Title: Improving the quality of Nanopore sequencing with Deep Learning

Supervisor: Alessio Ansuini
Affiliation: Research and Technology Institute, Area Science Park
Requirements: proficiency in python, knowledge of PyTorch is a plus
Keywords: Deep Learning, Nanopore sequencing, Genomics
Duration: 12 months
Validity period: 12 months since publication on web site

Project AIM

Next Generation Sequencing (NGS) techniques are a series of technologies that allow to massively sequence genomes in a relatively short time. They are essential to modern Science, being instrumental to numerous applications, ranging from precision medicine to the study of virus evolution, and the development of targeted vaccines.

Nanopore is a new technology of ultra-high throughput sequencing, that holds the promise to solve some of the problems of NGS sequencing, due to its much longer *reads* (a read is an inferred sequence of base pairs corresponding to a DNA fragment), that could greatly simplify whole genomes reconstructions, thus replacing NGS technologies it in the medium-long term [1][2]. The problem is that the current quality of the sequences obtained with Nanopore technologies are still below the NGS standards.

In this project we will develop new algorithms, trying to improve the quality of sequence callers obtained from Oxford Nanopore instruments. We will work on benchmark datasets and on new data coming from the PromethION sequencer [3] hosted in the LAGE laboratory in our Institute.
The methodology used will be a mixture of deep learning models for sequences and advanced clustering algorithms.

Bibliography

[1] https://nanoporetech.com/products/promethion
[2] https://www.nature.com/articles/s41587-021-01108-x
[3] https://genome.cshlp.org/content/30/9/1345