

EP 17

October 13, 2026 · 00:47:20

Math and Biological Data - Jingyuan Hu

Jingyuan Hu — Postdoctoral Researcher, Harvard Biostatistics



Jingyuan Hu on inferring where a cell is heading from one snapshot of single-cell data, why more cells of the same kind hit a ceiling for biology's AI models, and learning the mathematics on YouTube.

20,000

genes whose expression sequencing reports [3:33](#)

1 cell

is where a human body of trillions starts [12:25](#)

1 gene

at a time is often all validation can look at [16:20](#)

3

parameters in Newton's second law; AI has trillions [37:06](#)

KEY TAKEAWAYS

1 A snapshot holds the dynamics

Sequencing freezes each cell at one moment, yet where cells are heading can be inferred from it, like cars in a photograph of a highway. [6:42](#)

2 FlowMap infers the landscape

His method learns Waddington's landscape from the molecular profiles, to ask why two similar stem cells choose different fates. [13:00](#)

3 Validation is the bottleneck

Single-cell data makes hypotheses easy to generate, but checking them can mean looking at one gene at a time. [16:08](#)

4 More cells hit a hard ceiling

Random matrix theory shows a saturation point, past which more cells of the same kind add very little biological information. [25:01](#)

5 Expertise can be built on YouTube

With little mathematics in his formal training, he learned it course by course on YouTube, starting with Gilbert Strang's linear algebra. [32:08](#)

“If you are curious, you can build expertise outside your formal training.”

Jingyuan Hu · [33:07](#)

CHAPTERS

0:00	Intro	2 min	22:11	Random matrix theory in simple words	2 min
1:32	What single-cell data can show	2 min	24:33	The hard ceiling on more cells	2 min
3:57	Manifolds: the geometry of cell data	1 min	26:22	Is AI for biology overhyped?	3 min
5:20	Dynamics from a single snapshot	6 min	29:02	The virtual cell and treatment	2 min
11:00	FlowMap and a cell's fate	4 min	30:55	Learning math from YouTube	4 min
15:13	Validation, the field's bottleneck	4 min	35:17	Physics models vs. black-box AI	6 min
19:01	Does biology just need more data?	3 min	41:09	Creating new biology	6 min

TERMS EXPLAINED

Single-cell sequencing

Reading the molecular profile of each cell separately, not the total for a whole tissue.

RNA velocity

A method for inferring which way a cell's state is changing from a single measurement.

Random matrix theory

The study of large tables of random numbers, used to tell the signal in data from its noise.

Manifold

A curved surface of few dimensions that data with thousands of features lies close to.

Waddington's landscape

A metaphor for development: a ball rolling down a hill, ending as one cell type or another.

Virtual cell

A computer model that predicts how a cell responds to a drug or another intervention.

MENTIONED IN THIS EPISODE

[The FlowMap preprint](#)

[FlowMap documentation](#)

[FlowMap on GitHub](#)

[His random matrix theory series](#)

[Gilbert Strang's linear algebra](#)

Listen, watch or download the transcript: aarsentyev.com/podcast/ep17-jingyuan-hu

Each time opens the video at that moment. Hosted by Arseniy Arsenyev. A summary of a conversation, not medical advice.