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

***Module 1: GENOMICS***

**Proponenti:**

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

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

**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 primo modulo (“Genomics”), il dottorando sarà in grado di:

1) conoscere ed utilizzare i database genomici disponibili in rete;

2) conoscere gli strumenti informatici per la costruzione di modelli predittivi per l’interpretazione delle varianti genomiche;

3) affrontare problemi di annotazione di varianti genomiche identificate mediante next generation sequencing.

**Specifiche del corso:**

Numero di ore/lezioni: 20

Periodo di svolgimento:

* Mercoledi 19 giugno (tutta la giornata)
* Giovedi 20 giugno (tutta la giornata)
* Venerdi 21 giugno (solo mattina)

Docenti: Valente, Errichiello, Bellazzi, Limongelli, others to be defined

Comitato scientifico (da confermare): Valente, Pastore, Gherardi, Bellazzi, Zupan, Riva, Magni, Binda, Bione, Peverali

Modalità di verifica dell'apprendimento: MCQ e 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 genomics and bioinformatics databases and tools

In questa prima giornata, i dottorandi riceveranno lezioni frontali per conoscere i più importanti database genomici, 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 popolazione (gnomAD, EVS, dbSNP…) ed i database di patologia (ClinVar, OMIM, HGMD, LOVD, Decipher…).

- 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 e la costruzione di modelli predittivi, comprendenti la distinzione fra approcci supervisionati e non supervisionati, i principali metodi di clustering e di classificazione utilizzati in campo genomico e infine 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.