Ph.D. in Information Technology
Thesis Defense

October 27th, 2025
at 10:30 am
"Emilio Gatti" Conference Room – building 20

Aldo Sergi – XXXV Cycle

Computational approaches in oncogenomics: from synthetic tumor data generation to machine-learning classification for ovarian cancer

Supervisor: Prof. Marco Masseroli

Abstract:

Ovarian cancer, especially its high-grade serous subtype, remains a major challenge in oncology due to late diagnosis, tumor heterogeneity, and frequent resistance to chemotherapy. Recent advancements in genomics and Next-Generation Sequencing (NGS) technologies have revealed the intricate web of genetic mutations and alterations present in ovarian cancer. These findings underscore the urgent need for computational strategies to improve diagnostic accuracy and therapeutic decision-making.

This thesis focuses on two critical topics in oncogenomics. First, it focuses on detecting somatic variants at low fractions, a critical bottleneck for clinical applications such as circulating tumor DNA (ctDNA) analysis. A novel computational approach was developed to generate realistic synthetic tumor datasets as a general-purpose resource for developing and benchmarking computational methods in oncogenomics. Leveraging tools like BAMSurgeon and Nextflow, we created ultra-deep, low-fraction sequencing samples with a complete ground-truth catalogue of mutations. These datasets support a range of computational and analytical applications, including the development and evaluation of bioinformatic tools. Benchmarking and evaluation of state-of-the-art variant callers is presented here as a showcase, revealing concrete strategies to raise sensitivity and specificity at allele fractions below 1%. The same approach can be extended to other analytic tasks, including quality-control pipelines and machine-learning model training, thereby reducing experimental costs and enhancing reproducibility across laboratories.

Second, the thesis introduces Homologous Recombination-Signature Classifier (HR-SC), a new developed machine learning-based framework that predicts Homologous Recombination Deficiency (HRD). By integrating Copy Number signatures and BRCA mutation status, HR-SC distinguishes tumors as either HRD or Homologous Recombination Proficient (HRP). Its performance has been validated in two independent clinical trial datasets (PAOLA-1 and MITO16A/MaNGO-OV2), revealing a robust predictive value for response to Parp-inhibitors and a prognostic role under standard platinum-based therapy. This dual utility underscores the impact of Copy Number signatures and Machine Learning in refining the molecular characterization of ovarian cancer.

This work explores two distinct yet complementary research directions: enhancing low-fraction variant detection and improving HRD classification. Both areas leverage genomics, bioinformatics and computational approaches, with the latter also using machine learning to refine patient stratification for targeted therapies, highlighting the significant impact of data-driven innovation in precision oncology.

PhD Committee

Prof. Rosario Piro, Politecnico di Milano

Prof. Maria Coral del Val Muñoz, Universidad de Granada

Prof. Luigi Cerulo, Università degli Studi del Sannio