

Silvia Cascianelli

Research fellow

Adv: Marco Masseroli¹, Enzo Medico²



Computational Methods for Breast Cancer Stratification and Recurrence Risk Prediction

Relevant publications/submissions

- **Silvia Cascianelli**, Ivan Molineris, Claudio Isella, Marco Masseroli, Enzo Medico. *Machine learning for RNA sequencing-based intrinsic subtyping of breast cancer*. **Under review** @ Scientific Reports
- Francisco Cristovao, Arif Canakoglu, Mark Carman, **Silvia Cascianelli**, Luca Nanni, Pietro Pinoli, Marco Masseroli. *Comparing classic, deep and semi-supervised learning for whole-transcriptome breast cancer subtyping*. Presented at the 16th International Conference on Computational Intelligence methods for Bioinformatics and Biostatistics (CIBB2019).
- Lorenzo Perino, **Silvia Cascianelli**, Marco Masseroli. *Hybrid Evolutionary Framework for Selection of Genes Predicting Breast Cancer Relapse*. **Under review** @ IEEE World Congress on Computational Intelligence (WCCI) 2020, Special Session on Computational Intelligence for Bioinformatics and Computational Biology.

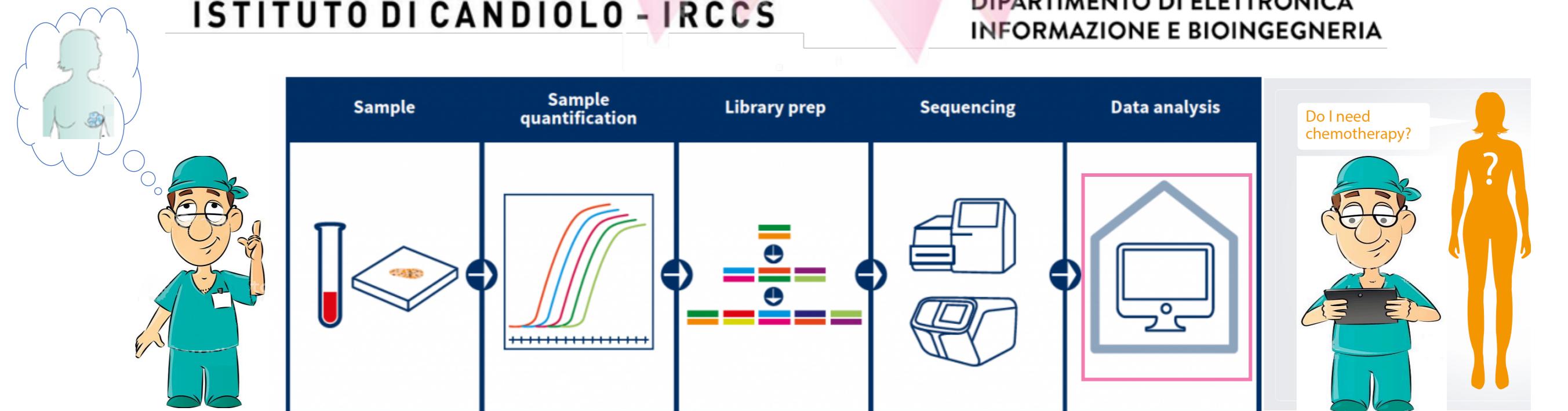
¹Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano; ²Candiolo Cancer Institute, FPO-IRCCS

JOINT PROJECT



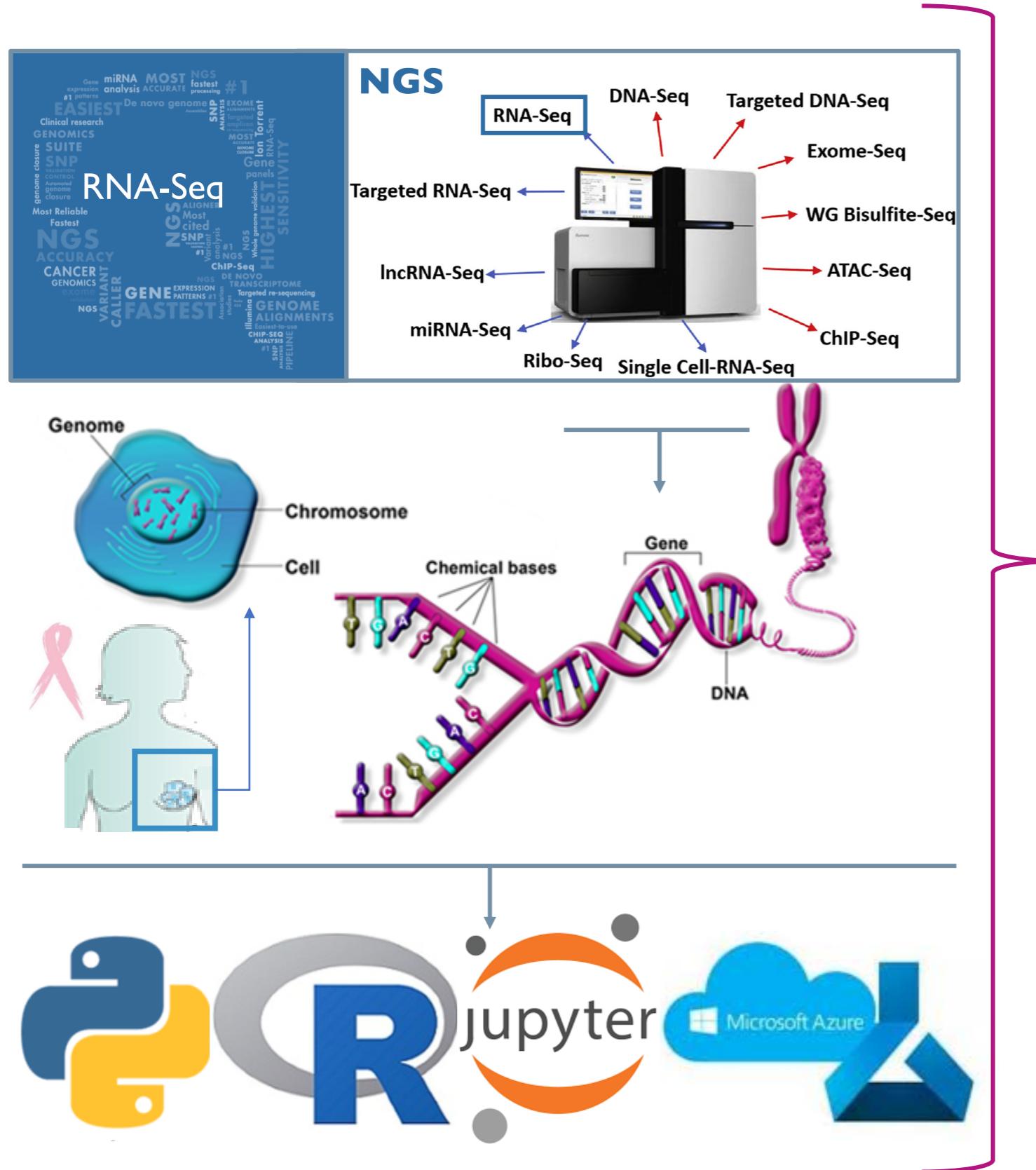
POLITECNICO
MILANO 1863

DIPARTIMENTO DI ELETTRONICA
INFORMAZIONE E BIOINGEGNERIA



INTRODUCTION OF RNA-SEQ TECHNOLOGY IN BREAST CANCER CLINICAL PRACTICE

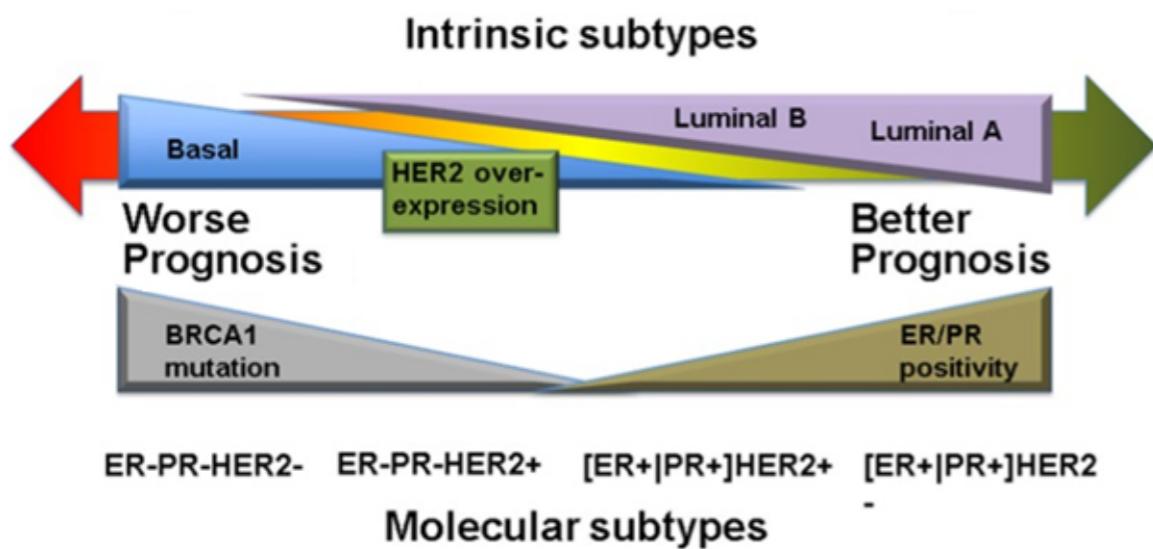
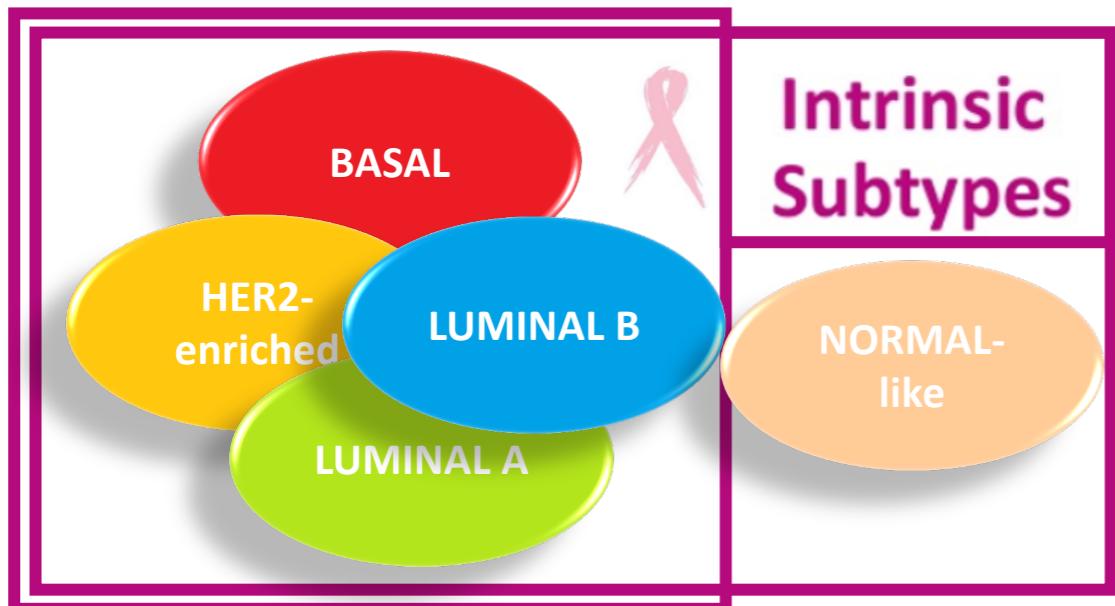
BREAST CANCER INVESTIGATION



Intrinsic subtypes
Recurrence risk



BREAST CANCER INTRINSIC MOLECULAR SUBTYPING

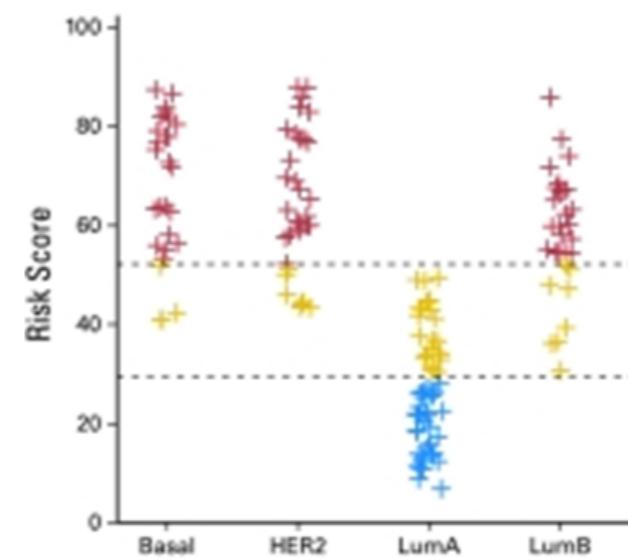


PAM50

Prediction Analysis of Microarray 50



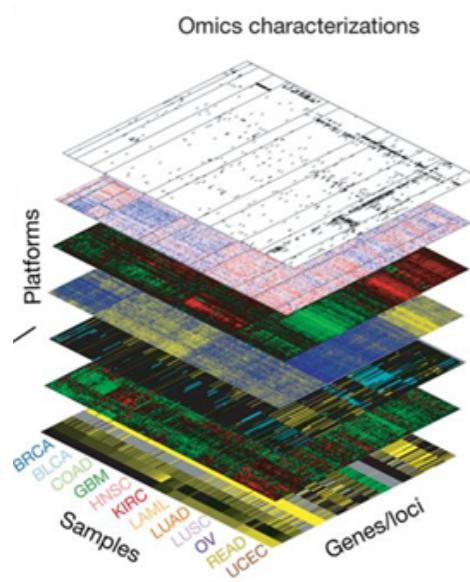
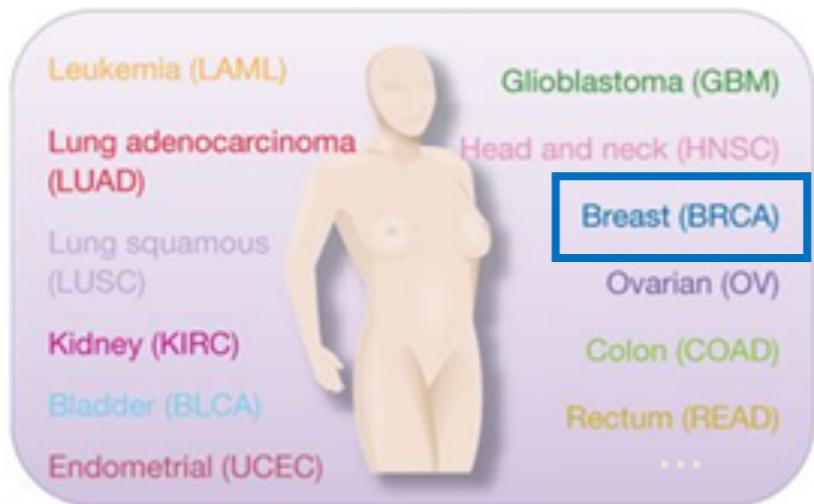
- Intrinsic subtypes are prognostically relevant as stand-alone indicators
- Also exploited in predictive models to estimate risk of distant recurrence



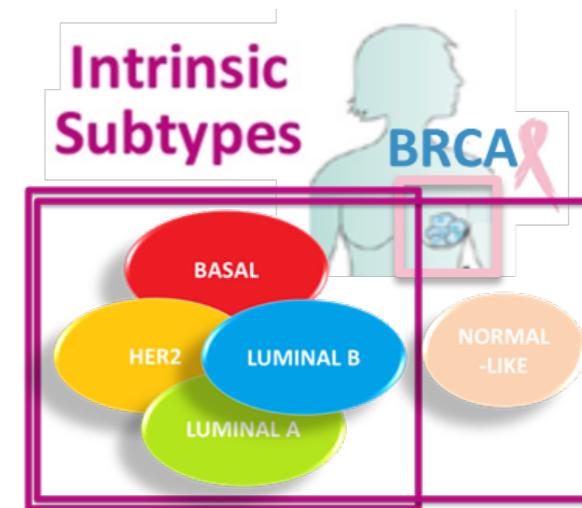
TCGA PUBLIC OMIC DATA FOR BREAST CANCER STRATIFICATION



TASK: Assign **breast cancer patients** to one of the prognosis-related “**intrinsic molecular subtypes**”



Mutation
Copy number
Gene expression
DNA methylation
MicroRNA
RPPA
Clinical data

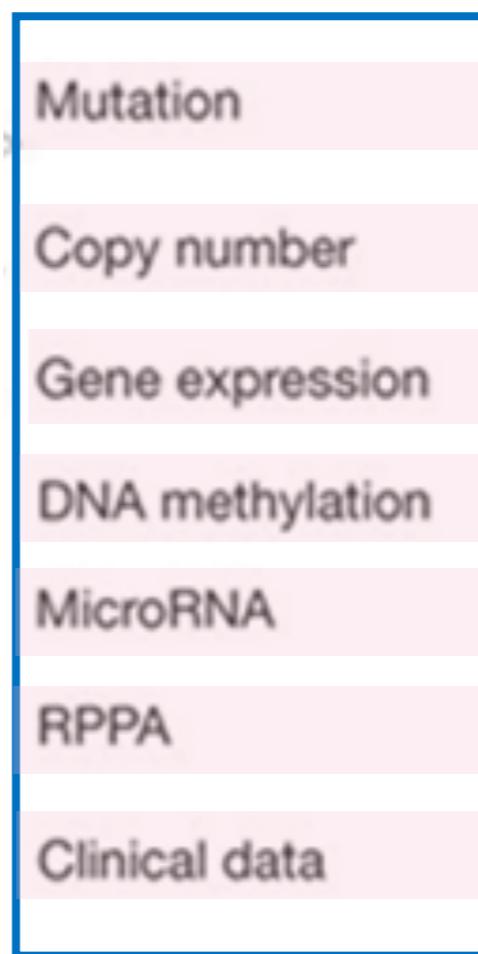
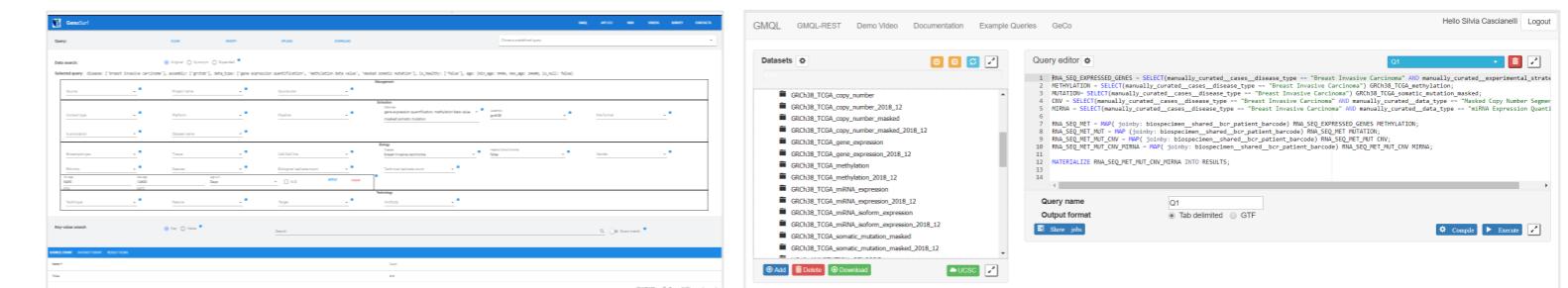
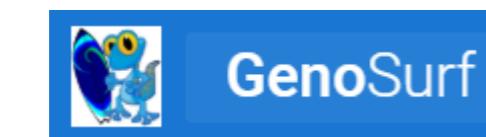


Computational methods:

The computational methods section lists several tools: jupyter, Microsoft Azure, R, Python, TensorFlow, and Keras. The text specifies:

- RNA-seq gene expression data as input
- Already published PAM50 intrinsic subtypes as labels/targets

TCGA PUBLIC OMIC DATA FOR BREAST CANCER STRATIFICATION



- TCGA multi-omic and clinical data are extracted from GMQL and GenoSurf
- Multi-omic profiles of BRCA patients are joined based on their TCGA patient ID
- Patients are labeled with the published subtype, if present
- All the collected data can be used to learn alternative ways of performing the **intrinsic suyping task**
- and **OTHER TASKS and DATA SOURCES** can be added!



COMPUTATIONAL METHODS



Breast cancer prognostic tests



mammaprint®

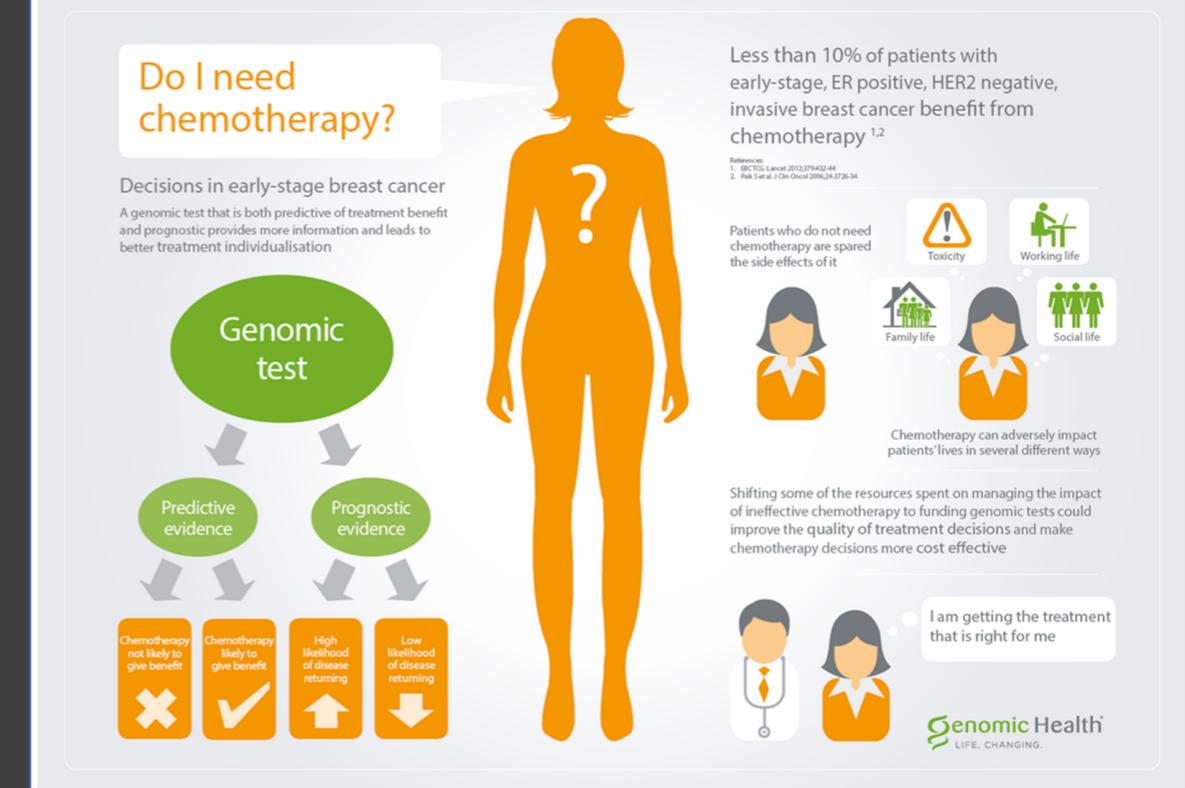
70-Gene Breast Cancer Recurrence Assay

prosigna®
Breast cancer
prognostic gene signature assay

PAM50

Prediction Analysis of
Microarray 50

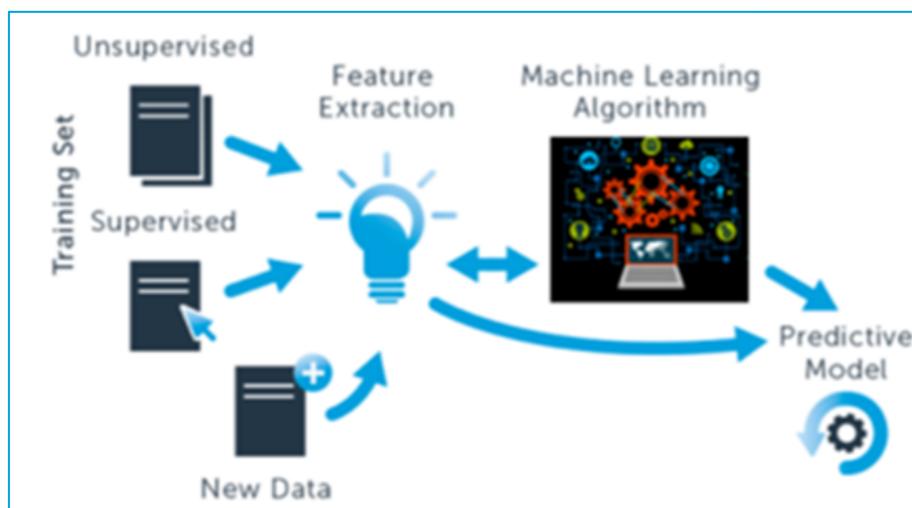
oncotype DX™
Breast Cancer Assay



COMPUTATIONAL METHODS



MACHINE LEARNING



Breast cancer

Intrinsic
subtyping

Recurrence
prediction

Data analysis :

PRINCIPAL COMPONENT ANALYSIS
DIFFERENTIAL EXPRESSION ANALYSIS
GENE SET ENRICHMENT ANALYSIS

Feature selection strategies:

EMBEDDED REGULARIZATIONS
FILTER-BASED METHODS
DEG-BASED METHODS
WRAPPER METHODS
EVOLUTIONARY ALGORITHMS

Alternative models:

UNSUPERVISED CLUSTERING
LINEAR MODELS
SUPPORT VECTOR MACHINES
ENSEMBLE METHODS
NON-PARAMETRIC MODELS
SEMI-SUPERVISED METHODS
DEEP APPROACHES

COMPUTATIONAL METHODS



NON-NEGATIVE MATRIX TRI- FACTORIZATION

$$R_{12\text{train}} = \begin{bmatrix} Pt\ 1 & Pt\ 2 & \dots & Pt\ n_2 \\ 0 & 0,5 & \dots & 0 \\ 1 & 0,5 & \dots & 0 \\ 0 & 0,5 & \dots & 0 \\ 0 & 0,5 & \dots & 1 \\ 0 & 0,5 & \dots & 0 \end{bmatrix}$$

Basal
Her2-e
LumA
LumB
Normal-I

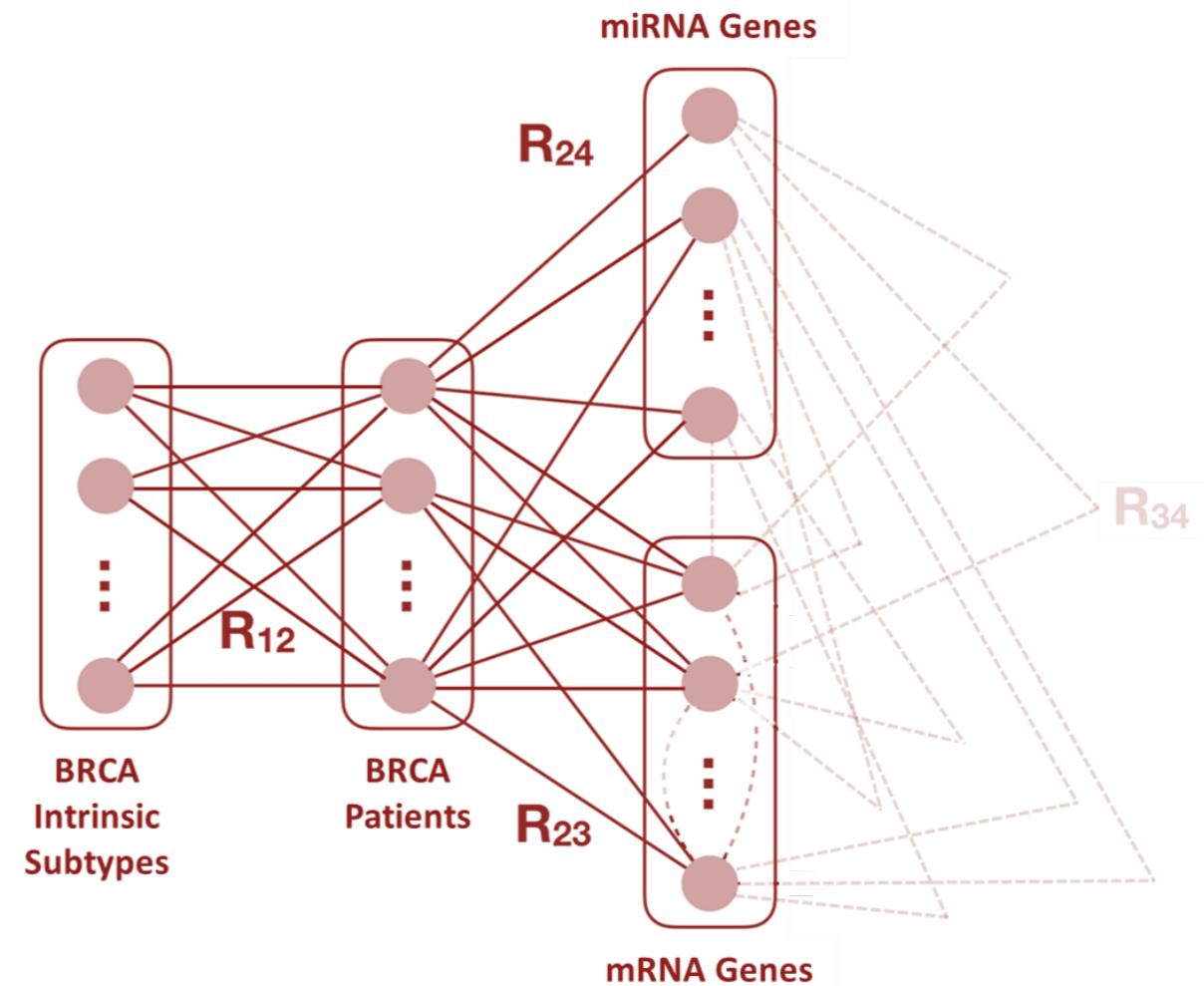
$$\tilde{R}_{12\text{train}}^{n_1 \times n_2} = G_1 * H_{12} * G_2^T$$



DATA-DRIVEN
GENOMIC
COMPUTING
(GECO)

Breast cancer
**Intrinsic
subtyping**

Multi-omic data integration



OPEN QUESTIONS

Is '***more***' really ***more*** to gain robustness, reliability and accuracy in subtyping or in clinical outcome predictions?



Which could be the most reasonable and useful way of
merging multiple approaches and omic data types?

1

