**Project title**: Novel methods and tools for integrating large single-cell datasets using machine learning and artificial intelligence approaches.

PI. Dr. Claudia Angelini (claudia.angelini@cnr.it),

Istituto per le Applicazioni del Calcolo-Consiglio Nazionale delle Ricerche

**Abstract**

Integrating omics and multi-omics data has shown great potential to improve inference and prediction in several contexts. On the one hand, the integration of bulk data in either supervised or unsupervised frameworks has already provided novel ways for studying cancer and developmental diseases and characterizing immune responses, demonstrating that data integration can extract hidden biological signals from complex datasets. Conversely, the rapid advancement of single-cell technologies has enabled unprecedented resolution in profiling cellular heterogeneity across diverse biological systems. However, integrating heterogeneous single-cell datasets presents several significant challenges. Nowadays, while several methods have already been proposed for batch removal and clustering, the post hoc analysis of single cells still requires the development of novel approaches. For example, inferring accurate, scalable, and robust GRNs from single-cell data must face data sparsity, noise, and variability across different experiments and modalities.

This project proposes a machine learning-driven framework for integrating multimodal single-cell datasets—including transcriptomics, chromatin accessibility, and others—to enable high-fidelity GRN inference and develop open-source, user-friendly computational tools and pipelines to support the broader research community in single-cell GRN analysis.

Overall, the project will combine advanced integration techniques such as dimensionality reduction techniques, Bayesian approaches, machine learning methods, variational autoencoders, and other artificial intelligence approaches to deliver robust computational tools for comprehensive single-cell data integration with a specific focus on GRN inference. This project will improve our understanding of complex biological systems and disease processes by benchmarking public datasets and validating the proposed approaches in biologically relevant contexts such as immunological disorders.