Toward intelligent pharmacology and toxicology assessment in drug discovery and early development

Jitao David Zhang on behalf of Predictive Modeling and Data Analytics (PMDA) Chapter, Pharmaceutical Sciences, Pharma Research and Early Development, Roche Innovation Center Basel, F. Hoffmann-La Roche

The Annual SSPT Spring Meeting 2024

At Roche, we combine human and machine intelligence to discover new drugs and bring them to patients

Pittigrilli, CC BY-SA 4.0, via Wikimedia Commons

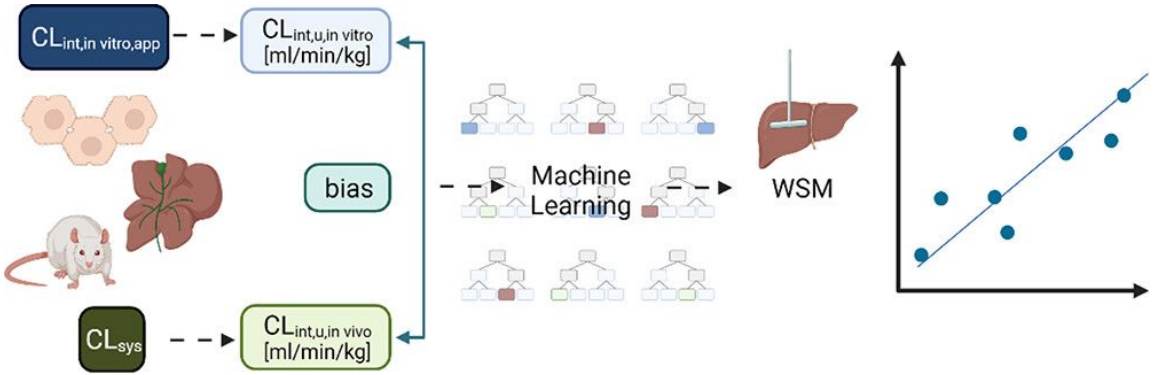
ML/AI as Inspiration
for decision and actions

ML/AI as Amplifier
for information integration

ML/AI as Microscopy
for biology-drug interaction

We predict pharmacokinetics (PK) parameters using and combining PBPK and machine-learning approaches

- With *SwiftPK*, we develop and implement a high-throughput physiologically-based pharmacokinetics modelling (HT-PBPK) framework to inform candidate selection in early discovery.
- We develop ML models for direct and indirect prediction of preclinical and clinical PK parameters.



We are exploring ways to combine HT-PBPK and ML approaches in order to

- ➢ Recognise liabilities, understand their causes, and optimise PK profiles;
- ➢ Reduce the need for rodent PK studies;
- ➢ Reduce cycle times.

Andrea Andrews-Morger, Davide Bassani, Neil Parrott, Andrés Olivares, Michael Reutlinger, Yaniv Cohen, Jitao David Zhang, Lisa Sach-Peltason, Leonid Komissarov, Nenad Manevski, Florian Spinelli, Torsten Schindler, the *SwiftPK* Team. See publications (1) Andrews-Morger *et al.*: A Machine Learning Framework to Improve Rat Clearance Predictions and Inform Physiologically Based Pharmacokinetic Modeling, and (2) Stoyanova et al.: Computational Predictions of Nonclinical pharmacokinetics at the Drug Design Stage.

We predict phototoxicity potential of drug candidates by combining Quantum Mechanics and Machine Learning

Context

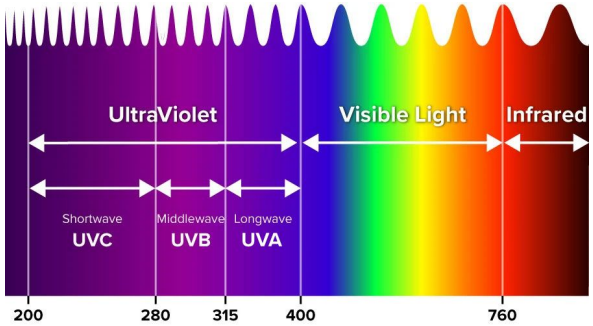
Phototoxic compounds cause skin toxicity when exposed to UV light. Existing *In vitro* models have limited hightroughtput. An *in silico* predictor is available; yet we see potential to improve.

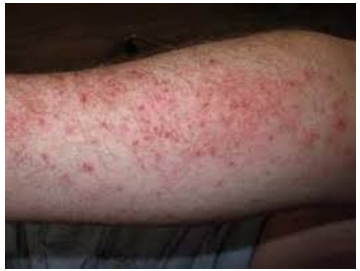
Method

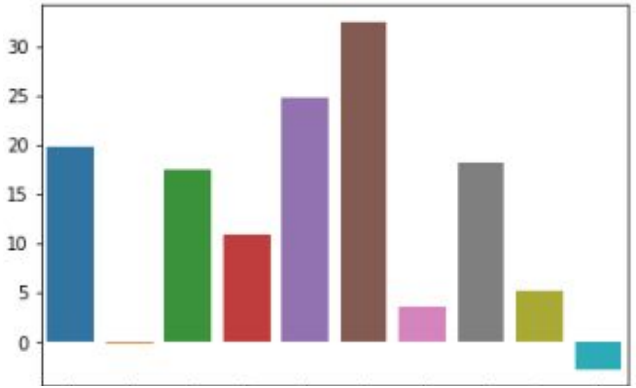
We combined Quantum Mechanics (conformer minimization, UV-vis spectra prediction) and Machine Learning (fingerprints, phys-chem properties, and UV-vis as input, an ensemble model, probability/binary output).

Outcome & Impact

The QM+ML model achieves comparable performance of with an experimental approach, and fares 10%-30% better than the existing model. It supports ongoing projects.







Percentage of improvement of prediction accuracy (y-axis), stratified by splits (x-axis).

Nicolò Milani, Andrea Anelli, Michael Reutlinger

We evaluated Federated Learning for *in vitro* toxicity assay prediction

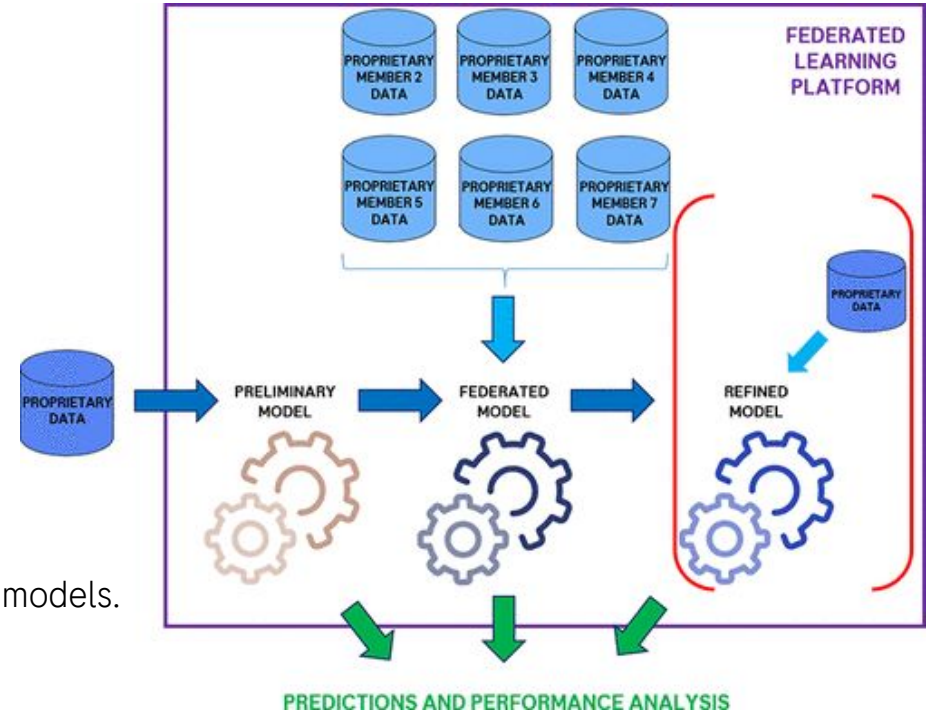
- *In vitro* toxicological datasets are typically small and imbalanced. Can we increase the chemical space of training data - without sharing sensitive data - with Federated Learning?
- Seven pharmaceutical companies participated the Effiris FL Hackathon organized by Lhasa Ltd.
- We tested Effiris on 5 internal datasets (ACh M_1, $GABA_A$ benzodiazepine binding site, 5-HT_{2B}, hERG, and COX-2), and benchmarked the performance.

➢ FL achieved a similar performance to that of internal models.

➢ FL expanded the applicability domain.

➢ Benefit of federated data varies by target.

➢ Whether FL offers benefit for a company depends on the availability of (internal) data and models



Davide Bassani, Andrea Andrews-Morger, Alessandro Brigo. Acknowledgements: Michael Reutlinger, Liudmila Polonchuk, Sonia Roberts, Angelo D'Annunzio, Lhasa. See details in the publication by Bassani, Brigo, and Andrews-Morger Federated Learning in Computational Toxicology: An Industrial Perspective on the Effiris Hackathon.

We use Large Language Models for information integration and retrieval

Problem: Terminated projects are annotated in an internal **Attrition Database.** It lacks **detailed information of attrition reasons, which hinders reverse translation,** *i.e.* re-use of knowledge/data to support projects.

Goal: To enrich project-specific discontinuation reasons (e.g. **toxicity type, main organ of toxicity, hypothesis of causes**) by integrating data from documents, presentations, reports and meeting minutes.

Results: In a pilot study, the **GPT-4-Turbo** model delivers detailed reasoning about discontinuation with an **accuracy ranging from 76 to 89%** depending on the parameters of retrieval.

Outlook: To improve robustness of the retrieval pipeline and to expand the pilot study, with the goal to support projects by **retrieving safety-relevant information**.

Ilya Schneider, Tatyana Doktorova, Julia Pletz, Alessandro Brigo, Igor Kulev, Torsten Schindler, Jingshu Sun, Elena Rivkin, Ercan Suekuer, Eunice Musvasva, Franziska Boess, Alessandro Brigo, Martin Bopst, Lutz Mueller, Susanne Mohr, Christian Freichel, Franziska Regenass, Christine Schubert, Sven Kronenberg, Barbara Lenz, Marlene Juedes, Tianyi Jiang, Nika Rack, Dragomir Draganov, Maddalena Marchesi

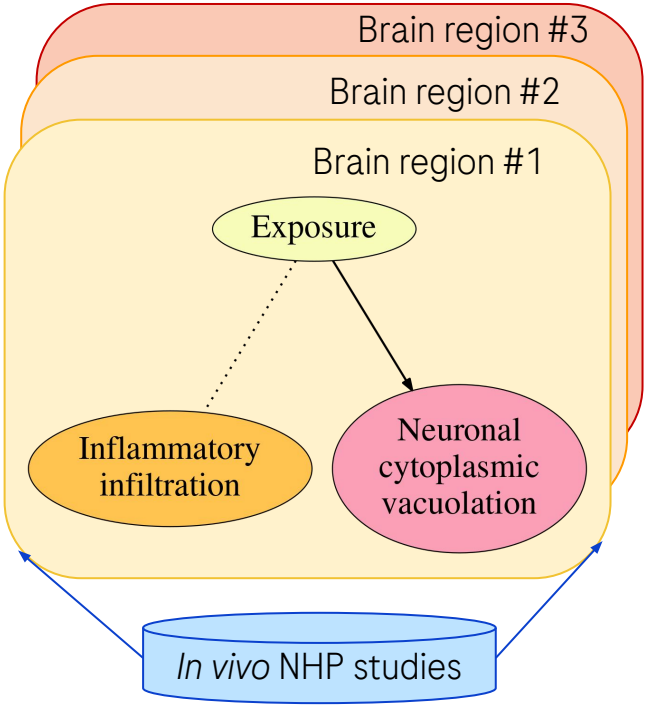
We integrate and analyse historical data to investigate relationships between exposure and pathological findings

Question: Four nucleic acid based medicine (NABM) compounds caused different types of findings in the central nervous system (CNS) in non-human primate (NHP) studies. *We wondered how individual pathological findings are correlated with exposure and with each other.*

Approach: Forming an expert team, we curated, integrated, and analysed brain-region-specific exposure and pathological findings from 9 *in vivo* studies.

Results: We observed meaningful correlation between exposure and neuronal cytoplasmic vacuolation in one brain region, yet no meaningful correlation between exposure and inflammatory infiltration. We consider that inflammatory infiltration and neuronal cytoplasmic vacuolation are likely caused by different mechanisms of action.

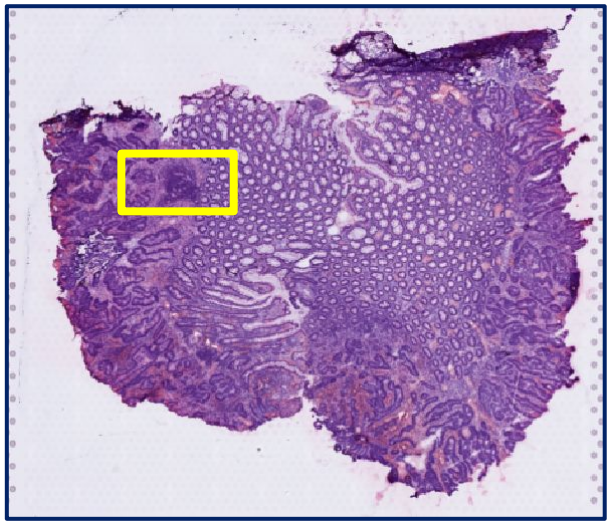
Outlook: We plan to integrate biomarker data in order to further improve our understanding.



Contributors: Zhiwen Jiang, Julia Pletz, Laura Polledo, Pierre Maliver, Dragomir Ivanov Draganov, Nadege Foiselle, Charlotte Bon, Yann Tessier

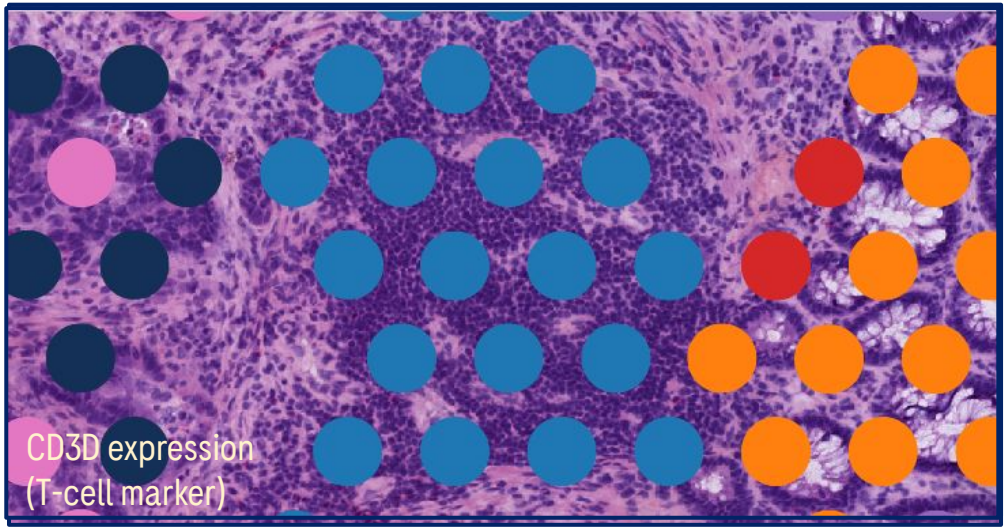
We support precision pathology with spatial omics

Spatial Transcriptomics brings together high-resolution omics data with detailed morphological information, and help us track transcriptomics changes in specific regions (e.g. drug-induced lesions).



Colorectal Cancer Sample

Pathological Annotations based on morphology and gene expression



Tumor | Immune cell aggregate | Normal mucosa

We are developing annotation transfer algorithms and *ActiveVisium,* an semi-supervised, active learning software as an AI assistant for rapid spot annotations to empower the pathologists.

Conclusions

We combine human and machine intelligence to generate insights and to support ongoing portfolio projects.

Computational pharmaceutical sciences shall become even more essential for drug discovery and development as (1) data awareness increases, (2) better tools become available to more people, and (3) more digitally savvy and enthusiastic people join us.

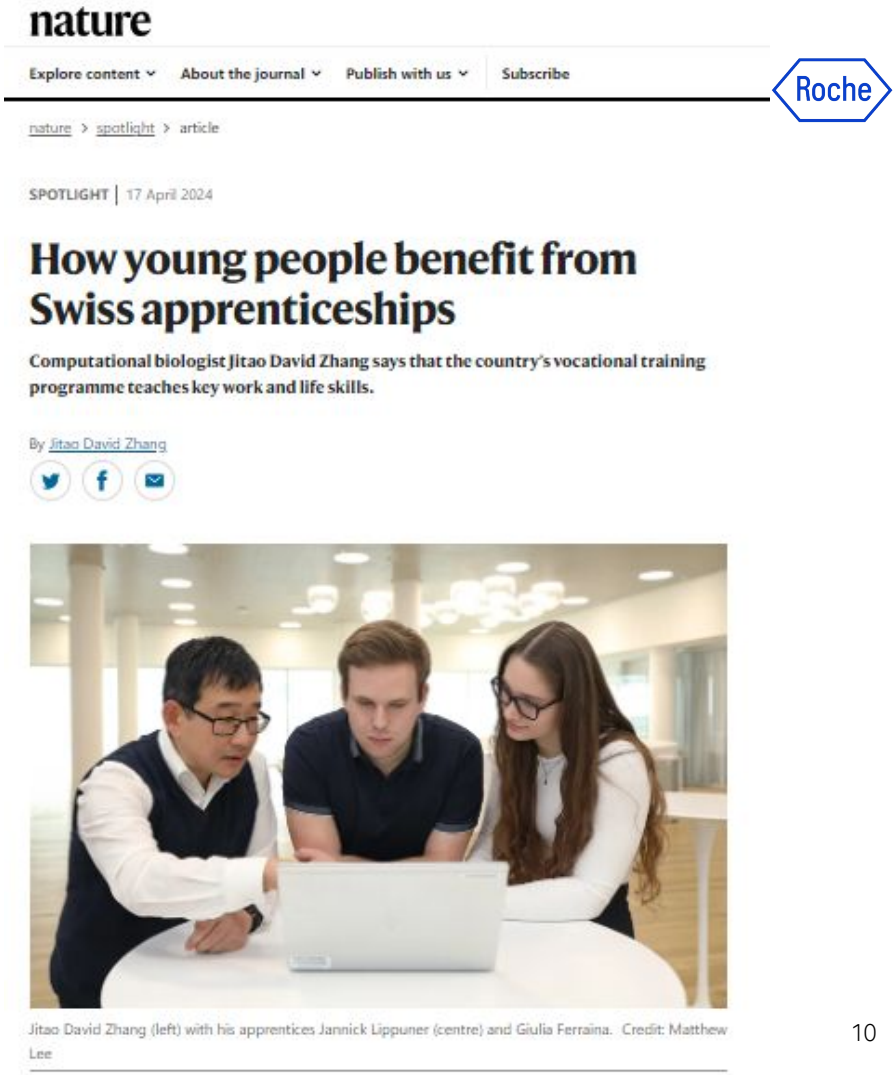
ML/AI as inspiration
Phototoxicity, PK

ML/AI as amplifier
Federated learning, LLM

ML/AI as microscopy
Exposure/pathology, spatial omics

We train people to create and leverage machine intelligence

- Apprentices in informatics and computer technology
- Internship and master thesis
- Co-supervision of PhD theses
- **Open postdoc position**: machine learning for PK prediction and prediction validity. Please contact me for more information.



nature

Explore content · About the journal · Publish with us · Subscribe

Roche

nature > spotlight > article

SPOTLIGHT | 17 April 2024

How young people benefit from Swiss apprenticeships

Computational biologist Jitao David Zhang says that the country's vocational training programme teaches key work and life skills.

By Jitao David Zhang

Jitao David Zhang (left) with his apprentices Jannick Lippuner (centre) and Giulia Ferraina. Credit: Matthew Lee

Doing now what patients need next

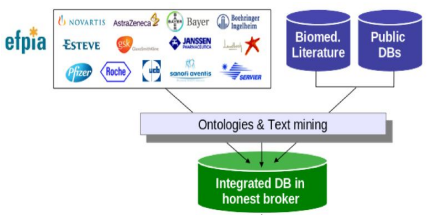
Backup slides

Biological read-across

Data from animal tox study

+

eTox database



Finding historical molecules with similar effect in eTox

Method:
Kolmogorov-Smirnov test

- Similar **histopathology** observations
- Similar **lab test** results
  - → Clinical chemistry
  - → Hematology
  - → Urinalysis

Molecules with similar effects

Analysis of mechanistic data

- Structurally similar molecules in the ToxCast database

Hypotheses of pathways that lead to the toxicity

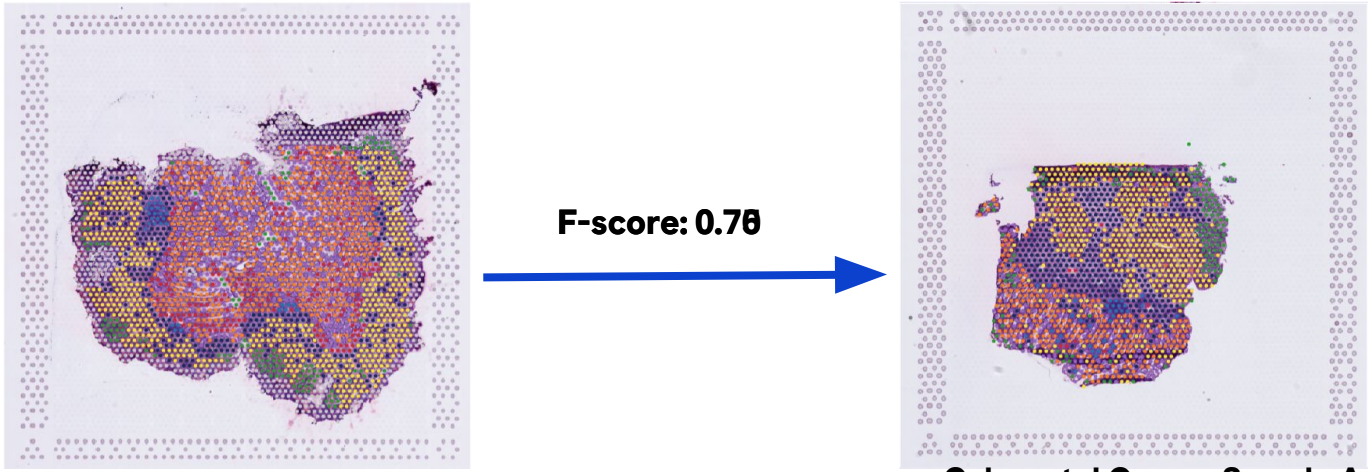
Hypothesis verification

- RNAseq data
- Literature review

Contributors: Elizaveta Chernova, Tatyana Doktorova, Dragomir Ivanov Draganov, Jitao David Zhang, Maddalena Marchesi, Zhiwen Jiang, Melanie Guerard

We work on an AI assistant for rapid spot annotations

1. **Foundation Models in Digital Pathology to extract knowledge based on the morphological features** of a small set of manually annotated locations by a pathologist (Tumor, stroma …).
2. **Active learning for selection of most informative, yet unannotated spots**, allowing fast pathological annotations of large samples.
3. **Annotation transfer** from an annotated sample to a replicate from the same sample or to another colorectal cancer sample → **Fast integration of histology and gene expression across samples.**



Structure of the talk

Abstract of the talk

The talk will be given internally in ToxAcademy and externally in SSPT

Abstract: In this talk, I will share insights derived from real-world applications of machine learning, causal inference, and computational methods to predict and explain pharmacology and toxicity profiles of drug candidates. Drawing learnings from undertaking three tasks in drug discovery, including pharmacokinetic parameter prediction, safety evaluation, and mode-of-action inference, I try to sketch a possible plan towards intelligent benefit-risk assessment of drug candidates in translational research, using fit-for-purpose models, causality, and combined human and machine intelligence.

Draft agenda of SSPT (25 minutes, likely I will use 18 minutes for the talk and 7 minutes for Q&A)

Talk time: 11:25-11:50

Before: Igor Tetko, Helmholtz Munich, AI meets toxicology

11:25 - 11:50 Jitao David Zhang, F. Hoffmann La-Roche Basel

After: 11:50 - 12:15 Maximilian Brackmann, Federal Department of Defence, Spiez

Draft agenda of ToxAcademy (20 minutes, likely I will use 15 minutes for the talk, and 5 minutes for Q&A)

Proposed structure: criticisms & feedback welcome

1. Combining human and machine intelligence for safety evaluation (1 slide, to be made by David)
    a. Opportunities: data awareness, better models, digitally-savvy people
    b. Dangers/illusions: AI as oracle, AI as surrogate, AI as quant, AI as arbiter.
    c. Our belief: AI as inspiration, AI as Organiser, AI as Microscope.
2. Our use cases: while I may not be able to cover them all in the talk in detail. The slides are made by individual teams, harmonized by David, and checked by the team together. Please name all contributors on each slide. One slide per topic in the main deck, backup slides if necessary. Please make sure the texts are all readable at one arm's length without using a glass:
    a. **Predicting off-target effects, toxicity endpoints, and PK (1 slide each topic)**
        i. Machine learning for tox-end prediction: example from Nicolo and Andrea and co
        ii. Federated learning: example from Davide, Alessandro, and Andrea and co
        iii. Please also prepare one slide for PK (Davide, Leonid, Andrea, Lisa, etc.) to be included in the backup.
    b. **Leveraging historical and heterogeneous data (1 slide each topic)**
        i. Using LLM to retrieve information from heterogeneous data sources: example from Ilya, Ercan, Tatyana and co. Or: the read-across example from Eliz, Zhiwen, Melanie, Tatyana and co. David recommends preparing both slides - they are beautiful work - and showing one due to time.
        ii. Leveraging historical data and connecting exposure with in-vivo data to understand causes of toxicity: example from Zhiwen, Julia, and the RNAhub team.
    c. **Omics and precision pathology (1 slide)**
        i. ML/AI in spatial omics and digital pathology: example from Alberto, Petra and the Pathology team.
3. Perspective of combing machine and human intelligence to support drug discovery and development (1 slide)

Doing now what patients need next