagli terapeutici e per sviluppare nuovi farmaci.

RCSB PDB
PROTEIN DATA BANK

LE BANCHE DATI



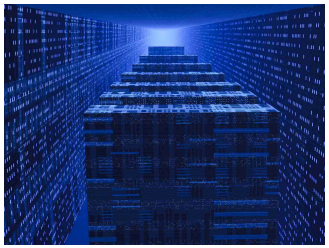
- Banche dati di strutture

Anche per l'RNA esistono banche dati di struttura, che permettono di accedere alle informazioni sulla struttura tridimensionale dell'RNA. Questi database includono anche informazioni sulle regioni non codificanti dell'RNA, che sono importanti per la regolazione dell'espressione genica.

Tra le più usate:

- RNAcentral https://rnacentral.org/
- PseudoBase++ https://rnavlab.utep.edu/database
- NONCODE v5.noncode.org

LE BANCHE DATI

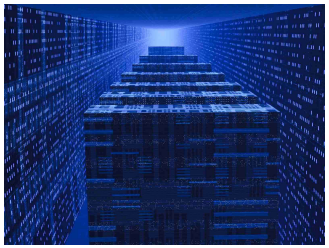


- Banche dati di strutture

RNAcentral

è una banca dati di riferimento per l'RNA che integra dati da diverse banche dati di RNA, tra cui Rfam, miRBase e l'archivio di strutture dell'RNA (RNA Strucure Atlas). RNAcentral fornisce accesso a sequenze, annotazioni, strutture e informazioni funzionali sull'RNA.

LE BANCHE DATI



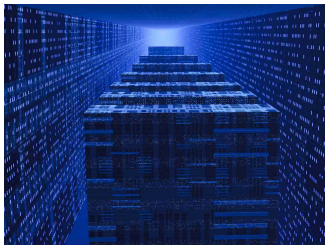
- Banche dati di strutture

PseudoBase++

è una banca dati di riferimento per gli RNA non codificanti (ncRNA) e contiene informazioni sulle loro strutture tridimensionali, interazioni con proteine e funzioni biologiche. PseudoBase++ fornisce anche strumenti per la previsione delle strutture degli ncRNA e l'analisi delle loro funzioni.

RNAVLab
RNA Virtual Laboratory

LE BANCHE DATI



- Banche dati di strutture

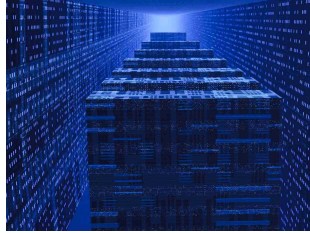
NONCODE

è una banca dati di riferimento per gli RNA non codificanti (ncRNA) e fornisce informazioni sulla loro espressione, funzione e regolazione. La banca dati contiene anche informazioni sulle loro strutture tridimensionali, interazioni con proteine e annotazioni genomiche

NONCODE

An integrated knowledge database dedicated to ncRNAs, especially lncRNAs.

LE BANCHE DATI



- Banche dati di funzione

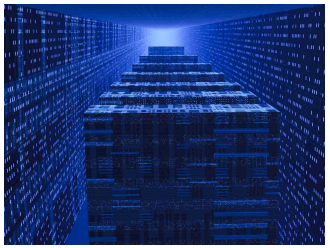
Una banca dati di funzione è un database che archivia informazioni sulle funzioni biologiche di una particolare sequenza di DNA, RNA o proteina.

Queste banche dati sono utilizzate per **annotare** le sequenze e **identificare** le loro **funzioni** biologiche, come ad esempio la catalisi di una reazione chimica, l'interazione con altre proteine o la regolazione dell'espressione genica.
Le informazioni contenute nelle banche dati di funzione sono spesso ottenute attraverso esperimenti di laboratorio o analisi bioinformatiche.

Tra le più importanti:

- UniProt
- KEGG

LE BANCHE DATI



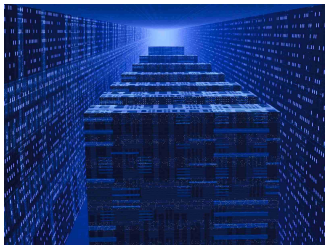
- Banche dati di funzione

UniProt

UniProt è un database di sequenze proteiche annotate, che fornisce informazioni dettagliate sulla funzione, la struttura e le interazioni di proteine provenienti da una vasta gamma di organismi. Le informazioni contenute in UniProt sono ottenute da una varietà di fonti, tra cui la letteratura scientifica, i depositi di sequenze, i laboratori di ricerca e la cura manuale.

UniProt integra informazioni provenienti da altre banche dati, come Pfam, InterPro, GO e OMIM, per fornire una visione completa della funzione delle proteine.

LE BANCHE DATI

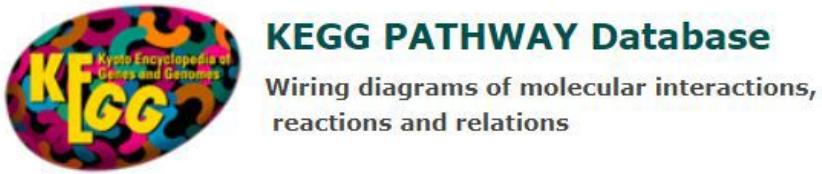


- Banche dati di funzione

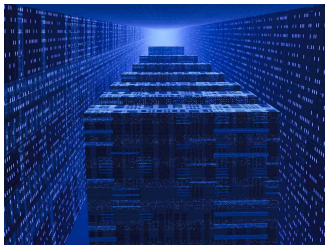
KEGG

(Kyoto Encyclopedia of Genes and Genomes) è un database di pathway metabolici e genetici, che fornisce informazioni sulle funzioni biologiche dei geni e delle proteine, nonché sui loro ruoli nei pathway metabolici e delle malattie.

KEGG integra informazioni provenienti da diverse fonti, tra cui sequenze genomiche, pathway, funzioni biologiche, espressione genica, malattie e droghe, per fornire una visione integrata della biologia dei sistemi.



LE BANCHE DATI



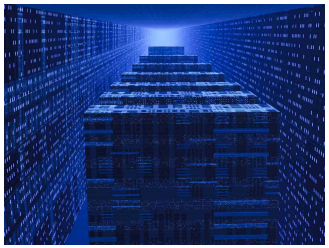
- Banche dati di espressione

Una banca dati biologica di espressione è un tipo di banca dati che contiene informazioni sul livello di espressione genica in diversi tessuti, cellule o condizioni sperimentali. Queste banche dati di solito contengono informazioni sulle **quantità** relative di mRNA prodotto da ogni gene in un determinato campione biologico, che possono essere misurate utilizzando tecnologie come la sequenza di RNA (RNA-Seq).

Inoltre, le banche dati di espressione possono contenere informazioni sulla regolazione dell'espressione genica, sui percorsi di segnalazione cellulare e su altri processi biologici che influenzano l'espressione genica.

Le banche dati di espressione sono un'importante risorsa per i ricercatori che cercano di comprendere la regolazione dell'espressione genica in diversi tessuti e condizioni sperimentali

LE BANCHE DATI



- Banche dati di espressione

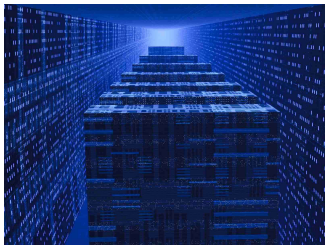
Tra le più importanti ricordiamo:

- Gene Expression Omnibus (GEO)
- ArrayExpress

Gene Expression Omnibus (GEO)
è un database curato dal National Center for Biotechnology Information (NCBI) che archivia dati di espressione genica provenienti da diversi tipi di esperimenti, come sequenziamento dell'RNA. GEO è uno dei principali repository di dati di espressione genica e permette agli utenti di accedere e scaricare facilmente i dati.

ArrayExpress
è un altro database di espressione genica curato dal European Bioinformatics Institute (EBI) che archivia dati di espressione genica provenienti da diversi organismi e tipi di esperimenti. ArrayExpress offre strumenti di analisi dei dati e visualizzazioni per aiutare gli utenti a comprendere e interpretare i dati di espressione genica.

LE BANCHE DATI



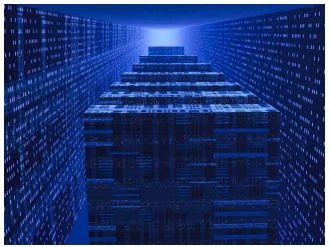
- Utilizzo pratico delle banche dati biologiche

Ad esempio …

1. Un ricercatore che sta studiando una proteina sconosciuta può utilizzare una banca dati di **sequenza** come GenBank o UniProt per cercare proteine simili o omologhe a quella di interesse. Ciò può fornire informazioni sulla funzione della proteina e sulla sua possibile interazione con altre proteine o composti.

2. I ricercatori possono utilizzare banche dati di espressione come il Gene Expression Omnibus (GEO) per confrontare i livelli di espressione genica in diversi tessuti o in diverse condizioni fisiologiche. Ciò può aiutare a identificare geni coinvolti in processi biologici specifici e a comprendere come la regolazione dell'espressione genica contribuisca alla fisiologia dell'organism

LE BANCHE DATI



QUINDI …

- Quali informazioni si possono ottenere da una banca dati di strutture come il Protein Data Bank e come si possono utilizzare tali informazioni per prevedere la funzione di una proteina?

LE BANCHE DATI

QUINDI …

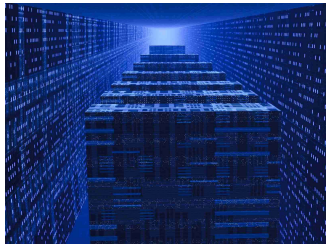
Utilizzando le informazioni presenti nel PDB, è possibile **prevedere** la **funzione** proteina in diversi modi.

Ad esempio, se si conosce la struttura tridimensionale di una proteina, è possibile confrontarla con quelle di altre proteine note per avere una funzione simile, individuando eventuali analogie strutturali.

Inoltre, l'analisi delle regioni attive o dei siti di legame di una proteina può fornire informazioni sulla sua funzione biochimica.

Infine, la combinazione di informazioni strutturali e funzionali può essere utilizzata per progettare **farmaci** o composti terapeutici mirati.

LE BANCHE DATI



- Utilizzo pratico delle banche dati biologiche

Un ricercatore vuole studiare una proteina coinvolta nella regolazione della crescita cellulare.

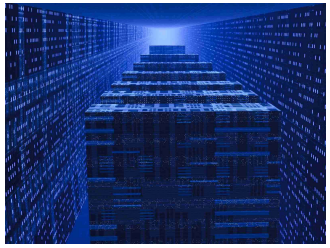
Per fare ciò, accede al database di sequenza UniProt per identificare **proteine omologhe** in diverse specie.

Successivamente, utilizza il database di struttura PDB per trovare informazioni sulla **struttura tridimensionale** della proteina in questione, che aiuteranno a comprendere come interagisce con altre molecole all'interno della cellula.

Infine, il ricercatore accede a banche dati di **espressione** come il Gene Expression Omnibus (GEO) per capire in quali tessuti e sotto quali **condizioni** la proteina viene espressa, al fine di comprendere il suo ruolo nella crescita cellulare.

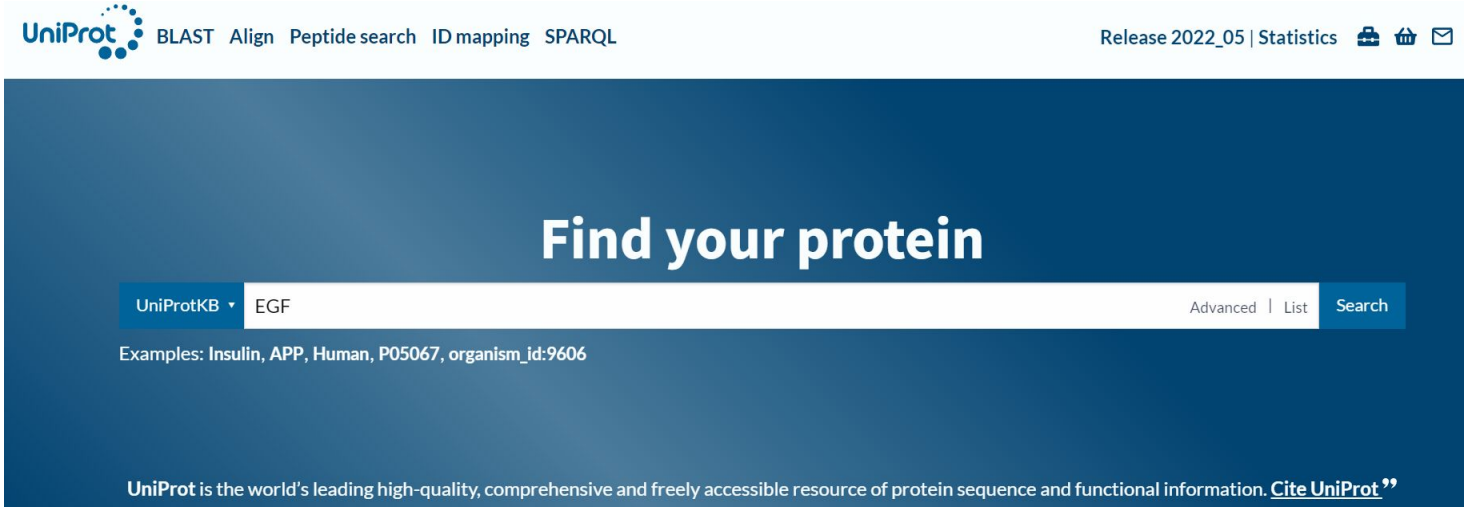
Grazie all'utilizzo di queste banche dati, il ricercatore riesce ad approfondire la sua conoscenza sulla proteina in questione e sviluppare nuove ipotesi di ricerca.

LE BANCHE DATI

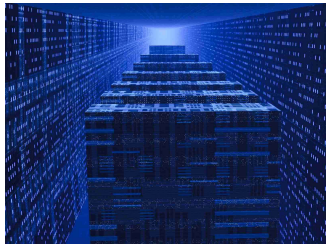


- Utilizzo pratico delle banche dati biologiche

Un esempio di proteina coinvolta nella regolazione della crescita cellulare è il fattore di crescita epidermico (EGF – Epidermal Growth Factor), una proteina segnale che lega il suo recettore sulla superficie delle cellule e attiva la cascata di segnalazione cellulare coinvolta nella crescita, proliferazione e differenziazione cellulare. https://www.uniprot.org/



LE BANCHE DATI



- Utilizzo pratico delle banche dati biologiche

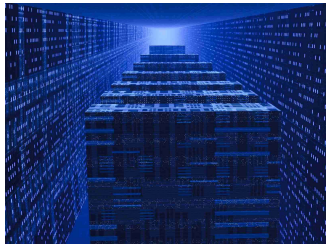
UniProtKB 443,094 results or search "EGF" as a Protein Name, Protein family, Catalytic Activity, Gene Name, Author[...]

BLAST Align Map IDs Download Add View: Cards ○ Table ◉ Customize columns Share

Entry	Entry Name	Protein Names	Gene Names	Organism	Length
P01133	EGF_HUMAN	Pro-epidermal growth factor[...]	EGF	Homo sapiens (Human)	1,207 AA
P01132	EGF_MOUSE	Pro-epidermal growth factor[...]	Egf	Mus musculus (Mouse)	1,217 AA
P07522	EGF_RAT	Pro-epidermal growth factor[...]	Egf	Rattus norvegicus (Rat)	1,133 AA
Q9BEA0	EGF_CANLF	Pro-epidermal growth factor[...]	EGF	Canis lupus familiaris (Dog) (Canis familiaris)	1,216 AA
Q00968	EGF_PIG	Pro-epidermal growth factor[...]	EGF	Sus scrofa (Pig)	1,214 AA
Q95ND4	EGF_FELCA	Pro-epidermal growth	EGF	Felis catus (Cat) (Felis	1,210 AA

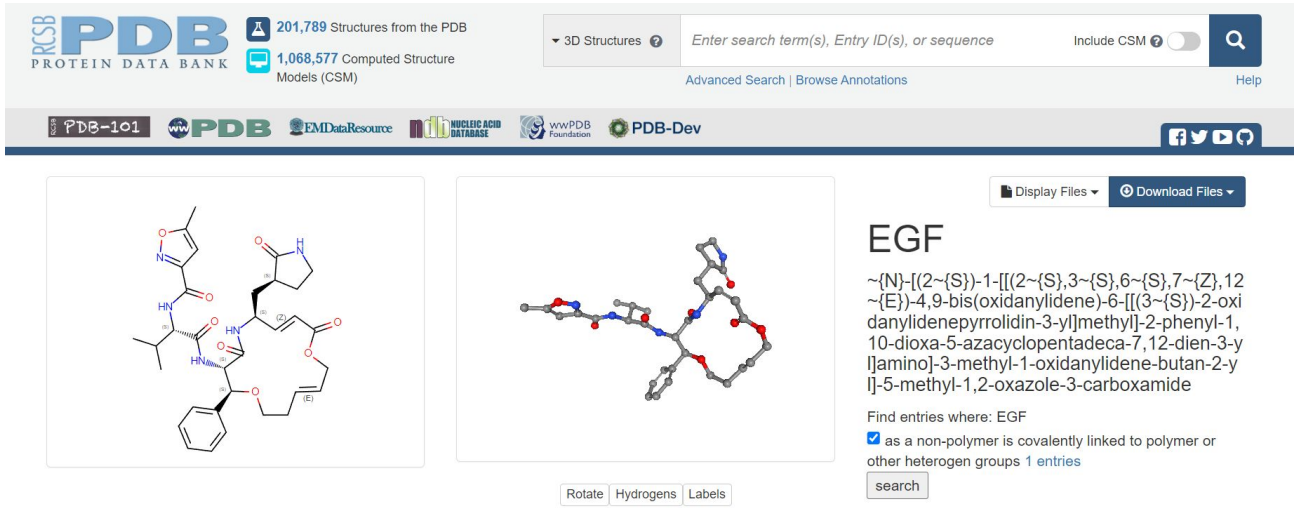
Mi permette di osservare se sono presenti proteine omologhe in altri organismi

LE BANCHE DATI

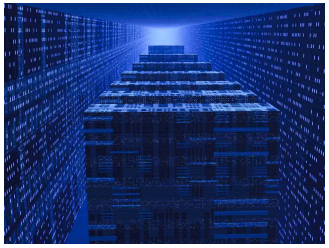


- Utilizzo pratico delle banche dati biologiche

Utilizza il database di struttura PDB per trovare informazioni sulla **struttura tridimensionale** della proteina in questione, che aiuteranno a comprendere come interagisce con altre molecole all'interno della cellula.
https://www.rcsb.org/ligand/EGF

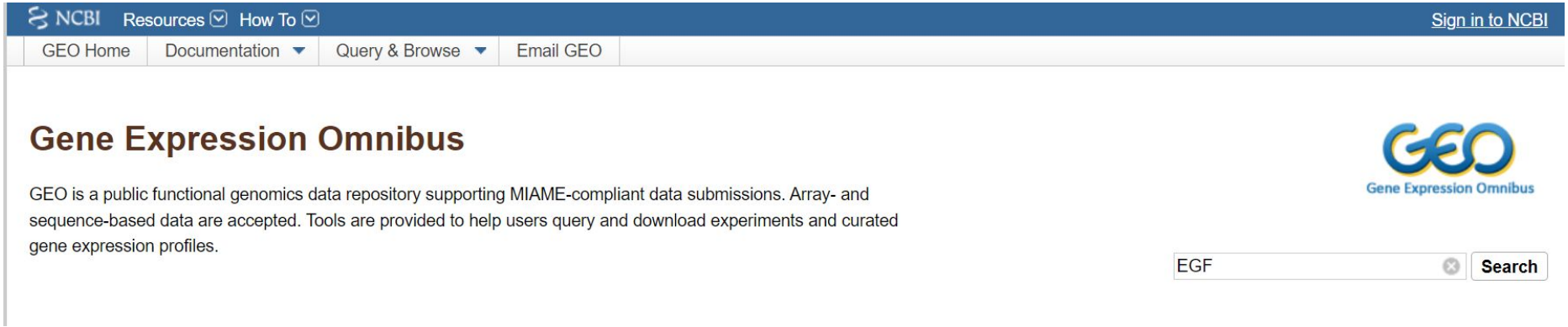


LE BANCHE DATI

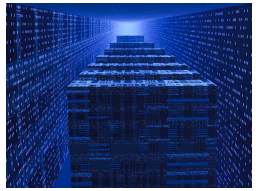


- Utilizzo pratico delle banche dati biologiche

Il ricercatore accede a banche dati di **espressione** come il Gene Expression Omnibus (GEO) per capire in quali tessuti e sotto quali **condizioni** la proteina viene espressa, al fine di comprendere il suo ruolo nella crescita cellulare.
https://www.ncbi.nlm.nih.gov/geo/



LE BANCHE DATI



- Utilizzo pratico delle banche dati biologiche

National Library of Medicine
National Center for Biotechnology Information

Log in

GEO DataSets | GEO DataSets | egf | Search | Create alert | Advanced | Help

Entry type
DataSets (45)
Series (2,837)
Samples (55,576)
Platforms (6)

Organism
Customize ...

Study type
Expression profiling by array
Methylation profiling by array
Customize ...

Author
Customize ...

Attribute name
tissue (10,962)
strain (8,313)
Customize ...

Publication dates
30 days

Summary | 20 per page | Sort by Default order | Send to:

Search results

Items: 1 to 20 of 58464

<< First | < Prev | Page 1 of 2924 | Next > | Last >>

1. Granulosa cell mevalonate pathway abnormalities contribute to oocyte meiotic defects and aneuploidy [HiC GGOH-treatment]

(Submitter supplied) The mechanisms of aging-related oocyte aneuploidy remain elusive. Hi-C and SMART-seq revealed aging-related decreases in chromosome condensation, particularly for genomic regions proximal to the centromeres, accompanied with disrupted meiosis-associated gene expression in metaphase I (MI) aged oocytes. Further transcriptomic analysis showed that oocyte meiotic maturation was correlated with robust increases in mevalonate (MVA) pathway gene expression in young oocyte-surrounding granulosa cells (GCs), which was largely downregulated in aged GCs. more...

Organism: Mus musculus
Type: Other
Platform: GPL24247 4 Samples
Download data: COOL
Series Accession: GSE212860 ID: 200212860
Analyze with GEO2R

Filters: Manage Filters

Top Organisms [Tree]
Homo sapiens *(43816)*
Mus musculus *(12710)*
Rattus norvegicus *(776)*
Macaca fascicularis *(674)*
Canis lupus *(149)*
More...

Find related data
Database: Select
Find items

Search details
egf[All Fields]