

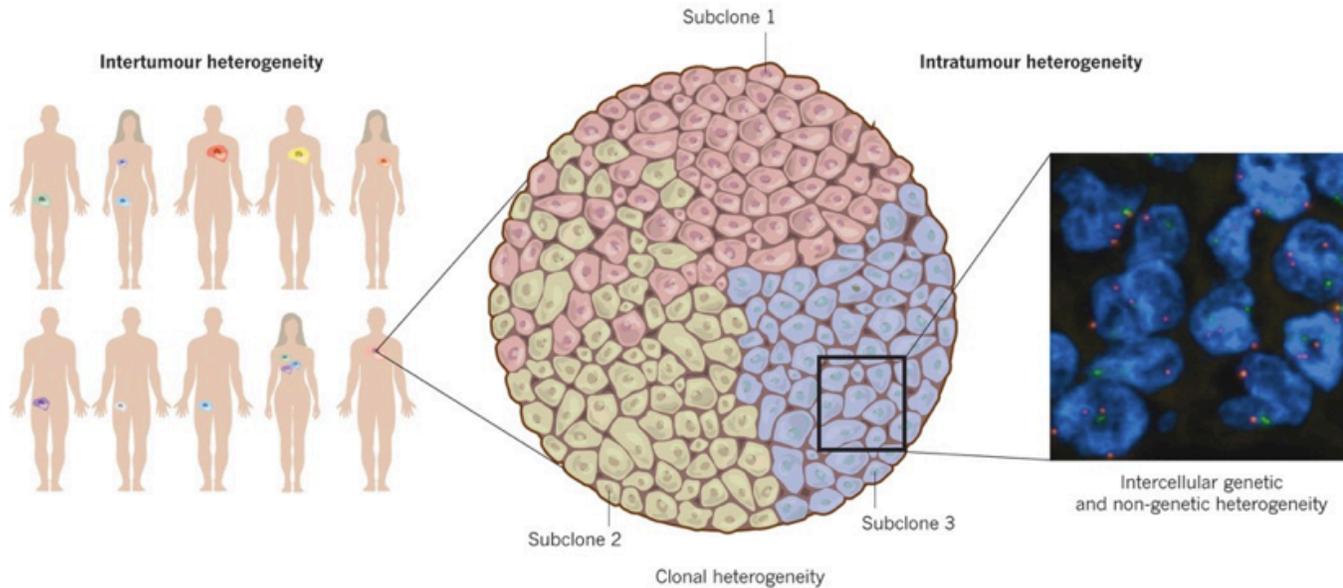
Borse di studio GARR
Orio Carlini

Quantum Machine Learning per l'identificazione di marker nei tumori rari

Valeria Repetto
IIT-CNR Pisa



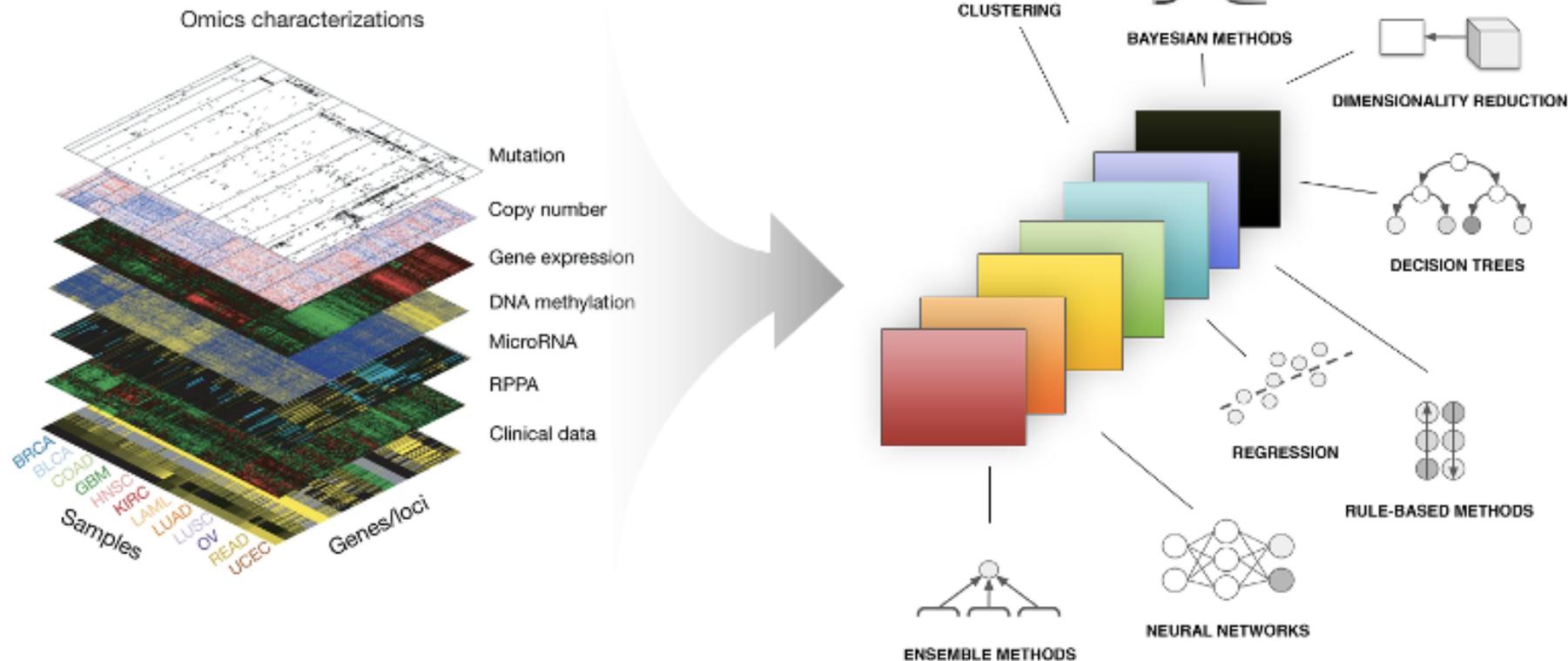
Oncologia di precisione



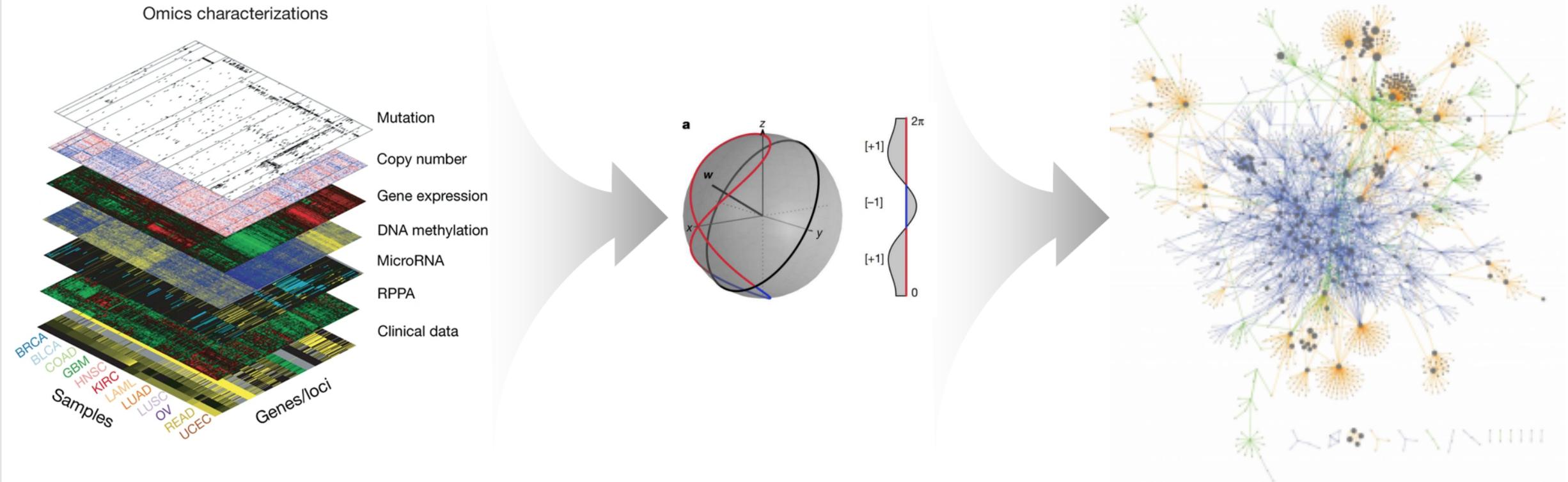
Terapie personalizzate sulle caratteristiche molecolari e cliniche del paziente

Metodi di Machine Learning necessitano di tanti dati

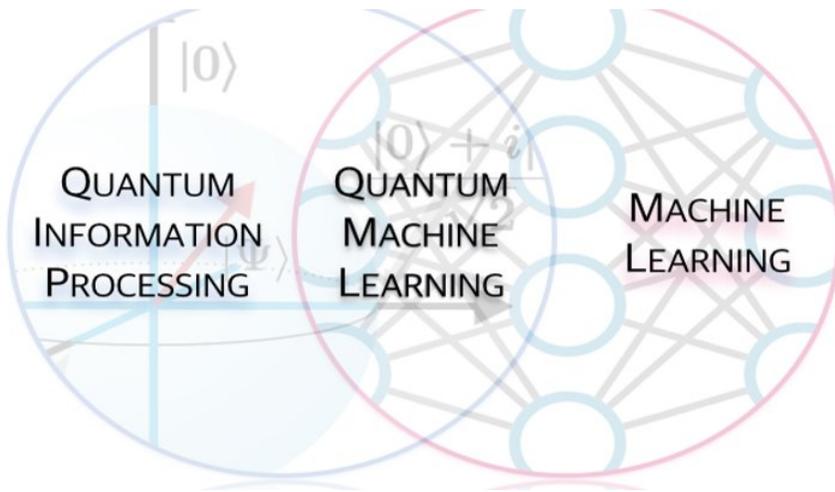
I tumori come sistemi complessi



È possibile sfruttare le Tecnologie Quantistiche per i sistemi biologici complessi?



Quantum Enhanced Machine Learning



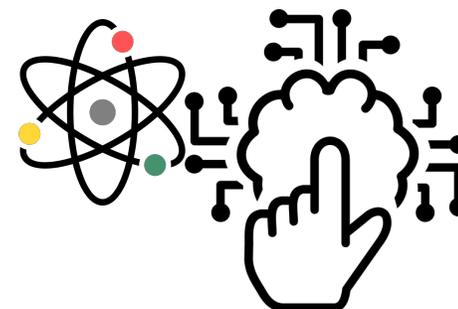
Metodi di **Machine Learning Classico** che possono essere potenziati dall'integrazione col **dominio quantistico**

Quali sono i vantaggi?

- **Model complexity:** Modelli con complessità non riproducibile classicamente
- **Sample Complexity:** Meno dati rispetto alla controparte classica

Scopo del progetto originale

I modelli **classici** spesso non riescono a catturare la **complessità** dei **dati oncologici**, in particolare nei tumori rari

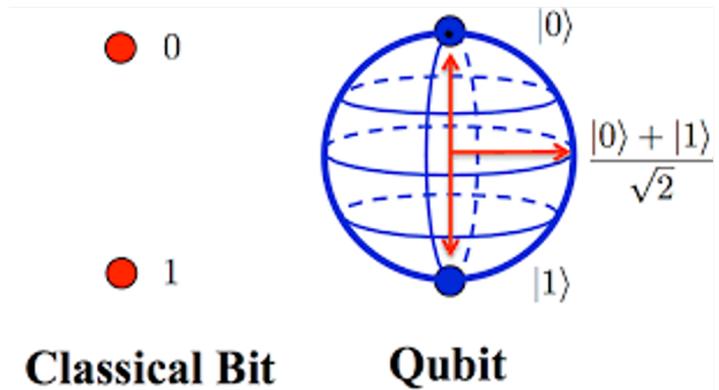
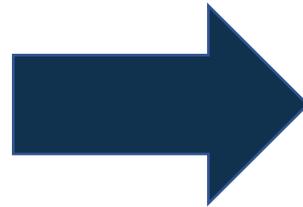


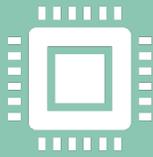
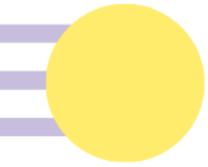
- Esplorare la **compatibilità** del Quantum Machine Learning (QML) con **dati biologici complessi**
- Sfruttare il Q per identificare **biomarcatori**

Il Quantum Computing

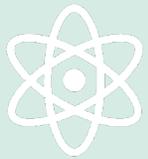
La computazione quantistica **cambia il paradigma** computazionale classico sfruttando i principi della **meccanica quantistica**

$$\frac{1}{\sqrt{2}}|\text{cat}\rangle + \frac{1}{\sqrt{2}}|\text{mouse}\rangle$$





La comunità scientifica è concorde sul riconoscere come questa tecnologia possa superare quella classica in specifiche applicazioni



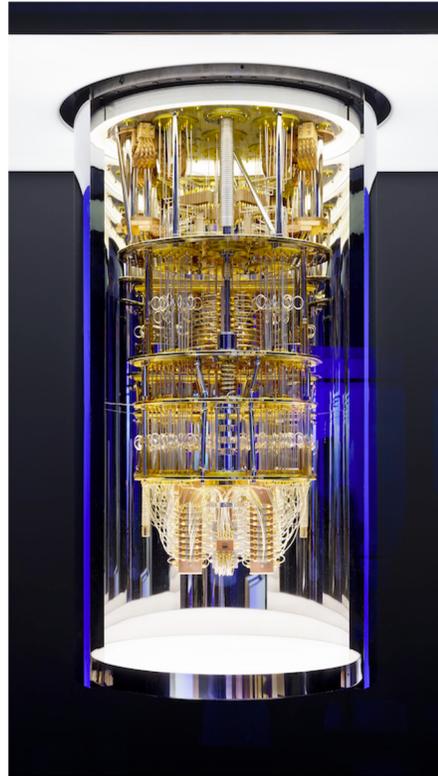
- Velocità • Problemi di ottimizzazione • Crittografia

Infrastrutture tecnologiche utilizzate

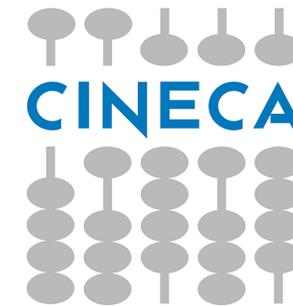
Quantum Hardware

IBM Quantum Network

- Gate Model
- Noisy Intermediate-Scale Quantum Device

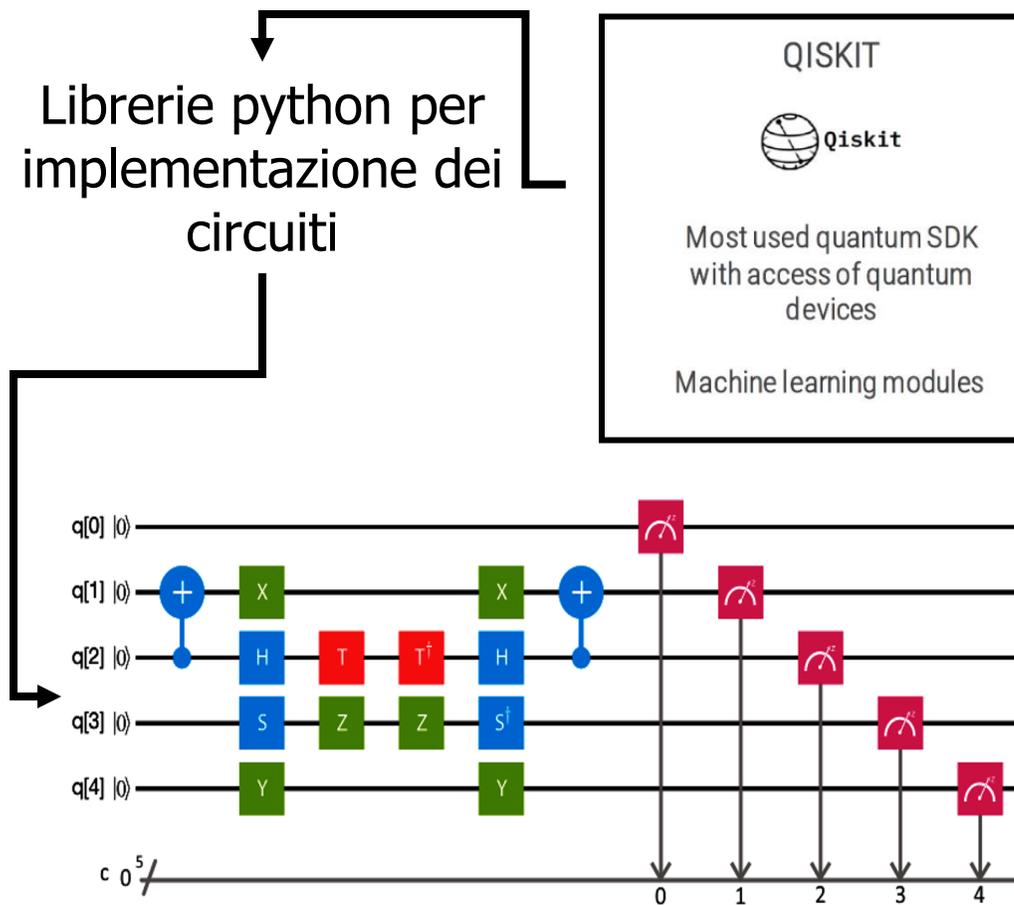
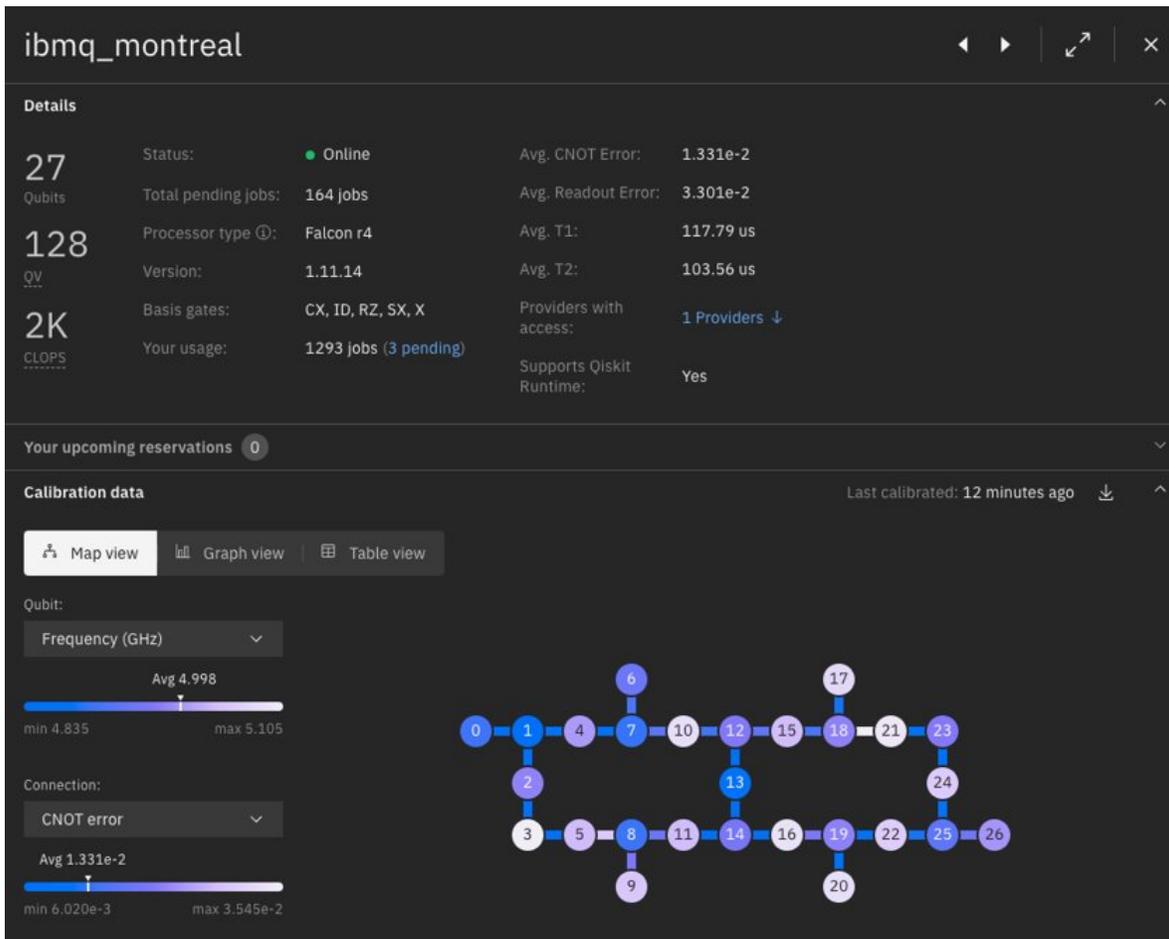


HPC Hardware



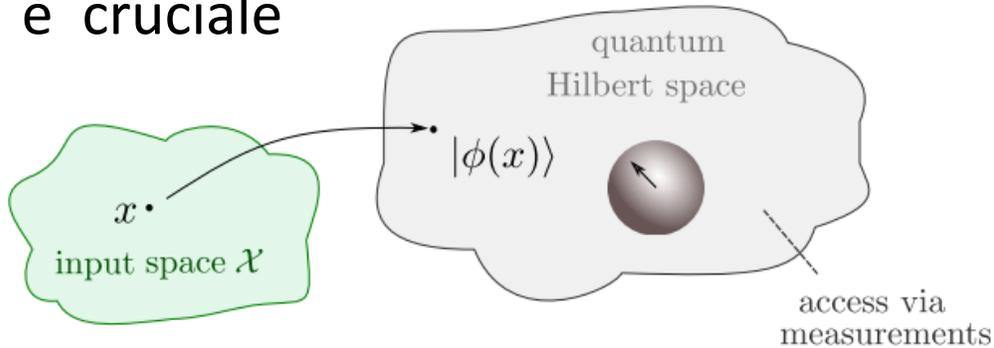
Simulazioni per **testare** gli algoritmi in una configurazione ideale

Implementazione del Gate model



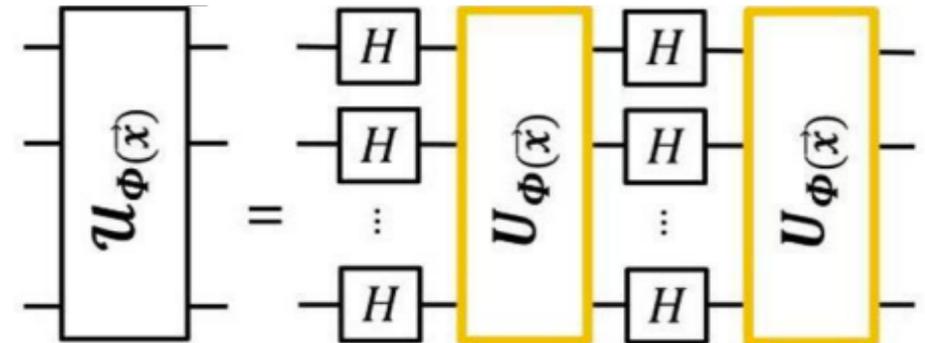
Importanza dell' embedding

La **codifica** dell'input classico nel dominio Quantistico è cruciale



Feature Map:

$$\vec{x} \mapsto |\Phi(\vec{x})\rangle = \mathcal{U}_{\Phi(\vec{x})} |0\rangle^{\otimes n}$$



Schul, Quantum models as kernel methods(2021)

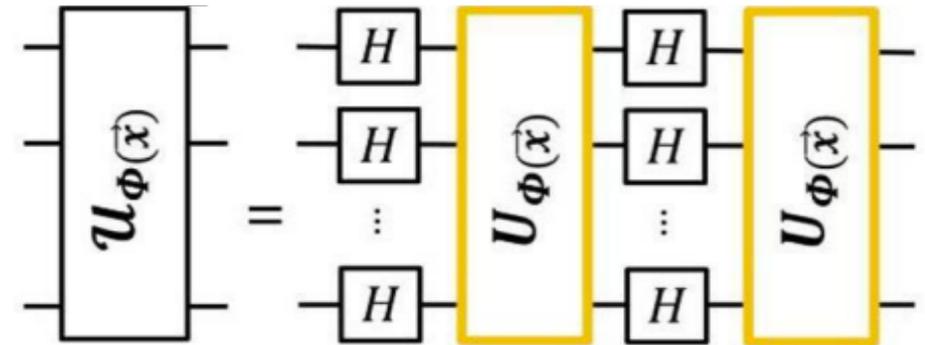
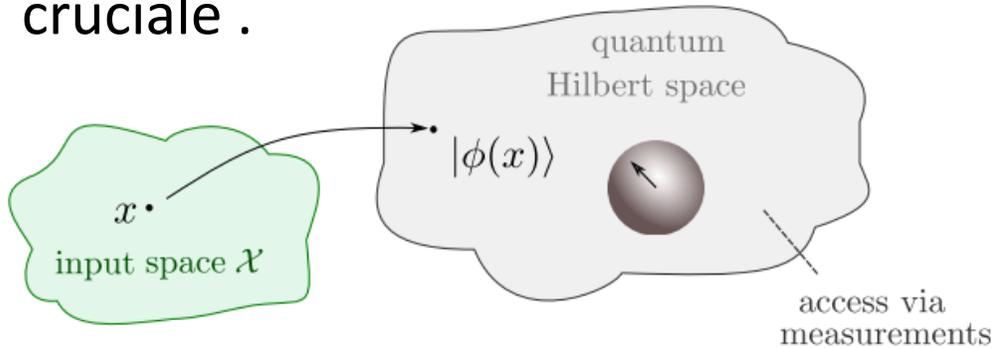
Importanza dell' embedding

La **codifica** dell'input classico nel dominio Quantistico è cruciale .



Feature Map:

$$\vec{x} \mapsto |\Phi(\vec{x})\rangle = \mathcal{U}_{\Phi(\vec{x})} |0\rangle^{\otimes n}$$

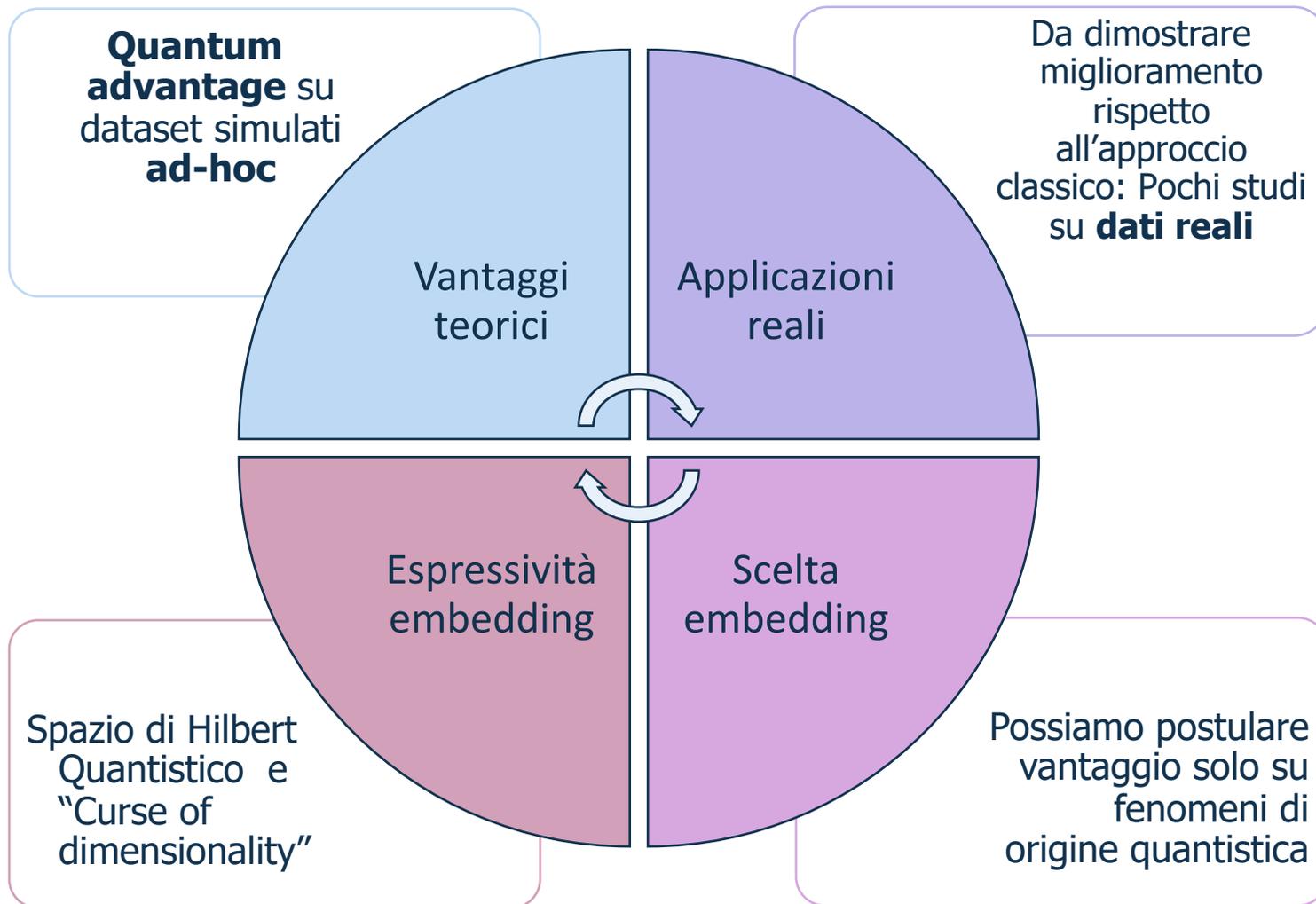


- **N Features = N Qubits**
- **Feature Map difficili da computare classicamente**



Schul, Quantum models as kernel methods(2021)

Sfide QML e dati reali



Jerbi, S. et al. (2023) Quantum machine learning beyond kernel methods.

Data set

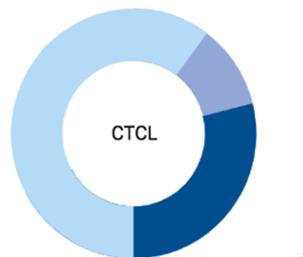
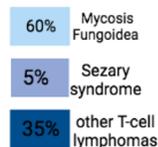
CTCL

- Tumore raro delle cellule T



6 studi
pubblici
raccolti e
processati

- Studi diversi con pochi campioni, data mancanti e rumorosi, elevata eterogeneità fenotipica, e effetti di batch non noti



Data set

CTCL



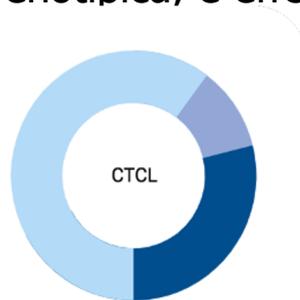
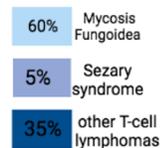
METABRIC

- Tumore raro cellule T



6 studi
pubblici
raccolti e
processati

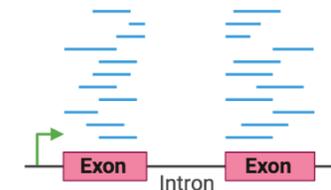
- Studi diversi con pochi campioni, data mancanti e rumorosi, elevata eterogeneità fenotipica, e effetti di batch non noti



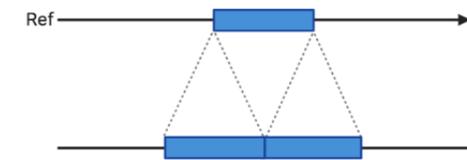
- Unico dataset **multi-omico** di Breast cancer



- Stratificazione clinicamente significativa:
Integrative Clustering (IC10)



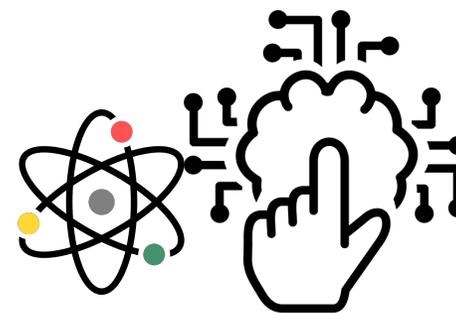
Espressione Genica
(~20 000 features)



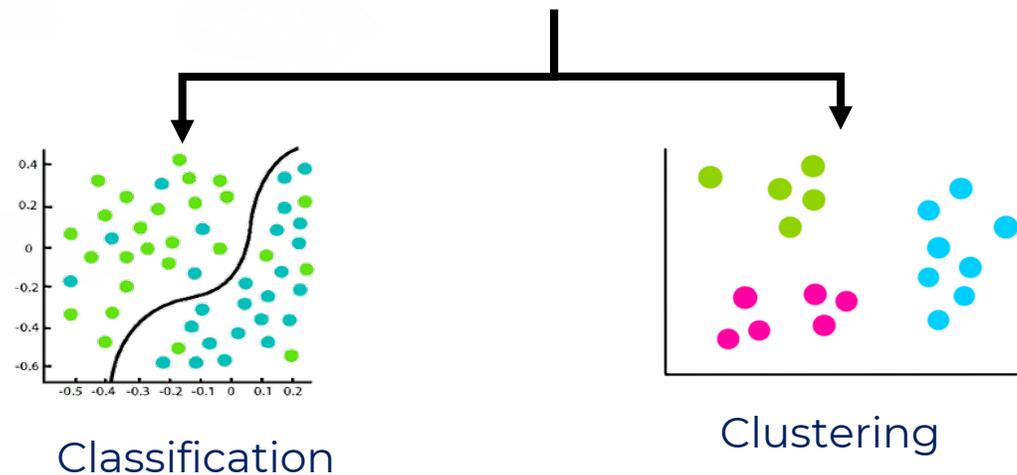
Variazione del numero
di copie
(~20 000 features)

Scopo del progetto sviluppato

I modelli **classici** spesso non riescono a catturare la **complessità** dei **dati oncologici**



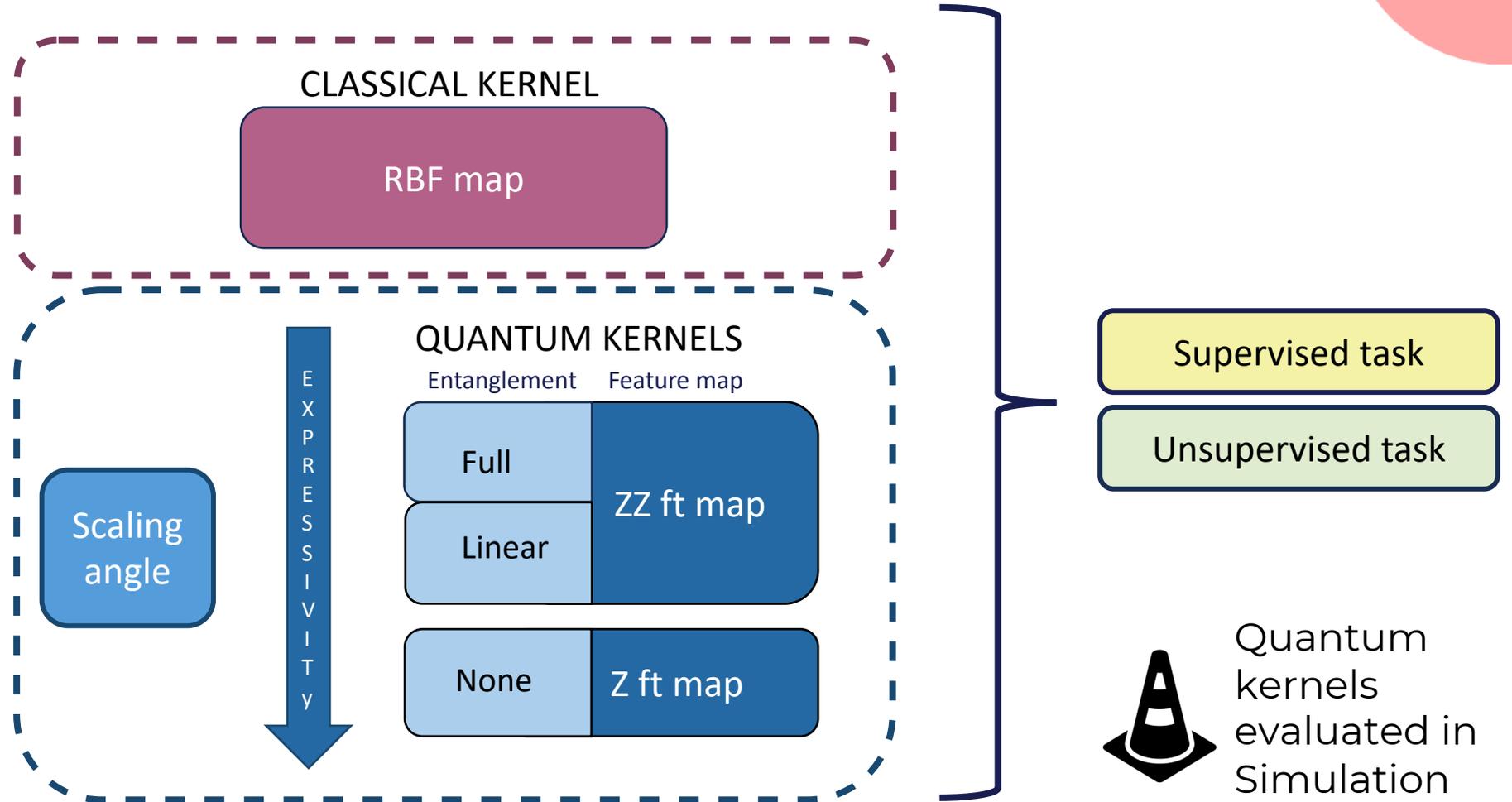
Esplorare la **compatibilità** del QML con **dati biologici complessi**



Design sperimentale

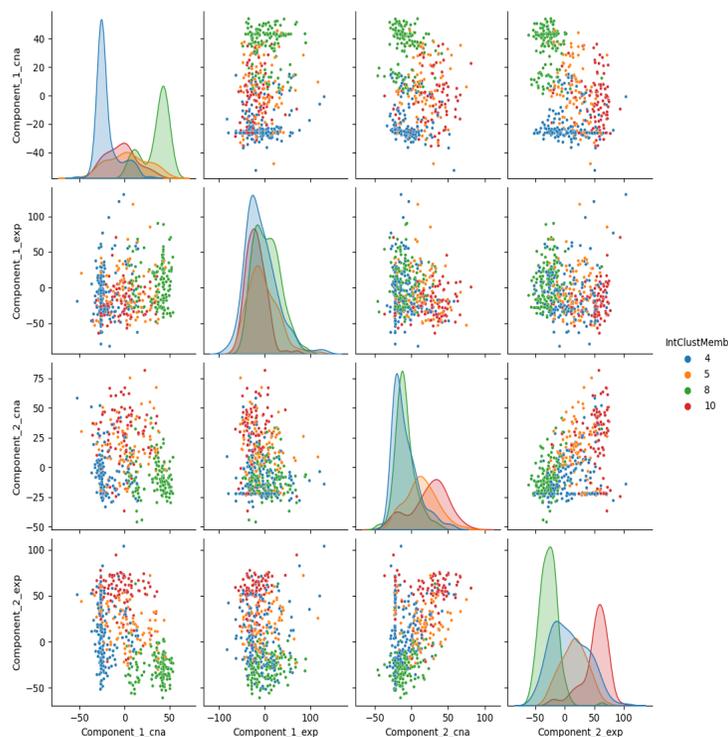


Dimensionality Reduction (4 PCs)



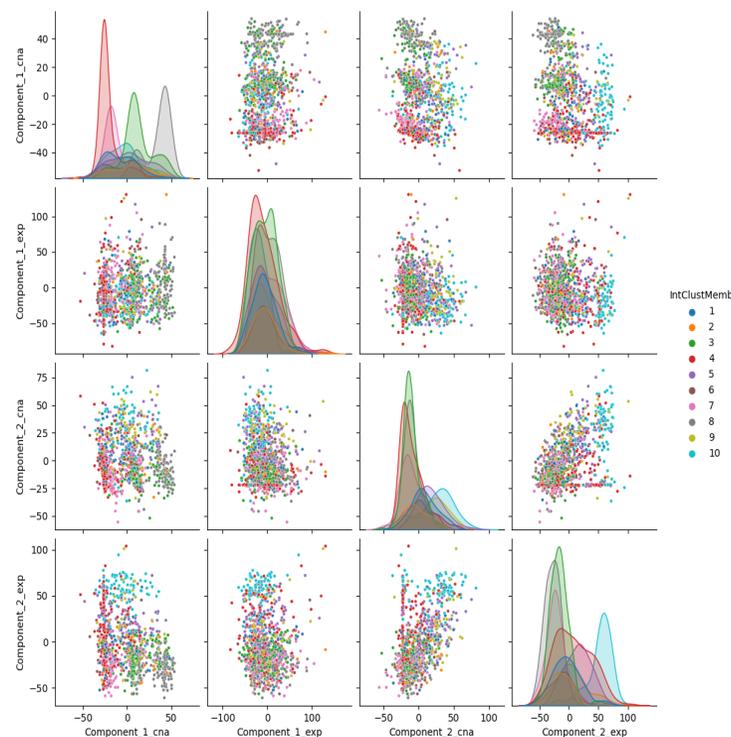
Caso Supervisionato

4 CLASSI



224 train + 152 test

10 CLASSI

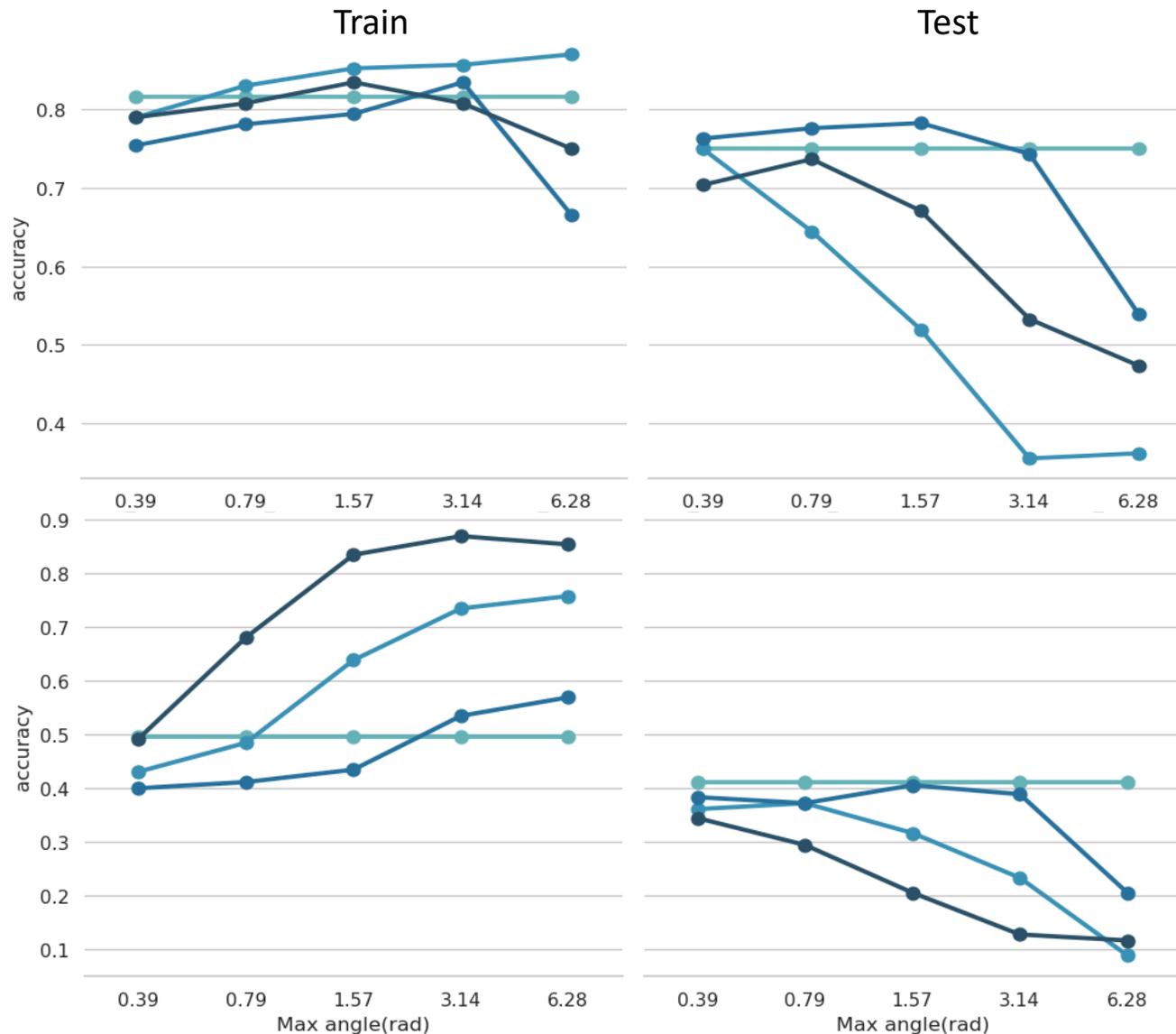


260 train + 180 test

- Due casi con **diversa difficoltà**
- **Quantum Support Vector Machine**
- Stratificazione di riferimento: **IntClust10**
- Bilanciamento dataset

Risultati

Kernel complessi limitano
l'apprendimento



4 classi

10 classi

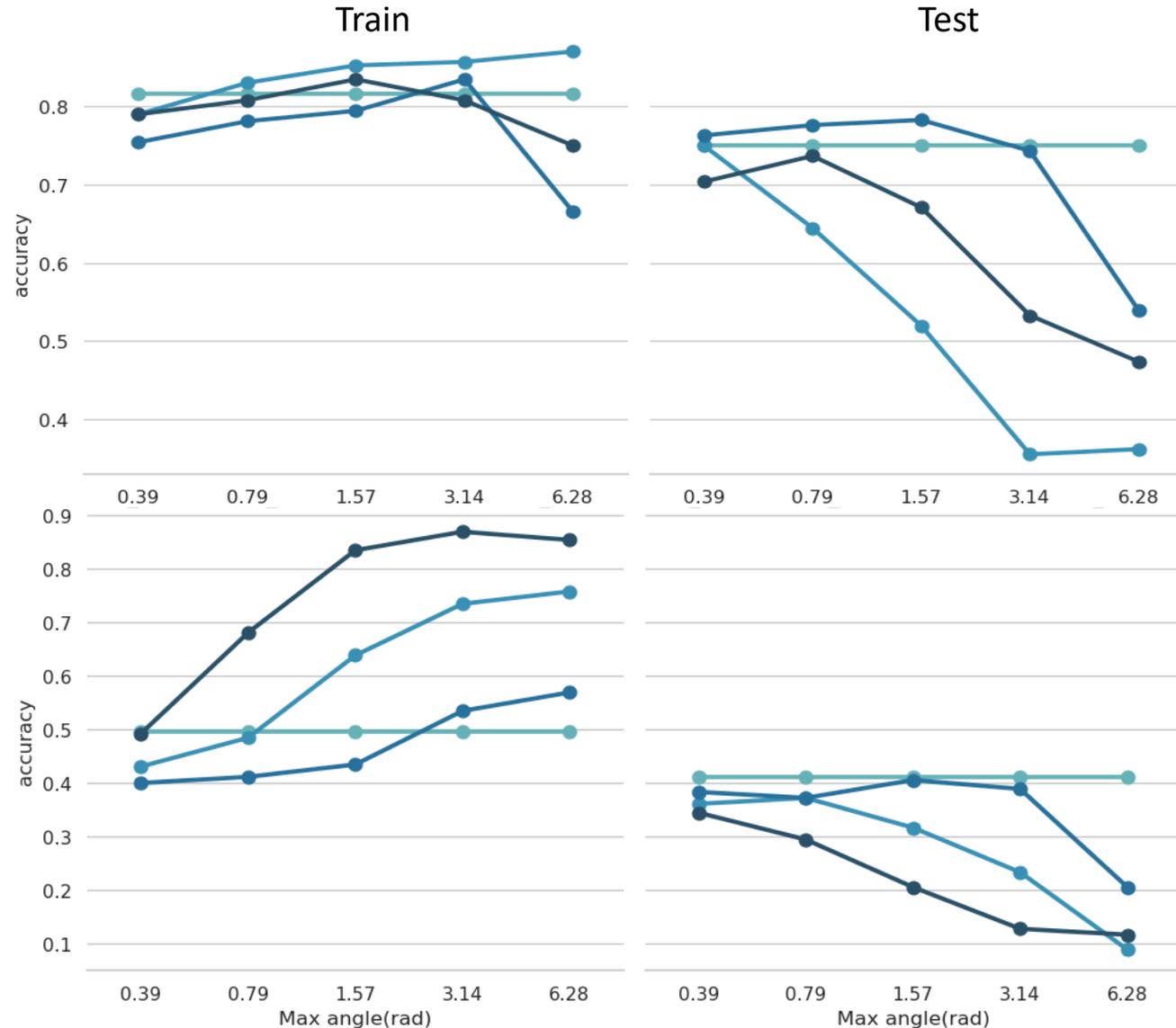
- ftmap
- rbf
 - ZZ_linear
 - Z_
 - ZZ_full

Risultati

Kernel complessi limitano
l'apprendimento



Complessità
modello
 \neq
Complessità dato



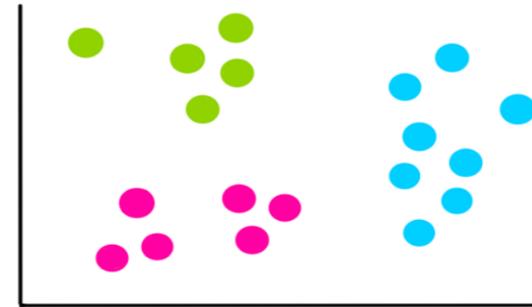
4 classi

10 classi

- ftmap
- rbf
 - ZZ_linear
 - Z_
 - ZZ_full

Caso Non supervisionato

- **Spectral Clustering** con kernel quantistici
- Test diverse numerosità del dato
- Analisi di sopravvivenza

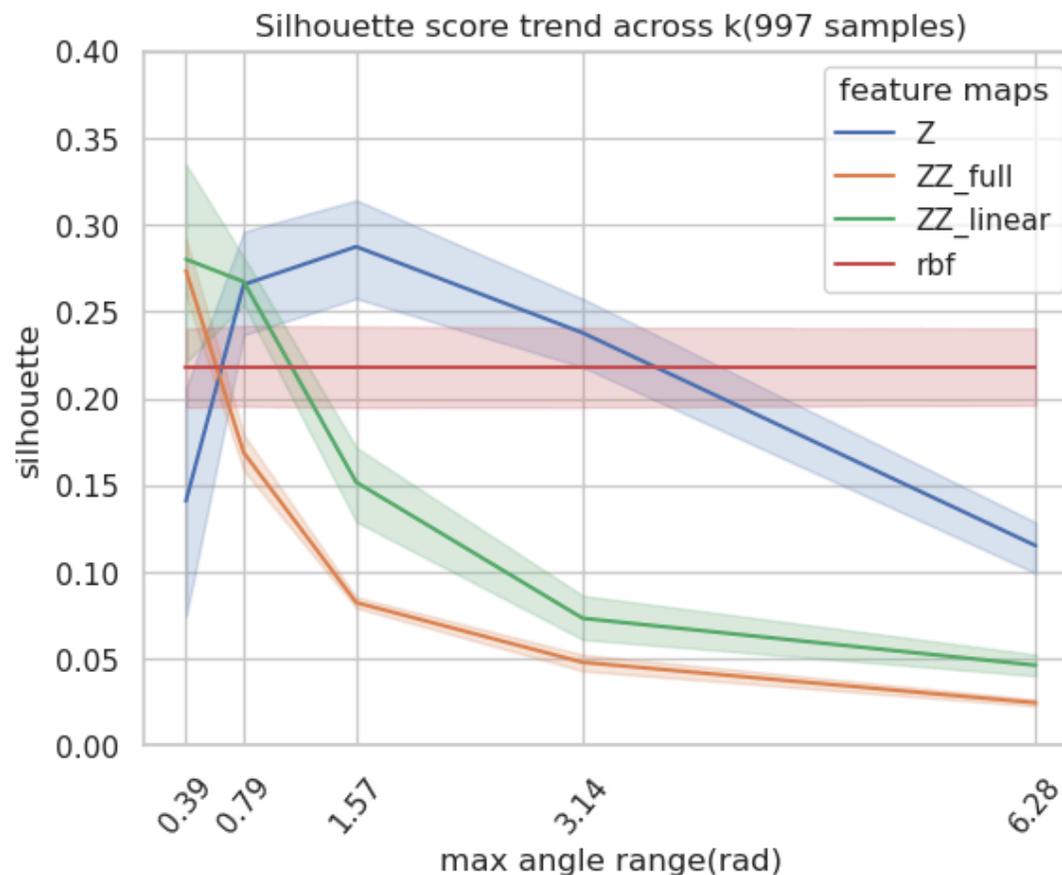
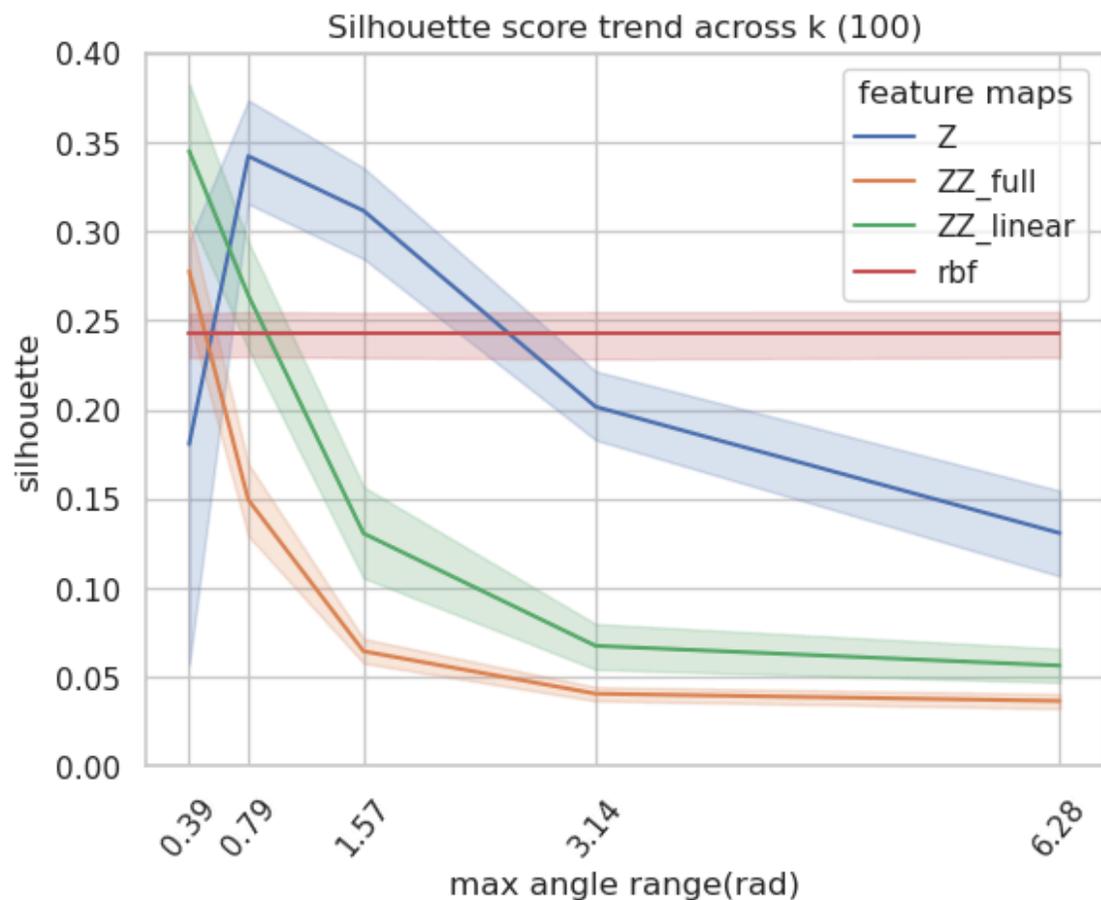


Clustering

Metrica:

Silhouette score $\rightarrow -1 \leq S \leq 1$

Trend stabile rispetto al campionamento



Conclusioni

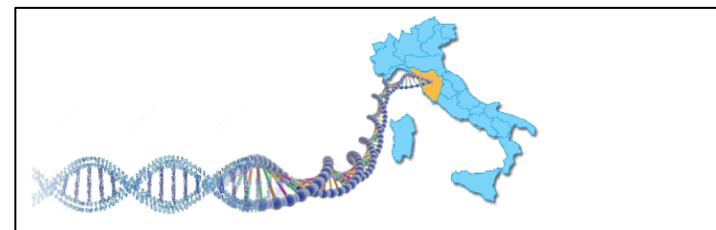
- L'alta **complessità** delle feature map ha un costo nella **capacità di generalizzazione**
- Alcuni embedding raggiungono **performance comparabili** agli approcci classici
- Nel clustering I risultati sono **indipendenti** dalla **numerosità del dato**
- **Stratificazione quantistica** è significativa in termini **di sopravvivenza**

Disseminazione dell'attività di ricerca



NUMerical computations: Theory and Algorithms (NUMTA)
The 4th international conference and summer school
Calabria, 14 – 20 giugno, 2023

BIOINFORMATIHA 10
X Ed. della Giornata Toscana di Bioinformatica e Systems Biology,
Siena, 7-8 settembre 2023



Attività di formazione



[Practical Quantum Computing School](#)
28 NOV-02 DEC 2022 , organizzata dal
CINECA

Qiskit Global Summer School 2023
17-28 Luglio 2023



Quantum Hackaton for Industrial Application
12-13 Settembre 2023, Bologna

Borse di studio GARR
Orio Carlini

**Grazie per
l'attenzione**