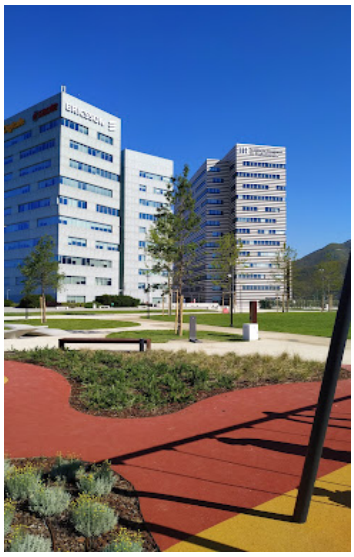
ISTITUTO ITALIANO DI TECNOLOGIA
INTEGRATIVE NUCLEAR ARCHITECTURE

the giovanni ARMENISE HARVARD foundation

Integrative Nuclear Architecture
Istituto Italiano di Tecnologia



Who we are:

The Armenise-Harvard Integrative Nuclear Architecture (INA) research line is located at the Center for Human Technologies (CHT@Erzelli), Istituto Italiano di Tecnologia in Genova. INA's mission is to investigate the 3D genome organisation plasticity and how multiple components of the nucleus influence gene regulation information across scales. To this end, we integrate imaging, genomics data, and bioinformatics by implementing ad-hoc quantitative analysis and modelling tools. INA is an active member of the FANTOM6 collaborative project and of the Center of Genome Imaging.
More information at:
https://www.farabellalab.org/
https://ina.iit.it/it/

You will be working in a multicultural and multi-disciplinary group, where computational and experimental researcher collaborate, each with their own expertise, to advance our knowledge in how the human genome is organized and regulated.

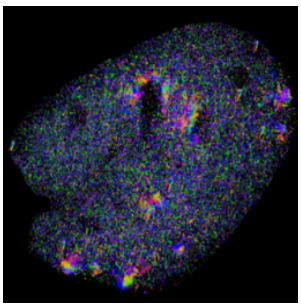
What we offer:

We can offer up to 2 or 3 internships (depending on duration).
Support: up to 800 euros gross monthly.
Requirements: machine learning, Python scientific programming.
Plus: Knowledge of version control systems (e.g., GitHub)

We seek highly motivated and enthusiastic candidates willing to challenge an innovative project by adopting an interdisciplinary approach. *Varied projects are available*, depending on candidate interests, all focusing on developing new computational methods for analyzing and interpreting multiplexed single-cell imaging to measure and predict genome folding patterns using a combination of data modelling, optimization, and machine learning techniques. Possible project topics:

Topic 1: Chromatin Imaging Data Analysis: detection

This research project is related to the extension and application of our python library Chromatin IMaging Analysis tool (CIMA, unpublished), aimed at automatic processing and analysis of chromatin imaging data. Specifically, the project will focus on point-cloud data as thus obtained with Single-molecule localization microscopy (SMLM) experiments as OligoSTORM (*Nir*, Farabella* et al. PlosGen 2018*), aiming at developing and testing automated methods for the identification and decoding of imaged chromatin loci. Due to the *Big Data* nature of the chromatin imaging dataset, the project will require improving and optimising the CIMA clustering pipeline. Possible directions might include hyperparameter optimisation for density-based algorithms (e.g., DBSCAN and HDBSCAN) or the application of network-based (e.g. Louvain method) community detection algorithms. The project will require coding the algorithmic implementation and analysis of both synthetic and real data.



17.5Mb of Chr21 and 7.5Mb of Chr1 in male fibroblast cell line. Image acquired by Sarah Aufmkolk and Antonios Lioutas at Harvard Medical School.

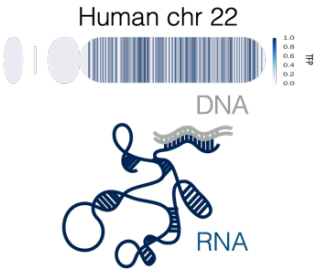
Topic 2: Chromatin Imaging Data Analysis

This research project will aim to identify 3D organisation patterns based on chromatin imaging data (generally 3D point-cloud data; see Flores, Farabella, Nir Curr Opin Cell Biol. 2023). Specifically, the project aims to implement unsupervised learning techniques to infer chromatin loci sub-type based on morphological and spatial features. The project will require testing appropriate representation methods and embedding strategies (e.g. UMAP,

PaCMAP, and Autoencoders), as well as measuring algorithm performance. The algorithmic implementation will be part of our Python library Chromatin IMaging Analysis tool (CIMA, unpublished).

Topic 3: Computational Genomics

This research project will focus on the study of statistical preferences of Chromatin-associated RNAs (caRNAs), particularly those that bind the genome via triplex-formation (*Farabella et al. Nat. Struct. Mol. Biol. 2021; Morf et al. Nat Biotechnol. 2019*). Specifically, the project aims to acquire a multi-omics view of the lncRNA-chromatin interactome, integrating bioinformatic predictions, RADICL-seq and publicly available chromatin conformation capture (3C-based) and chromatin tracing experiments. The creation of this common framework across varied omic *Big Data* of will serve as the starting point to investigate changes in the network of interaction between lncRNAs and the chromatin (lncRNA-chromatin interactome) during neural differentiation. The project is part of the FANTOM6 collaborative efforts, a worldwide collaborative project aiming at identifying all functional elements in mammalian genomes.



Requirements and Further Actions:

If interested, contact irene.farabella@iit.it