Allegato 2-ELIXIR-IT

Portfolio servizi ELIXIR-IT

PIATTAFORMA OMICS

La piattaforma per l'accesso ai servizi della piattaforma OMICS di ELIXIR-IT è stata implementata grazie ai progetti di potenziamento infrastrutturale CNRBiomics (PON PIR00017, CIR00017) ed ELIXIRxNextGenIT (PNRR IR0000010), finalizzati allo sviluppo delle regioni del Sud Italia. Coinvolge sia il **Consiglio Nazionale delle Ricerche (CNR)** che l'**Università degli Studi di Bari (UNIBA)** fornendo accesso alle tecnologie più avanzate per il sequenziamento massivo/parallelo del DNA (Next Generation Sequencing, NGS), anche a livello di singola cellula, oltre a servizi di trascrittomica spaziale, proteomica, metabolomica e supporto bioinformatico specializzato.

La piattaforma OMICS è supportata dalla piattaforma COMPUTE (vedi di seguito), dotata di capacità elevate di calcolo e archiviazione, per fornire agli utenti finali sia la produzione che l'analisi e la gestione dei dati.

La piattaforma OMICS è equipaggiata con strumentazione all'avanguardia, tra cui:
- Piattaforme di sequenziamento Illumina: Novaseq X Plus, Novaseq 6000, MiSeq, NextSeq 2000
- Illumina iScan
- PacBio Sequel IIe, PacBio Vega
- Oxford Nanopore MinION, GridION, PromethION 24
- Sistemi automatizzati per la dispensazione dei liquidi (Hamilton, Masmec)
- Agilent Fragment Analyzer, Agilent Femto pulse
- Bionano Genomics
- 10X Visium Cytassist, 10X Xenium, Vizgen Merscope
- ViewSizer Horiba
- Sistema UHPLC Vanquish Flex (Thermo Fisher Scientific) e spettrometro di massa Orbitrap Fusion Tribrid (Thermo Fisher Scientific)
- SCIEX TripleTOF 6600+
- Orbitrap Exploris 240 e 120 – Thermo Scientific
- LTQ-Orbitrap – Thermo Scientific
- Risonanza Magnetica 3T (Discovery MR-750, General Electric).

Il portfolio dei servizi Omics include, ma non si limita a:

Sequenziamento dell'intero genoma/esoma
Il sequenziamento completo del genoma individua mutazioni comuni e rare, rilevando diversi tipi di mutazioni, inclusi polimorfismi a singolo nucleotide (SNP), inserzioni/delezioni (InDel), variazioni strutturali (SV) e variazioni del numero di copie (CNV).
Il sequenziamento dell' intero genoma (Whole Genome resequencing) è ampiamente utilizzato in studi clinici, come per la definizione della patogenesi delle malattie genetiche, la profilazione molecolare del cancro, la valutazione del rischio associato all'insorgenza di patologie acute e croniche, ecc.

Analisi epigenetiche per valutare la struttura/accessibilità della cromatina e rilevare modifiche del DNA/RNA e RNA Editing
- ATAC-seq (Assay for Transposase Accessible Chromatin using sequencing)
- Organizzazione spaziale della cromatina: identificazione delle regioni di cromatina in prossimità spaziale tramite procedure di ligazione delle regioni prossimali (Hi-C).

- Analisi delle interazioni DNA-proteina (ChIP-seq): combina l'immunoprecipitazione della cromatina (ChIP) con il sequenziamento ad elevata profondità. È un metodo potente per identificare a livello genomico i siti di legame del DNA con fattori di trascrizione e altre proteine.
- Analisi del metiloma: mediante array (MethylomeEPIC 2.0) o sequenziamento del DNA nativo con tecnologia long-read Nanopore.

Metagenomica
- Shotgun Metagenomic: sequenziamento casuale del DNA totale presente in un campione, rappresentativo di tutti i genomi delle specie contenute (metagenoma).
- DNA metabarcoding: permette la caratterizzazione tassonomica da un campione misto attraverso il sequenziamento ad elevata profondità di specifici marcatori del DNA.

Sequenziamento del trascrittoma
- **mRNA-Seq e total RNAseq**: offrono una panoramica completa del profilo trascrizionale cellulare senza filtri legati a conoscenze pregresse. Rilevano isoforme di trascritti noti e nuovi, espressione differenziale, espressione allele-specifica, fusioni geniche, isoforme, SNP in regioni codificanti (cSNP) e giunzioni di splicing.
- **Small RNA sequencing**: sequenziamento massivo della frazione di RNA a basso peso molecolare (ad esempio miRNA e altri snRNA).

Single Cell Omics
Il sequenziamento del genoma e del trascrittoma a livello di singola cellula è cruciale per rilevare l'eterogeneità delle popolazioni cellulari, identificare sottopopolazioni minoritarie di interesse e scoprire caratteristiche uniche delle singole cellule.

Trascrittomica spaziale
Le piattaforme di trascrittomica spaziale (10X Visium HD, 10X Xenium, Vizgen Merscope) permettono di mappare la localizzazione dell'espressione di geni e proteine all'interno di sezioni di tessuto, catturando la distribuzione subcellulare dei trascritti di RNA a risoluzione di singola molecola e definendo l'architettura cellulare in tessuti sani e patologici di diverse specie.

Caratterizzazione delle vescicole extracellulari
Analisi qualitativa e quantitativa delle vescicole extracellulari presenti nei fluidi biologici.

Mappatura ottica genome-wide
L'imaging ottico diretto delle molecole di DNA linearizzato tramite NanoCanali paralleli colma le lacune dei dati ottenuti con il sequenziamento, identificando grandi variazioni strutturali (da 500 bp fino a lunghezze di megabasi), come delezioni, duplicazioni, traslocazioni e inversioni, con una sensibilità fino al 99%.

Validazione dei dati genomici o trascrittomici ottenuti con piattaforme NGS
Sequenziamento Sanger a bassa produttività o sistema QuantStudio™ 12K Flex per qRT-PCR.

Metabolomics targeted and untargeted profiling: estrazione dei metaboliti e separazione tramite cromatografia UHPLC (RP e HILIC), analisi ad alta risoluzione con frammentazione MS2 e MS3, analisi dati.

Analisi di Proteomica: analisi avanzate mediante nano/microLC accoppiata a spettrometria di massa tandem, con applicazioni che spaziano dalla caratterizzazione proteica allo studio delle vie metaboliche e delle interazioni molecolari.

CONNETTOMICA CEREBRALE UMANA
Il servizio di analisi del connettoma cerebrale umano rappresenta un *unicum* nell'offerta dell'infrastruttura BEST. Nell'ottica di ampliare e potenziare i servizi della piattaforma, grazie alle potenzialità della Risonanza

Magnetica 3T, attraverso l'impiego delle più moderne tecnologie di neuroimaging e algoritmi avanzati di analisi dei network cerebrali, il servizio restituisce marcatori quantitativi di integrità cerebrale, offrendo una visione oggettiva dello stato funzionale e strutturale del cervello umano. Questi dati rappresentano uno strumento prezioso non solo per avanzare la ricerca scientifica, ma anche per contribuire alla valutazione dello stato di malattia in ambito neurologico e psichiatrico, aprendo nuove prospettive nella diagnosi e nel monitoraggio clinico.

PIATTAFORMA TRAINING

La piattaforma TRAINING di ELIXIR-IT (TP) sviluppa programmi formativi internazionali per rafforzare le competenze in bioinformatica, calcolo e gestione dei dati, sia in Italia sia a livello globale. Composta da discenti e docenti, promuove un'istruzione di qualità come base per l'eccellenza nella ricerca nelle scienze della vita. La TP collabora con altri nodi ELIXIR europei in progetti formativi e realizza iniziative nazionali e internazionali, inclusi percorsi e-learning e attività di valutazione della qualità e dell'impatto formativo.
La piattaforma organizza workshop, corsi ed eventi nei seguenti ambiti tematici:

Competenze computazionali
Le competenze computazionali sono essenziali per organizzare, archiviare, gestire e analizzare i dati e sono fondamentali per una ricerca efficiente, aperta e riproducibile. I nostri corsi sono pensati per i ricercatori delle Scienze della Vita, spesso esperti nella loro disciplina ma alle prime armi con la programmazione e l'uso di strumenti computazionali specialistici.
Argomenti trattati:
- Shell, Python, R
- Git, GitHub, Docker, Galaxy
- Machine Learning

Competenze nella gestione dei dati
Acquisire solide competenze nella gestione, FAIRification (aderenza ai principi F.A.I.R.) e cura dei dati è cruciale nel campo delle Scienze della Vita. Le competenze sui dati sono oggi fondamentali per condurre una ricerca di alta qualità e riproducibile. I corsi coprono l'intero ciclo di vita della ricerca guidata dai dati.
Argomenti trattati:
- Gestione dei dati e Data Stewardship
- Interoperabilità dei dati e FAIRification
- Ontologie e Bioschemas

Analisi di dati omici
L'analisi dei dati omici e gli studi multi-omici che combinano genomica, trascrittomica, proteomica e altri "omics" sono sempre più diffusi nella ricerca biologica e biomedica. I corsi ELIXIR-IT coprono molteplici approcci all'analisi di tali dati.
Argomenti trattati:
- RNA-seq / scRNA-seq / WGS / WES / ChIP-seq
- Genomica delle popolazioni e pangenomica
- Metagenomica
- Metabolomica
- Proteomica
- Assemblaggio genomico
- Integrazione di dati multi-omici

Risorse e strumenti bioinformatici
Corsi su come utilizzare in modo efficace i database e gli strumenti bioinformatici per svolgere una ricerca riproducibile.

Argomenti trattati:
- Struttura delle proteine
- Interazioni proteiche
- Reti biologiche
- Malattie rare

Competenze pedagogiche
Il corso *Train The Trainer* è pensato per aiutare formatori e insegnanti a motivare i propri studenti, insegnando diverse modalità di erogazione didattica per coinvolgere efficacemente i partecipanti. Si affrontano anche lo sviluppo di corsi, percorsi formativi e materiali didattici, nonché la loro FAIRification.
Argomenti trattati:
- Train the Trainer
- Sviluppo e FAIRificazione dei materiali didattici
- Progettazione di corsi e percorsi formativi

PIATTAFORMA INTEROPERABILITY

ELIXIR-IT promuove e sostiene l'adozione delle migliori pratiche di interoperabilità da parte di enti pubblici e privati attivi nelle scienze della vita, offrendo supporto e attività formative per facilitare l'accesso ai servizi di interoperabilità di ELIXIR.

La Piattaforma di Interoperabilità di ELIXIR ha l'obiettivo di facilitare la scoperta, l'accesso, l'integrazione e l'analisi dei dati biologici da parte di persone e macchine. Essa incoraggia la comunità accademica e industriale operanti nel settore delle scienze della vita a utilizzare formati di file standardizzati, metadati strutturati, vocabolari controllati e identificatori persistenti.

La crescente produzione di dati in formati eterogenei rappresenta una sfida, rendendo complessa la ricerca e il confronto tra dataset provenienti da fonti diverse e ostacolando l'avanzamento scientifico.

L'obiettivo finale è rendere dati e strumenti bioinformatici:

- Findable (rintracciabili)
- Accessible (accessibili)
- Interoperable (interoperabili)
- Reusable (riutilizzabili)

in linea con i principi FAIR.

PIATTAFORMA DATA

L'obiettivo della Piattaforma Data di ELIXIR è quello di potenziare l'uso, il riutilizzo e il valore dei dati nel campo delle scienze della vita. Per raggiungere questo scopo, la piattaforma mette a disposizione degli utenti risorse dati di alta qualità e sostenibili, all'interno di un ecosistema connesso e scalabile, governato da solidi standard. I ricercatori delle scienze della vita, sia accademici che industriali, fanno affidamento su risorse dati affidabili, gestite attraverso una governance solida, che seguono un ciclo di vita sostenibile e garantiscono un impegno a lungo termine nei confronti degli utenti. Inoltre, i ricercatori e gli utilizzatori dei dati necessitano di accesso aperto a risorse tecnicamente e scientificamente eccellenti per supportare la scoperta, la conservazione e il riutilizzo efficace dei dati. In qualità di promotore dell' Accesso Aperto (Open Access), la Piattaforma Data di ELIXIR dà priorità alla ricerca finanziata con fondi pubblici e promuove il riutilizzo e la rielaborazione dei dati tramite licenze aperte e condizioni d'uso coerenti con questo impegno (si veda la Open Definition per l'elenco delle licenze aperte). Il Research Data Management Kit (RDMkit) di ELIXIR è una risorsa online che contiene linee guida per la gestione dei dati, applicabili a progetti di ricerca,

in particolare nel campo delle scienze della vita. Il Nodo Italiano di ELIXIR contribuisce attivamente a questa risorsa, supportando la definizione di linee guida e strumenti di riferimento per aiutare i ricercatori ad affrontare le problematiche legate all'intero ciclo di vita dei dati, migliorando così la gestione dei dati di ricerca affinché siano Findable, Accessible, Interoperable e Reusable (FAIR). Inoltre, ELIXIR-IT può fornire consulenza e collaborare nell'istituzione di repository dati specifici per le scienze della vita, così come nello sviluppo di sistemi di recupero delle informazioni per garantirne l'accessibilità.

Sito web: https://rdmkit.elixir-europe.org/

PIATTAFORMA TOOLS

Per ottenere una comprensione più approfondita dei dati delle scienze della vita, gli scienziati hanno bisogno di strumenti software per accedere, studiare e confrontare i dati.
L'obiettivo principale della Piattaforma Tools di ELIXIR è migliorare la scoperta, qualità, disponibilità e sostenibilità delle risorse software bioinformatiche. La comunità scientifica italiana vanta una lunga e consolidata tradizione nello sviluppo di strumenti innovativi per l'analisi computazionale e statistica di diversi tipi di dati biologici, molti dei quali sono considerati all'avanguardia nei rispettivi ambiti di applicazione. Tutti gli strumenti sviluppati vengono aggiunti al repository bio.tools, che ad oggi conta oltre 300 strumenti sviluppati da membri della comunità ELIXIR-IT.
Inoltre, la piattaforma Tools di ELIXIR-IT ha selezionato 27 strumenti "core", sulla base di parametri chiave come la base di utenti, la novità e unicità dell'approccio, le prospettive di sviluppo e la quota di mercato attuale/attesa. Questi strumenti coprono praticamente ogni ambito della bioinformatica moderna.

Principali ambiti coperti:

- **Protein analysis, modeling and annotation**
Strumenti per l'analisi proteica, dalla sequenza primaria alla modellazione strutturale, previsione delle interazioni e annotazione funzionale.

- **Genomic Sequencing and Assembly:** Strumenti per l'assemblaggio di sequenze genomiche a partire da letture prodotte da diverse piattaforme di sequenziamento.

- **Analisi omiche/genomiche**
Strumenti per l'assemblaggio degli aplotipi, identificazione delle varianti e previsione delle varianti associate a malattie.
Gli strumenti di bioinformatica dell'RNA includono: analisi e caratterizzazione del trascrittoma ed epitrascrittoma, progettazione e valutazione di RNA terapeutici (inclusi vaccini, siRNA, miRNA e oligonucleotidi antisenso - ASO).
- **Strumenti per l'annotazione di elementi regolatori sia nel genoma** (es. siti di legame di fattori di trascrizione) sia nelle sequenze RNA (es. previsioni dei target di miRNA).

- **Servizi Galaxy**
ELIXIR-IT mantiene diversi servizi Galaxy specifici per dominio, tra cui:
- ARIES: per la caratterizzazione genomica dei microrganismi
- BioMaS e ORIONE: per l'analisi di dati di metabarcoding e metagenomica
- CorGAT: per l'annotazione delle varianti genomiche di SARS-CoV-2
- VINYL: per la prioritizzazione delle varianti genomiche umane

Inoltre, molte altre istanze Galaxy pubbliche e private sono mantenute da istituzioni esterne che sfruttano il servizio Laniakea@ReCaS, una piattaforma Galaxy on-demand resa disponibile dalla piattaforma di calcolo ELIXIR-IT:
https://galaxyproject.org/elixir-it/#communities-laniakeas-public-servers

PIATTAFORMA COMPUTE

La Piattaforma di Calcolo di ELIXIR-IT è composta da servizi eterogenei di calcolo ad alte prestazioni (HPC), cloud e storage, distribuiti tra diverse istituzioni, tra cui INFN, GARR, CINECA, CNR, CRS4 ed ENEA. Ogni servizio disponibile copre diversi aspetti delle attività tipicamente richieste dalla comunità ELIXIR. Inoltre, per migliorare l'uso, il riutilizzo e il valore dei dati nel campo delle Scienze della Vita, la piattaforma supporta risorse dati di alta qualità e sostenibili, all'interno di un ecosistema connesso e scalabile, governato da standard solidi. L'approccio Cloud consente una maggiore flessibilità nell'offerta di cluster di calcolo e nel supporto agli sviluppatori nella costruzione di servizi avanzati. È disponibile una soluzione di tipo Infrastructure as a Service (IaaS) che offre cluster di calcolo altamente personalizzabili. Insieme, i servizi della Piattaforma di Calcolo ELIXIR e i partner tecnologici e scientifici offrono una soluzione completa per l'accesso e l'analisi dei dati ELIXIR.

Le risorse di calcolo e storage disponibili, ospitate presso il Data Center ReCaS-Bari, includono:
- un elevato numero di core CPU con ampie capacità di memoria,
- schede GPU di ultima generazione in grado di accelerare algoritmi di machine learning (ML), deep learning (DL) e intelligenza artificiale (AI)
- una soluzione di archiviazione veloce basata su tecnologie SSD, per ridurre la latenza dei dati.

Questa architettura consente l'implementazione rapida ed efficiente di:
- Macchine Virtuali (VM) preconfigurate con Ambienti Virtuali (VE) per eseguire analisi bioinformatiche senza passare per la fase di installazione e configurazione, garantendo l'integrità e la sicurezza dei dati;
- una soluzione IaaS per offrire agli utenti la massima flessibilità possibile;
- una soluzione PaaS per il deployment di applicazioni on-demand;
- un ambiente HPC dedicato alla formazione, test e validazione di modelli ML/DL/AI;
- soluzioni di archiviazione per la conservazione dei dati a medio e lungo termine.

Inoltre, ELIXIR-IT offre il servizio Laniakea@ReCaS (https://laniakea-elixir-it.github.io/), una soluzione PaaS in grado di distribuire servizi bioinformatici comunemente utilizzati, come Galaxy, RStudio e JupyterLab, sfruttando risorse Cloud. Nascondendo la complessità tecnica dell'infrastruttura dietro una interfaccia web user-friendly, Laniakea consente agli utenti di configurare e distribuire applicazioni virtuali con pochi clic. Tra le funzionalità integrate vi è anche la crittografia dello storage su richiesta, rendendolo ideale per scenari in cui è necessario un pieno controllo amministrativo su un'istanza privata di Galaxy.

Infine, ELIXIR-IT supporta lo sviluppo di UseGalaxy.it, un server Galaxy nazionale a carattere generale, liberamente accessibile dalla comunità scientifica. Supporta un'ampia gamma di flussi di lavoro scientifici in diversi ambiti, offrendo un ambiente accessibile e intuitivo per l'analisi dei dati.