*Proposta di corso trasversale:*

**Data analysis and interpretation in molecular medicine: from databases to artificial intelligence**

***Module 2: TRANSCRIPTOMICS***

**Proponenti:**

- Prof. Enza Maria Valente (coordinatore, corso di dottorato in Translational Medicine)

- Prof. Riccardo Bellazzi (corso di dottorato in Bioingegneria, Bioinformatica e Tecnologie per la Salute)

**Obiettivi formativi:**

Questo corso trasversale si articolerà in tre moduli indipendenti (Genomics, Transcriptomics, Proteomics), che saranno svolti a rotazione su tre anni. In questo modo il corso sarà fruibile ogni anno a tutti i dottorandi dei cicli attivi, senza ripetizioni.

Alla fine del secondo modulo (“Transcriptomics”), il dottorando sarà in grado di:

1) conoscere ed utilizzare i database di espressione genica e trascrittomica disponibili in rete;

2) conoscere gli strumenti informatici per l’interpretazione dei dati di espressione genica ed esperimenti di transcriptomics and single-cell transcriptomics e per la costruzione di modelli predittivi;

3) affrontare problemi di analisi di dati di espressione genica.

**Specifiche del corso:**

Numero di ore/lezioni: 20

Periodo di svolgimento: giugno 2020

Docenti: Valente, Bellazzi, Cereda, Zupan, others to be defined

Comitato scientifico: Valente, Gherardi, Bellazzi, Zupan, Lolicato

Modalità di verifica dell'apprendimento: sessione pratica finale di esercitazione

**Potenziali dottorati interessati:**

*Macroarea di Scienze della Vita:* Translational Medicine; Genetica, Biologia Molecolare e Cellulare

*Macro-area di Scienze e Tecnologie:* Tecnologie per la Salute, Bioingegneria e Bioinformatica; Ingegneria Elettronica, Informatica ed Elettrica

**Programma preliminare:**

- Session 1 (8 h): Introduction to transcriptomics and bioinformatics databases and tools

In questa prima giornata, i dottorandi riceveranno lezioni frontali per conoscere i più importanti database di espressione genica e trascrittomica, confrontarli ed imparare ad utilizzarne le funzioni più importanti. In particolare, saranno discussi i database di sequenza (NCBI genome, RefSeq, Ensembl, UCSC…), i database di espressione genica (es. GTEX, GEO, BrainTx, GO ontology, KEGG) e nuovi databases di single cell transcriptomics (es. mouse organogenesis cell atlas - MOCA)

- Session 2 (8 h): Data analytics and artificial intelligence tools

In questa seconda giornata, i dottorandi riceveranno lezioni frontali che introdurranno diversi approcci metodologici per l’analisi dei dati di trascrittomica e per la loro interpretazione. Verranno presentati approcci di gene-set enrichment, di costruzione e analisi di reti di trascrittomi e si introdurranno metodi di machine learning applicati alla trascrittomica, includendo sia gli approcci non supervisionati, presentando i principali metodi di clustering e di data projection, sia i metodi supervisionati e di classificazione utilizzati in campo genomico. In conclusione, si presenteranno i moderni metodi di pre-processing e embedding basati su deep learning.

- Session 3 (4 h): Hands-on practice

In questa mezza giornata finale del corso, che si svolgerà in aula multimediale, i dottorandi avranno la possibilità di mettere in pratica quanto appreso utilizzando pacchetti software per il machine learning (R, Python) e risolvendo alcuni quesiti che richiedono l’utilizzo dei database genomici.