Politecnico di Torino
Dipartimento di Automatica e Informatica

PhD in Computer and Control Engineering
34 ° cycle

Supervisor
Prof. Elisa Ficarra

AI techniques applied to human genome analysis in cancer research

PhD Candidate: *Marilisa Montemurro*

ISTITUTO DI CANDIOLO-IRCCS
SmartData

1. Context

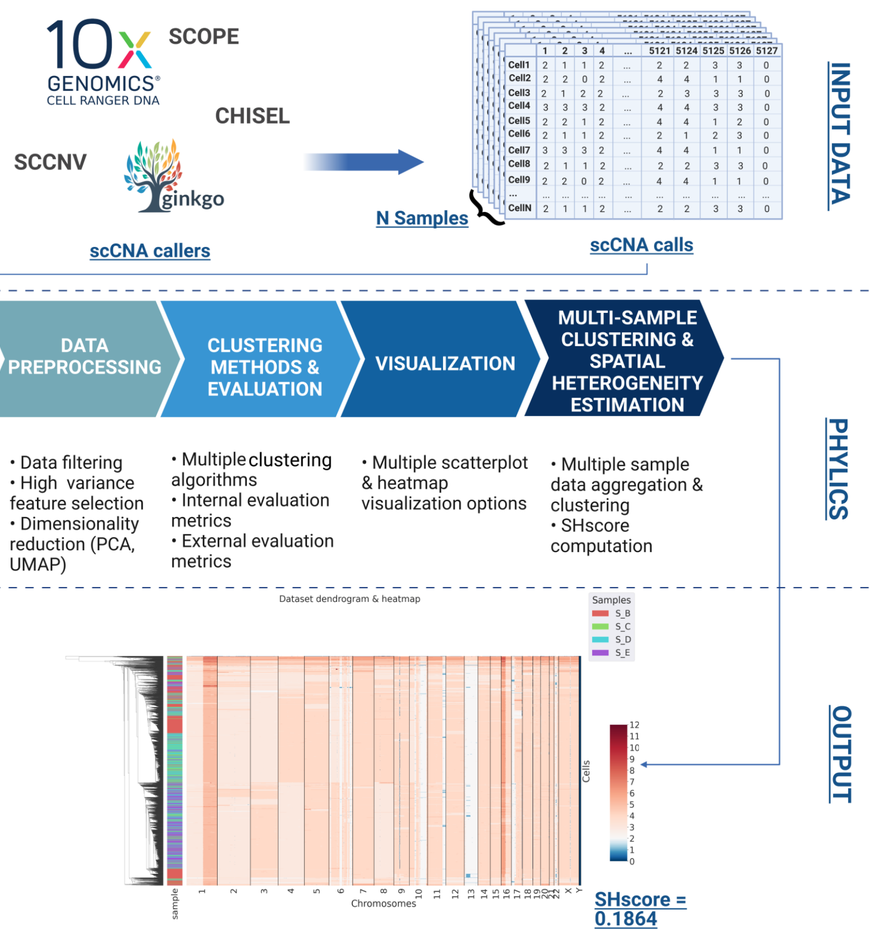
Next generation sequencing (NGS) is a high-throughput sequencing technology that allows to determine the order of nucleotides in entire genomes or targeted regions of DNA or RNA. NGS data are **high dimensional** and **complex**. Additionally, high-throughput NGS technologies are a source of "**big data".**

2. Goal

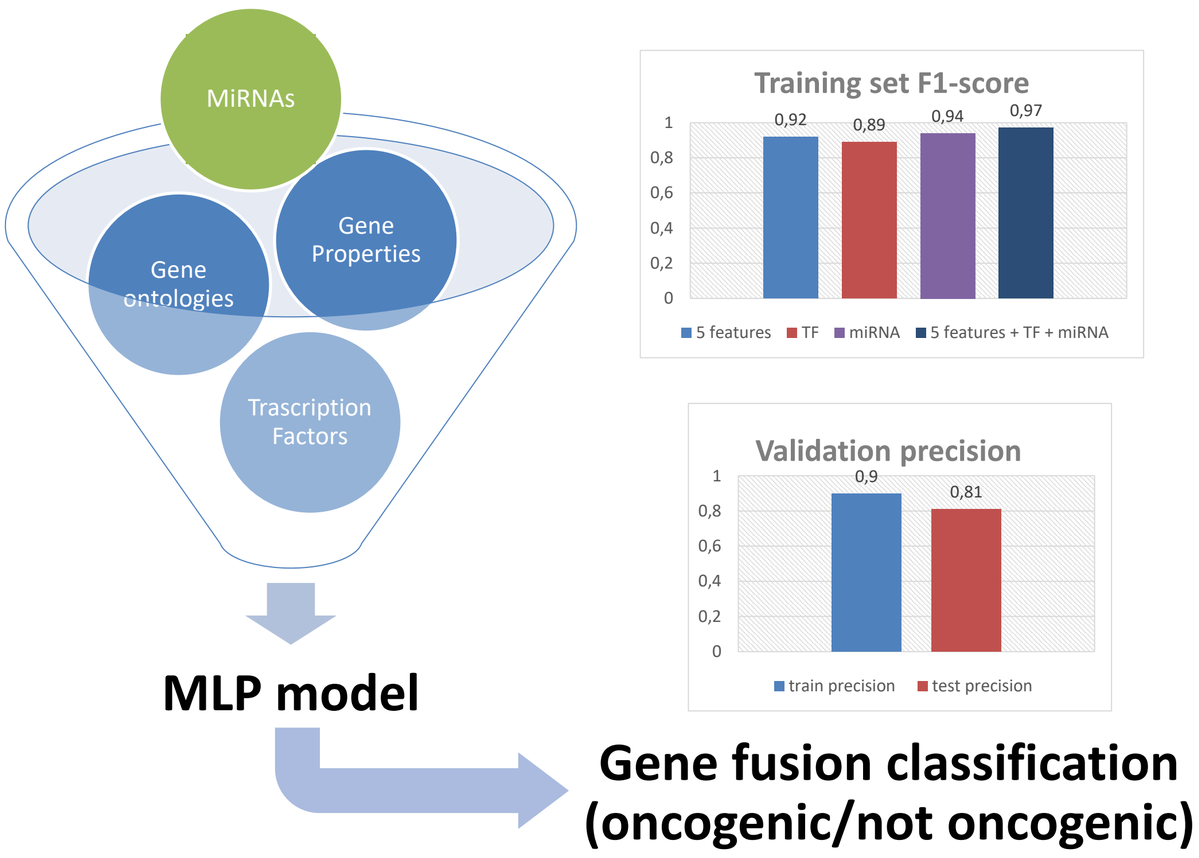
This PhD goal is to develop **AI-based methods** to explore, classify and model human genomic data.

3. Methods and Results

PhyliCS [1]: a flexible **Python library** that for, the first time, combines **multi-sample analysis** and **scDNA** data to perform a comparative analysis between multiple regions of the same tumor and estimate spatial **intra-tumor heterogeneity** through a new metric, the **Spatial-Heterogeneity score (SHscore)**.



ChimerDriver [2]: a **MLP-based** technique to classify **gene fusions** as **oncogenic/not-oncogenic** which, for the first time, combines gene-level features with context information regarding some molecules involved in the regulatory processes of fused protein synthesis.



4. Conclusions

PhyliCS and **ChimerDriver** represent two valuable instruments for research laboratories, allowing for better diagnosis, classification and treatment of cancer patients.

5. References

1. Montemurro, M., *et al.* PhyliCS: a Python library to explore scCNA data and quantify spatial tumor heterogeneity. *BMC Bioinformatics* **22,** 360 (2021).
2. Barrese, V. S., *et al.*, Identifying the oncogenic potential of gene fusions exploiting miRNAs. *bioRxiv* 2021.10.20.46512.