

# From Data to Discovery: How GenAI is Changing Bioinformatics

Iris Wing To Lam  
Co-founder and CEO, Genie TechBio

1. **Why Bioinformatics (BFX)?**
2. **My personal experience**
3. **Applications of GenAI in BFX**
4. **Future Potential of GenAI in BFX**
5. **Conclusion**

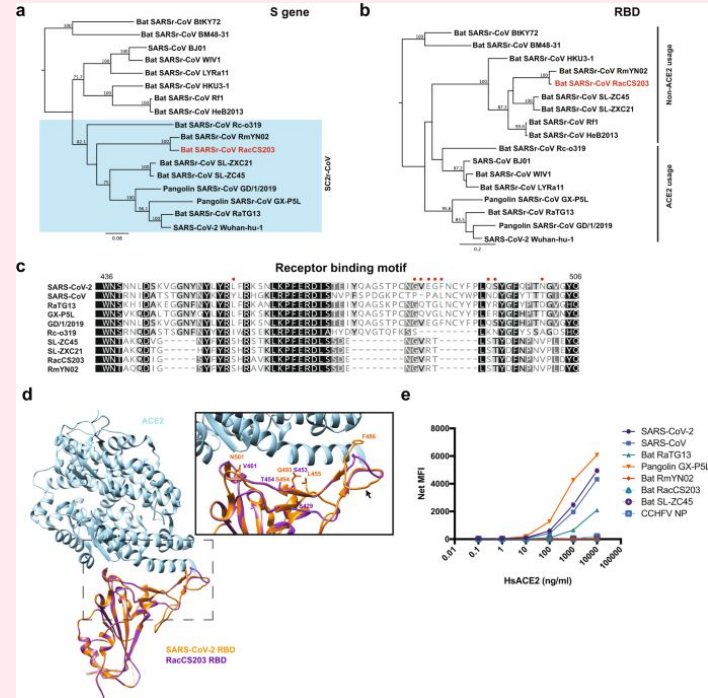
1. **Why Bioinformatics (BFX)?**
2. **My personal experience**
3. **Applications of GenAI in BFX**
4. **Outlook and Conclusion**

Bioinformatics is an **interdisciplinary** field that combines biology, computer science, and mathematics to **analyze and interpret complex biological data.**

Think of bioinformatics as the **Google Maps for biological research** – it helps **scientists navigate through vast amounts of data to find their destination: understanding diseases.**

# Understanding Virus Structure and Function

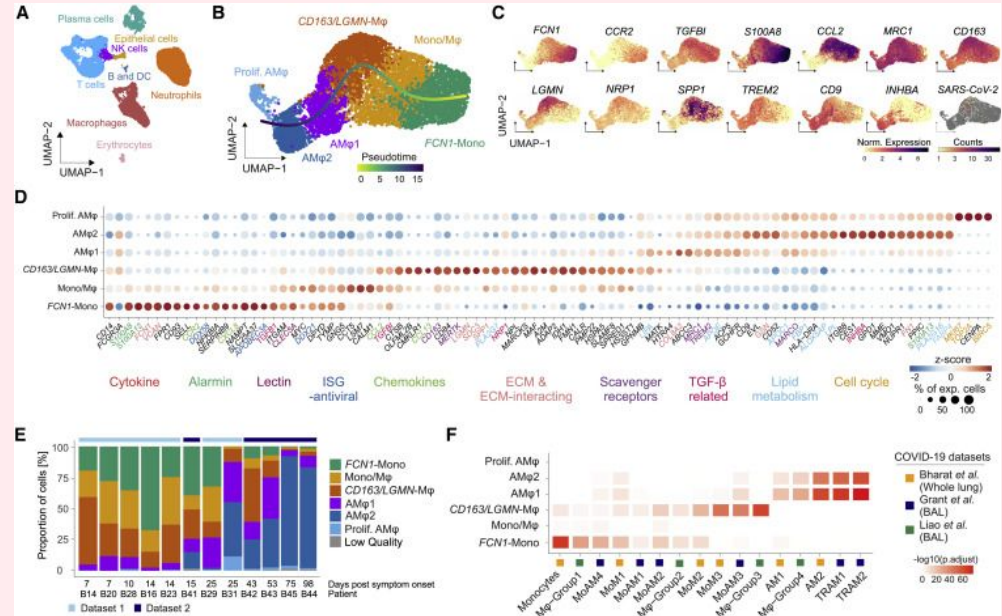
By utilizing genomic sequencing, researchers were able to identify viral genes and proteins that could serve as targets for vaccine development.



Wacharapluesadee, S., Tan, C.W., Maneerom, P. et al. Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. Nat Commun 12, 972 (2021). <https://doi.org/10.1038/s41467-021-21240-1>

# Identifying Immune Responses

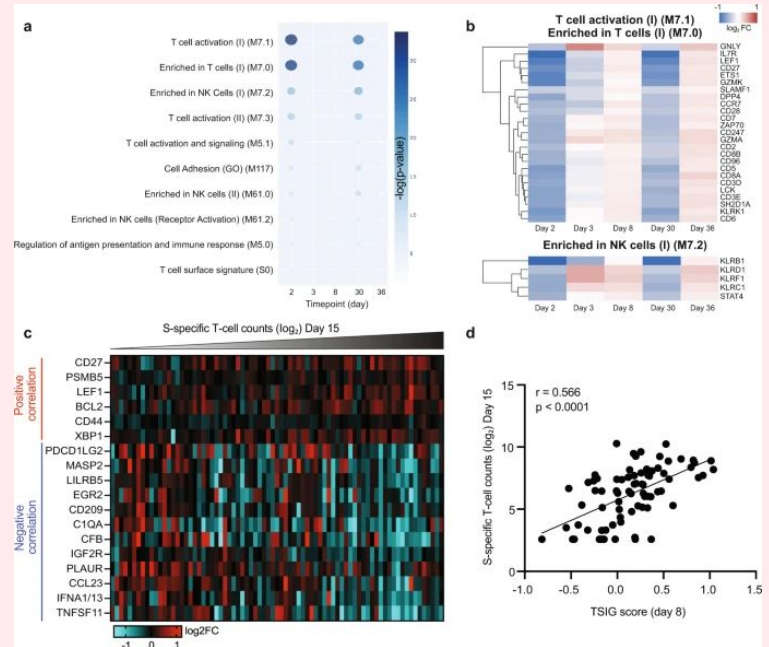
Transcriptomics and proteomics provided insights into how the immune system responds to vaccination.



Wendisch D, Dietrich O, Mari T, et al. SARS-CoV-2 infection triggers profibrotic macrophage responses and lung fibrosis. *Cell*. 2021;184(26):6243–6261.e27. doi:10.1016/j.cell.2021.11.033

# Accelerating Antigen Discovery

Viral genome sequences, transcriptomic and proteomic data allowed researchers to identify specific antigens from the SARS-CoV-2 virus that could elicit a strong immune response.

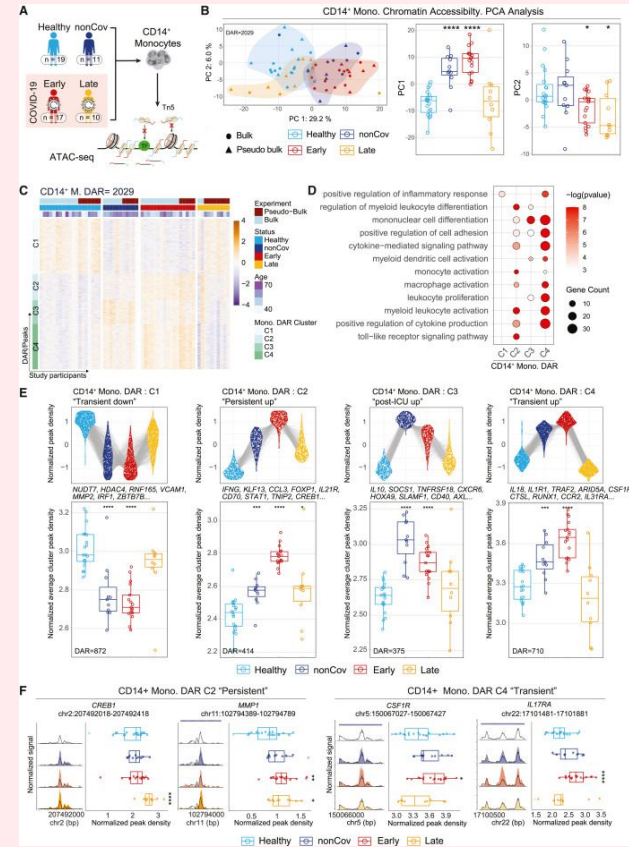


Ong, E.Z., Yee, J.X., Ooi, J.S.G. et al. Immune gene expression analysis indicates the potential of a self-amplifying Covid-19 mRNA vaccine. *npj Vaccines* 7, 154 (2022). <https://doi.org/10.1038/s41541-022-00573-y>



# Guiding Vaccine Design

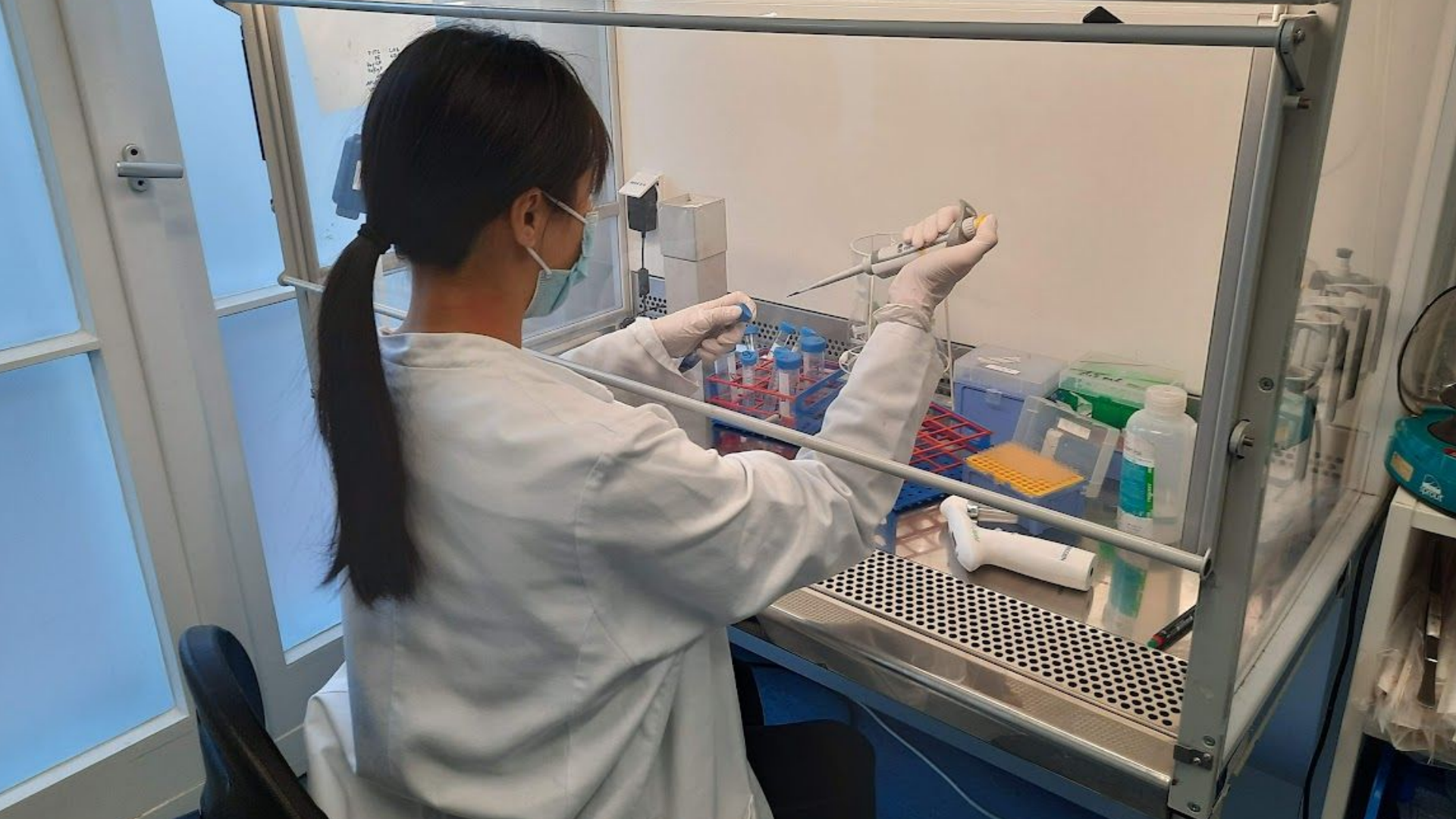
Omics analysis helped researchers understand how to induce long-lasting immunity, which is essential for effective vaccination strategies.



Cheong JG, Ravishankar A, Sharma S, et al. Epigenetic memory of coronavirus infection in innate immune cells and their progenitors. *Cell*. 2023;186(18):3882–3902.e24. doi:10.1016/j.cell.2023.07.019

**Bioinformatics enables faster, more accurate vaccine and drug design, leading to improved patient outcomes.**

1. Why Bioinformatics (BFX)?
2. **My personal experience**
3. Applications of GenAI in BFX
4. Outlook and Conclusion



There is a **shortage** of skilled **bioinformaticians**, and it's delaying the progress of biomedical research.

## Genie! TechBio

Welcome to Genie!

I have single-cell data of brain organoids and want to compare its cell types to existing published data.

Sure! Please upload your data set and I will help you analyze it.

 GSE109564...



## Genie! TechBio

Welcome to Genie!

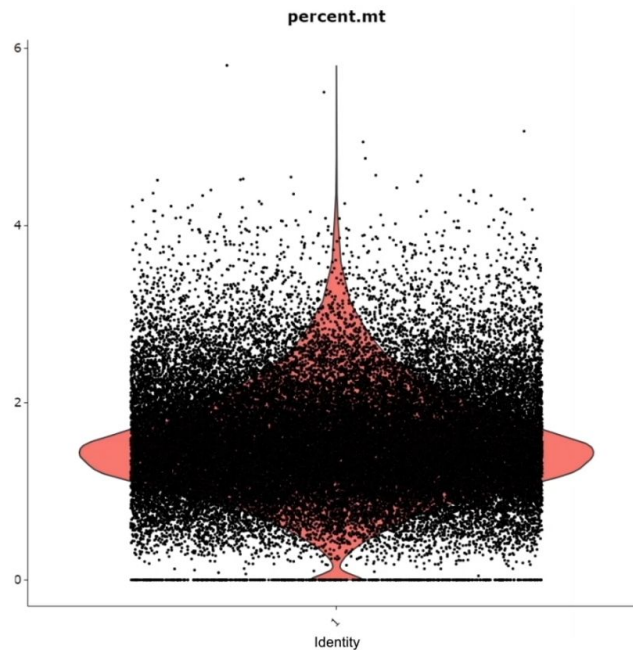
I have single-cell data of brain organoids and want to compare its cell types to existing published data.

Sure! Please upload your data set and I will help you analyze it.



I see you uploaded a count matrix of single-cell RNA-seq data. The data consists of 36,321 genes across 47,024 samples. I suggest we first do some basic quality control checks, such as determining the fraction of mitochondrial genes and the relationship between the number of unique molecules and genes in each cell.

Ok.



## Genie! TechBio

Welcome to Genie!

I have single-cell data of brain organoids and want to compare its cell types to existing published data.

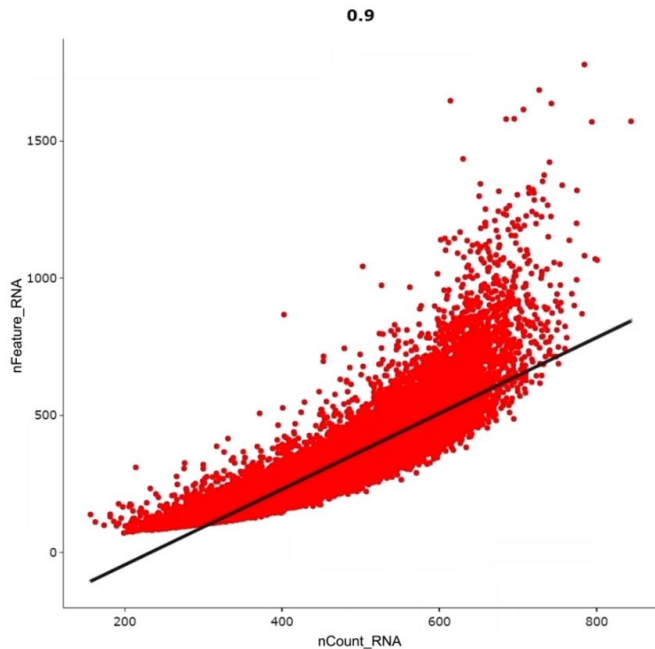
Sure! Please upload your data set and I will help you analyze it.



I see you uploaded a count matrix of single-cell RNA-seq data. The data consists of 36,321 genes across 47,024 samples. I suggest we first do some basic quality control checks, such as determining the fraction of mitochondrial genes and the relationship between the number of unique molecules and genes in each cell.

Ok.

The data quality looks very good! Cells seem to be healthy, and there are very few doublets or other technical artefacts. Please let me know if you have any questions about these plots.





## Genie! TechBio

Welcome to Genie!

checks, such as determining the fraction of mitochondrial genes and the relationship between the number of unique molecules and genes in each cell.

Ok.

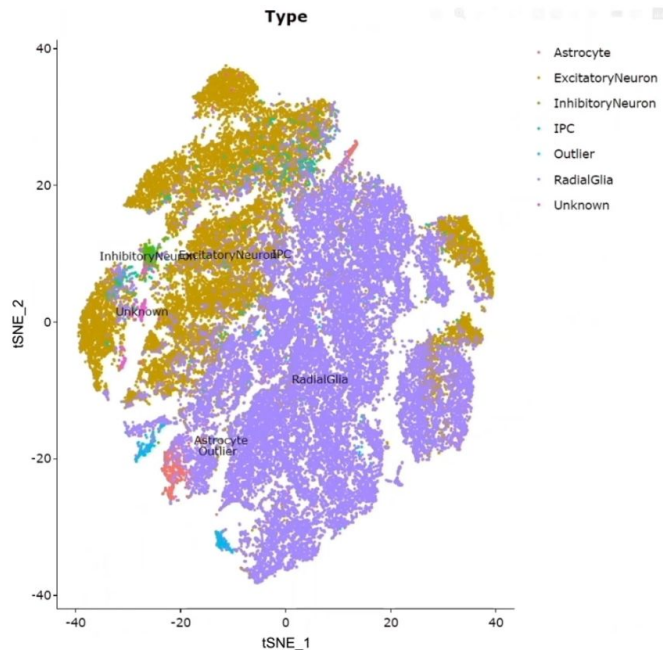
The data quality looks very good! Cells seem to be healthy, and there are very few doublets or other technical artefacts. Please let me know if you have any questions about these plots.

Please compare the cell types in my data to those in Bhaduri et al. 2020.

Sure! I just scanned that paper and found the relevant data sets. Since this data set already includes cell type annotations, I suggest using their organoid data set to build a model for assigning cell type labels to your data. This will allow us to compare the cell types in your data sets to those from the study. What do you think?

Great!

I ha ●



## Genie! TechBio

Welcome to Genie!

Great!

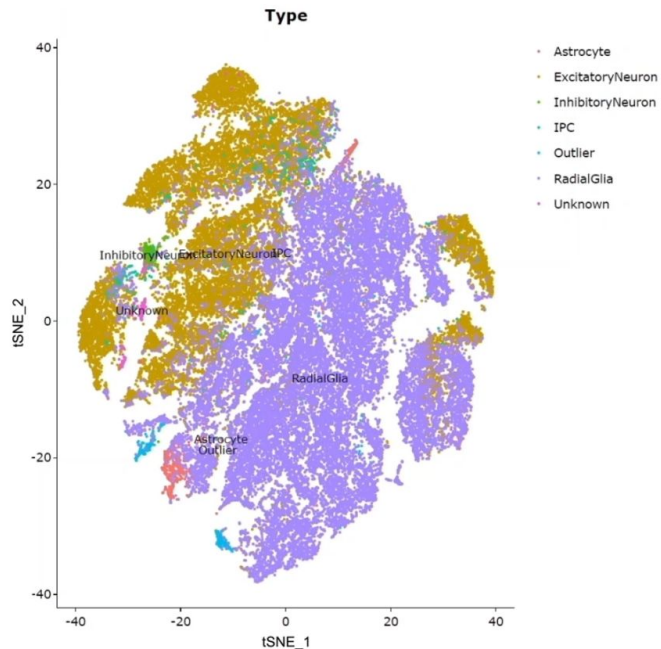
I have built a cell type model using the organoid data set from Bhaduri et al. 2020 consisting of 235,121 cells. Then, I assigned cell types to your data using this model. You see a UMAP plot showing the separation of cells in your data. The cluster labels correspond to the newly assigned cell types.

Can you explain more?

Sure! First, I downloaded the "Aparna Bhaduri (2019-06-01), Version 3" data set from Bhaduri et al. 2020 from the UCSC web portal. Next, I used CellTypist to build a cell type model using the labelled cell types in that data set. I then calculated the principal components in the data set you uploaded, which allowed me to perform k-nearest neighbor embedding and clustering of your cells. Finally, I used the CellTypist model to infer the cell types in your data.

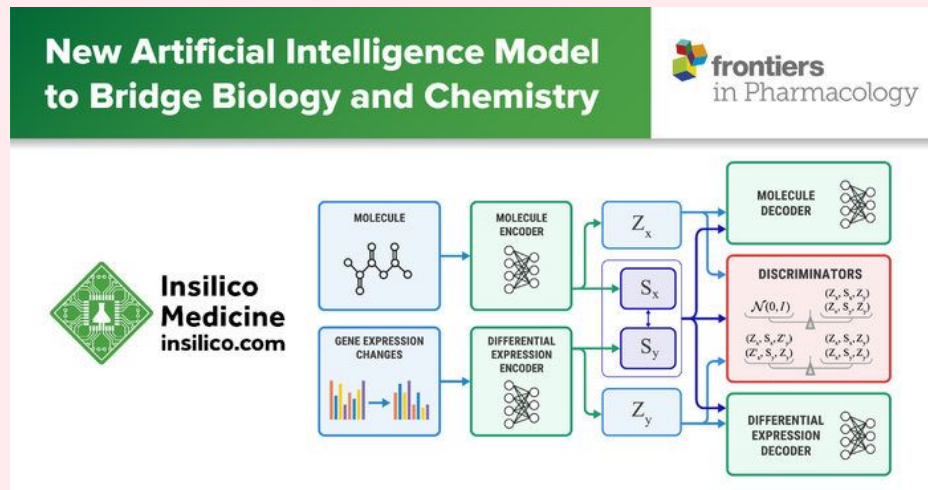
The plot shows the cells in your data, projected on two dimensions using the UMAP method. The cells are colored using the cell types assigned to your data by the model. You can compare this plot to Fig. 1b in Bhaduri et al. 2020. Please let me know if you have any more questions!

🔊 📄 👍 🗑️ ↻



**Genie TechBio harnesses the power of GenAI to unlock access to complex yet powerful bioinformatics tools to make research faster, cheaper and more reproducible.**

1. Why Bioinformatics (BFX)?
2. My personal experience
3. **Applications of GenAI in BFX**
4. Outlook and Conclusion

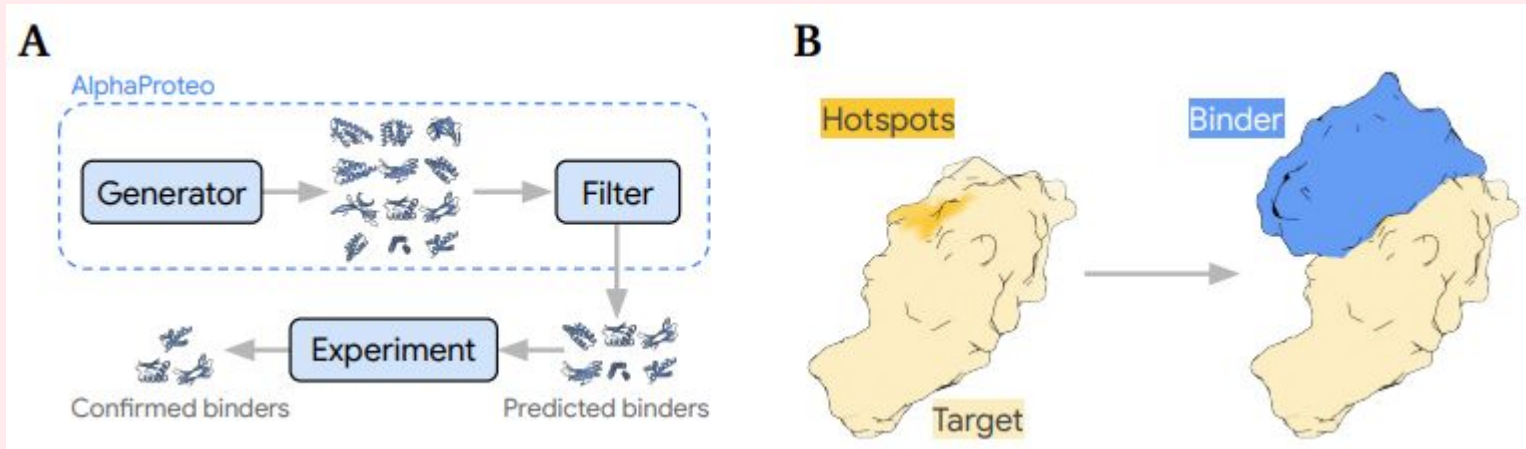


# Drug Discovery and Development

Insilico Medicine has pioneered the use of generative adversarial networks (GANs) for de novo drug design, integrating advanced AI methodologies into their drug discovery processes. They consist of two neural networks, the generator and the discriminator, which engage in a competitive process to improve their performance. This architecture allows GANs to generate realistic data that closely resembles a given training dataset. This approach allows for the rapid identification and synthesis of novel drug candidates, significantly accelerating the timeline and reducing costs associated with traditional drug development.

# Protein Engineering

Google DeepMind has introduced AlphaProteo, a tool for designing novel proteins that can bind effectively to target molecules. This innovation illustrates how GenAI can facilitate the engineering of proteins for therapeutic purposes.



1. **Why Bioinformatics (BFX)?**
2. **My personal experience**
3. **Applications of GenAI in BFX**
4. **Outlook and Conclusion**

We'll see **drug candidates** generated in mere **days instead of years**, and clinical trials that adapt in real-time based on patient responses.



**This accessibility** will fundamentally speed up biomedical research, enabling scientists to **focus on discovery** rather than data bottlenecks.

Iris Wing To Lam  
Co-founder and CEO, Genie TechBio

[iris@genietechbio.com](mailto:iris@genietechbio.com)