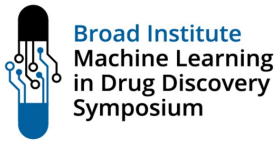
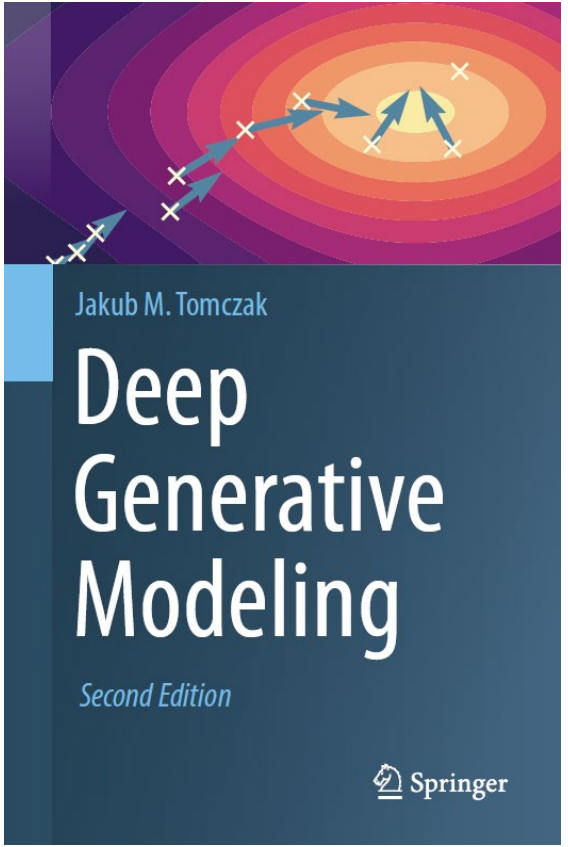
Accelerating drug discovery with Generative AI

Jakub M. Tomczak
Eindhoven University of Technology



Why GenAI in drug discovery?

GenAI = Generative Modeling with Deep Neural Networks



Drug discovery: R&D is about 4-7y

Research & Development

Research

1-3 years

Target identification & validation

Understanding of molecular and biochemical mechanisms

Identification of putatative causes of disease or conditions

Identification of drug candidates

Drug development

3-4 years

Lead identification & optimization
Compound screening
Optimization by analyzing discovered mechanisms

***In vitro* studies**
Experiments on isolated targets, cells and organoids.

***In vivo* studies**
Experiments on living organisms (animal models).

Trials & Approval

Clinical trials Review & Approval

4-7 years | **1-2 years**

Phase 1-3 trials
20-80 people
100-300 people
1,000-3,000 people

Evaluation & Approval
The regulating body reviews the evidence

Post-release monitoring

Drug discovery 2.0: The premise of GenAI is to speed up the process (and make it cheaper)

Research & Development

Research

~~1-3 years~~ 1 year

Target identification & validation

Understanding of molecular and biochemical mechanisms

Identification of putatative causes of disease or conditions

Identification of drug candidates

Drug development

~~3-4 years~~ 1-2 years

Lead identification & optimization
Compound screening
Optimization by analyzing discovered mechanisms

***In vitro* studies**
Experiments on isolated targets, cells and organoids.

***In vivo* studies**
Experiments on living organisms (animal models).

Trials & Approval

Clinical trials Review & Approval

4-7 years | 1-2 years

Phase 1-3 trials
20-80 people
100-300 people
1,000-3,000 people

Evaluation & Approval
The regulating body reviews the evidence

Post-release monitoring

For instance:
Open databases (e.g.: see CZI's data: link)
GenAI for understanding mechanisms

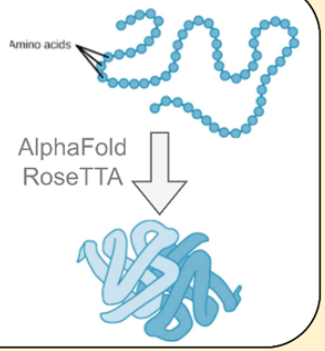
For instance:
GenAI for (*de novo*) drug design
GenAI for phenotypic profiling

What can we do with GenAI in drug discovery?

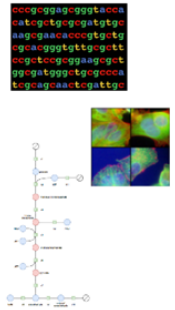
(Selected) Tasks that can be enhanced with GenAI

Target Aquisition

Structure Analysis



Data Analysis



Sequencing data

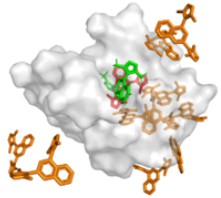
Imaging data

Network data

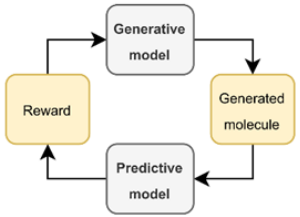
Simulation data

Lead/Hit identification

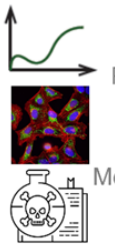
Molecular docking



De novo drug design



Properties



Pharmacokinetics

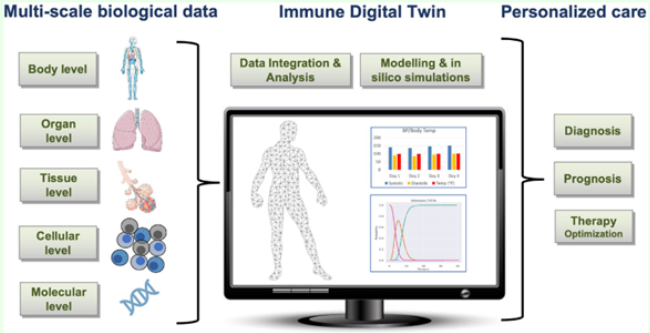
Pharmacodynamics

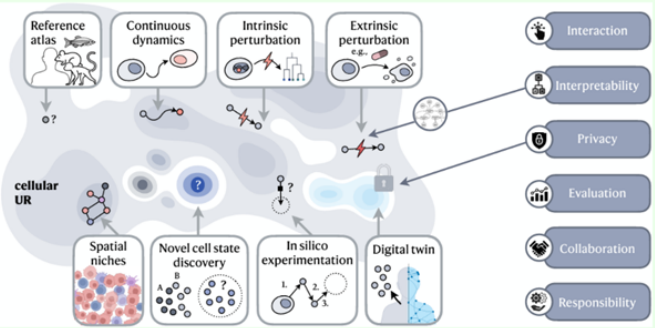
Enzymatics

Mechanism of action

Toxicology, etc.

Virtual Cells/Organisms

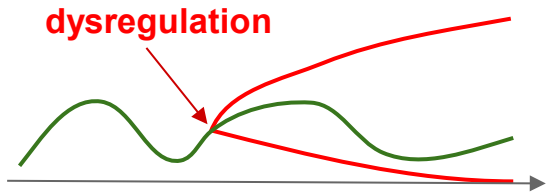




How can we use GenAI in drug discovery?

Regulators are **natural compounds** that **control** biochemical reactions.

A dysregulation results in a misbehavior of a biological system.



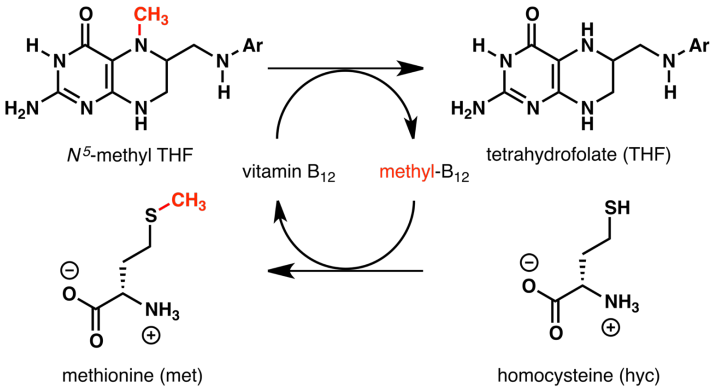
EXAMPLE: Vitamin B_{12}

In folate methionine cycle: Methionine synthase transfers the methyl group to the vitamin and then transfers the methyl group to homocysteine, converting that to methionine.

Vitamin B_{12} deficiency results in in an increased homocysteine level and the trapping of folate as 5-methyl-tetrahydrofolate, from which THF (the active form of folate) **cannot be recovered**.

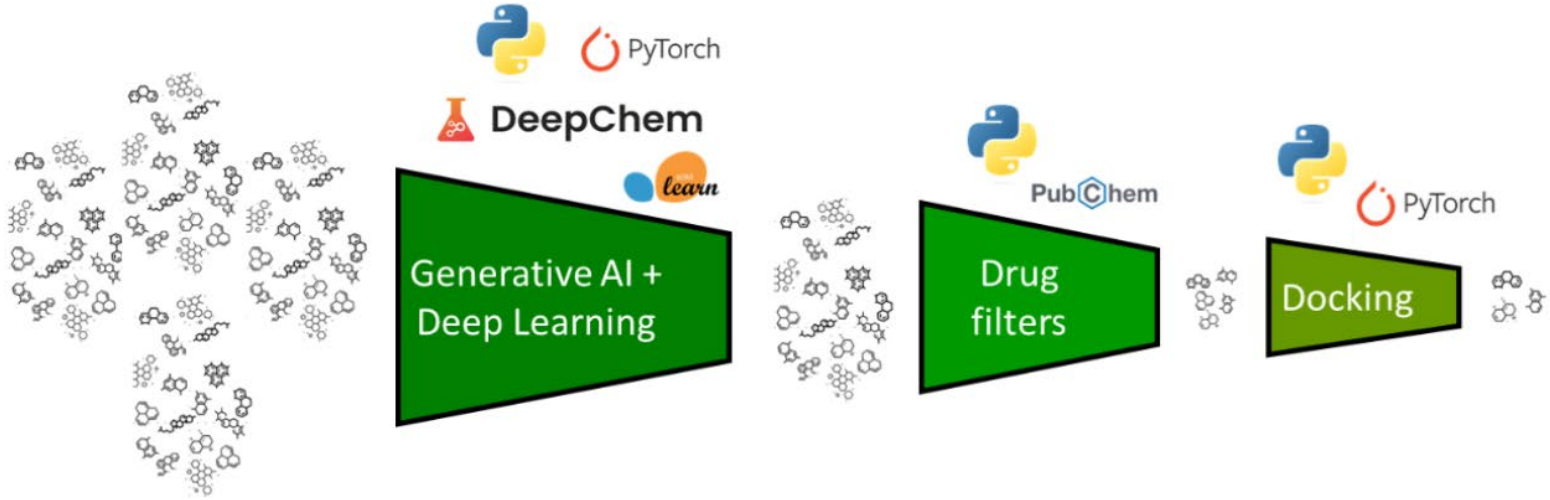
THF plays an important role in DNA synthesis.

As a result, vitamin B_{12} deficiency causes megaloblastic anemia.



GenAI for screening regulators of biochemical processes

NatInLab developed a GenAI-based in-house platform to screen natural regulators for a target of **Alzheimer's disease**.



>220k candidates

natural regulators

10k candidates

5k candidates

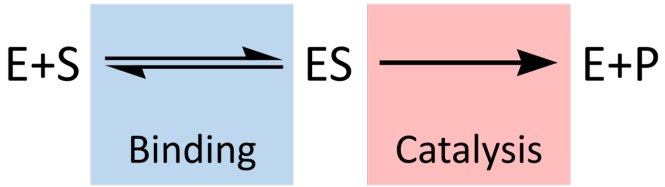
1.4% of the original candidates

Candidates for LEADS!

Enzyme kinetics: Do it fast and accurately!

Enzyme kinetics the discipline that studies

- how enzymatic reactions take place,
- the rate at which they occur,
- and the influence of environmental conditions in the reaction process.

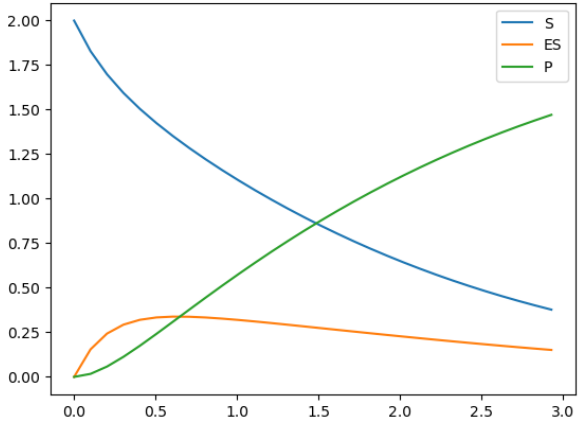


EXAMPLE

Michaelis-Menten model describes how the (initial) reaction rate depends on the position of the substrate-binding equilibrium and the rate constant:

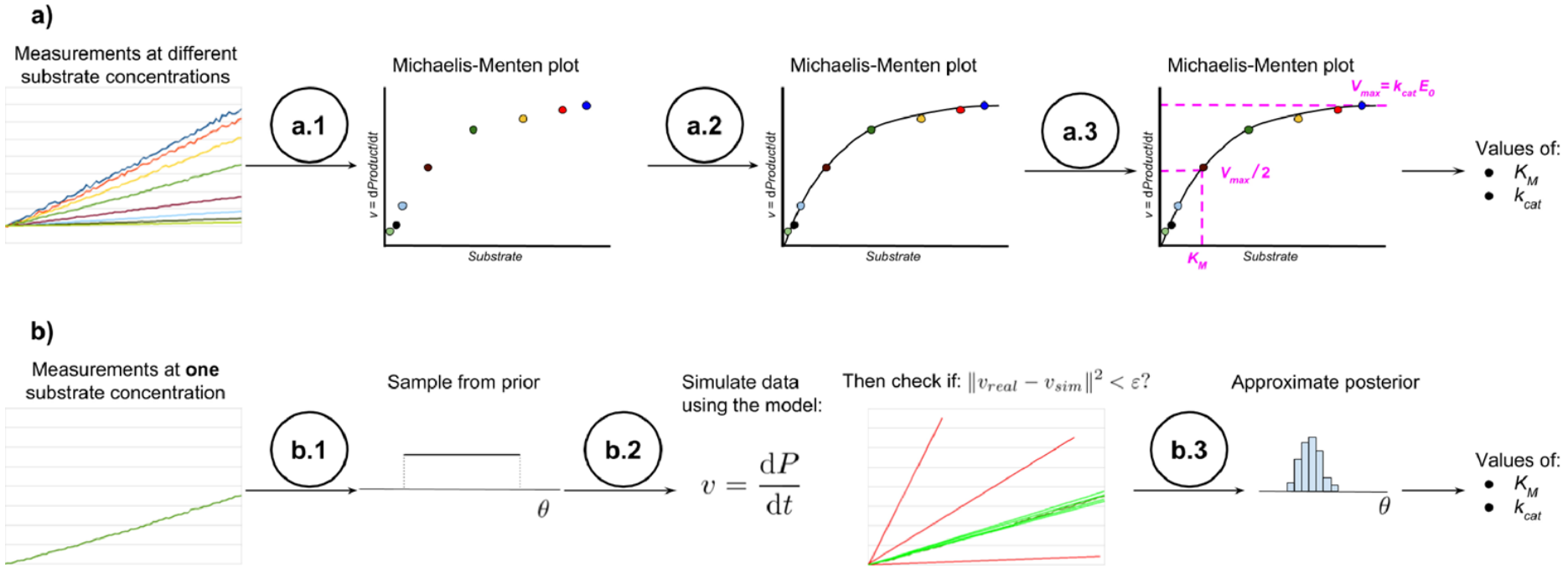
$$v_0 = \frac{V_{\max}[\mathrm{S}]}{K_M + [\mathrm{S}]}$$

$$V_{\max} \overset{\text{def}}{=} k_{cat}[\mathrm{E}]_{tot}$$



Q: How to calculate K_M and k_{cat} in an efficient way?

GenAI for enzyme kinetics: A local model



a. The *standard* approach using multiple measurements and the Michaelis-Menten plot.

b. Our proposed computational method: Use a single measurement and a simulator to identify parameters.

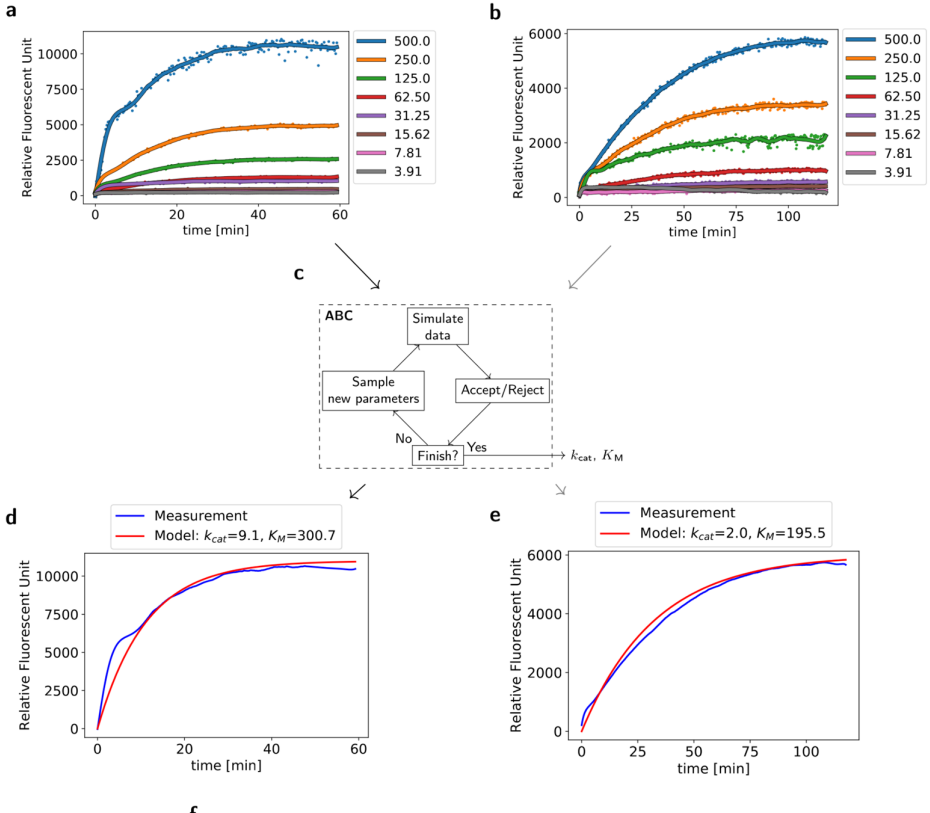
Tomczak, J. M., & Węglarz-Tomczak, E. (2019). Estimating kinetic constants in the Michaelis–Menten model from one enzymatic assay using Approximate Bayesian Computation. FEBS letters, 593(19), 2742-2750.

GenAI for enzyme kinetics: COVID-19

During **COVID-19**, we used a modified version of our previously proposed method to estimate the enzyme kinetics parameters.

It greatly helped us to speed up the process!

Our first findings on May 17, 2020 (**on bioRxiv ~2 months after first infections in the Netherlands**).



f

	k_{cat} $[s^{-1}]$	K_M $[\mu M]$	k_{cat}/K_M $[s^{-1}M^{-1}]$
PL^{pro}SARS	9.1 ± 0.5	300.7 ± 20.2	**0.030** ± 0.003
PL^{pro}CoV2	2.0 ± 0.3	195.5 ± 5.2	**0.010** ± 0.001

Weglarz-Tomczak, E., Tomczak, J. M., Talma, M., Burda-Grabowska, M., Giurg, M., & Brul, S. (2021). Identification of ebselen and its analogues as potent covalent inhibitors of papain-like protease from SARS-CoV-2. Scientific reports, 11(1), 3640.

GenAI for enzyme kinetics: A global model

Q: Is it possible to learn an AI model that mimics enzyme kinetics?

Given:

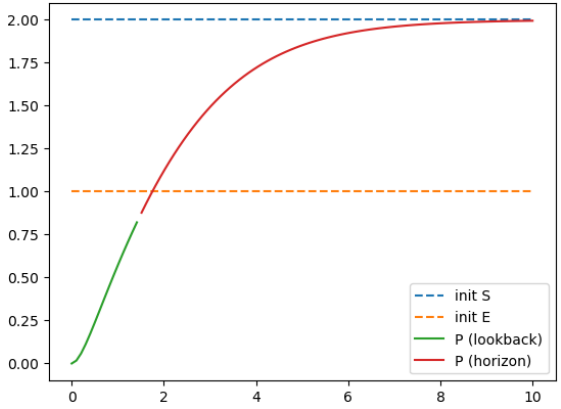
- exogenous: the initial concentrations of S and E
- (scenario 1) a few seconds of measurements of P
- (scenario 2) only the initial concentration of P

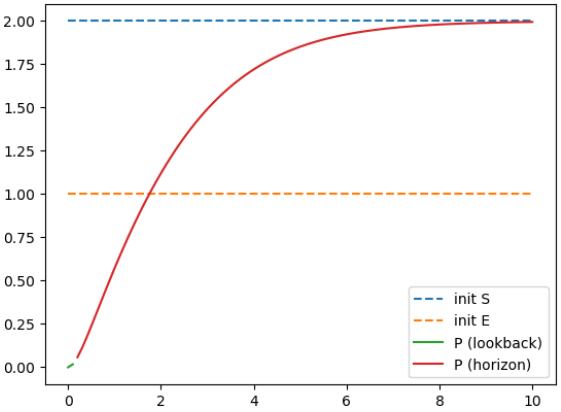
GOAL: Generate the remaining of P (horizon)

Baseline: TiDE with and without ex (Das et al., 2023)

Our approach: A non-linear extension of TiDE with ex

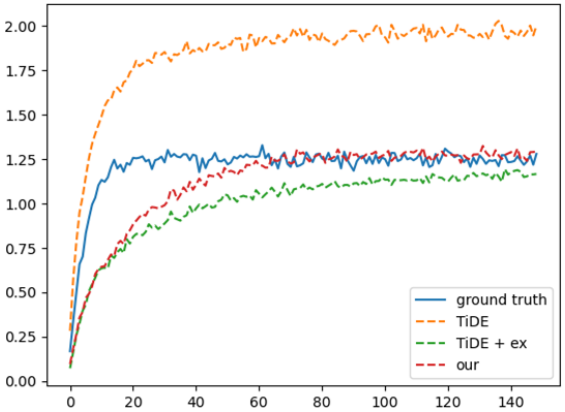
Preliminary work (unpublished!)

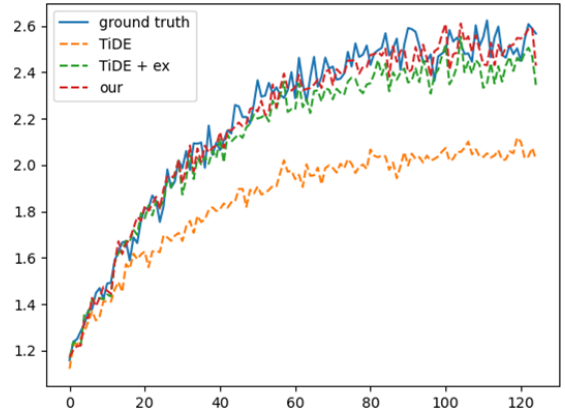




GenAI for enzyme kinetics: A global model

Approach	Scenario 1	Scenario 2
TiDE	2.607±0.103	46.392±0.13
TiDE + ex	0.601±0.024	9.999±0.519
Our + ex	**0.494**±0.017	**9.305**±0.289





Take-aways!

First: Using exogenous information is crucial (as expected.

Second: Using exogenous allows generating a signal for given initial conditions pretty well!

Third: Our approach gives a slight boost!

Fourth: Very promising results, more to come!

Phenotypic profiling: Getting insights into Mechanism of Action (MOA) of drugs

Image-based **phenotypic profiling** of small molecules can be used for:

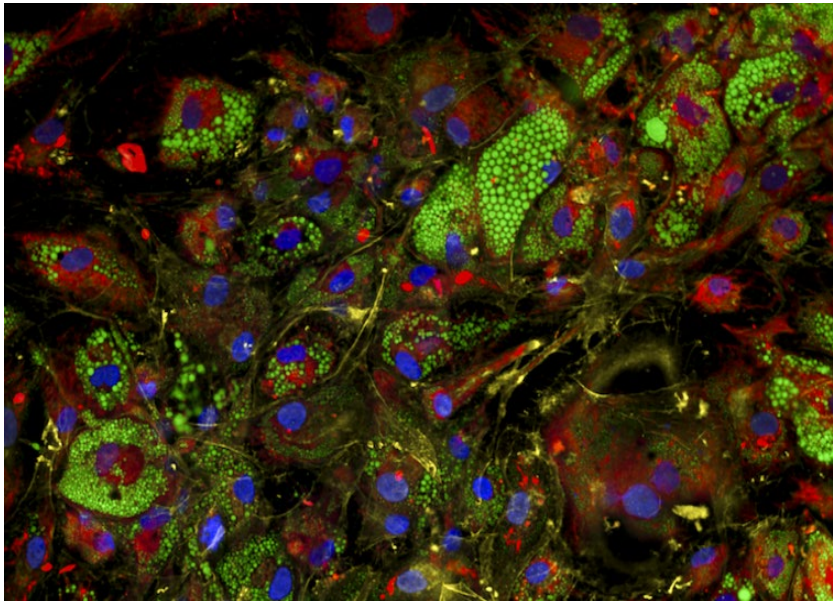
- identification and characterisation of small molecules in drug discovery
- Getting important insights into their mechanisms of action (MOA).

EXAMPLE: BBBC021

We used the **BBBC021** dataset containing microscopy images of MCF7 breast cancer cell lines treated with 113 compounds for 24 hours.

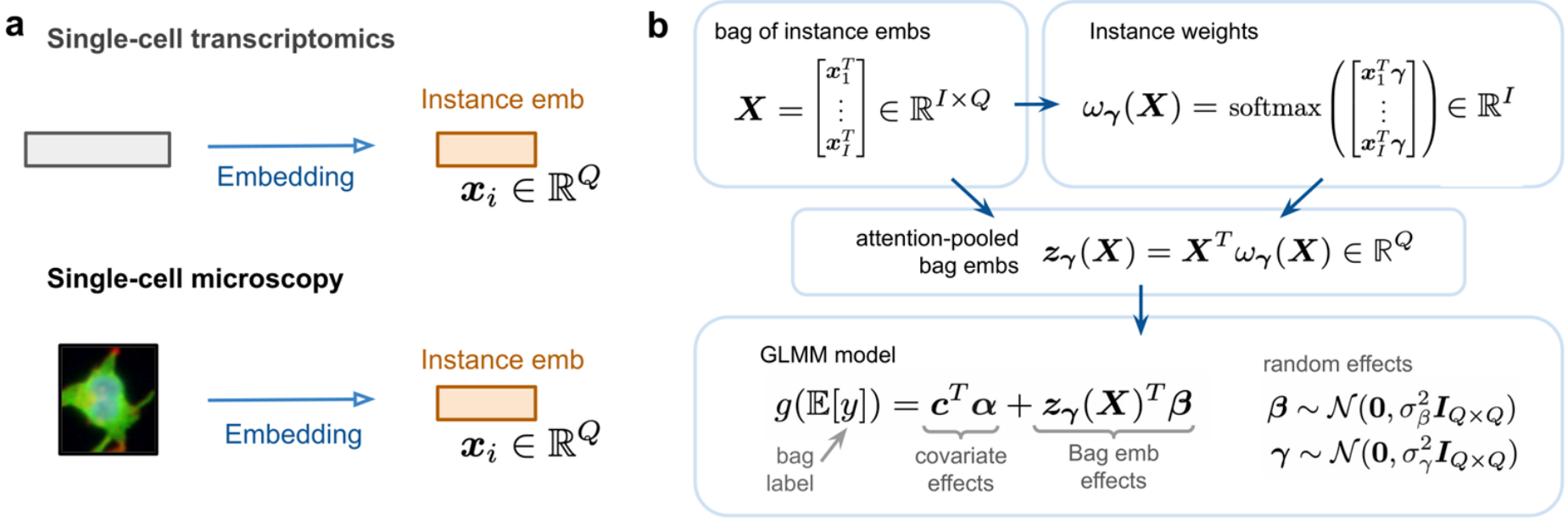
We focus on 39 compounds with a visible impact on cell morphology, which was associated with 12 distinct MoA labels

Eventually, we got 2,526 wells (bags), 133,628 cells (total number of instances), and 12 MoAs (labels).



https://www.broadinstitute.org/news/lipocyte-profiler-metabolic-biology-tool

MixMIL: A probabilistic model with attention mechanism



a. MixMIL uses predefined instance embeddings from domain-specific unsupervised models.

b. Generalized multi-instance mixed model framework defining MixMIL.

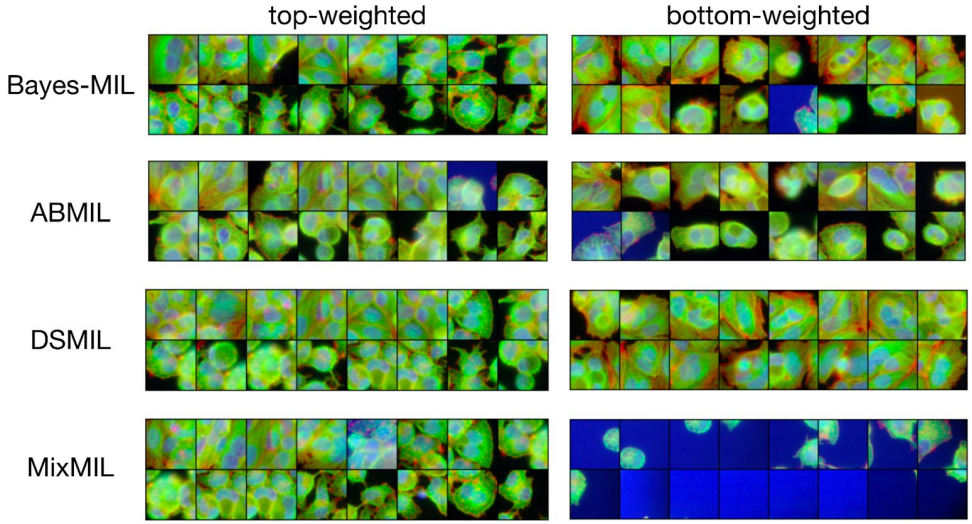
Engelmann, J. P., Palma, A., Tomczak, J. M., Theis, F., & Casale, F. P. (2024, April). Mixed Models with Multiple Instance Learning. In International Conference on Artificial Intelligence and Statistics (pp. 3664-3672). PMLR. (**oral**)

MixMIL for Mechanism of Action Prediction

Method	Bal. Accuracy	F1 Macro	F1 Micro
Bayes-MIL	0.63 ± 0.02	0.63 ± 0.02	0.70 ± 0.01
ABMIL	0.72 ± 0.02	0.73 ± 0.01	0.76 ± 0.01
Gated ABMIL	0.67 ± 0.03	0.65 ± 0.03	0.70 ± 0.03
Additive ABMIL	0.41 ± 0.00	0.34 ± 0.00	0.47 ± 0.02
DSMIL	0.89 ± 0.02	0.89 ± 0.02	0.90 ± 0.01
MixMIL	$\mathbf{0.94 \pm 0.02}$	$\mathbf{0.94 \pm 0.01}$	$\mathbf{0.95 \pm 0.01}$

Our approach achieves SOTA results on the multi-label classification problem!

94% of images are properly assigned to a MOA!



Additionally, our approach properly identifies less important images by assigning them low attention weight.

Engelmann, J. P., Palma, A., Tomczak, J. M., Theis, F., & Casale, F. P. (2024, April). Mixed Models with Multiple Instance Learning. In International Conference on Artificial Intelligence and Statistics (pp. 3664-3672). PMLR.

GenAI for drug discovery: Conclusion

Conclusion

GenAI offers more than LLMs

GenAI can (should!) be used for **computational chemistry** and **drug discovery**

GenAI can **drastically speed up the R&D process**

GenAI beyond tasks like generating drugs (drug design), molecular docking, 3D structure generation

GenAI can be useful in:

- understanding biochemical mechanisms,
- pharmacokinetics/dynamics,
- mechanism of action,
- enzyme kinetics,
- and many more!

Future: GenAI for **virtual cells/organisms**

Thank you!
Questions?

Jakub M. Tomczak, Ph.D.
Group Leader, Associate Professor
Eindhoven University of Technology
Director/Founder Amsterdam AI Solutions

jmk.tomczak@gmail.com

https://jmtomczak.github.io/

Mahdi Mehmanchi

Haotian Chan

Ewelina Weglarz-Tomczak, Ph.D.

Jan Engelmann

Alessandro Palma

F. Paolo Casale, Ph.D.

Generativ/e

NatInLab

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

HELMHOLTZ MUNICH

Amsterdam AI Solutions