

RAZVOJNA POT: ZNANOST

Partner razvojne poti: SLAiF

Naslednja predstavitev:

Genomski jezikovni modeli v raziskavah raka

dr. Fran Supek,

Head of Genome Data Science Laboratory, University of Copenhagen

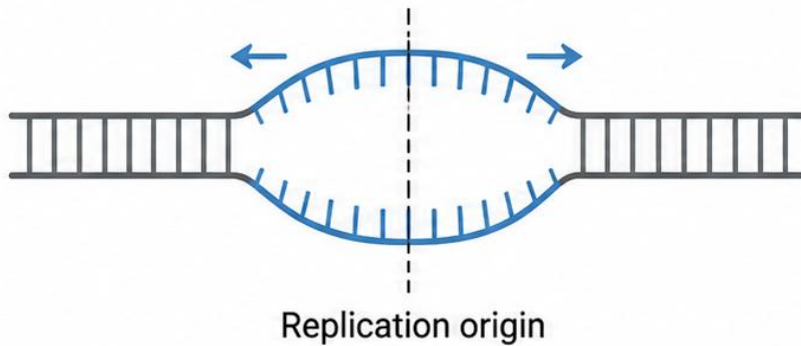
Learning the hidden instructions in our genomes

Genomic language models for elucidating DNA replication, mutagenesis and RNA turnover

Fran Supek
BRIC, University of Copenhagen
Genome Data Science laboratory

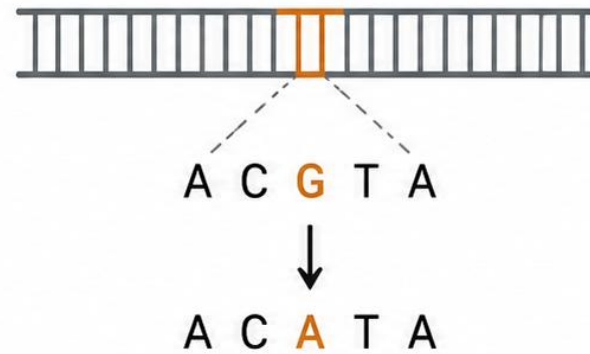
Three questions about genetic information

Where does copying begin?



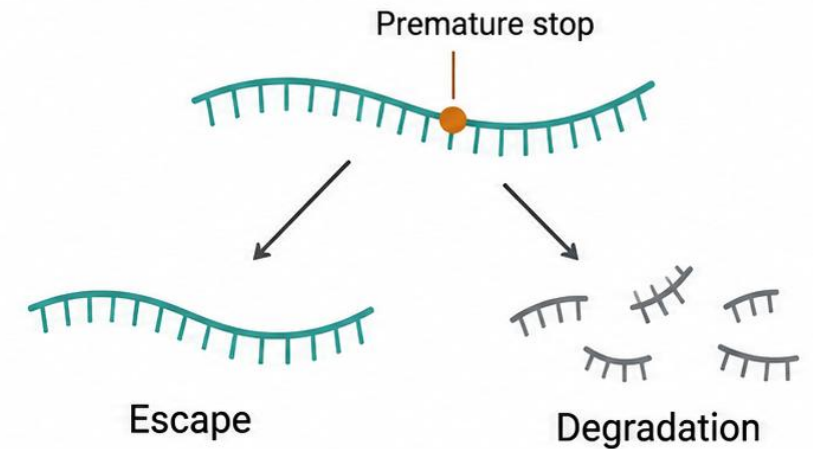
ORIFormer

Where do errors occur?



MutFormer

Does the altered RNA survive?



NMDetective-AI

Learning biological properties from sequence

Input

DNA or RNA sequence, with task-relevant context

Learning signal

1

2

3

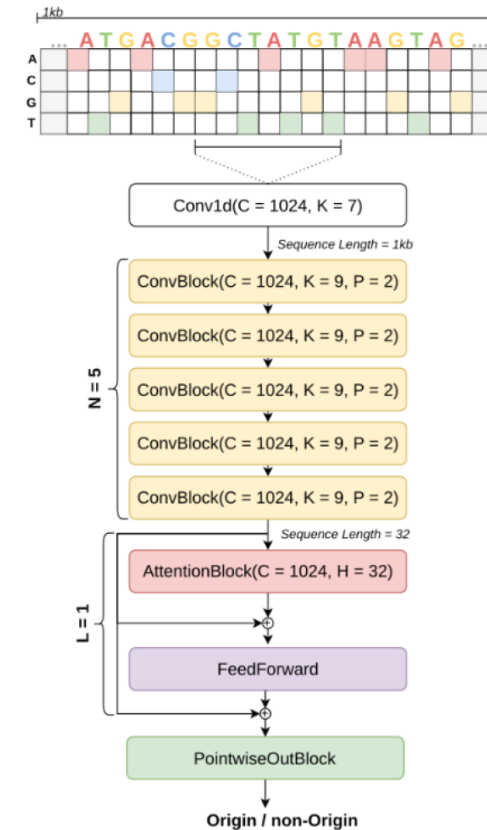
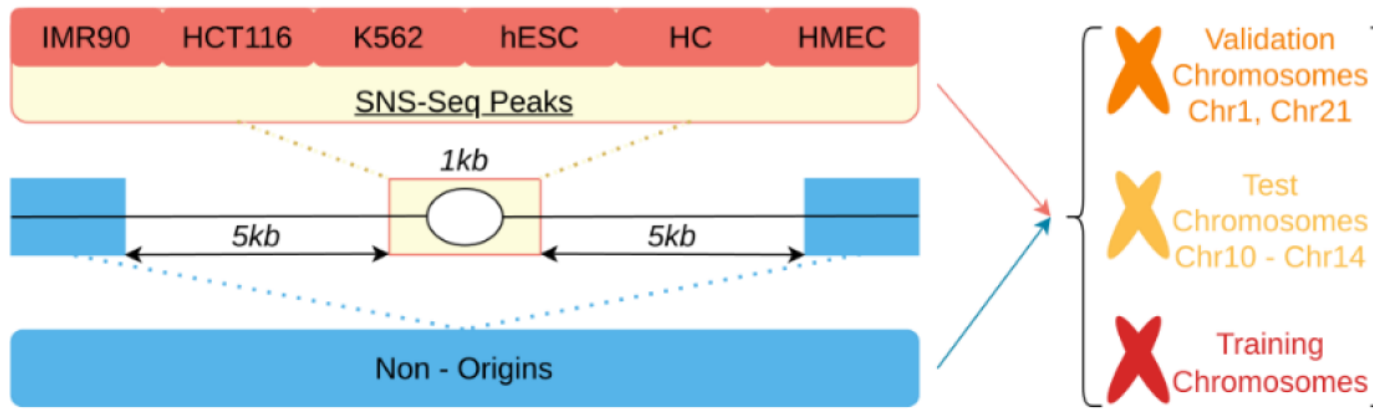
Measured replication origins, somatic mutations or mRNA abundance

Output

A biological prediction that we can interpret and test

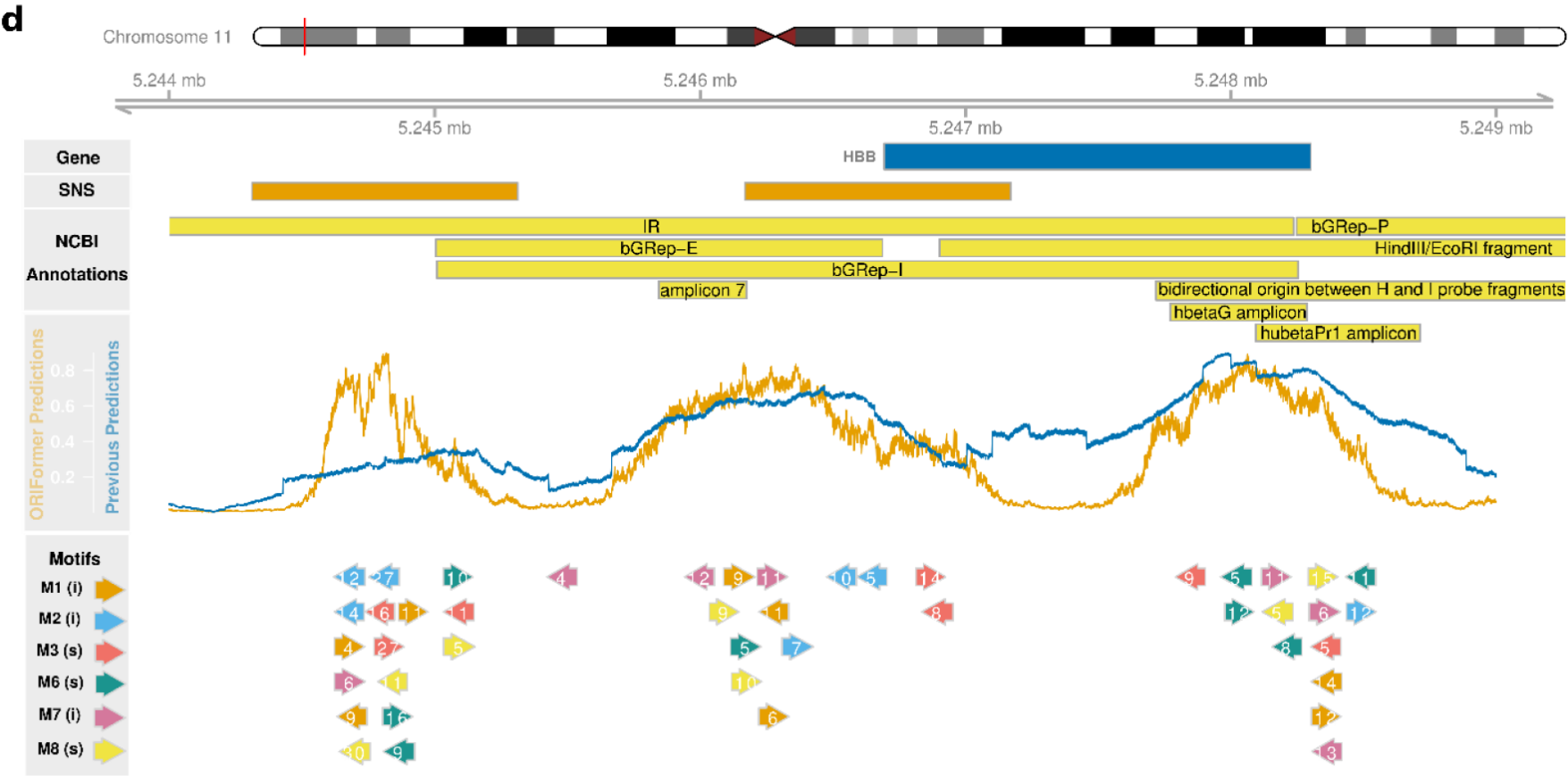
Where DNA copying begins

ORFormer predicts replication initiation potential from DNA sequence



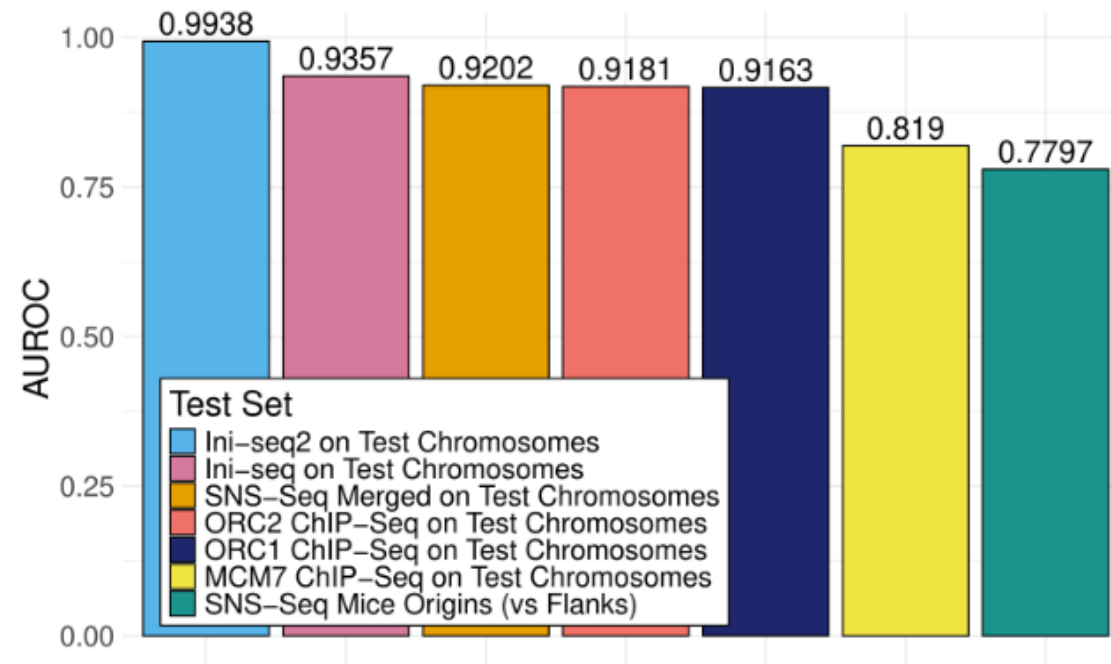
A sequence model locates replication origins

The HBB locus makes the prediction visible along a stretch of DNA



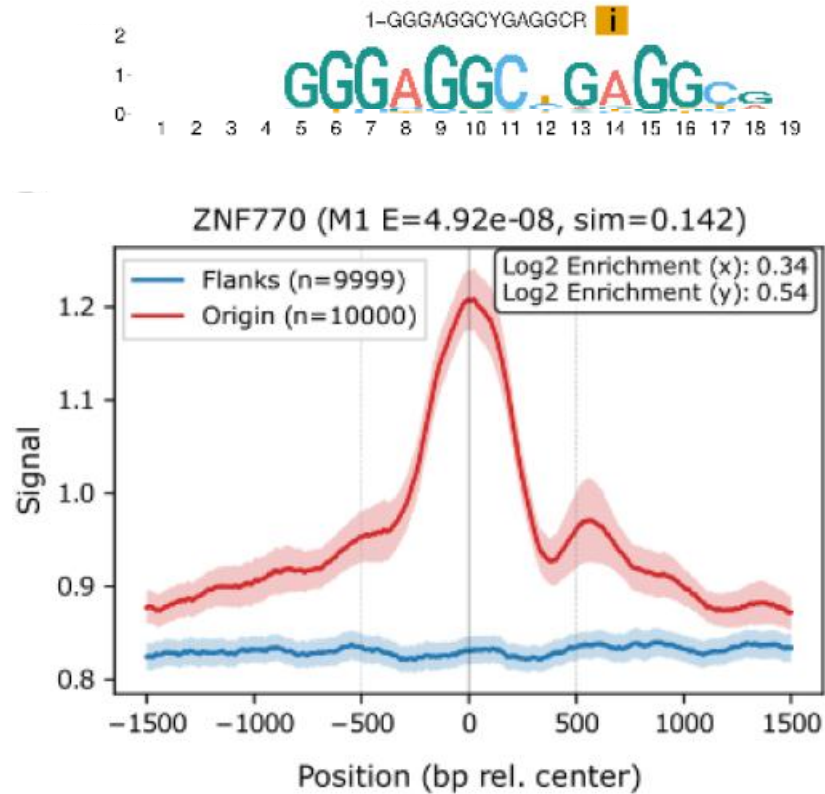
Origin predictions generalise across assays

Independent assays and mouse origins support a conserved sequence signal

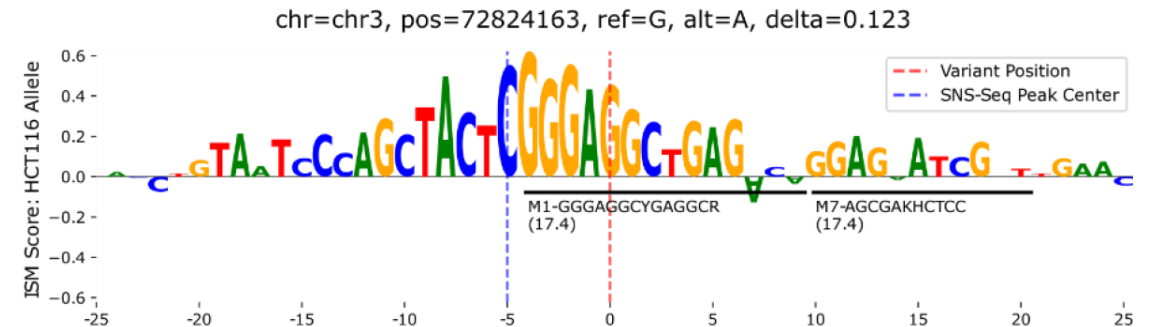


Learned motifs point to origin-associated proteins

A prominent ORFormer motif matches binding preferences of ZNF770 protein



Mutation effects perturb origins



Why the same mutation can recur



High mutation risk

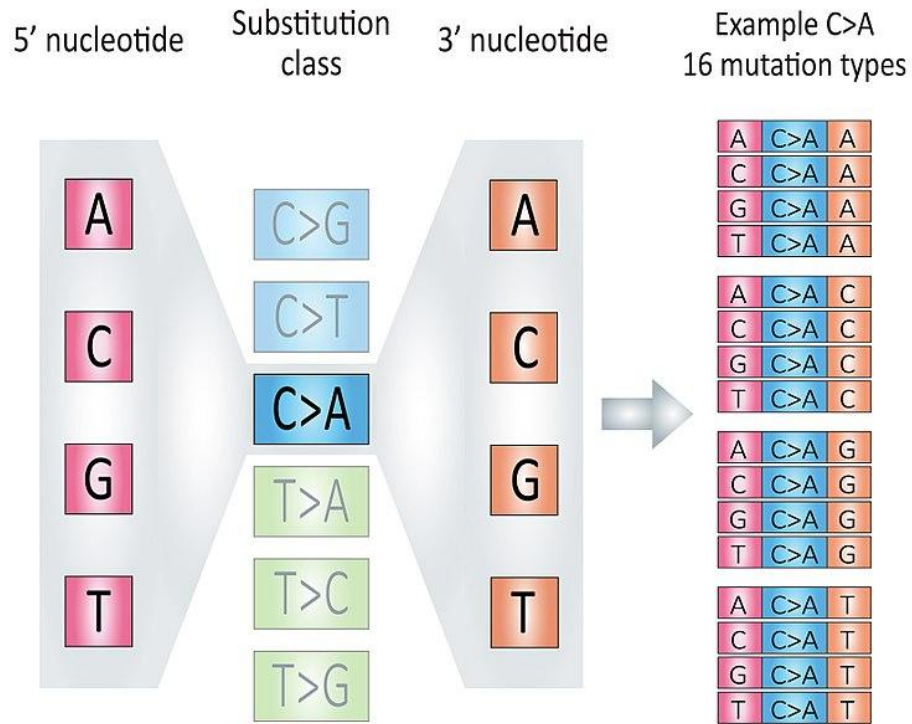
Some DNA sites acquire mutations unusually often

Selection in the tumour

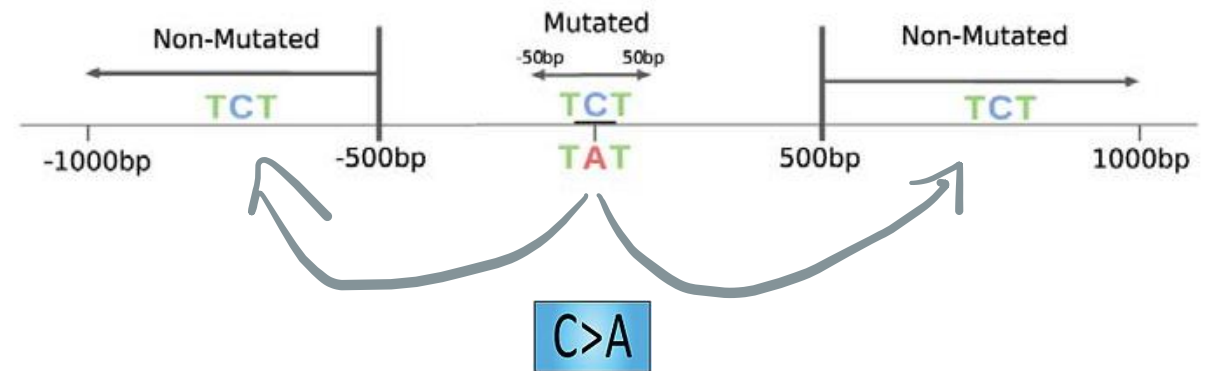
Some mutations help the affected cell expand

Interpreting recurrence requires an estimate of local mutation risk

MutFormer learns local mutation risk



Matched controls isolate information beyond the immediate three-letter context



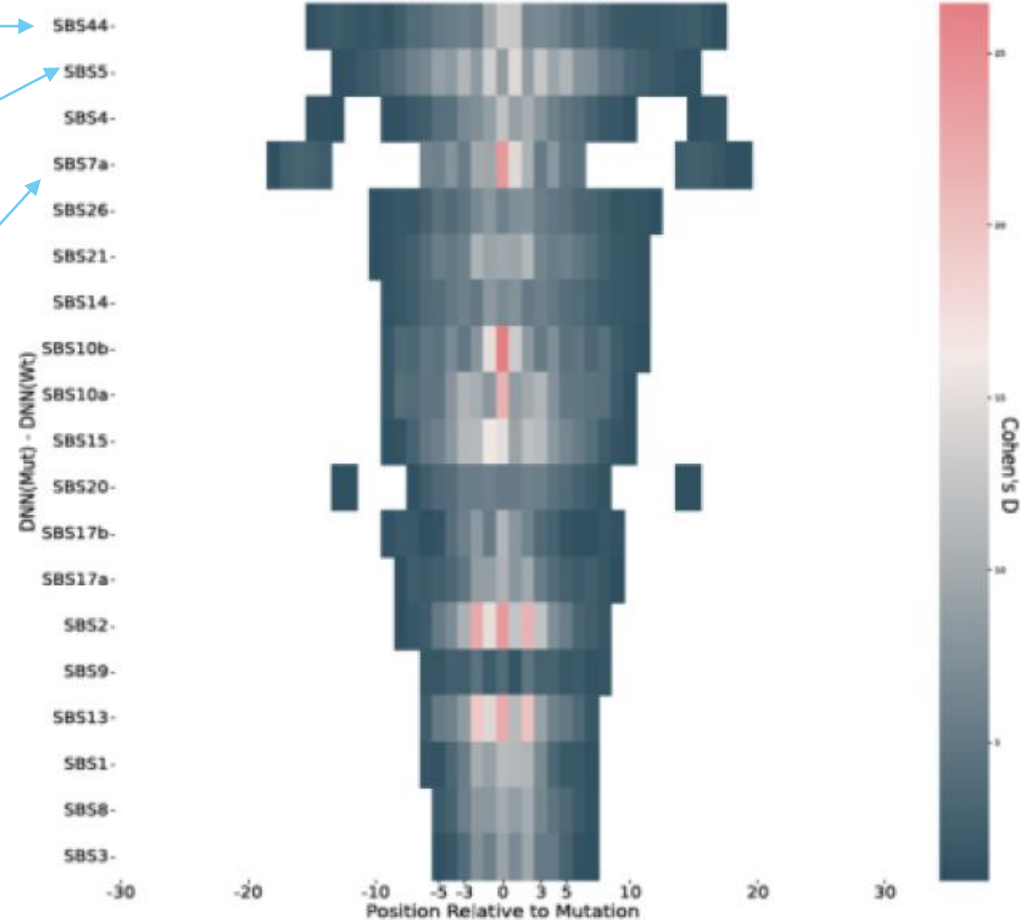
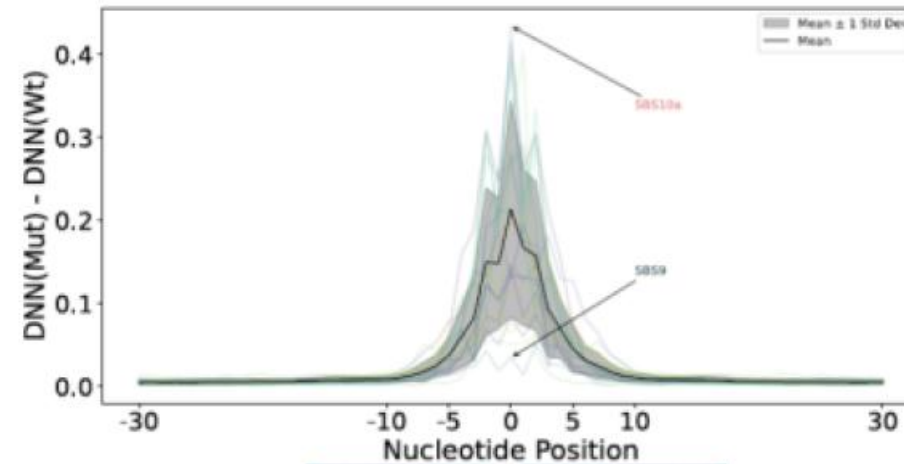
The surrounding DNA influences mutation risk

Different mutational processes recognise different extended DNA sequence contexts

broken
DNA
repair

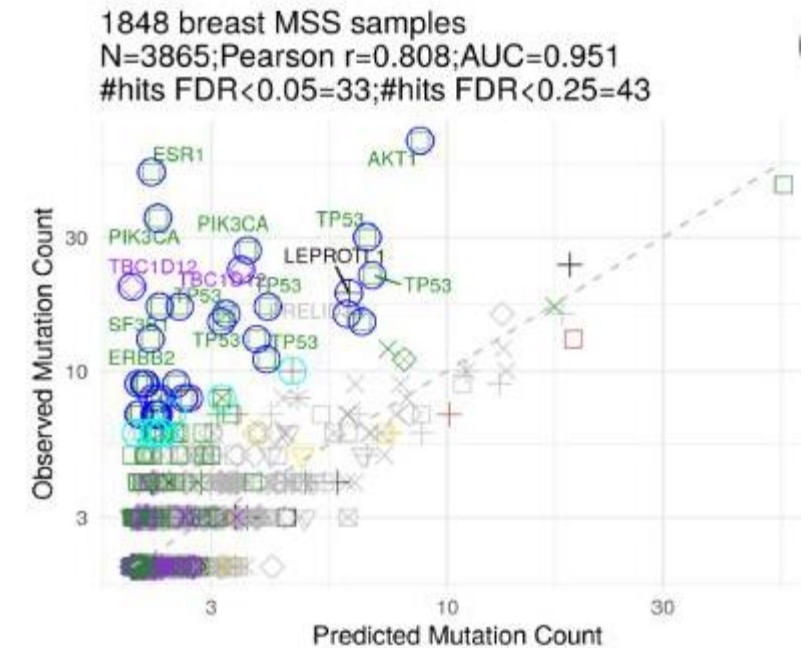
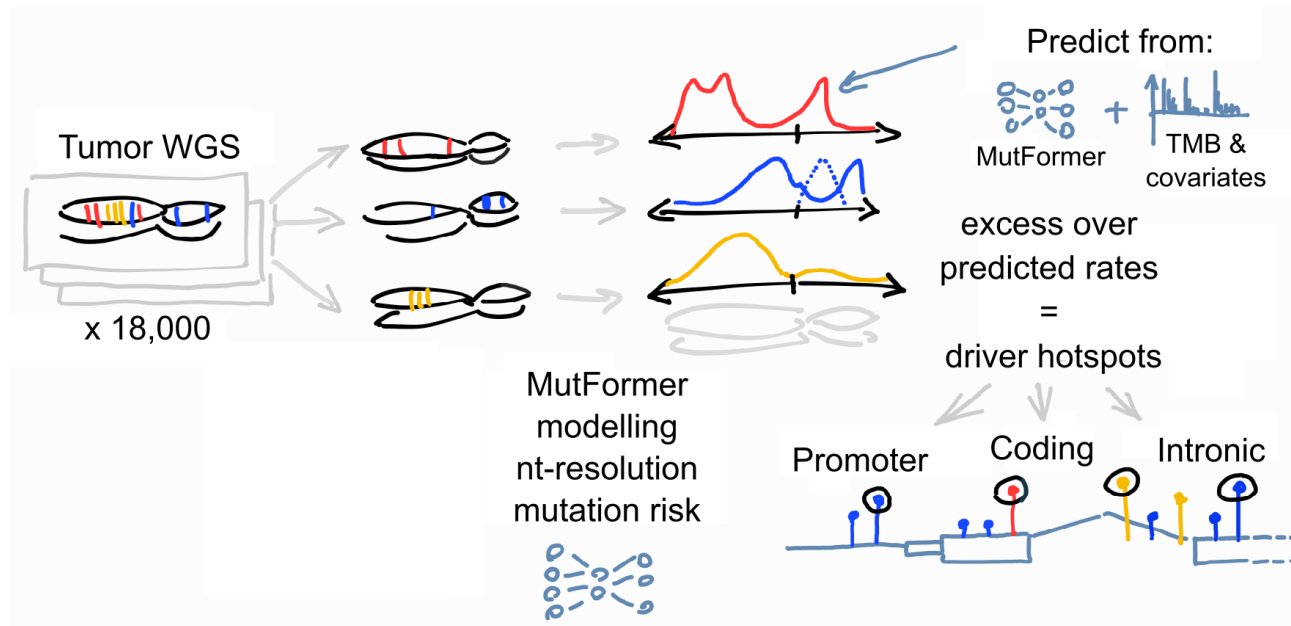
aging

UV
radiation



Expected recurrence helps identify selected hotspots

Many recurrent sites fit the mutational baseline, while selected candidates stand out



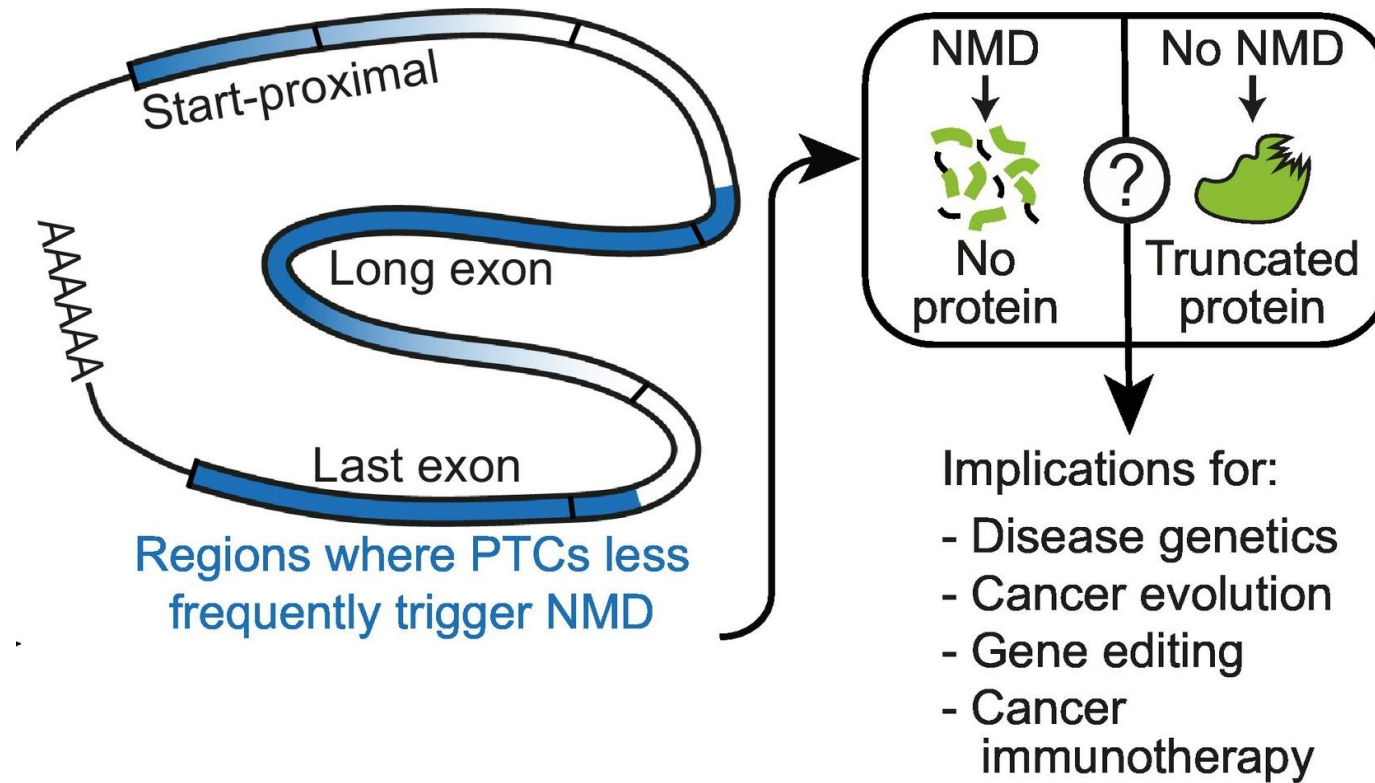
Mutpanning & CGC lists

- common mutational cancer driver
- rare mutational cancer driver
- likely driver genes
- possible driver genes
- passenger genes
- in vitro hairpin passenger

The fate of an altered RNA message

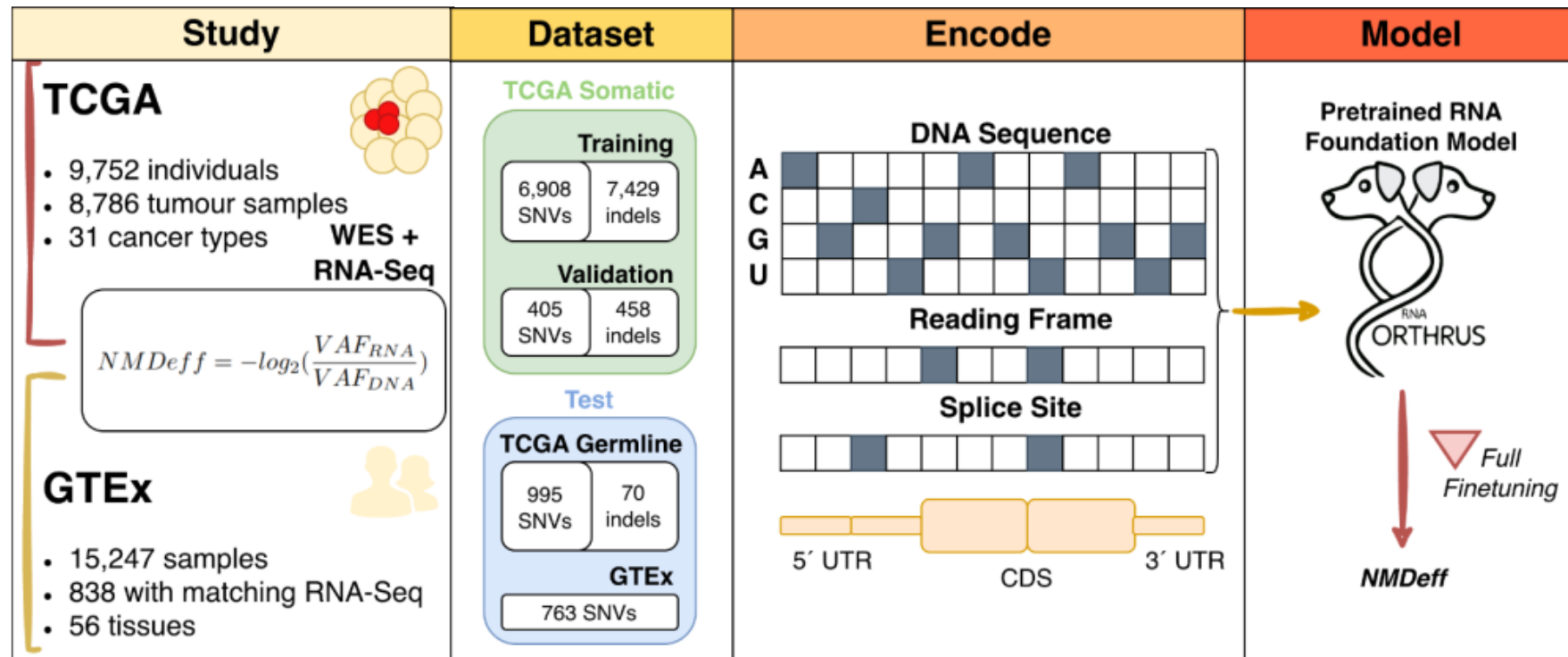
3

Nonsense-mediated decay can remove RNA containing a premature stop mutation



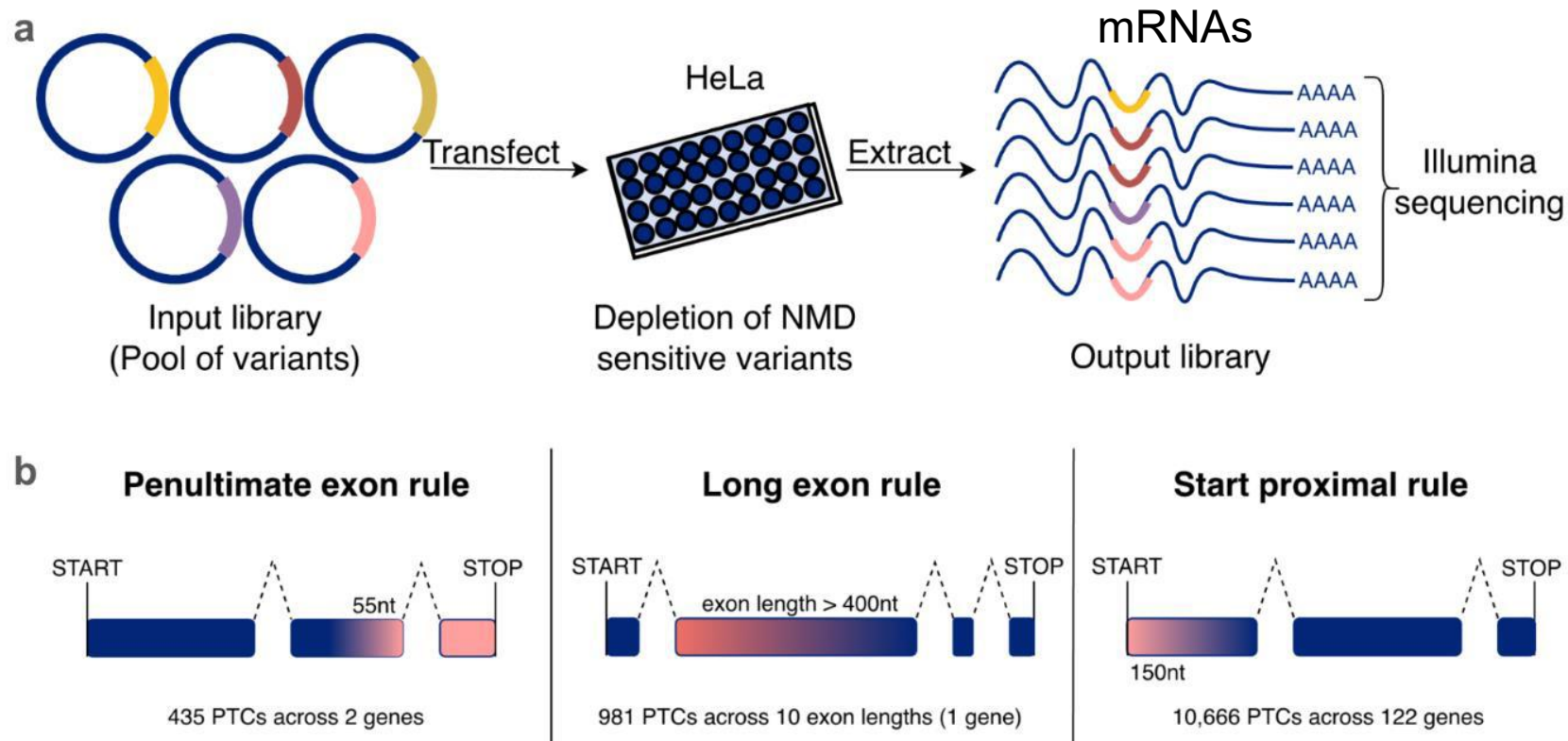
NMDetective-AI predicts the degree of RNA decay

A pretrained RNA model adapts to quantitative measurements from human variants (mutations)



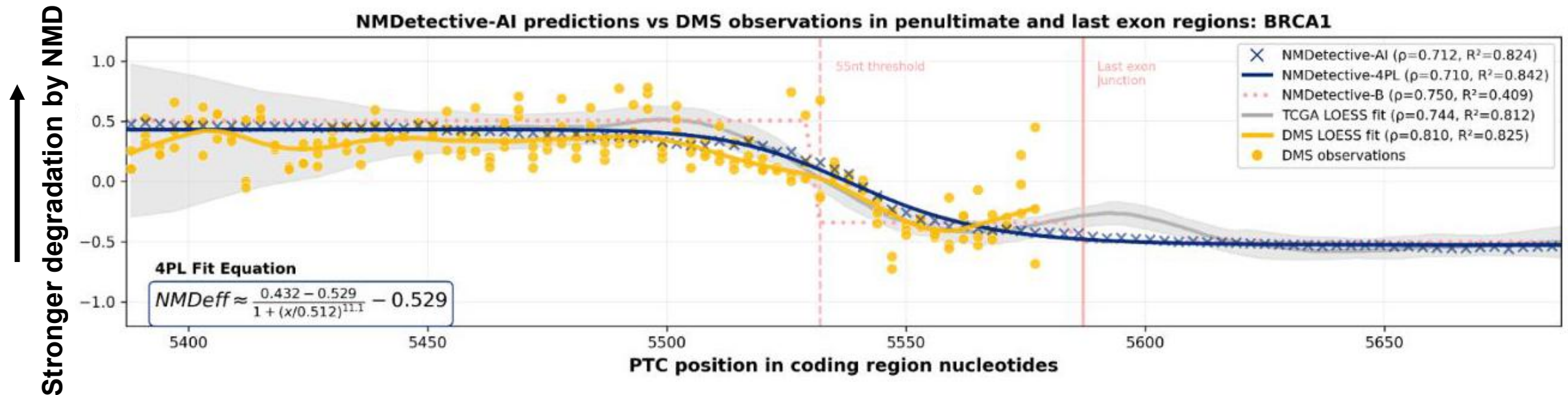
Systematic experiments test the predictions

Mutational scanning measures RNA abundance across engineered variants



A binary NMD rule resolves into a graded response

BRCA1 mutagenesis supports a quantitative transition in RNA degradation



Quantitative RNA fate informs variant interpretation

Molecular consequence

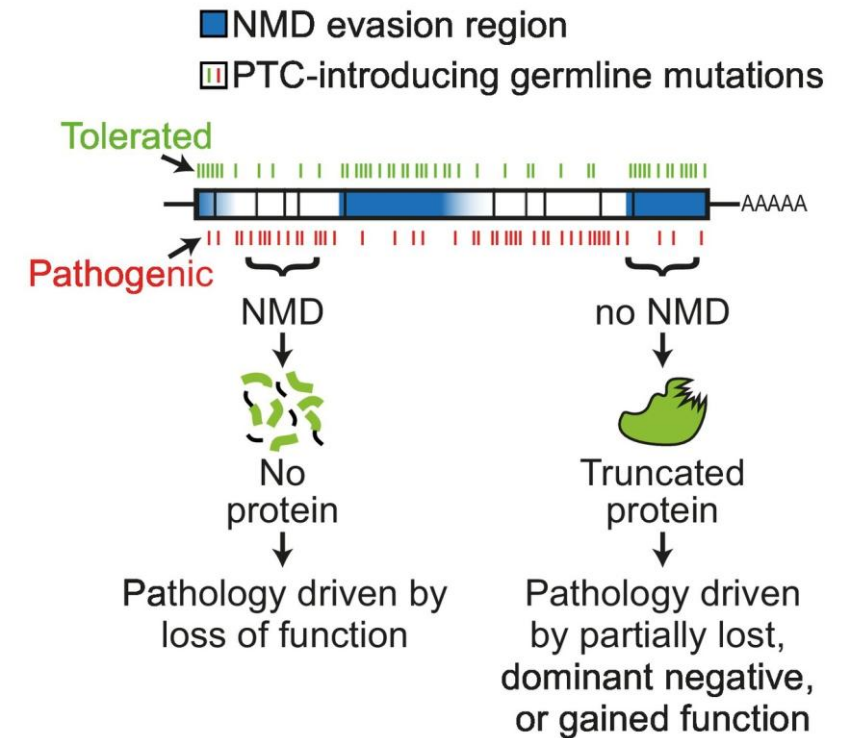
Estimate how strongly a premature stop reduces its RNA

Cancer relevance

Consider RNA loss alongside the gene and altered protein

Research use

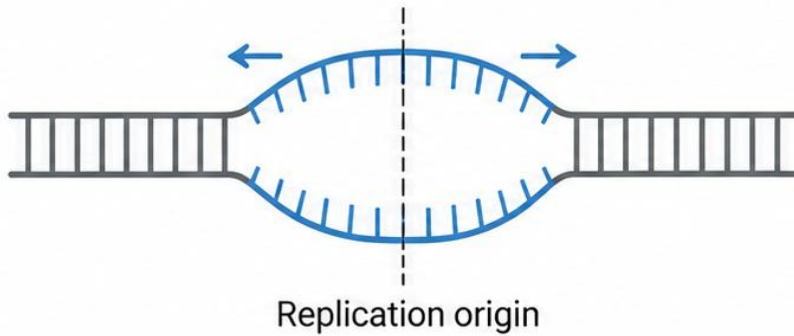
Prioritise variants and hypotheses for functional testing



Sequence models reveal different layers of genome function

Replication potential, **mutation risk** and **RNA fate** each contain learnable sequence information

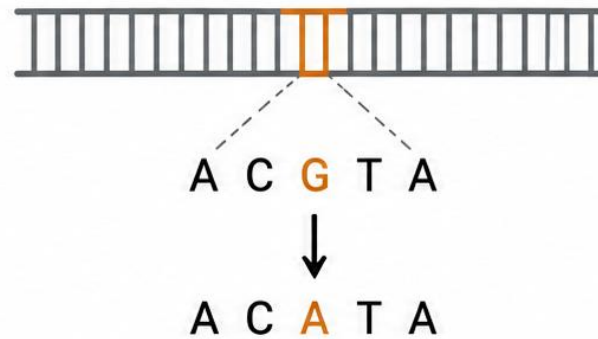
Sequence helps specify
replication origins



ORIFormer

Evidence: independent assays
and protein-binding data

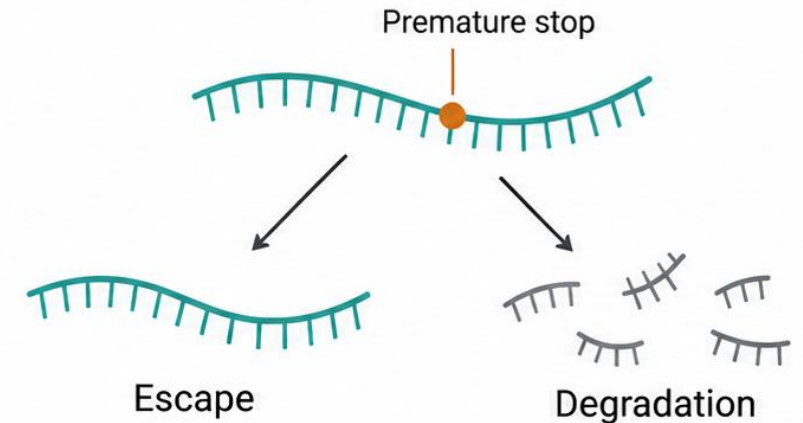
Extended sequence context
helps explain mutation hotspots



MutFormer

Evidence: observed versus
expected mutation recurrence

RNA decay follows
quantitative sequence rules



NMDetective-AI

Evidence: controlled
mutagenesis experiments

The people behind the work

ORIFormer and MuSAS

Marcell Veiner, Marina Salvadores, Iván Galván-Femenía

<https://doi.org/10.64898/2025.12.28.696672>

MutFormer

Iván Galván-Femenía, Marcell Veiner, Daniel Naro

<https://doi.org/10.64898/2026.07.07.736824>

NMDetective-AI

Marcell Veiner, Ignasi Toledano, Guillermo Palou-Márquez, Ben Lehner

<https://doi.org/10.64898/2026.03.24.71400>

Genome Data Science laboratory, BRIC / University of Copenhagen, and IRB Barcelona

Thank you! Questions?



Get in touch! fran.supek@bric.ku.dk
<https://www.genomedatalab.org/>
X: @FranSupek @GenomeDataLab



Naslednja predstavitev:
**Slovenska tovarna UI (SLAIF) kot infrastruktura za znanost, gospodarstvo
in javni sektor**

**prof. dr. Sašo Džeroski, Institut Jožef Stefan,
Odsek za tehnologije znanja,**



SLAIF – Slovenska tovarna UI

**Umetna inteligenca kot infrastruktura za znanost ,
gospodarstvo in javni sektor**

Brdo pri Kranju, 11. september 2026

Prof. Sašo Džeroski – tehnični vodja HE projekta SLAIF

KAJ SO TOVARNE UI?

*Kratek odgovor: **Tovarne UI (AI factories) so infrastruktura za UI***

- V 20. stoletju so države gradile ceste in električna omrežja
- V 21. stoletju morajo graditi tudi infrastrukturo za UI (ker zahtevata gradnja in prilagajanje generativnih modelov visokokozmogljive računalnike=HPC)
- To je vprašanje tehnološke suverenosti in globalne konkurenčnosti

Slovenska tovarna UI (SLAIF) nam bo omogočila,
da Slovenija ni zgolj uporabnica umetne inteligence,
temveč **aktivna soustvarjalka evropske tehnološke suverenosti**
in globalne konkurenčnosti

KAJ SO TOVARNE UI?

Tovarne UI znižujejo ovire za uvajanje UI v prakso

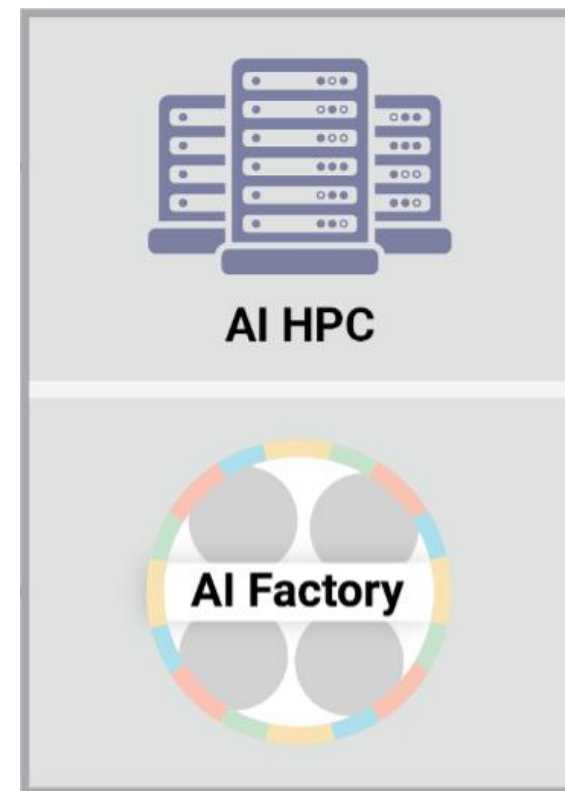
- Ponujajo širok in celovit nabor storitev za industrijo in raziskovalce, ki so potrebne za razvoj evropskih modelov (generativne) UI in njihovo uporabo

Tovarne imajo dve glavni komponenti:

- Superračunalnik, optimiziran za UI
- Storitve in dejavnosti tovarne UI

Tovarne UI združujejo:

- Računsko moč
- Algoritme in delovne tokove
- Podatke in modele
- Veščine, znanje in talente



EVROPSKI EKOSISTEM TOVARN UI

- EuroHPC in države članice sofinancirajo postavitev **tovarn UI** (s superračunalniki) in **anten** (uporabljajo računalnike tovarn UI)
- Skupaj 19 tovarn UI
- Skupaj 13 anten
- A le 9 tovarn nabavlja in postavlja nove superračunalnike
- Nekaj držav ima deleže v tovarnah UI, ki se vzpostavljajo v drugih državah (npr. Portugalska in Turčija pri BSC AIF)
- Slovenija bo gostila nov superračunalnik, ima pa tudi delež v IT4LIA (10 mio EUR)



ZAKAJ IMA LAHKO SLOVENIJA TOVARNO UI?

Čeprav bo imela Evropa številne tovarne UI, jih ne bo imela vsaka država EU. Potrebna je vizija. Prav tako politična volja in finančna podpora. In nenazadnje: znanstvena in tehnična podkovanost ter človeški viri

- **Slovenija ima močno skupnost na področju UI**
- **Slovenija ima močno skupnost na področju visokozmogljivega računalništva:**
Vega je prvi delujoči sistem EuroHPC
- **Skupnosti odlično sodelujeta** (*kar ni samoumevno*)

Potreben je širši pogled na infrastrukturo

- Infrastruktura ni le strojna oprema (računalniki)
- Podatki so infrastruktura (Slovenija vodi CLARIN, jezikovni podatki)
- UI je infrastruktura (vsekakor za znanost, primer infrastrukturnega centra IJS)

SLAIF – SLOVENSKA TOVARNA UI

Dva projekta, skupna vrednost 135 mio EUR, sofinancerki sta EK in RS (MVZI)

- Superračunalnik, optimiziran za UI, 123.32 mio EUR
 - Sofinanciran iz programa Digital Europe
 - Konzorcij projekta sestavljajo



- Tovarna UI, storitve in dejavnosti, 11.68 mio EUR
 - Sofinanciran iz programa Horizon Europe
 - Konzorcij projekta sestavljajo



SLAIF – SUPERRAČUNALNIK

Kombinacija CPU in GPU enot za podporo tako klasičnemu kot UI-računanju

- CPU (cca 10 petaFLOPS)
- GPU (cca 30 – 1. faza +70 – 2. faza = 100 petaFLOPS)

Dve fazi nabave za zagotovitev daljšega obdobja delovanja

- Superračunalnik zagnan do konca leta 2027
- Superračunalnik nadgrajen v letu 2029
- V primerjavi s superračunalnikom VEGA: 3× več moči CPU in 32× več moči GPU

Trenutno se na superračunalniku LEONARDO v Bologni trenirajo:

- Veliki jezikovni model za slovenščino – GaMS (kolegi z UL FRI)
- Prvi vizualno-jezikovni model za slovenščino – SViLa (kolegi z IJS)
- Takšne naloge bo podpiral novi superračunalnik optimiziran za UI

Generativni model
za slovenščino
Generative Model
for Slovene

GaMS: “

SLAIF – CILJNE SKUPINE UPORABNIKOV

Različne skupine uporabnikov

- Raziskovalna sfera (univerze, instituti)
- Gospodarstvo/ industrija
 - Zagonska podjetja
 - Mala in srednja podjetja
 - Velika podjetja
- Javni sektor
 - Zdravstvo
 - Državna uprava
 - Javna podjetja, javni zavodi in agencije

Različne stopnje UI-pripravljenosti

SLAIF – STORITVE IN DEJAVNOSTI

Dostop do računskih in podatkovnih zmogljivosti

Splošni delotoki za učenje in prilagajanje modelov

- Veliki jezikovni modeli
- Veliki govorni modeli
- Vizualno-jezikovni modeli
- Večmodalni modeli

Domensko-prilagojeni delotoki (in modeli) za različne vertikale

- Zeleni prehod
- Zdravje in botehnologija
- Digitalna družba
- UI za znanost

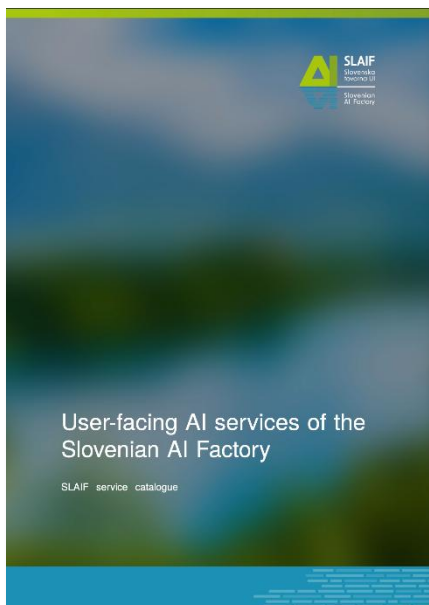


Izobraževanje in pridobivanje veščin UI

SLAIF – STORITVE

Več vrst storitev

- Podpora uporabnikom
- Infrastruktura in podatki
- Razvoj in uvajanje UI
- Produkti UI



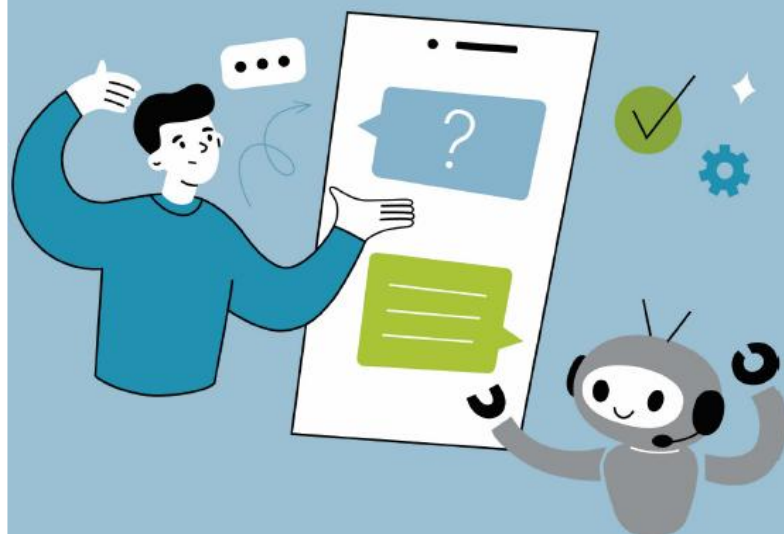
Contents

User support	9
Business Processes Review and Automation Using AI	11
Support at Distributed Training of Large-Scale AI Models	12
Support in Model Optimisation: Pruning, Quantisation, and Distillation	14
Support in Proof-of-Concept (PoC) Application Development	16
Technical User Support	18
Infrastructure and data-related services	19
Cloud Platform Service (OpenStack PaaS)	21
Container Image Registry	22
Container Platform Service (Kubernetes)	23
Data Access Services	24
Data Analytics Services	26
Semantic Data Discovery over the SLAIF Catalogue	28
Data Governance Services	30
Data Management / Sharing Services – Connectors for Trusted Sharing	31
Data preparation services	33
Data Processing Services	35
Exploration and Investigation of Disaster Events through the Slovenian EO Data Cube	37
Facebook PE-Video with Slovenian Annotations	39
FAIR-Compliant Metadata Management and Discovery Service	40
Fast Parallel Storage	42
Foundation-Model-Assisted Dataset Bootstrapping from Uncurated Visual Data (Auto-Annotation)	43
HPC Accelerated Compute Service	45
HPC Compute Service	47
Mid- and Long-Term Storage	48
Mid- and Long-Term Remote Access Storage	49
Monitoring Services	50
Orchestration Services	51
Synthetic Data Services	52
Synthetic Genetic Material Data Generation	54
Transformation and Processing Pipeline for Genetic Material Data	55
User Software Repository	57
AI development and deployment	58
Foundation Models for Human Sensing with Support for Fine-Tuning	60
Adaptation of a Large Language Model for Domain-Specific Question Answering	62
Adaptation of a Large Language Model for Domain-Specific Summarization	63
Adaptation of a Large Language Model for a Specific Task using Parameter Efficient Fine-Tuning	64
Adaptation of a Large Language Model for Sentiment and Emotion Classification using Parameter Efficient Fine-Tuning	65
Adaptation of a Large Language Model to a Specific Domain using Instruction Following	66
AI-Based Image Analysis for Industrial Anomaly Detection and Quality Control	67
AI-Based Microscopy Image Analysis	69
Automated Construction of Domain-Specific Knowledge Graphs	71
Change Detection and Description from Remotely Sensed Data	73
Embedding Selection – Intelligent Embedding Selection for Scalable AI Systems	75
ML Pipeline Exploiting Known Attribute Value Uncertainties	77
Model Evaluation for the Local Context Using Linguistic and Cultural Competence Measures	78
Per-Prompt LLM Selection – Automatically Choosing the Best Open Access LLM	80
Preference Optimization Pipeline and LLM-Based Machine Translator for English and Slovene Pairs	82
Satellite and Aerial Image Analysis	83
Temporal, Multi-Modal Foundation Models from Remotely Sensed Information	85
Visual and Interactive AI Workflows (OrangeWeb)	87
Visual and Interactive AI Workflows (OrangeWeb) for Green Transition	88
Visual and Interactive AI Workflows (OrangeWeb) for Health and Biotechnology	89
AI products	90
Access to a Multimodal Language/Sensor Model for Data Analysis	92
Agentic RAG for Administration	94
AI-Assisted Workflow for Grading Scanned Handwritten STEM Exams	96
AI-Based Human Pose Estimation and Kinematic Analysis	98
AI-Based Predictive Analytics for Strategic Resource and Trend Forecasting	100
AI-Based Visual Defect Detection and Quality Assessment for Industrial Production	102
AI-Driven Predictive Maintenance and Performance Optimization for Energy Systems	104
AI-Enhanced Weather Forecasting for Power Grid Operations	106
AI for Health and Biotechnology – AI-Assisted Lab Blood Test Interpretation	108
AI Framework for ECG Analysis	110
AI-Powered Intelligent Document Analysis and Semantic Retrieval	112
AIDA – Agricultural Intelligence and Decision Assistant	114
Automated Classification and Feature Extraction for Experimental Materials Data	116
Crop Growth Monitoring and Analytics Services	118
Domain-Agnostic Object Counting, Detection and Tracking in Images	120
Energy Distribution Optimisation	122
FoodyLLM – AI Copilot for Food Nutrition Intelligence	123
Irradiance Prediction Service for Energy Modelling	125
Multi-Label Classification for Microbial Data	127
Multilingual Idiom Semantic Search and Translation System	128
Observable-Cue Mental Health Risk Prediction & Screening Toolbox	130
Open Model for Slovenian Acoustic Prosody Annotation	132
Pre- and Post-Surgery Primary Brain Tumour Volume Estimation	133

SLAIF – STORITVE

AI development and deployment

Services for adapting, fine-tuning, integrating, scaling, and operationalising AI workflows and models for real use cases.



SLAIF service catalogue



SLAIF service catalogue



AI development and deployment

Adaptation of a Large Language Model for Domain-Specific Question Answering

Responsible institution
University of Ljubljana

Vertical Cross-domain Availability Q3/2027 Type Pipeline

Target users: Industry, Large Enterprises, Scientific Institutions, Public Institutions

Overview

This pipeline enables the adaptation of a large language model to answer domain-specific questions accurately and efficiently. It is designed to support expert systems, knowledge bases, or FAQ systems in specialized fields. The prerequisites are a sufficient amount of domain-specific data and domain know-how. The pipeline first helps users in preparing data for question answering based on the domain data and provides code and know-how for synthetically expanding the training data based on the provided domain data. Next, it provides the know-how and code for fine-tuning the model with the created instruction-tuning dataset. Finally, it provides code for efficient inference using the adapted model. The pipeline leverages the Slovene LLM GaMS and similar open-source models, tools for generating domain-specific question-answer pairs using NeMo Curator in combination with the GaMS-Instruct approach, LoRA or QLoRA-based supervised fine-tuning using Hugging Face and NeMo, and vLLM for fast domain-specific question answering inference.

Expected outcome

A step-by-step guide for domain-specific question-answering training, along with scripts for data generation, model training, and inference.

Requirements

Requires several NVIDIA GPUs.

Service output

Pipeline for question answering adaptation, including data generation scripts, model training scripts, and inference code.

Specific limitations

Licensing

MIT



AI development and deployment

Adaptation of a Large Language Model for Domain-Specific Summarization

Responsible institution
University of Ljubljana

Vertical Cross-domain Availability Q3/2027 Type Pipeline

Target users: Industry, Large Enterprises, Scientific Institutions, Public Institutions

Overview

This pipeline enables the adaptation of a large language model (LLM) to generate high-quality, domain-specific summaries for technical or scientific documents, reports, research papers, or industry-specific content. It is designed to enhance the accuracy and relevance of summaries within a specific domain. The pipeline first helps users in preparing data for summarization (OCR of PDF documents) and converting it into a training dataset. Next, it provides the know-how and code for fine-tuning the model with the created dataset. Finally, it provides code for efficient inference using the adapted model. The pipeline leverages the Slovene LLM GaMS (for Slovene-language summarization) and other open VLMs (e.g., Gemma), open-source OCR libraries such as Marker or small open-source VLMs such as Nanonets, supervised fine-tuning or preference tuning using GRPO with LoRA or QLoRA-based fine-tuning, and vLLM for efficient, low-latency inference of domain-specific summarization tasks.

Expected outcome

A step-by-step guide for domain-specific summarization adaptation, along with scripts for data preparation, model training, and inference.

Requirements

Requires several NVIDIA GPUs.

Service output

Pipeline for summarization adaptation, including data preparation scripts, model training scripts, and inference code.

Specific limitations

Licensing

MIT

SLAIF – IZOBRAŽEVANJE IN USPOSABLJANJE



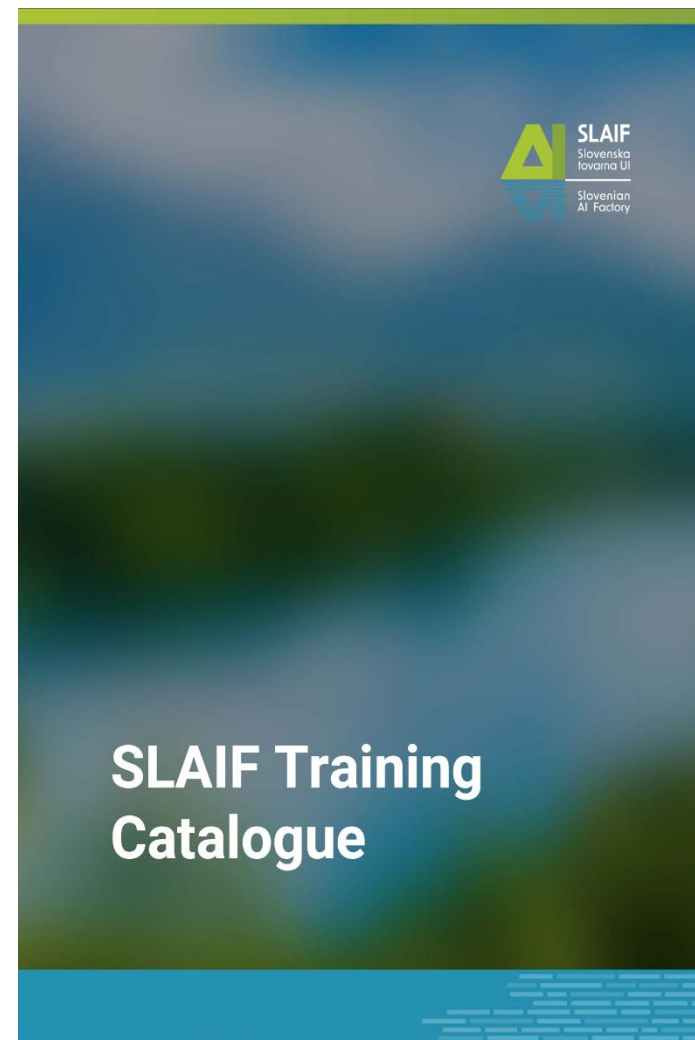
Različne vrste izobraževanj

- Delavnice
- Webinarji
- Izobraževanja z mikro-certifikati

Različne tematike izobraževanj

- UI metode
- Visokozmogljivo računalništvo
- Področja uporabe UI

Različne ciljne skupine



SLAIF Training Catalogue



Large language models and domain-specific AI for the Slovenian agricultural sector

Course provider:

Jožef Stefan Institute (JSI)

Instructors:

Marko Debeljak (JSI) and colleagues

Content: The event aims to provide agricultural stakeholders with a clear, realistic, and application-oriented understanding of large language models (LLMs) and their potential role in Slovenian agriculture. Participants will be introduced to the rationale for focusing on high-quality, domain-specific data and benchmarks rather than individual model versions, and to the role of the national AI infrastructure (SLAIF) in supporting domain- and language-specific AI solutions. A key objective is to foster informed expectations, trust, and early engagement of future users by openly addressing both the capabilities and limitations of current AI technologies in regulated and practice-oriented agricultural contexts.

The event presents current activities related to the development and application of large language models within the Slovenian agricultural domain in the context of SLAIF. It introduces the fundamentals of LLMs in an accessible manner, with a focus on their applicability, risks, and constraints in agriculture. Participants will be presented with a proposed taxonomy of Slovenian agriculture, including key sectors and subdomains, along with example benchmark tasks designed to evaluate model performance in Slovenian. Preliminary benchmark results will be demonstrated to illustrate how existing models perform on selected agricultural use cases. The event further provides a structured forum for validating assumptions about users, use cases, and sectoral coverage, and introduces a dedicated web application for the collection of high-quality question-answer pairs and other expert inputs to support model adaptation and evaluation.

After the event, participants will understand the importance of domain-specific data, benchmarks, and expert validation for the sustainable development of AI solutions in agriculture. They will have a realistic understanding of what current large language models can and cannot do in the Slovenian agricultural context, and how national AI infrastructure supports their adaptation. Participants will be able to critically assess proposed use cases and benchmarks from a practitioner's perspective and will be equipped to contribute expert knowledge, feedback, and structured data inputs for the continued development and evaluation of AI tools for Slovenian agriculture.

Prerequisites: No technical prerequisites are required. The event is intended for practitioners and experts with domain knowledge in agriculture; prior experience with AI technologies is not required.

Target audience: Agricultural advisers and extension services; Farmers and agricultural practitioners; Agri-food stakeholders and industry representatives; Researchers and developers working on AI for agriculture; Public-sector stakeholders and policy makers

Language: Slovenian; English

Tentative date: Autumn 2026

Venue: ISI Ljubljana/Murska Sobota

Format: Lectures; Hands-on sessions

Duration: 6h

No. of ECTS credits: N/A

SLAIF: VERTIKALNE STORITVE ZA UPORABO UI NA RAZLIČNIH PODROČJIH

Zeleni prehod

- *Kmetijstvo, podnebje, okolje, energija, pametna proizvodnja, ...*
- *Multimodalni temeljni modeli za daljinsko zaznavanje, relevantni za kmetijstvo in okolje*

Zdravje in biotehnologija

- *Diagnostika, razvoj zdravil, prehrana, personalizirana medicina*
- *Temeljni modeli za relevantne senzorske podatke (npr. nosljive naprave)*

Digitalna družba

- *Jezikovne tehnologije (veliki jezikovni modeli za slovenščino)*
- *Generativna UI za medijski sektor, javno upravo, izobraževanje*

UI za znanost

- *Vede o živem, znanosti o materialih, okoljske znanosti*
- *Interdisciplinarne raziskave*

ZELENI PREHOD: Določanje rabe tal iz satelitskih posnetkov

Na IJS razvijamo metode za polnadzorovano razvrščanje satelitskih slik

- Satelitski posnetki programa Copernicus so obsežni, vendar je označenih podatkov malo in njihovo pridobivanje je drago
- Oznake rabe tal so naravno večoznačne in hierarhične (npr. Corine Land Cover)

SLAIF + Copernicus Data Space =

Računska moč & Delotoki + Velike količine satelitskih posnetkov

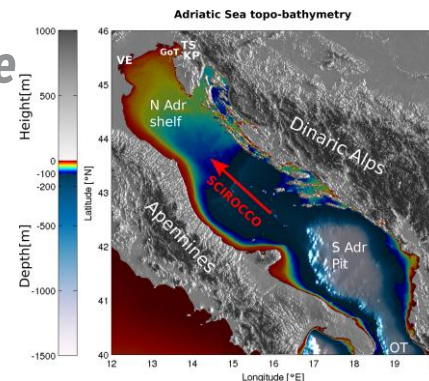
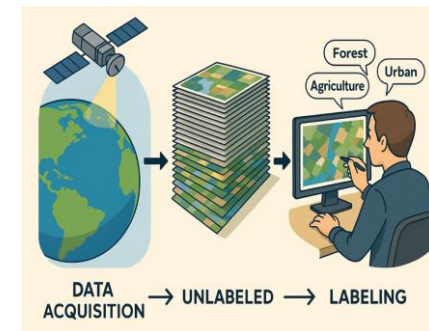
Omogoča veliko različnih storitev za kmetijstvo in okolje

Podnebje in okolje

Podnebne spremembe povzročajo dvig morske gladine in bolj ekstremne vremenske pojave

- Do leta 2100, +30 cm to +1 m (IPCC)
- Pri +1 m, deli Pirana bi lahko bili poplavljeni dvakrat dnevno

HIDRA (razvita na Univerzi v Ljubljani – FRI in ARSO) uporablja umetno inteligenco za napovedovanje višine morske gladine in tveganja obalnih poplav.



ZDRAVJE IN BIOTEHNOLOGIJA:

Prilagajanje VJM za reševanje nalog iz nutricionistike

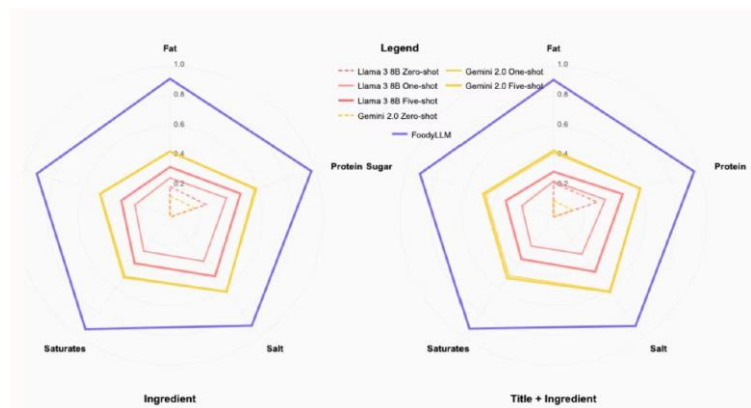


Naloge:

- Semafor zdrave prehrane: zdrava (zeleno), zmerno (oranžno), omeji uživanje (rdeča)
- Ocenjevanje vsebnosti hranil (sol, sladkor, maščobe, ...)

Pristop: Uglasovanje modela Llama3 z lastnimi podatki

Izid: Uglaseni modeli delajo veliko bolje (spodaj performance pri napovedovanju vsebnosti hranil)



DIGITALNA DRUŽBA: VJM za slovenščino

GaMS – slovenski generativni jezikovni model

- *GaMS = Generative Model for Slovene – generativni model za slovenščino*
- *Zasnovan za zapolnjevanje vrzeli, ki jih puščajo globalni modeli, kjer je slovenščini*
- *Odprtokoden in razvit v okviru projekta PoVeJMo (UL CJVT, UL FRI, INZ, ZRC SAŽU)*
- *Načrtovane ali že dostopne so različne velikosti modelov (od 1B do več kot 27B parametrov)*
- *Veliki jezikovni model, prilagojen slovenskemu jeziku in kulturi*

Generativni model
za slovenščino
Generative Model
for Slovene

GaMS: “

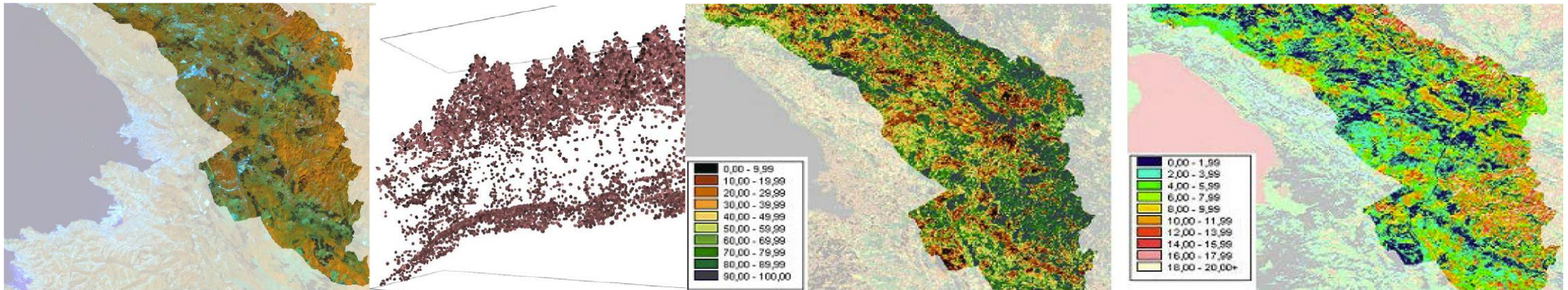
Specializirani VJM za področje telekomunikacij

IJS skupaj s Telekomom Slovenije sodeluje pri razvoju evropskega velikega jezikovnega modela (LLM), prilagojenega potrebam telekomunikacijske industrije: Večjezična podpora, vključno s slovenščino, francoščino, italijanščino, nizozemščino in poljščino.

Ključni primeri uporabe:

- *Prilagoditev domeni (globoko razumevanje telekomunikacijskega znanja)*
- *Analiza klicev (samodejno povzemanje, analiza sentimenta, določanje nujnosti klica)*
- *Zasebnost (napredna anonimizacija in psevdonimizacija uporabnikov)*

- Spremljanje stanja okolja (env. monitoring) z daljinskim zaznavanjem
 - Strojno učenje iz kombinacije 2D (satelitskih) in 3D (LIDAR) posnetkov
 - Ocena višine in gostote gozdov (tudi biomase)

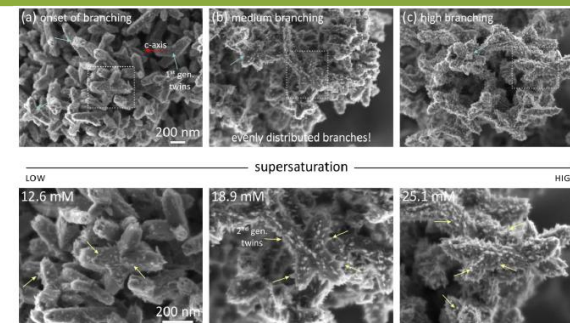


- Ocena požarne ogroženosti naravnega okolja
- Ocena učinkov podnebnih sprememb (npr. spremenjenih vzorcev padavin) in načrtovanje ustreznih ukrepov

UI V ZNANOSTI O MATERIALIH

- **Materiali so izjemnega pomena za industrijo**
 - So ozko grlo zelenega prehoda
 - Določajo zmogljivost, stroške in trajnostnost industrijskih izdelkov
 - Shranjevanje/pretvorba energije; Proizvodnja/mobilnost; Elektronika/senzorji
- **Uporaba UI lahko bistveno skrajša cikel načrtovanja novih materialov**
 - Usmerjanje eksperimentov in simulacij
 - Samovozeči laboratoriji

- **Podatki**
 - Parametri sinteze materialov (sestava, temperatura, čas obdelave)
- **Kvantne lastnosti**
 - Rezultati simulacij
- **Zgradba materialov (slike)**
- **Lastnosti materialov**
 - Električna prevodnost
 - Katalitična aktivnost
 - Magnetne lastnosti
 - Anti-korozivne lastnosti



Pristop

- Kombinacija strojnega učenja in optimizacije za načrtovanje novih materialov
- Uporaba strojnega učenja za izgradnjo modela, ki napoveduje lastnosti materialov iz parametrov načrtovanja/sinteze.
- Uporaba naučenega modela v kombinaciji z optimizacijskim pristopom za iskanje novega načrta materiala, ki dosega želene lastnosti materialov

Primer iz IJS: Načrtovanje penjenih stekel s kombinacijo SU in optimizacije

- Poceni, trpežen, nestrupeni, ognjevarni in izolacijski material
- Proizvodnja v velikem obsegu – pogosto iz recikliranega odpadnega stekla
- Lastnosti za optimizacijo: gostota, poroznost
- Novi materiali, sintetizirani na podlagi predlogov UI, so v laboratoriju želene lastnosti



Hvala za udeležbo!



REPUBLIKA SLOVENIJA
MINISTRSTVO ZA VISOKO ŠOLSTVO,
ZNANOST IN INOVACIJE

SLAIF: Slovenska tovarna umetne inteligence je prejela financiranje Skupnega podjetja za evropsko visokozmogljivo računalništvo v okviru sporazuma o nepovratnih sredstvih št. 101254461. Projekt financirata Ministrstvo za visoko šolstvo, znanost in inovacije Republike Slovenije ter program Horizon Europe Evropske unije.

SLAIF: Slovenian AI Factory has received funding by the European High Performance Computing Joint Undertaking under the grant agreement no. 101254461. The project is funded by the Ministry of Higher Education, Science and Innovation of Republic of Slovenia and by Horizon Europe Programme of the European Union.

Naslednja predstavitev:

HIDRA – Model globokega učenja za napovedovanje obalnih poplav

prof. dr. Danijel Skočaj,
Univerza v Ljubljani, Fakulteta za računalništvo in informatiko



FRI

Fakulteta za računalništvo
in informatiko

ViCOS
sualgnitive
ystemslab

HIDRA – model globokega učenja za napovedovanje obalnih poplav

Danijel Skočaj

(Matej Kristan)

Univerza v Ljubljani

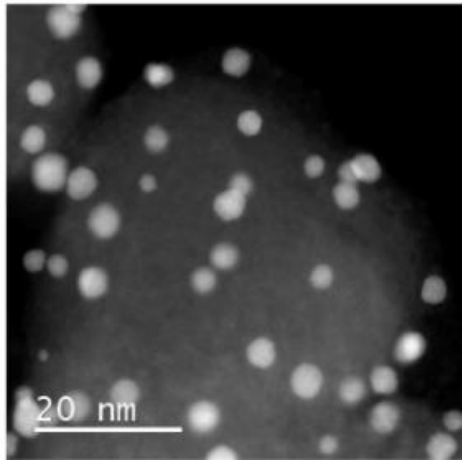
Fakulteta za računalništvo in informatiko

H I D R A

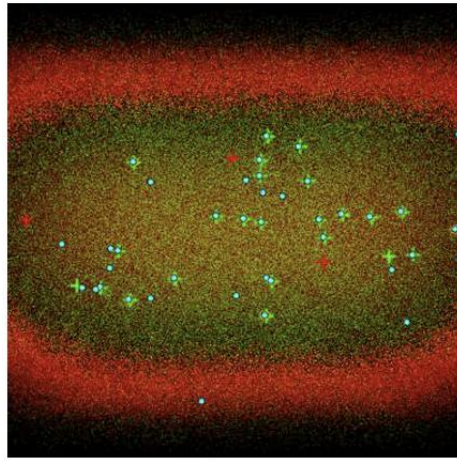
Evolucija.AI , 11. 9. 2026

Računalniški vid za znanost

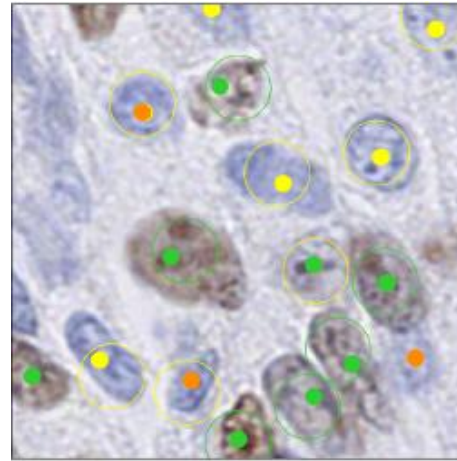
- Neposredno uporabljen v mnogih znanostih
- Procesiranje vizualne informacije



Materiali



Astrofizika



Medicina

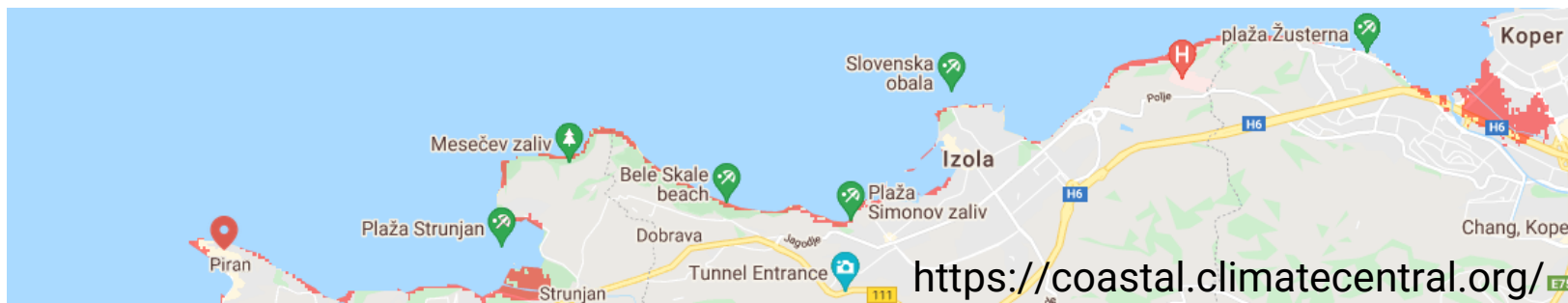
idr.



Opazovanje Zemlje

Napovedovanje obalnih poplav

- Globalno segrevanje povzroča **dvig gladine oceanov** in **večjo vremensko spremenljivost**
- Do konca 2100: **30cm – 1m** dvig gladine
- Slovenija @1m: **Piran pod vodo 2x na dan**
- Obalne **poplave bodo vse pogostejše**
- **Nujnost napovedovanja** več dni v prihodnost
- Napovedovanje višine morske gladine



Družina metod HIDRA

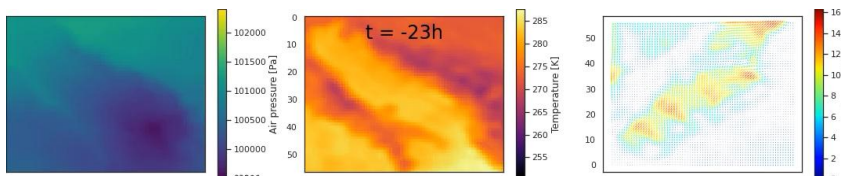
- Globoki model za napovedovanje dinamike morja v sodelovanju z ARSO

(doc. ddr. Matjaž Ličer)

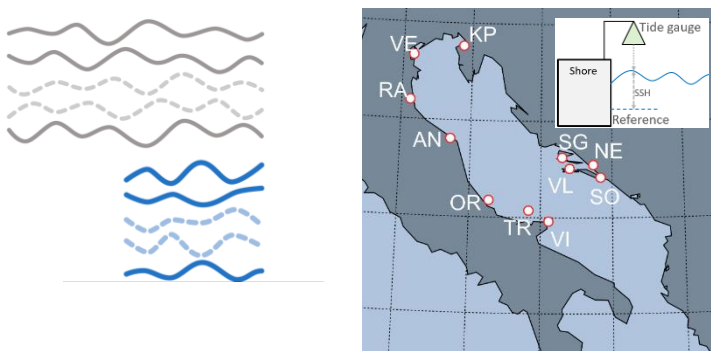


ARSO METEO
Slovenian Environment Agency

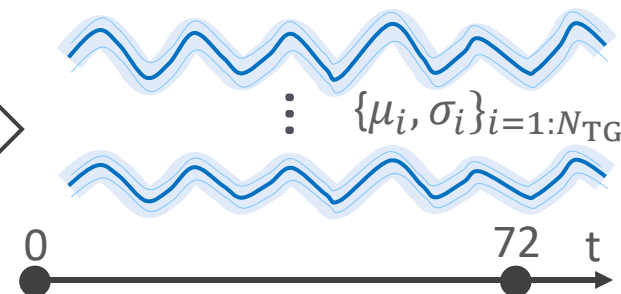
Atmosferske spremenljivke (ECMWF):



Pretekle meritve postaj + plima



Napoved višin morja za 72h



- Prekaša NEMO v natančnosti in pol milijonkrat hitrejša (16 ms za napoved)

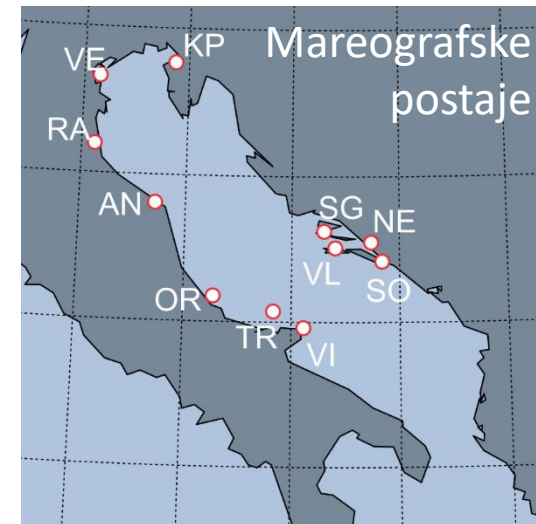
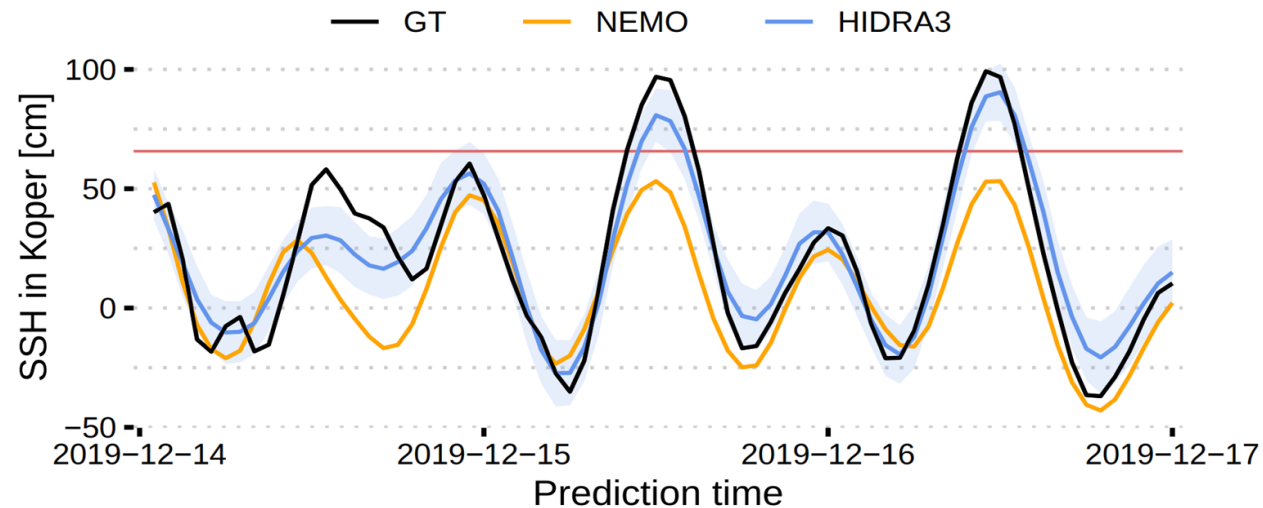
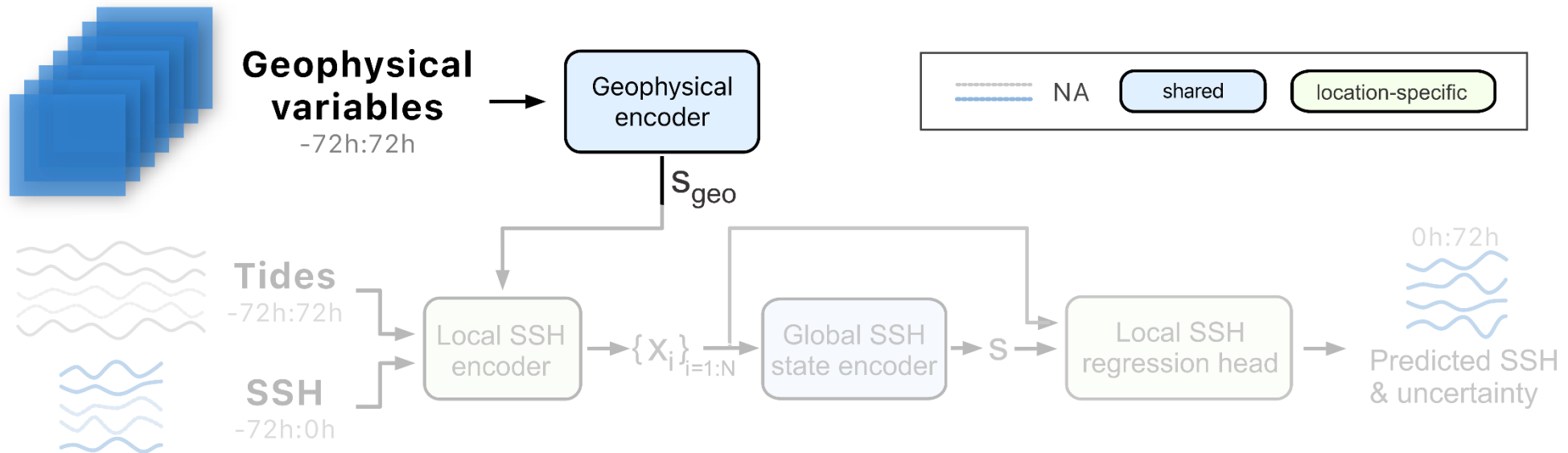
[1] Žust, Fettich, Kristan, Ličer, HIDRA 1.0: deep-learning-based ensemble sea level forecasting in the northern Adriatic, Geosci. Model Dev., 2021

[2] Rus, Fettich, Kristan, Ličer, HIDRA2: deep-learning ensemble storm surge forecasting in the presence of seiches – the case of Northern Adriatic, Geosci. Model Dev., 2023

[3] Rus, Mihanović, Ličer, Kristan, HIDRA3: a robust deep-learning model for multi-point ensemble sea level forecasting, Geosci. Model Dev., 2025

[4] Rus, Ličer, Kristan, HIDRA-D: deep-learning model for dense sea level forecasting using sparse altimetry and tide gauge data, Geosci. Model Dev., 2026

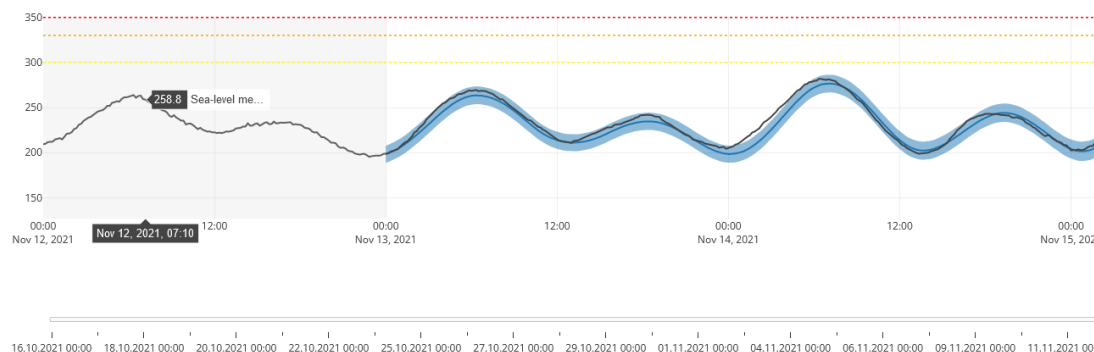
HIDRA3 ^[1]



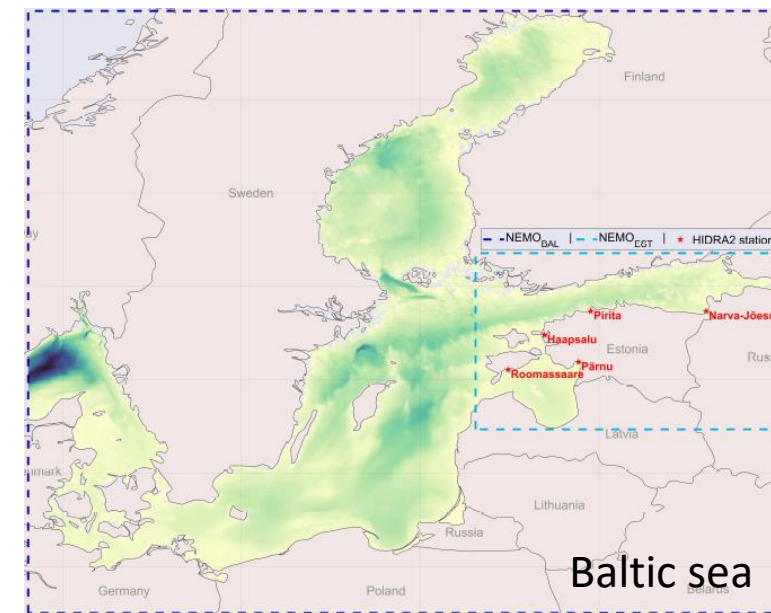
^[1]Rus et al. HIDRA3: a robust deep-learning model for multi-point ensemble sea level forecasting, GMD 2025

HIDRA že prenesena v prakso

- HIDRA2 integrirana v prognostični paket ECMWF^[1] (Destination Earth digital twin)
- HIDRA3^[3] v operativni uporabi:
 - V Sloveniji (ARSO), Estonija (4 točke), Danska (>50 točk)
- HIDRA3^[3] trenutno javno testirana:



<https://bit.ly/hidra-forecast>



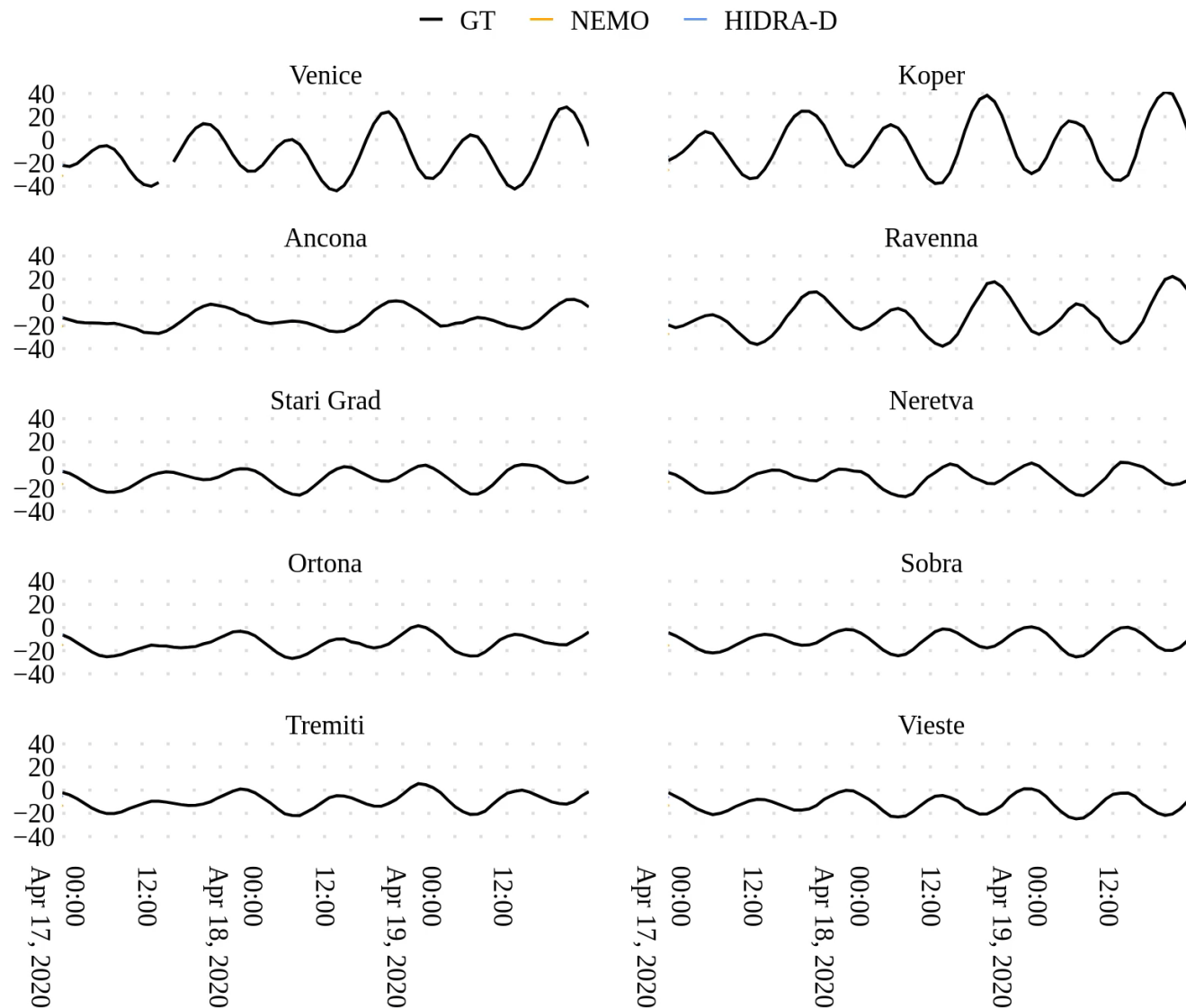
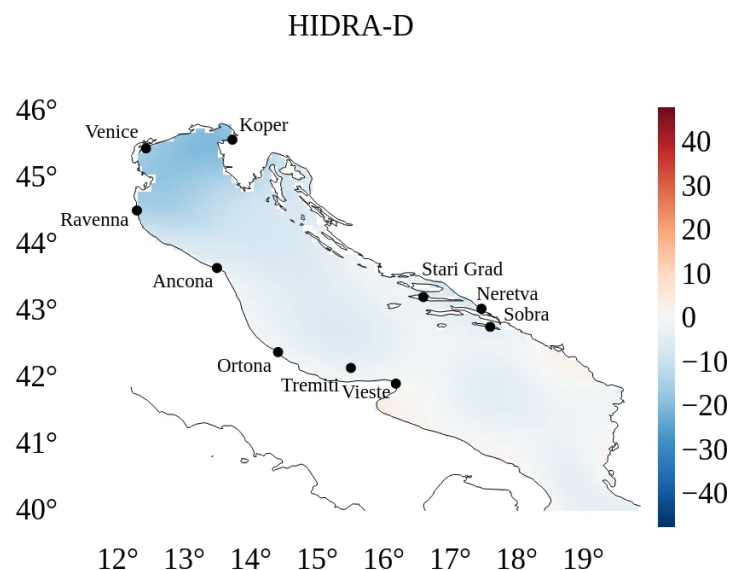
^[1]ECMWF (European Centre for Medium-Range Weather Forecasts)

^[2] Rus, Fettich, Kristan, Ličer, HIDRA2, Geosci. Model Dev., 2023

^[3] Rus, Mihanović, Ličer, Kristan, HIDRA3, Geosci. Model Dev., 2025

HIDRA-D ^[1] – gosta napoved

2020-04-17 00:00, T+0

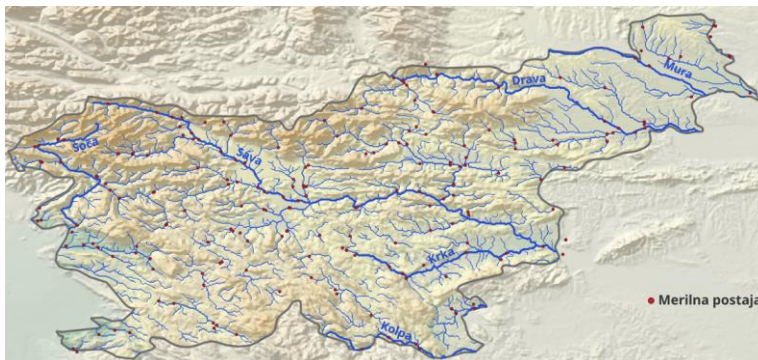
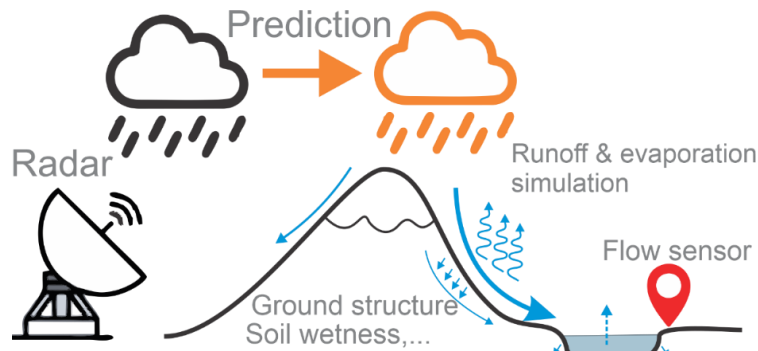


Each plot was made with HIDRA-D trained WITHOUT the corresponding tide gauge.

^[1] Rus, Ličer, Kristan, HIDRA-D: deep-learning model for dense sea level forecasting using sparse altimetry and tide gauge data, GMD 2026

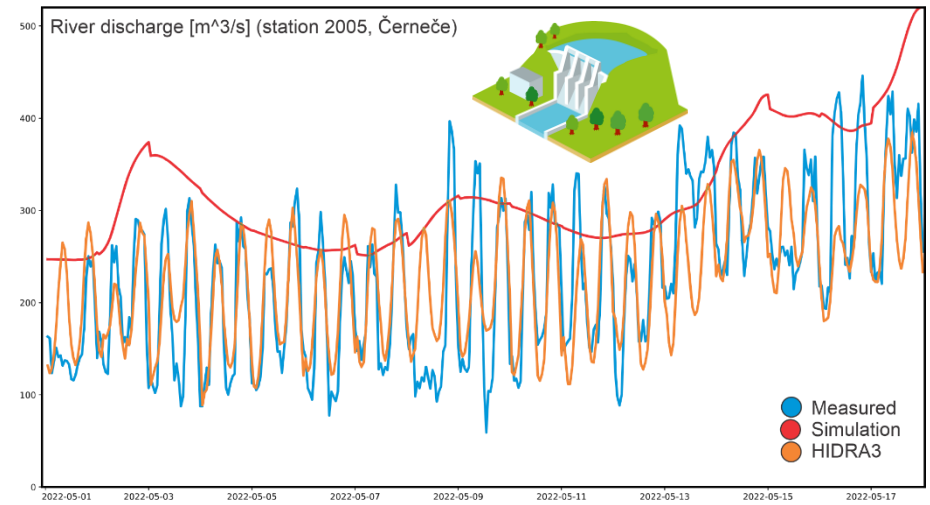
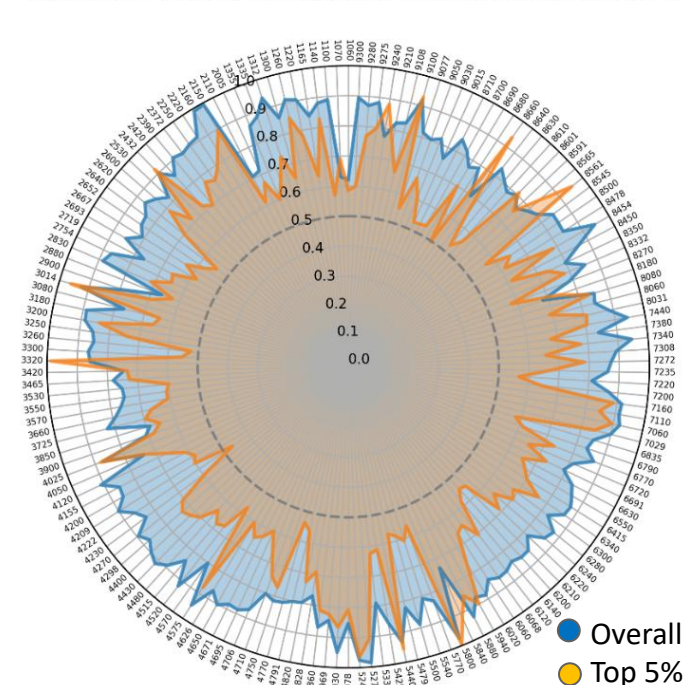
HIDRA za rečne pretoke in poplave

- Testirana na 216 rečnih postajah
- Bistveno prekaša najnaprednejši ARSOv model
- Modelira tudi nevremenske vplive



ARSO: dr. Sašo Petan, Jure Cedilnik, doc. ddr. Matjaž Ličer

Proportion of observations where HIDRA outperforms simulation



Zahvala

Interdisciplinarni pristop



Lojze Žust
(FRI-UL)



Marko Rus
(FRI-UL)



Matej Kristan
(FRI-UL)



Matjaž Ličer
(ARSO/NIB)



Anja Fettich
(ARSO)



Jure Cedilnik
(ARSO)



Sašo Petan
(ARSO)



Petja Furlan
(FRI-UL)



FRI

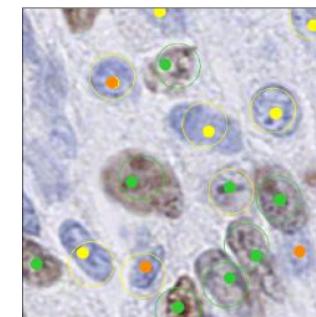
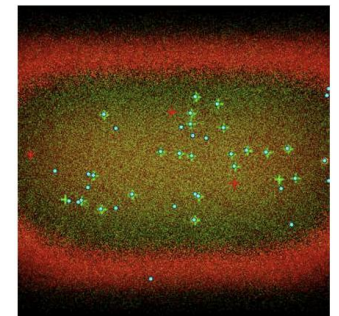
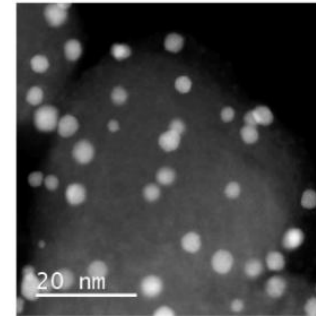
Fakulteta za računalništvo
in informatiko



ARSO METEO
Slovenian Environment Agency

Zaključek

- HIDRA: natančna, robustna in izjemno hitra napoved
- Podatkovni modeli lahko učinkovito dopolnijo fizikalne modele
- Učeče se rešitve temelječe na podatkih
- UI/SU/RV – ključne omogočitvene tehnologije prihodnosti
- Izjemno široka uporabnost, interdisciplinarnost
- Robustnost, nove priložnosti, novi izzivi
- Izreden pospešek znanosti
- Sodelovanje akademije in industrije
- Umetna inteligenca – tehnologija prihodnosti!



Naslednja predstavitev:

Uporaba umetne inteligence pri razvoju zdravil:

Odkrivanje učinkovin in njihovih bioloških tarč

izr. prof. dr. Marko Jukić,

Univerza v Mariboru, Fakulteta za kemijo in kemijsko tehnologijo;

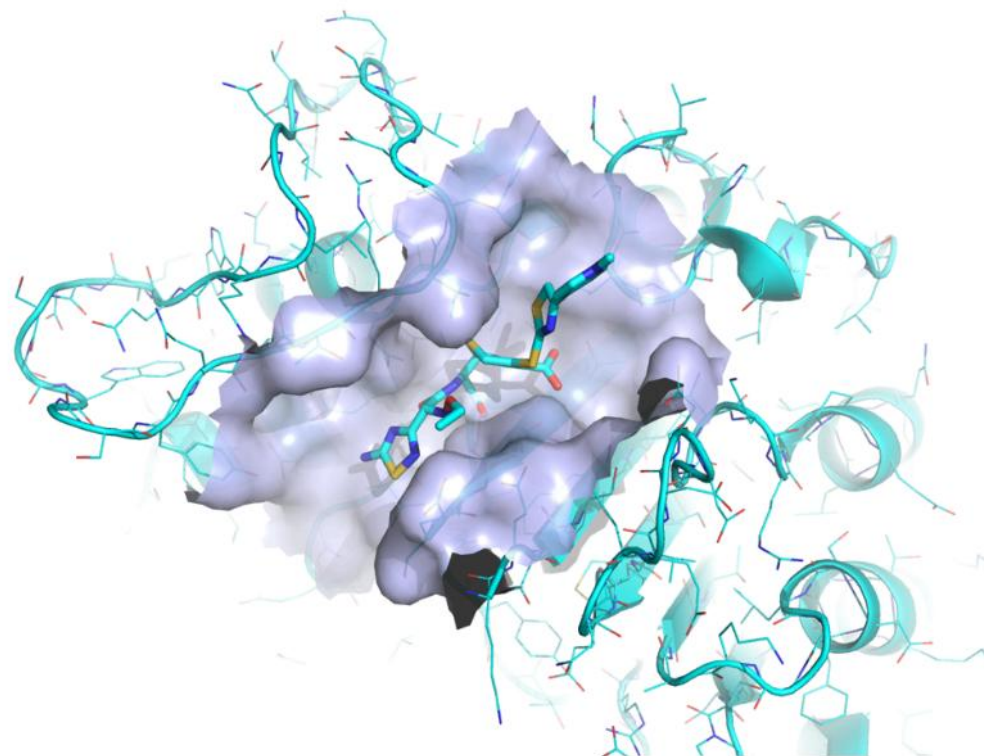
Umetna inteligenca pri razvoju zdravil:

od prepoznavanja potencialnih učinkovin do njihovih bioloških tarč

evolucija.ai

Marko Jukič

11.9.26



active compound from a registered medicine (drug) in its active site



Fakulteta za kemijo
in kemijsko tehnologijo

Razvoj učinkovin je predolg, predrag in preveč tvegan

- Od zamisli do registrirane učinkovine mine **več kot desetletje**.
- Velik del stroška nastane, **preden** sploh vemo, ali je tarča prava.
- Eksperimentalno določanje vezavnih mest je počasno, drago in za mnoge tarče neizvedljivo.
- Računske metode teh vprašanj ne odpravijo — lahko pa **dramatično zožijo prostor**, ki ga je treba eksperimentalno preveriti.

Tri ozka grla, ki jih naslavljamo

Terapevtske tarče

Katere tarče moduliramo z namenom poseganja v patologijo bolezni?

Vezava učinkovin?

Kam na tarčah se vežejo učinkovine – naj si gre za velike ali male molekule. Ali lahko povzročajo **stranske učinke** – **off-target effects**?

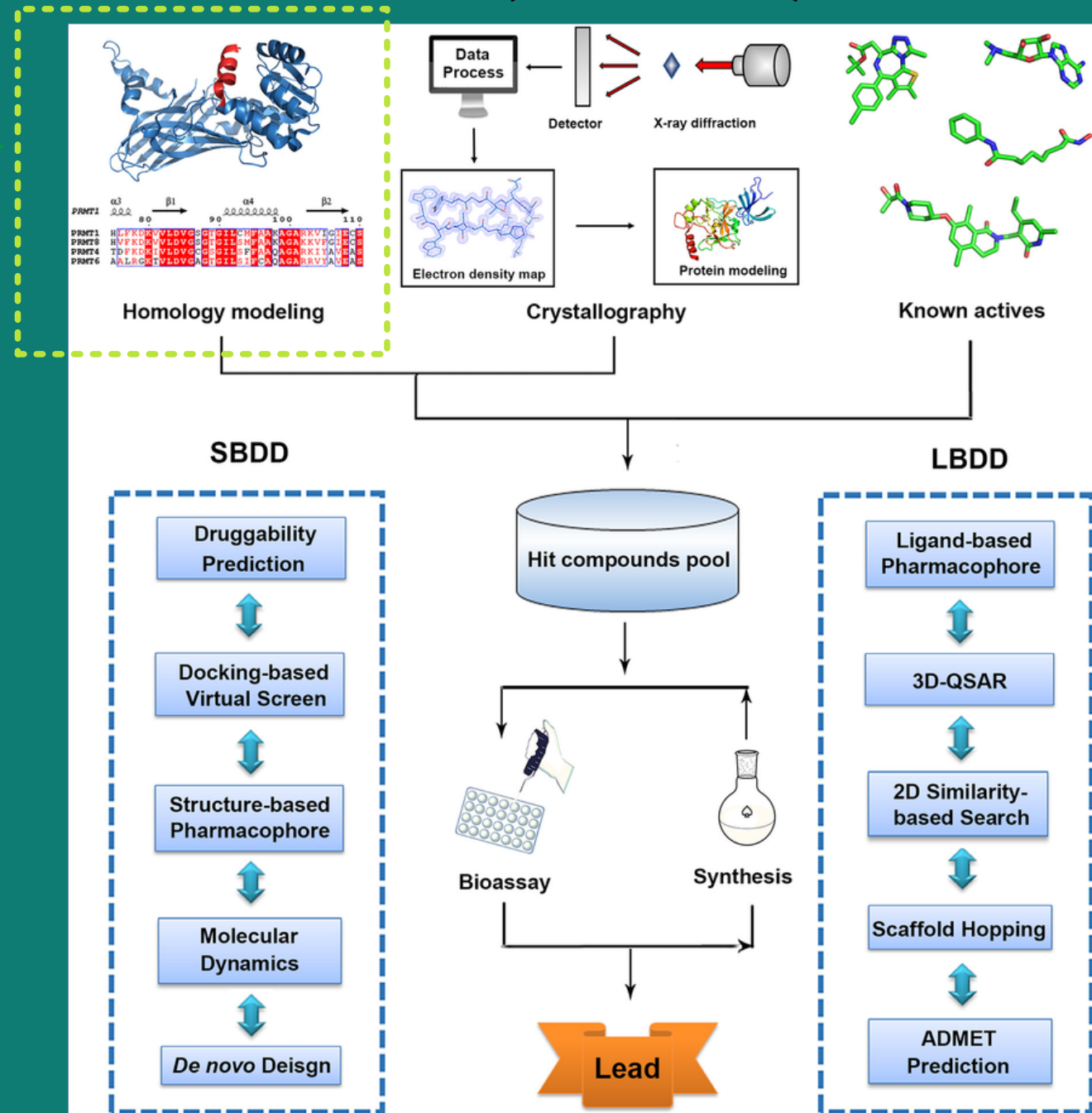
Zadetki

Kako sploh najdemo zadetke – molekule, ki interagirajo s tarčami?

PRVI DEL

Tarče

Kako do strukture tarč ?

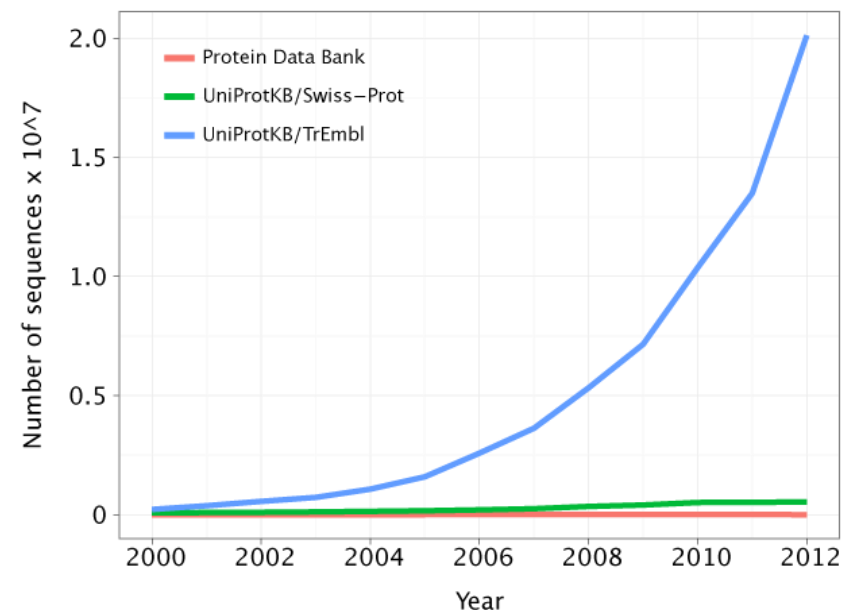
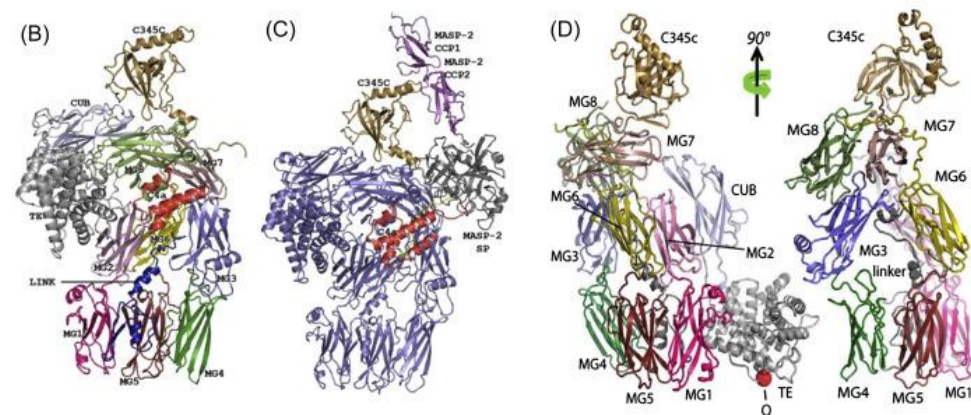
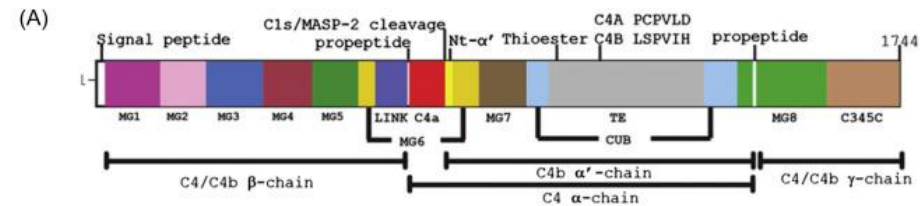


Kaj je tarča za zdravilo

- **Tarča** je biološka molekula — najpogosteje protein — katere delovanje želimo spremeniti.
- Učinkovina se nanjo veže in jo **zavre, aktivira... (modulira, označi za razgradnjo, ...)**
- Da bi tako vezavo načrtovali, moramo poznati **tridimenzionalno zgradbo** tarče.
- Izbira tarče je torej odločitev, ki določi usodo celotnega projekta.

Eksperimentalno določanje struktur je počasno

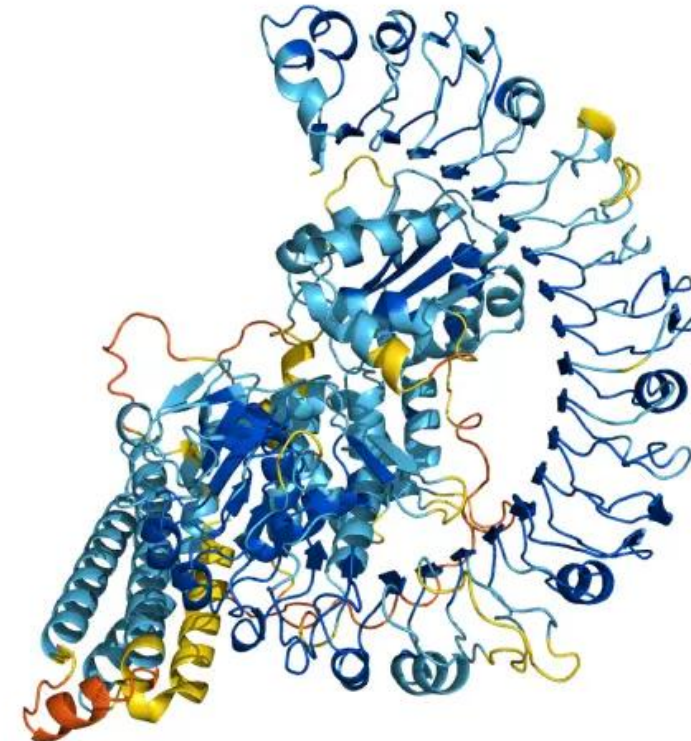
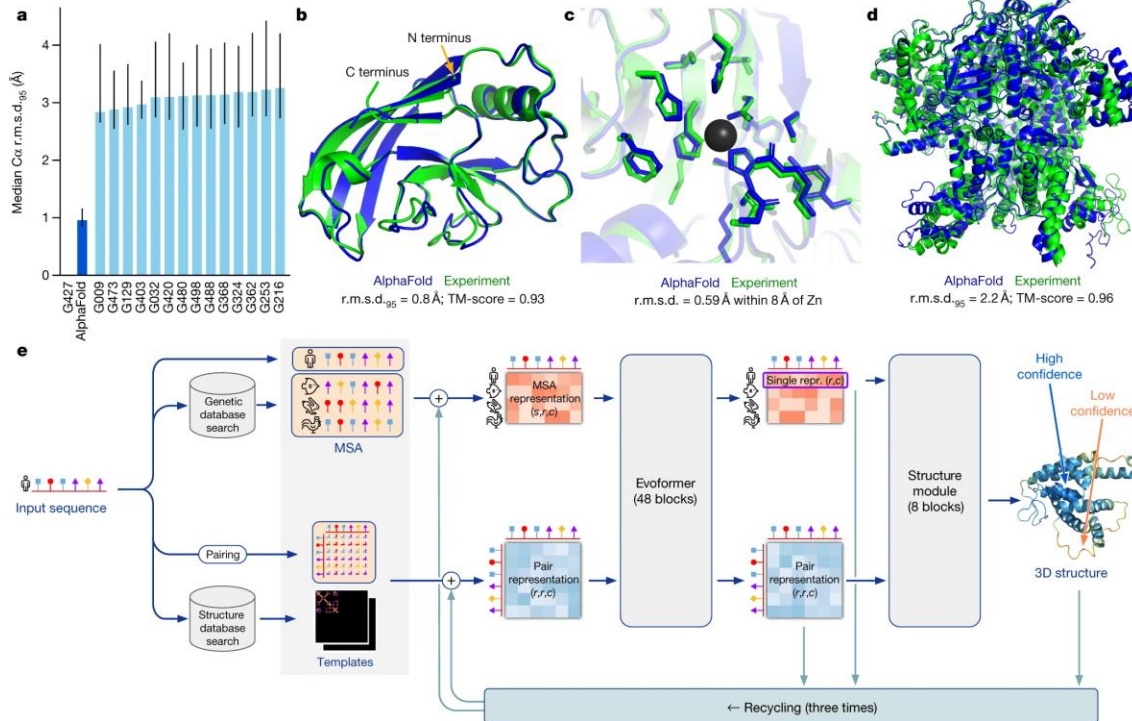
- **Rentgenska kristalografija** zahteva kristal, logistično in časovno zahtevna.
- **Krio-elektronska mikroskopija** je zmogljiva, a draga in zahtevna.
- **NMR** je omejen na manjše proteine, časovno zahteven.
- Rezultat: Protein Data Bank vsebuje **nekaj sto tisoč** struktur, znanih proteinskih zaporedij pa je **na stotine milijonov**.
- Za veliko večino proteinov strukture preprosto **ni bilo**.



Napoved strukture iz zaporedja, ključni napredek s pomočjo AI

- **AlphaFold 2** (DeepMind, 2021)
- Vhod je **aminokislinsko zaporedje**; izhod je tridimenzionalni model s koordinatami.
- **AlphaFold DB** danes vsebuje **več kot 200 milijonov** napovedanih struktur — praktično celoten znani proteinski svet.
- **AlphaFold 3** (2024) napoveduje tudi komplekse z ligandi, nukleinskimi kislinami in ioni.
- *Nobelova nagrada za kemijo 2024 je šla Bakerju, Hassabisu in Jumperju prav za to področje.*

Napovedana struktura ni eksperiment. Je hipoteza — a hipoteza, ki jo lahko dobimo za skoraj vsak protein, takoj in brezplačno.

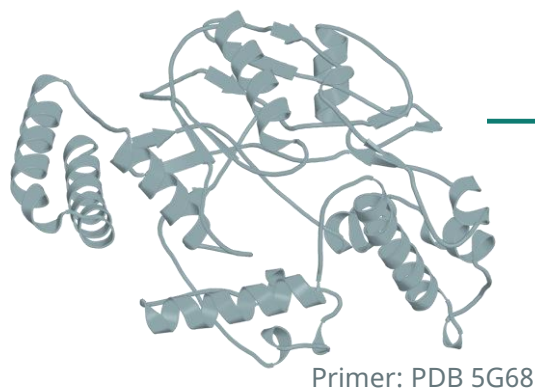


Od strukture do gibljivosti proteinske tarče

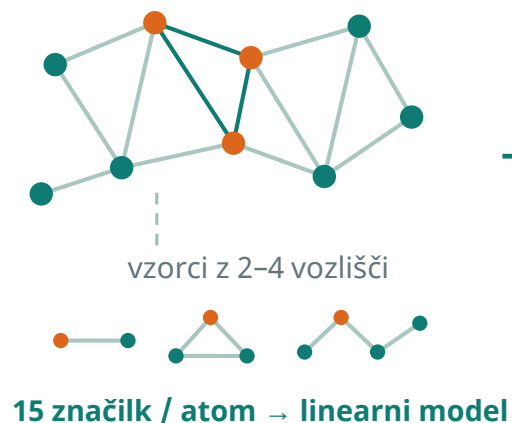
Lokalno atomsko okolje kot informacija za strukturno načrtovanje učinkovin.

1 3D-struktura

Vhod: atomske koordinate

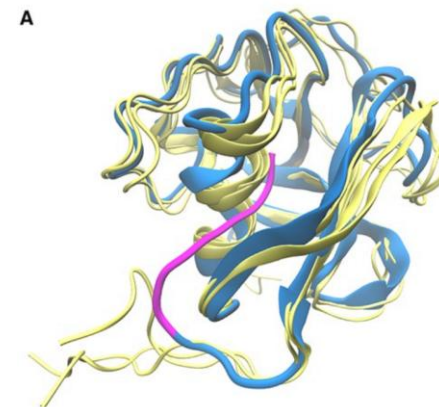
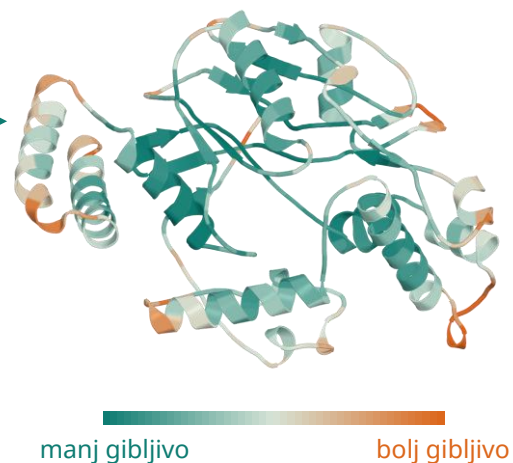


2 Graf atomskih stikov



3 Lokalna gibljivost

Izhod: normalizirani RMSF



POMEN ZA NAČRTOVANJE

Upoštevamo gibljivost tarče

- 01 Presoja gibljivosti vezavne okolice
- 02 Izbor struktur za sidranje in nadaljnje simulacije

PREDLAGANA UPORABA

Ob napovedi žepov in validaciji tarče.

Podpora strukturni opredelitvi potencialnih tarč; sama napoved gibljivosti ne identificira ali potrdi nove tarče.

2026 · FastProtFlex

$\rho \approx 0,79$

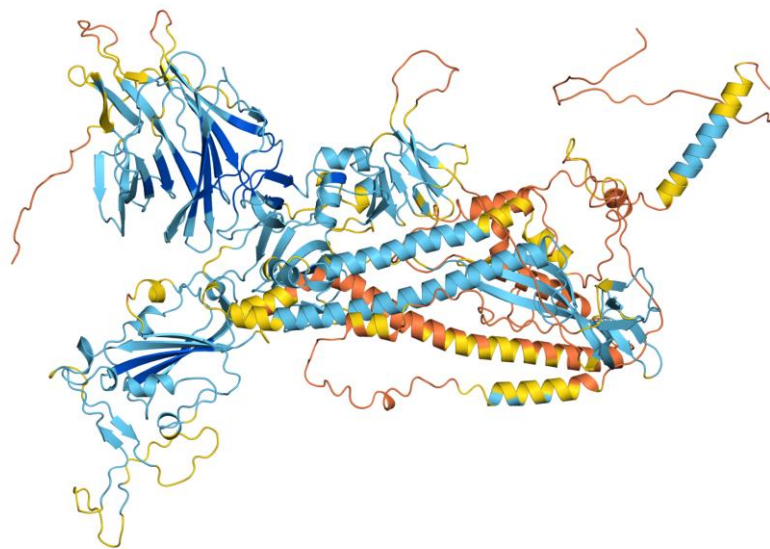
≈ 10 s

20.000 atomov



Struktura še ni vezavno mesto

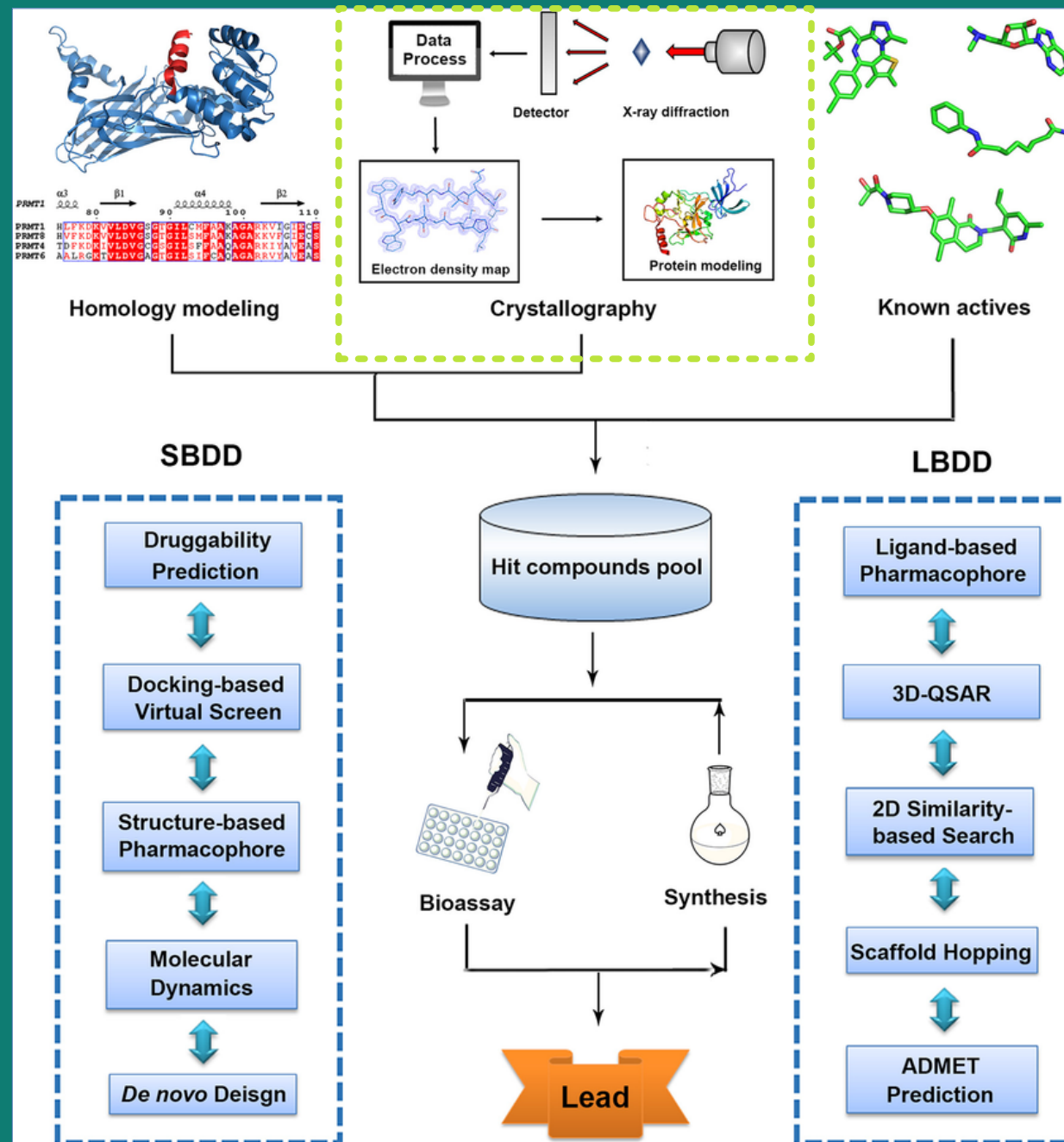
AlphaFold nam je dal *koordinate za tako rekoč vsak protein*. Ni pa nam povedal, *kje na tej površini se lahko veže učinkovina*.



DRUGI DEL

Vezavna mesta

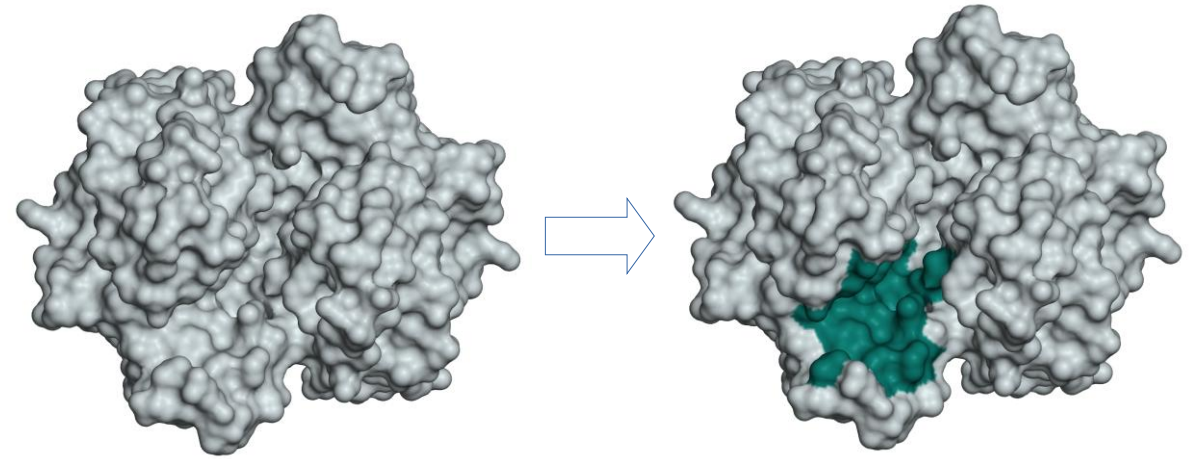
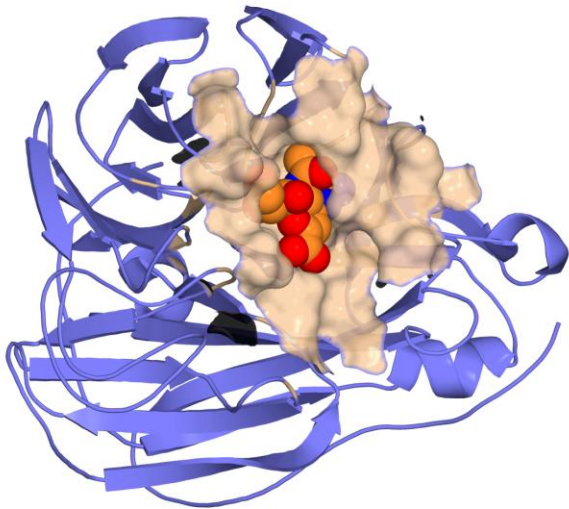
Kje na proteinu se učinkovina sploh lahko veže?



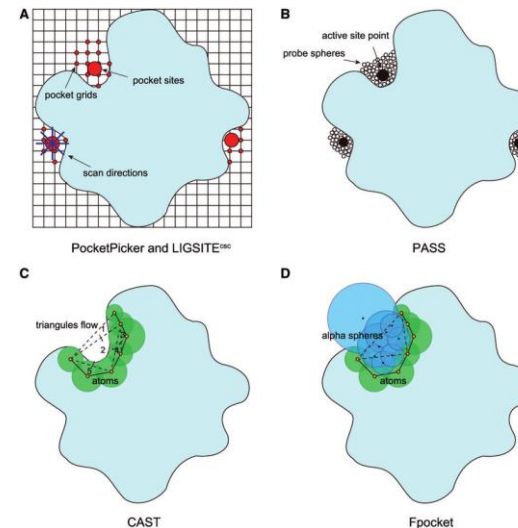
PROBLEM

Vsaka točka na površini je enako verjeten kandidat

- Površina povprečnega proteina lahko obsega **več tisoč kvadratnih angstromov**.
- Brez dodatne informacije je vsako mesto na površini verjetno za vezavo ligandov
- Eksperimentalno preverjanje vsakega mesta je nemogoče.
- Potrebujemo metodo, ki **oceni kje se lahko vežejo učinkovine**



obarvani so ostanki v stiku z ligandom. Naloga metode je najti prav to območje — brez liganda.

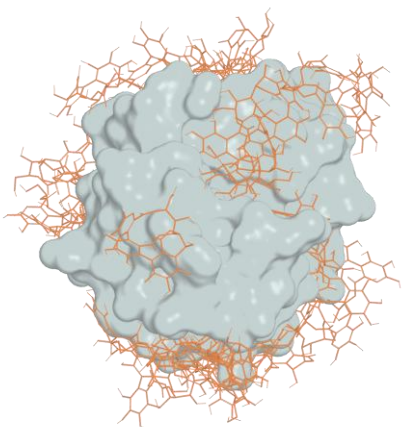


Sidranje na celotni površini razkrije prednostna vezavna območja

Primer: Lizocim + polifenolne probe:
ponovljene poze povežemo s proteinskimi stiki in podatki NMR.

1 Sidranje po površini

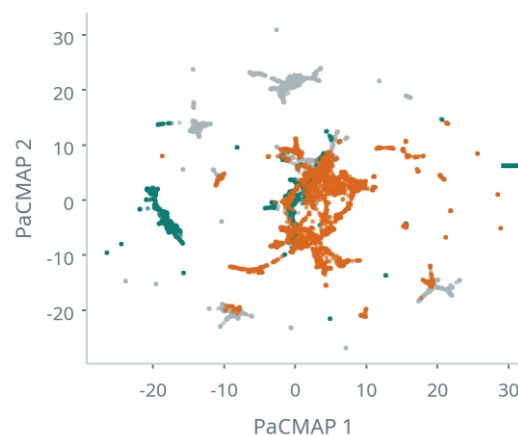
Veskalin · prikazan izbor poz



Celotna površina · CmDock

2 Profili stikov in gruče

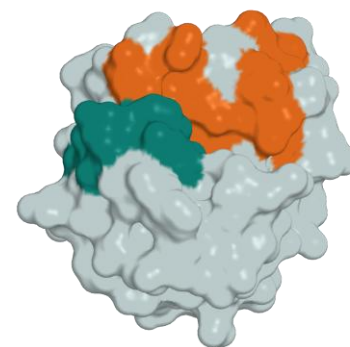
Interakcijski prstni odtisi (IFP)



PaCMAP → DBSCAN

3 Primerjava z NMR

Podobmočji glavne glavnega vezavnega mesta



● distalno

● proksimalno

GLAVNA UGOTOVITEV

Vezava se osredotoči v NMR podprta mesta!

Usmeritve pri nadaljnjem sidranju!

Delež poz ni afiniteta.

PREVERJANJE OBSEGA VZORČENJA

100.000 + sidranj

Identifikacija sekundarnih vezavnih mest !

EKSPERIMENTALNA OPORA

NMR: Trp63 in Trp108

Družine mest interpretiramo s perturbacijami kemijskih premikov.

Barve v projekciji povezujejo poze z območji NMR; sidranje predlaga vezavna mesta, NMR podpira njihovo interpretacijo.

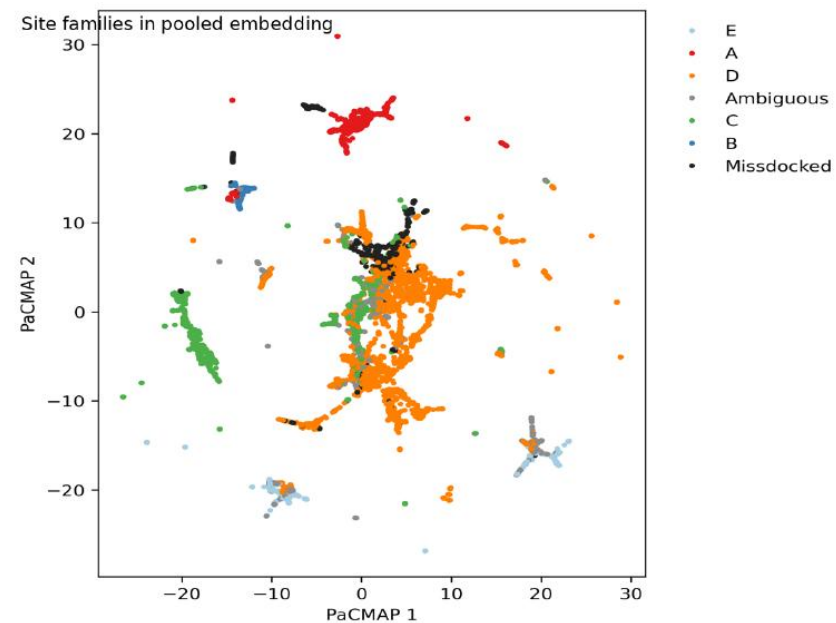
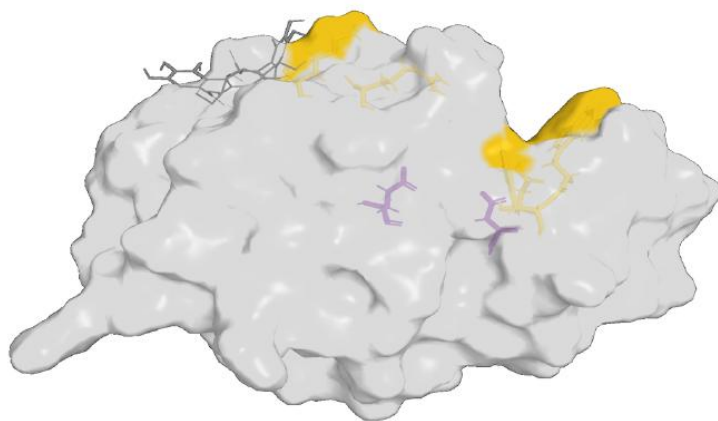


Univerza v Mariboru

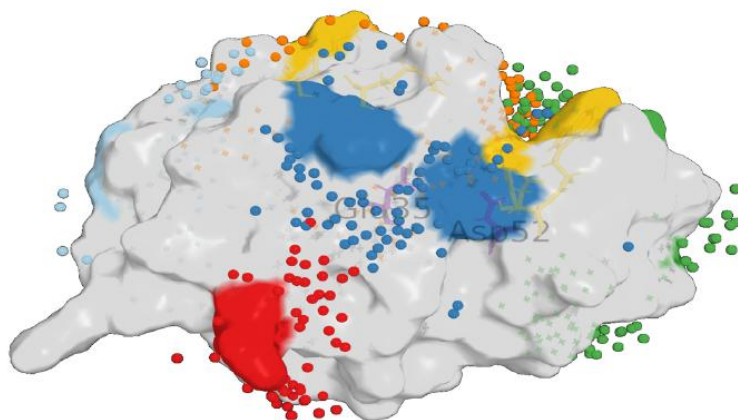
Fakulteta za kemijo
in kemijsko tehnologijo

AI-podprta identifikacija mest

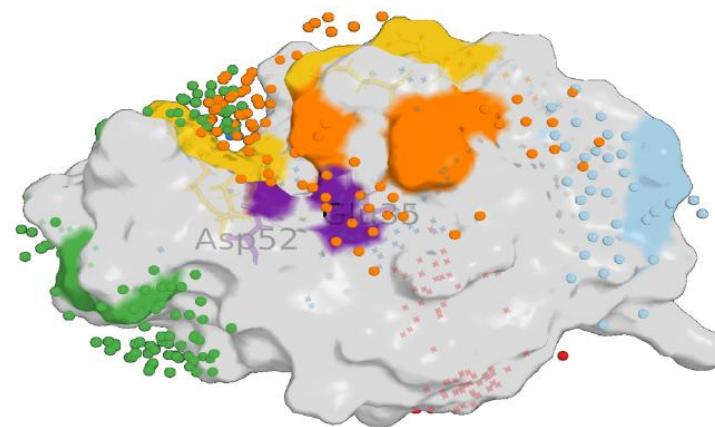
Sampled docked conformers from the 10k ensemble



Protein surface FRONT



Protein surface BACK (y 180)

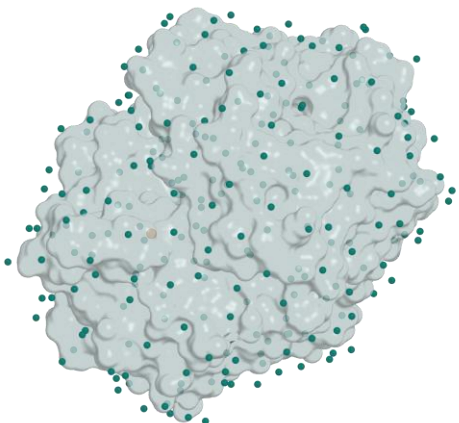


Vsako površinsko točko opišemo z lokalnim grafom

Najprej zgradimo enako predstavitev vseh točk, nato model naučimo prepoznati vezavno okolico.

1 Vzorčimo površino

Protein brez liganda · 10GS



Površinske točke: razmik $\geq 3,5 \text{ \AA}$

2 Zgradimo lokalni graf

45 najbližjih aminokislinskih ostankov



Povezava: razdalja $\text{Ca-Ca} \leq 8 \text{ \AA}$

3 Ocenimo

Identiteta ostankov
in prostorske razdalje

GINE-RBF

Ocena vezavnega okolja

POVRŠINSKA ZBIRKA

1,72 milijona grafov

6.730 proteinskih struktur

Značilka, ki nosi največ signala, je natanko tista, ki je iz zaporedja ni mogoče rekonstruirati

KLJUČ ZASNOVE

Uspešna klasifikacija vezavnih mest!

Znane ligande uporabimo le za oznake učnih primerov.

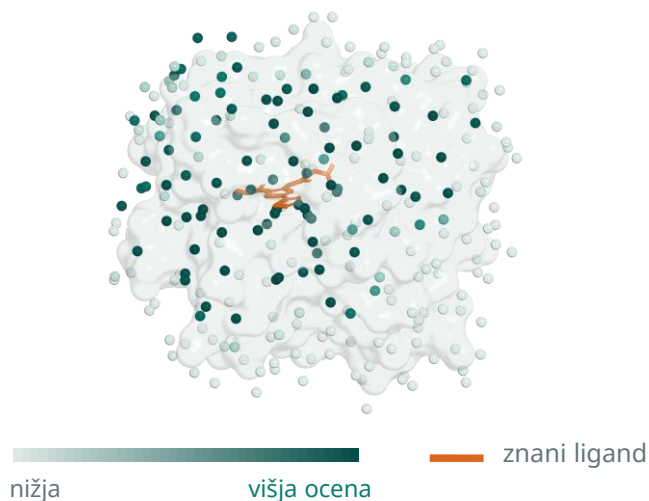
- vezavna okolica
- oddaljena površina
- dvoumno območje

Model uvrsti vezavno okolico med prve kandidate

Vsaka površinska točka dobi oceno; nato vse točke na posameznem proteinu razvrstimo.

Zemljevid ocen na površini

3HK1 · primer iz validacijske množice

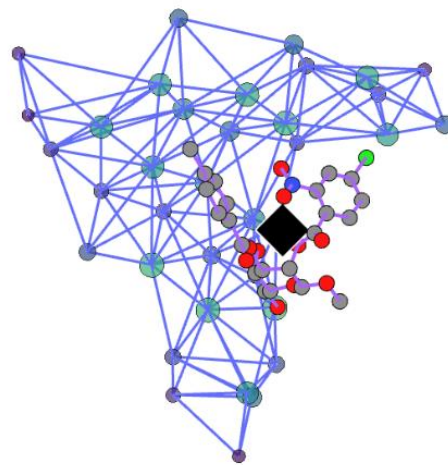


JASNO RAZLIKOVANJE MEST

ROC AUC = 0,888

GINE-RBF · 158.070 jasno označenih testnih točk.

Graf žepa 7rh3 – barvanje po razdalji do sonde

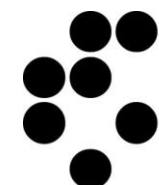


USPEŠNA IDENTIFIKACIJA IN KLASIFIKACIJA
VEZAVNIH MEST !

POMEN ZA NADALJNJE DELO

**Narava
vezavnih mest
ne načrtuje
naključno!**

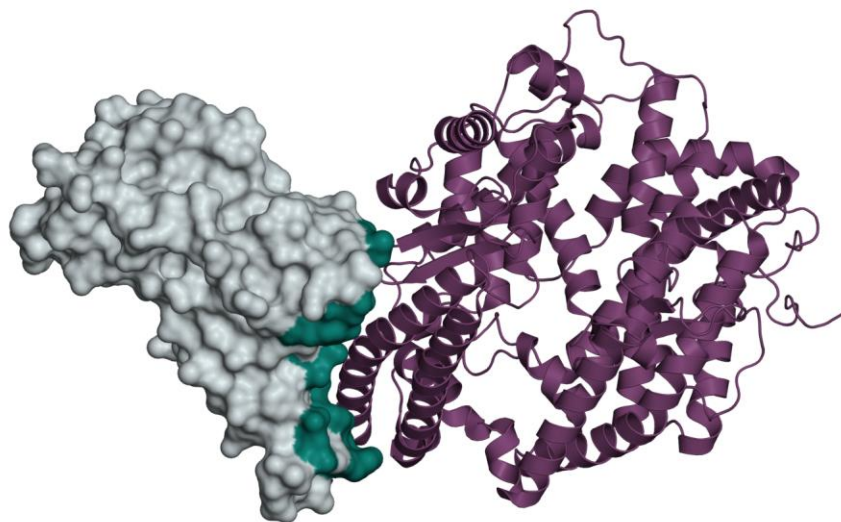


 **Institut
"Jožef Stefan"
Ljubljana, Slovenija**

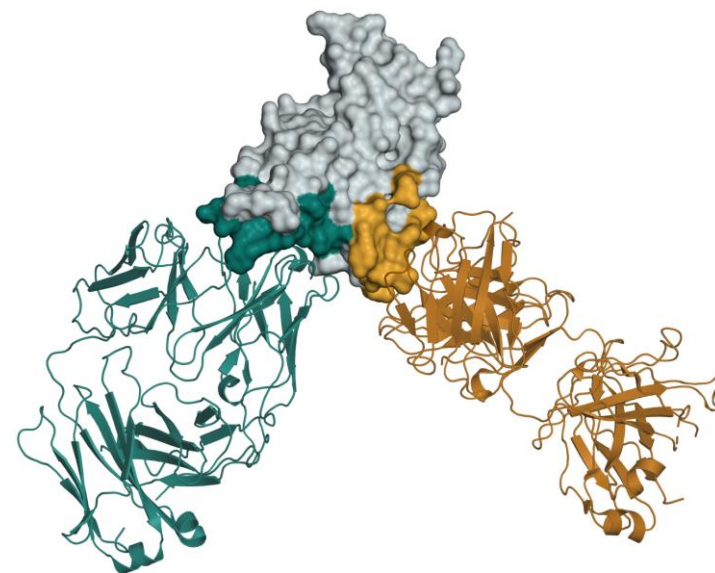
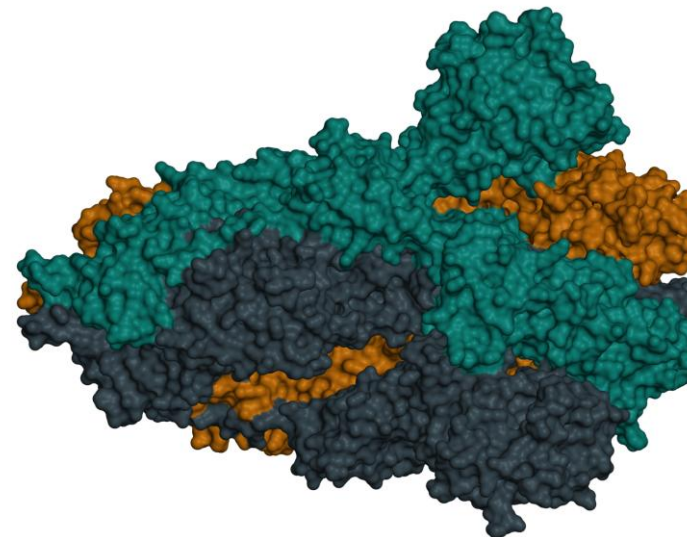
Rezultati poleg aplikativne vrednosti razkrivajo osnovne mehanizme konstrukcije vezavnih mest za male in velike molekule.

SARS-CoV-2: S^{pro}

- Protein bodice je **trimer** — tri enake podenote, obarvane različno.
- Prikazana je zaprta konformacija
- Na vrhu vsake podenote je **receptorsko vezavna domena (RBD)**.
- To je glavna tarča nevtralizirajočih protiteles in cepiv.
- Za majhno molekulo tu skoraj ni primernega žepa.



RBD (siva površina) se veže na človeški receptor **ACE2** (vijolično).

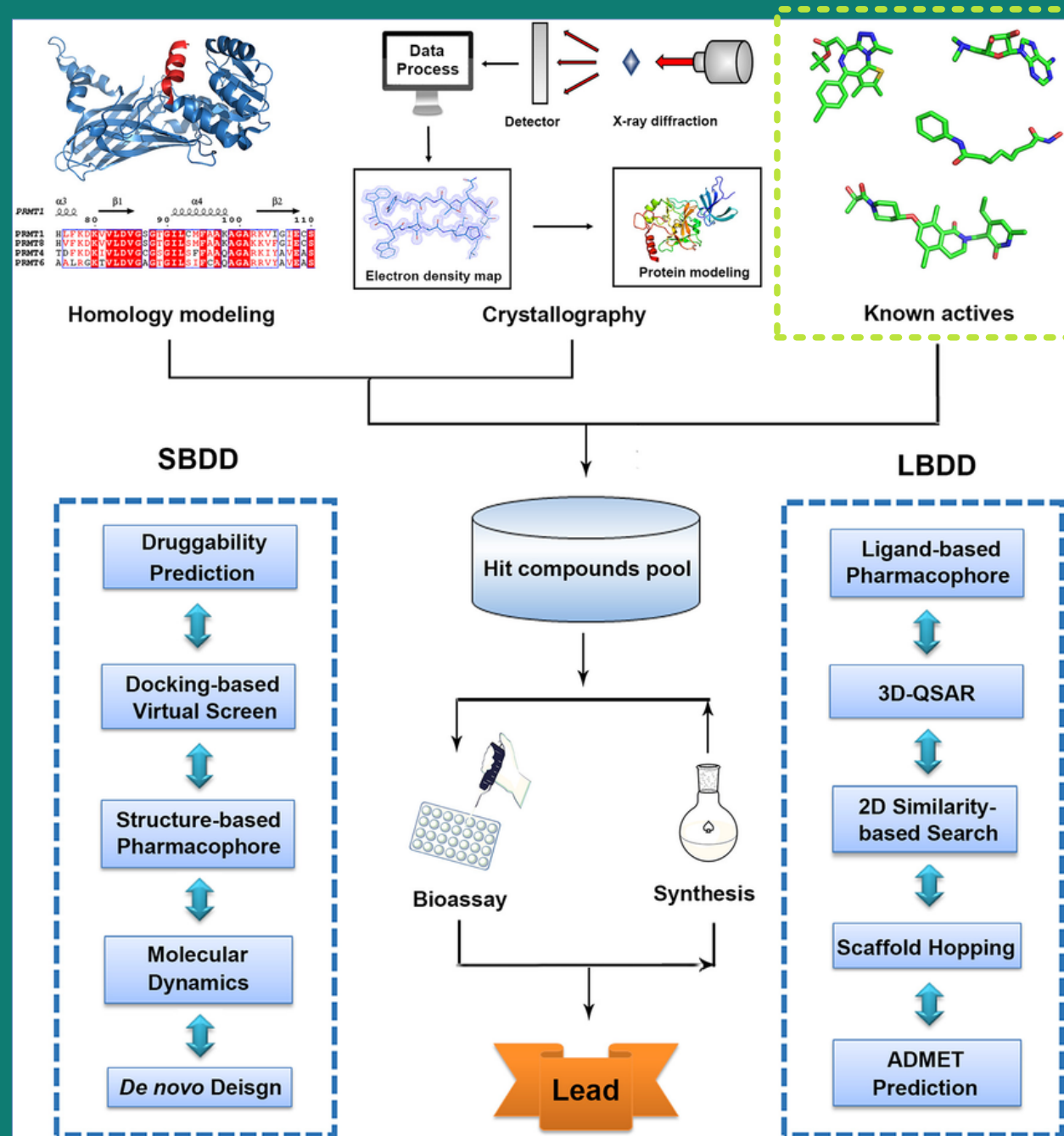


Vezava dveh protiteles na RBD ! – ključna so vezavna mesta !

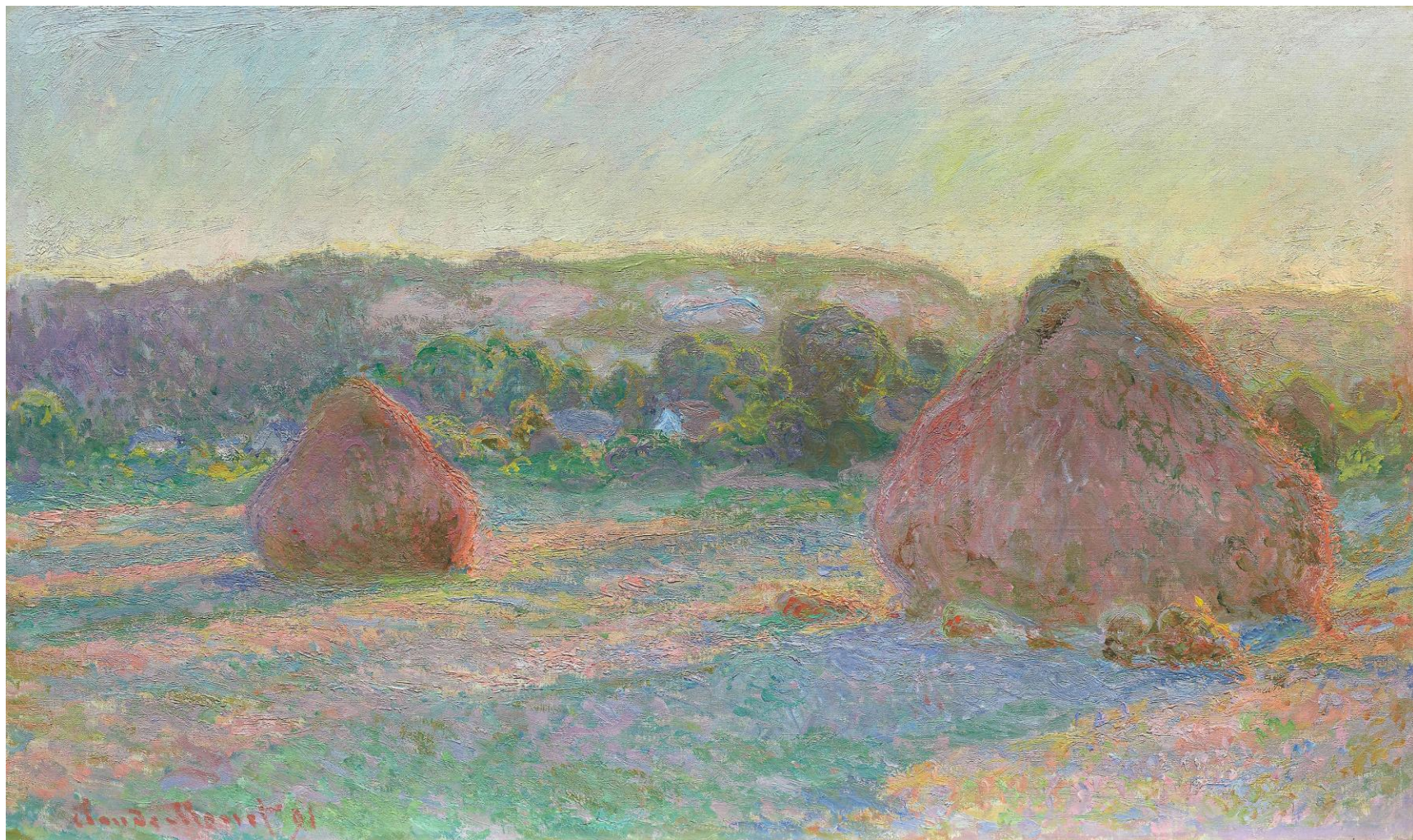
TRETJI DEL

Učinkovine ?

Kaj se sploh lahko veže na protein?



Poznamo tarčo ter njeno vezavno mesto, kako pa do učinkovine?



Kopice sena - Monet

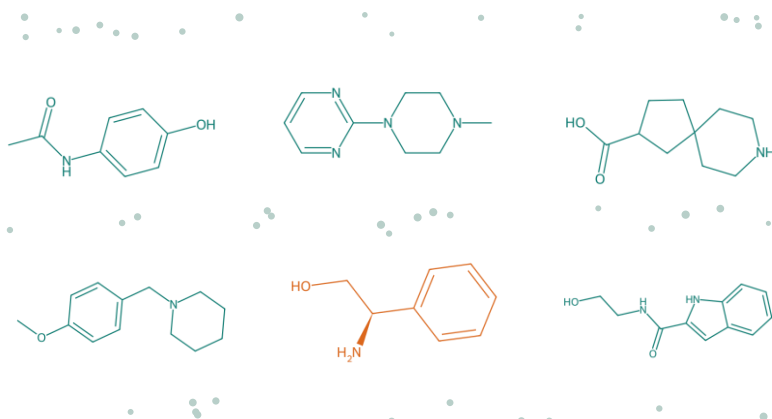
Kemijski prostor je prevelik za izčrpno presejanje

Kombinacije atomov, vezi, obročev in stereokemije ustvarjajo ogromno možnih majhnih molekul.

UČINKOVINAM PODOBNE MOLEKULE

10^{33} – 10^{60}

Ocene so močno odvisne od definicije in omejitev kemijskega prostora. [1]



Različni skeleti, substituenti in stereoizomeri

Izbor obetavnih območij → sidranje kandidatov → sinteza in preizkus

KONKRETNA PRIMERA OBSEGA

166,4 milijarde

GDB-17: računalniško naštete strukture z največ 17 atomi, ki niso vodik. [2]

PRESEJANJE · RECEPTOR D₄ · 2019 [3]

138 milijonov

sidranih

549

↓
sintetiziranih
in preizkušenih

OPTIMISTIČNI
MISELNI POSKUS

1 milijarda
molekul / sekundo

Za presejanje 10^{33} molekul:

$\approx 3 \times 10^{16}$
let

Ena ocena za vsako molekulo;
brez vseh drugih korakov.

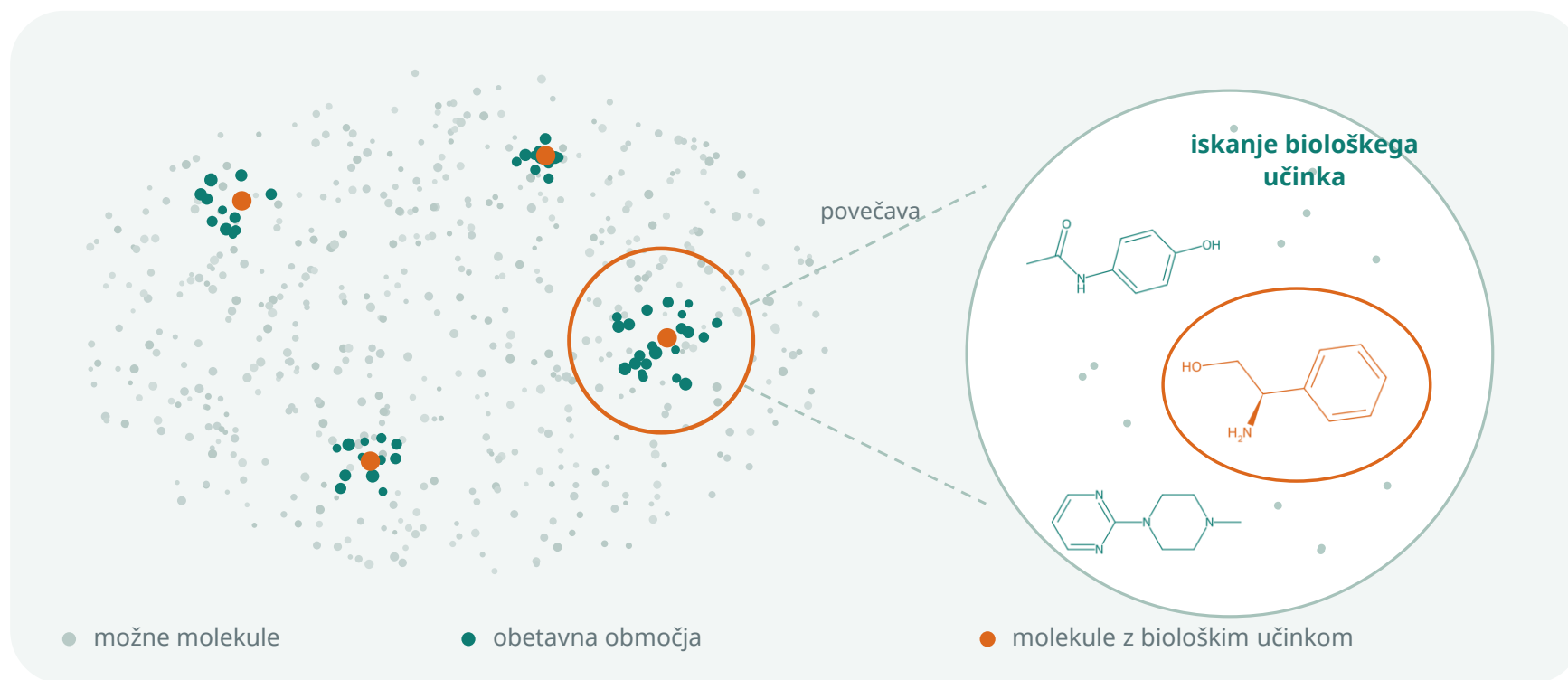
Konformacije, vezavne lege, sinteza in ponovitve testov dodatno povečajo obseg dela.

Kje v kemijskem prostoru iskati učinkovine?

Podobno ko redke zvezde v prostranem kemijskem vesolju.

PROSTOR MOŽNIH MOLEKUL

OBETAVNO OBMOČJE



KLJUČNO VPRAŠANJE

Kje je največ možnosti za uporaben biološki učinek?

Usmerjeno iskanje obetavnih območij

Izbrane molekule preverimo s poskusi.

Znanje o biološki tarči pomaga usmeriti izbor molekul za preizkus.

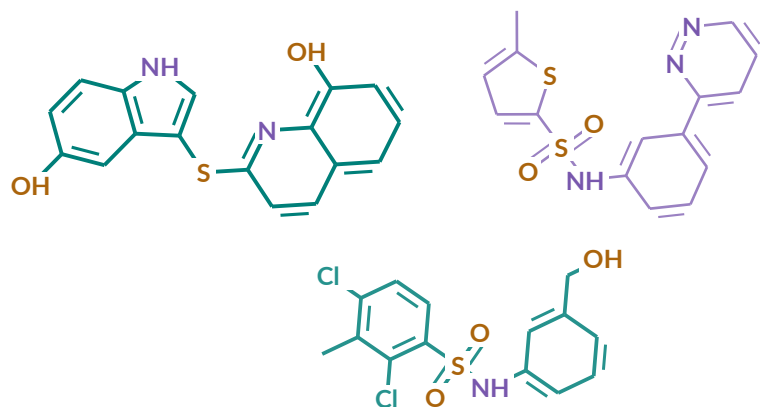
Generativno sidranje

Ogromen kemijski prostor raziskujemo s sprotnim generiranjem molekul.

1

Generiraj

Novi molekularni predlogi



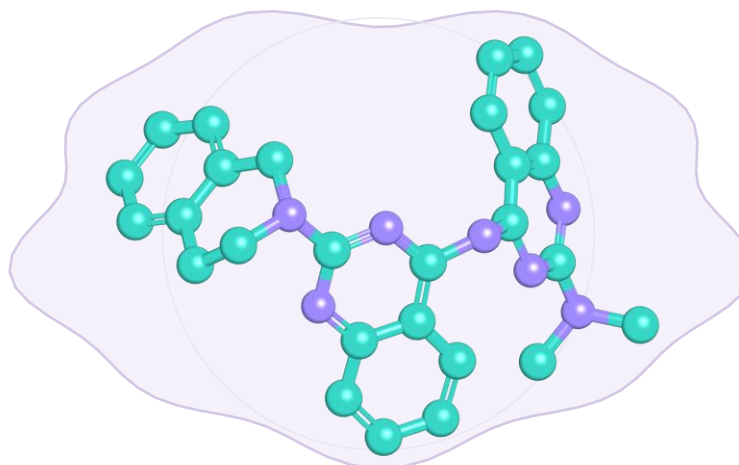
primeri iz generiranega nabora

FASMIFRA • Jensen GB-GA • NEAT

2

Sidraj

Poza v vezavnem mestu



arhivirana poza • shematsko vezavno mesto

CmDock • 3D-poza + ocena

3

Ocena !



shematski prikaz meril izbora

Npr. 250 staršev + 50 elit

POVRATNA INFORMACIJA → NOVA GENERACIJA

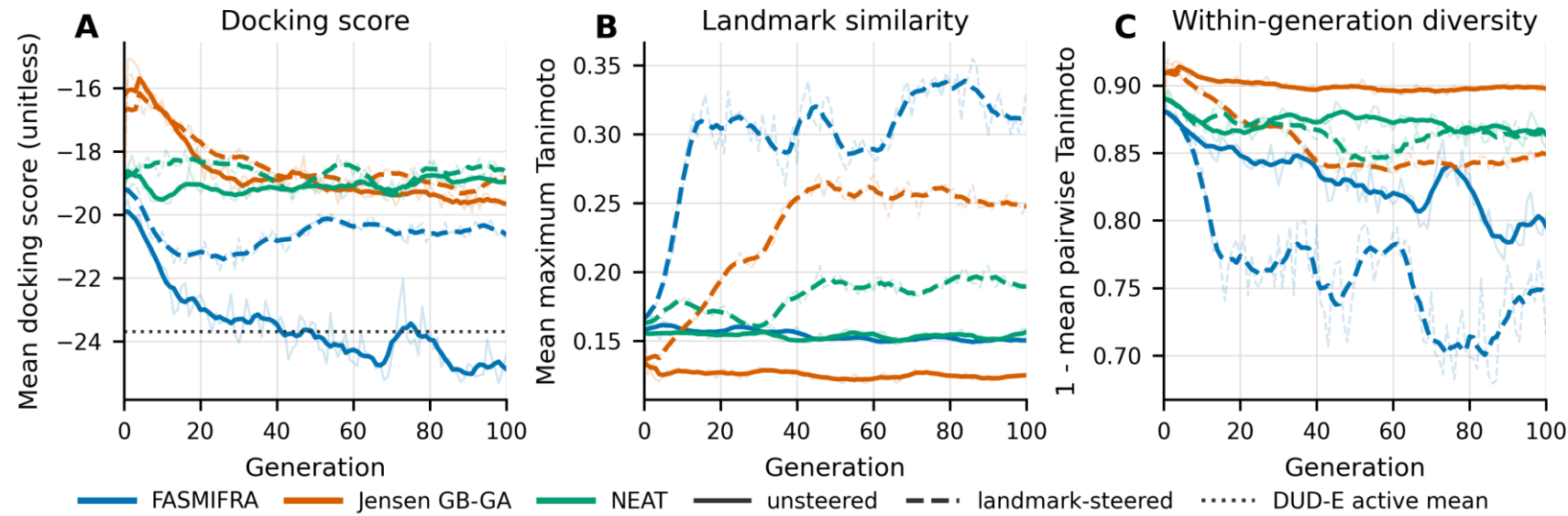
Enoten krmilnik • zamenljivi generatorji

Ocena sidranja vodi iskanje v obetavne regije kemijskega prostora.

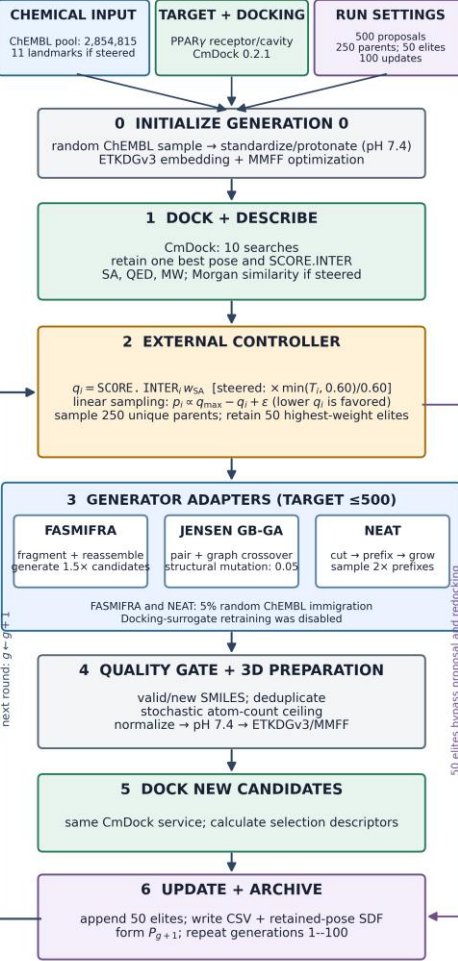
Hitra konvergenca. Velik kemijski prostor.

Zgodnje izboljšanje cenilne funkcije in sprotno širjenje raziskane kemije.

PRIMER KONVERGENCA CENILNE FUNKCIJE, PPARy



ARCHIVED PPARy GENERATIVE-DOCKING CAMPAIGN



Hitro izboljšanje ocene + raziskovanje obsežnega kemijskega prostora.

Širok kemijski prostor – doseg različnih molekul

63.567 različnih molekulskih ogrodi

Molekule nastajajo po potrebi.
MED SAMIM REŠETANJEM!

Lastno odprtokodno orodje za sidranje

— CmDock

- Odprtokodno (LGPL-3.0) sidranje na **proteine in nukleinske kisline**; različica 0.2.2.
- Iskanje združuje **genetski algoritem**, **nizkotemperaturni Monte Carlo** in **simpleksno minimizacijo**.
- prstni odtisi interakcij, kontrolne točke, ponovljiv generator naključnih števil in **slepo sidranje** čez celotno površino.



Kaj si je vredno zapomniti

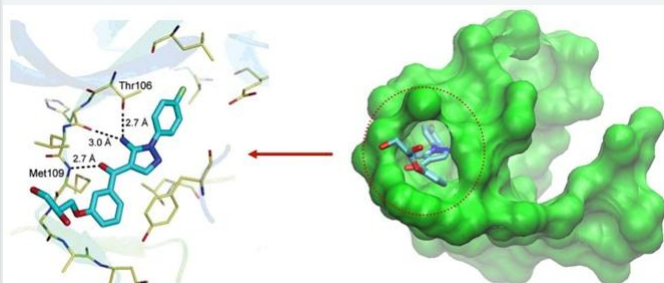
Tarče so postale dostopne

A bodimo kritični in uprabljamo ensambel metod.



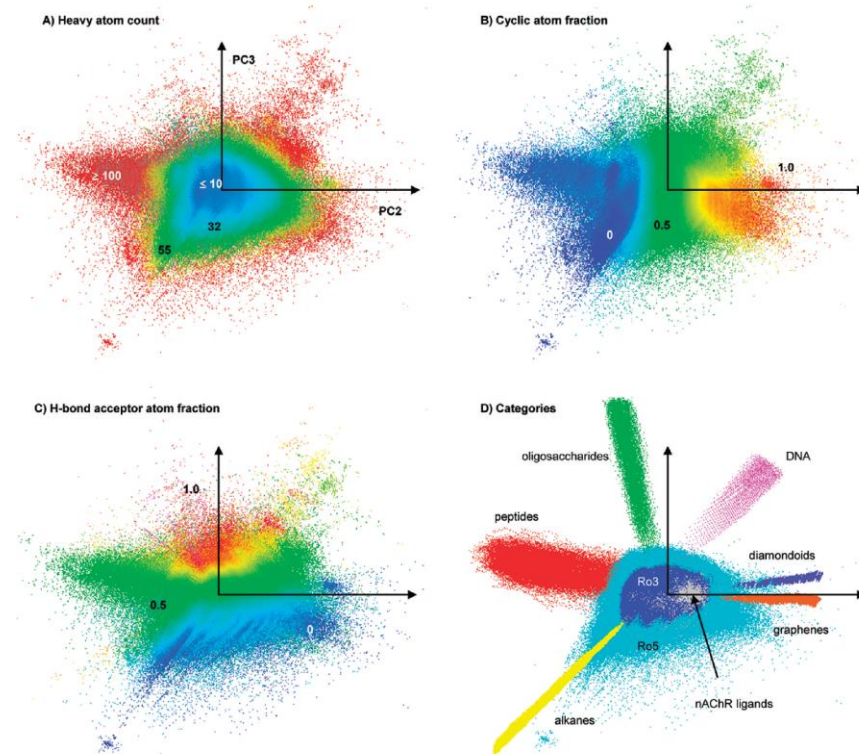
Vezavna mesta lahko poiščemo

Uspešno lahko poiščemo vezavna mesta, jih klasificiramo in upravimo za nove študije.



Zadetke lahko ustvarjamo

Generativna zanka širi kemijski prostor, namesto da bi ga le pregledovala.



Nove AI-podprte metode omogočajo in bodo omogočile pot do novih spoznanj in ključno prispevale k načrtovanju novih zdravil, a le tam kjer jih bomo znali učinkovito uporabiti !

Hvala za pozornost

Vprašanja so dobrodošla.



Klimt

Marko Jukič

Pišite mi na:

marko.jukic@um.si / marko.jukic@famnit.upr.si

Naslednja predstavitev:
Umetna inteligenca za učinkovitejše zdravljenje raka

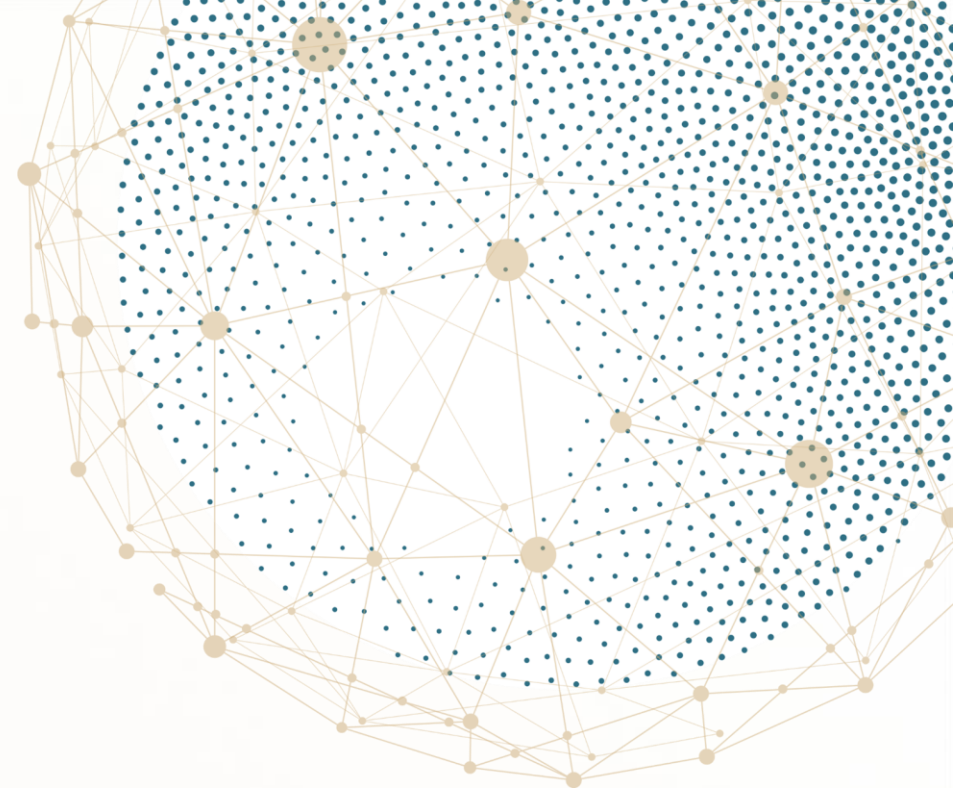
dr. Miha Štajdohar,
Soustanovitelj in tehnični direktor, Genialis



Genialis™ Supermodel: The AI platform for therapeutic intelligence to benefit every cancer drug and cancer patient

Dr. Miha Štajdohar, Co-founder & CTO

11 September 2026



PROBLEM

Most cancer drugs don't work for most patients

Despite spending \$ billions...

<45% of patients benefit
from expensive medicines

>90% of new drugs fail
in clinical trials

10-15 years

Sunk development **TIME**

\$2-5 billion

COST not counting cost of capital

> \$90 billion

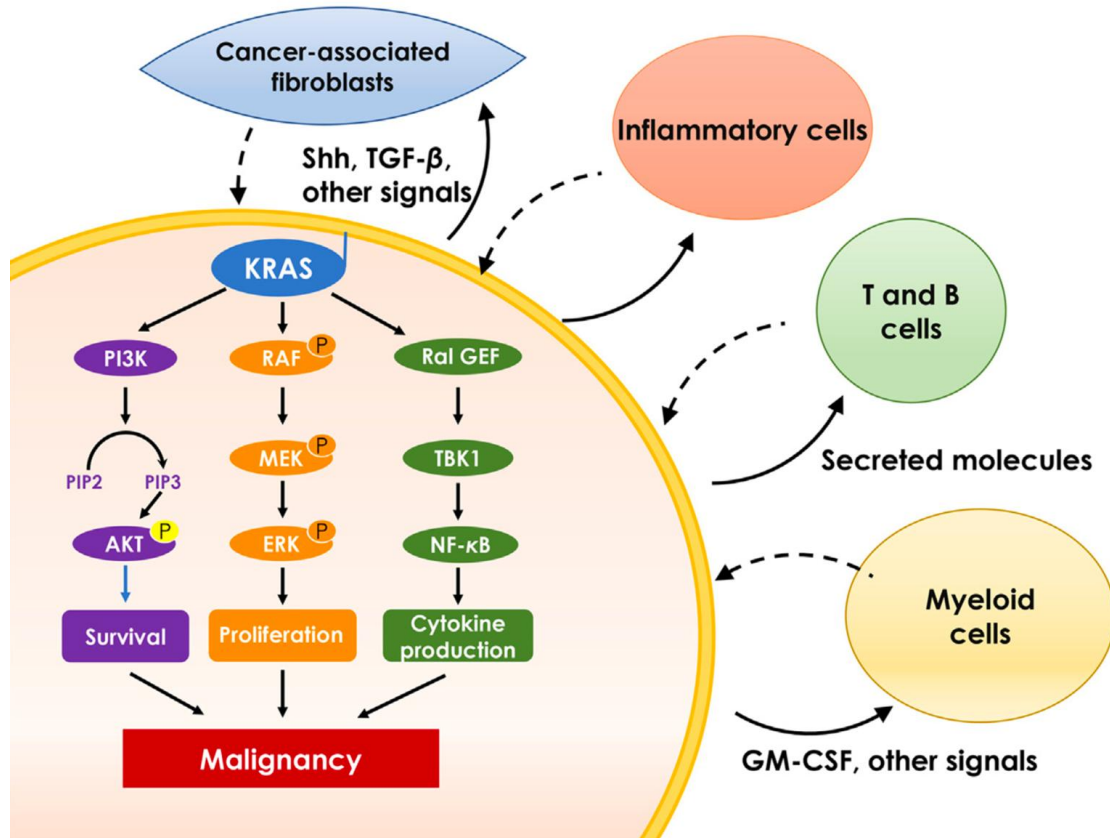
SPENT on top 10 cancer drugs

< 45% patients

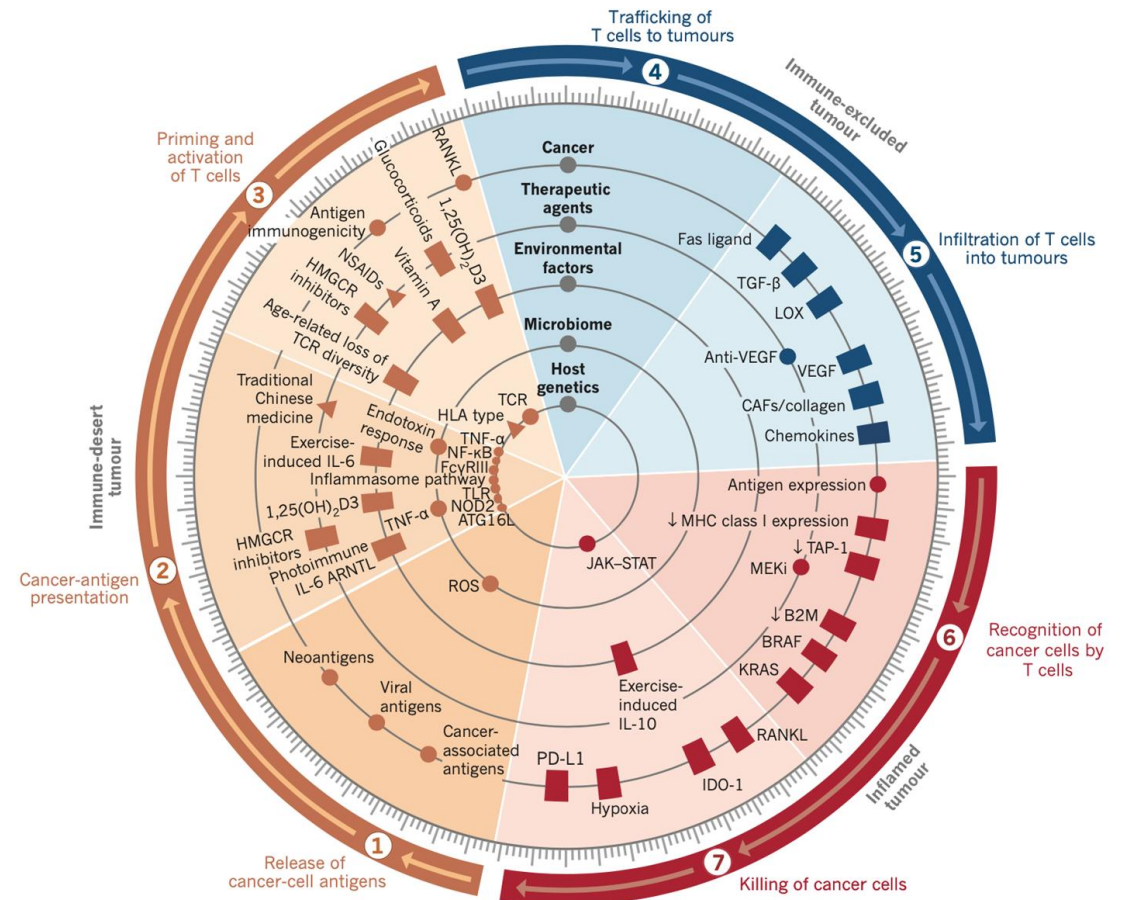
Demonstrate **CLINICAL RESPONSE**

PROBLEM

A single mutation cannot explain how a patient will respond to treatment



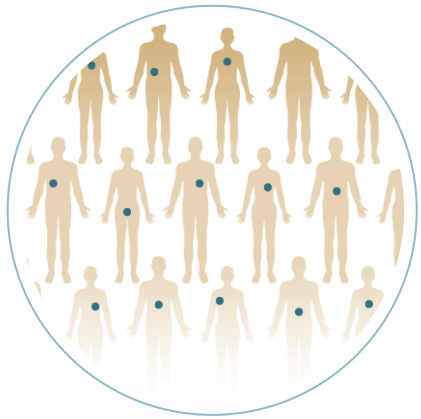
Liu et al. Acta Pharm Sin B Vol 9, 2019



Chen & Mellman. Nature Vol 541, 2017

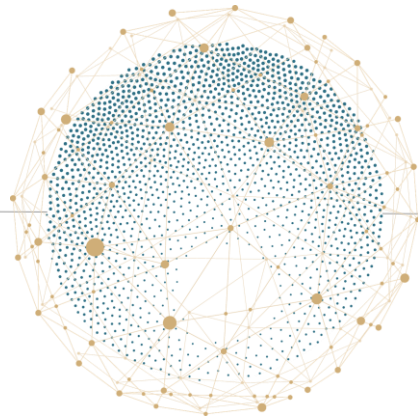
SOLUTION

Supermodel turns RNA into an interpretable map of cancer biology



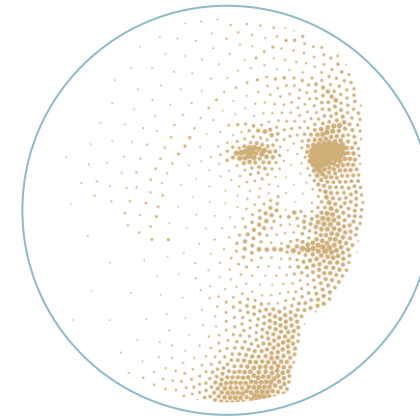
Proprietary harmonized dataset

Built on >1Bn data points from
100s of thousands of *globally
diverse* RNA-seq records



Groundbreaking AI technology

Genialis Supermodel is a large
molecular model (*LMM*) that has
learned *ground truth cancer biology*



Confident outcomes across geographic borders

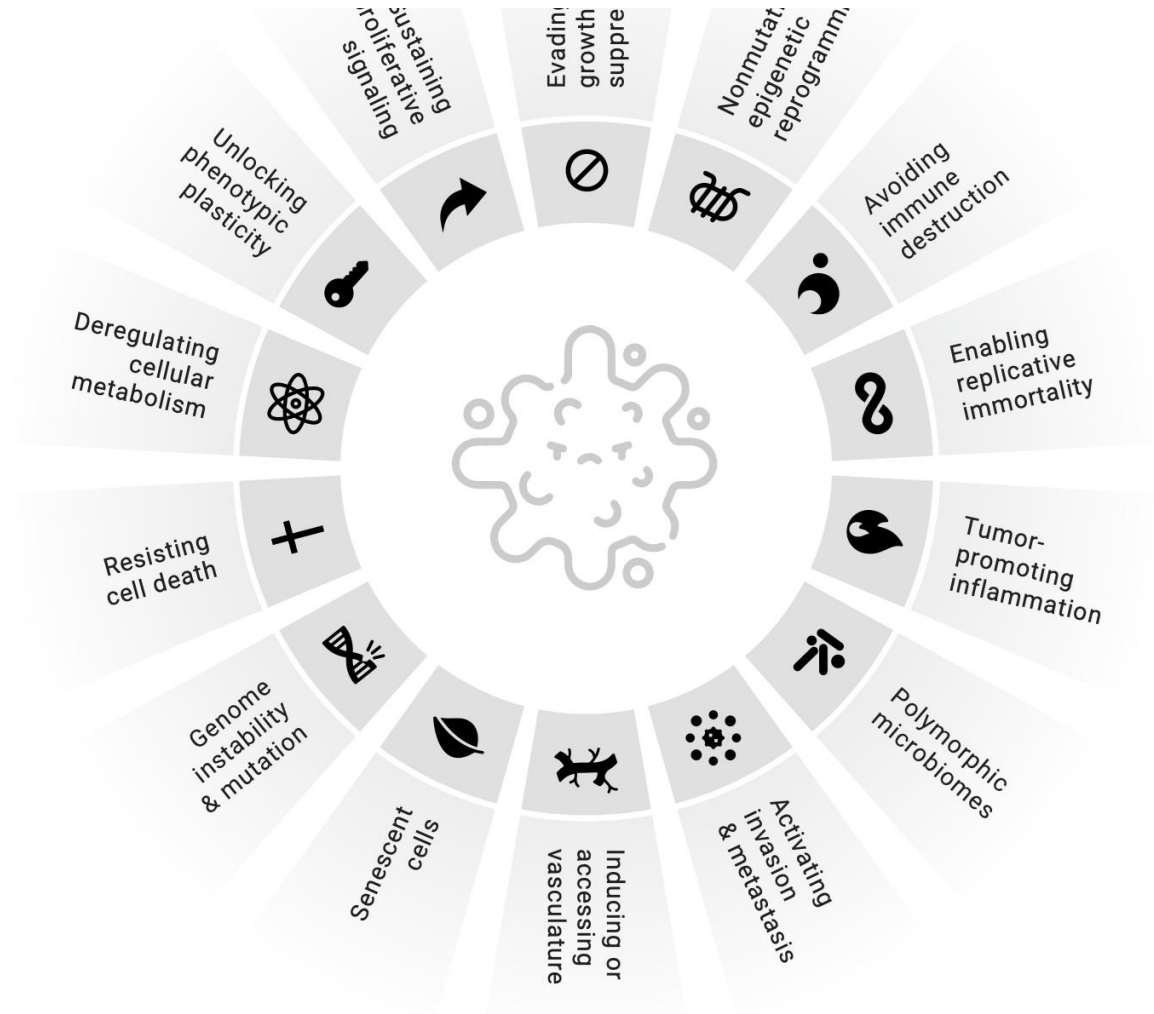
Designed to be consistent across
sequencing platforms, genome
builds, and ethnogeography

SOLUTION

Biological signatures reveal what is driving each patient's tumor

Hallmark biologies can be represented as multi-gene signatures, measured by gene expression.

Genialis has a library of over **150 validated signatures**, and 1000s more in progress, **for modeling cancer-related biologies**



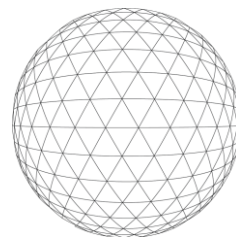
SOLUTION

Supermodel is a foundation model of cancer biology

RNA-Seq Assay (Gene Expression Profile)

~60,000 genes

Supermodel



Large Molecular Model (LMM)



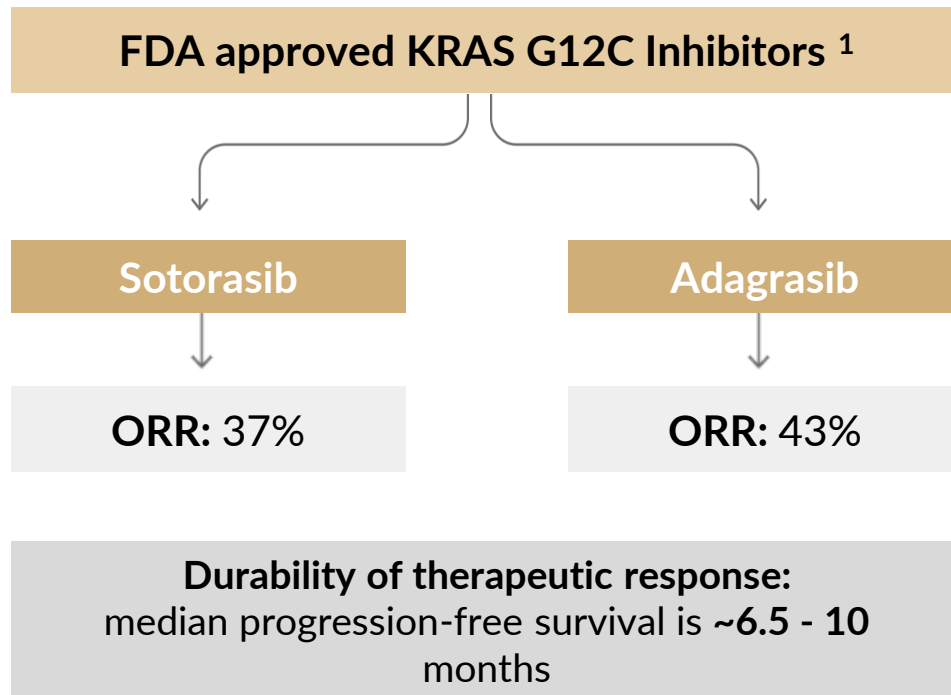
Cancer Biology

100s of validated signatures

CASE STUDY

KRAS mutations identify eligible patients, not responders

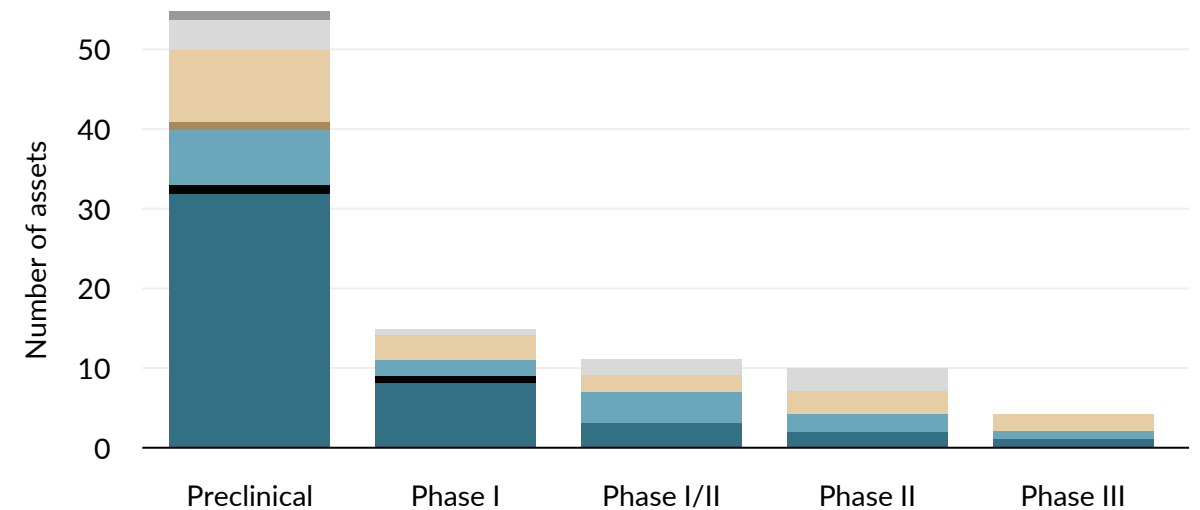
Two drugs currently on market



Competition to get to market

>50 companies competing across ~7 different solid tumor indications ²

■ Unspecified Solid Tumor ■ Breast Cancer ■ Colorectal Cancer
■ Liver Cancer ■ NSCLC ■ Pancreatic Cancer ■ Bladder Cancer



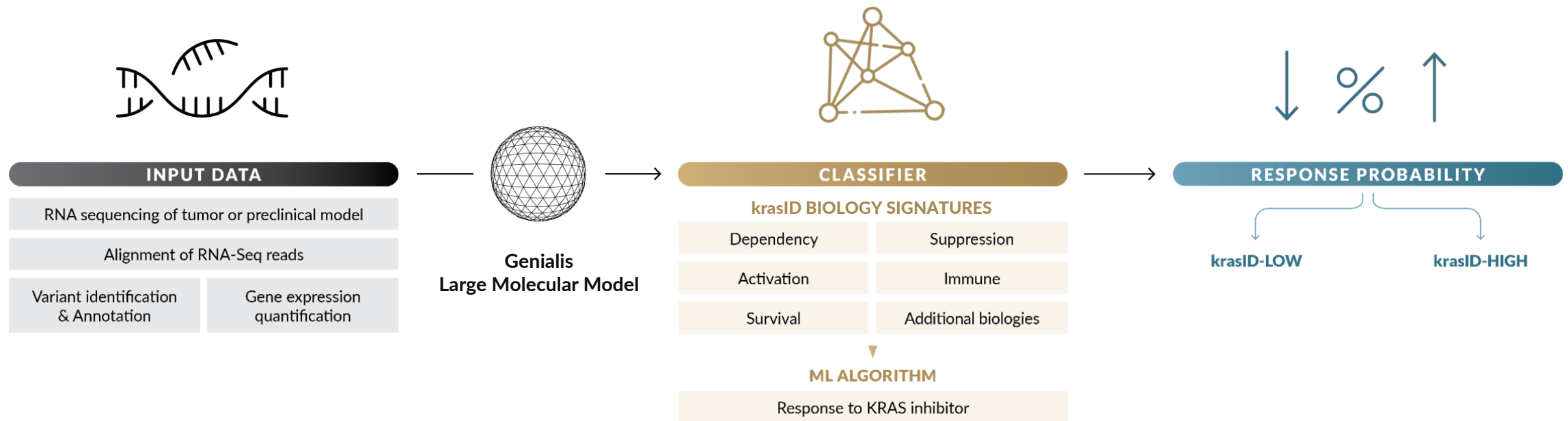
¹ Tian H, Yang Z, He J. Adagrasib: A landmark in the KRASG12C-mutated NSCLC. MedComm (2020). 2022 Nov 25;3(4):e190. doi: 10.1002/mco2.190. PMID: 36448054; PMCID: PMC9697582.

² Biocentury Inc. "KRAS Pipeline" (2024)

CASE STUDY

krasID combines multiple biological signals to predict response

RNA-seq → cancer biology signatures → response probability



RESULTS

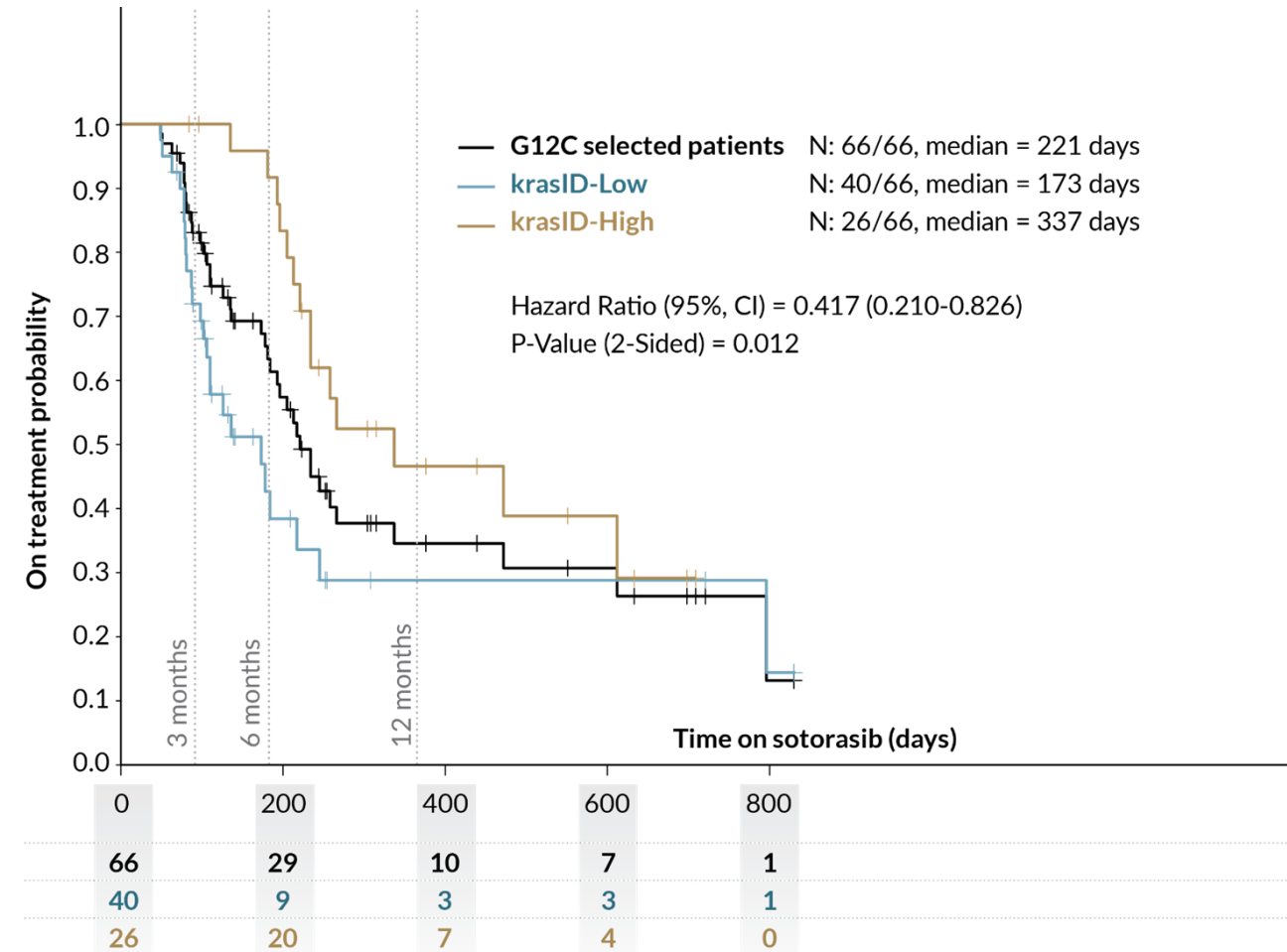
krasID identifies patients with more durable response

Patients stratified by time-on-therapy

Dramatically different time on treatment observed between krasID-LOW and -HIGH cohorts and between krasID-selected and mutation-selected (baseline) cohorts

krasID-HIGH captures majority of responders

Patients stratified by time-on-therapy reveals that krasID-HIGH stay on treatment 2x longer with 92% and 47% remaining responsive at 6 months and 12 months



IMPACT

Better patient selection makes trials smaller, faster, and cheaper

Patients required

Retrospective power analysis

333



KRAS
mutation

82



krasID

4× fewer patients

Enrollment duration

U.S. oncology benchmark

~29 mo



Oncology
median

~17 mo



krasID

~1y faster time to market

Development cost

U.S. oncology phases 1–3

\$60M



Oncology
median

<\$20M



krasID

>\$40M estimated savings

VALIDATION

Our models are already used by leading pharmaceutical companies

Biomarker model	Collaborators and commercial partners	
Tumor Micro Environment <i>Xerna™ TME</i>		
KRAS inhibitors <i>Genialis™ krasID</i>	Multiple top 10 Pharma & biotech (undisclosed)	"TEMPUS
DNA Damage Response <i>Genialis™ ddrID</i>	 Debiopharm	
Antibody Drug Conjugates <i>Genialis™ adclID</i>	Multiple top 10 Pharma & biotech (undisclosed)	"TEMPUS
Tissue-centric, pan-target: PDAC, Lung, Ovarian		

Select papers and posters:

- Žibera et al, ASCO (2026)
- Pavlic et al, AACR (2026)
- Špendl et al, Artificial Intelligence in Medicine (September 2025)
- Fuchs et al, ESMO AI (2025)
- Uhlik et al, ASCO (2025)
- Lovše et al, AACR (2025)
- Fuchs et al, AACR (2025)
- Žibera et al, AACR (2024)
- Basu et al, SITC (2023)
- Uhlik et al, Frontiers in Oncology (May 2023)
- Chen et al, Journal of Personalized Medicine (November 2022)

VALIDATION

From academic research to FDA-accepted clinical use

IDE accepted by the FDA



Validated as a clinical trial assay



Licensed by Qiagen for CDx development



Out-licensed for commercialization as an LDT



Deployed for retrospective analysis for various biotech/pharma



Xerna™ TME panel developed by Genialis with OncXerna Therapeutics (now Feng Bio)