Pacientes Virtuales

Generando datos biomédicos con IA

Francisco Carrillo Pérez - PyData Granada

Who am I?

BSc. Computer Science

UNIVERSIDAD DE GRANADA

2013-2018

MSc. in Data Science / Data Scientist

UNIVERSIDAD DE GRANADA

mDurance

2018-2019

Ph.D. in Machine Learning

UNIVERSIDAD DE GRANADA

2019-2023

Fulbright Scholarship

2021-2022

Postdoctoral Researcher

2023-2024

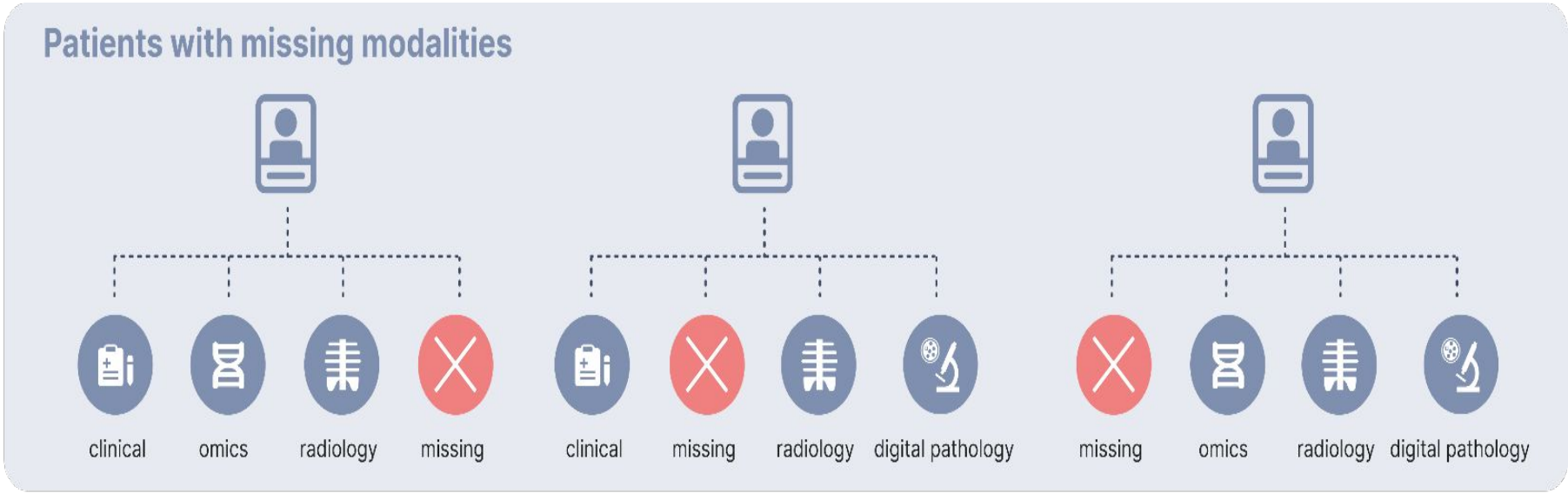
Senior Scientist, Imaging AI

Bristol Myers Squibb

2024-

The multimodal nature of biomedical data

- Biomedical data is, by nature, **multimodal and incomplete**
- We have non-invasive modalities (clinical data, radiology, CT Scans, etc.) and also invasive modalities (biopsies, transcriptomics sequencing, etc.)
- Furthermore, not all hospitals have access to the equipment to perform these tests

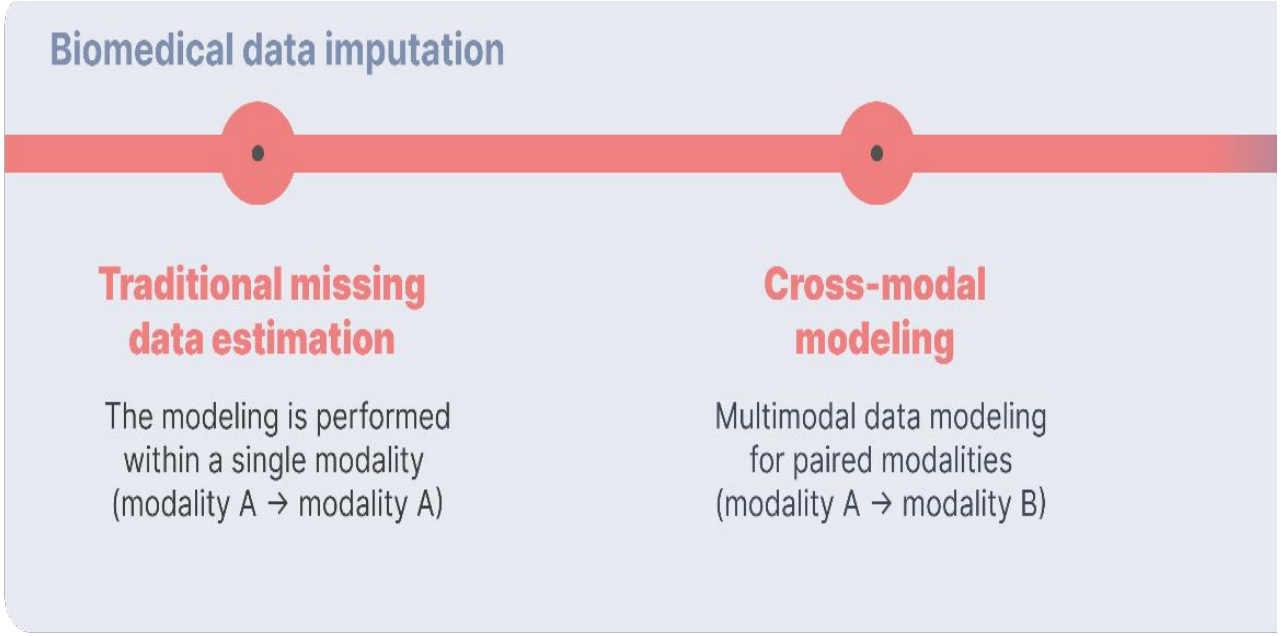


The multimodal nature of biomedical data

- Biomedical data is, by nature, **multimodal and incomplete**
- We have non-invasive modalities (clinical data, radiology, CT Scans, etc.) and also invasive modalities (biopsies, transcriptomics sequencing, etc.)
- Furthermore, not all hospitals have access to the equipment to perform these tests
- **Solution? Generate/Impute that data using AI!**

The timeline of biomedical data imputation

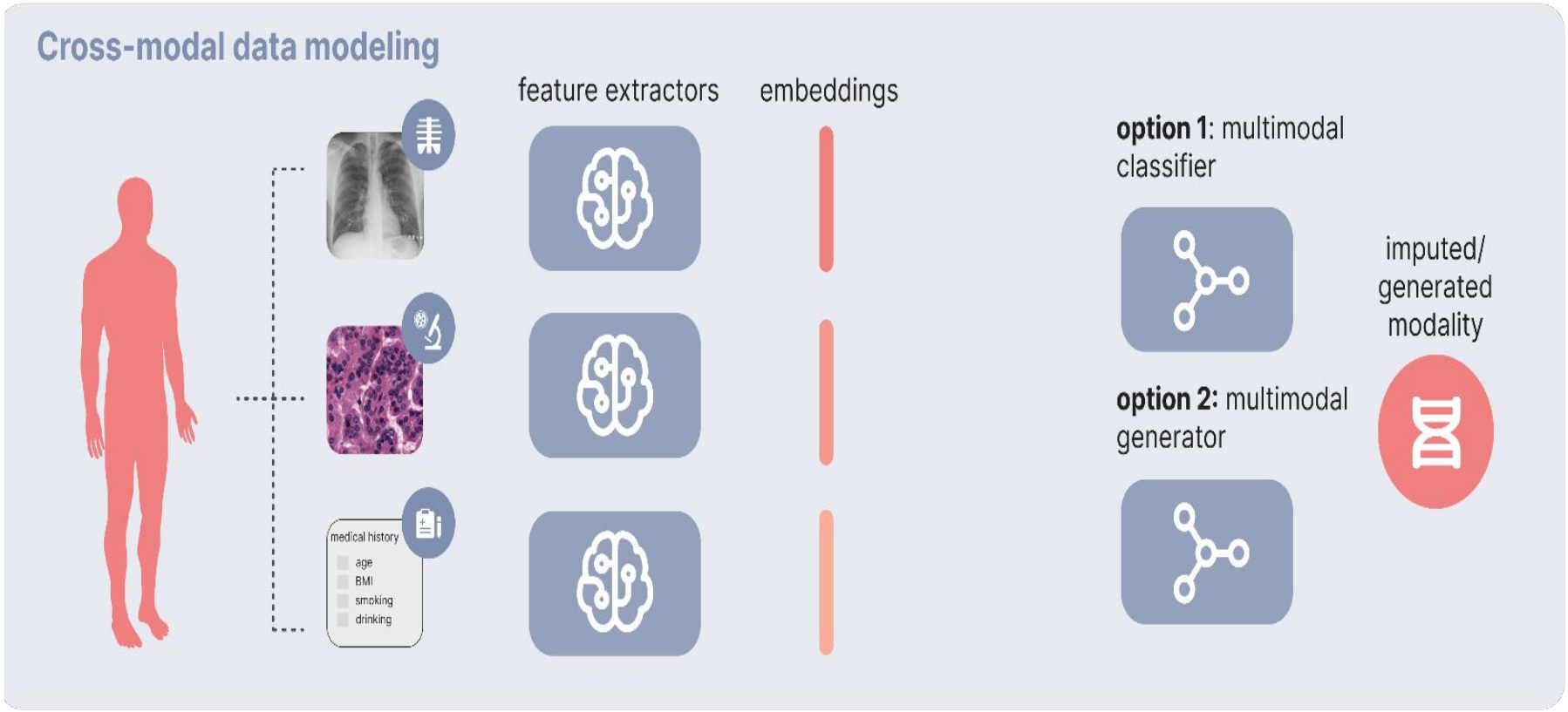
- **Traditional data imputation:** we impute missing values based on the distribution of the modality
- **Cross-modal modeling:** we use modality A to impute modality B



Future?

Cross-modal data imputation

- **Cross-modal modeling:** we use modality A to impute modality B using feature extractors and AI



Cross-modal data imputation: Two examples

Article | Open access | Published: 14 November 2024

Digital profiling of gene expression from histology images with linearized attention

Marija Pizurica, Yuanning Zheng, Francisco Carrillo-Perez, Humaira Noor, Wei Yao, Christian Wohlfart, Antoaneta Vladimirova, Kathleen Marchal & Olivier Gevaert

Nature Communications **15**, Article number: 9886 (2024) | Cite this article

IMAGE -> TABULAR

Everything implemented, tested, and loved in Python!

Article | Published: 21 March 2024

Generation of synthetic whole-slide image tiles of tumours from RNA-sequencing data via cascaded diffusion models

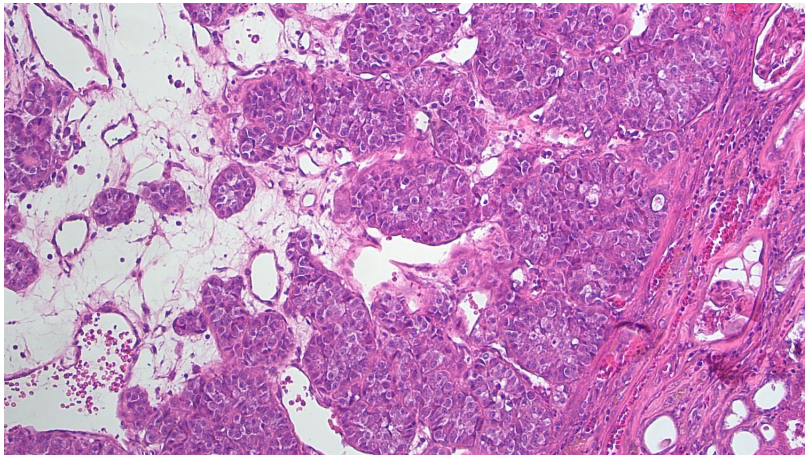
Francisco Carrillo-Perez, Marija Pizurica, Yuanning Zheng, Tarak Nath Nandi, Ravi Madduri, Jeanne Shen & Olivier Gevaert

Nature Biomedical Engineering (2024) | Cite this article

TABULAR -> IMAGE

Digital Pathology and RNA-Seq sequencing

Cancer tissue stained with Hematoxylin and Eosin (H&E)



https://focusontoxpath.com/wp-content/uploads/toxicologic-pathology-tissue-slide.jpg

RNA-Seq Sequencing

0.5	0.6	0.9	0.1	.....	1.1
TP53	*EGFR*	*KRAS*	*ALB*		*IDH1*

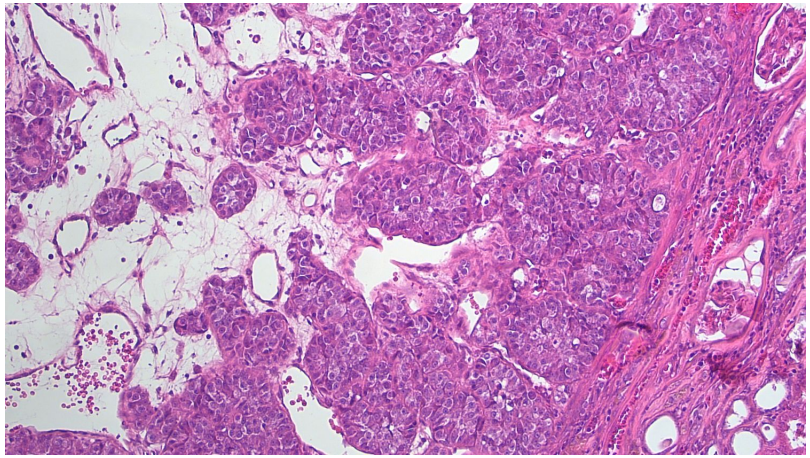
H&E to RNA-Seq

Article | Open access | Published: 14 November 2024

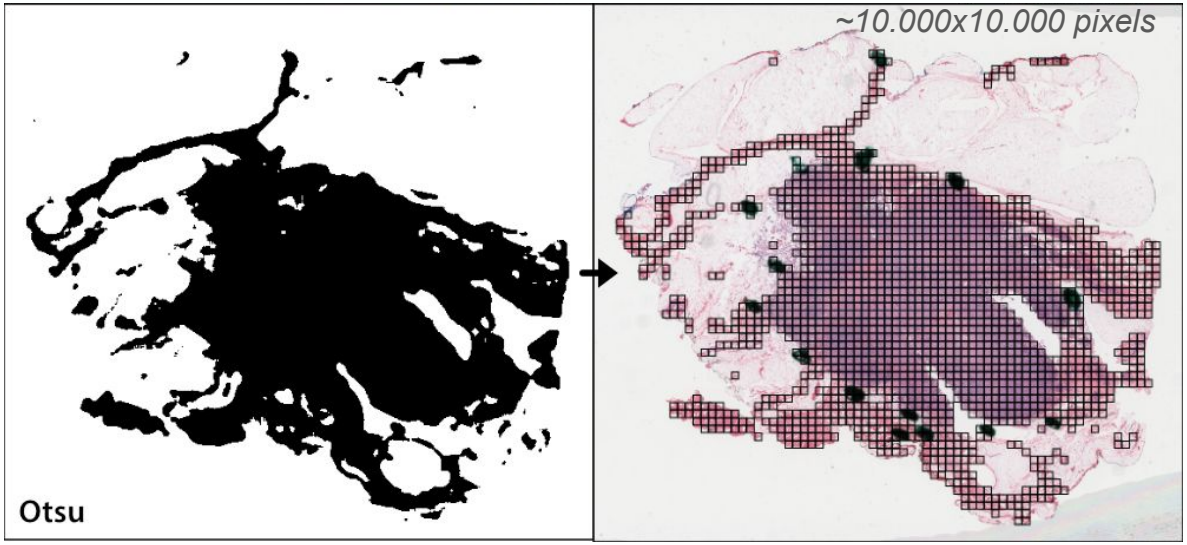
Digital profiling of gene expression from histology images with linearized attention

Marija Pizurica, Yuanning Zheng, Francisco Carrillo-Perez, Humaira Noor, Wei Yao, Christian Wohlfart, Antoaneta Vladimirova, Kathleen Marchal & Olivier Gevaert

Nature Communications **15**, Article number: 9886 (2024) | Cite this article

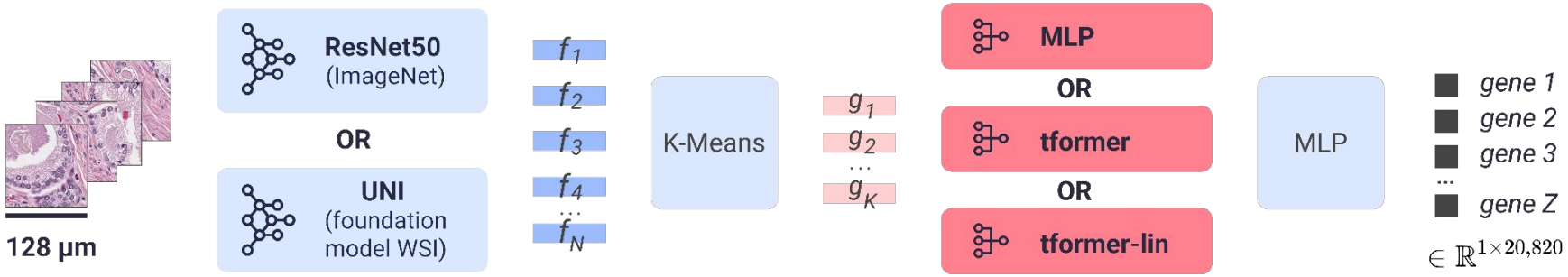


SEQUOIA

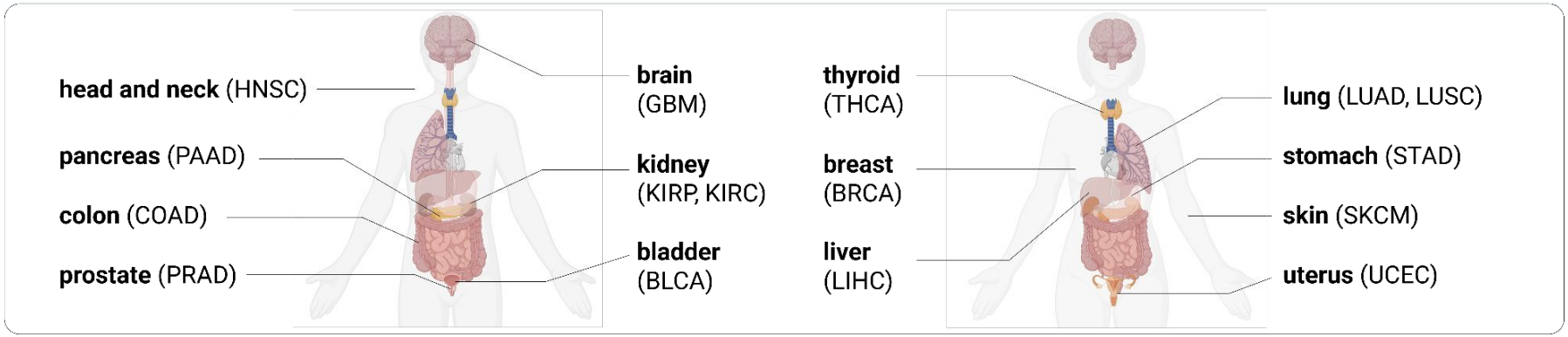


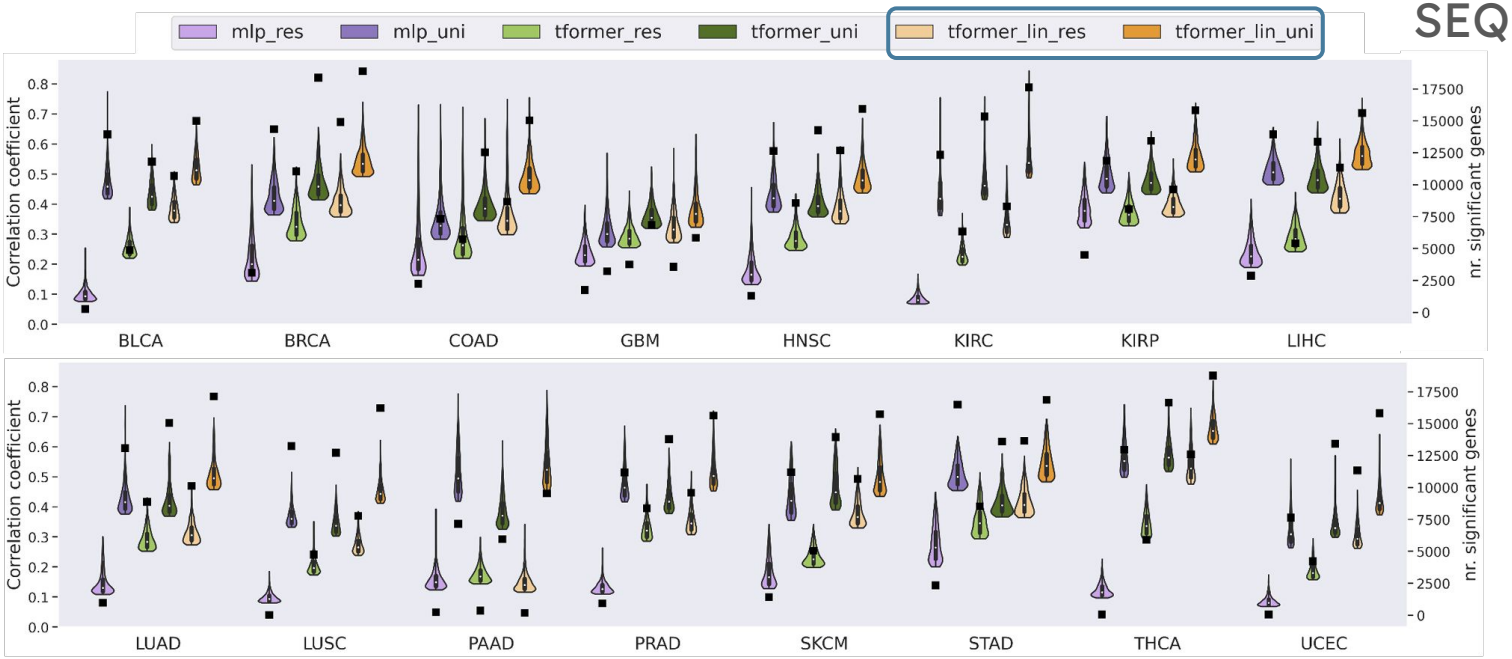
https://slideflow.dev/slide_processing/

Example from Slideflow, a Python library for digital pathology data https://github.com/slideflow/slideflow

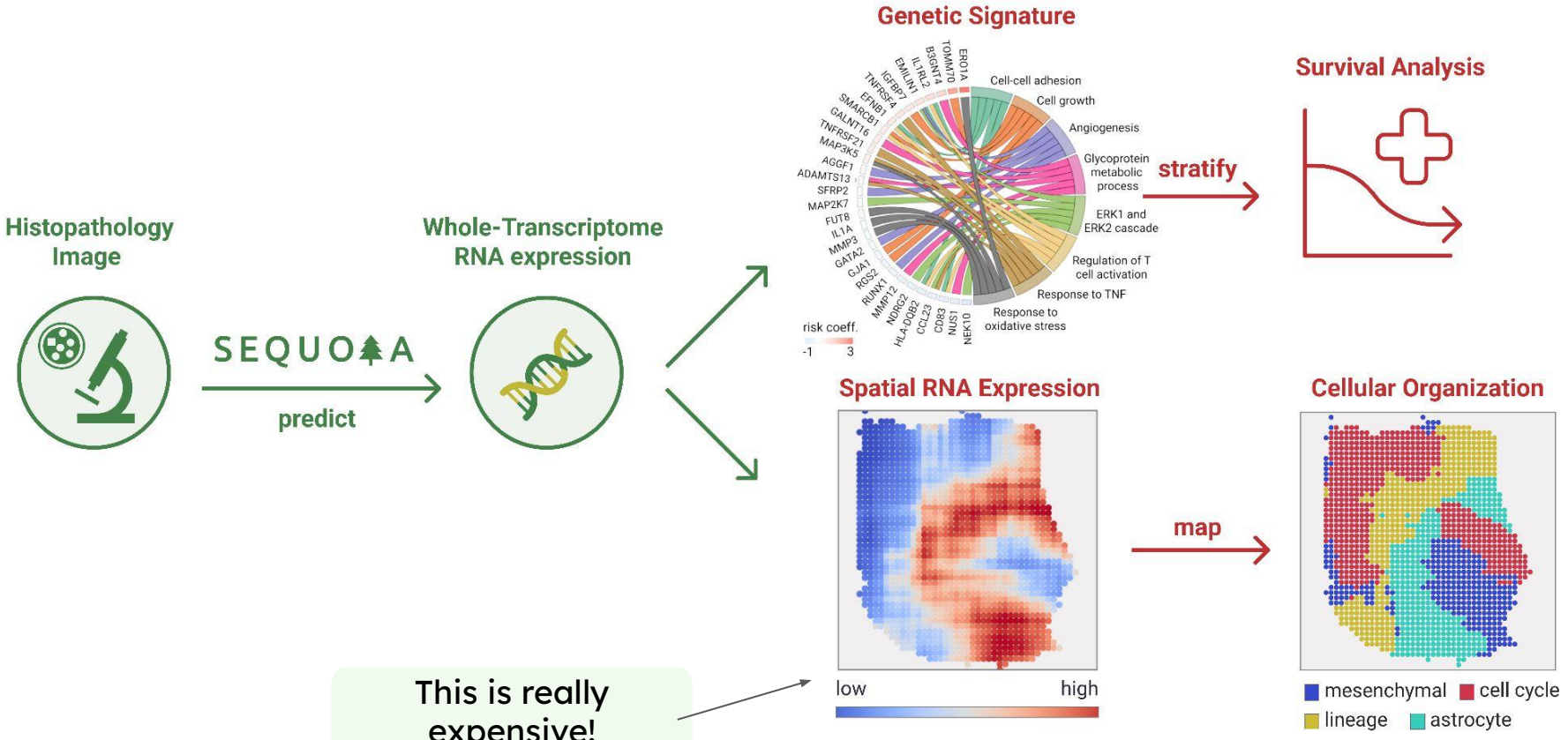


SEQUOIA outperforms other models in literature





What is this useful for?



This is really expensive!

Code: https://github.com/gevaertlab/sequoia-pub
Models: https://huggingface.co/gevaertlab
Demo: https://sequoia.stanford.edu/

Made with:

Streamlit

RNA-Seq to H&E

Article | Published: 21 March 2024

Generation of synthetic whole-slide image tiles of tumours from RNA-sequencing data via cascaded diffusion models

Francisco Carrillo-Perez, Marija Pizurica, Yuanning Zheng, Tarak Nath Nandi, Ravi Madduri, Jeanne Shen & Olivier Gevaert

Nature Biomedical Engineering (2024) | Cite this article

Text-To-Image models in Biomedicine?

Text prompt

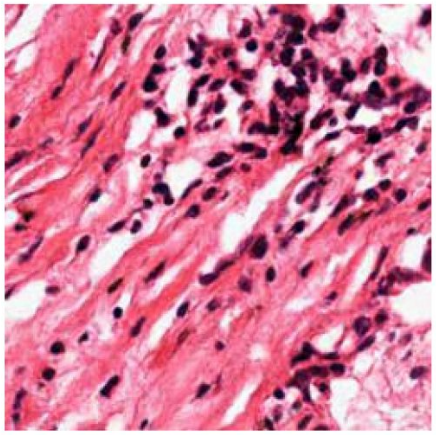
"A painting of the Stanford Hoover Tower in the style of Monet"

→ DALL-E 2 →

RNA-Seq

[0.64,0.75,0.25, 0.80,...0.93]

→ RNA-CDM →



RNA-CDM Architecture

RNA-Seq reduction

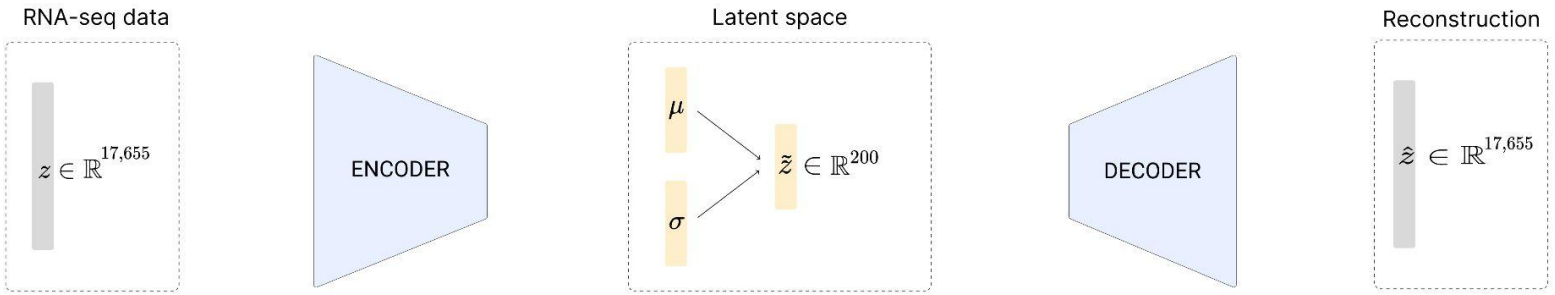
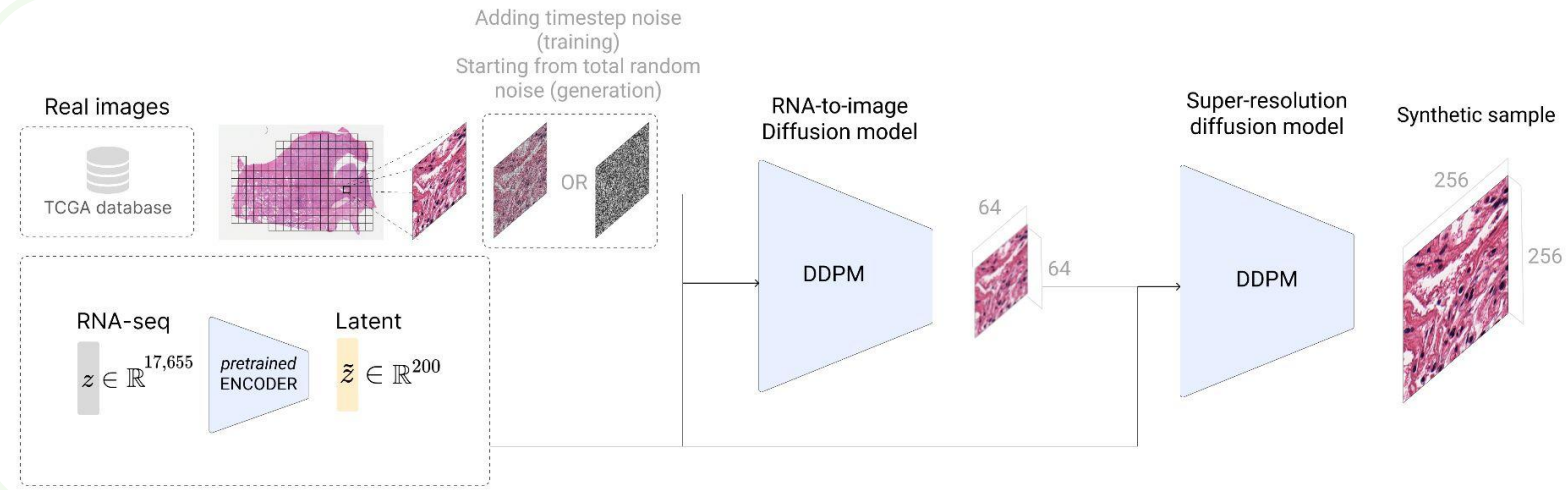
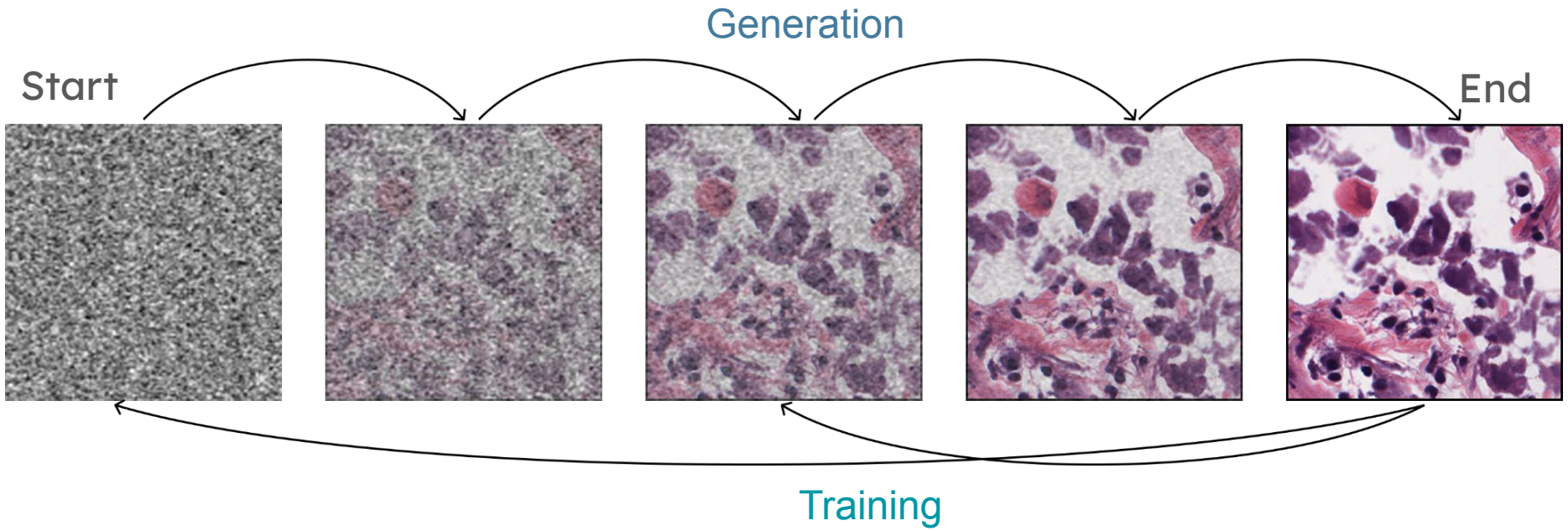


Image generation



Diffusion models

- During training, noise is added to the image and the model learns to predict that noise
- During the generation, we start from total noise, and the model reduces the noise one step at a time until we have the final image

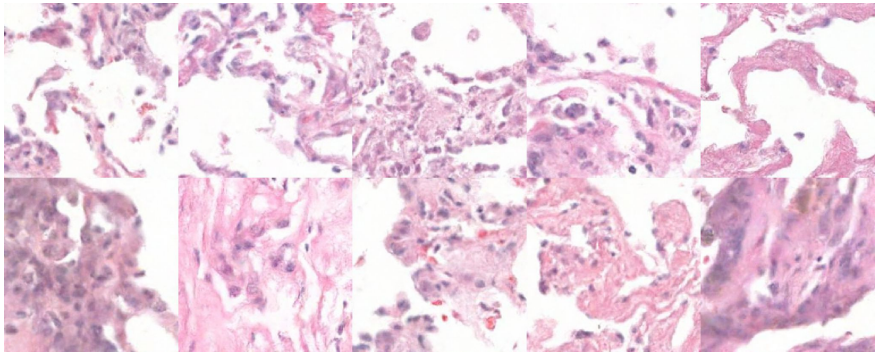


Model size matters

Simple Linear Regression

$$Y = a + bX$$

~500.000 parameters

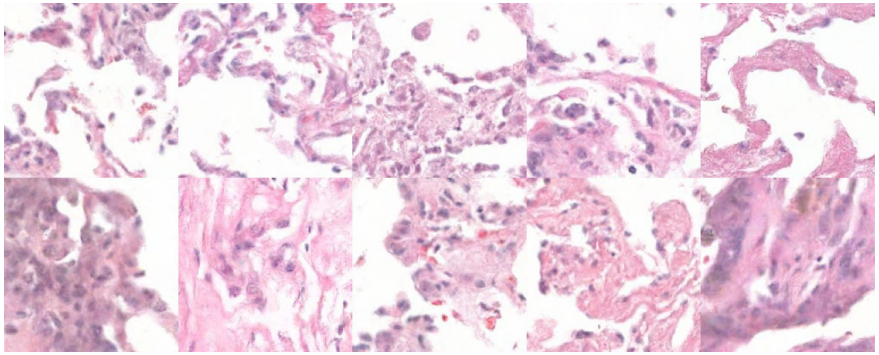


Model size matters

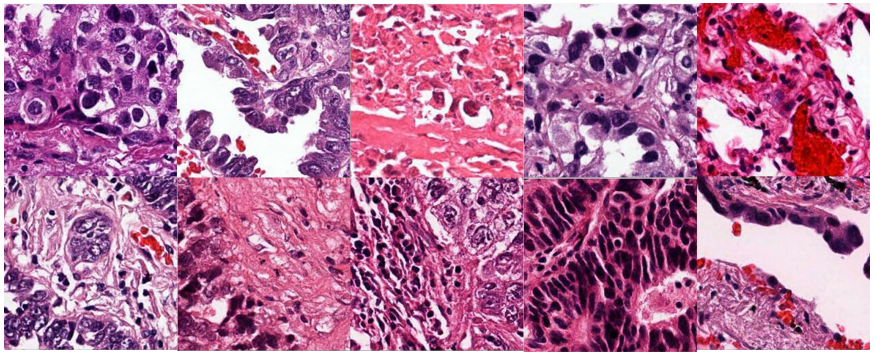
Simple Linear Regression

$$Y = a + bX$$

~500.000 parameters

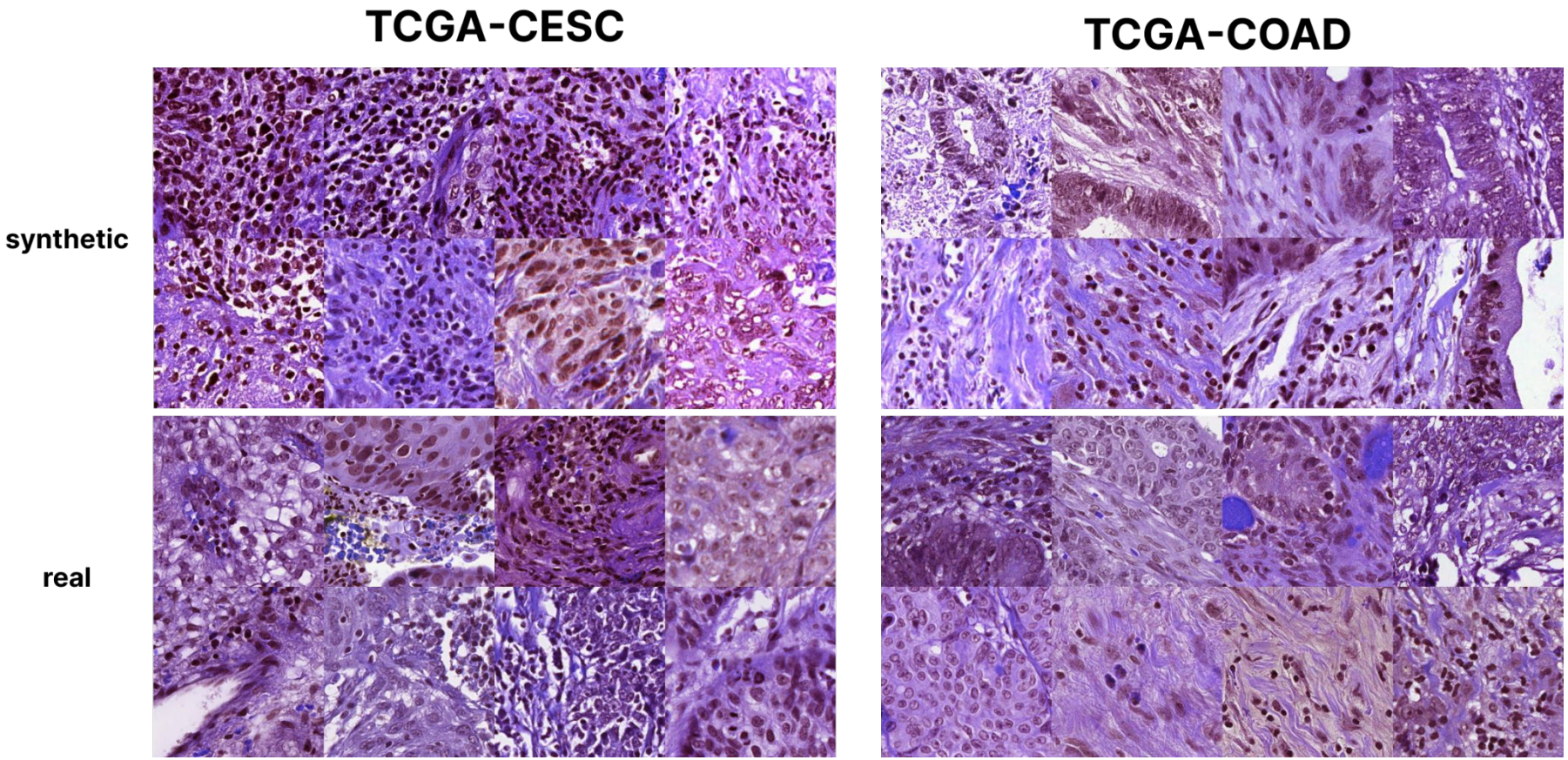


~1.9 billion parameters

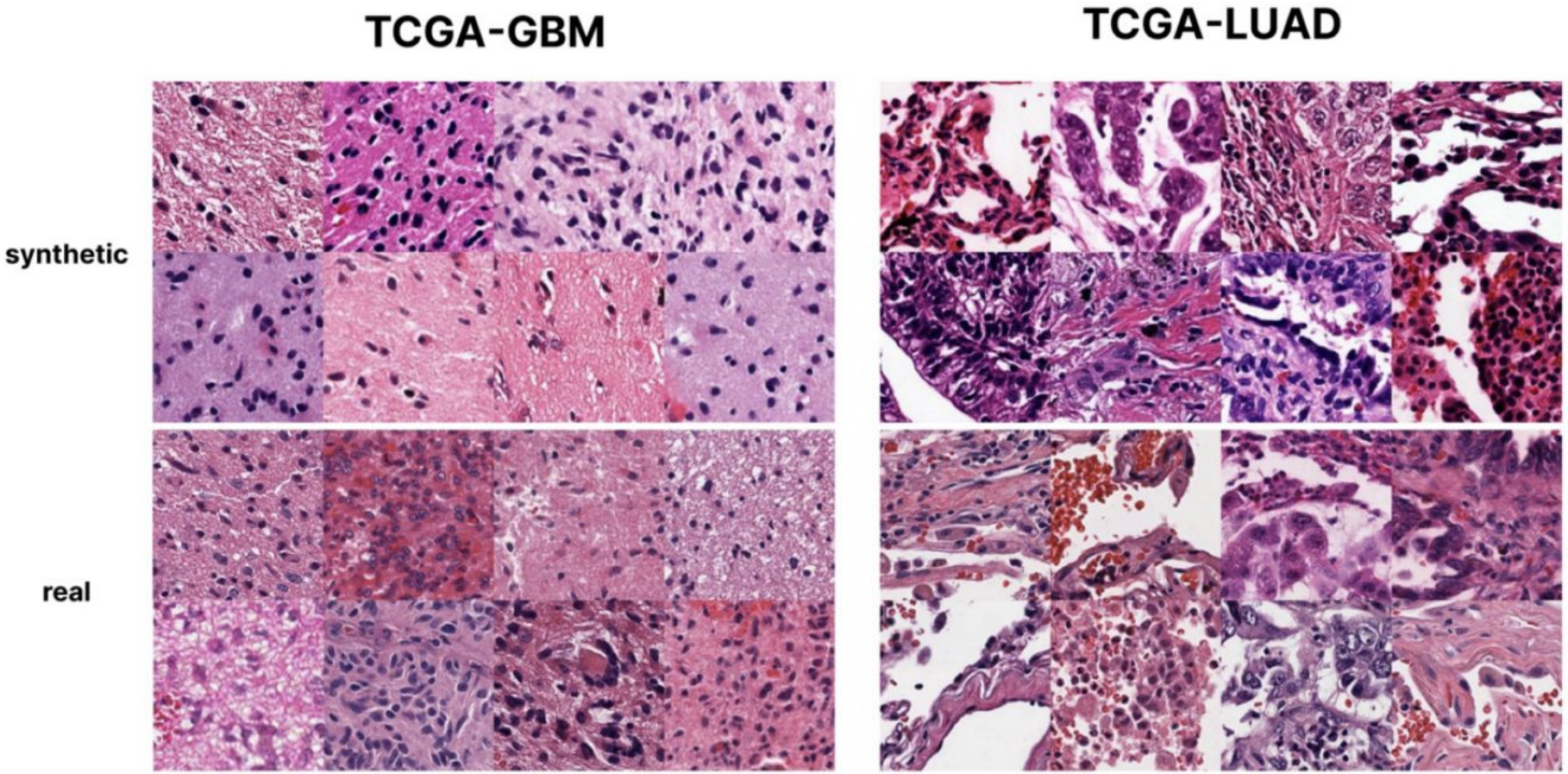


Trained in a distributed setting (8 GPUs) using the Hugging Face Accelerate Python library!

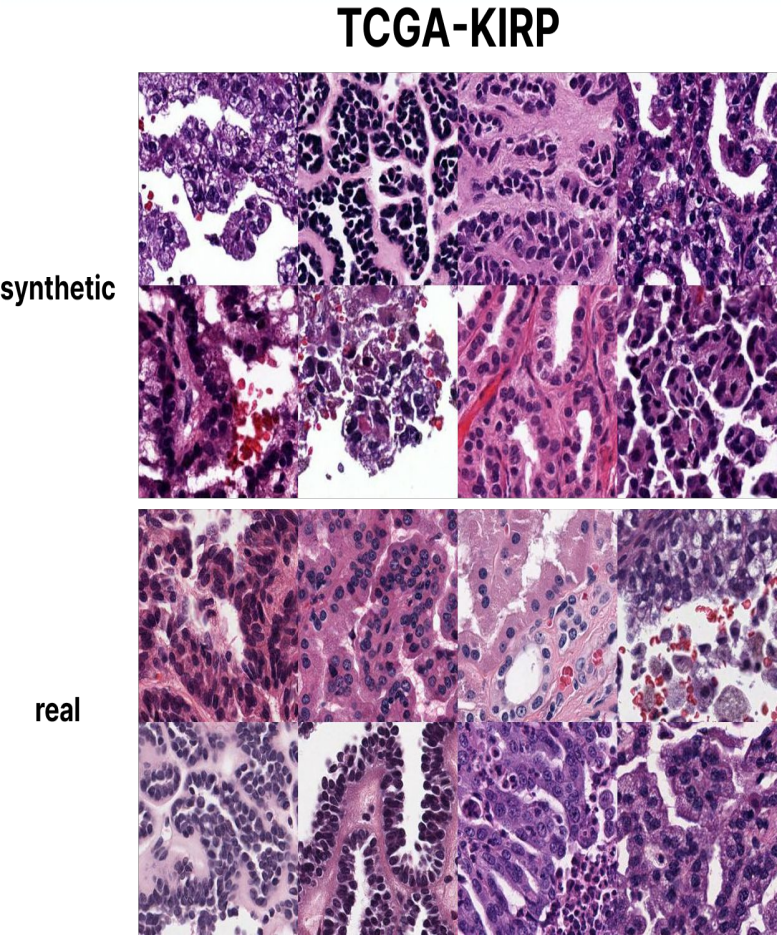
Synthetic tissue images



Synthetic tissue images

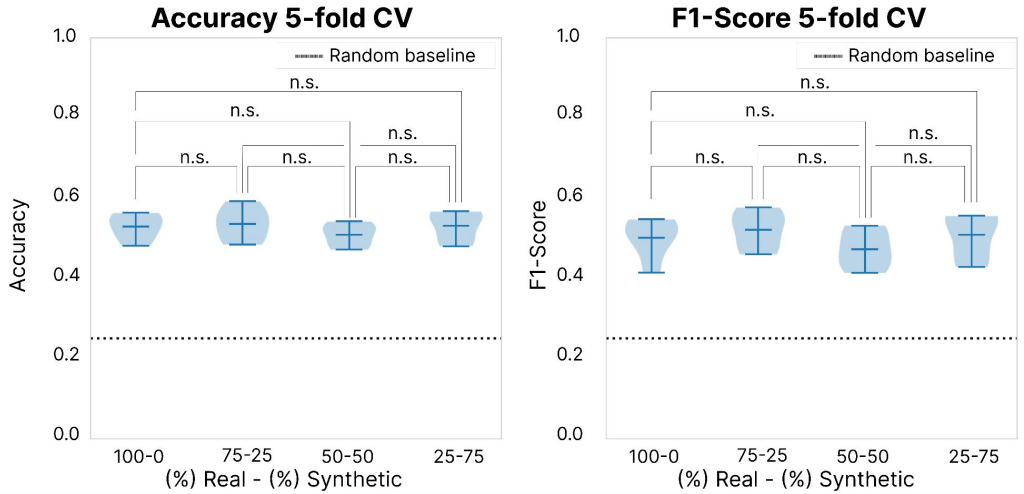


Synthetic tissue images



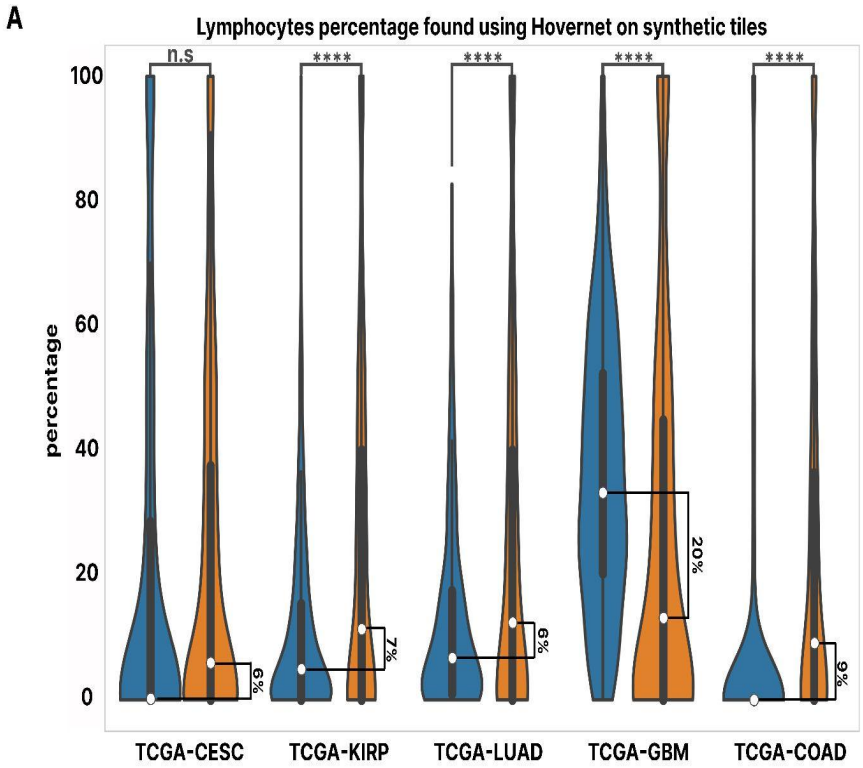
Model	Train CS (mean ± std)	Val CS (mean ± std)	Test CS
Steyaert et al[43]	0.900 ± 0.010	0.792 ± 0.070	0.854
Ours	0.806 ± 0.029	0.805 ± 0.058	**0.871**

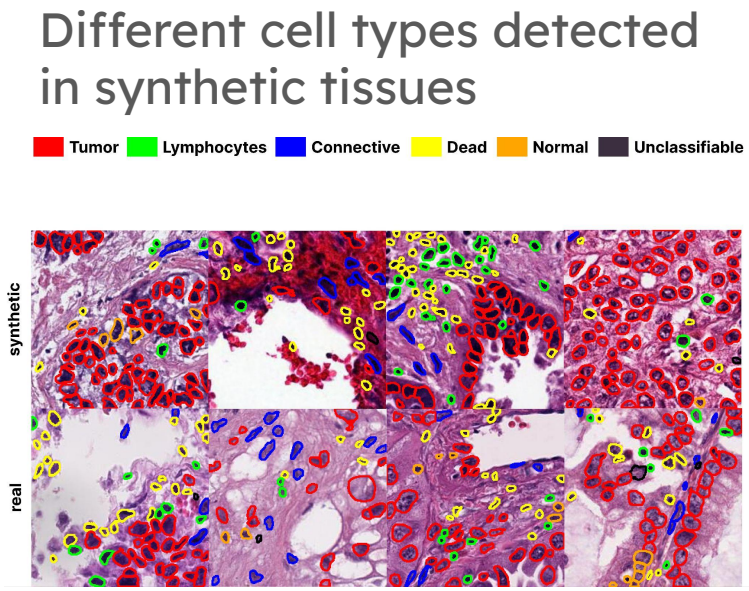
Using synthetic images as pre-training improves the results of ML models



We can substitute real images with synthetic ones without affecting the performance of machine learning models

RNA-Seq affects tissue morphology





Code and demo: https://rna-cdm.stanford.edu/

Examples across modalities!

Article | Open access | Published: 07 November 2024

Collaboration between clinicians and vision–language models in radiology report generation

Ryutaro Tanno, David G. T. Barrett, Andrew Sellergren, Sumedh Ghaisas, Sumanth Dathathri, Abigail See, Johannes Welbl, Charles Lau, Tao Tu, Shekoofeh Azizi, Karan Singhal, Mike Schaekermann, Rhys May, Roy Lee, SiWai Man, Sara Mahdavi, Zahra Ahmed, Yossi Matias, Joelle Barral, S. M. Ali Eslami, Danielle Belgrave, Yun Liu, Sreenivasa Raju Kalidindi, Shravya Shetty, ... Ira Ktena

+ Show authors

Nature Medicine **31**, 599–608 (2025) | Cite this article

Radiology-Report

Article | Published: 11 December 2024

Self-improving generative foundation model for synthetic medical image generation and clinical applications

Jinzhuo Wang, Kai Wang, Yunfang Yu, Yuxing Lu, Wenchao Xiao, Zhuo Sun, Fei Liu, Zixing Zou, Yuanxu Gao, Lei Yang, Hong-Yu Zhou, Hanpei Miao, Wenting Zhao, Lisha Huang, Lingchao Zeng, Rui Guo, Ieng Chong, Boyu Deng, Linling Cheng, Xiaoniao Chen, Jing Luo, Meng-Hua Zhu, Daniel Baptista-Hon, Olivia Monteiro, ... Jia Qu

+ Show authors

Text-Images

Article | Open access | Published: 10 February 2024

Regression-based Deep-Learning predicts molecular biomarkers from pathology slides

Omar S. M. El Nahhas, Chiara M. L. Loeffler, Zunamys I. Carrero, Marko van Treeck, Fiona R. Kolbinger, Katherine J. Hewitt, Hannah S. Muti, Mara Graziani, Qinghe Zeng, Julien Calderaro, Nadina Ortiz-Brüchle, Tanwei Yuan, Michael Hoffmeister, Hermann Brenner, Alexander Brobeil, Jorge S. Reis-Filho & Jakob Nikolas Kather

Nature Communications **15**, Article number: 1253 (2024) | Cite this article

Digital Pathology-Biomarkers

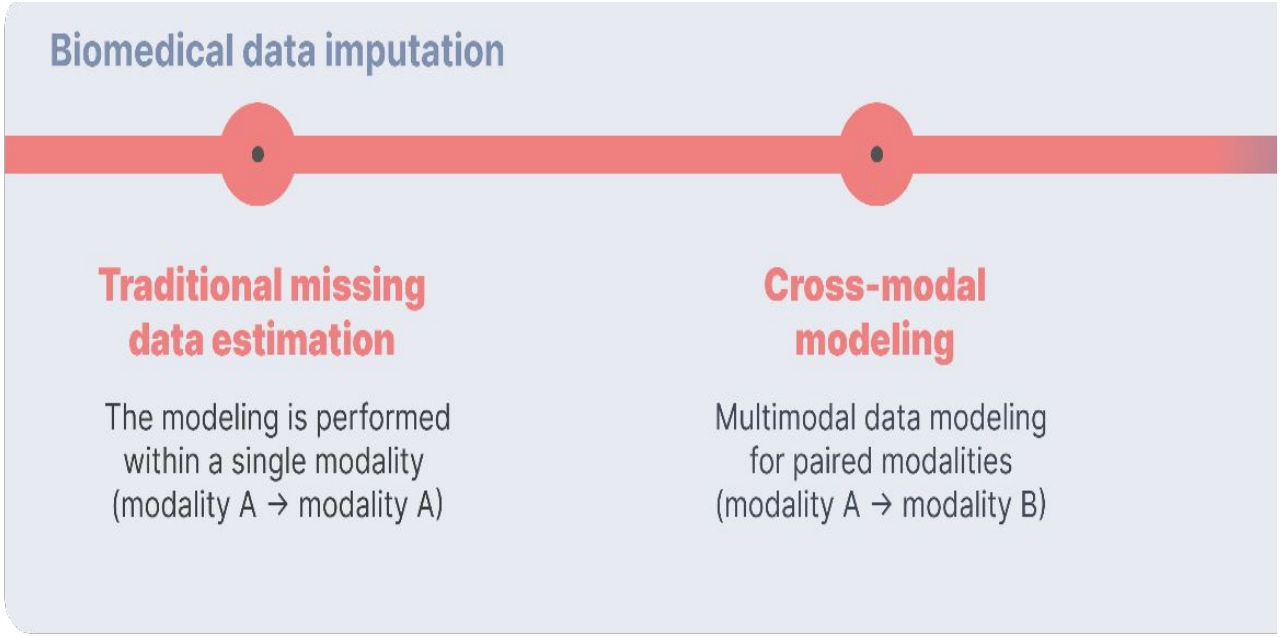
A multimodal generative AI copilot for human pathology

Ming Y. Lu, Bowen Chen, Drew F. K. Williamson, Richard J. Chen, Melissa Zhao, Aaron K. Chow, Kenji Ikemura, Ahrong Kim, Dimitra Pouli, Ankush Patel, Amr Soliman, Chengkuan Chen, Tong Ding, Judy J. Wang, Georg Gerber, Ivy Liang, Long Phi Le, Anil V. Parwani, Luca L. Weishaupt & Faisal Mahmood

Nature **634**, 466–473 (2024) | Cite this article

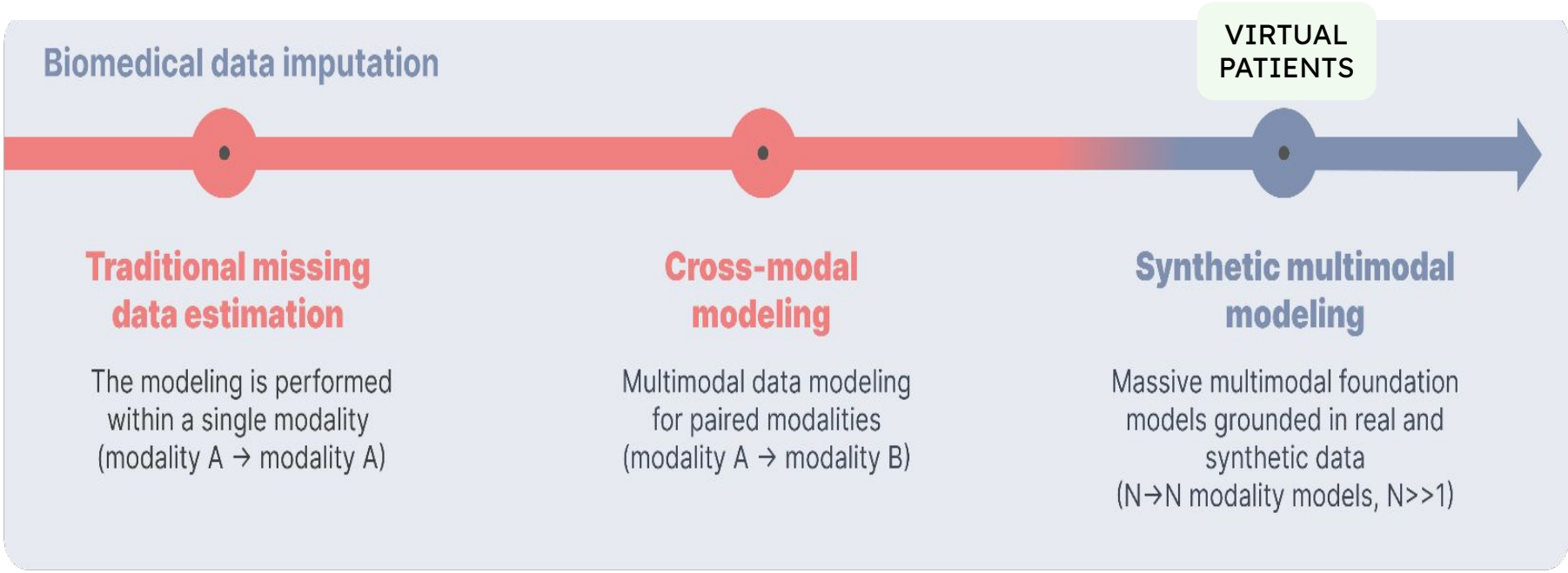
Digital Pathology-Report

Where are we going from here?



Future?
Virtual patients

Where are we going from here?



Cell
Leading Edge

Perspective

How to build the virtual cell with artificial intelligence: Priorities and opportunities

Charlotte Bunne,[1,2,3,4,50] Yusuf Roohani,[1,3,5,50] Yanay Rosen,[1,3,50] Ankit Gupta,[3,6] Xikun Zhang,[1,3,7] Marcel Roed,[1,3] Theo Alexandrov,[8,9] Mohammed AlQuraishi,[9] Patricia Brennan,[3] Daniel B. Burkhardt,[11] Andrea Califano,[10,12,13]

CellPress
OPEN ACCESS

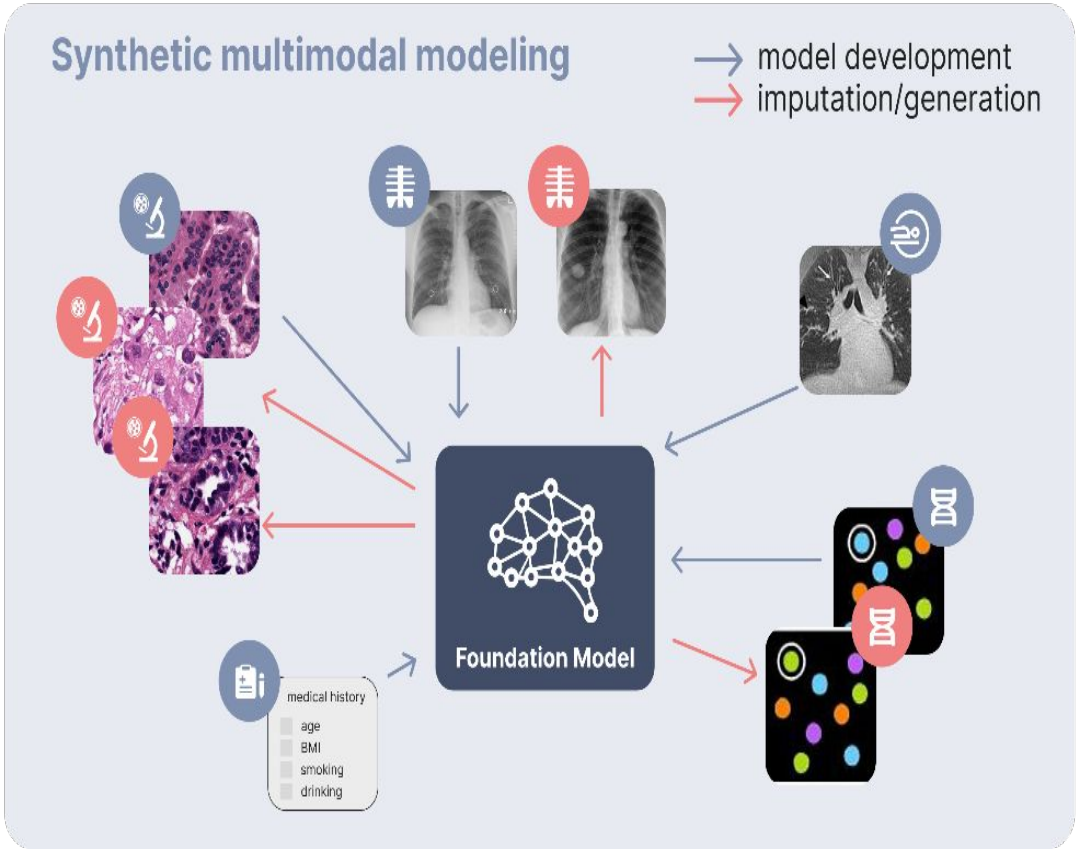
Comment | Published: 23 December 2024

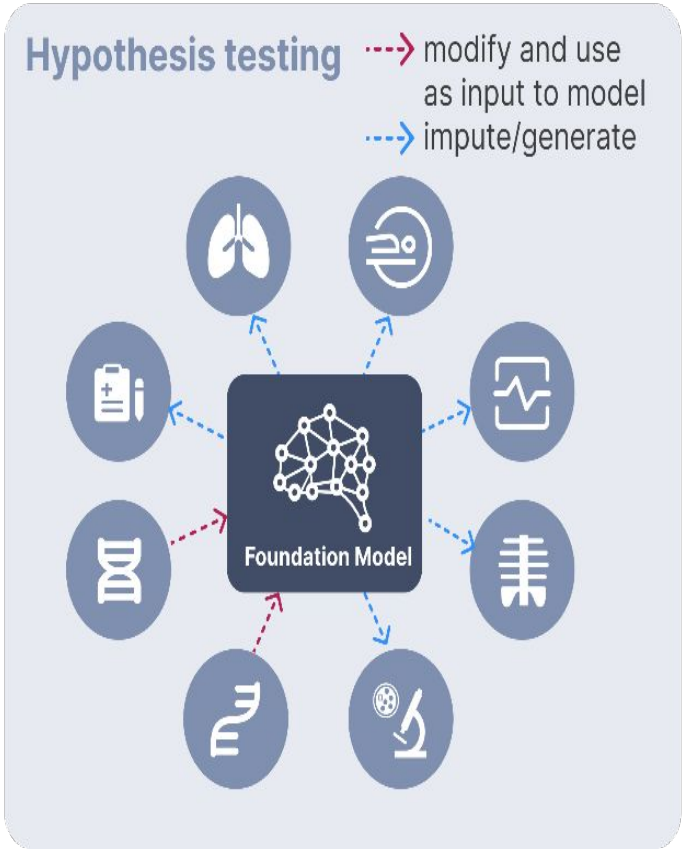
Synthetic multimodal data modelling for data imputation

Francisco Carrillo-Perez, Marija Pizurica, Kathleen Marchal & Olivier Gevaert

Nature Biomedical Engineering (2024) | Cite this article

Future of generative AI models in medicine?





Thanks for your attention!

carrilloperezfrancisco@gmail.com
BSKY: @pacocp.bsky.social
Github: @pacocp
Linkedin: Francisco Carrillo Pérez