Exploring Computational Oncology:
A Journey from evolutionary biology to Drug Discovery

Martha Serrano,
Principal Scientist
Center of Excellence, Computational Biology & Medicine
Oncology Research

Guest lecture in Applied Mathematics and Informatics in Drug Discovery
University of Basel

My background and current role

Martha Liliana Serrano Serrano

Journey from evolutionary biologist to clinical research and development

Colombia, Zapatoca

2016 — **PhD in Life Sciences, UNIL - SIB**

Computational biology in plant evolution

Nicolas Salamin

2011 — **Bachelor and Master in Biology in Colombia**

Domestication and genetic diversity of Lima beans in the Americas

Maria Isabel Chacón

Martha Liliana Serrano Serrano

Journey from evolutionary biologist to clinical research and development

Colombia, Zapatoca

2025 — **Principal Scientist in Roche**

Data & Analytics for preclinical and clinical project support (biomarker analysis)

Jan Attig, Nicolas Städler

2018 — **Senior Scientist in Roche**

Preclinical studies using bulk and single cell transcriptomics

Petra Schwalie, Iakov Davydov

2018 — **Postdoctoral fellow in Roche (RPF)**

Target discovery using CRISPR pooled genetic screens

Andreas Roller, Tobias Schmidt

*A key lesson: the principles of modeling, data analysis, and understanding complex, adaptive systems—which I learned in evolutionary biology—are the **exact same principles** I apply today to human cancer*

Pharma research and early development pRED

A global team of scientists dedicated to translating science into medicines

2,900

Employees around the world

Five innovation centres

Basel, Zurich, Munich, New York and Welwyn

Our pipeline consists of approximately **50% LMs[1], 35% SMs[2] and 15% RMs[3], ADCs[4] and GTs[5]**

pRED has delivered many **"firsts"** for Roche and with that created **new opportunities**

155 publications

in 2024 with at least one of the main authors being from pRED

Since 2014, **> 8,000,000 patients** have been treated with pRED molecules

We are uniquely positioned to apply **artificial intelligence and machine learning**

Multiple disease areas

Cardiovascular and Metabolism, Immunology, Infectious Diseases, Neuroscience, Oncology, Ophthalmology and Rare Diseases

96 patent applications

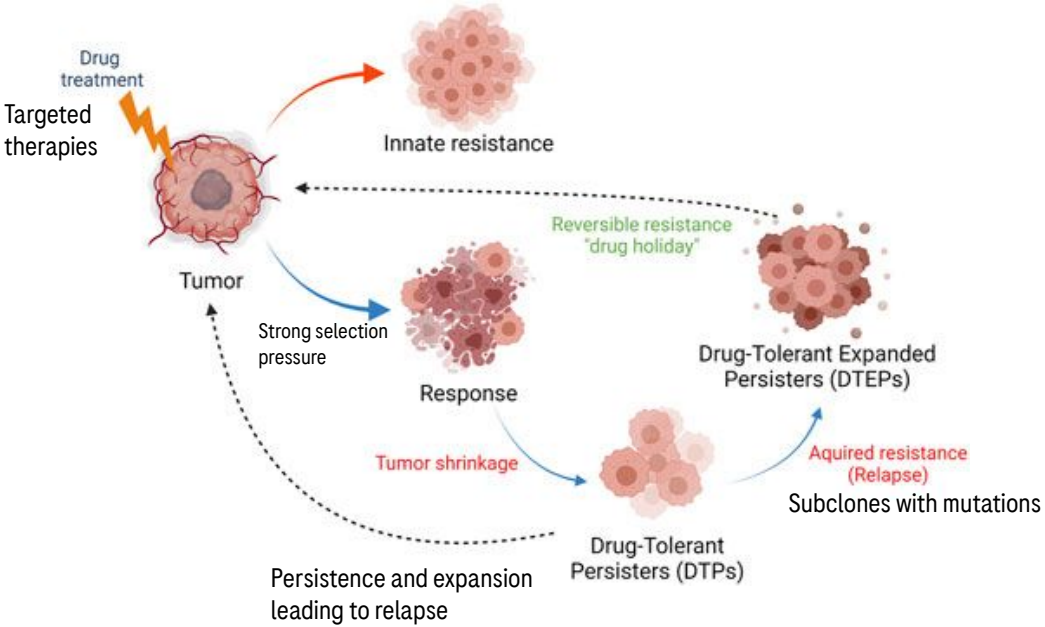
for new inventions in 2024

1: Large Molecules, 2: Small Molecules, 3: RNA Molecules, 4: Antibody-Drug Conjugates, 5: Gene Therapies

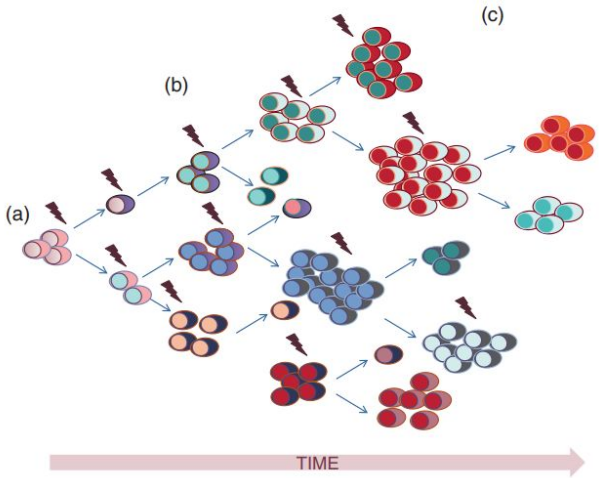
What drives me, what is the challenge?

My motivation and the grand challenge

The evolutionary arms race of cancer



Tumor heterogeneity (Vickers 2024)



Innate and acquired drug resistance (Kermi et al. 2022)

My motivation and the grand challenge

The evolutionary arms race of cancer

Innate and acquired drug resistance (Kermi et al. 2022)

How [computational] oncology drug discovery works?

What do we do at Roche?

How drug discovery works and what is my contribution… and how our skills are essential

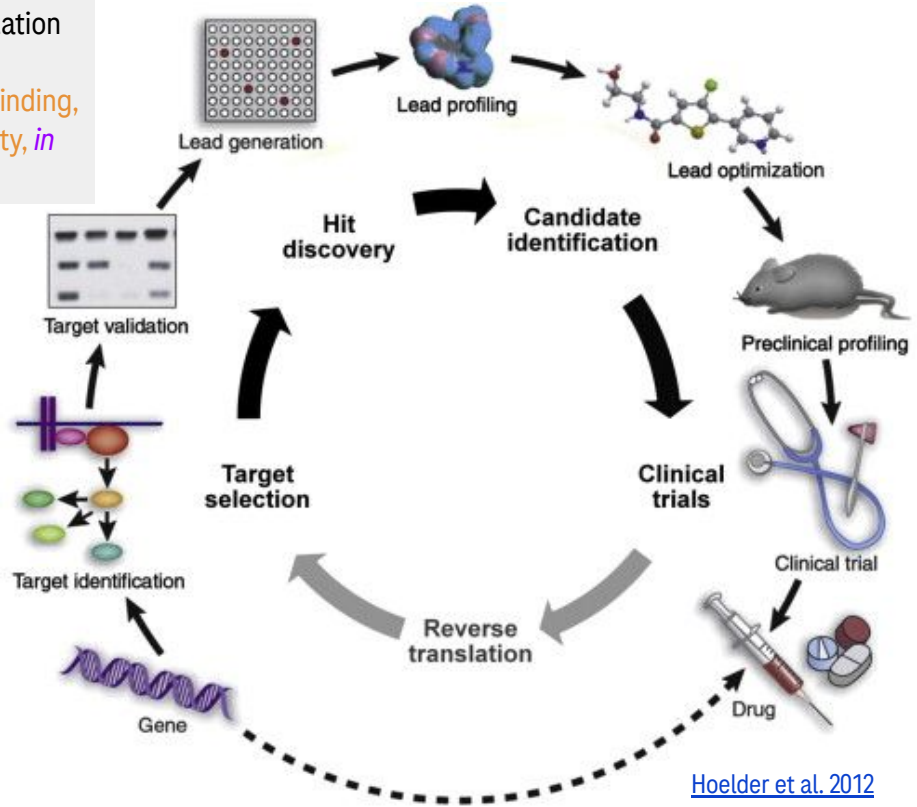
Natural Sciences

Mathematics/ chemoinformatics

Computer Sciences

Roche

Molecule production, characterization and performance:
Compound screening for protein binding, compound stability and selectivity, *in silico* binding prediction.



Preclinical models
Confidence on growth inhibition, target modulation, safety

Target Identification and validation
requires disease understanding, hypothesis generation and validation

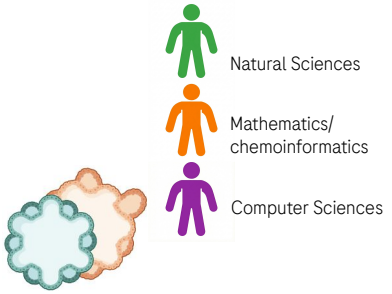
Does target contribute to cancer growth? Will inhibition lead to tumor regression or increased survival?

Clinical trial execution, patient monitoring, reverse translation
Hypothesis-driven trials
Understanding of resistance and finding predictive biomarkers

Hoelder et al. 2012

What can we do next?

The power of scale (data), modeling, and hypothesis generation

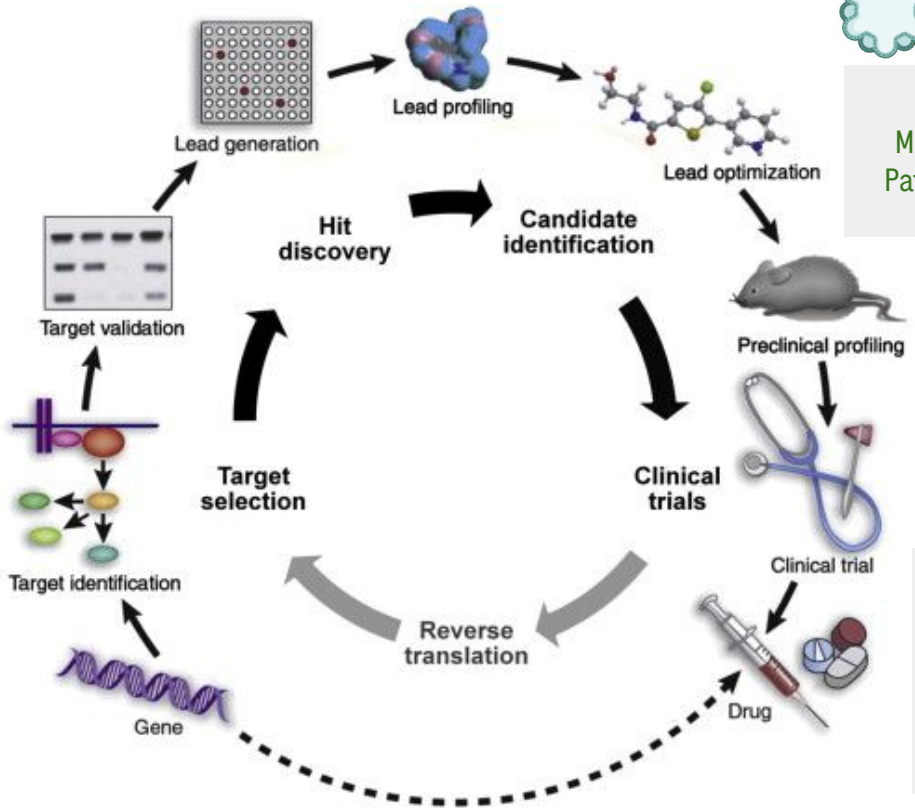


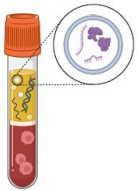
Roche

Preclinical models
More realistic 3D models
Patient-derived tumoroids

Target Identification and validation

Drug repurposing
Understanding disease states (scRNA)
In-silico perturbations



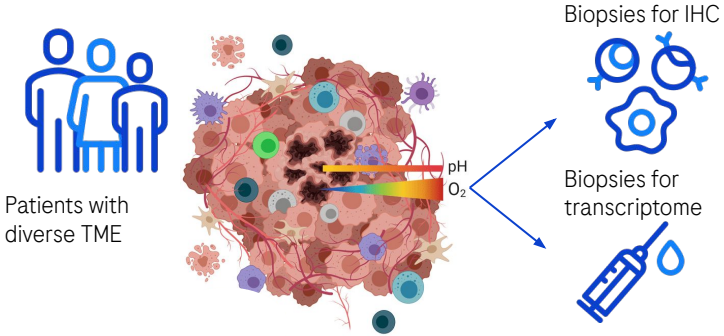


Clinical trials
Liquid biopsies and adaptive trails
Models for reverse translation
Digital pathology and FM for new biomarkers

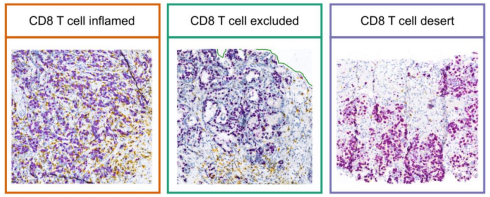
Innovation

Transcriptome-based classifier for TME and patients outcome

Roller *et al.* 2024



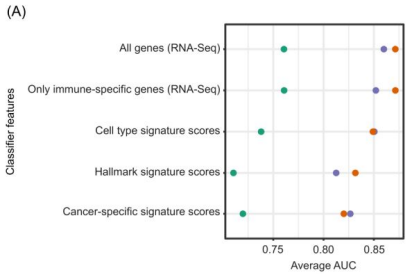
We developed a transcriptome-based classifier that could accurately **predict different spatial CD8+T cell infiltration** patterns in the TME

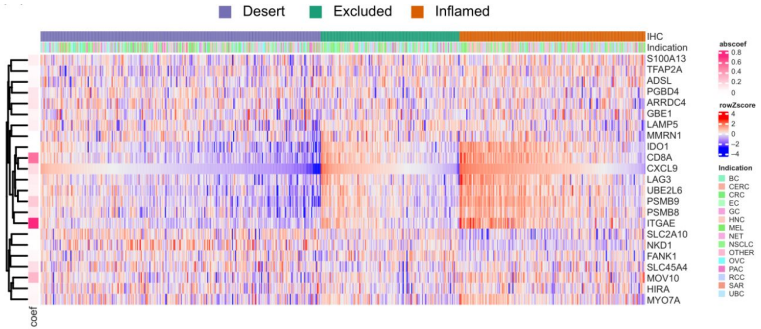


High diversity of immune cells and high infiltration

T cells are retained in the stroma

Limited intra-tumoral infiltration





Inflamed phenotypes -> immune cells, fibroblasts, endothelial and mast cells.
Desert phenotypes → stromal genes and immunosuppressive environment
Classifier has lower performance for **excluded** in which spatial relationships are more difficult to capture by bulk transcriptomics.

TME: Tumor micro environment
IHC: Immunohistochemistry staining

Lessons learned and tips

My key lessons learned

What are good skills?

Learn to speak multiple languages

Biology, code, maths, communication between collaborators in a project is key

High-performing algorithms don't matter if they are biologically implausible, or based on biased-data

Power of collaboration

Team sport, only the synergy of disciplines: computational prediction and experimental findings bring us success

Tips for our path

My favorite to go *Ten simple rules for doing a postdoc in pharma (Zhang 2021)*

Embrace the mess	Real world & experimental data is dirty, master data wrangling
80/20 rule	80% of our impact comes from 20% of our data, models and code. Focus on interpretability and robustness
Take a biology class	Focus on systems/molecular biology, genomics, built your biological intuition
Learn R and Python, code is our new pipette	Learn version control (Git), command-line environment, workflow engines
Get comfortable with probabilities	Computational biology needs understanding p-values, confidence intervals and null hypotheses

We need the rigor (math), the efficiency (CS), and the domain knowledge (Nat Sci) to solve the world's most critical health problems, only in that way we can build the future of medicine.

“"She stood in the storm, and when the wind did not blow her way, she adjusted her sails,"”

- Elizabeth Edwards

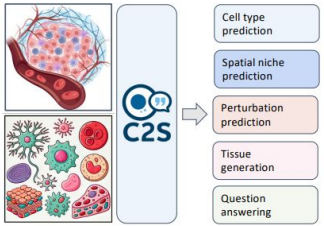
Doing now what patients need next

For what do we use LLM?

- [Scaling Large Language Models For Next-Generation Single-Cell Analysis](#) Van Dijk lab @ Yale

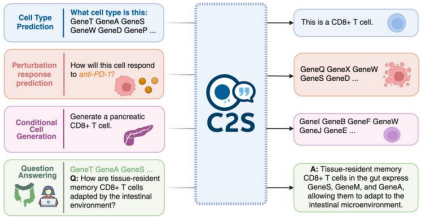
Vision: Teaching LLMs to Simulate the Virtual Cell

- Model **predicts** cellular responses under context
- **Integrates** multimodal data (genomics, text, spatial, drugs)
- Conditions on **microenvironment** (cytokines, neighbors, tissue)
- **Natural-language**e interface for reasoning & design



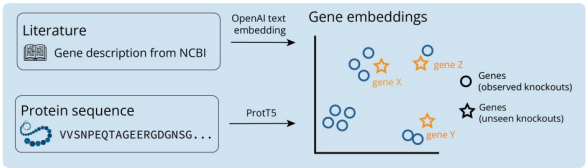
Cell2Sentence: LLMs that Learn Single-Cell Biology

- **Analogy:** Genes ≈ Words | Cells ≈ Sentences | Tissues ≈ Paragraphs.
- Cell2Sentence (***ICML*, 2024**): Training LLMs directly on massive single-cell datasets (>50M cells).
- **Capabilities:** Deep semantic understanding of cell types, states, gene programs. Enables "reading" and "writing" biology.



- LLM for unseen perturbation prediction [Kaspar Märtens](#)

LLM embedding approach



Our last year's work [1] showed that **LLM-derived gene embeddings**

- **Literature** embeddings, using NCBI gene descriptions [2]
- **Protein sequence** embeddings

are informative for unseen perturbation outcome prediction, and outperformed existing methods

[1] Märtens et al (2024), *Enhancing generative perturbation models with LLM-informed gene embeddings*

[2] Chen, & Zou (2024). *GenePT: a simple but effective foundation model for genes and cells built from ChatGPT*

Slides taken from a workshop, please do not distribute, see at the material published.

How to prepare for this career path?

Pointers to Pharma postdoc programs

- Take every opportunity to learn
- Always balance decisions on what makes the most impact (for your career and others) but mostly what makes you happy!
- *Ten simple rules for doing a postdoc in pharma*

"Change is constant": reorganization or reprioritization

"Collaboration, collaboration" … common goal

"Final reward or incentive" in industry requires extensive disciplines collaboration.

Expansion of your personal network (interdisciplinary)