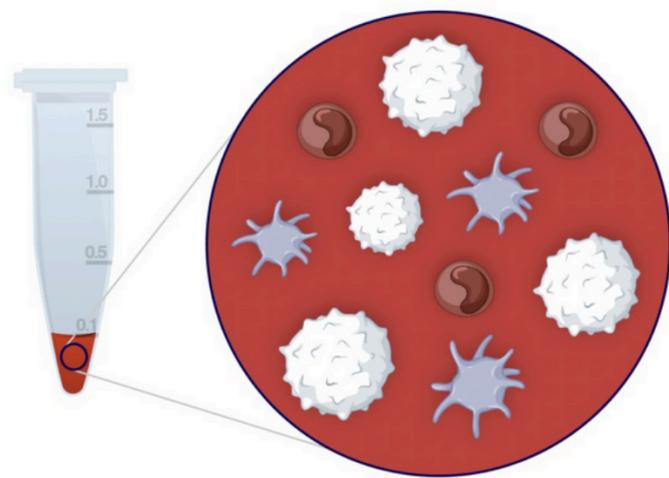


# Single Cell Omics e Medicina de Precisão na era da Inteligência Artificial



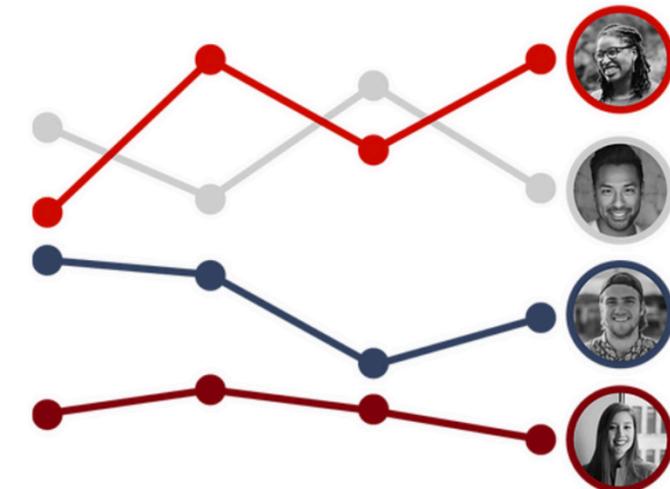
single cell omics

genes

0	4	0	0	2	0	0	3	0
0	0	0	0	0	0	0	0	1
0	0	0	1	0	0	2	0	0
2	0	0	0	0	0	0	0	0
0	0	1	0	0	1	0	4	0
0	1	0	0	1	0	0	0	1
0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	3	0	0
0	1	0	0	1	0	0	0	0

cells

matriz de contagem de genes



perfis de precisão

**Eduardo da Veiga Beltrame**

*Assistant Professor of Computational Biology, MBZUAI*

# Tópicos de hoje...

1. Trajetória pessoal
2. A MBZUAI e a nova escola de saúde pública digital
3. Introdução a single cell omics
4. Exemplos de novos estudos sendo realizados
5. Machine Learning e análise de dados
6. O panorama atual em IA e single cell

***Anotem as perguntas, discutiremos no final***

# Breve trajetória

Técnico Integrado em Eletrônica



B.S. Biological Physics



PhD, Bioengineering



Founding Scientist



Bioinformatics Lead

Visiting Scholar at UC Berkeley



Assistant Professor of Computational Biology



# *Emirados Árabes Unidos*

## Um país nascido em 1971



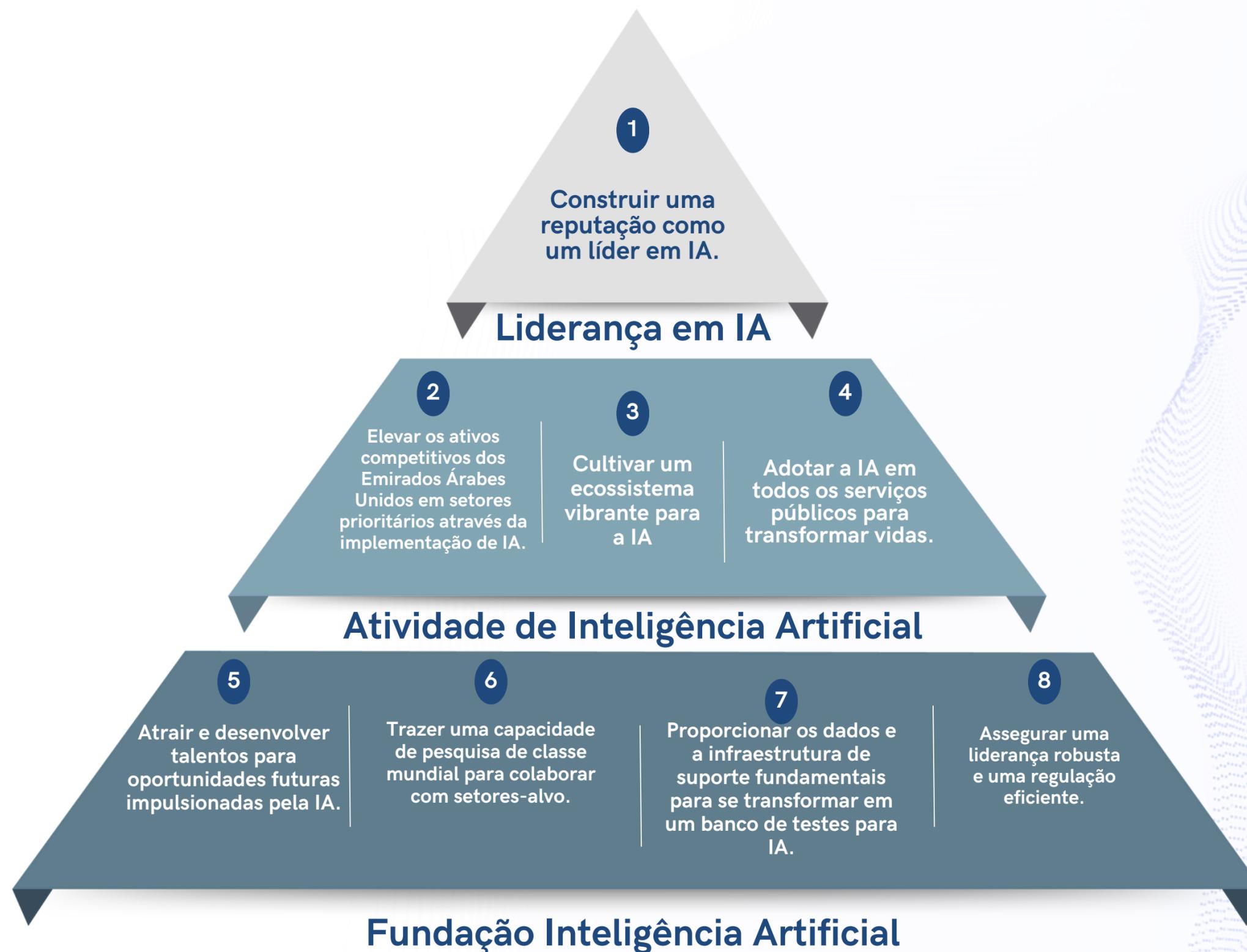
Estratégia de Inteligência Artificial dos Emirados Árabes Unidos até 2031



A inteligência artificial ocupa um lugar de destaque nas prioridades nacionais dos Emirados Árabes Unidos.

# Visão em IA dos Emirados Árabes Unidos

Transformarem-se em uma das nações pioneiras em IA até 2031.





# A missão da MBZUAI



Transformar-se numa instituição global de referência que revolucione a educação, promova a inovação em pesquisa e molde o panorama tecnológico mundial.  
Uma "Stanford" do Oriente Médio



**Paraíso para pesquisa transformadora**



**Lar para os mais brilhantes cientistas e engenheiros computacionais**



**Centro regional para o ecossistema de start-ups e inovação tecnológica**

# Abu Dhabi



**MBZUAI**



Mohamed bin Zayed  
University of  
Artificial Intelligence



Masdar City Campus



# MBZUAI em números

Uma nova universidade criada em 2019 focada em inteligência artificial

# 84

professores

# 365

alunos de  
pós-graduação\*

# 200

pesquisadores

## Programas de Mestrado e doutorado

- Processamento de Linguagem
- Machine Learning
- Visão Computacional
- Ciências da Computação
- Robótica
- Estatística e Análise de Dados

Abrindo em 2026:

- Biologia Computacional
- Interação Homem-Máquina

## Programa de Graduação

### Bacharelado em Inteligência Artificial

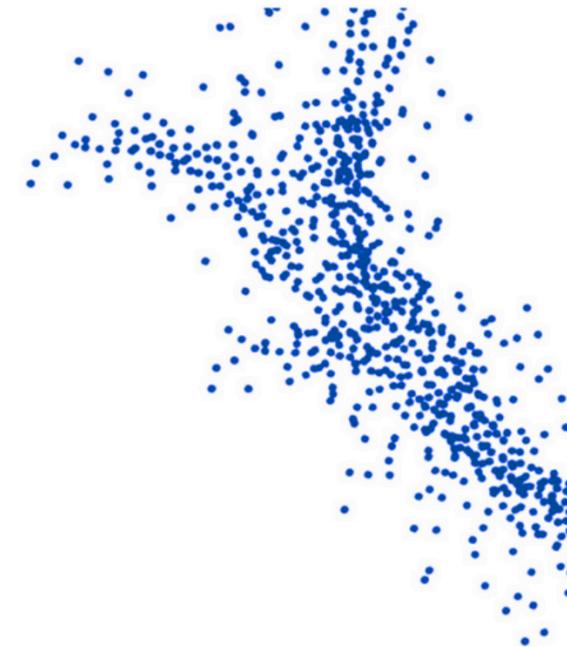
10º lugar no CSRankings a nível global  
em IA, CV, ML, PNL e Robótica.

1º na região do Oriente Médio e África

*\*Todos os estudantes do MBZUAI têm  
bolsas integrais e alojamento no  
campus.*



# Conquistas da MBZUAI



## EXCELÊNCIA ACADÊMICA

**RANKING:**  
**10°**

Nos Rankings Globais CSR em IA, Visão Computacional, Aprendizagem de Máquina, Processamento de Linguagem Natural e Robótica.

**80°-85°**  
Ranking geral da CS. A única universidade no MENA (excluindo 2 de Israel) entre as 100 melhores

**ARTIGOS:**  
**1200+**  
Publicados em conferências de topo até o momento



## DESENVOLVIMENTO DE TALENTOS

**GRADUADOS:**  
**213**  
Graduados até o momento

**EMPREGO:**  
**+80%**  
Mantidos no mercado de trabalho nos EAU

**TALENTO EMIRATI:**  
**20%**  
Naturais dos EAU no corpo docente



## IMPACTO NO ECOSSISTEMA

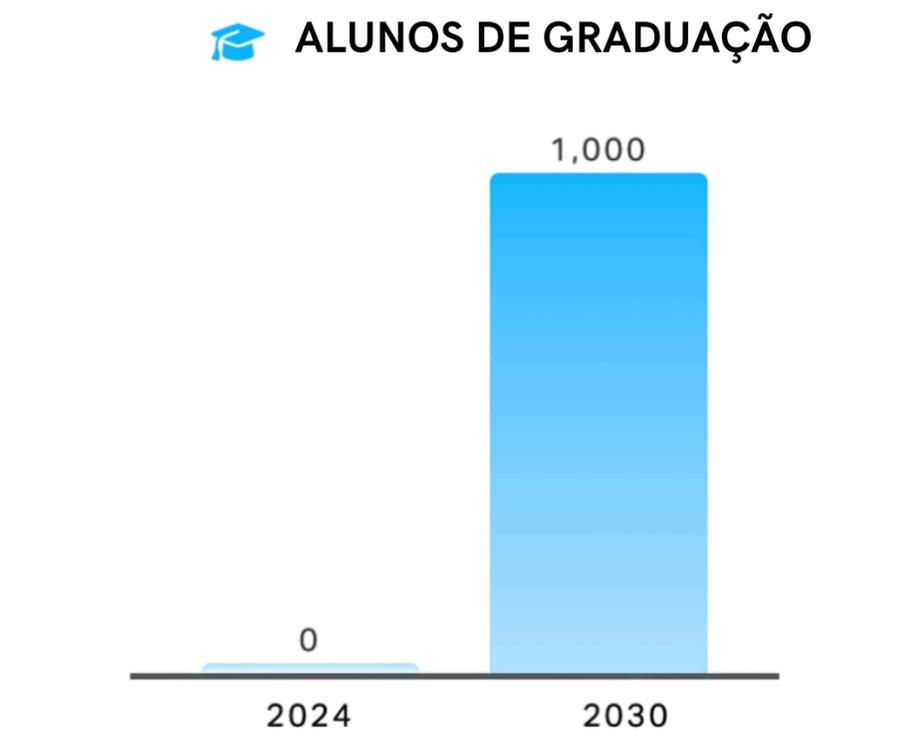
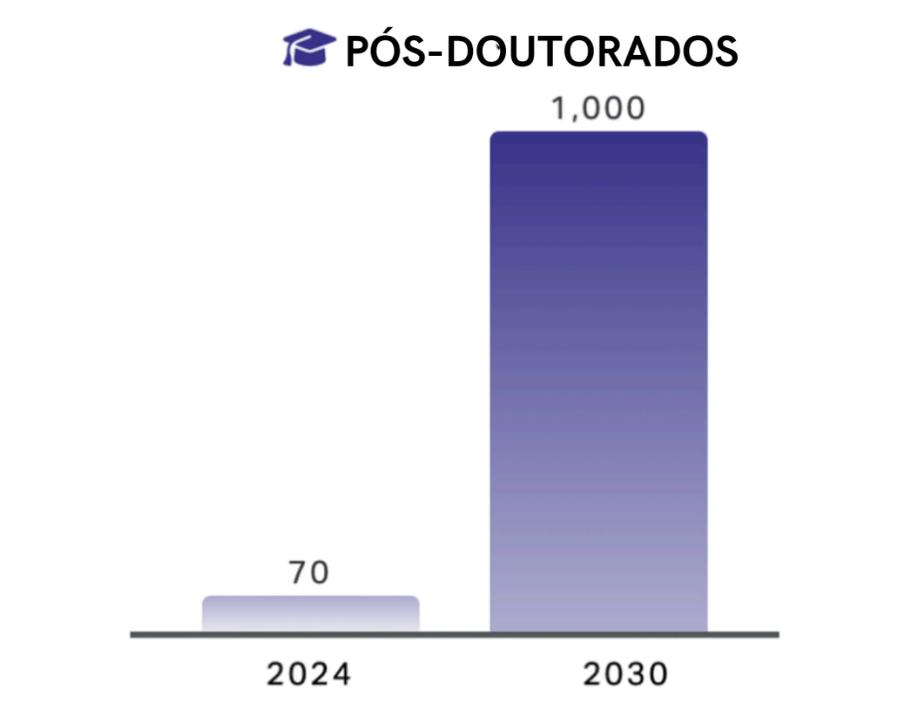
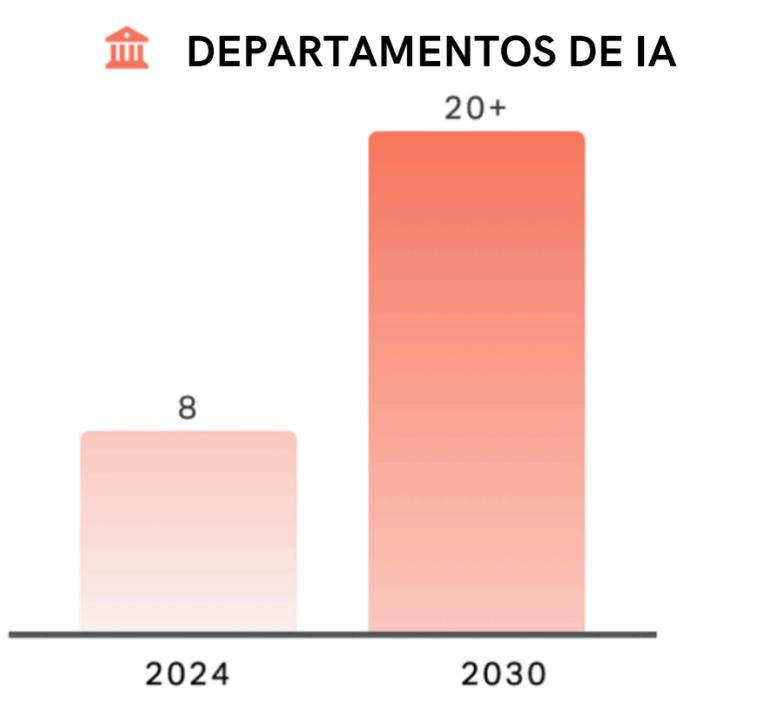
**STARTUPS DE IA**  
**10**  
Startups fundadas por docentes e estudantes

**P&D NA INDÚSTRIA:**  
Jais, Jais Climate  
K2, Nanda

**CONSCIENTIZAÇÃO E CAPACITAÇÃO EM IA NA COMUNIDADE:**  
**1000+**  
Capacitados em IA



# Planos de crescimento para a MBZUAI



# Novas direções

## IA Generativa e Modelos de Fundacionais



---

Criar sistemas versáteis para enfrentar desafios complexos e multidomínios

## AI para Medicina Personalizada



---

Acelerar o controle de doenças e o design de medicamentos por meio de predição biológica, programação e simulação em múltiplas escalas

## AI pela Sustentabilidade



---

Modelar padrões climáticos e energéticos para otimizar recursos e conservação



# MBZUAI School of Digital Public Health

*A Escola de Saúde Pública Digital da MBZUAI*

## Departamentos

- Biologia Computacional
- Medicina Personalizada
- Epidemiologia e Doenças infecciosas

## Vagas em Aberto:



## Áreas de Foco

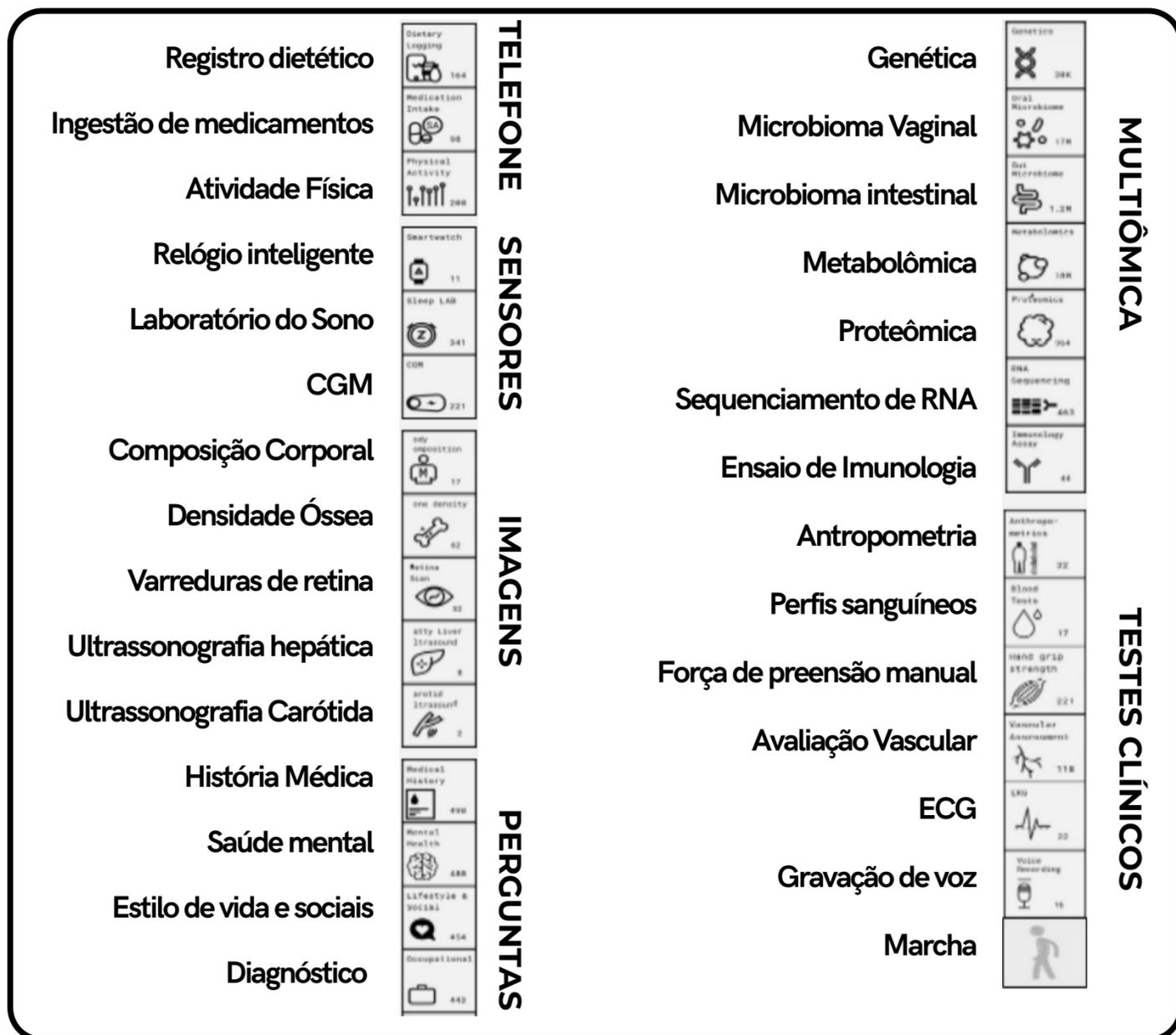
- Biologia Computacional e de Sistemas
- Informática médica e de saúde
- Medicina personalizada
- Modelos básicos de IA
- Análise de dados multiômica e integrativa
- Análise de nível populacional e EHR
- Epidemiologia
- Pandemias
- Imagens biomédicas

# Principais Projetos da *School of Digital Public Health*

## *The Human Phenotype Project*

*O Projeto Fenótipo Humano*

**13.000 pessoas com perfis detalhados**



## *Análise de dados do Programa de Genoma Emirati*

*Medicina personalizada está um passo mais perto com 800.000 pessoas contribuindo para o Emirati Genome Project*



O Programa Genoma dos Emirados foi lançado em Abu Dhabi em 2019 antes de ser expandido para todos os Emirados



Até o momento, foram coletadas amostras de sangue e amostras de DNA de cerca de 815.000 emiratis



Os Emirados Árabes Unidos pretendem coletar dados genéticos de 1 milhão de emiratis para promover tratamentos para doenças raras e crônicas



Os emiratis podem fornecer amostras em dezenas de centros de coleta aprovados em hospitais e clínicas em todos os sete emirados

# Apoio aos alunos

## Bolsas Integrais

**Mestrado:** AED 8,000/mês  
(USD 2,178)

**Doutorado:** AED 10,000/mês  
(USD 2,723)

**Graduação:** Caso a caso.

## Mais...

- Acomodação;
- Passagem aérea anual de retorno ao país de origem;
- Seguro de saúde e visto;
- Custos para participação em conferências.

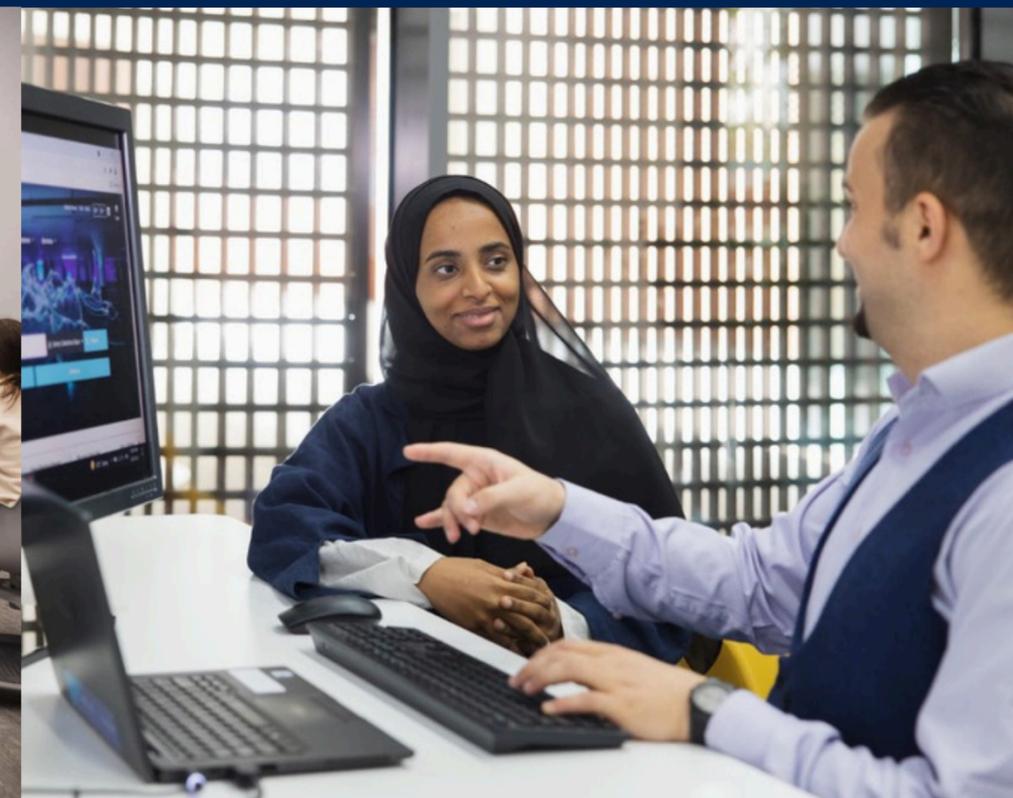
## Programas de Estudantes visitantes

### Graduação

4 semanas no verão

### Pós-Graduação

3 meses, ano todo



# Estudantes: Uma Universidade Global

## DEMOGRAFIA

(estudantes totais)



49

Nacionalidades



365

Estudantes no total



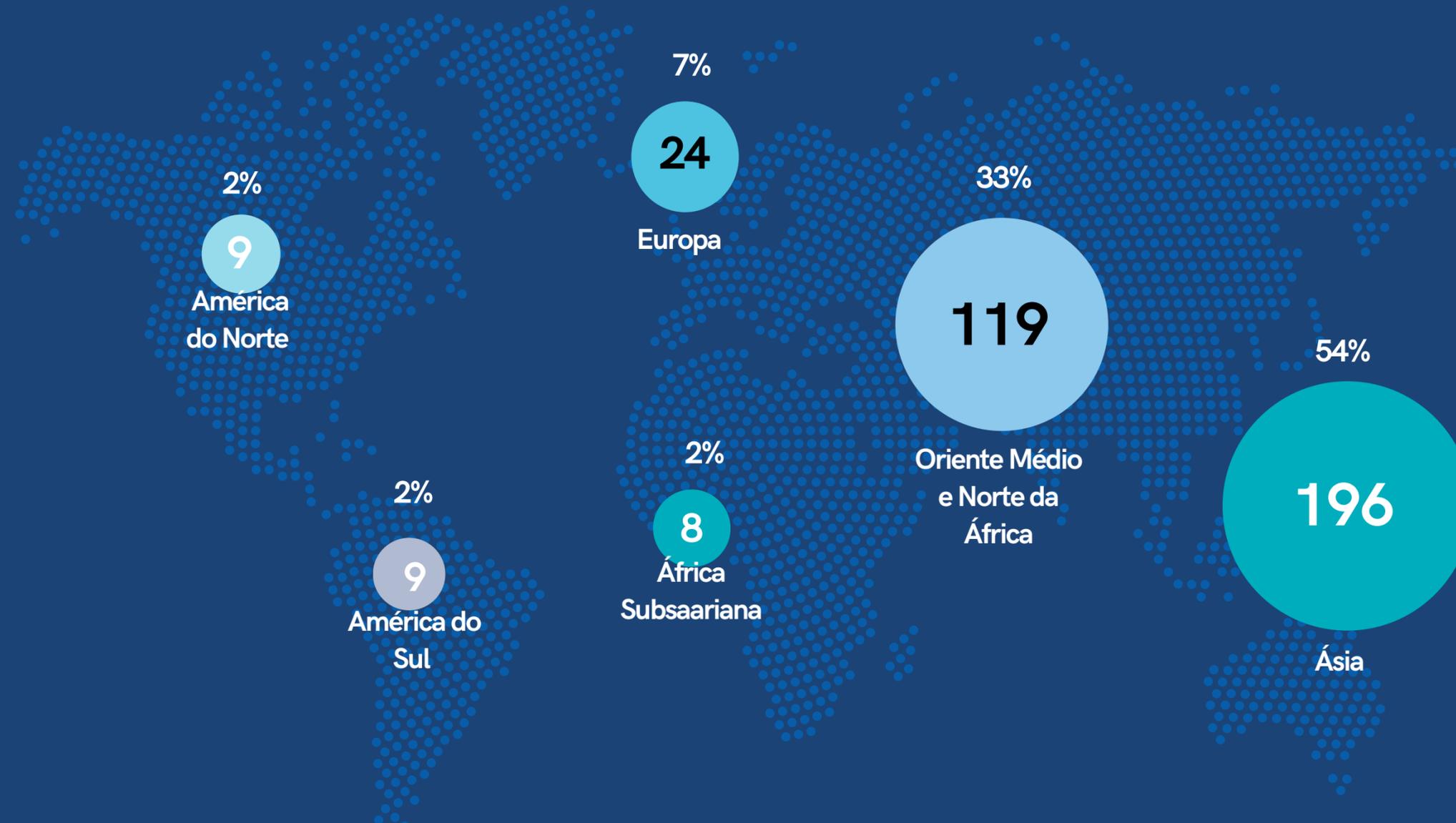
~30%

Estudantes Mulheres



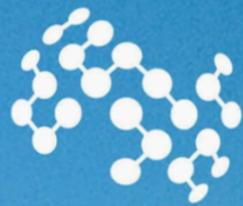
4.2:1

Relação aluno-professor



Composição de estudantes em Agosto de 2024





MOHAMED BIN ZAYED  
UNIVERSITY OF  
ARTIFICIAL INTELLIGENCE

Estude em **Abu Dhabi** na  
primeira **universidade  
especializada em IA**  
do mundo



 **Top 20  
global em IA\***

\*Classificações do CSRankings para IA, VC,  
AM, PLN e robótica.

 **Bolsas integrais  
e moradia**

**Mestrado: USD 2,178/mês**  
**Doutorado: USD 2,722/mês**

**INSCREVA-SE  
AGORA**

**Programas de  
pós-graduação**

**Inscriva-se até 31  
de maio!**

[mbzuai.ac.ae/study](https://mbzuai.ac.ae/study)



# Bacharelado em Ciências na área de Inteligência Artificial

Inscrições abertas até 31 de maio!



*O Bacharelado em Ciências em IA foi criado para o futuro.*

Nosso currículo alinhado à indústria garante que os graduados estejam preparados para o mercado de trabalho, motivados pela inovação e focados no impacto.

## 1) Fundamentos básicos de IA (todos os alunos)

- Aprendizado de máquina e aprendizado profundo
- Processamento de linguagem natural
- AI Ética e AI Responsável
- IA Generativa e Visão Computacional

## 2) Fluxos de especialização:

- Engenharia: Para quem deseja projetar soluções de IA. Algoritmos avançados de IA, redes neurais, sistemas inteligentes.
- Negócios: Para quem deseja aproveitar o poder da IA para a transformação empresarial. Tomada de decisões orientada por IA, IA em finanças, empreendedorismo com IA.

## 3) Aprendizagem prática:

- Estágios na indústria com líderes de IA
- Estágios em tempo integral na indústria, no mundo real
- Colaboração em pesquisa com pioneiros globais em IA

**Duração: 3 anos + 1 ano de estágio na indústria.**

**Formato: No campus com aprendizado integrado à indústria.**

**Um sonho  
inspirador:  
a medicina  
personalizada**



# O espectro de modelos

**Modelos  
analíticos**

**Grandes modelos  
*black box***



Pensamento  
da física

Pensamento  
de ML/AI

# Não vamos conseguir modelar toda a medicina analiticamente....



# O dogma central da Biologia Molecular

**DNA**

muito fácil de medir, mas é o mesmo em toda célula

*polimerase reversa*  
 **transcrição reversa**



**transcrição**

*RNA polimerase*

**RNA**

fácil de medir, diferente em cada célula



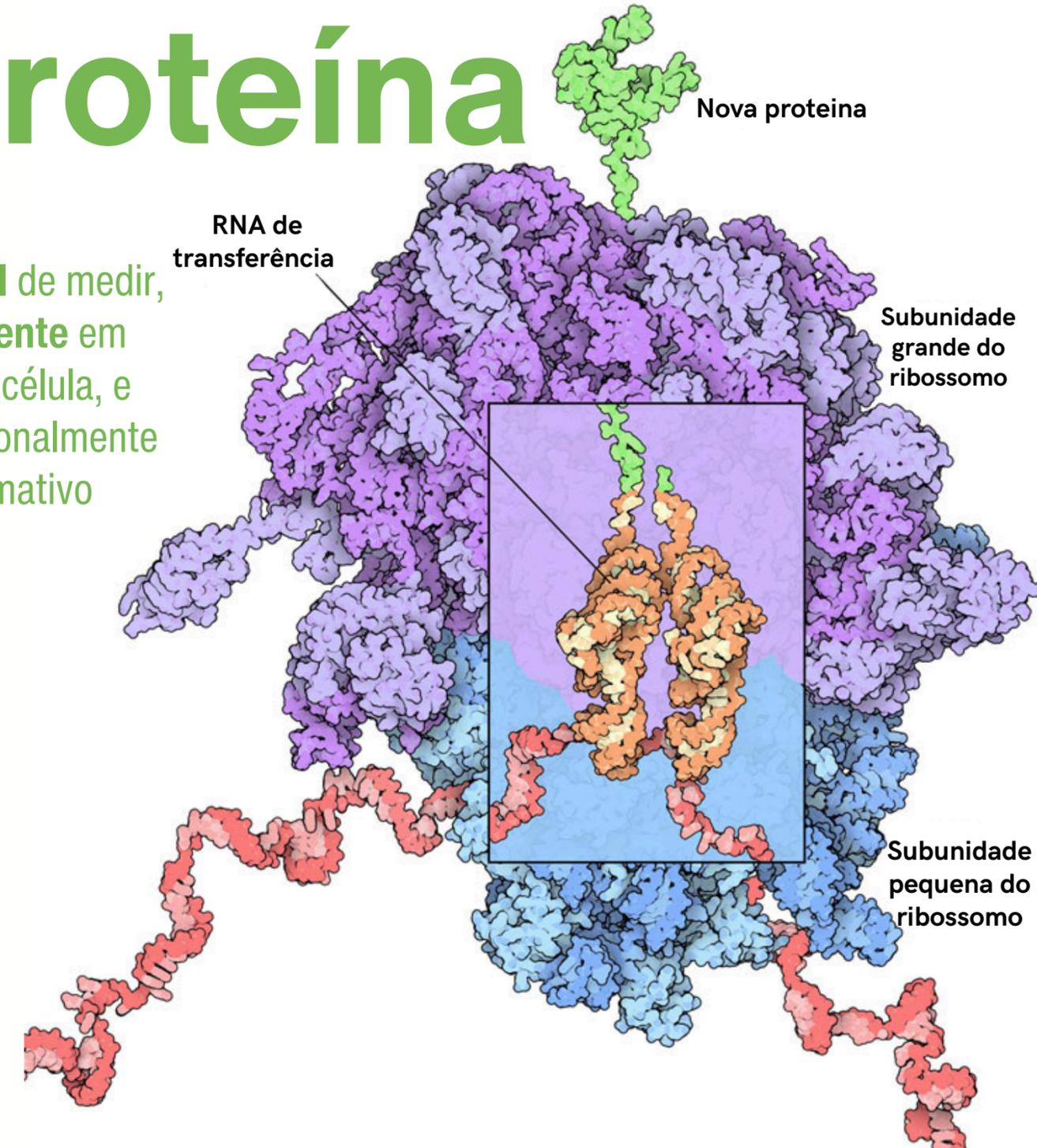
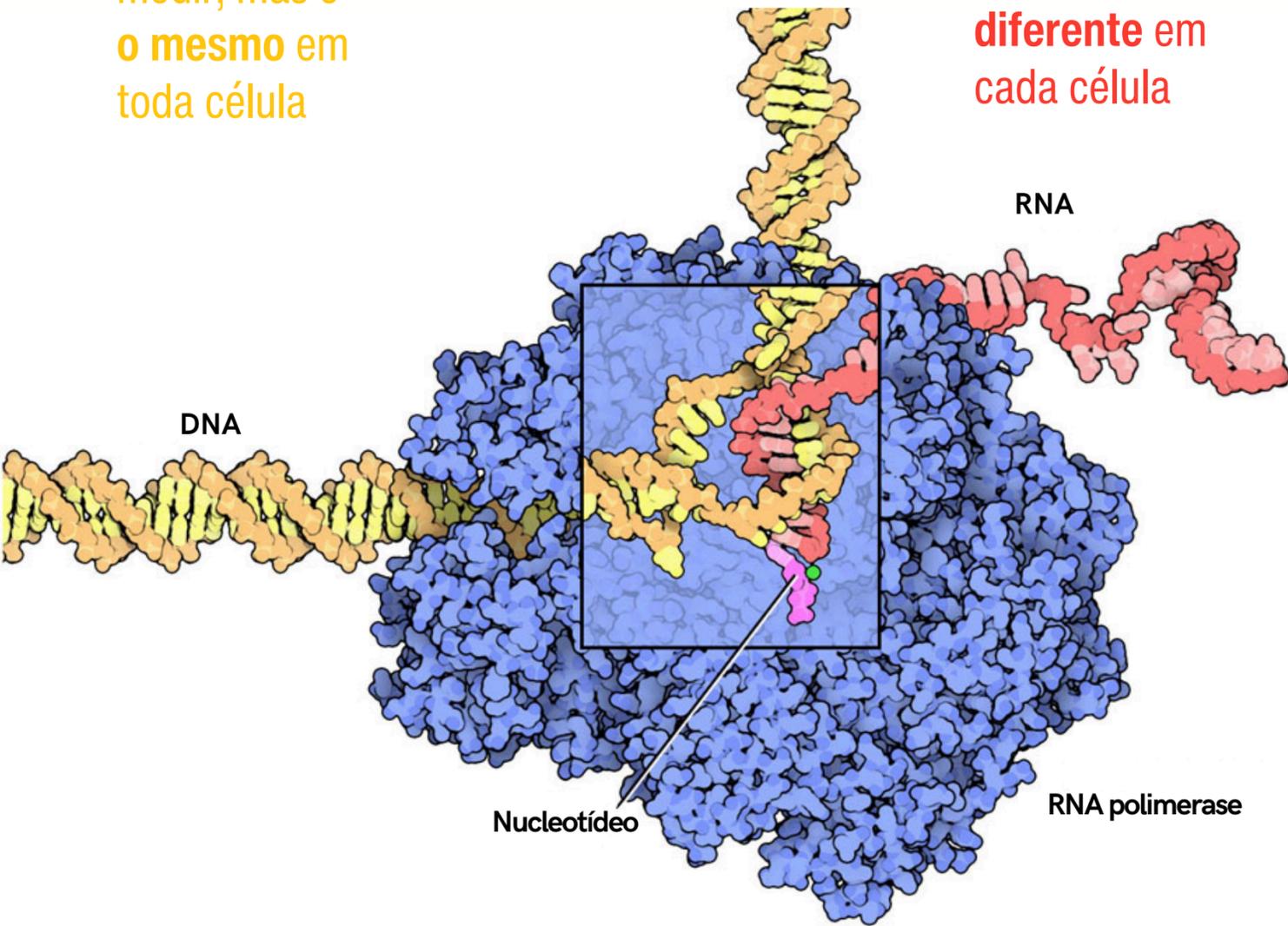
**tradução**

*Ribossomo*

**Proteína**

difícil de medir, diferente em cada célula, e funcionalmente informativo

Nova proteína



# RNA-seq: single cell vs bulk

RNA-seq (bulk)

Tecido heterogêneo



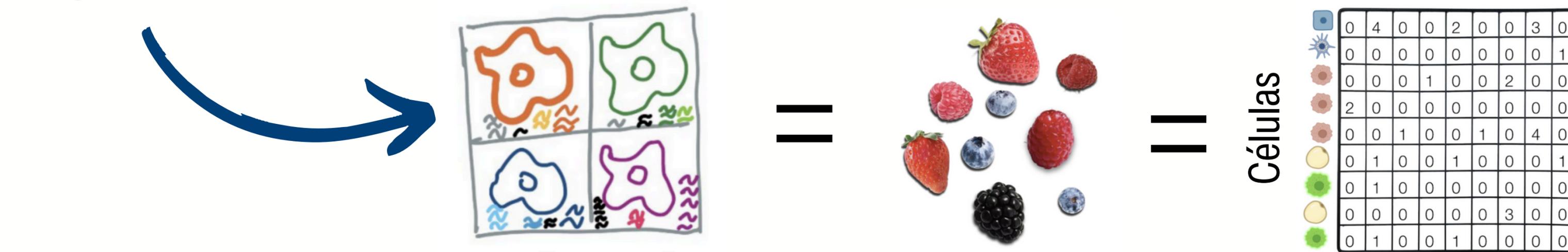
Vetor de médias

Amostra

genes



single cell RNA-seq



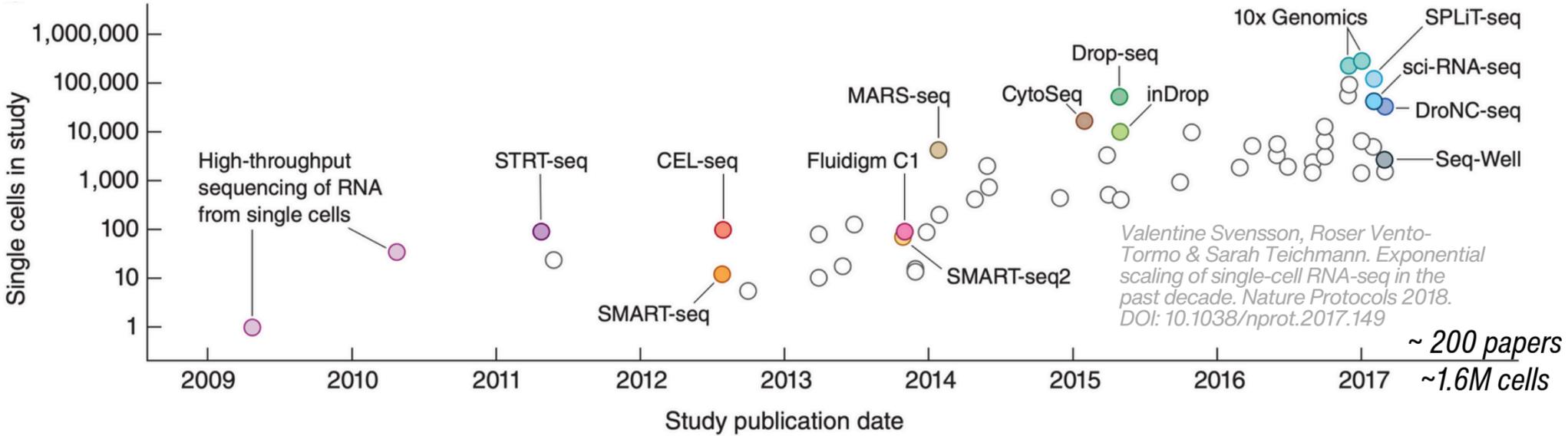
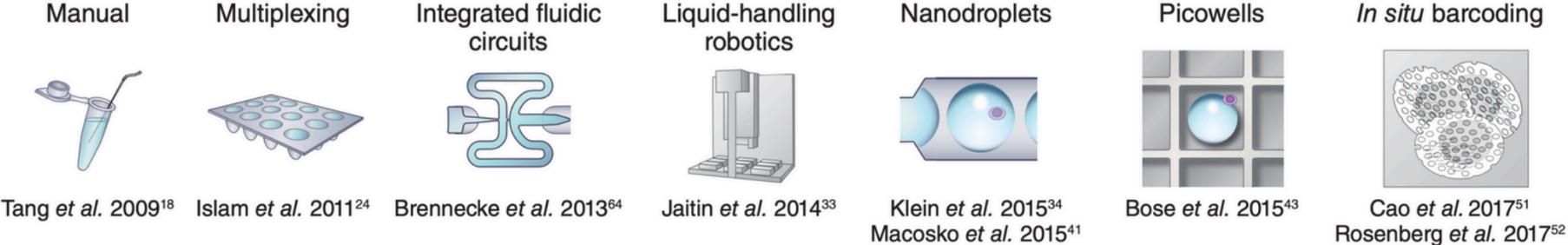
Matriz de contagem de genes

Genes

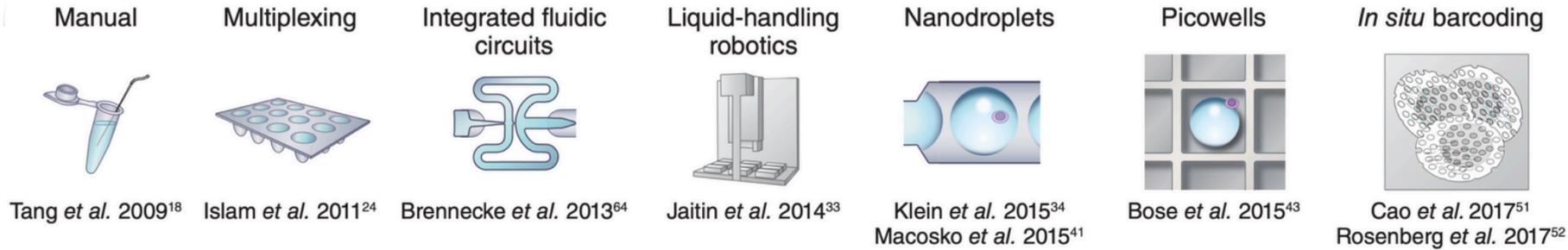
Células

	0	4	0	0	2	0	0	3	0
	0	0	0	0	0	0	0	0	1
	0	0	0	1	0	0	2	0	0
	2	0	0	0	0	0	0	0	0
	0	0	1	0	0	1	0	4	0
	0	1	0	0	1	0	0	0	1
	0	1	0	0	0	0	0	0	0
	0	0	0	0	0	0	3	0	0
	0	1	0	0	1	0	0	0	0

# Crecimiento exponencial de scRNA-seq

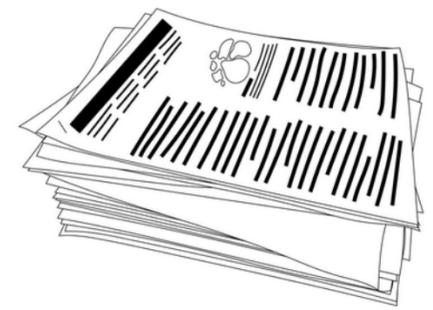
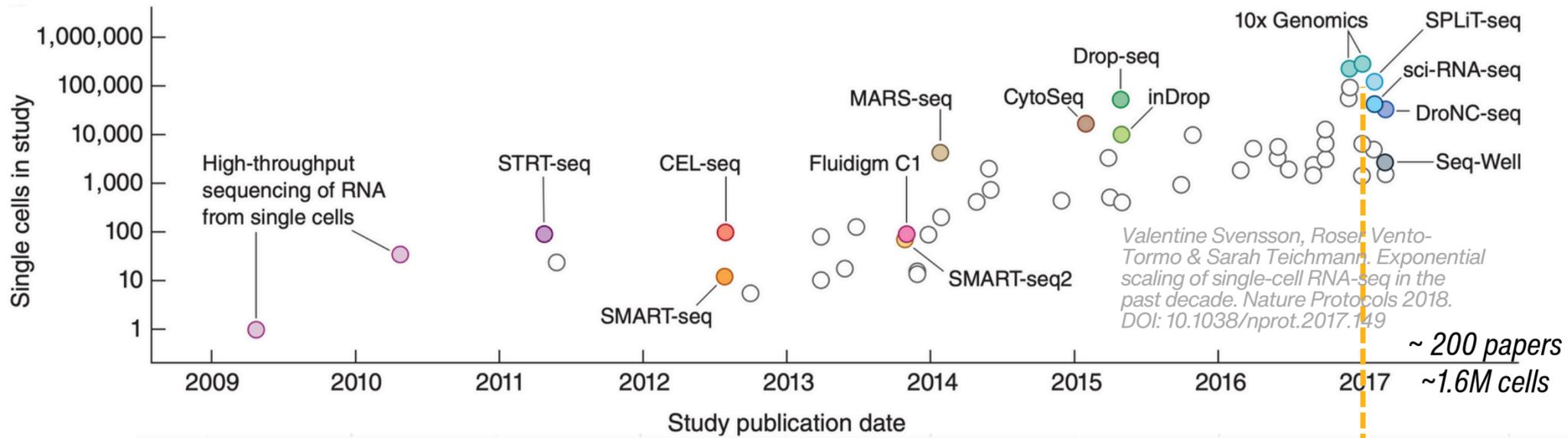


# Crescimento exponencial de scRNA-seq



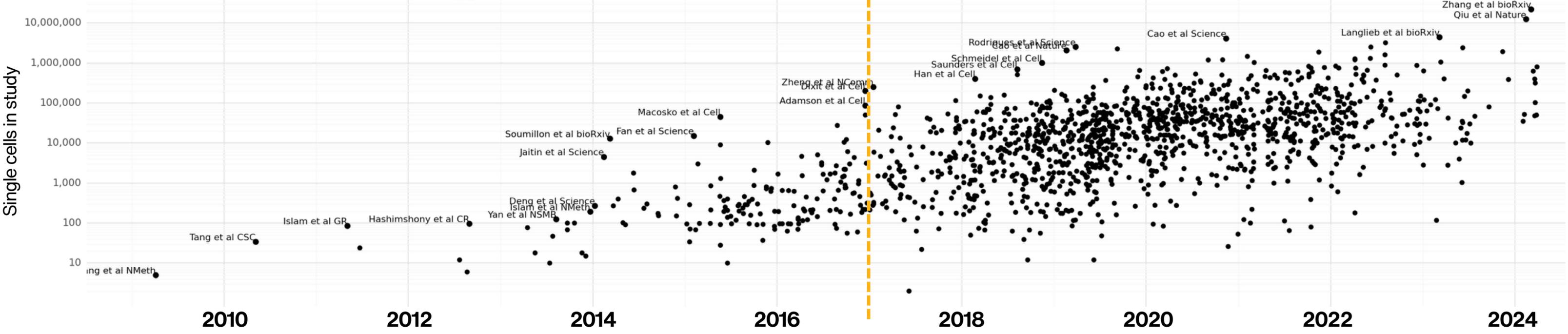
## The single cell studies database

<https://nxn.se/single-cell-studies>



Valentine Svensson, Eduardo da Veiga Beltrame, Lior Pachter, *A curated database reveals trends in single-cell transcriptomics. Database* 2020. DOI: 10.1093/database/baaa073

~2000 artigos  
~160M celulas



# Tahoe Open Sources Tahoe-100M, the World's Largest Single-Cell Dataset, as the Inaugural Contribution to Arc Institute's New Virtual Cell Atlas

300 million single cell atlas now accessible to the scientific community comprised of Tahoe's Tahoe-100M, mapping 60,000 drug-cell interactions, and Arc's AI-curated scBaseCount 200 million cell dataset

Generated using Tahoe's Mosaic platform, Tahoe-100M leveraged Parse Biosciences' GigaLab for single cell sample preparation and Ultima Genomics for sequencing.

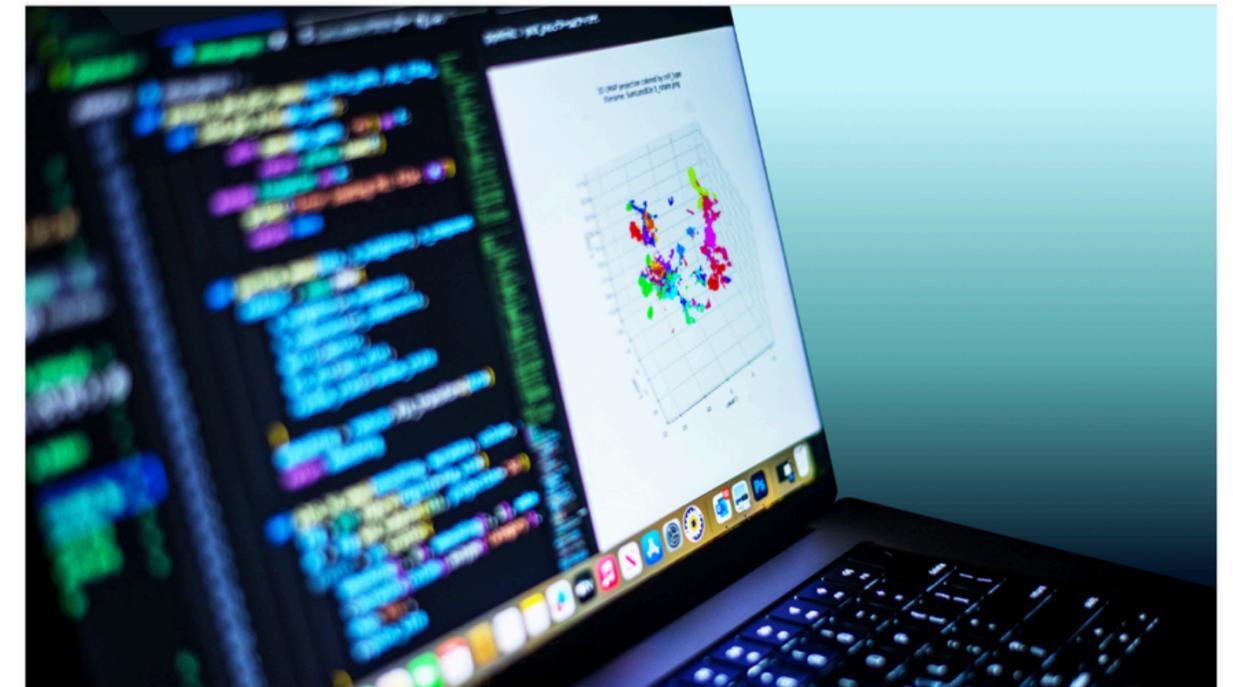


Arc Institute

## CZI Launches Billion Cells Project With 10x Genomics and Ultima Genomics To Advance AI in Biology

Landmark single-cell dataset of one billion cells will be used to train new AI models to advance researchers' understanding of cellular behavior and gene function.

Tags: [Science](#), [Technology](#)



• Photo by Zachary DeBruine.

T

oday, the Chan Zuckerberg Initiative (CZI) launched the Billion Cells Project, an effort to generate an unprecedented one billion cell dataset to fuel rapid progress in AI model development in biology. CZI is embarking on this landmark initiative in collaboration with partners 10x Genomics, Ultima Genomics, and a group of leading researchers. Industry partners are enabling data generation at an unprecedented scale by increasing access to cutting-edge technologies and lowering costs.

Once completed, this single-cell data set will bring critical new data and resolution to multiple domains of biology that need comprehensive resources such as mapping genetic perturbations across diverse cell types and tissues. Critically, the scale and cohesion of the data made possible by current and future contributions to the Billion Cells Project will deliver the volume of data with greater consistency than past efforts to better enable researchers to train AI models and make transformative discoveries across precision medicine and functional genomics.

# Panorada de monitoramento biomolecular

Riqueza do que se mede

#Características

(genes, proteínas...)

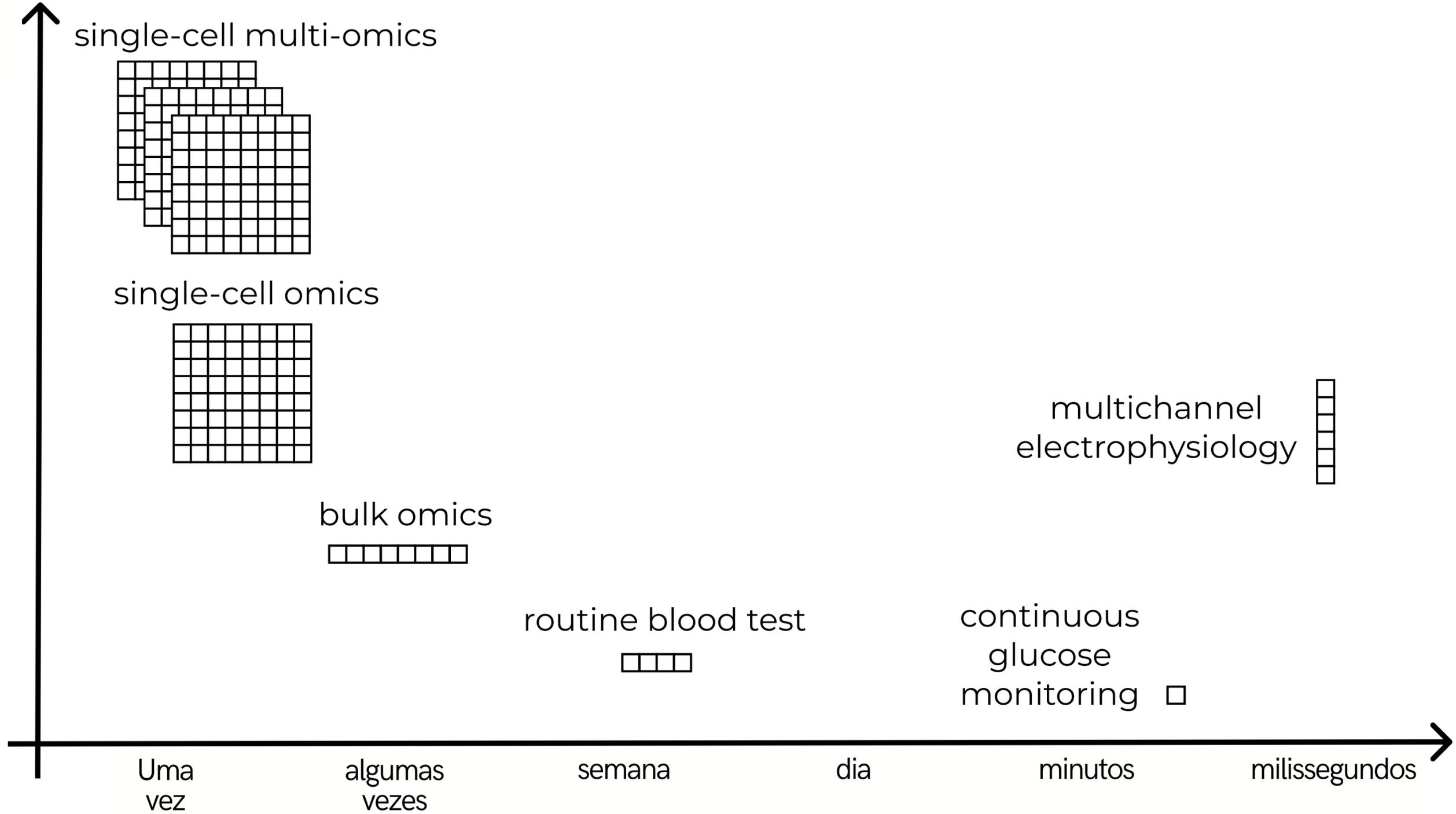
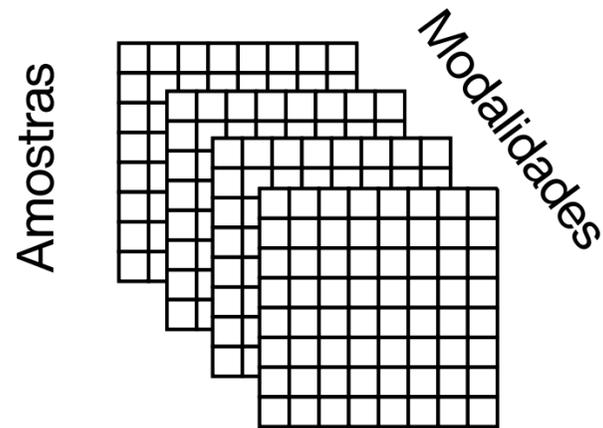
#Modalidades

(RNA, DNA, proteína...)

Amostras

(Células, eletrodos...)

Características



Frequência do que se mede

# Panorâmica de monitoramento biomolecular

Riqueza do que se mede

#Características

(genes, proteínas...)

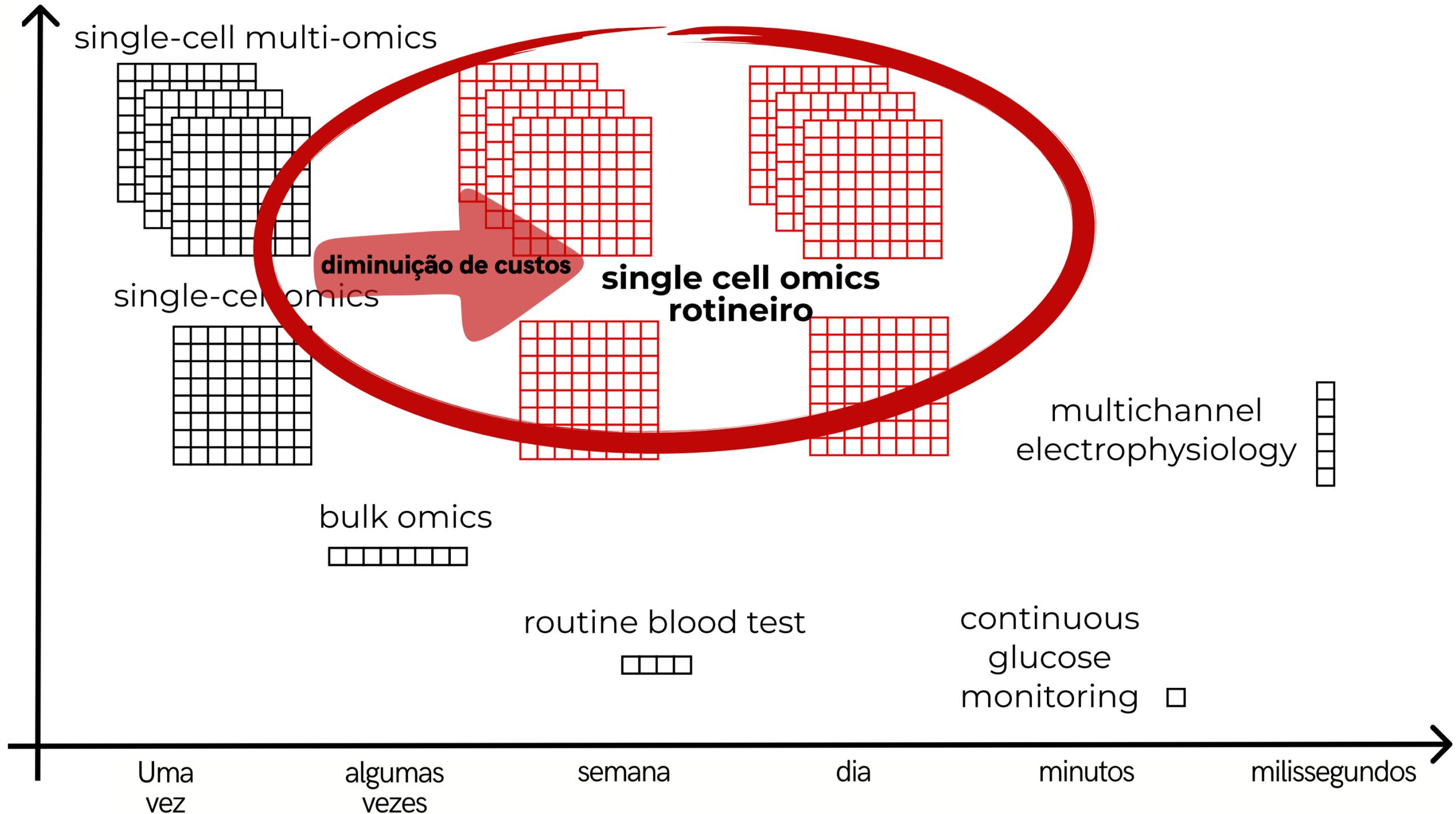
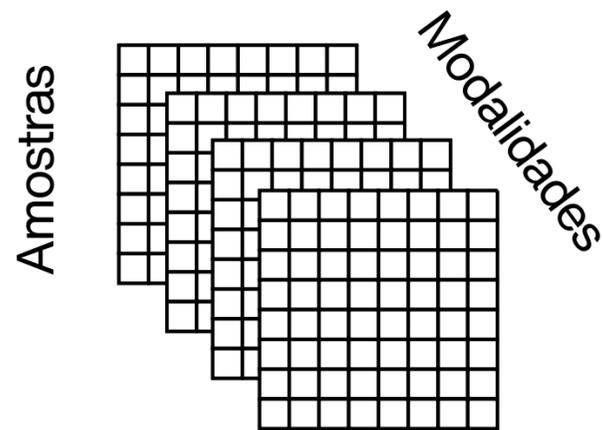
#Modalidades

(RNA, DNA, proteína...)

Amostras

(Células, eletrodos...)

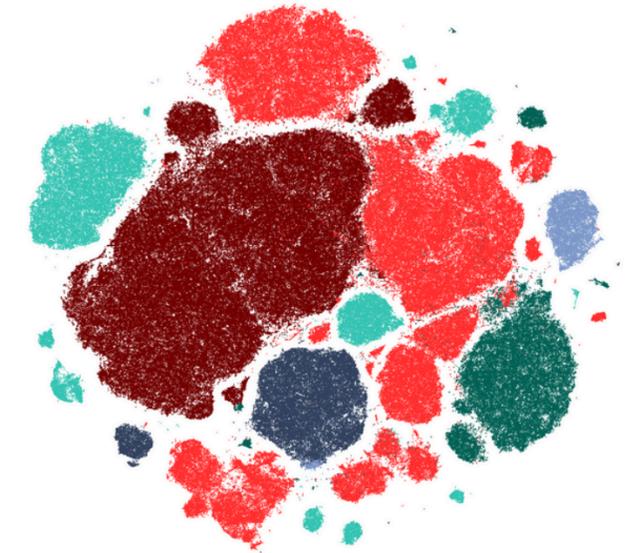
Características



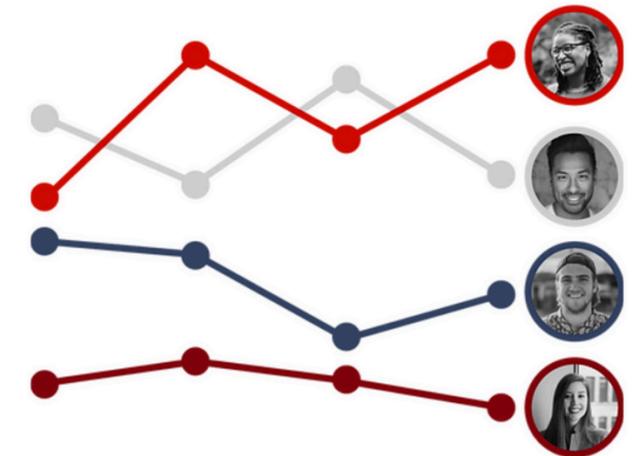
Frequência do que se mede

# Estudos descentralizados com *single cell*

*Atlas de referência*



*Estudos longitudinais*



**Imyoo**  
imyoo.com

*Coleta própria de 100ul  
de sangue capilar,  
envio pelo correio*

**Pontos críticos**  
**Longitudinal**  
**Alta resolução**  
**Populações diversas**

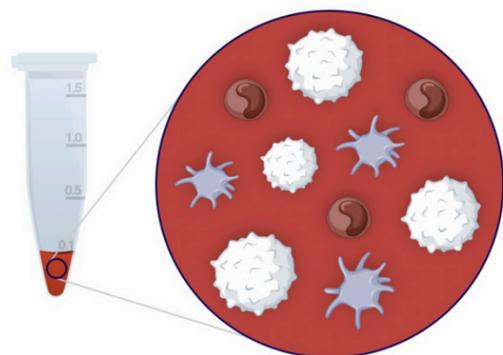


*Dispositivo TAP II  
(yourbiohealth.com)*

*amostragem em casa*

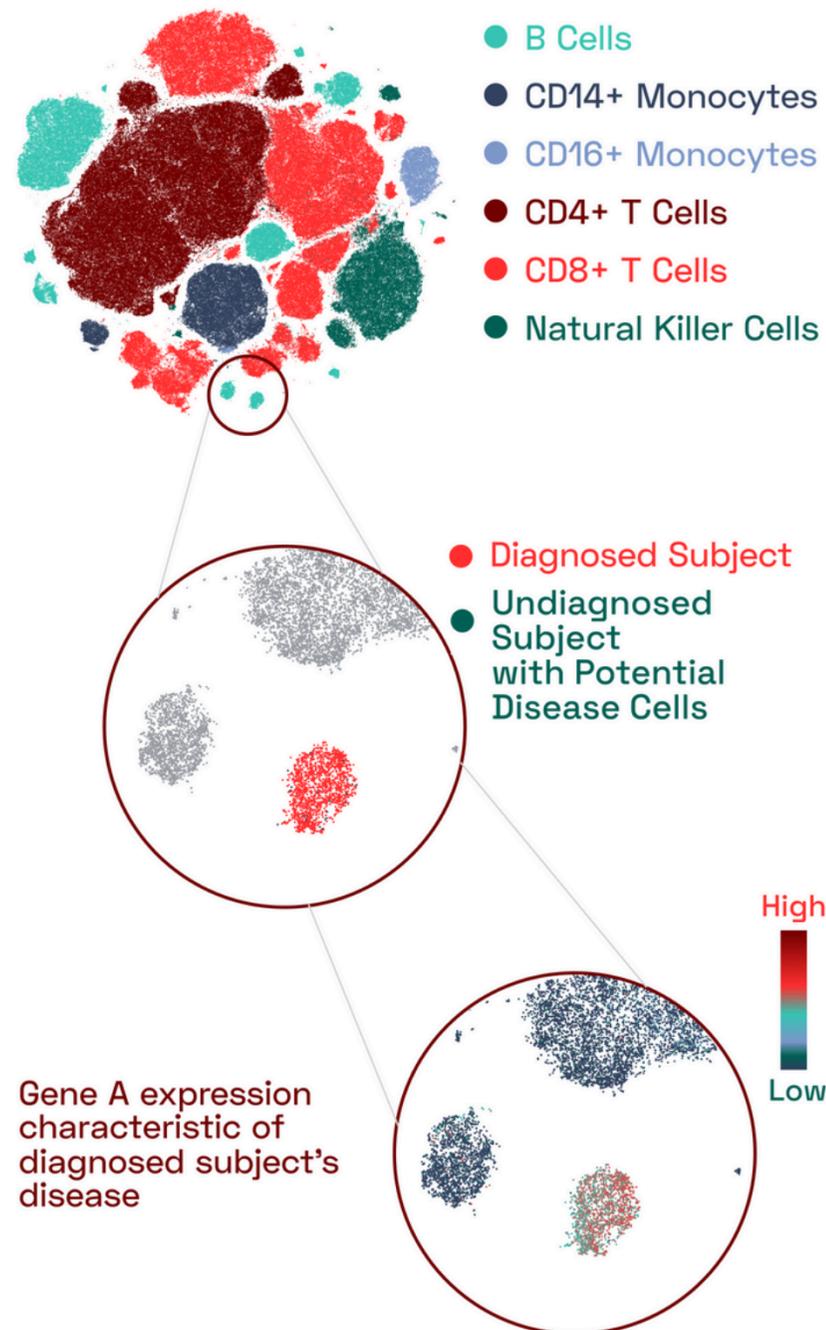
+

*single cell omics*

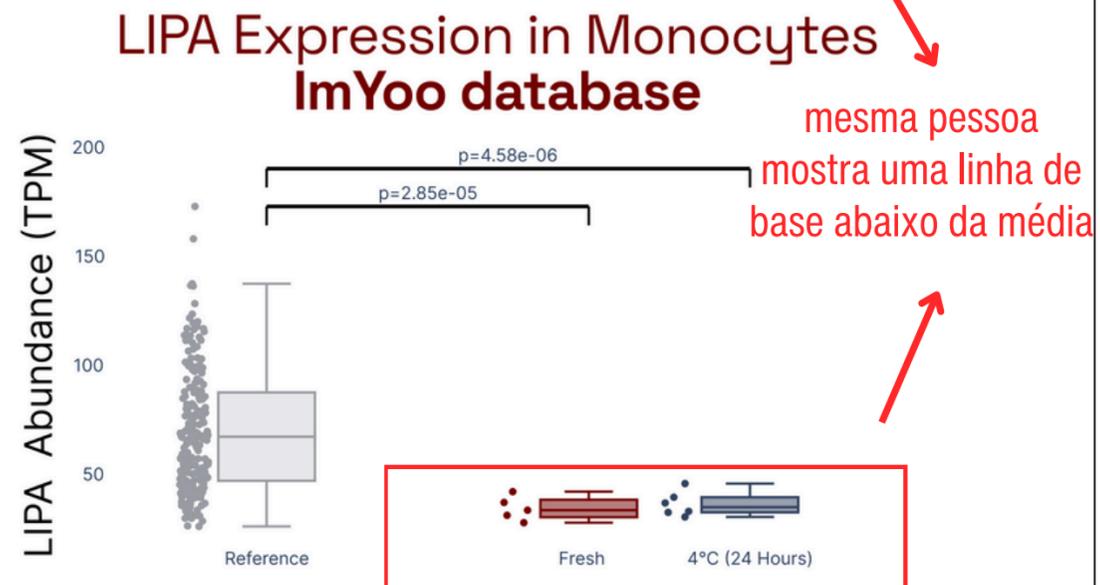
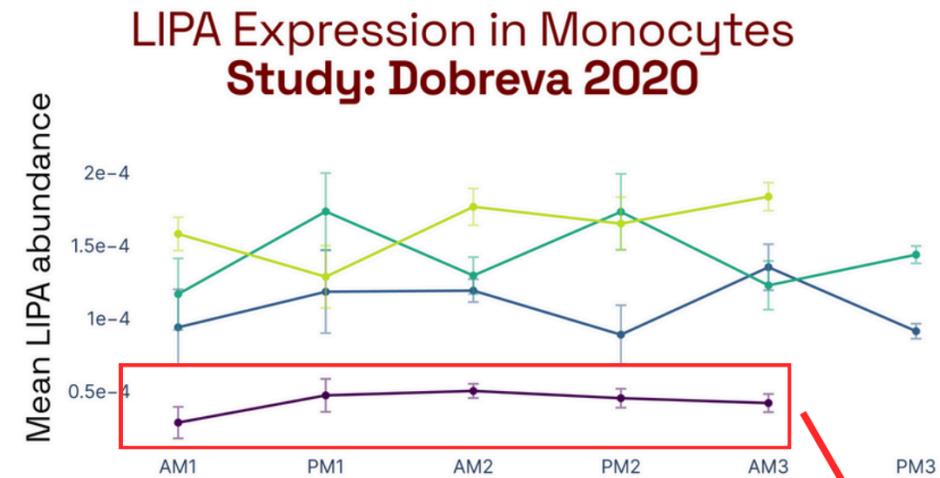


# Exemplos de descobertas em single cell

## Tipos de células raras e agrupamento anormal



## Assinaturas de genes específicas da pessoa e do tipo de célula



Single cell profiling of capillary blood enables out of clinic human immunity studies

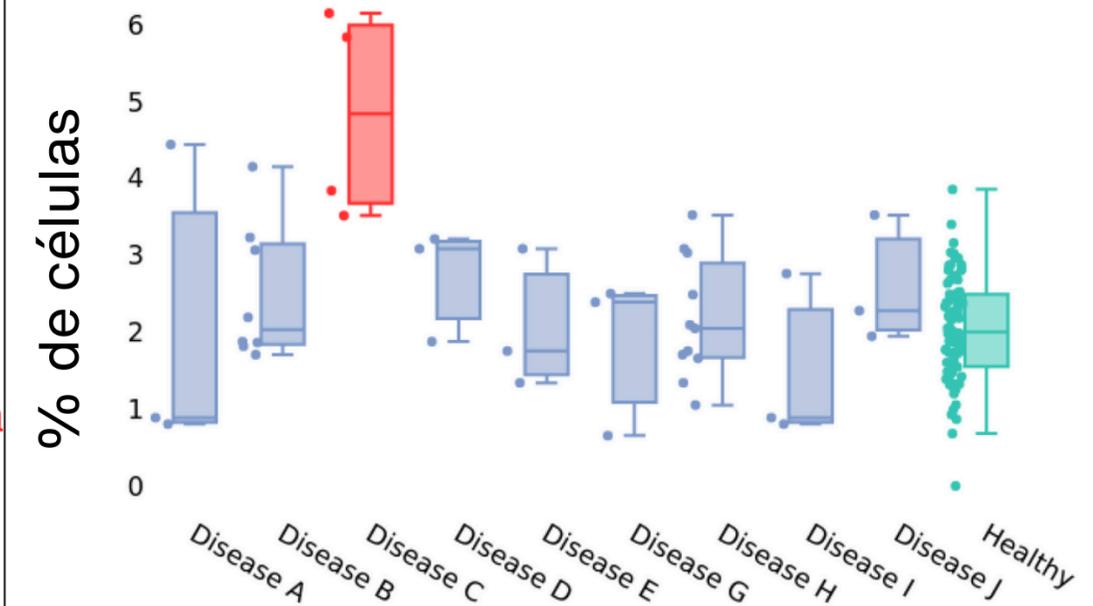
Tatyana Dobрева, David Brown, Jong Hwee Park & Matt Thomson

Scientific Reports volume 10, Article number: 20540 (2020)

DOI: 10.1038/s41598-020-77073-3

## Assinaturas de abundância de tipos de células específicas de doenças

### Abundância de células T de memória CD8+ por doença

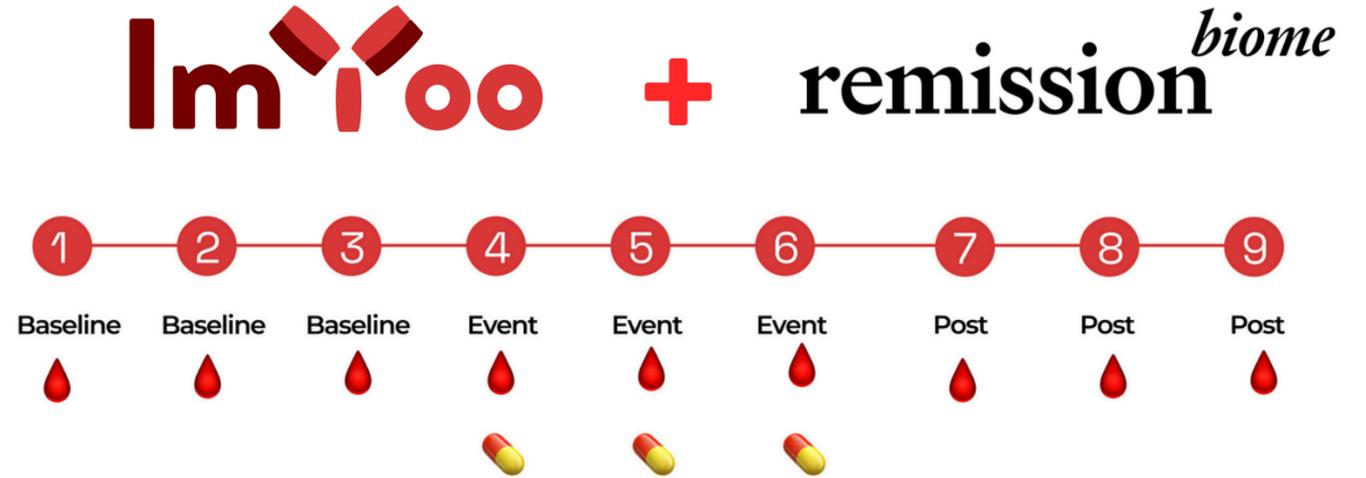


# N = 1: rumo à medicina personalizada

*A medicina personalizada exige abordagens centradas no paciente*

**Investigando a encefalomielite málgica/síndrome da fadiga crônica (EM/SFC)**

**Sequenciamento longitudinal de RNA de single cell de PBMCs de um paciente com EM/SFC antes, durante e depois do tratamento com antibióticos**



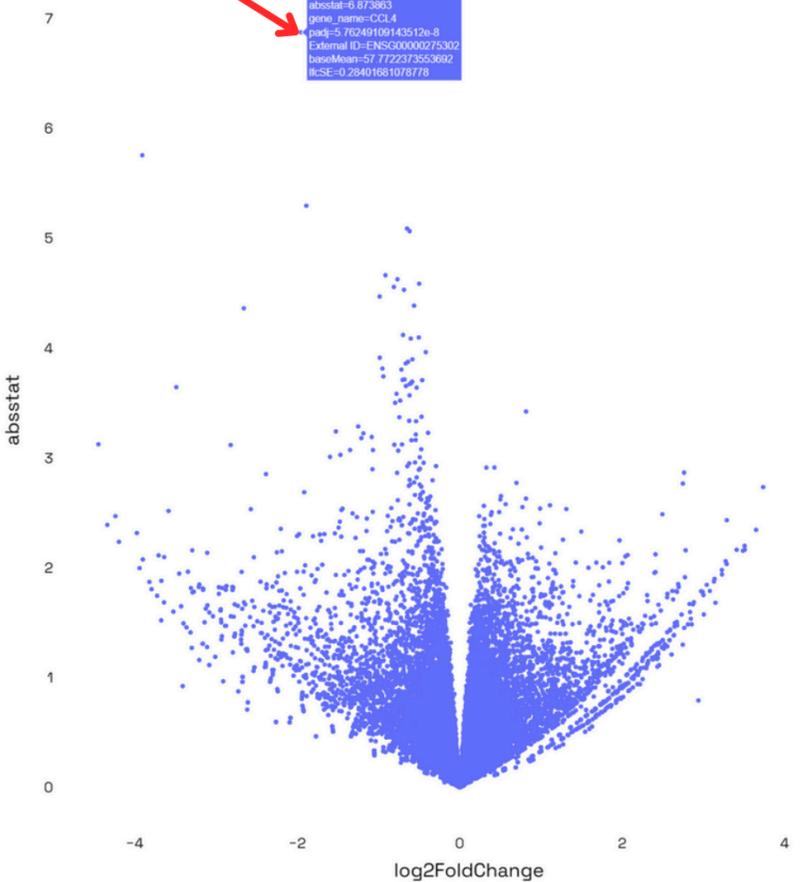
**Exemplo: O CCL4 surgiu como uma regulação negativa nos monócitos durante a remissão da EM/SFC**

*e continua a diminuir sua expressão durante meses após a intervenção.*

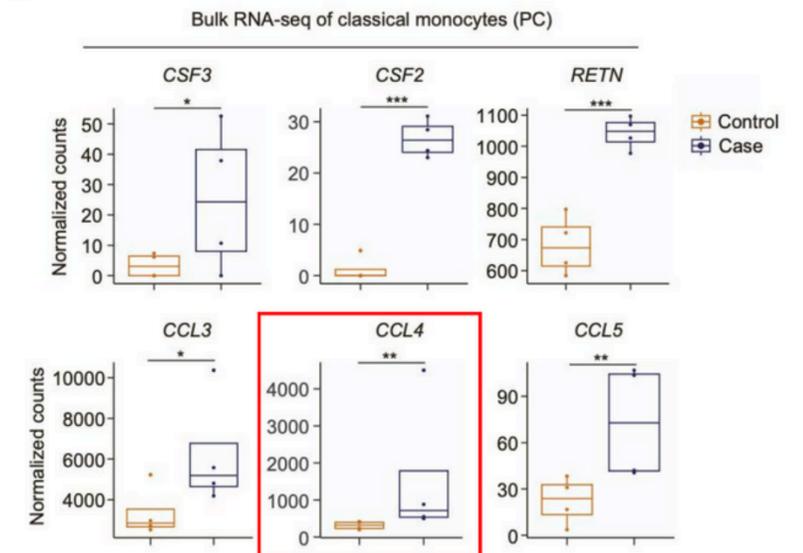
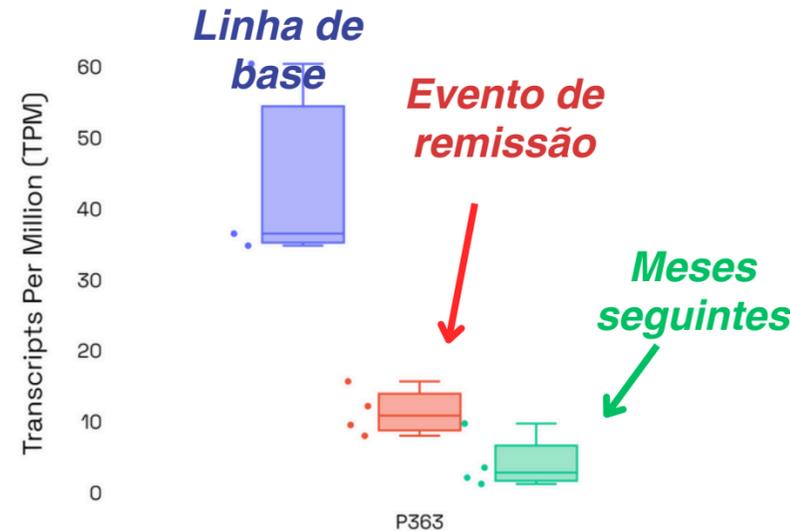
*Essa descoberta foi apoiada por um trabalho publicado recentemente usando RNA-seq em massa*

**CCL4**

log2FoldChange: -1.952293  
absstat: 0.873083  
gene\_name: CCL4  
padj: 5.76249109143512e-8  
External ID: ENSG00000275302  
baseMean: 57.7722373553692  
lfcSE: 0.28401681078778



CCL4 Expression in CD14 Monocytes



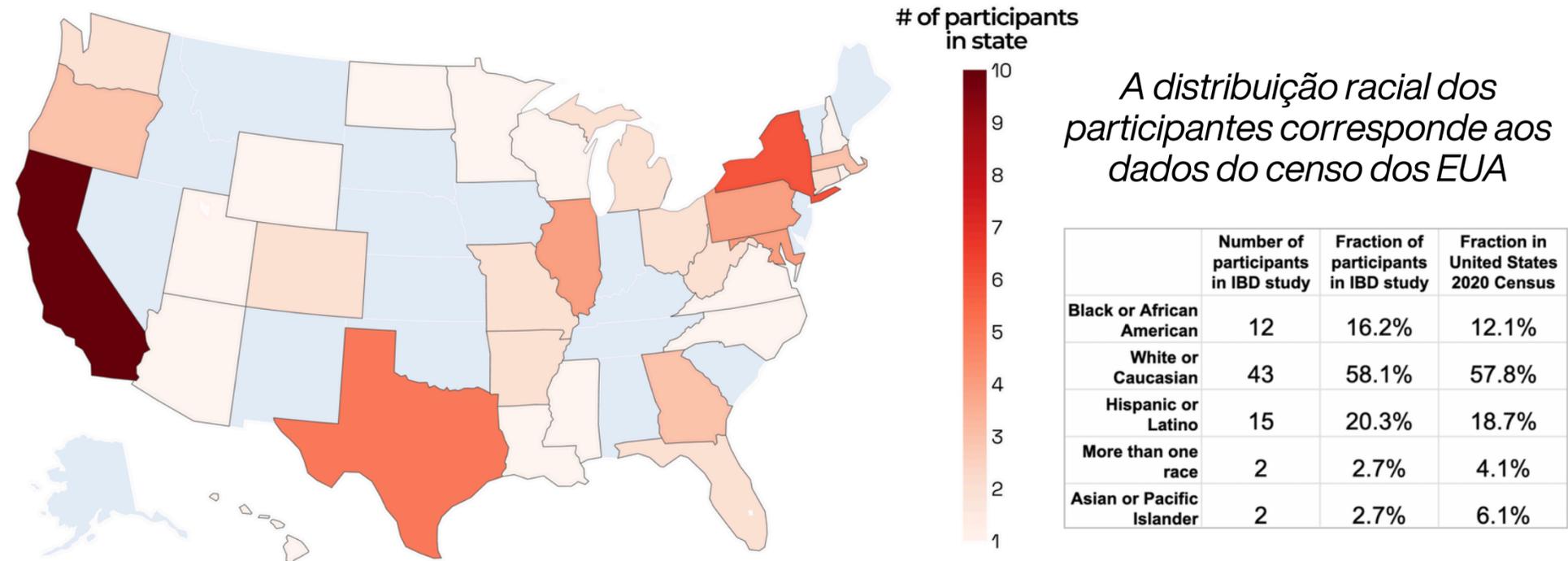
Single-cell transcriptomics of the immune system in ME/CFS at baseline and following symptom provocation.

Vu et al. Cell Reports Medicine 2024. DOI: 10.1016/j.xcrm.2023.101373

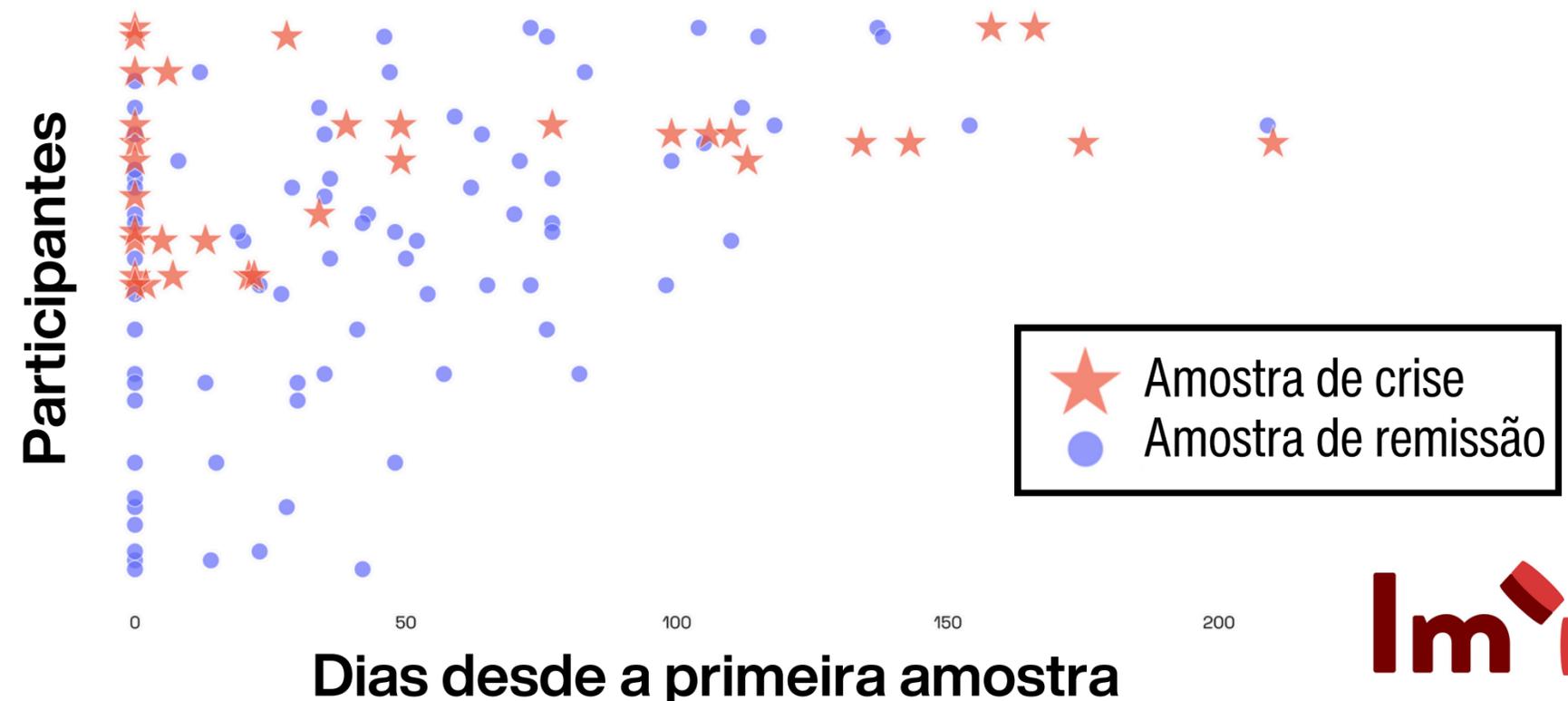
# Estudo descentralizado em andamento sobre IBD



Distribuição geográfica de 216 amostras enviadas pelo correio por 74 participantes

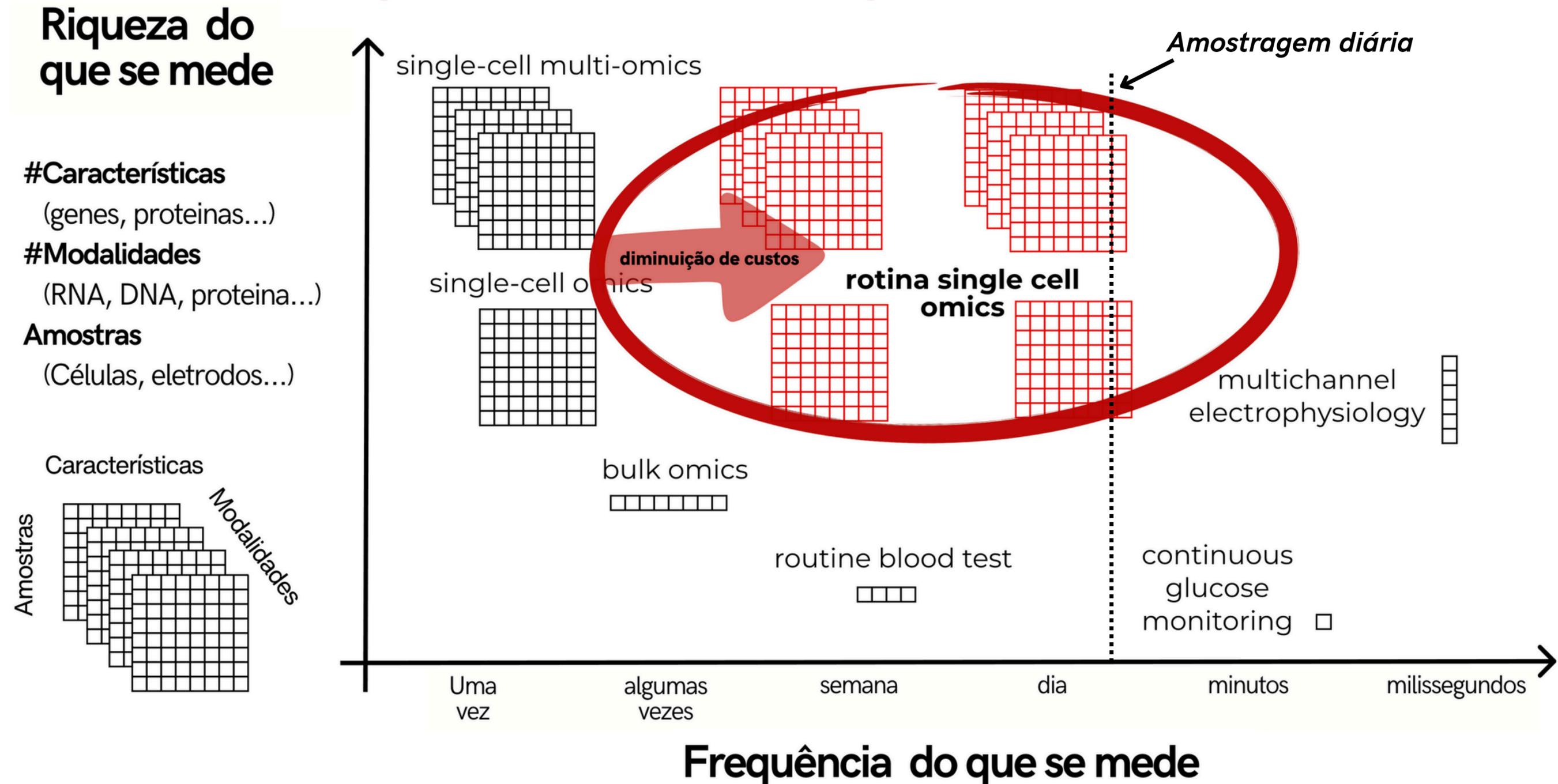


Distribuição temporal das amostras enviadas pelo correio, por participante

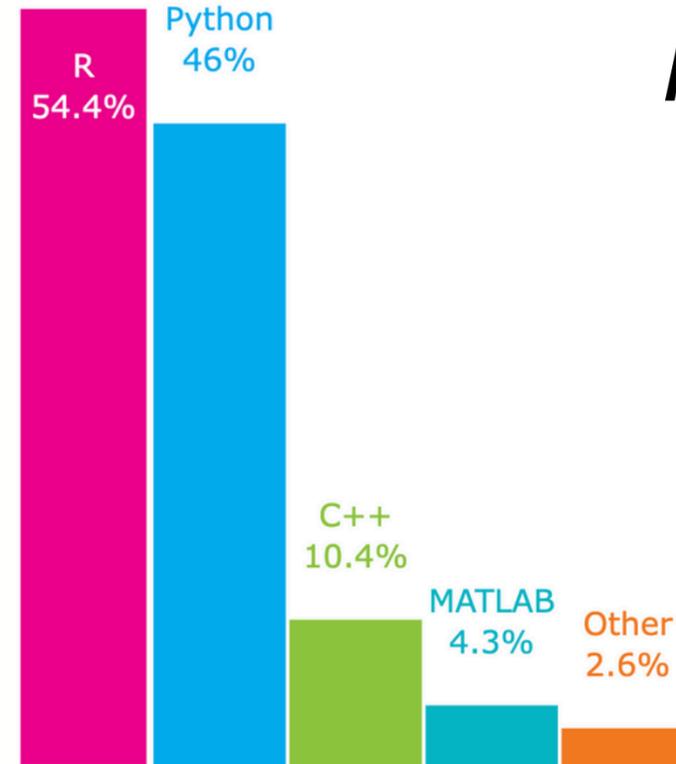


# Novas direções de pesquisa

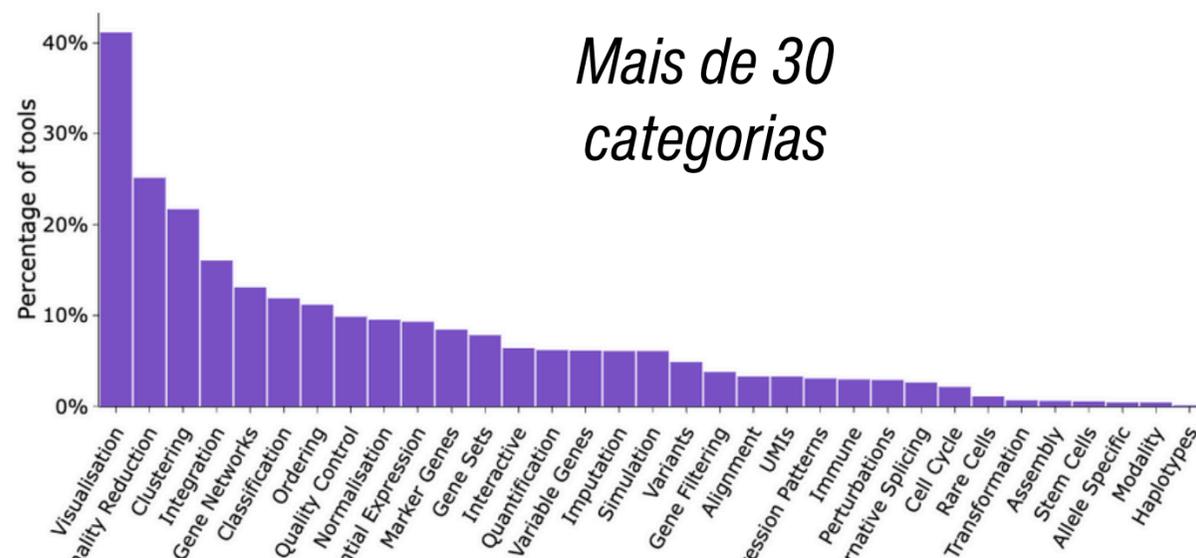
## Ferramentas e métodos para análise de dados longitudinais de single cell omics



# Explosão de ferramentas computacionais



<https://scrna-tools.org>



RESEARCH ARTICLE

## Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database

**Luke Zappia<sup>1,2</sup>, Belinda Phipson<sup>1</sup>, Alicia Oshlack<sup>1,2\*</sup>**

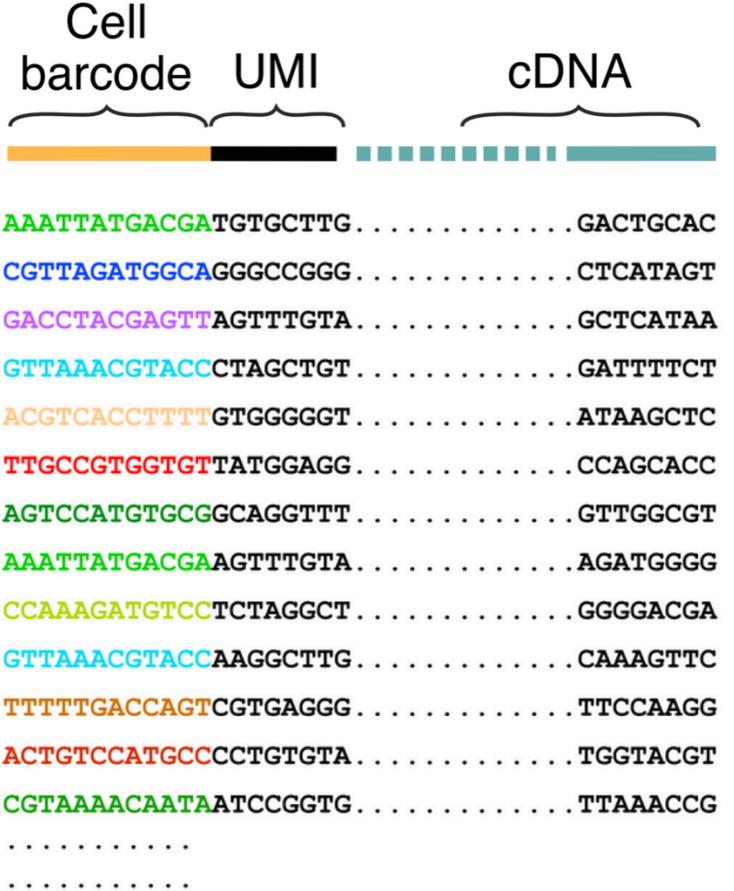
**1** Bioinformatics, Murdoch Children's Research Institute, Melbourne, Victoria, Australia, **2** School of Biosciences, Faculty of Science, University of Melbourne, Melbourne, Victoria, Australia

# **Processamento de dados de single cell RNA sequencing**

**Machine learning  
(aprendizagem de máquina)  
nos fornece métodos  
escaláveis para lidar com  
single cell omics**

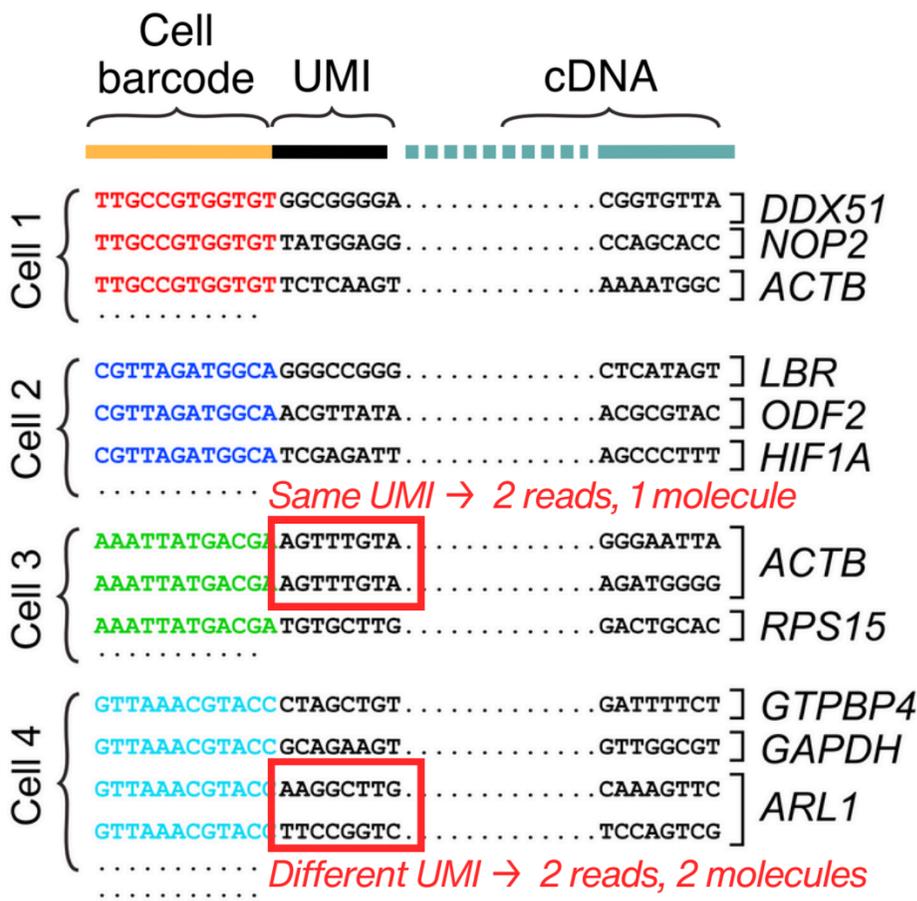
# Pré-processamento de dados de scRNAseq

## Arquivos FASTQ



Arquivos FASTQ com milhões de reads

## Alinhamento e contagem



Alinhar as reads válidas ao transcriptoma de referência

## Matriz de expressão gênica

	Cell: 1	2	...	N
GENE 1	1	2		14
GENE 2	4	27		8
GENE 3	0	0		1
⋮	⋮	⋮		⋮
GENE M	6	2		0

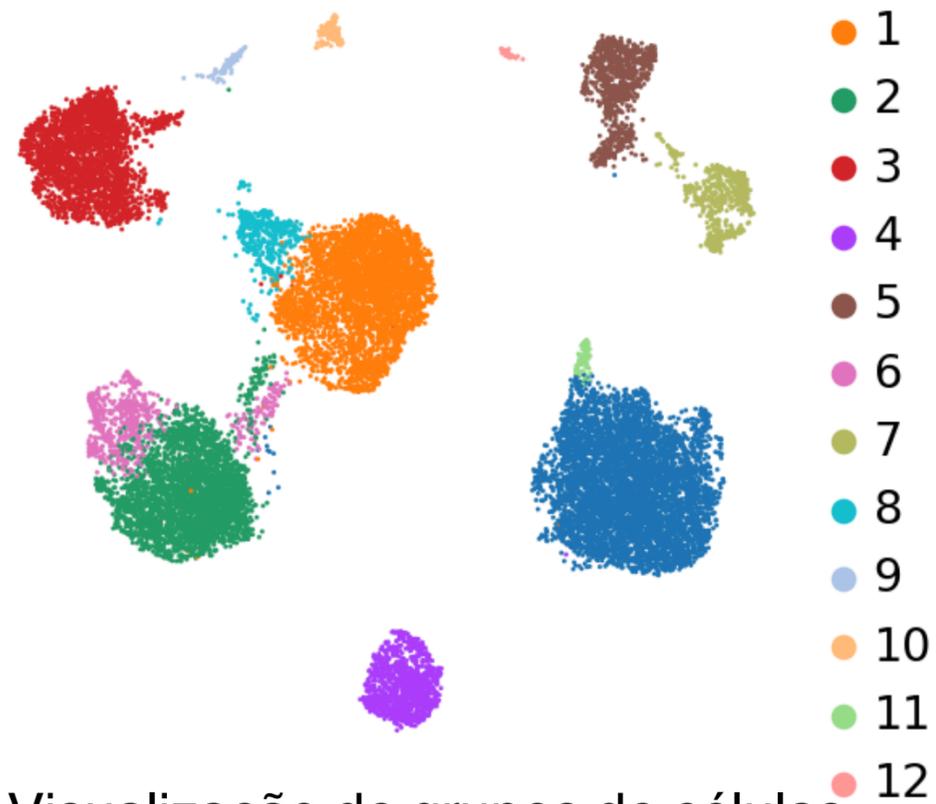
Contar o número de cópias de cada gene em cada célula para gerar a matriz de expressão gênica.

Figura adaptada de: EZ Macosko et al., Cell, 2015. DOI: 10.1016/j.cell.2015.05.002

# Tarefas de processamento de dados de scRNAseq

# Tarefas de processamento de dados de scRNAseq

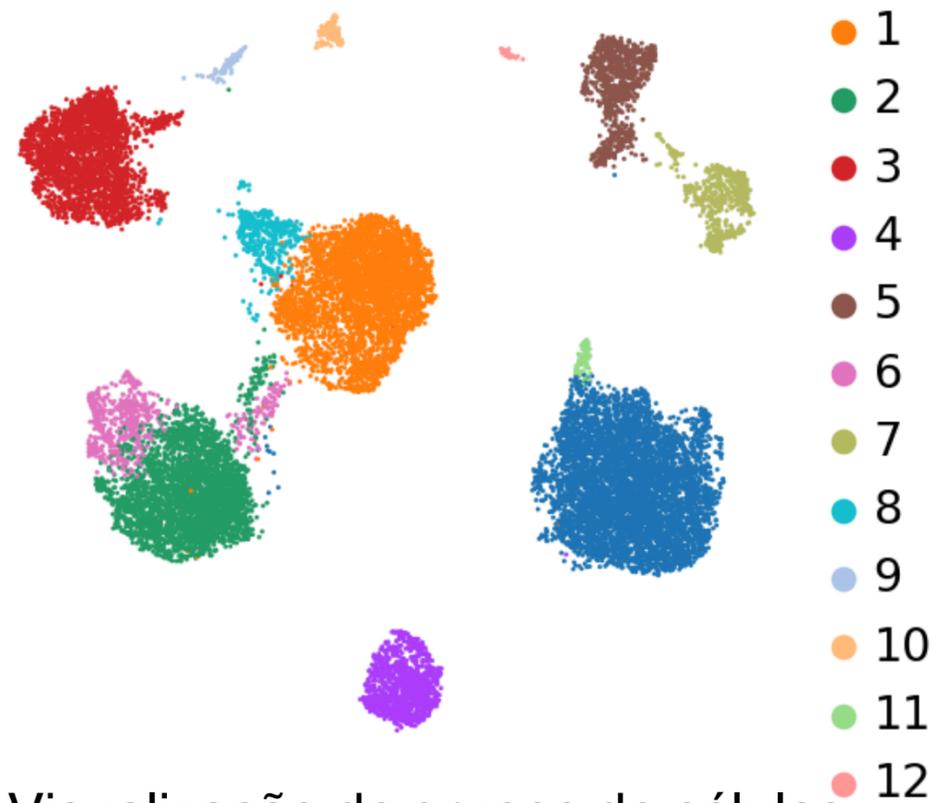
## *Agrupamento e identificação do tipo de célula*



Visualização de grupos de células

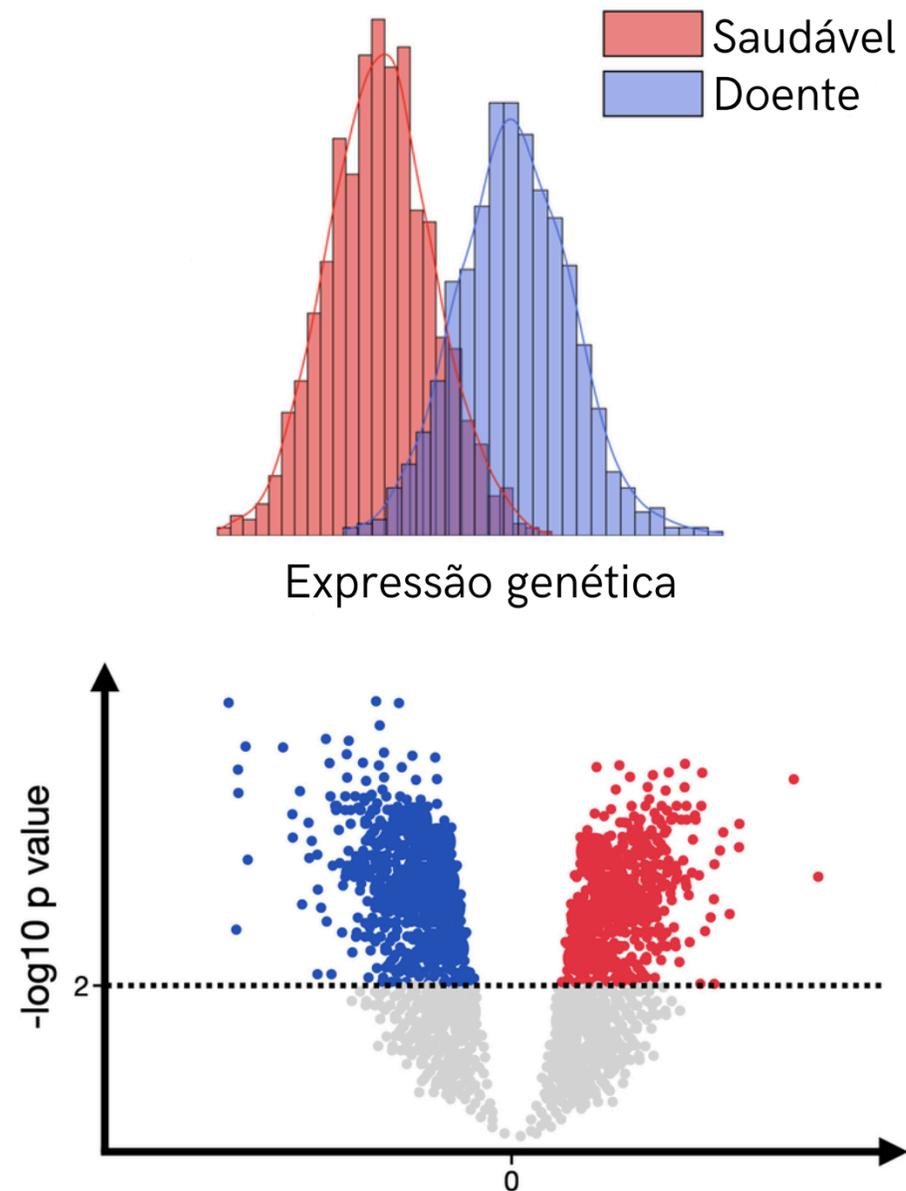
# Tarefas de processamento de dados de scRNAseq

## *Agrupamento e identificação do tipo de célula*



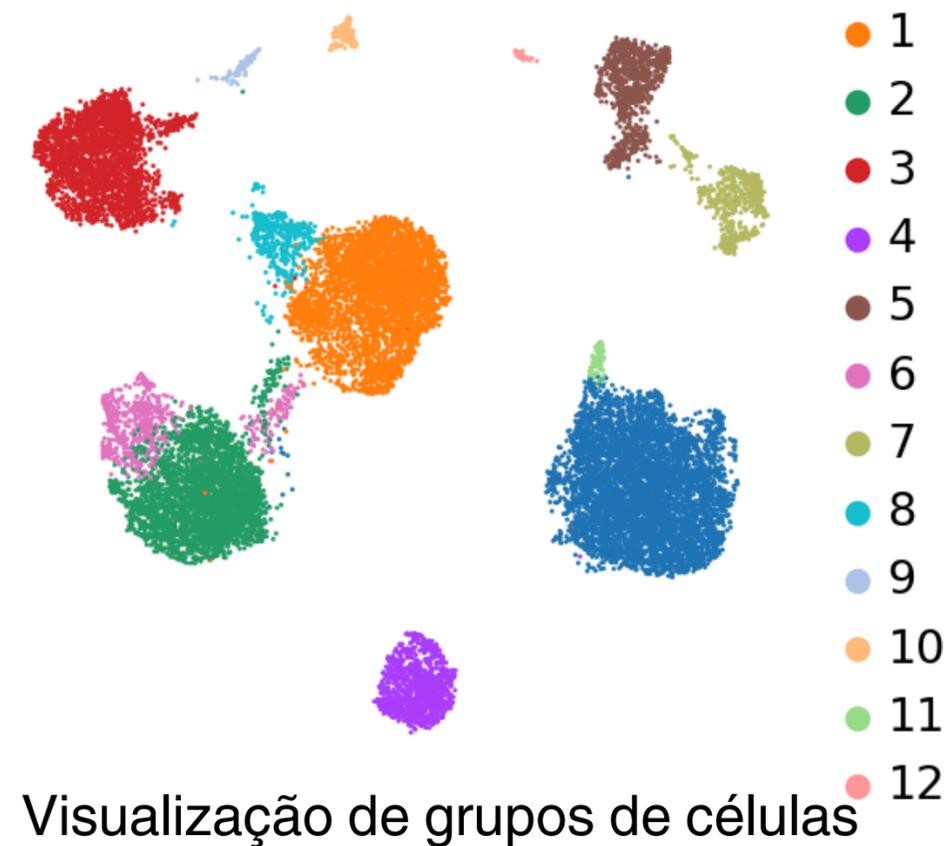
Visualização de grupos de células

## *Expressão gênica diferencial*

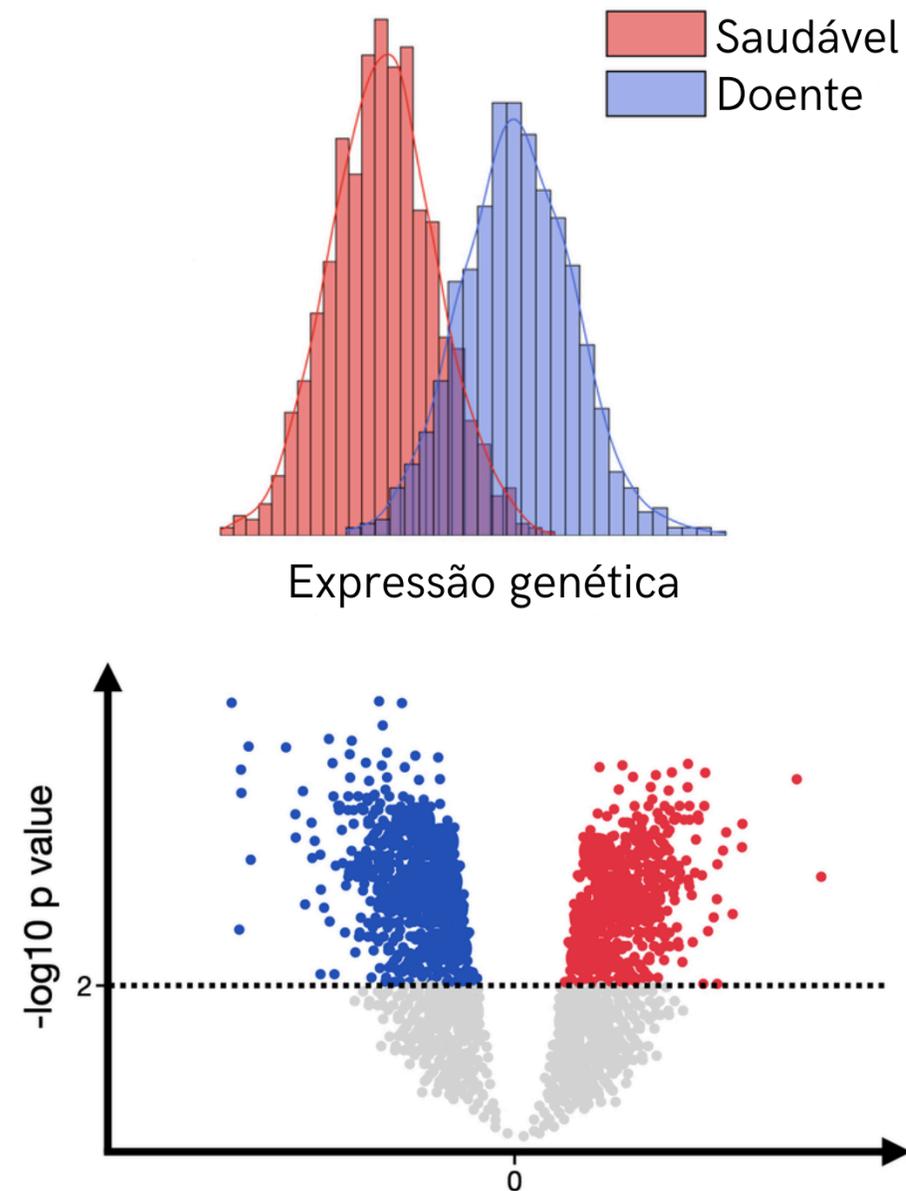


# Tarefas de processamento de dados de scRNAseq

## Agrupamento e identificação do tipo de célula



## Expressão gênica diferencial



## Integração de datasets

Sem integração



Integração excessiva



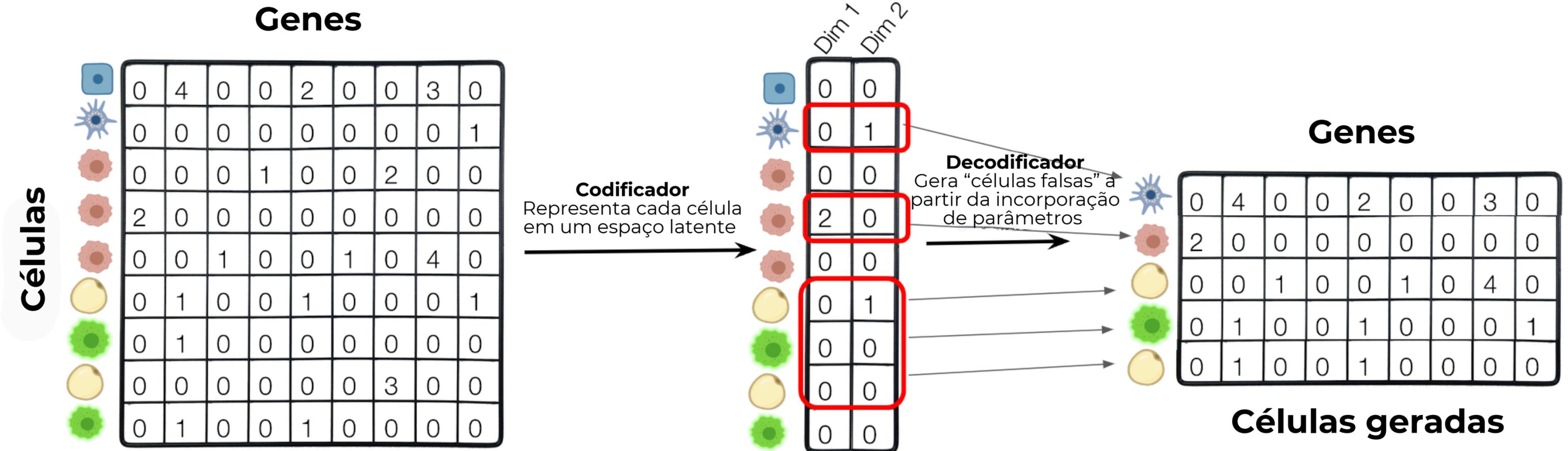
Integração adequada



Dataset

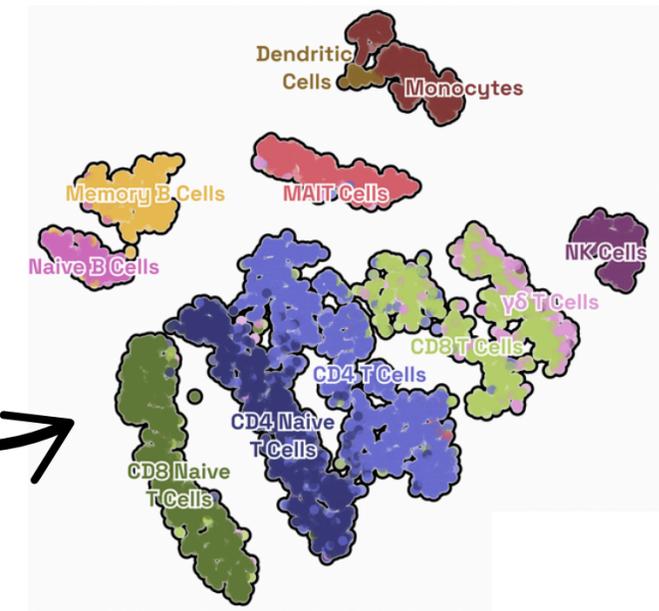
Tipos de células

# Modelos generativos de single cell



**Espaço latente**

**clustering,  
visualização...**



# Variational Autoencoders

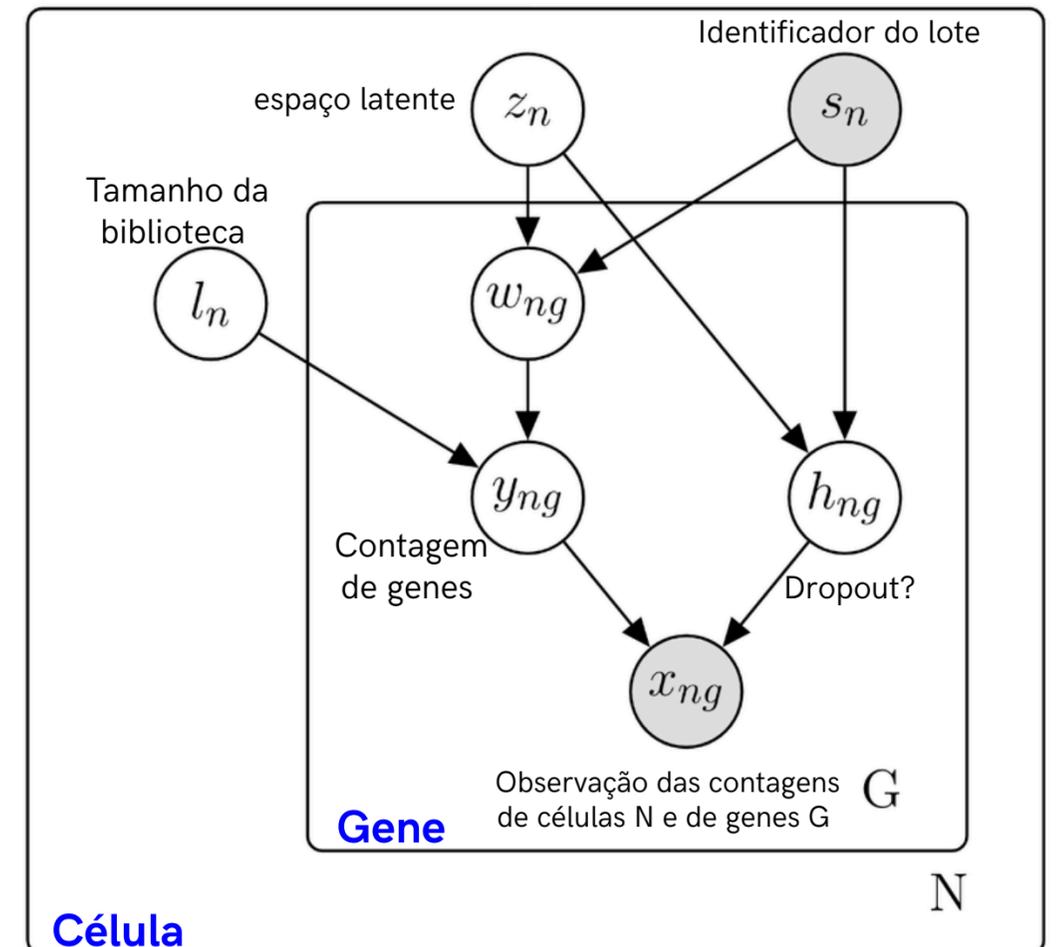
*Emergiram como uma abordagem escalável para o processamento de dados em single cell omics*

Article | Published: 30 November 2018

## Deep generative modeling for single-cell transcriptomics

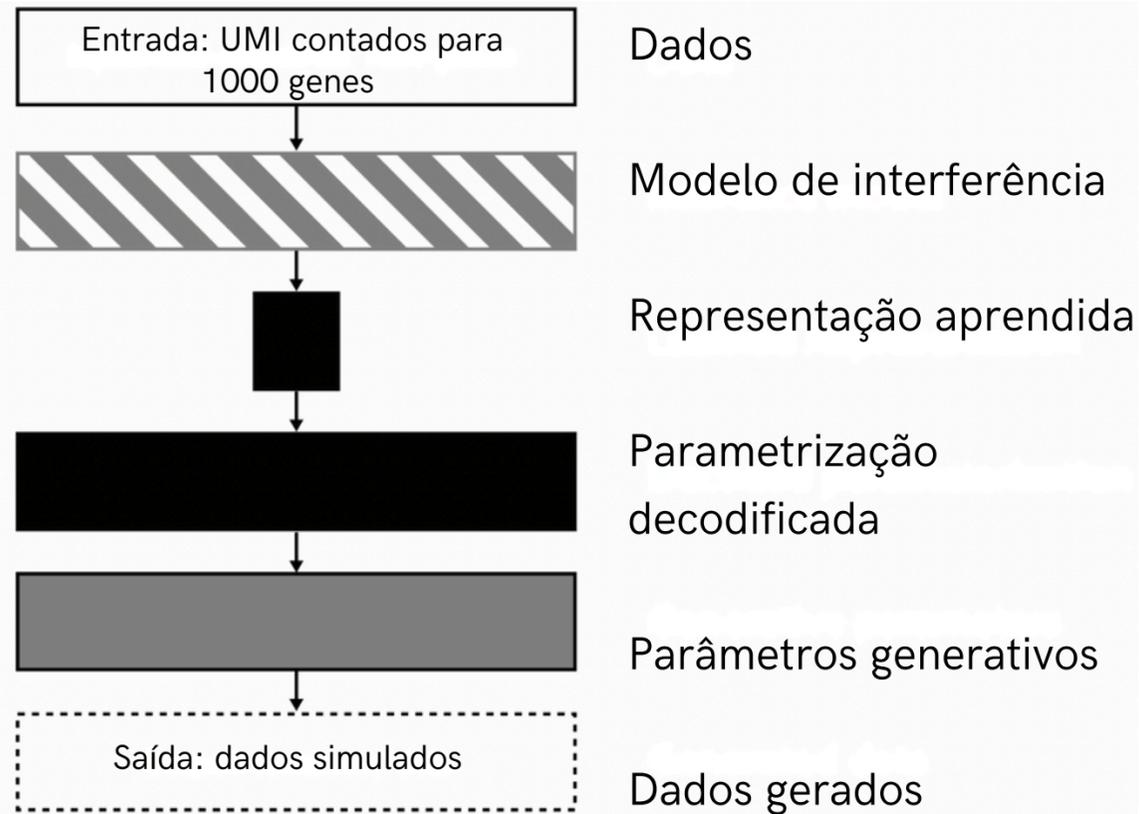
[Romain Lopez](#), [Jeffrey Regier](#), [Michael B. Cole](#), [Michael I. Jordan](#) & [Nir Yosef](#) 

[Nature Methods](#) **15**, 1053–1058 (2018) | [Cite this article](#)

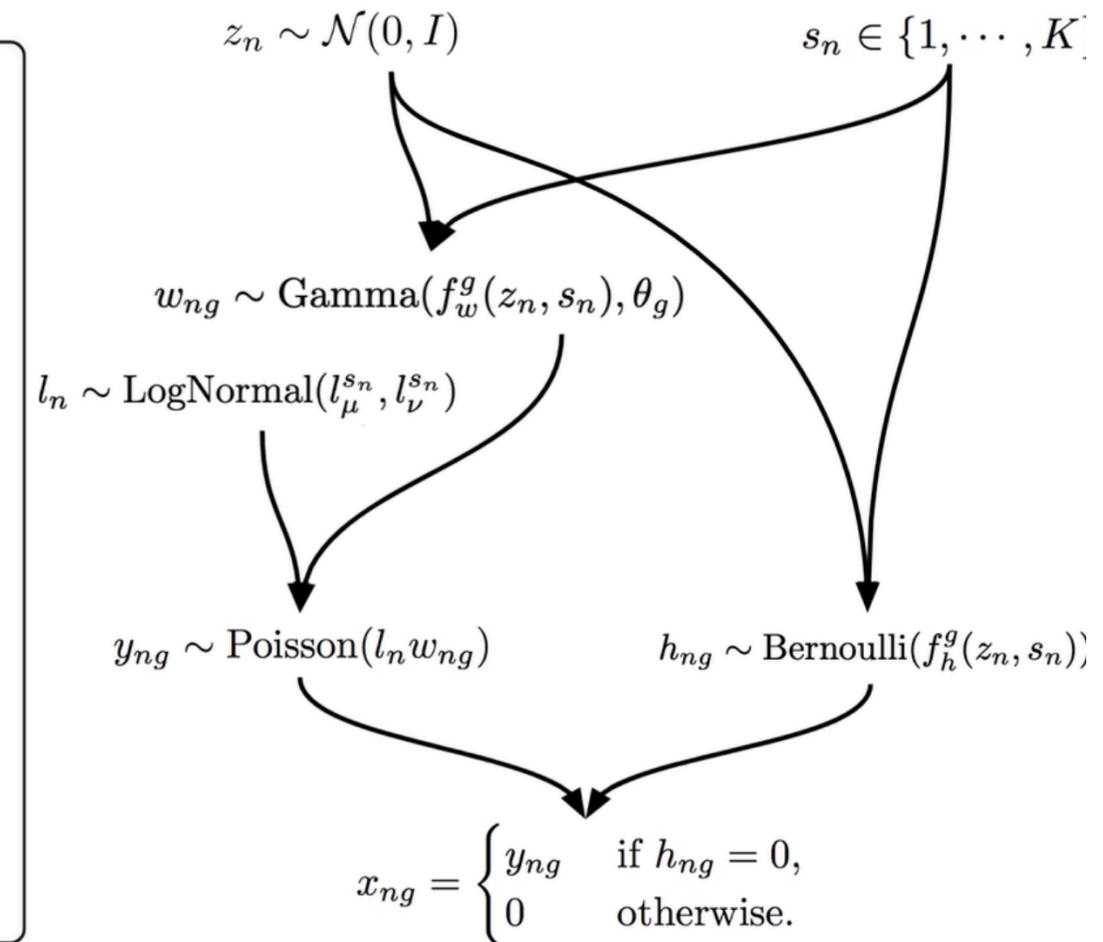
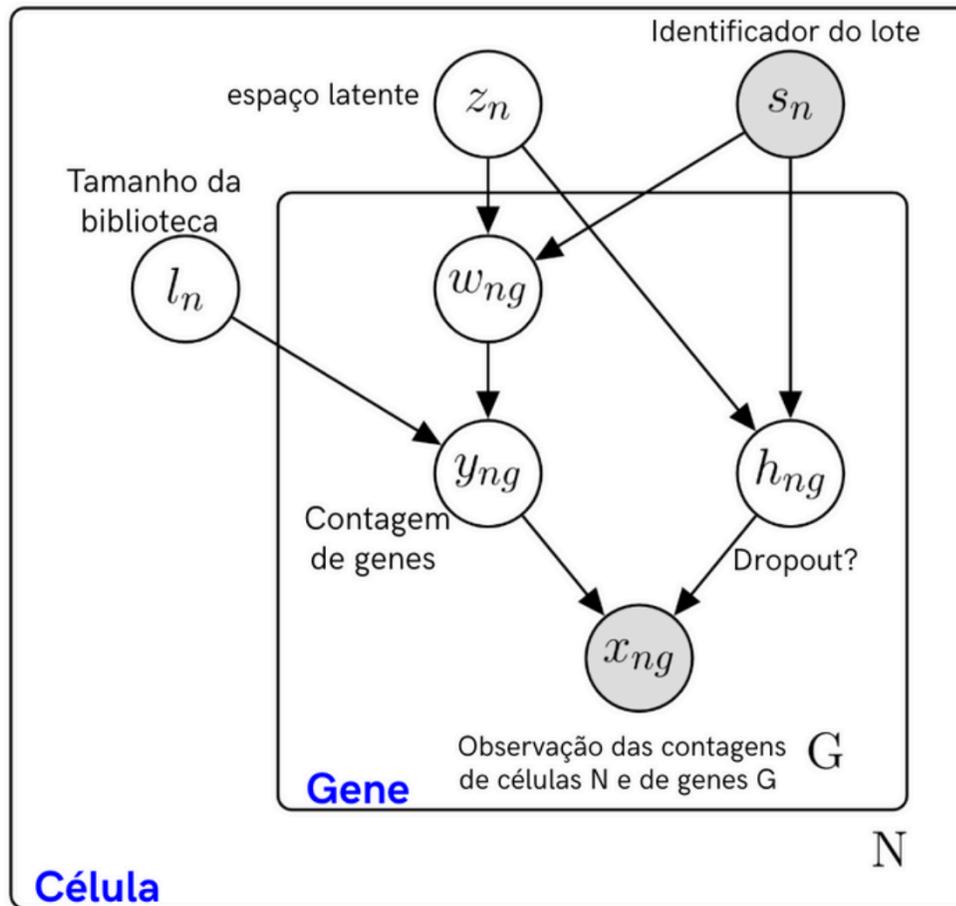


# Modelagem generativa de single cell omics

## scVI: Um VAE para single cell

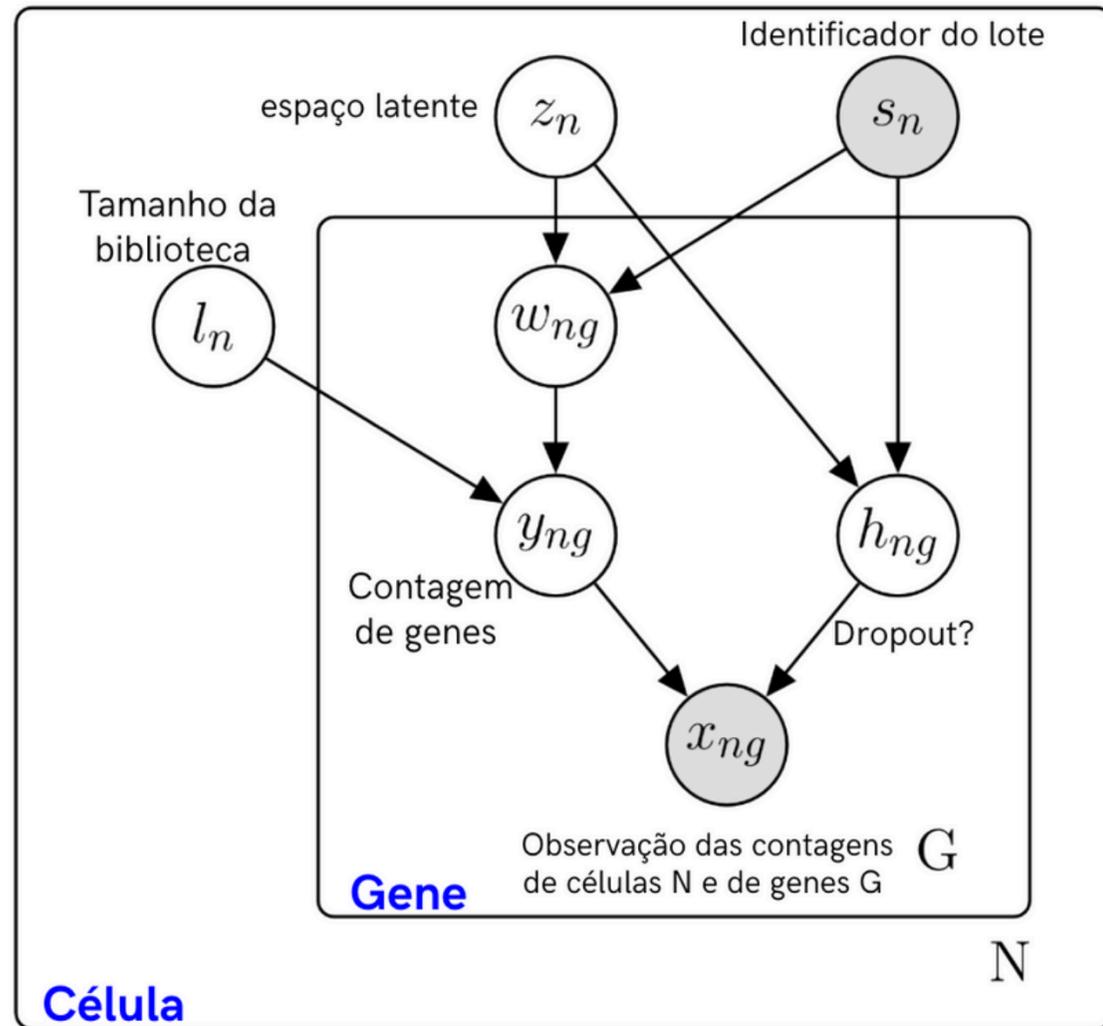


## Modelo gráfico do scVI

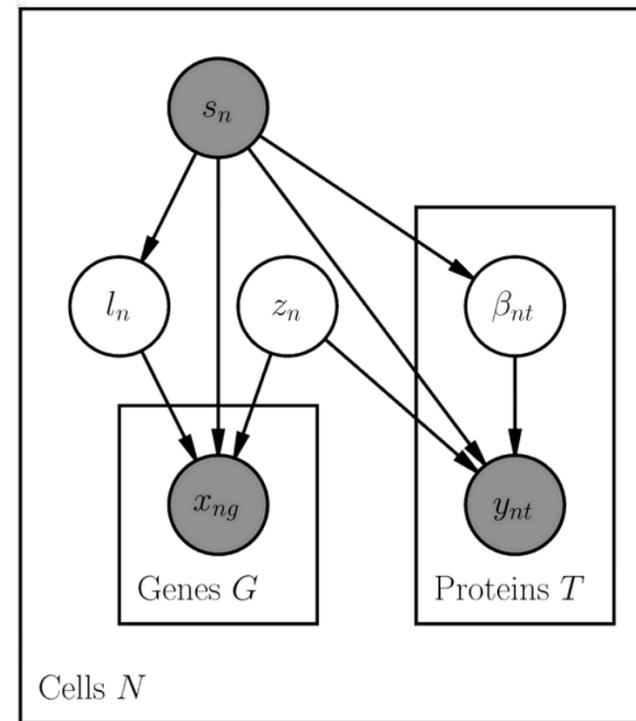


# Modelos gráficos de VAEs para single cell

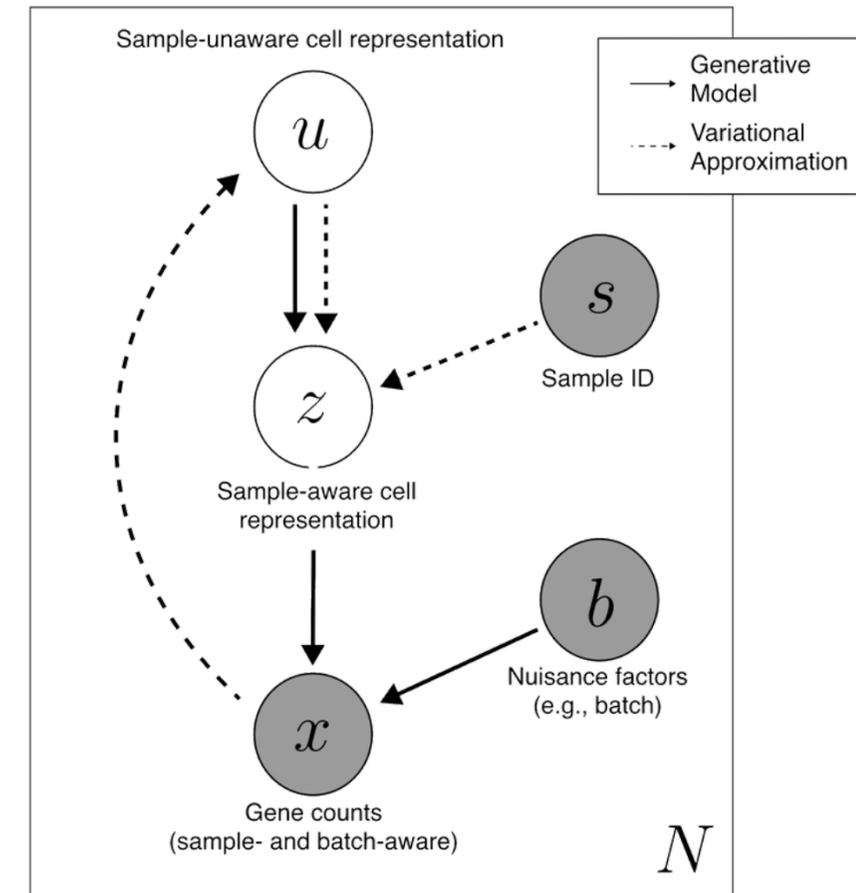
**scVI**  
RNA



**totalVI**  
RNA + Proteína  
(CITE-seq)



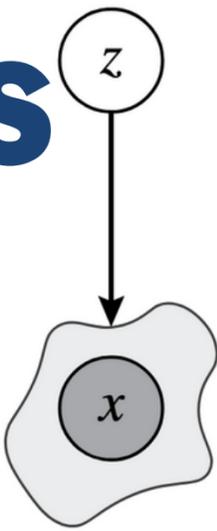
**MrVI**  
RNA



# Análise em Python com scvi-tools

## Modelos probabilísticos para single-cell omics

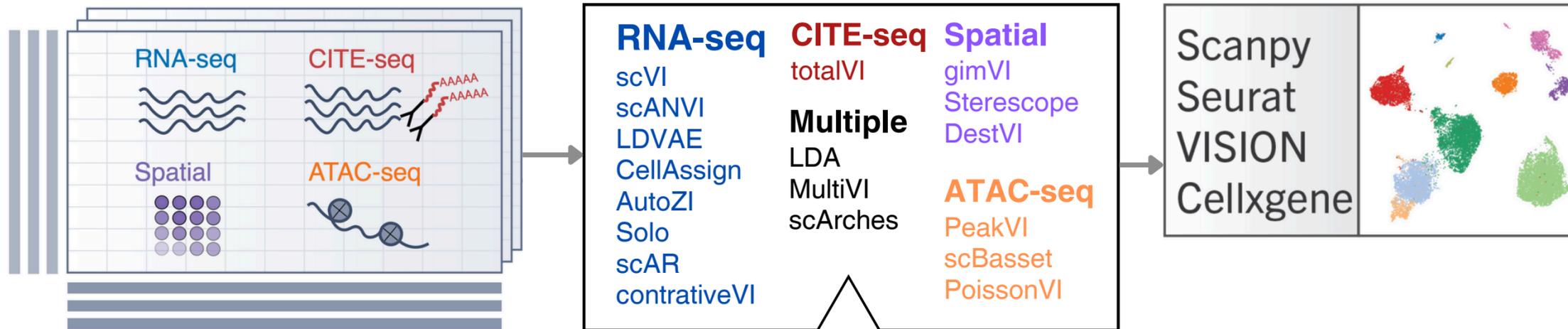
<https://scvi-tools.org>



Dados de Single cell omics

modelos modais

Visualização



nature biotechnology

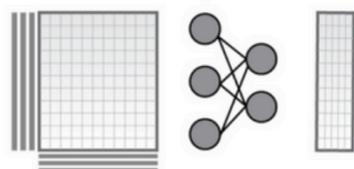
DOI: 10.1038/s41587-021-01206-w

Correspondence | Published: 07 February 2022

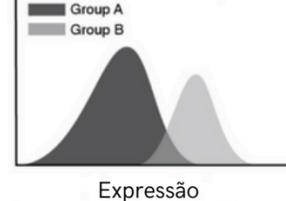
**A Python library for probabilistic analysis of single-cell omics data**

[Adam Gayoso](#), [Romain Lopez](#), [Galen Xing](#), [Pierre Boyeau](#), [Valeh Valiollah Pour Amiri](#), [Justin Hong](#), [Katherine Wu](#), [Michael Jayasuriya](#), [Edouard Mehlman](#), [Maxime Langevin](#), [Yining Liu](#), [Jules Samaran](#), [Gabriel Misrachi](#), [Achille Nazaret](#), [Oscar Clivio](#), [Chenling Xu](#), [Tal Ashuach](#), [Mariano Gabitto](#), [Mohammad Lotfollahi](#), [Valentine Svensson](#), [Eduardo da Veiga Beltrame](#), [Vitalii Kleshchevnikov](#), [Carlos Talavera-López](#), [Lior Pachter](#), [Fabian J. Theis](#), [Aaron Streets](#), [Michael I. Jordan](#), [Jeffrey Regier](#) & [Nir Yosef](#) ✉

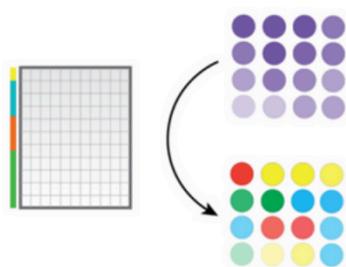
Redução da dimensionalidade



Compreensão diferencial

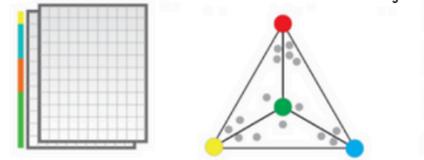


Deconvolução

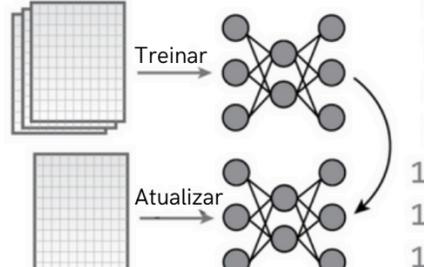


Anotação automatizada

Probabilidade de atribuição



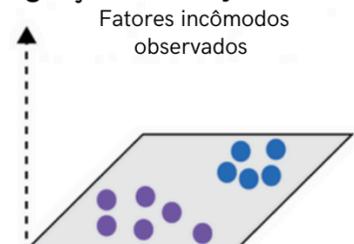
Transferir aprendizagem



Experiência de usuário consistente

```
1 import scvi
2 adata = scvi.data.read_h5ad("pbmc.h5ad")
3 scvi.data.setup_anndata(adata)
4 model = scvi.model.SCVI(adata)
5 model.train()
6 latent = model.get_latent_representation()
7 de_df = model.differential_expression(
8     groupby="cell_type",
9     group1="CD4",
10    group2="CD8",
11 )
12 model.save("save_dir")
```

Remoção de variação indesejada  
(integração de conjunto de dados)



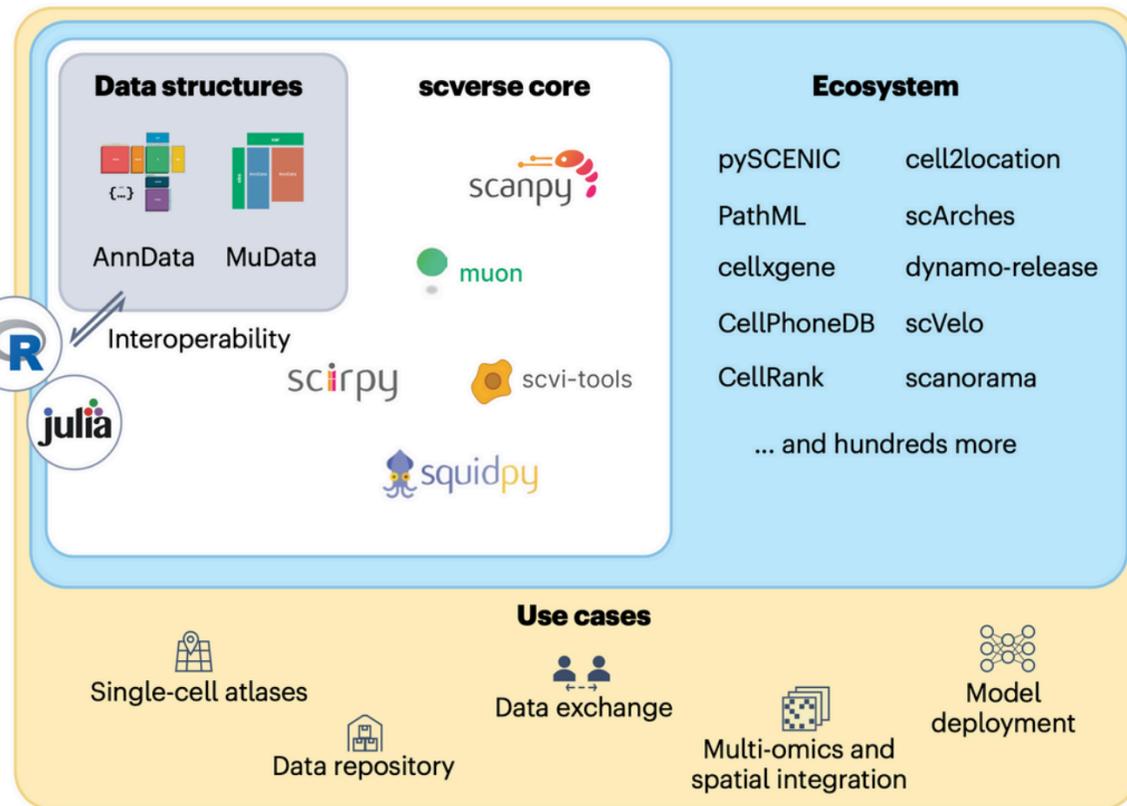
**Adam Gayoso**  
UC Berkeley,  
now Google Deepmind



**Romain Lopez**  
UC Berkeley,  
now Stanford/Genentech

# Desenvolvimento comunitário

## scverse



### Participation

Hackathons



Tutorials and documentation



Chat and forum



### Joint development

GitHub



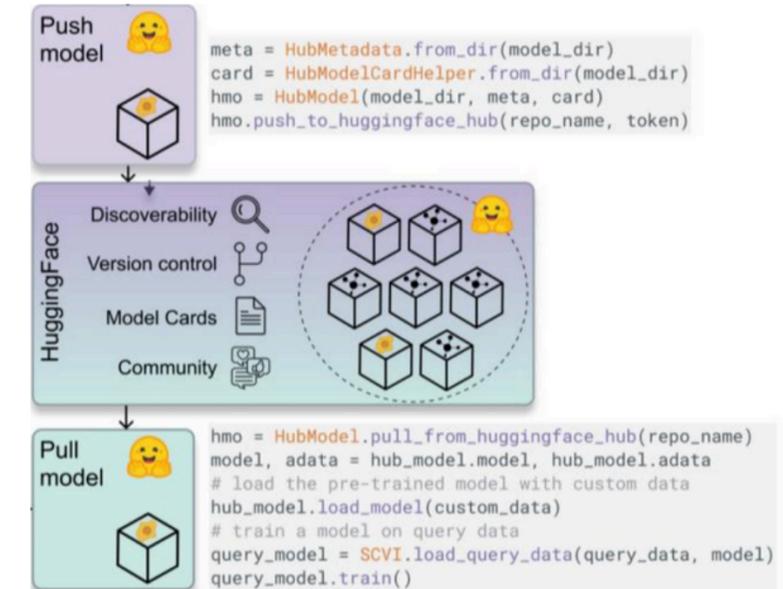
Extensible APIs, abstractions, best practices



Automated tests and continuous integration across projects



## scvi-hub



**Scvi-hub: an actionable repository for model-driven single cell analysis**

Can Ergen, Valeh Valiollah Pour Amiri, Martin Kim, Aaron Streets, Adam Gayoso, Nir Yosef  
bioRxiv 2024 doi: 10.1101/2024.03.01.582887

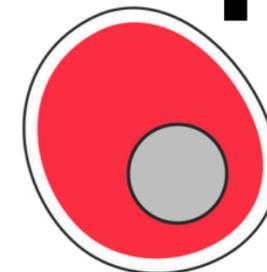
**The scverse project provides a computational ecosystem for single-cell omics data analysis**

Isaac Virshup, Danila Bredikhin, Lukas Heumos, Giovanni Palla, Gregor Sturm, Adam Gayoso, Iliia Kats, Mikaela Koutrouli, Scverse Community, Bonnie Berger, Dana Pe'er, Aviv Regev, Sarah A. Teichmann, Francesca Finotello, F. Alexander Wolf, Nir Yosef, Oliver Stegle & Fabian J. Theis  
Nature Biotechnology 2023. doi: 10.1038/s41587-023-01733-8

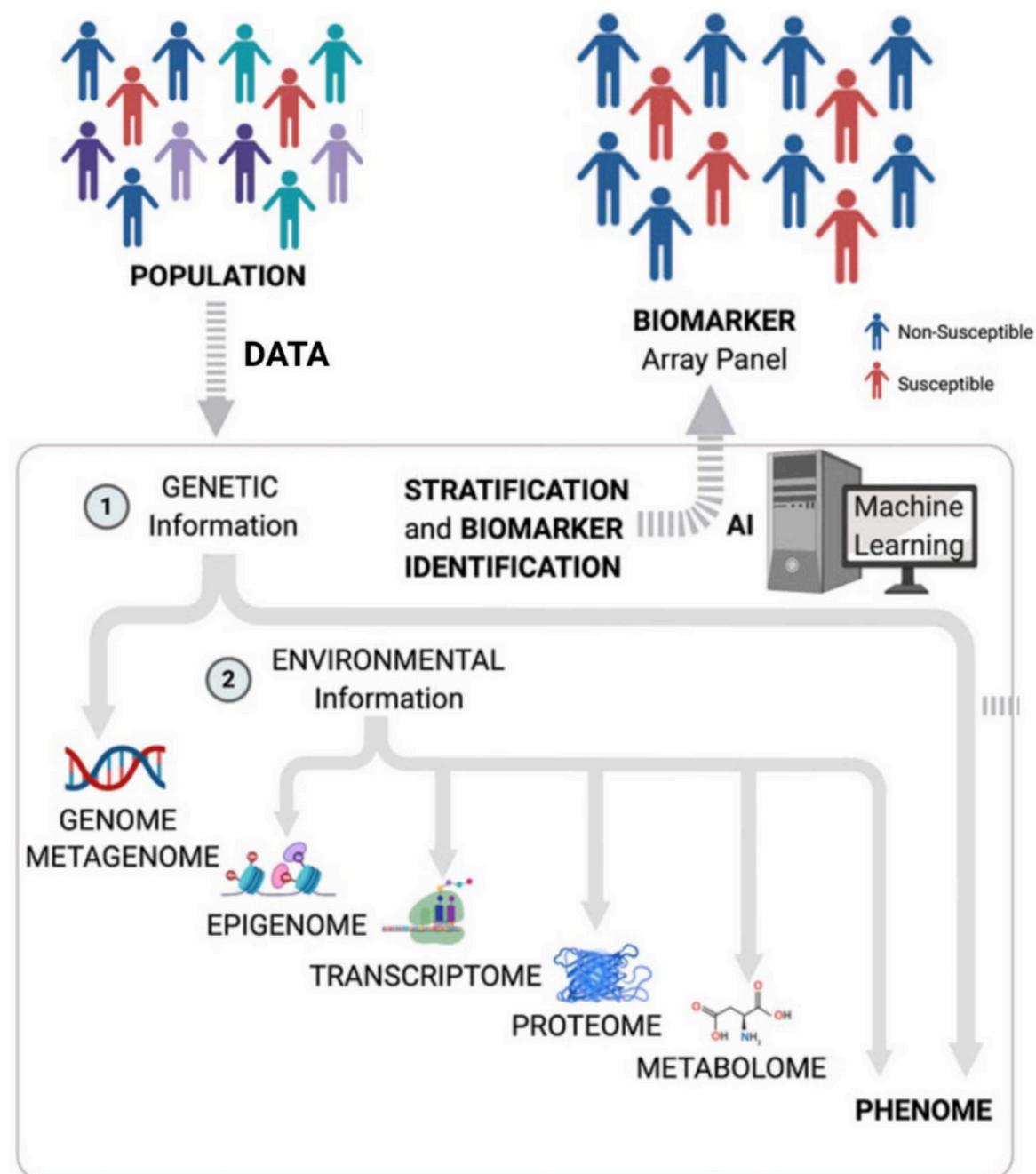
## Open Problems

**Open Problems** in  
**Single-Cell Analysis**

<https://openproblems.bio/>



# O futuro próximo: Estratificação de pacientes em alta resolução, melhores diagnósticos moleculares



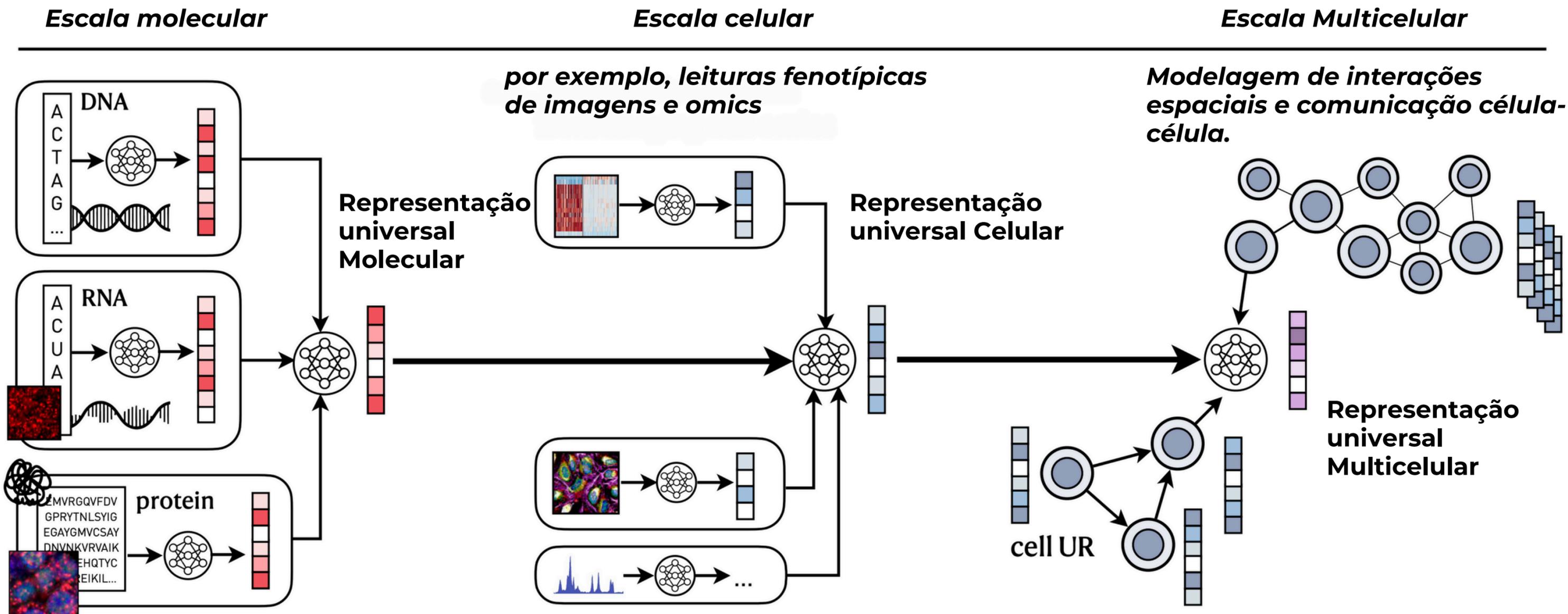
*Modelos para dados multimodais*

*Diagnóstico molecular de alta resolução*

*Identificação de intervenções que funcionam*

- Estratificação de pacientes
- Correspondência de tratamento
- Tratamento personalizado

# Integração de modelos de base biológica em uma “célula virtual de IA”



# AIDO: AI-Digital Organism

<https://genbio.ai>

Toward AI-Driven Digital Organism:  
A System of Multiscale Foundation Models for Predicting,  
Simulating and Programming Biology at All Levels

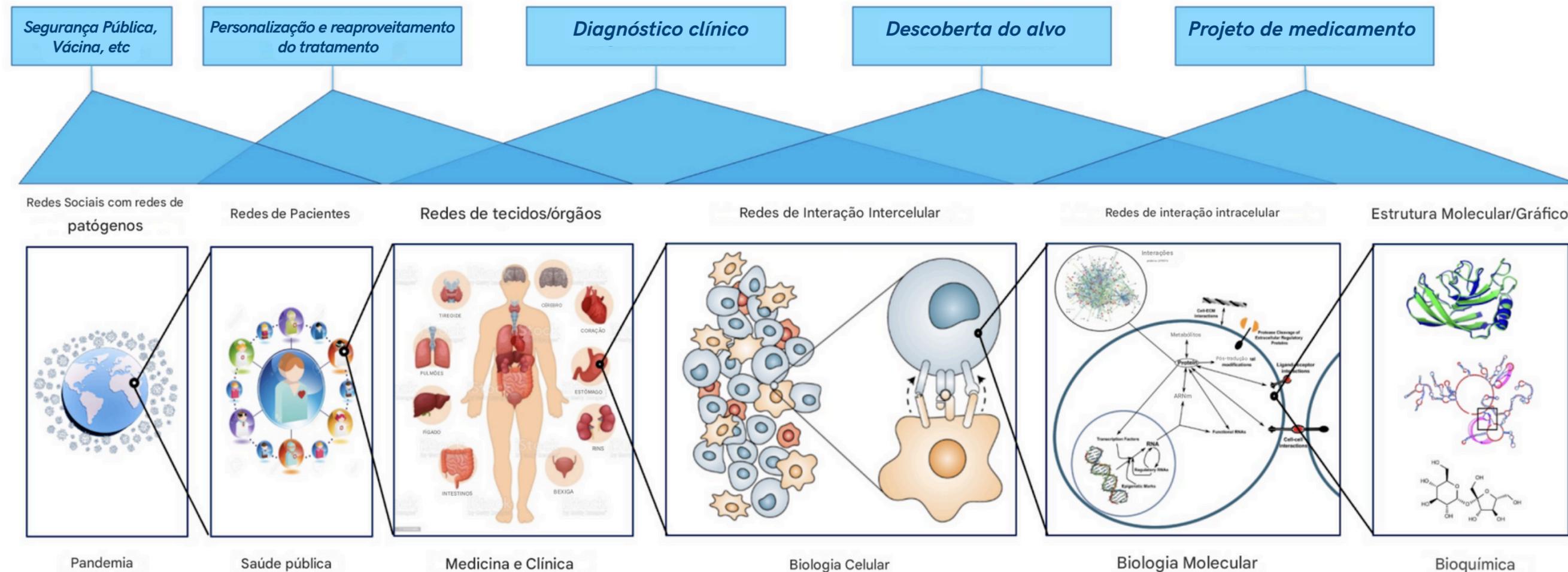
Le Song<sup>\*,◇</sup>, Eran Segal<sup>\*,◇,†</sup>, Eric Xing<sup>\*,◇,‡</sup>

<sup>\*</sup>GenBio AI

<sup>◇</sup>Mohamed bin Zayed University of Artificial Intelligence

<sup>†</sup>Weizmann Institute of Science

<sup>‡</sup>Carnegie Mellon University



# AIDO: AI-Digital Organism - <https://genbio.ai>



Um conjunto amplo e representativo de tarefas de previsão e geração de benchmark em todas as escalas

Inferência

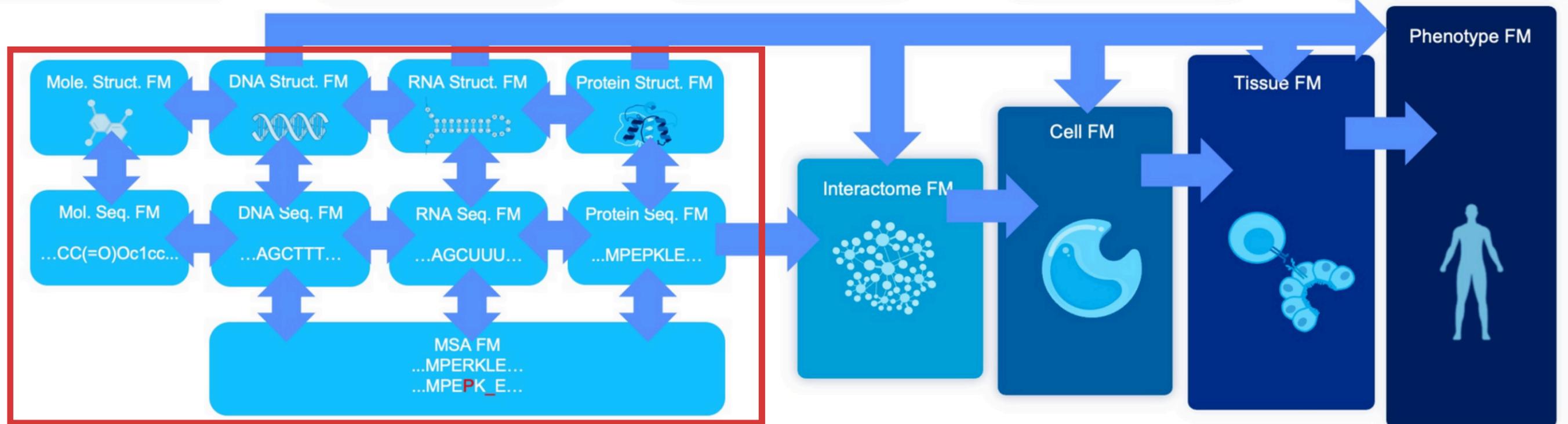
Afinar

Gerar

Projeto

Explicar

Modalidades  
lançadas  
pelo gen.ai



Integrated Billion-Scale Public & Proprietary Databases

Genebank, Refseq, IMG, Uniprot, PDB, AFDB, StringDB, KEGG, Reactome, GTRD, ENCODE, TCGA, Human Protein Atlas, CellxGene, Human Cell Atlas, Human Phenotype Project

# O Projeto Fenótipo Humano (Human Phenotype Project)

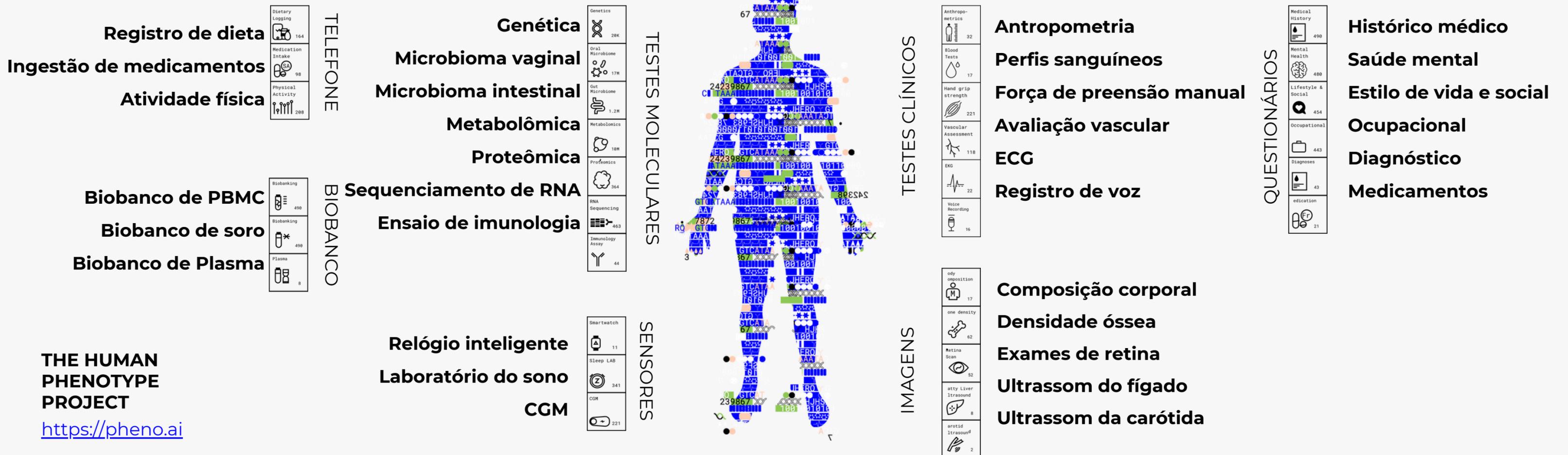
## 12.000 pessoas com perfil

- Em Israel
- Perfil longitudinal a cada 2 anos
- 17 sistemas corporais diferentes
- 40 doenças (7.700 pacientes no total)
- 30 condições médicas

Agora expandindo para

## 100.000 pessoas

- 4 países adicionais
- Mais doenças
- Mais medidas (ômicas de célula única!)
- Coortes nos Emirados Árabes Unidos e no Japão
- Iniciativa importante na MBZUAI



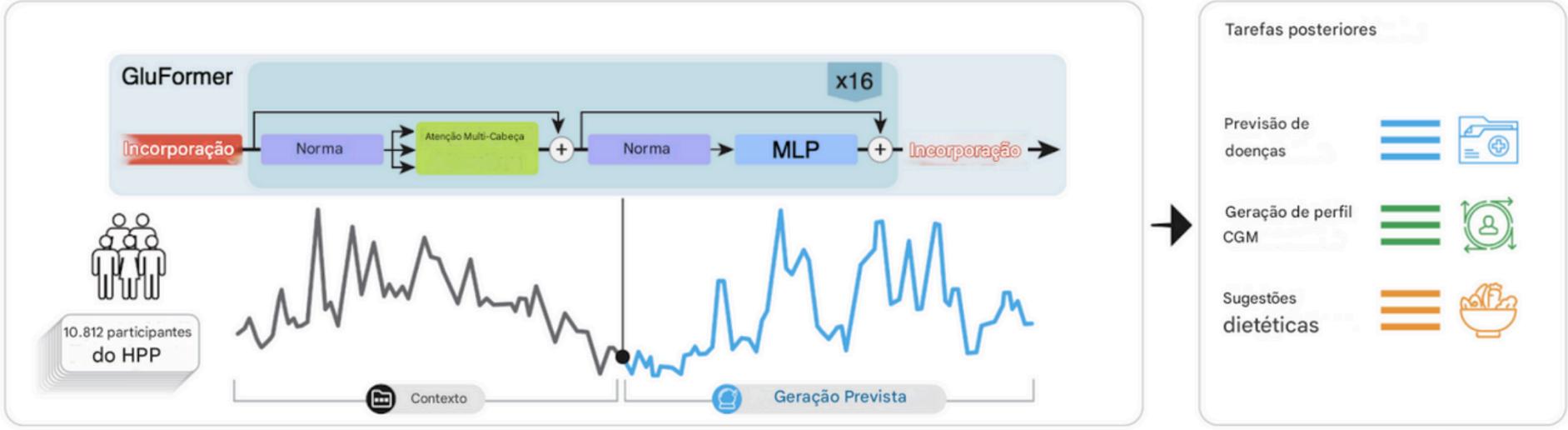
# Desenvolvido com os dados do HPP:

## Dietas

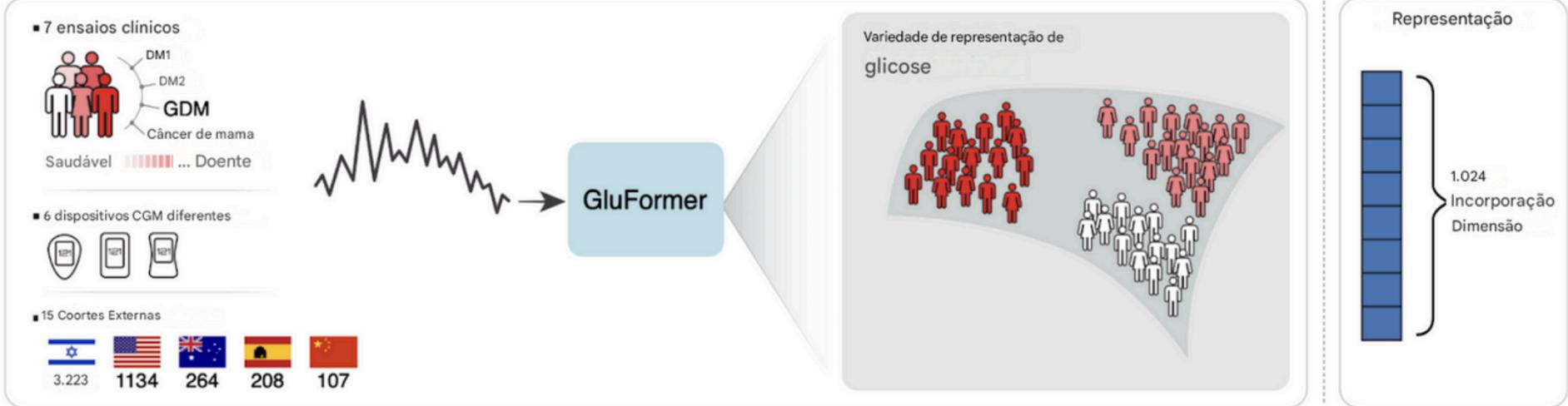
# personalizadas com um modelo de base para CGM

**From Glucose Patterns to Health Outcomes: A Generalizable Foundation Model for Continuous Glucose Monitor Data Analysis**  
 Lutsker *et al.* (2024)  
[arXiv:2408.11876](https://arxiv.org/abs/2408.11876)

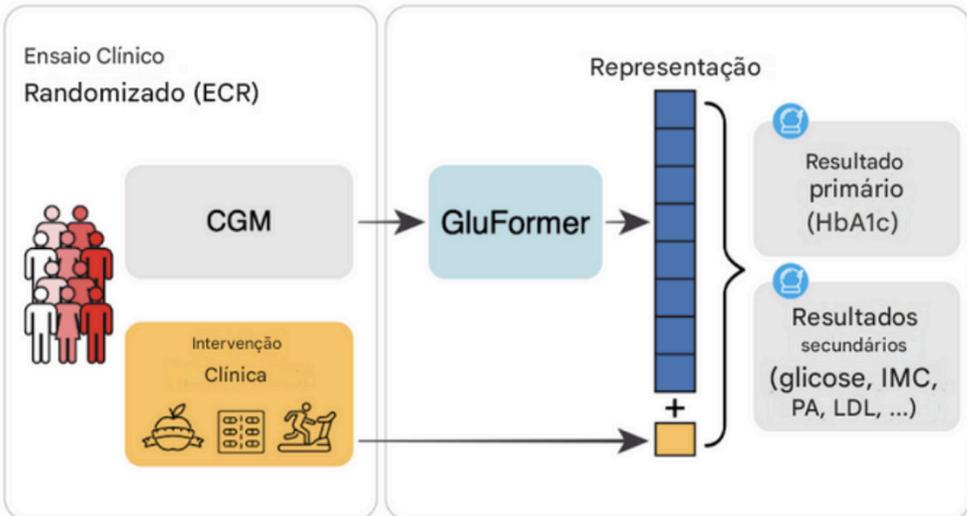
A. GluFormer - um modelo de base para dados CGM



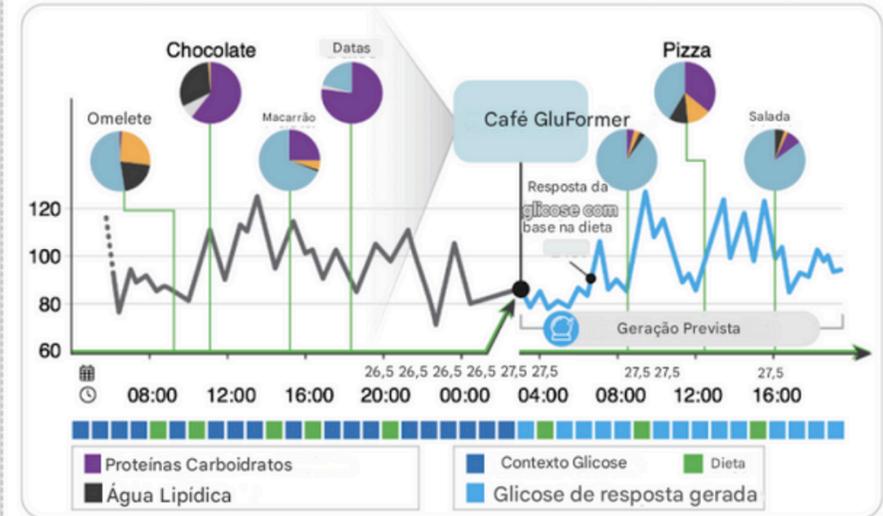
B. Generalizando para diversas patologias, dispositivos CGM e nacionalidades



C. Previsão de resultados de ECR



D. Adicionar dieta ao GluFormer para prever a resposta da glicose



# Modelos grandes e pequenos em IA para biologia

- **Com dados suficientes\***, os modelos grandes tendem a vencer (ensaio "The Bitter Lesson", de Rich Sutton)
- **Na maioria dos domínios biológicos, não temos dados\*** suficientes para criar bons modelos grandes.
- **Sem dados suficientes\***, geramos muito mais dados ou desenvolvemos modelos menores baseados em princípios que podem funcionar com os poucos dados que temos e podem utilizar os dados que podemos gerar sem grandes esforços.
- **Os modelos pequenos são úteis e adequados para esforços acadêmicos de pequenos grupos. Os modelos grandes exigem esforços e engenharia em escala industrial.**

***\* Não sabemos realmente o quanto é "suficiente"***

# O espectro de modelos

**Modelos  
analíticos**

**Modelos  
biofísicamente  
inspirados que  
aprendem com  
os dados**

**Grandes modelos  
*black box***

Pensamento  
da física

Modelagem bayesiana generativa  
Autoencodificadores variacionais  
Modelos gráficos probabilísticos

Pensamento  
de ML/AI



# Agentes de IA como assistentes científicos

## Busca e sumarização

## Referências da literatura

Platform <

New Task +

Recent Tasks [view all](#)

- Tacrolimus e superior a ci...
- nanobody
- Has anyone Tell me about ...
- Tell me about HSP1 and to...
- Tell me about HSP1 and to...

**Task:** Tell me about HSP1 and tomatoes heat resistance

Falcon - Deep Search **SUCCESS**

Task Details

Share

**Results** Reasoning References 11

This report details the current understanding of the tomato small heat shock protein HSP1 and its putative role in mediating heat resistance. Although integrated target-disease mapping platforms such as the Open Targets Platform do not yet document a clear association between HSP1 and heat stress (1.1, Abiotic Stress-HSP1), independent proteomic and transcriptomic studies provide evidence that small heat shock proteins (sHSPs) such as HSP1 contribute to cellular protection under elevated temperatures.

### I. Introduction

Heat stress imposes severe limitations on plant growth and crop yield, and tomato (*Solanum lycopersicum*) is particularly susceptible during reproductive and developmental phases. To mitigate the deleterious effects of high temperature, plants employ a sophisticated network of heat shock proteins (HSPs) that function as molecular chaperones to stabilize, refold, or facilitate the degradation of proteins that become misfolded or aggregated during stress (2.1). In this context, HSP1 is presumed to be a member of the sHSP family, implicated in the response to heat stress through its role in maintaining protein homeostasis. The following sections review the molecular features of HSP1, summarize proteomic findings supporting its upregulation in response to heat stress, and discuss the potential mechanisms by which HSP1 contributes to tomato thermotolerance.

### II. Molecular Function and Family Context

Small HSPs have long been recognized as first-line defense proteins that bind to unfolded polypeptides in an ATP-independent manner, effectively preventing irreversible aggregation (2.1). HSP1, as a putative member of this family in tomato, is expected to share similar properties. Its primary function is to act as a molecular chaperone activity, preventing the aggregation of proteins and assisting in their denaturation where rapid stabilization of non-native proteins is required until ATP-

**Task:** Tell me about HSP1 and tomatoes heat resistance

Falcon - Deep Search **SUCCESS**

Task Details

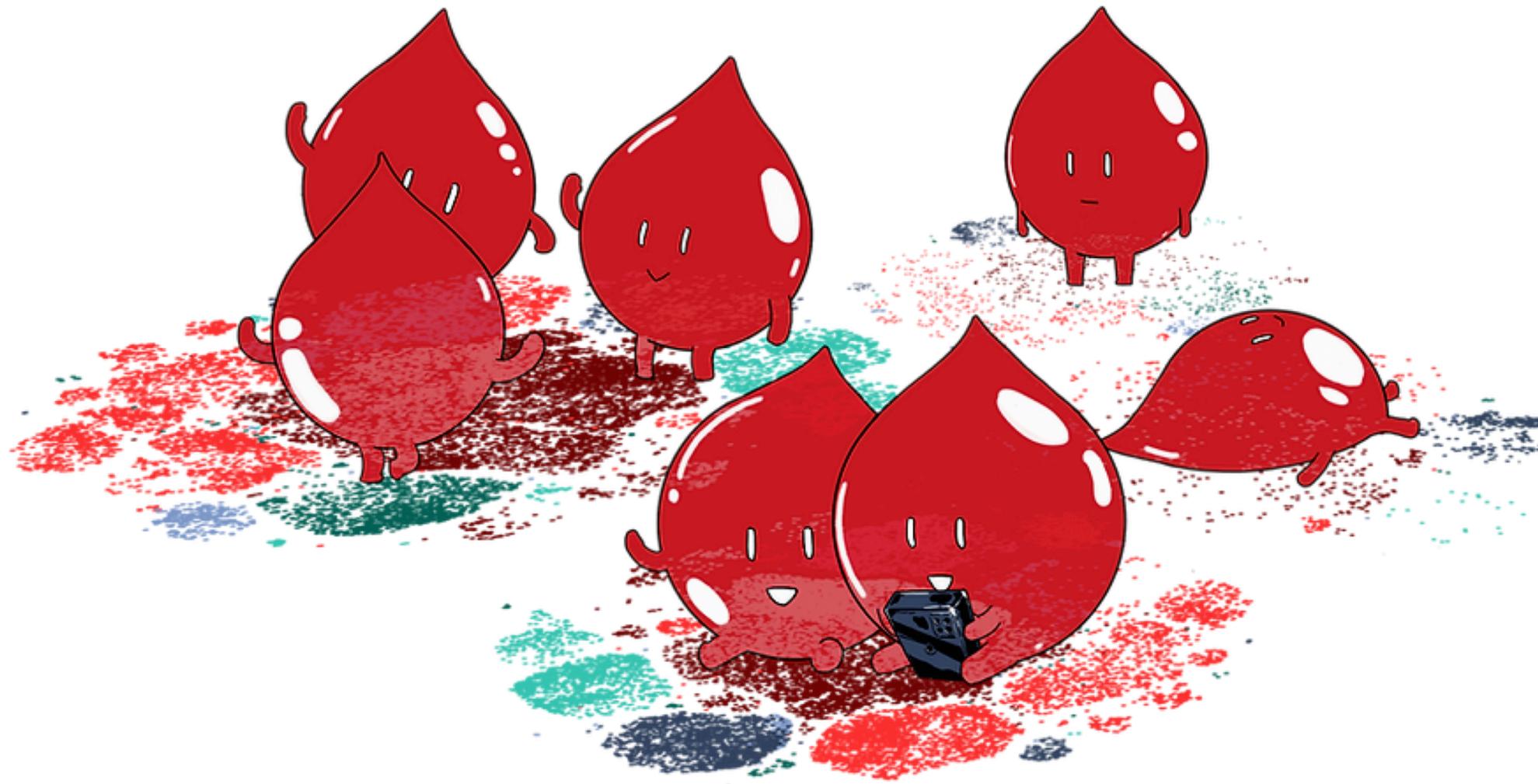
Share

Results Reasoning **References 11**

- Open Targets Disease-Associations: Heat Stress, Abiotic Stress-HSP1** [↗](#)  
Open Targets 2025  
Contexts: Used 1.1
- Transcriptomics of Heat Stress in Plants**  
Boghiredy Sailaja, Satendra K. Mangrauthia, Neelamraju Sarla, Sitapati Rao Voleti July 2014 • citations 9  
Contexts: Used 2.1 2.2 Unused 2.3
- Melatonin enhances thermotolerance by promoting cellular protein protection in tomato plants**  
Wen Xu, Shu-Yu Cai, Yun Zhang, Yu Wang, Golam Jalal Ahammed, Xiao-Jian Xia, Kai Shi, Yan-Hong Zhou, Jing-Quan Yu, Russel J. Reiter, Jie Zhou • Journal of Pineal Research, Nov 2016 **DOMAIN LEADING** • citations 271  
Contexts: Used 3.1 3.2 3.3 3.4 Unused 3.5 3.6
- Response mechanisms induced by exposure to high temperature in anthers from thermo-tolerant and thermo-sensitive tomato plants: A proteomic perspective** [↗](#)  
Maria Fiorella Mazzeo, Giuseppina Cacace, Paolo Iovieno, Immacolata Massarelli, Stefania Grillo, Rosa Anna Siciliano • PLOS ONE, July 2018 **PEER REVIEWED** • citations 41  
Contexts: Used 4.1 Unused 4.2
- The  $\alpha$ -Crystallin Domain Containing Genes: Identification, Phylogeny and Expression Profiling in Abiotic Stress, Phytohormone Response and Development in Tomato (*Solanum lycopersicum*)** [↗](#)  
Asosii Paul, Sombir Rao, Saloni Mathur • Frontiers in Plant Science, Mar 2016 **PEER REVIEWED** • citations 32  
Contexts: Used 5.1 Unused 5.2
- Developmental and heat stress-regulated expression of HsfA2 and small heat shock proteins in tomato anther** [↗](#)  
Filomena Giorno, Mieke Wolters-Arts, Stefania Grillo, Klaus-Dieter Scharf, Wim H. Vriezen, Celestina Mariani • Journal of Experimental Botany, Oct 2009 **DOMAIN LEADING** • citations 179  
Contexts: Used 6.1 6.2 Unused 6.3 6.4 6.5
- Enhancing the thermotolerance of tomato seedlings by heat shock treatment** [↗](#)  
Zhenrui Zhang, M. A. Umutoni • Photosynthetica, Nov 2019 **PEER REVIEWED** • citations 1  
Contexts: Used 7.1 7.2 Unused 7.3

<https://futurehouse.org>

# Obrigado!



Eduardo da Veiga Beltrame  
eduardo.beltrame@mbzuai.ac.ae

Gostaria de fazer um workshop online sobre análise de dados de single cell em Python com scvi-tools?



Registre seu interesse no formulário de feedback e entraremos em contato!