

# Integrating machine learning methods to single-cell experimental workflow increases throughput and accuracy for target identification in immuno-oncology

Marielle Péré<sup>1,2</sup>, Asma Chalabi<sup>1,4</sup>, Diego Oyarzun<sup>3</sup>, Madalena Chaves<sup>1</sup>, and Jérémie Roux<sup>1,2</sup>

<sup>1</sup>Université Côte d'Azur, Inria, INRAE, CNRS, Sorbonne Université, Biocore team, Sophia Antipolis, France

<sup>2</sup> Université Côte d'Azur, CNRS UMR 7284, Inserm U 1081, Institut de Recherche sur le Cancer et le Vieillessement de Nice, Centre Antoine Lacassagne, 06107 Nice, France

<sup>3</sup>School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, United Kingdom

<sup>4</sup>Université Côte d'Azur, Inria, Morpheme team, Sophia Antipolis, France

**Cell response heterogeneity** upon treatment is a main obstacle in preclinical development of efficacious cancer drugs, due to the emergence of drug-tolerant cells. We have previously developed a **single-cell workflow, *Fate-Seq* [2], to profile drug-tolerant persisters**, based on predictions of their drug response. To achieve this goal, *Fate-Seq* couples 3 single-cell techniques [1]: first the prediction of the cell response phenotype (resistant or sensitive) for clonal cancer cells treated with a chosen drug, then the separation of the predicted resistant cells from the predicted sensitive ones and finally the RNA sequencing of the cells.

To **automatize and increase the prediction throughput**, we present 3 major improvements in our workflow using **machine learning models to classify cell drug response and determine the molecular factors of non-genetic resistance to a drug**. These molecular factors represent good candidates to be **targeted during a co-treatment**, in combination with the first drug analyzed with our pipeline.

First, we present how we **combine image processes and machine learning classification models** to automatically **track cells overtime and detects important cellular events** like division or death. The output of this first technique are short and sparse fluorescent time-trajectories, that represents the transcriptomic activity in response to the drug, with a unique signal for each cell.

We then introduce our **eDRUGs (early Drug Response UpGraded) classifier**, that combines **mechanistic modeling of apoptosis** (cell death) and **machine learning classification models to predict cell drug response**, within an hour, for a maximum number of cells, using the fluorescent time-trajectories as input. This new method is twice as accurate as our previous prediction method [3].

Finally, we will also propose a **novel analysis method of sc-RNA-seq data** obtained with *Fate-Seq*. This method consists in training binary classifiers on the scRNAseq expression data obtained from the pipeline, using a range of models and **explainable AI techniques** such as DeepLift [4], in addition to clustering techniques, to obtain attribution scores for each gene. These scores are expected to reveal a reduced gene set, possibly containing only tens of genes, that are predictive of drug resistance.

**Key-words:** Single-cell, Machine learning models, Cancer, Time-series classification, Pharmacodynamics, therapeutic targets.

## References

- [1] Benjamin Bian, Agnès Paquet, Marie-Jeanne Arguel, Mickael Meyer, Ludovic Peyre, Asma Chababi, Marielle Péré, Kevin Lebrigand, Rainer Waldmann, Pascal Barbry, et al. Coupling live-cell imaging and in situ isolation of the same single cell to profile the transient states of predicted drug-tolerant cells. *STAR protocols*, 3(3):101600, 2022.
- [2] Mickael Meyer, Agnès Paquet, Marie-Jeanne Arguel, Ludovic Peyre, Luis C Gomes-Pereira, Kevin Lebrigand, Baharia Mograbi, Patrick Brest, Rainer Waldmann, Pascal Barbry, et al. Profiling the non-genetic origins of cancer drug resistance with a single-cell functional genomics approach using predictive cell dynamics. *Cell Systems*, 11(4):367–374, 2020.
- [3] Marielle Péré, Diego Oyarzun, Jérémie Roux, and Madalena Chaves. Modeling isogenic cancer cell response upon varying trail stimulations to decipher the kinetic determinants of cell fate decision. Elsevier, 2022.
- [4] Avanti Shrikumar, Peyton Greenside, and Anshul Kundaje. Learning important features through propagating activation differences. In *International conference on machine learning*, pages 3145–3153. PMLR, 2017.