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

***Module 1: GENOMICS***

**Organizers:**

- Prof. Enza Maria Valente (coordinator, PhD course Translational Medicine)

- Prof. Silvana Quaglini, Prof. Riccardo Bellazzi (coordinator and former coordinator, PhD course Tecnologie per la Salute, Bioingegneria e Bioinformatica)

**Objectives of the course**

Participants will learn how to evaluate NGS (e.g. whole exome) data by bio-informatic inspection of NGS files, how to interrogate online genomic databases and interpret variants for clinical purposes. Some popular and freely-available tools will be used; examples of variants identified in real clinical cases will be provided.

Participants will then learn the basics of machine learning and data analytics. Teaching will be performed following a learning-by-doing approach. A focus will be performed on representing genomics information as features of machine learning algorithms to build predictive models for genomic variant interpretation.

**Course specifics:**

How long: 20 hours

When: July 2022

Instructors: Proff. Valente, Errichiello, Giorgio, Bellazzi, Pasotti, Lescai, others to be defined

Scientific committee: Proff. Valente, Bellazzi, Giorgio, Lescai, Pasotti.

Evaluation test: hands-on practical tests

**PhD involved:**

*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

**Preliminary program:**

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

In the first day, PhD students will receive frontal lessons to learn the commonest and most important genomic databases, compare them and learn to utilize their most relevant functions. This will cover sequence dabatases (NCBI genome, RefSeq, Ensembl, UCSC…), population databases (gnomAD, EVS, dbSNP…) and disease databases (ClinVar, OMIM, HGMD, LOVD, Decipher…).

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

PhD students will be introduced to a range of methodological strategies for data analysis and building of predictive models, including distinction between supervised and non-supervised approaches, main strategies of clustering and classification adopted in genomic analysis and modern methods of pre-processing and embedding based on deep learning.

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

In the last half-day of the course, in a computer classroom, PhD students will have the chance to put in practice what they have learnt by confronting themselves with a practical exercise session: this will include using software for machine learning (R, Python) and solving problems which require the use of online tools and databases.