QUESITI COLLOQUIO – SELEZIONE N. 2025N14

I testi dei quesiti predisposti dalla Commissione sono i seguenti:

Domanda N. 1
Le moderne tecnologie di sequenziamento di nuova generazione (NGS) producono enormi quantità di dati. Qual è un esempio specifico di tecnologia NGS e quali sono le principali sfide computazionali associate all'allineamento e all'analisi di questi dati? Quali approcci algoritmici o software sono comunemente utilizzati per affrontare queste sfide?

The zinc finger (ZF) is the most frequently found sequence-specific DNA-binding domain in eukaryotic proteins. The ZF's modular protein–DNA interface has also served as a platform for genome engineering applications. Despite decades of intense study, a predictive understanding of the DNA-binding specificities of either natural or engineered ZF domains remains elusive. To help fill this gap, we developed an integrated experimental-computational approach to enrich and recover distinct groups of ZFs that bind common targets.

Domanda N. 2

Nel contesto di un'analisi bioinformatica complessa, che cosa si intende per "pipeline"? Spieghi perché l'utilizzo di pipeline automatizzate è cruciale per garantire la riproducibilità dei risultati, quali difficoltà si possono incontrare nella loro creazione e manutenzione, e quali strumenti software possono essere utilizzati per implementarle.

Single-cell transcriptomics (scRNA-seq) has become essential for biomedical research over the past decade, particularly in developmental biology, cancer, immunology, and neuroscience. Most commercially available scRNA-seq protocols require cells to be recovered intact and viable from tissue. This has precluded many cell types from study and largely destroys the spatial context that could otherwise inform analyses of cell identity and function. An increasing number of commercially available platforms now facilitate spatially resolved, high-dimensional assessment of gene transcription, known as 'spatial transcriptomics'

Domanda N. 3

La correzione degli errori è una fase fondamentale nell'analisi di dati biologici, in particolare quelli provenienti da tecnologie di sequenziamento. Descriva i tipi di errori più comuni che si possono riscontrare, e spieghi quali approcci computazionali sono utilizzati per identificarli e correggerli.

Systems-level analyses, such as differential gene expression analysis, co-expression analysis, and metabolic pathway reconstruction, depend on the accuracy of the transcriptome. Multiple tools exist to perform transcriptome assembly from RNAseq data. However, assembling high quality transcriptomes is still not a trivial problem. This is especially the case for non-model organisms where adequate reference genomes are often not available. Different methods produce different transcriptome models and there is no easy way to determine which are more accurate.

Domanda N. 4

Descriva il concetto di "cloud computing" e spieghi perché è sempre più utilizzato in ambito bioinformatico. Quali sono i principali vantaggi e gli svantaggi dell'utilizzo di risorse cloud rispetto a un'infrastruttura locale?

Single-cell (SC) gene expression analysis is crucial to dissect the complex cellular heterogeneity of solid tumors, which is one of the main obstacles for the development of effective cancer treatments. Such tumors typically contain a mixture of cells with aberrant genomic and transcriptomic profiles affecting specific sub-populations that might have a pivotal role in cancer progression, whose identification eludes bulk RNA-sequencing approaches. We present an R package that enables the characterization of cell identity in solid tumors on the basis of a various and complementary analyses on SC gene expression data.

Domanda N. 5

Un ricercatore le chiede di avviare un job di analisi genomica su un cluster di calcolo gestito da Slurm. Descriva le caratteristiche essenziali di uno script di job Slurm, inclusi i parametri fondamentali per la gestione delle risorse (es. memoria, tempo di esecuzione, numero di core), e spieghi come questi parametri influenzano l'esecuzione del job.

While DNA barcoding methods are an increasingly important tool in biological conservation, the resource requirements of constructing reference libraries frequently reduce their efficacy. One efficient way of sourcing taxonomically validated DNA for reference libraries is to use museum collections. However, DNA degradation intrinsic to historical museum specimens can, if not addressed in the wet lab, lead to low quality data generation and severely limit scientific output. Several DNA extraction and library build methods that are designed to work with degraded DNA have been developed, although the ability to implement these methods at scale and at low cost has yet to be formally addressed.