Bridging Machine Learning and Mathematical Modeling to Decipher Tumor heterogeneity and Therapy Resistance

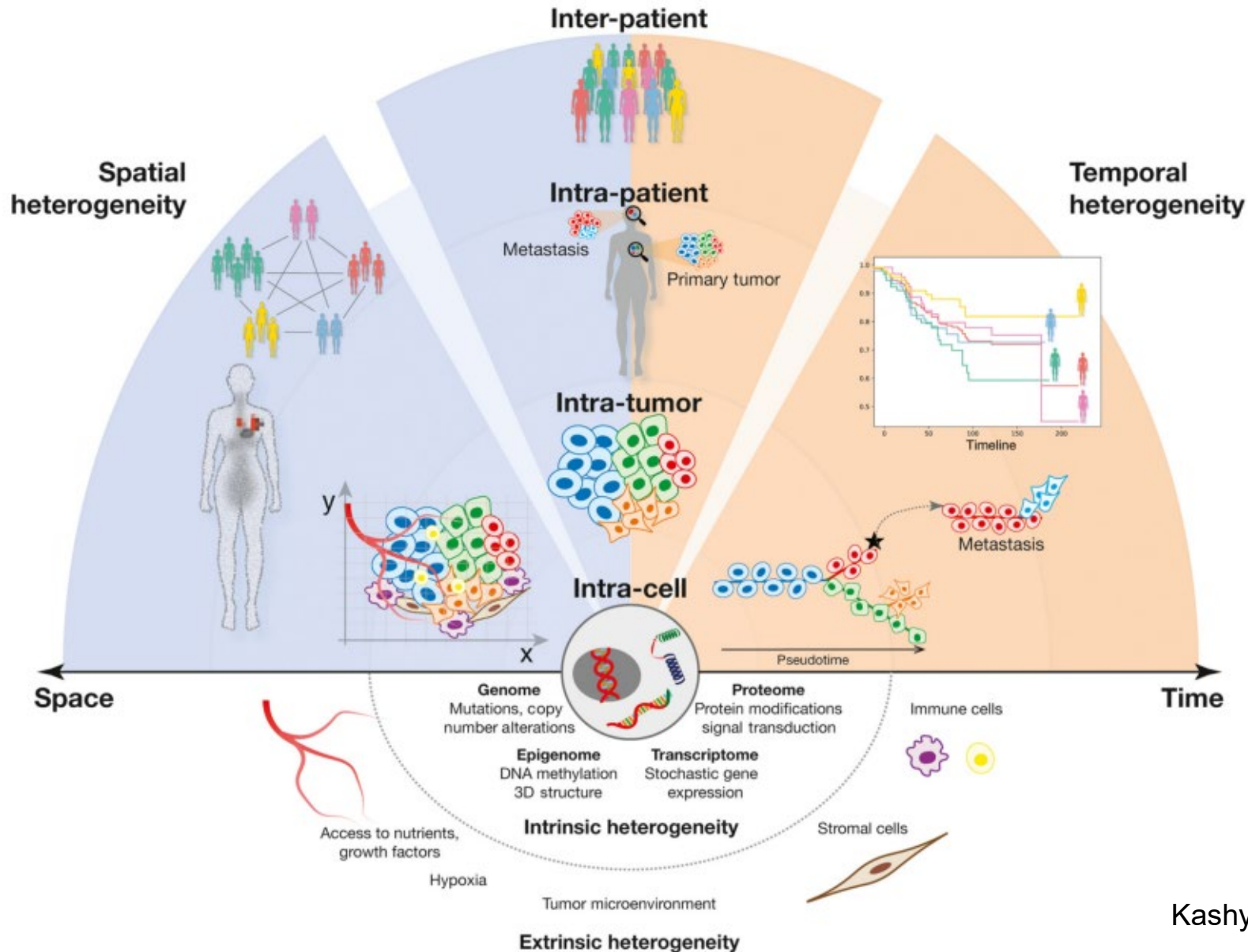## Xiufen Zou

School of Mathematics and Statistics, Wuhan University, China

xfzou@whu.edu.cn
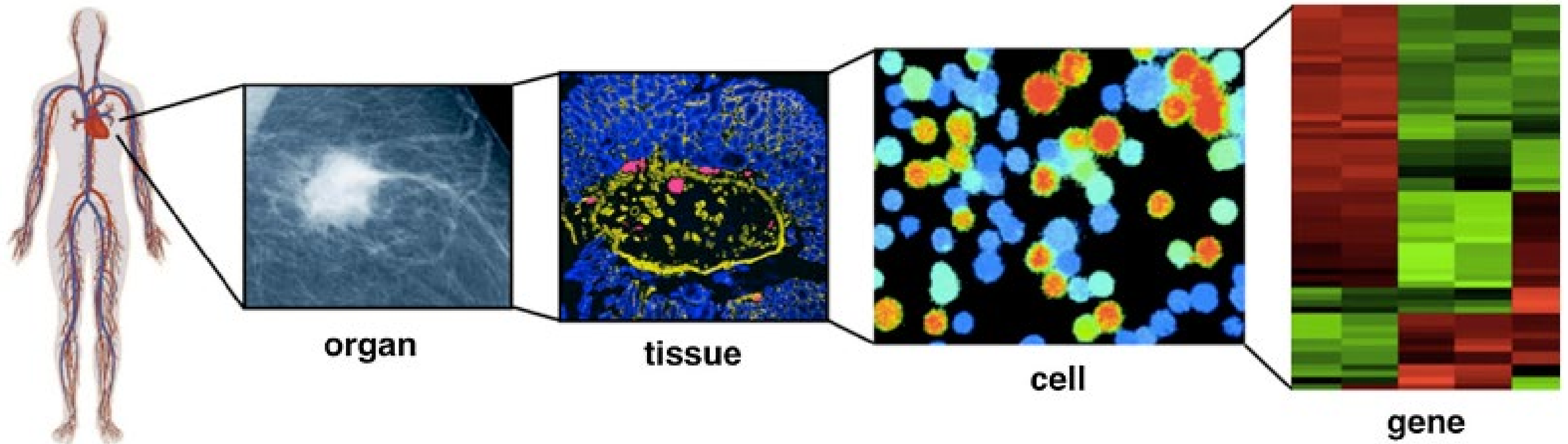
July 28-August 01, 2025, ICERM at Brown University

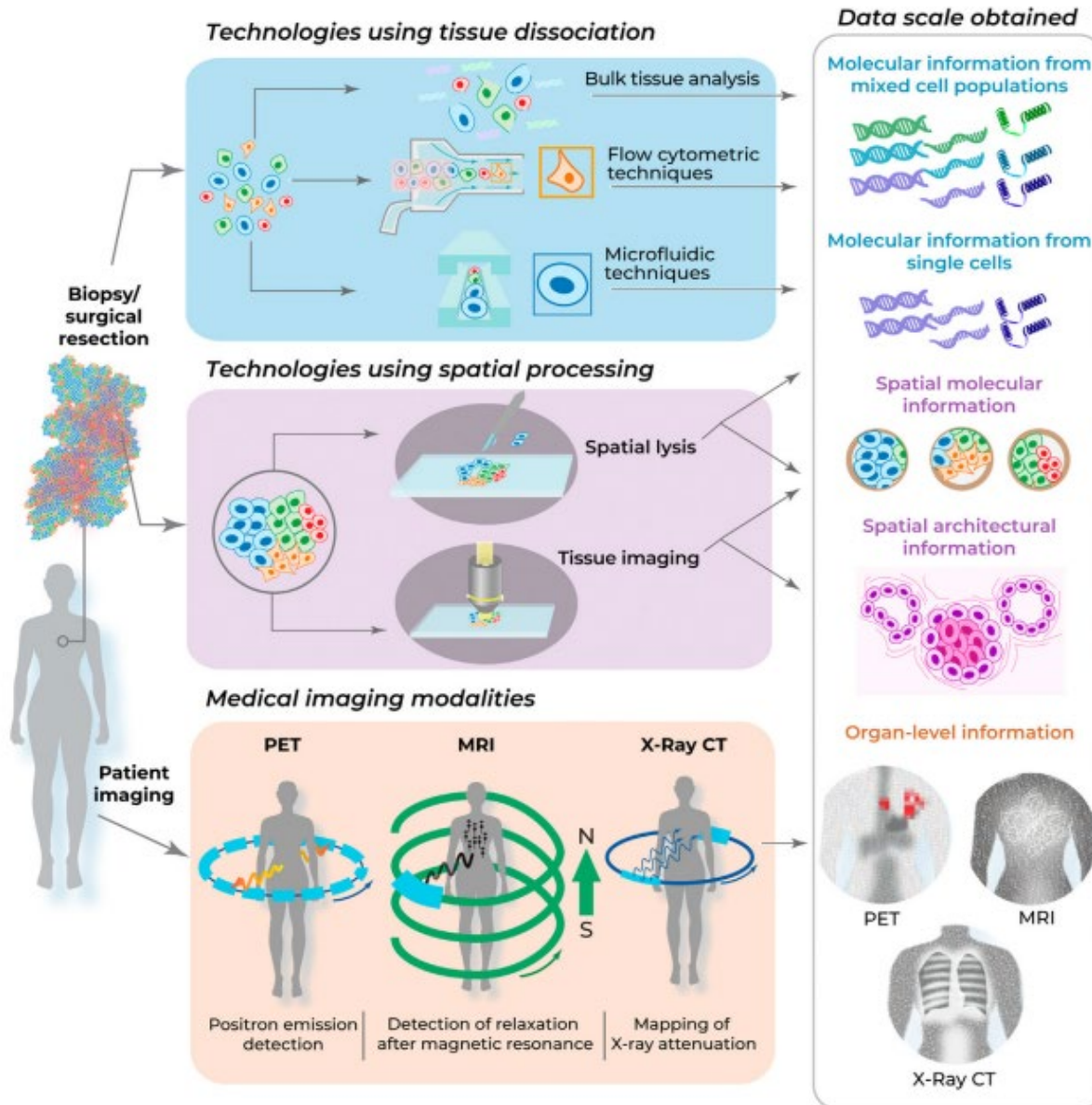# Introduction: Tumor heterogeneity

- Tumor heterogeneity across scales and dimensions

- Tumor heterogeneity affects disease outcome and treatment response

Introduction: Tumor evolution is multiscale

Tumor evolution involves changes at the molecular, cellular, and tissue scales, which can ultimately harm organs and the entire body.

organ   tissue   cell   gene

Anderson A R A, Maini P K. Mathematical Oncology. Bulletin of Mathematical Biology, 2018, 80: 945-953.
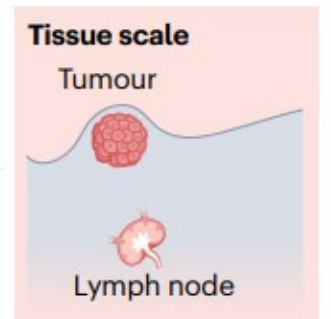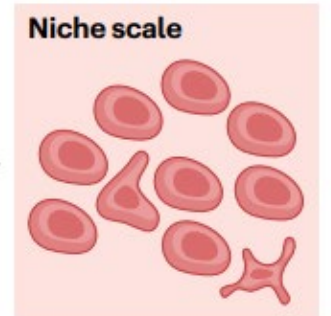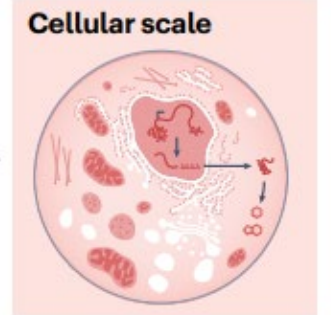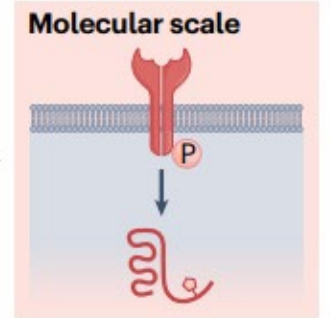
Introduction: Multidimensional Data

Bulk omics data

Single cell sequencing data

Spatial data

Medical imaging

Clinical data

Kashyap et al., Trends in Biotechnology 2022

David S. Fischer et al., Nature reviews genetics, 2025,

Introduction: Outstanding questions
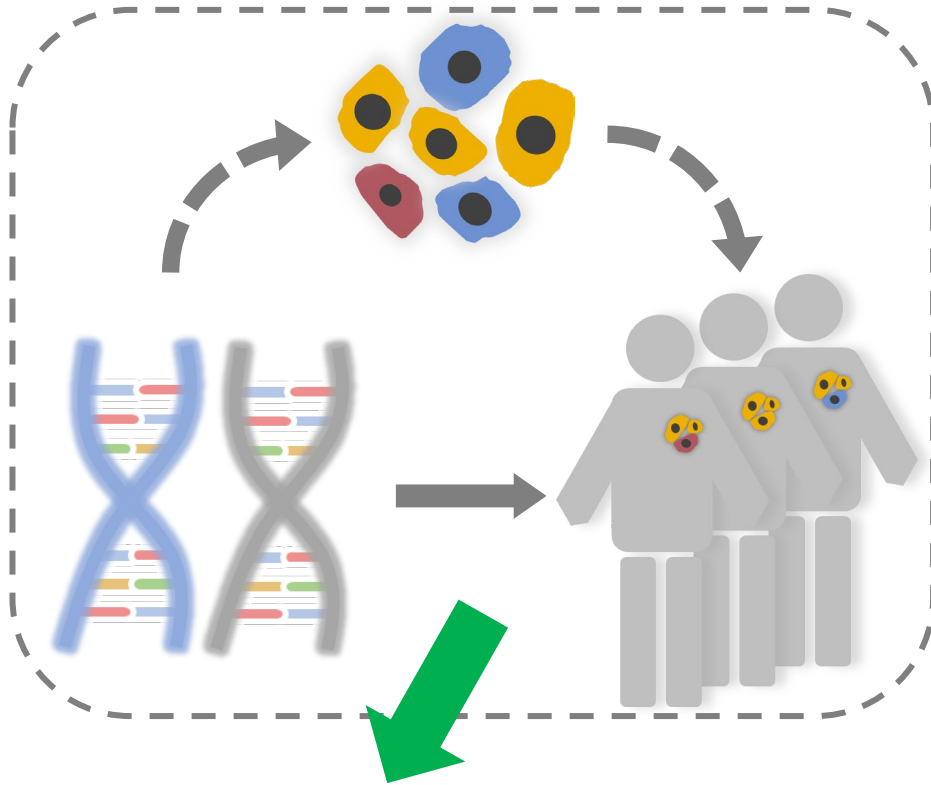
Several important outstanding questions in precision medicine require resolution through quantitative analysis of multidimensional datasets.

➢ How can we quantify tumor heterogeneity?

➢ Can short-term measurements predict long-term outcomes?

➢ How can we predict drug efficacy and disease prognosis?

➢ ... and other related questions.

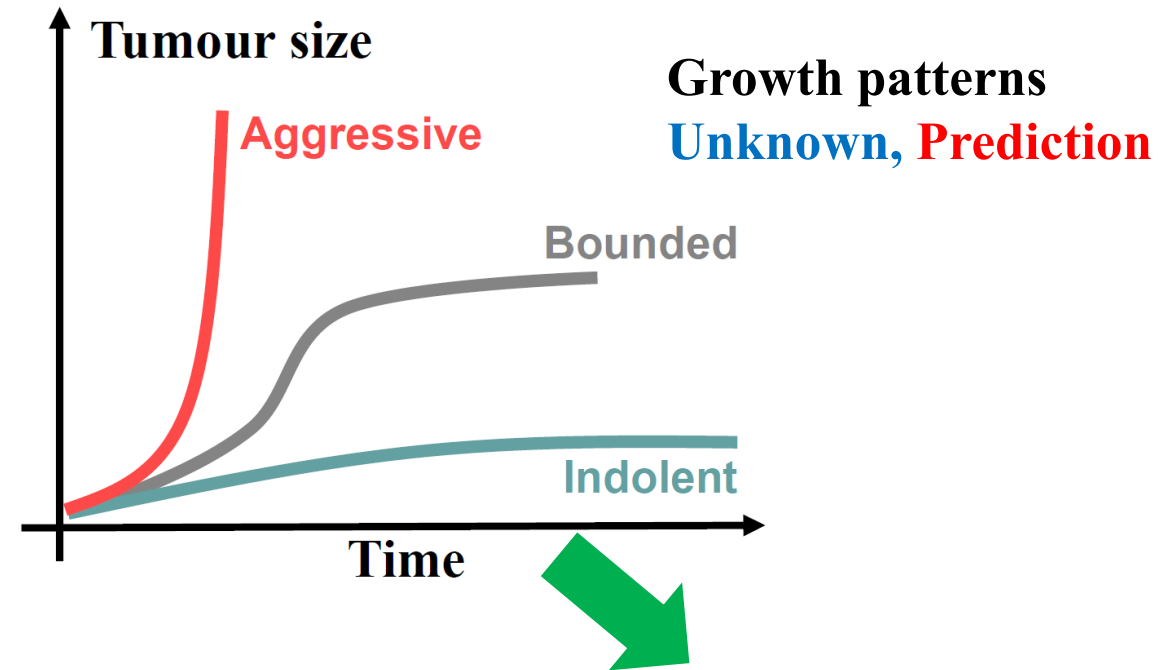# Introduction：Challenging

Available multiscale data

Disease progression

Tumour size

Aggressive

Growth patterns
Unknown, Prediction

Bounded

Indolent

Time

- Short term
- Discrete
- Static

Gap

- Long term
- Continuous
- Dynamics

Introduction: Integration of mathematical modeling with AI

Using mathematical modelling and AI to improve delivery and efficacy of therapies in cancer

Constantinos Harkos, Andreas G. Hadjigeorgiou, Chrysovalantis Voutouri, Ashwin S. Kumar, Triantafyllos Stylianopoulos & Rakesh K. Jain

A more effective way is to combine mathematical modelling approaches with artificial intelligence (AI) , including machine learning, to enabling efficient analysis of large datasets and enhance the accuracy of cancer prognosis and diagnostics.

Introduction: Why do combine AI and MM?

Mathematical modeling (MM)

Advantages

- Nonlinearity
- Dynamics
- The results are easily interpretable.
- Model predictions can be used to generate new hypotheses

Artificial intelligence (AI)

- A data-driven powerful tool
- Incorporate multidimensional data from the preclinical or clinical setting.
- Finding patterns without an understanding of the underlying biological mechanisms.

> Fortunately, integrating AI with MM can bridge these gaps, not only advancing our understanding of the biological mechanisms underlying cancer but also opening new ways for early detection, diagnosis, and personalized treatment.

Limitations

- ➤ Require the prior knowledge of biological mechanisms
- Model structure requires careful consideration
- Model calibration must avoid over-fitting to training data

- ➤ Require a large amount of detailed, high-resolution data
- ➤ The results may not be easily interpretable.
- ➤ Are computationally intensive

Introduction: How to integrate MM and AI

➢ To develop dynamics-informed and explainable AI approaches that advance both the understanding and treatment of complex diseases

|			
A Multiscale Framework Linking Cancer Growth Dynamics to Subclonal Evolution *Journal of Theoretical Biology 582 ,111743, 2024*	Integrating mathematical modeling and data analysis to reveal the dynamics of cancer evolution ➢ Journal of Mathematical Biology, 89:34, 2024 ➢ Inverse Problems, 40 (9), 2024	Multiscale Modeling and Multidimensional Data Reveal Epigenetic Instability's Role in Cancer Drug Resistance ➢ SIAM Journal on Applied Mathematics, 85(1), 2025. ➢ PLoS Comput Biol 21(2): e1012815, 2025	Linking Immune Patterns to Prognosis via Tumor Modeling and Spatial Transcriptomic Data (*In preparation*)

Work I

> A Multiscale Framework Linking Cancer Growth Dynamics to Cellular Subclonal Evolution

➢ Bridging diverse cellular-subclonal dynamics with heterogeneous population dynamics

Connection
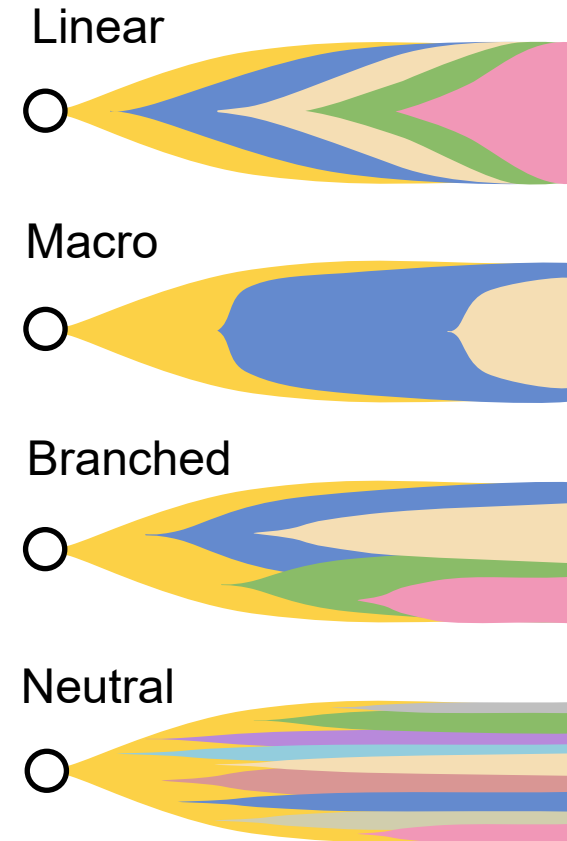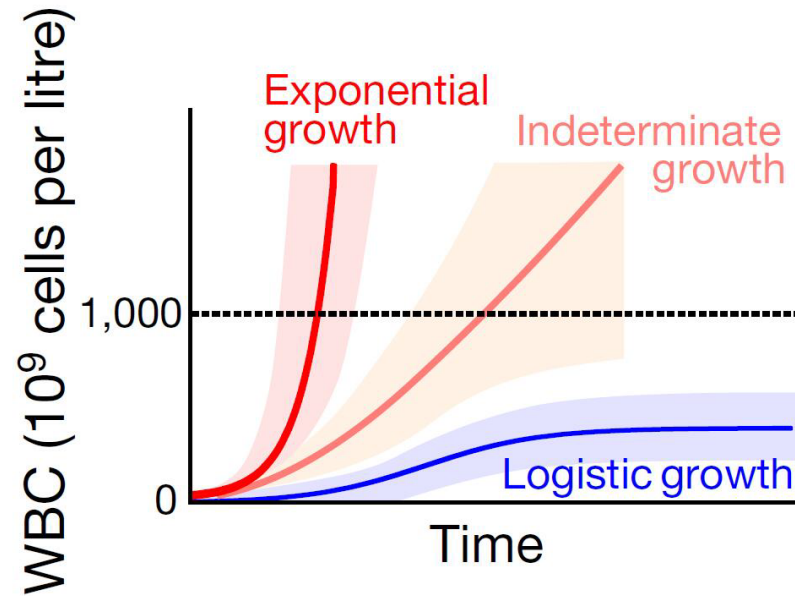
Heterogeneous growth patterns

Linear

Macro

Branched

Neutral
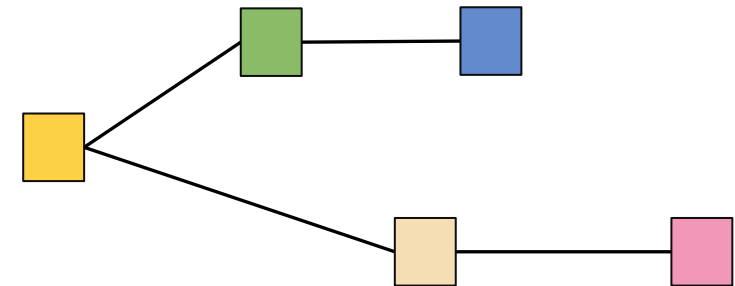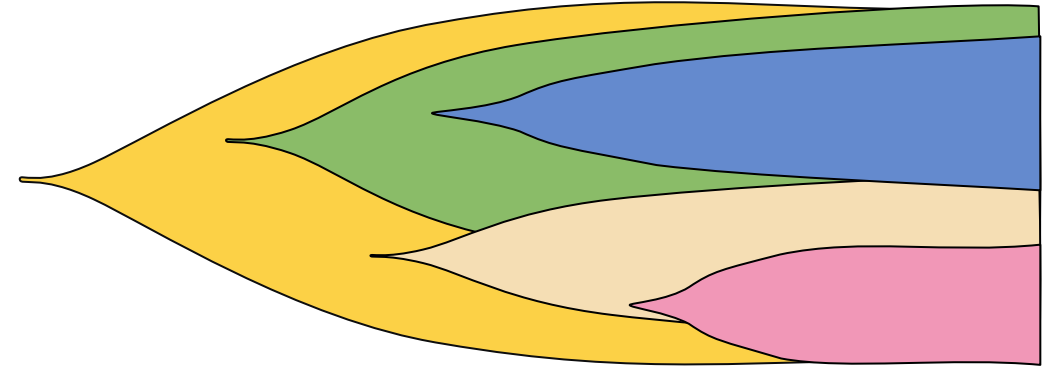
Diverse cellular-subclones

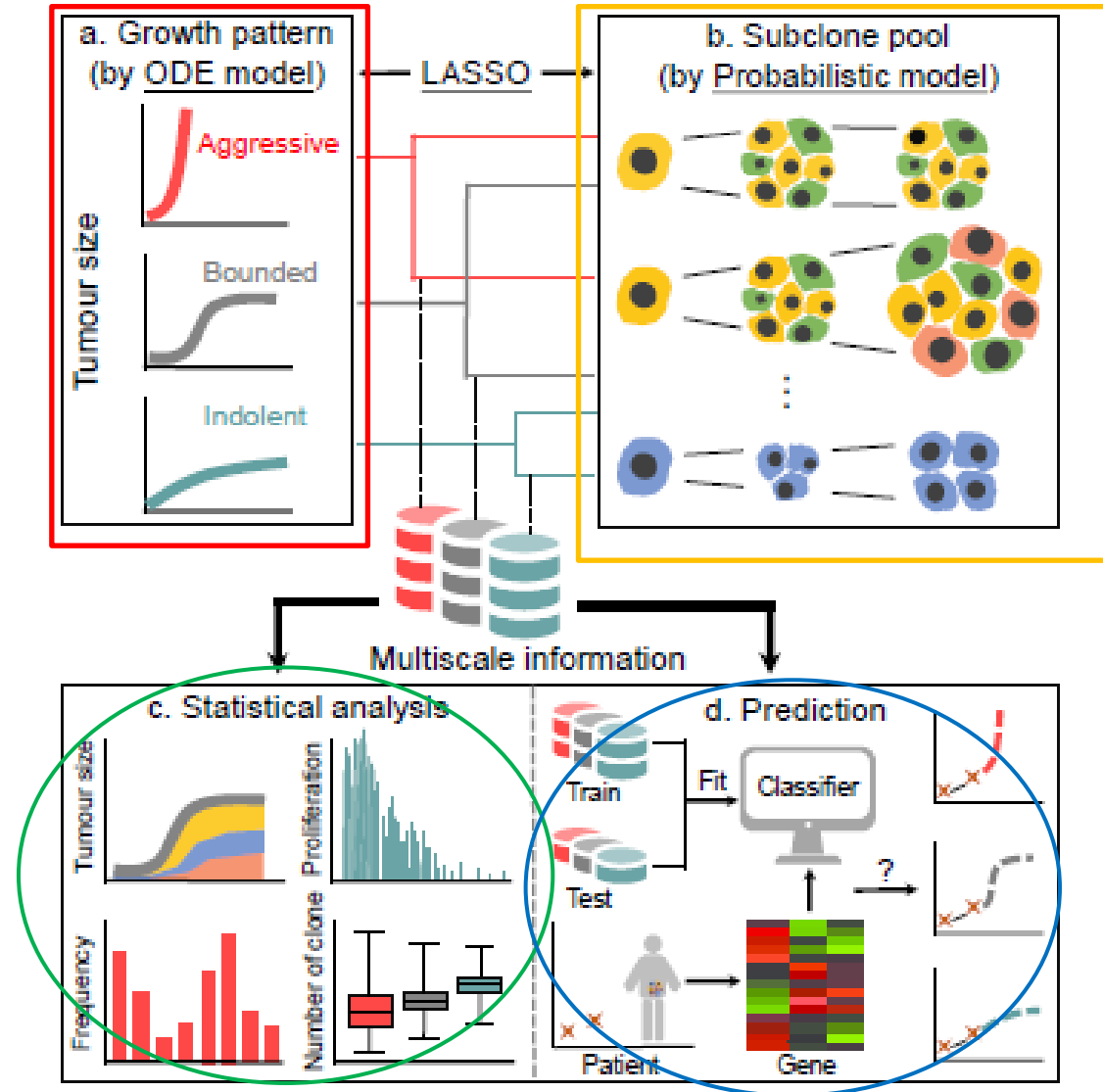# Introduction: Diversity of subclones

Combination 1
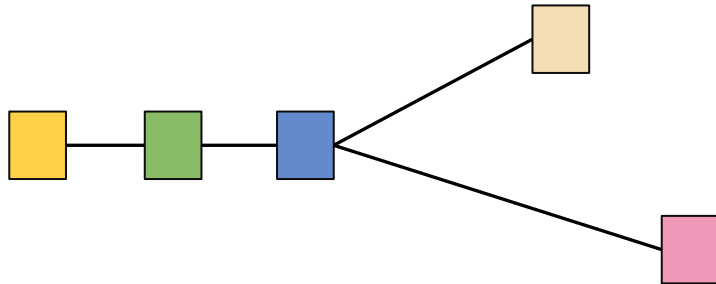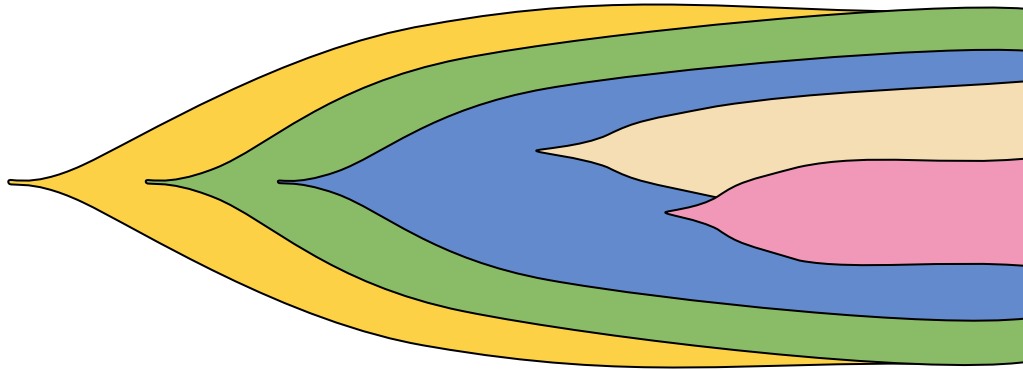
Combination 2

Multiscale computational framework

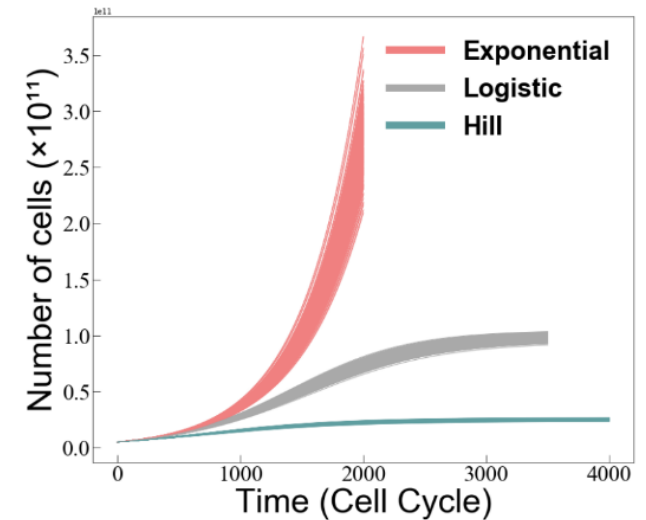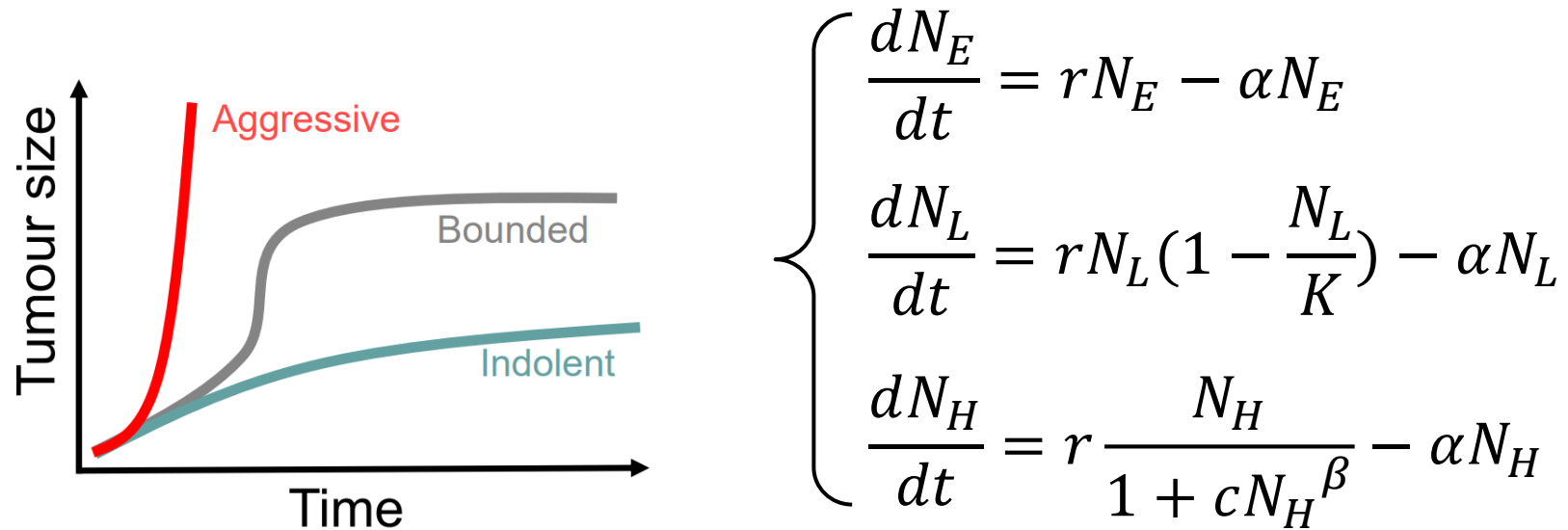

at macro scale
$N(r, \alpha, \beta, \gamma, K)$

at cellular scale
$A_{(T+1) \times m} \cdot x$

Multiscale mathematical modeling

➢ Dynamical model of tumor growth at population level

Three popular ODE (Exponential, Logistic, and Hill models) were used to describe three typical growth patterns: aggressive, bounded, and indolent, respectively..

$$\begin{cases} \dfrac{dN_E}{dt} = rN_E - \alpha N_E \\[2mm] \dfrac{dN_L}{dt} = rN_L(1 - \dfrac{N_L}{K}) - \alpha N_L \\[2mm] \dfrac{dN_H}{dt} = r\dfrac{N_H}{1 + cN_H{}^{\beta}} - \alpha N_H \end{cases}$$
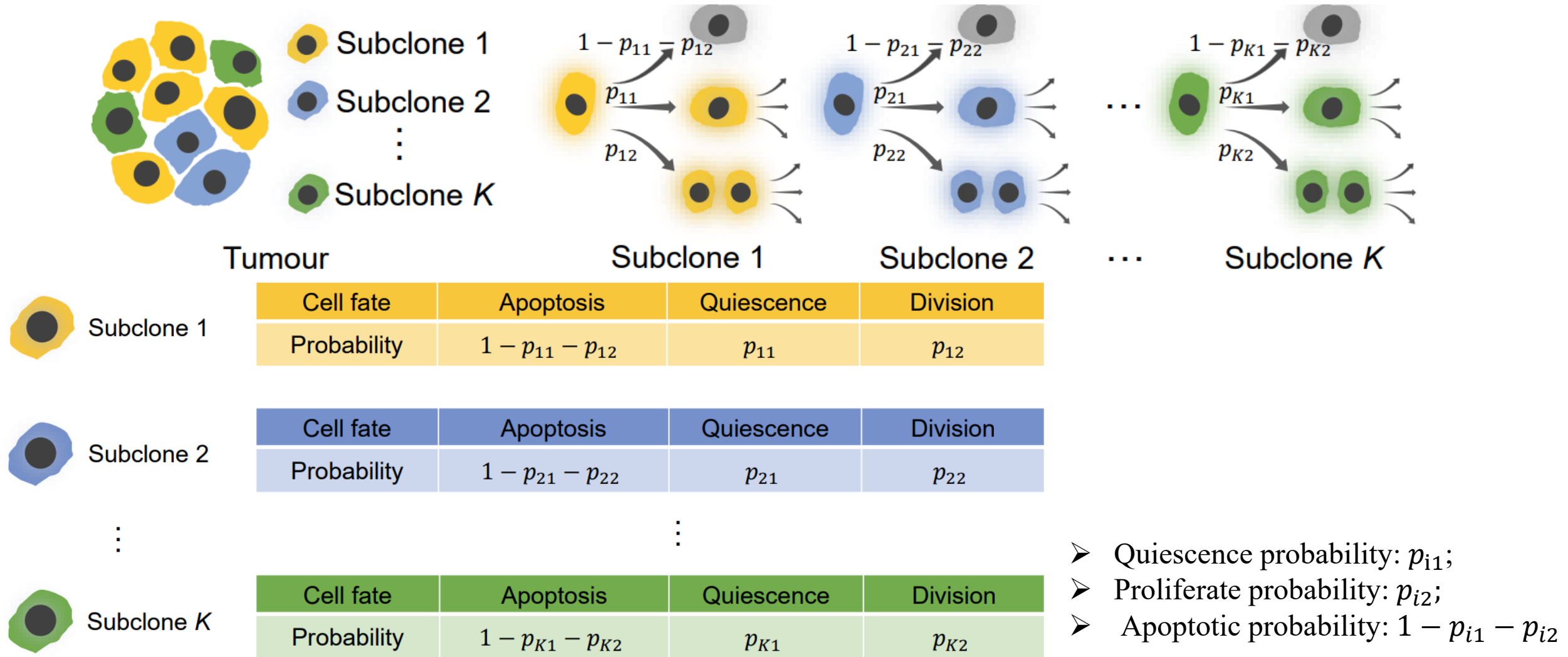
Norton, L. A gompertzian model of human breast cancer growth. *Cancer research* 48, 7067{7071 (1988).
Sachs, R., Hlatky, L. & Hahnfeldt, P. Simple ODE models of tumor growth and anti-angiogenic or radiation treatment. *Mathematical and Computer Modelling* 33, 12970-1305 (2001).
Gerlee, P. The model muddle: in search of tumor growth laws. *Cancer research* 73, 2407-2411 (2013).

Multiscale mathematical modeling

➢ Probabilistic model at cellular level

Cell fate	Apoptosis	Quiescence	Division
Probability	$1 - p_{11} - p_{12}$	p_{11}	p_{12}

Cell fate	Apoptosis	Quiescence	Division
Probability	$1 - p_{21} - p_{22}$	p_{21}	p_{22}

Cell fate	Apoptosis	Quiescence	Division
Probability	$1 - p_{K1} - p_{K2}$	p_{K1}	p_{K2}

➢ Quiescence probability: p_{i1};
➢ Proliferate probability: p_{i2};
➢ Apoptotic probability: $1 - p_{i1} - p_{i2}$

Multiscale mathematical modeling

➢ **Probabilistic model at cellular level**

We constructed a subclone pool (represented as matrix A) to store all possible subclones and their evolutionary trajectories

$$A = \begin{pmatrix} 1 & & 1 \\ R(0, D_{0,1}) & & R(2, D_{2,1}) \\ \vdots & \ddots & \vdots \\ (R(E^{(0)}, D^{(0)}))^{T-t_0}(R(0, D_{0,1}))^{t_0} & \cdots & (R(E^{(T-1)}, D^{(T-1)}))^{T-t_{T-1}}(R(2, D_{2,1}))^{t_{T-1}} \end{pmatrix}_{(T+1) \times m}$$
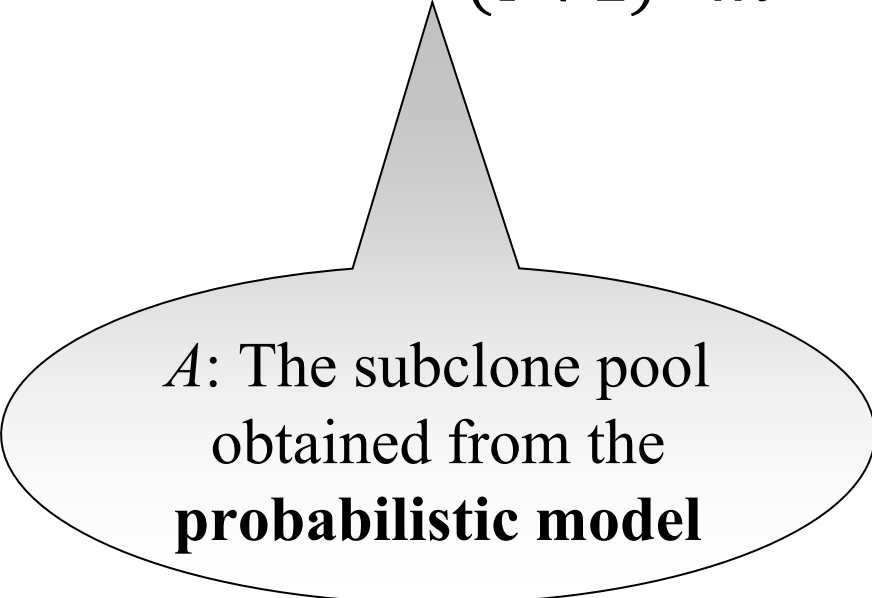
For each column, we assume that there is only one cell at the beginning, row t of this column tells us how many cells that one cell became over t cell cycles.
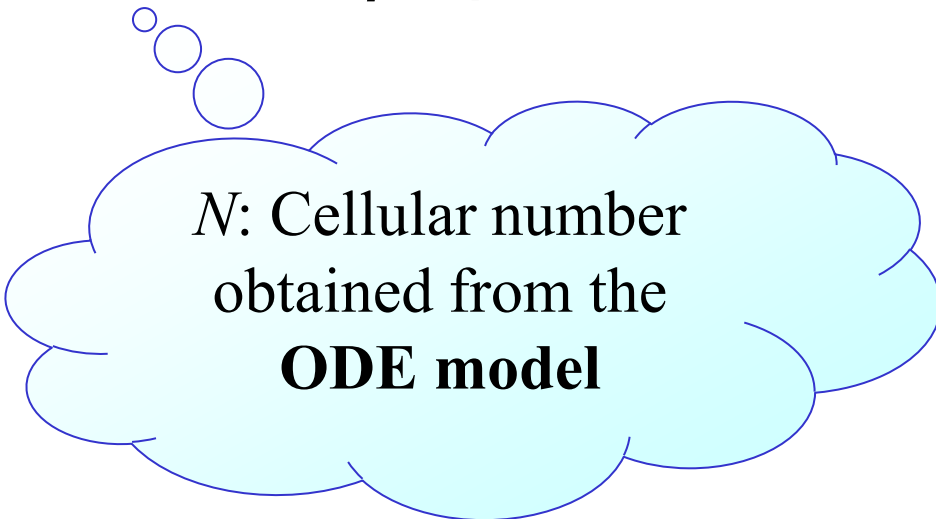
Bridging between different scales

➢ Linking population level (inter-patient heterogeneity) with cellular level (intra-tumor heterogeneity)

$$A_{(T+1) \times m} \cdot x = N(r, \alpha, \beta, \gamma, K)$$

A: The subclone pool obtained from the **probabilistic model**

N: Cellular number obtained from the **ODE model**

The solution x represents the number and combination of different subclones and evolutions at the initial time.

Bridging between different scales

➤ **Non-negative LASSO is used to obtain x**

$$x = arg\min_{x_i \geq 0} \|Ax - N\|_2^2 + \lambda\|x\|_1$$

$$x = \begin{pmatrix} x_{10} \\ \vdots \\ x_{m0} \end{pmatrix}$$ ⟶ Calculating the subclone composition at different time
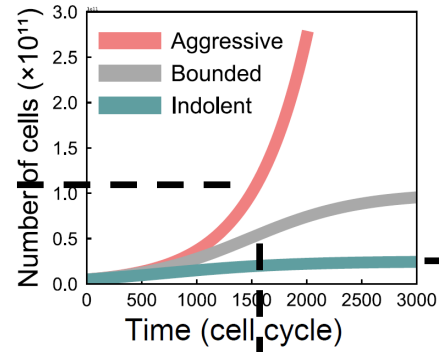
Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58, 267-288 (1996).

Results： Delineating cancer evolutionary trajectory

Compared to a single trajectory describing tumor size in population level, this multiscale model allows us to simulate and analyze dynamic changes in subclone composition.
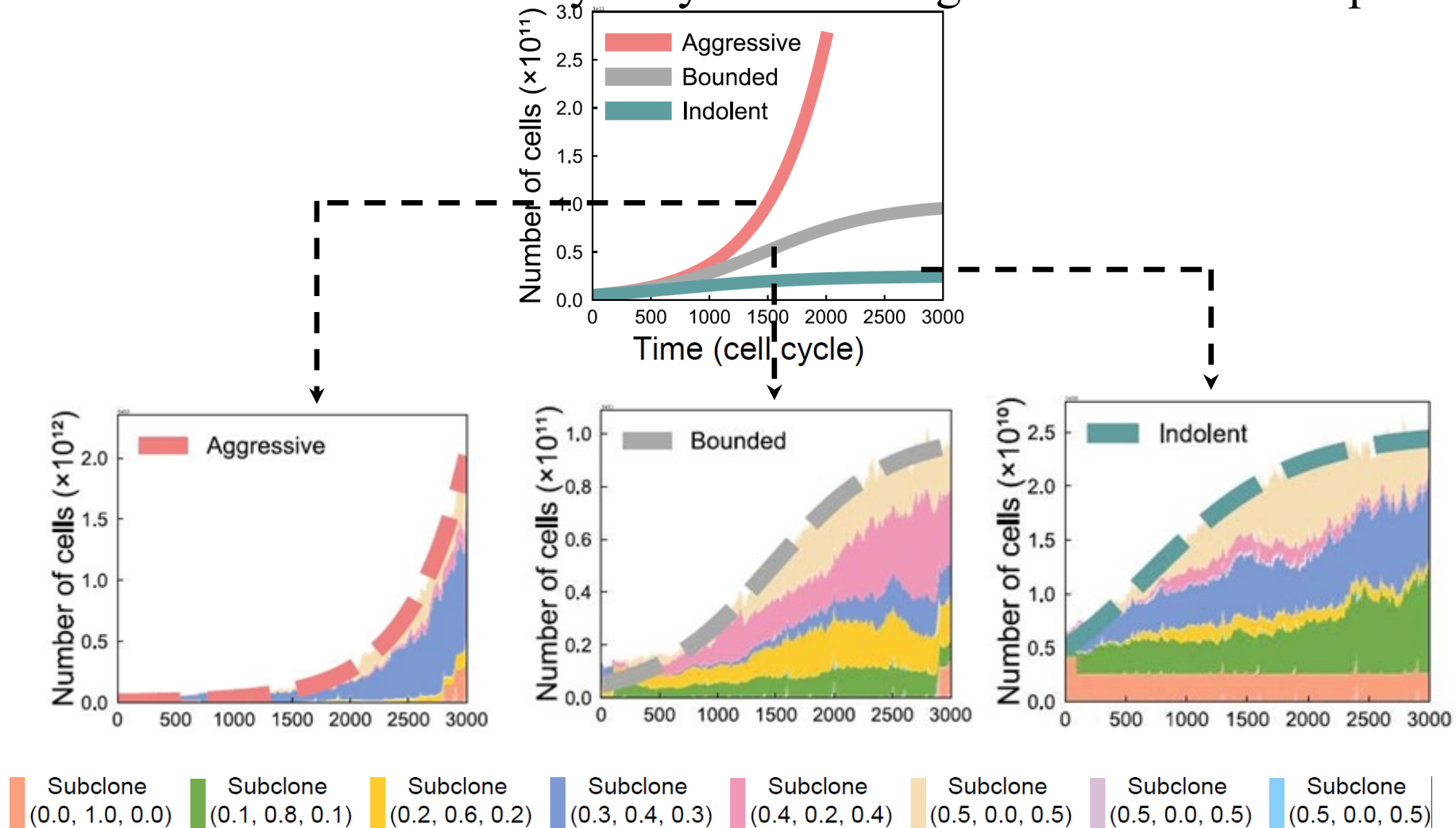
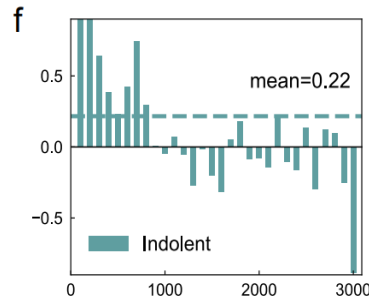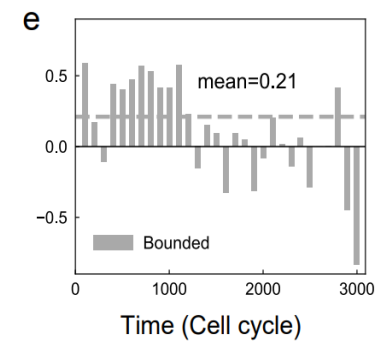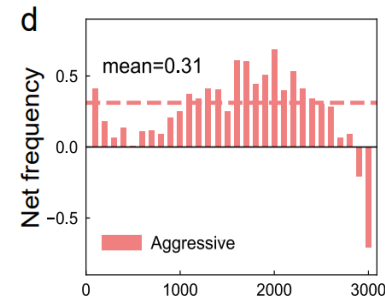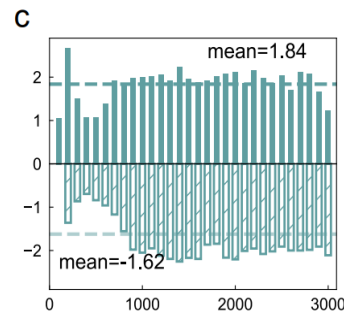# Results: Landscape of subclonal dynamics provide details of cancer evolution

Number of cells ($\times 10^{11}$) vs Time (cell cycle)
- Aggressive
- Bounded
- Indolent

a. Aggressive ($\times 10^{11}$)

Evolution

b. Bounded ($\times 10^{10}$)

Aggressive subclone

c. Indolent ($\times 10^{10}$)

Indolent tumors actually have the most intense internal competition!

a — Frequency; mean=0.97; mean=-0.66
b — mean=1.56; mean=-1.35
c — mean=1.84; mean=-1.62
d — Net frequency; mean=0.31; Aggressive
e — mean=0.21; Bounded
f — mean=0.22; Indolent

Time (Cell cycle)
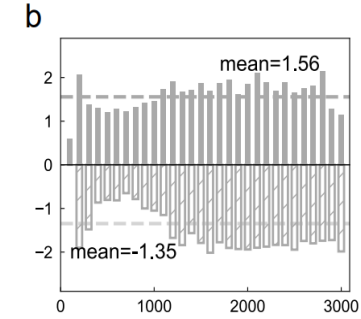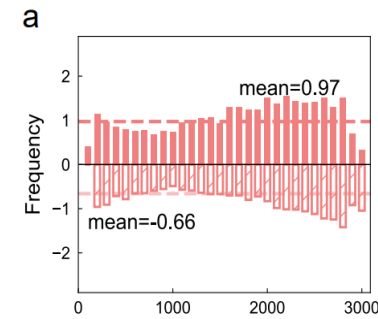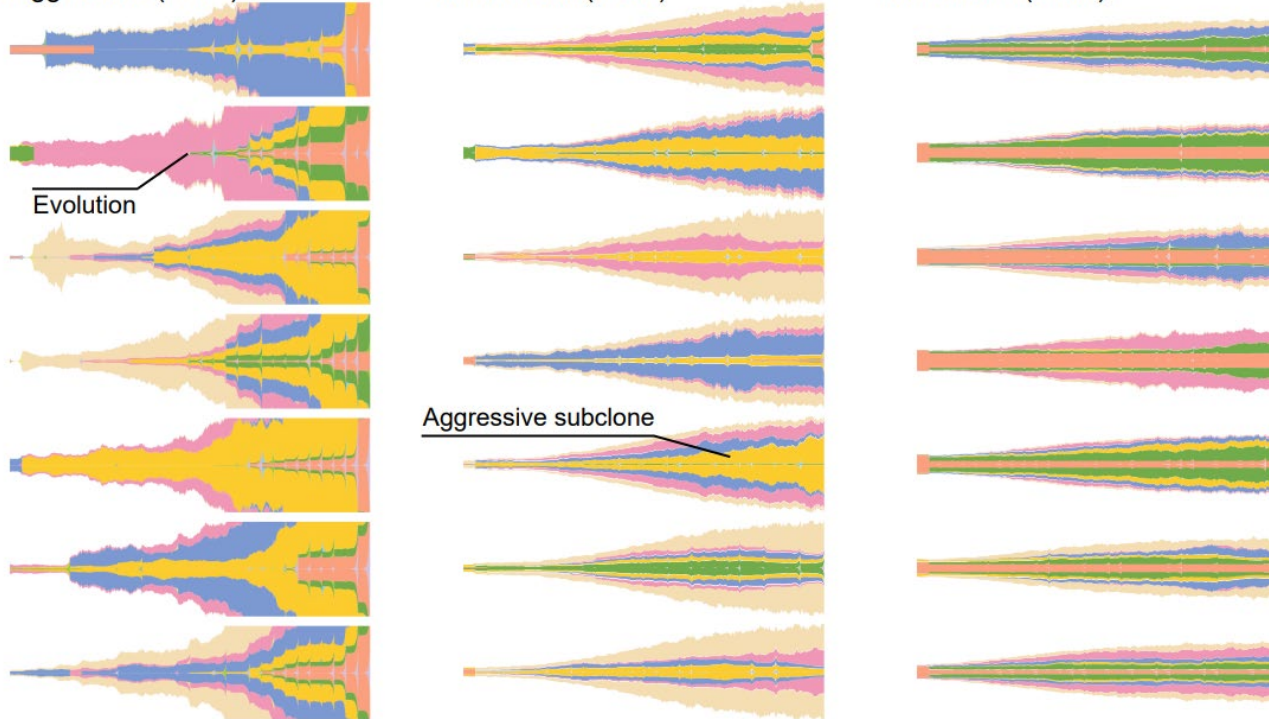
# Real application of the proposed framework

➤ Application to chronic lymphocytic leukaemia (CLL) patients (whole-exome sequencing (WES))

➤ Application to triple-negative breast cancer (TNBC) patients (Single cell RNA-seq data)
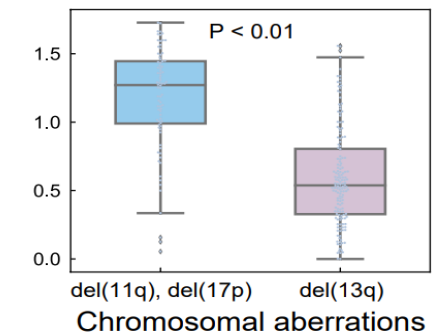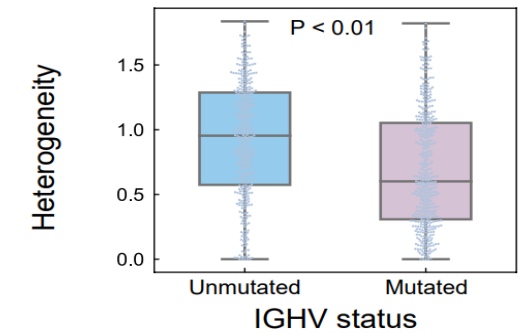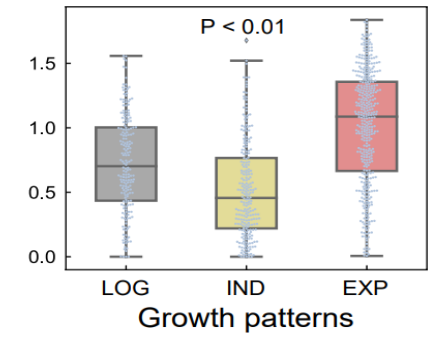
Data for CLL patients

21 CLL patients:

- time-serial white blood cell (WBC) counts (population level)
- whole-exome sequencing (WES) (molecular level)

Gruber, M. et al. Growth dynamics in naturally progressing chronic lymphocytic leukaemia. *Nature* 570, 474-479 (2019).
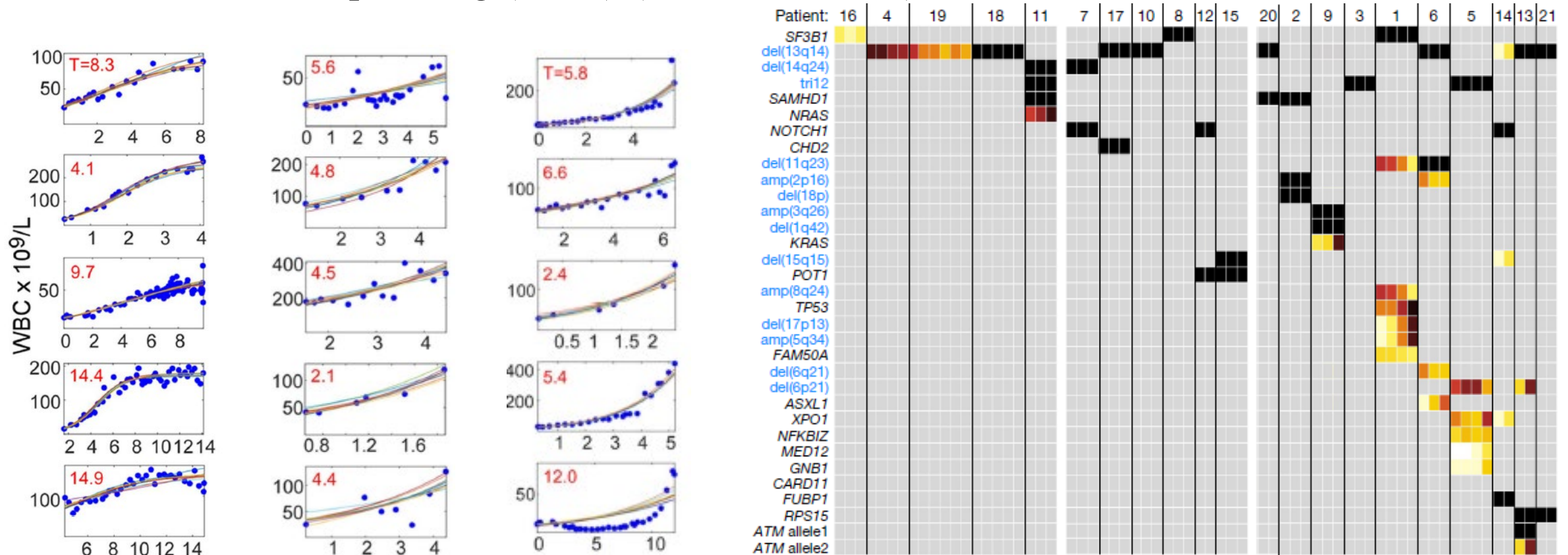
Predicting evolutionary trajectories of CLL patients
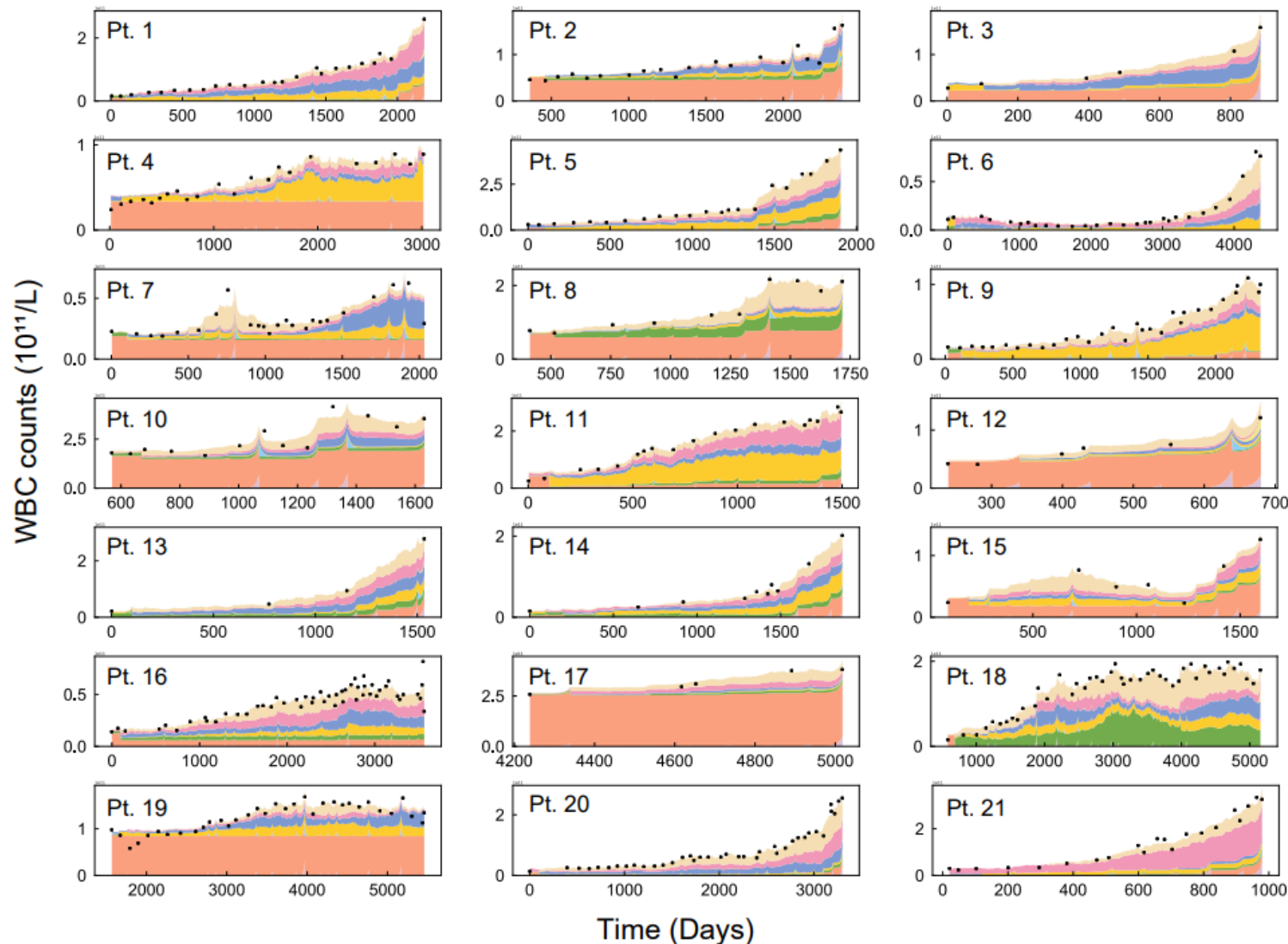
Hallek M, Cheson B D, Catovsky D, et al.. *Blood, The Journal of the American Society of Hematology*, 2018, 131(25): 2745-2760.

Döhner H, Stilgenbauer S, Benner A, et al.. *New England Journal of Medicine*, 2000, 343(26): 1910-1916.

Data for breast cancer patients

Data：triple-negative breast cancer (TNBC)

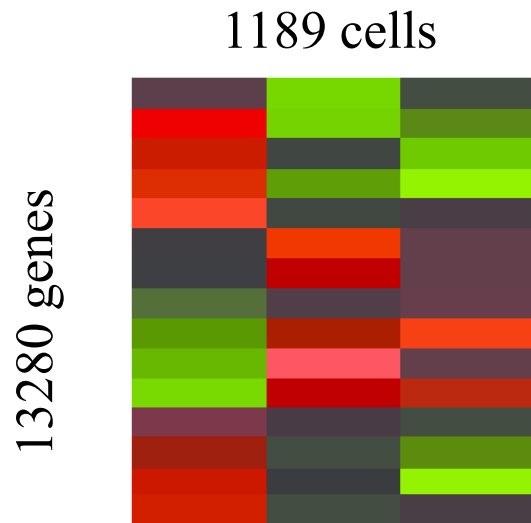- Tumor size (population level)

- scRNA-seq: GSE118390 (molecular level)

Patient ID	39	58	81	84	89	126
Tumour size (cm)	9.5	1.5	2.3	2.3	1.5	1.7

1189 cells

13280 genes

Computing "Gene Scoring Analysis" for connecting molecular with cellular.

The probability of death, resting and proliferation could be estimated by

- Apoptosis pathway: 136 genes

- Proliferation pathway: 97 genes (G2, M, S of cell cycle)

Karaayvaz, M. et al. *Nature communications* 9, 1-10, 2018.
Andreatta M, Carmona S J. Comput Struct Biotechnol J, 19:3796-3798, 2021.

Application to TNBC patients

➤ Using scRNA-seq data of TNBC (triple-negative breast cancer) patients to predict the patients' growth pattern
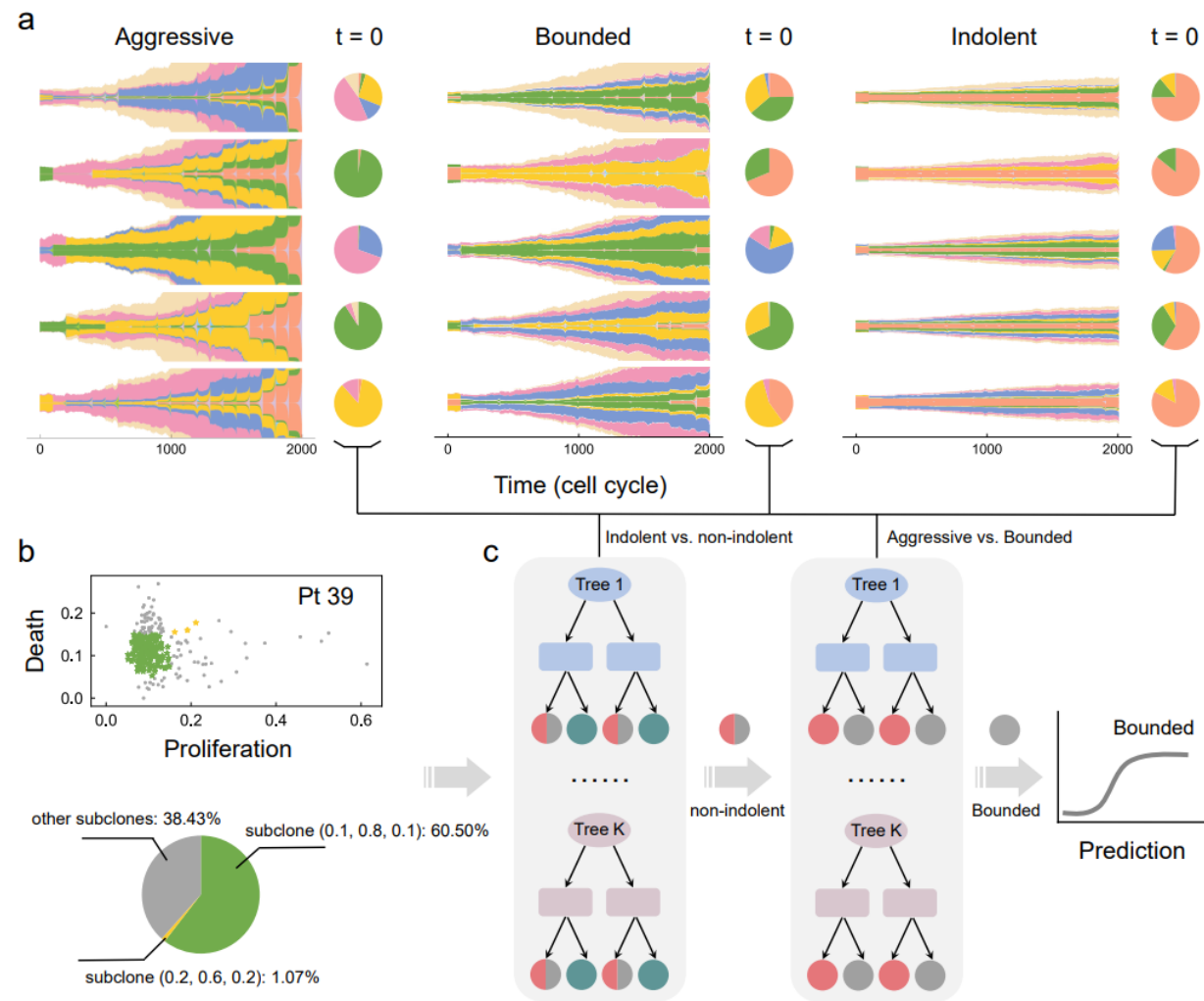
| Patient ID	Ki-67 (IHC Score)	BCL2 (TPM)	Prediction
39	35%	155.51	Bounded
58	<10%	60.31	Aggressive
81	>50%	95	Bounded
84	50%	121.25	Bounded
89	15%	243.88	Aggressive
126	65%	43.33	Aggressive

Zhihao Yao et al., *Journal of Theoretical Biology*, 2024

Karaayvaz M et al. ***Nature communications***, 2018, 9(1): 1-10.

Urruticoechea A et al., ***Journal of clinical oncology***, 2005, 23(28): 7212-7220.

Eom Y Het al. ***Journal of breast cancer***, 2016, 19(3): 252-260.

25

■ Summary I

Aggressive

Bounded

Indolent

t = 0

scRNA-seq   Composition

Bounded
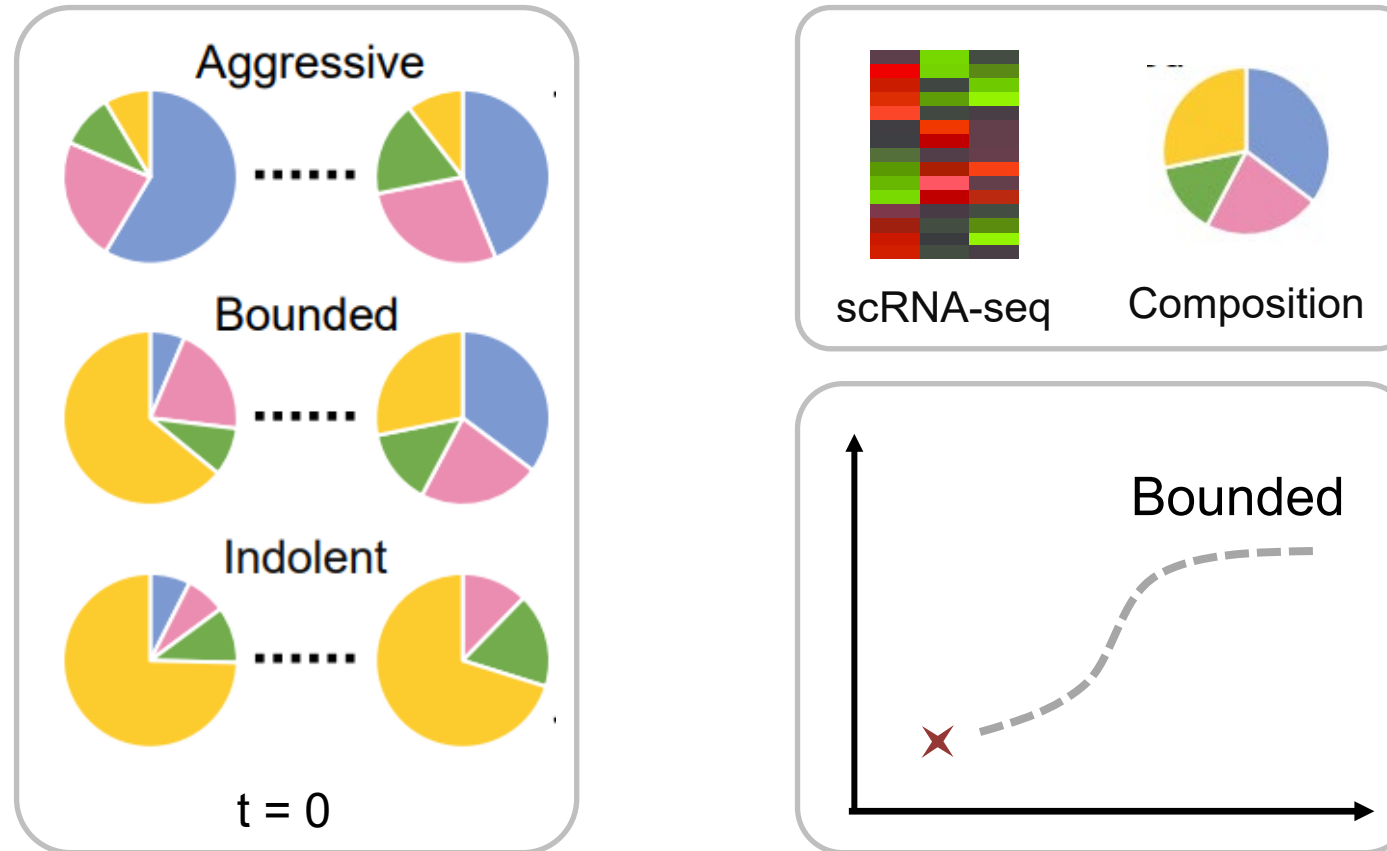
➢ **We identified the prognostic value of subclonal proportions at initial timepoints, enabling prediction of long-term growth patterns using short-term data.**

Journal of Theoretical Biology 582 ,111743, 2024

Work II

➢ **Linking Immune Patterns to Prognosis via Tumor Modeling and Spatial Transcriptomic Data**

To investigate and quantify the relationship between the Turing pattern in the model, the immune pattern in the tumor ecosystem, and the spatial pattern in the spatial data
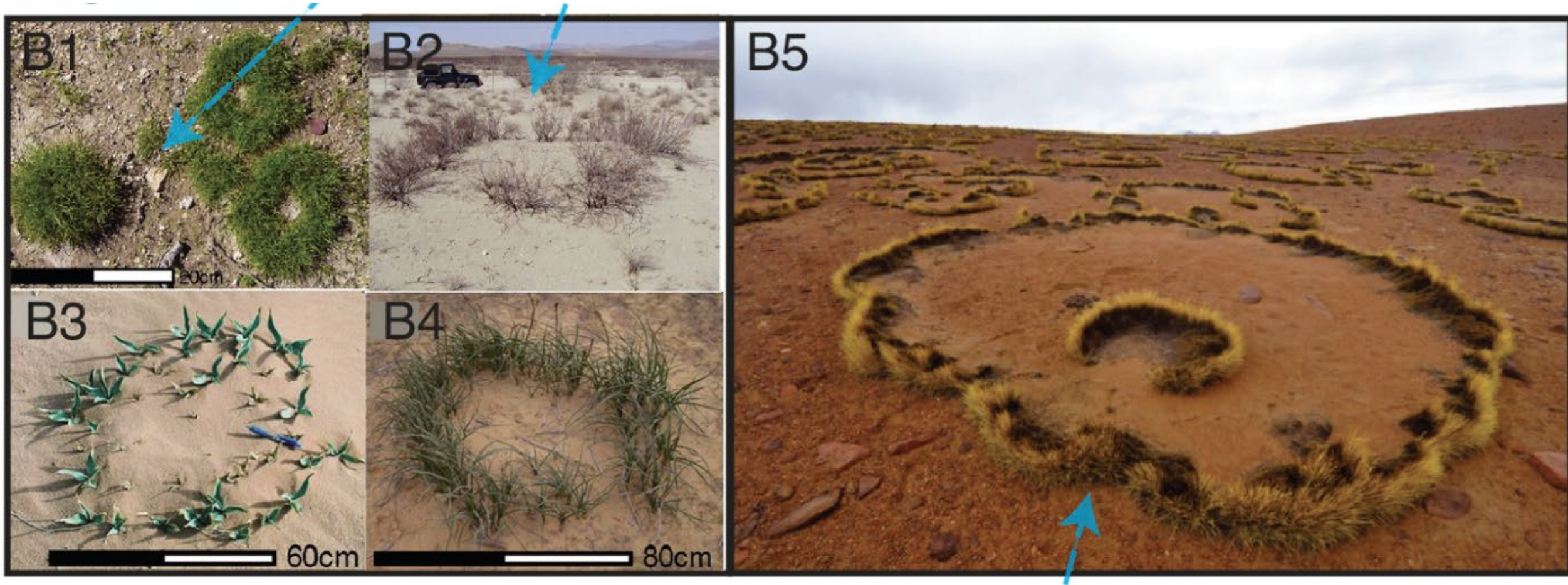
Turing Pattern in Animals

Animals sport a variety of patterns

How do animals get their spots and stripes? A Turing mechanism holds clues to patterns, Knowable Magazine, published June 1, 2024
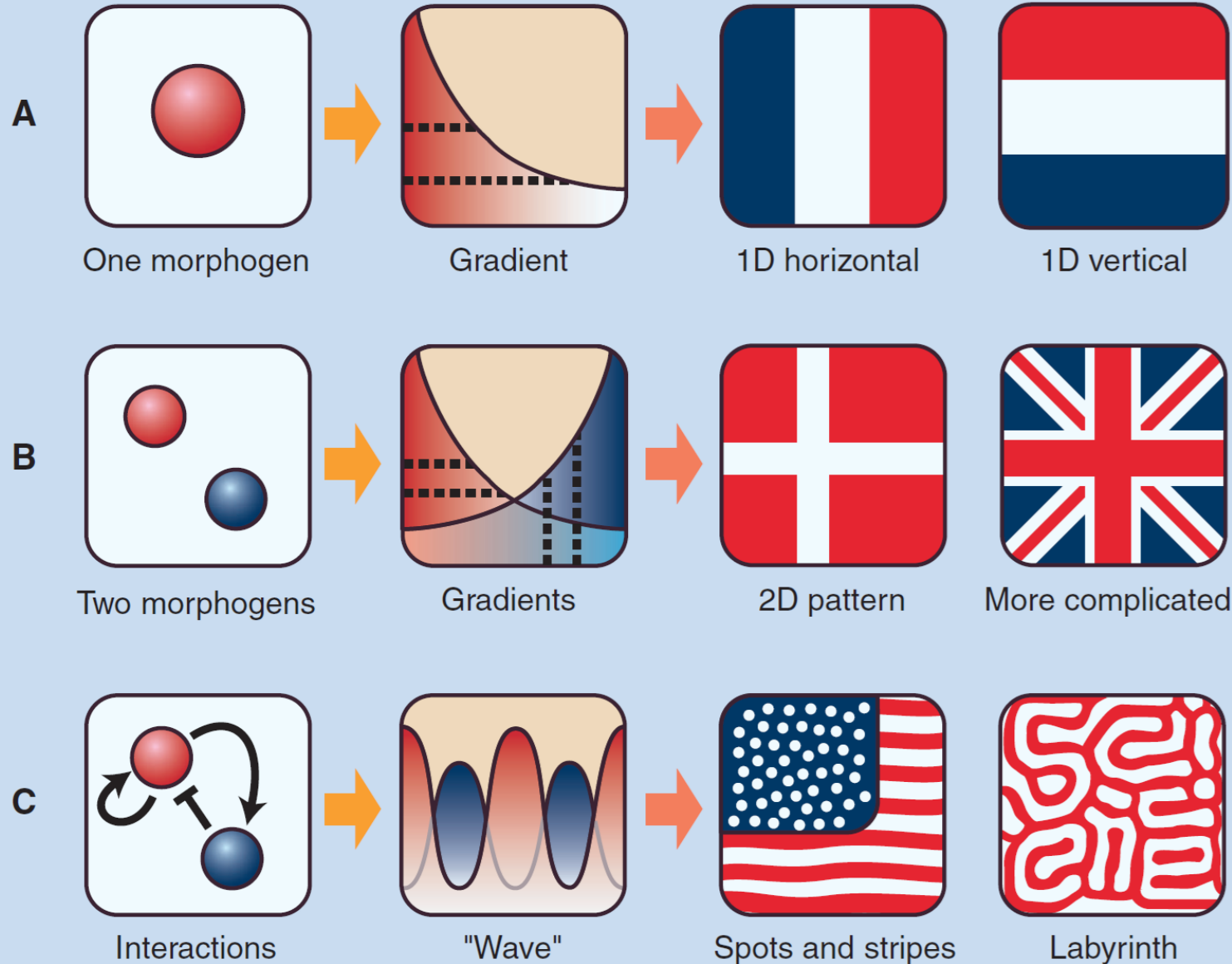
The typical transient Turing patterns of vegetation ecosystems

Zhenpeng Ge, PNAS, 120 (42) e2306514120, 2023

Turing patterns in biological systems

A
One morphogen → Gradient → 1D horizontal / 1D vertical

B
Two morphogens → Gradients → 2D pattern / More complicated

C
Interactions → "Wave" → Spots and stripes / Labyrinth

➢ Morphogen gradient and Turing pattern

Shigeru Kondo and Takashi Miura.
Science 329: 1616-1620, 2010.

Understanding the tumor–immune ecosystem

➢ **What are the patterns in tumor–immune ecosystem?**
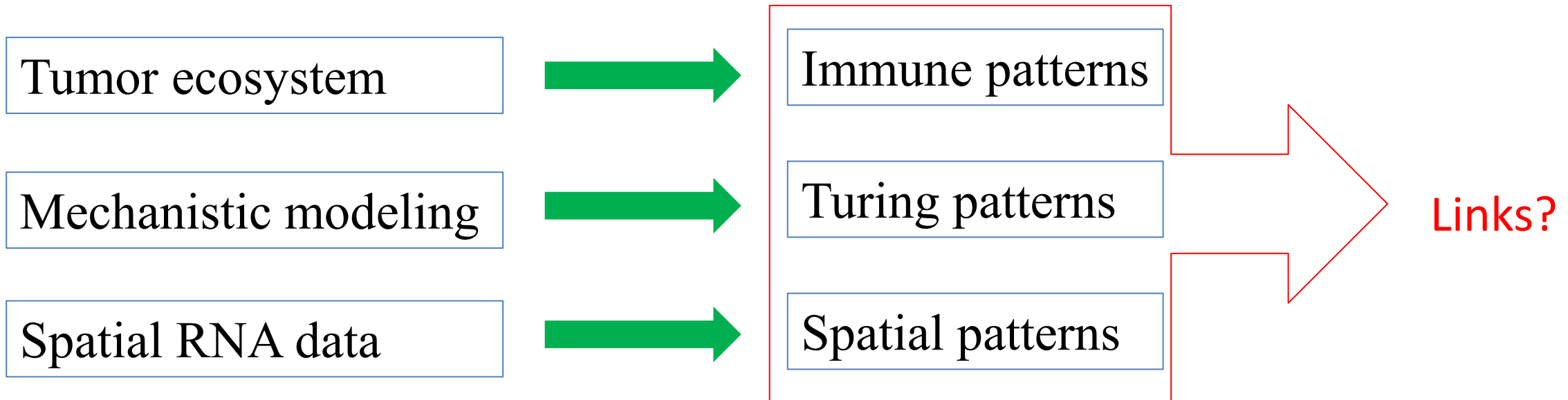
Tumor ecosystem	➡	Immune patterns
Mechanistic modeling	➡	Turing patterns
Spatial RNA data	➡	Spatial patterns

Links?

To investigate and quantify the relationship between the Turing pattern in the model, the immune pattern in the tumor ecosystem, and the spatial pattern in the spatial data

A reaction-diffusion tumor-immune model
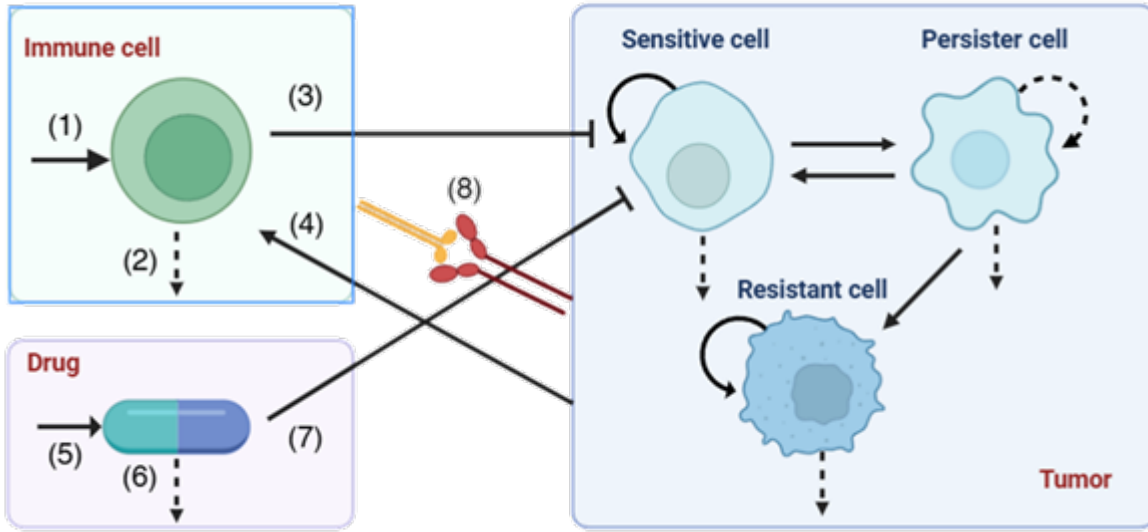
(1) Immune cell recruitment

(2) Immune cell apoptosis

(3) Immune cell clearance and deactivation

(4) Tumor-induced immune activation

(5) Drug delivery

(6) Drug degradation

(7) Drug-induced cell destruction

(8) PD-L1 binding to PD-1 receptor

$$\frac{\partial \tilde{S}}{\partial \tau} = \underbrace{D_S \Delta \tilde{S}}_{diffusion} + \underbrace{r_S \frac{\tilde{S}^n}{M^n + \tilde{S}^n}}_{proliferation} - \underbrace{q_{sp} \frac{\tilde{S}\tilde{P}}{c + \tilde{S}} + q_{ps}\tilde{P}\tilde{S}}_{\substack{transformation \\ between \ \tilde{S} \ and \ \tilde{P}}} - \underbrace{q_{es}\tilde{E}\tilde{S}}_{(3)} - \underbrace{q_{gs}\frac{\tilde{S}\tilde{G}}{\tilde{J_0} + \tilde{G}}}_{(7)} - \underbrace{u_1\tilde{S}}_{apoptosis},$$

$$\frac{\partial \tilde{P}}{\partial \tau} = \underbrace{D_P \Delta \tilde{P}}_{diffusion} + \underbrace{r_P \tilde{P}(1 - \frac{\tilde{N}}{K_1})}_{proliferation} + \underbrace{q_{sp} \frac{\tilde{S}\tilde{P}}{c + \tilde{S}} - q_{ps}\tilde{P}\tilde{S}}_{\substack{transformation \\ between \ \tilde{S} \ and \ \tilde{P}}} - \underbrace{q_{pr}\tilde{P}}_{\substack{transformation \\ from \ \tilde{P} \ to \ \tilde{R}}} - \underbrace{u_2\tilde{P}}_{apoptosis},$$

$$\frac{\partial \tilde{R}}{\partial \tau} = \underbrace{D_R \Delta \tilde{R}}_{diffusion} + \underbrace{r_R \tilde{R}(1 - \frac{\tilde{N}}{K_2})}_{proliferation} + \underbrace{q_{pr}\tilde{P}}_{\substack{transformation \\ from \ \tilde{P} \ to \ \tilde{R}}} - \underbrace{u_3\tilde{R}}_{apoptosis},$$

$$\frac{\partial \tilde{E}}{\partial \tau} = \underbrace{D_E \Delta \tilde{E}}_{diffusion} + (\underbrace{\Gamma}_{(1)} + \underbrace{q_{ne}\tilde{N}\tilde{E}}_{(4)})\underbrace{\frac{1}{1 + \varrho\tilde{E}(\tilde{E} + \epsilon_n\tilde{N})}}_{(8)} - \underbrace{q_{se}\tilde{S}\tilde{E}}_{(3)} - \underbrace{u_4\tilde{E}}_{(2)},$$

$$\frac{\partial \tilde{G}}{\partial \tau} = \underbrace{D_G \Delta \tilde{G}}_{diffusion} + \underbrace{\Lambda}_{(5)} - \underbrace{u_5\tilde{G}}_{(6)}.$$

Nondimensionalized model

$$S = \frac{\widetilde{S}}{K_1}, \ P = \frac{\widetilde{P}}{K_1}, \ R = \frac{\widetilde{R}}{K_1}, \ W = \frac{r_P \widetilde{W}}{\Gamma}, \ G = \frac{r_P \widetilde{G}}{\Lambda},$$

time scale with $t = r_P \tau$, and space scale with $x = \widetilde{x}/L, \ y = \widetilde{y}/L,$

$$
\begin{cases}
\dfrac{\partial S}{\partial t} = D_s \Delta S + r_s \dfrac{S^n}{1 + mS^n} - k_{sp} \dfrac{SP}{1 + \alpha S} + k_{ps} PS - k_{es} ES - k_{gs} \dfrac{SG}{J_0 + G} - \mu_1 S, \\[2mm]
\dfrac{\partial P}{\partial t} = D_p \Delta P + P(1 - N) + k_{sp} \dfrac{SP}{1 + \alpha S} - k_{ps} PS - k_{pr} P - \mu_2 P, \\[2mm]
\dfrac{\partial R}{\partial t} = D_r \Delta R + r_r R(1 - \kappa N) + k_{pr} P - \mu_3 R, \\[2mm]
\dfrac{\partial E}{\partial t} = D_e \Delta E + (1 + k_{ne} NE) \dfrac{1}{1 + \rho E(E + \varepsilon_n N)} - b_{se} SE - \mu_4 E, \\[2mm]
\dfrac{\partial G}{\partial t} = D_g \Delta G + 1 - \mu_5 G,
\end{cases}
$$

(1)

$$D_s = \frac{D_S}{K_1^2 L^2 r_P}, \ D_p = \frac{D_P}{K_1^2 L^2 r_P}, \ D_r = \frac{D_R}{K_1^2 L^2 r_P},$$

$$D_e = \frac{D_E}{K_1^2 L^2 r_P}, \ D_g = \frac{D_G}{K_1^2 L^2 r_P}, \ r_s = \frac{K_1^{\eta-1} r_S}{\widetilde{M}^\eta r_y},$$

$$m = \frac{K_1}{\widetilde{M}}, \ k_{sp} = \frac{q_{sp} K_1}{c r_P}, \ \alpha = \frac{K_1}{c}, \ k_{ps} = \frac{q_{ps} K_1}{r_P},$$

$$k_{es} = \frac{q_{es} \Gamma}{K_1 r_P^2}, \ k_{gs} = \frac{q_{gs}}{r_P}, \ J_0 = \frac{\widetilde{J_0} r_P}{\Lambda}, \ k_{pr} = \frac{q_{pr}}{r_P},$$

$$\kappa = \frac{K_1}{K_2}, \ r_R = \frac{r_R}{r_P}, \ k_{ne} = \frac{q_{ne} K_1}{r_P}, \ \rho = \frac{\varrho \Gamma^2}{r_P^2},$$

$$\varepsilon_n = \frac{\epsilon_N K_1 r_P}{\Gamma}, \ b_{se} = \frac{b_{SE} K_1 \Gamma}{r_P}, \ \mu_i = \frac{u_i}{r_P} \ (i = 1, \cdots, 5).$$
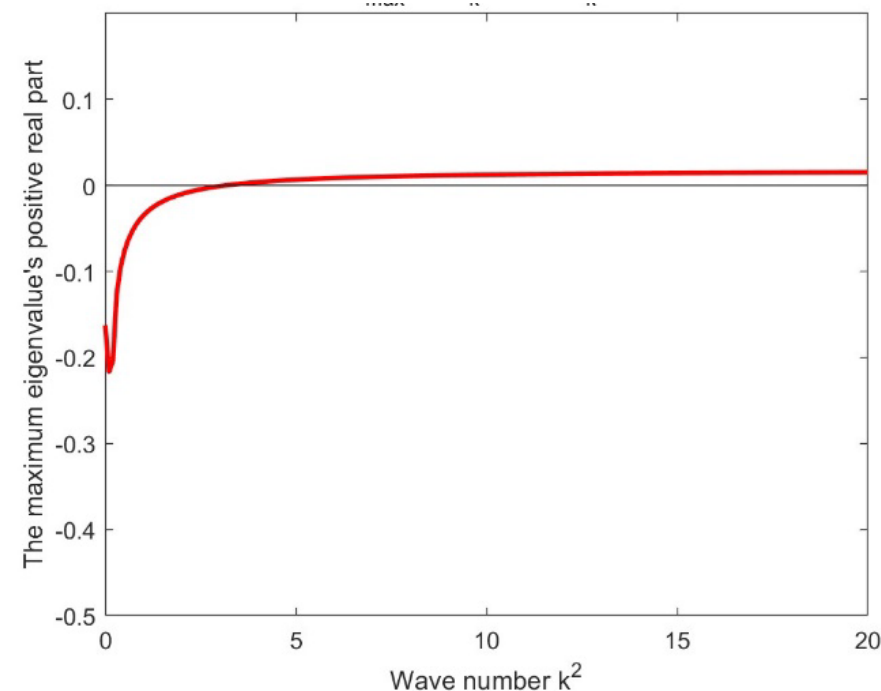
Boundary condition: Homogeneous neumann boundary

$$\frac{\partial S}{\partial \nu} = \frac{\partial P}{\partial \nu} = \frac{\partial R}{\partial \nu} = \frac{\partial E}{\partial \nu} = \frac{\partial G}{\partial \nu} = 0.$$

$$\dot{U} = D\Delta U + JU$$

$$U = \begin{pmatrix} S_k \\ P_k \\ R_k \\ E_k \\ G_k \end{pmatrix} e^{\lambda t + i\mathbf{k}\cdot r} \Longrightarrow \lambda \begin{pmatrix} S_k \\ P_k \\ R_k \\ E_k \\ G_k \end{pmatrix} = J_k \begin{pmatrix} S_k \\ P_k \\ R_k \\ E_k \\ G_k \end{pmatrix}$$

General conditions of Turing bifurcation

1. $eig(J)|_{E1} < 0$

2. Exist a wave number k, $s.t.$ $eig(J_k)|_{E1} > 0$

$$\lambda_{max} = \max_{k}\{Re(eig(J_k))\}$$

Diambra, Luis, et al. ACS synthetic biology 4.2 (2015): 177-186.
J.D. Murray, Mathematical Biology II: Spatial Models and Biomedical Applications, Third Edition, Springer, 2003
Edgardo Villar-Sepúlveda and Alan R. Champneys, Journal of Mathematical Biology (2023) 86:39

34

Transient Turing patterns in the bistable system

A Mathetical model

RDE model $\dot{U} = D\Delta U + JU$

$U = U_k \cdot e^{\lambda t + i\vec{k} \cdot \mathbf{r}}$

Characteristic Eq. $\lambda U_k = J_k U_k$

Sufficient conditions of transient Turing pattern

1) ODE $eig\,(J)\,|_{E1\ and\ E2} < 0$

2) RDE $\exists$ a wave number k, s.t. $eig\,(J\kappa)\,|_{E1\ or\ E2} > 0$

B Cells / Tipping point / Parameter
— Stable OSS
···· Unstable OSS

C Cells / Turing bifurcation / Tipping point / Parameter
— Stable OSS
– – TTP
···· Unstable OSS
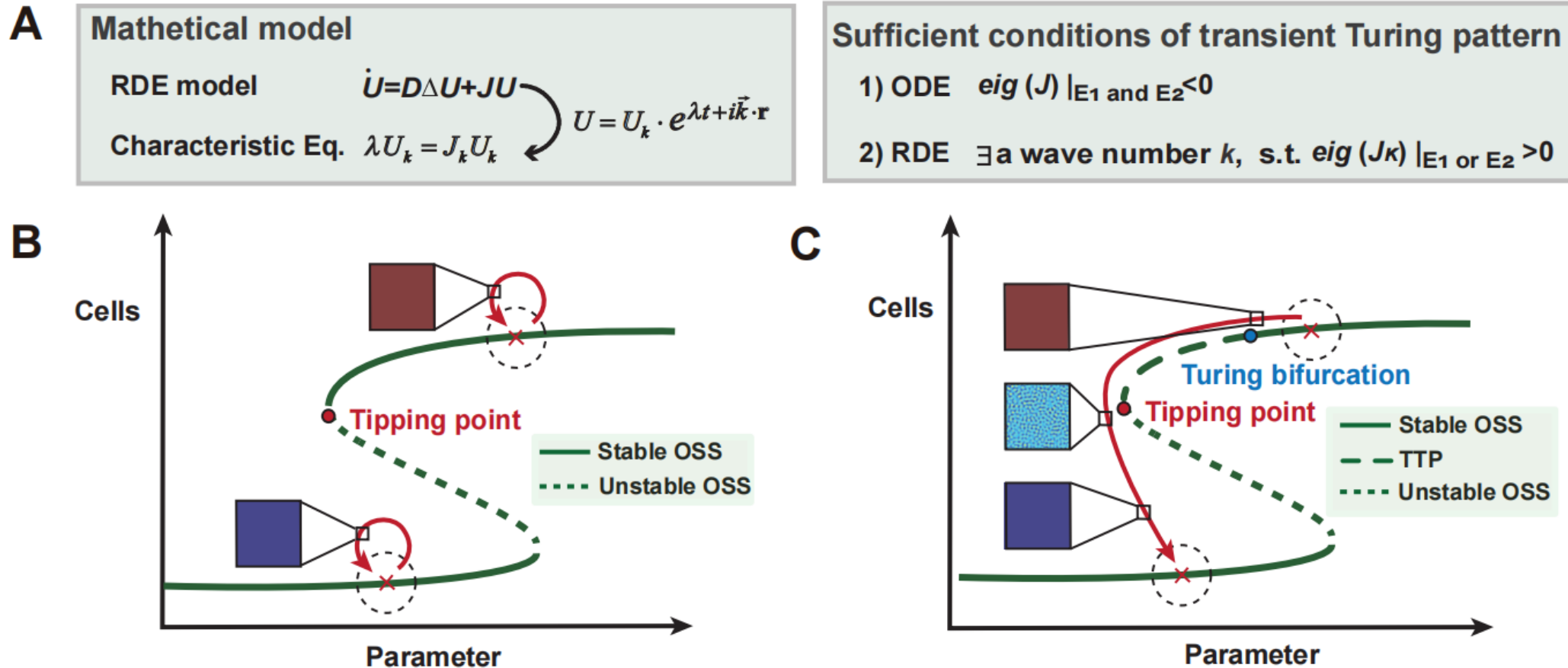
We derived sufficient conditions for transient Turing patterns in the bistable system, extending prior work on 1D RDE models in synthetic biology, where patterns eventually decay to homogeneous steady states.

A Morozov, et al., Long transients in ecology: Theory and applications. *Phys. life reviews* **32**,1–40 (2020).
N Guisoni, L Diambra, Transient turing patterns in a morphogenetic model. *Front. Phys.* **10**, 927152 (2022).

The simulated spatiotemporal dynamics
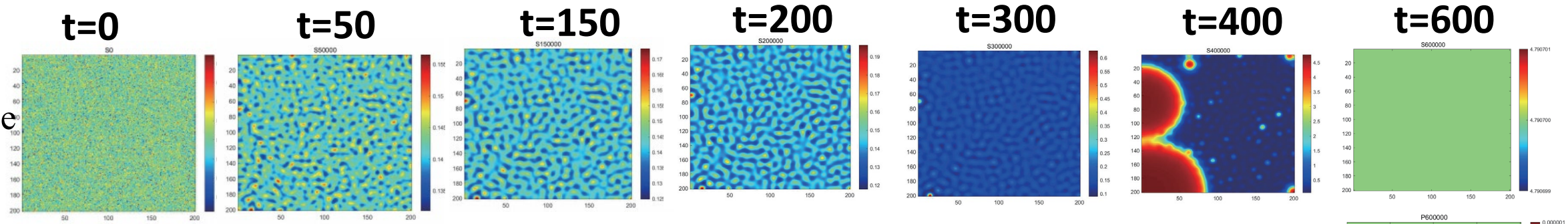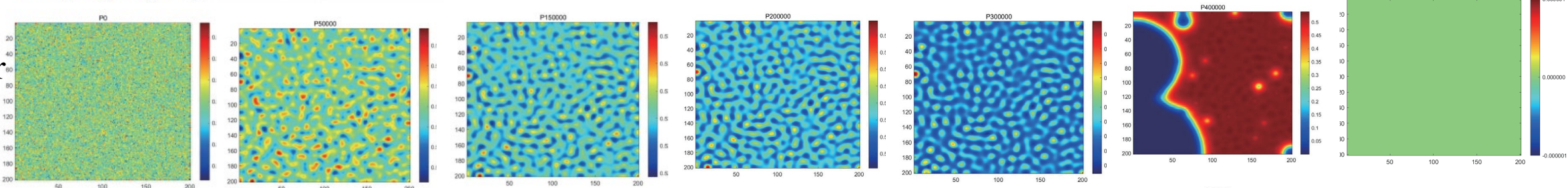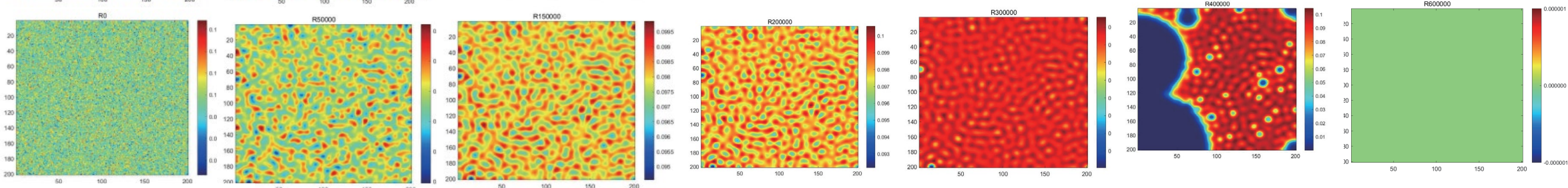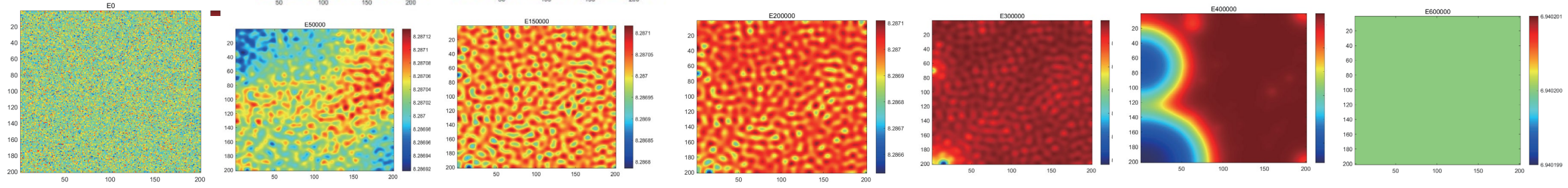
Quantitative index: Immune Dominance Ratio (IDR)

$$\text{IDR}(t) = \frac{\sum_{i,j} \mathbb{1}\left(E_{i,j}(t) > N_{i,j}(t)\right)}{M \times N},$$

$$\widehat{N} = \frac{N - N_{\min}}{N_{max} - N_{\min}}$$

$$\widehat{E} = \frac{E - E_{\min}}{E_{max} - E_{\min}}$$

Legend:
- $k_{gs} = 1.1$
- $k_{gs} = 0.6$
- $k_{gs} = 0.5$
- $k_{gs} = 0.4$
- $k_{gs} = 0.3$
- $k_{gs} = 0.2$
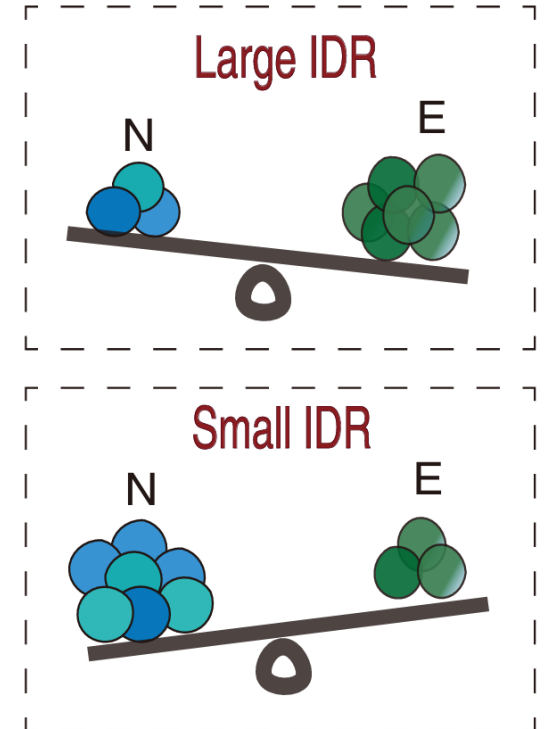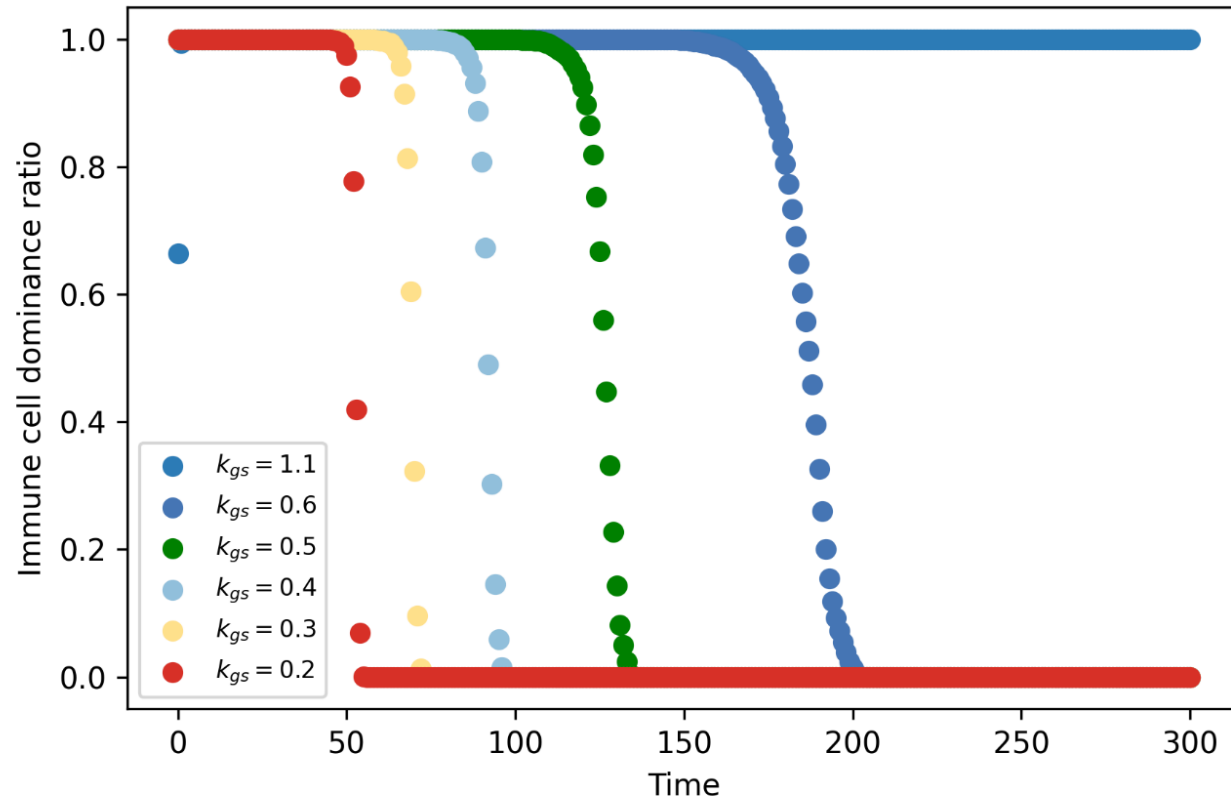
Kgs represents the killing rate of the drug on sensitive cells. The higher the killing rate, the fewer the tumor cells and the longer the immune dominance duration.

Transient Turing pattern V.S. Immune Dominance Ratio (IDR)
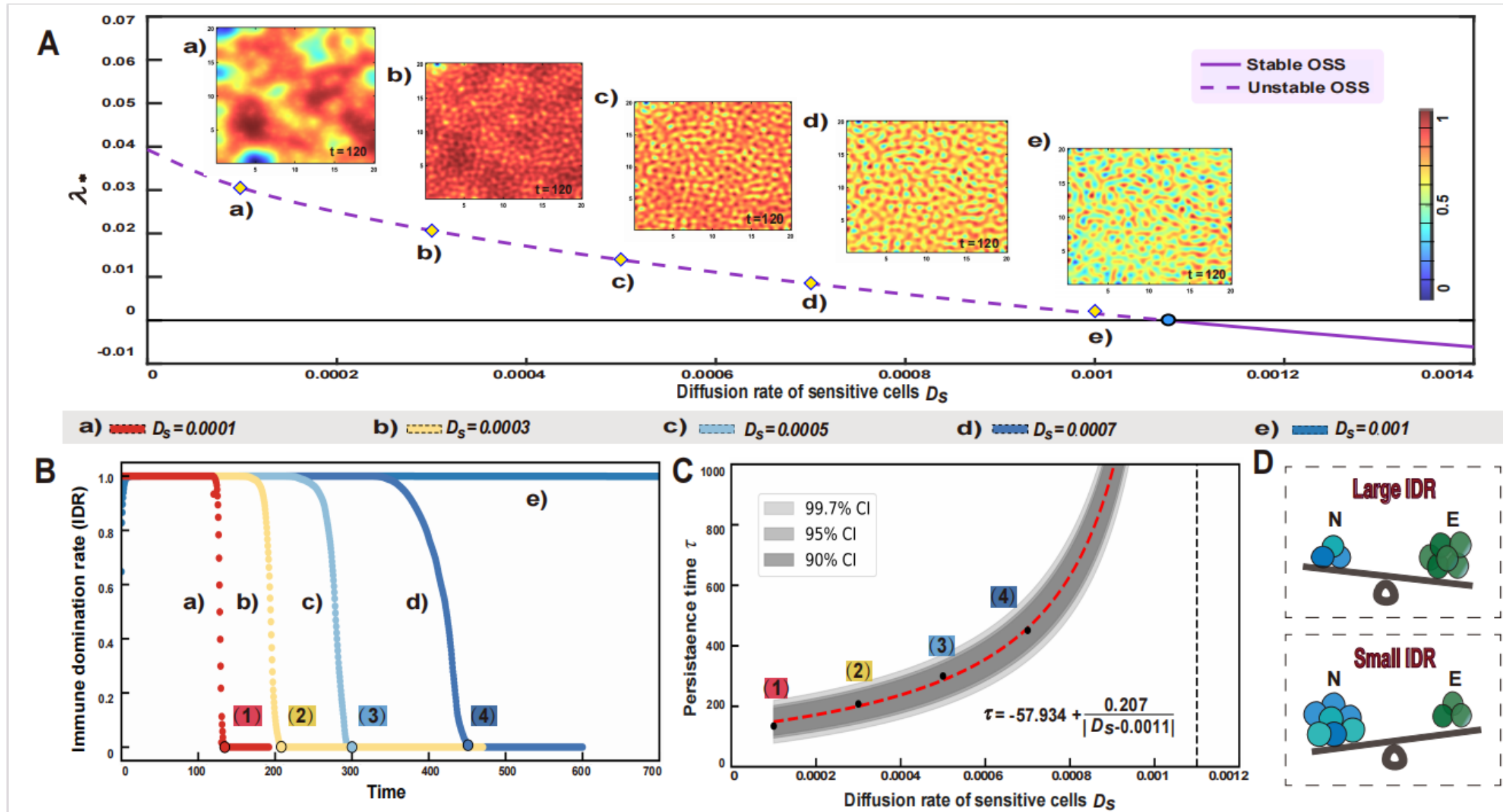
Fig. Temporal Evolution and Immune Dominance Rate (IDR) Analysis of Immune Cells with Varying Diffusion Rates.

Spot pattern V.S. Stripe pattern
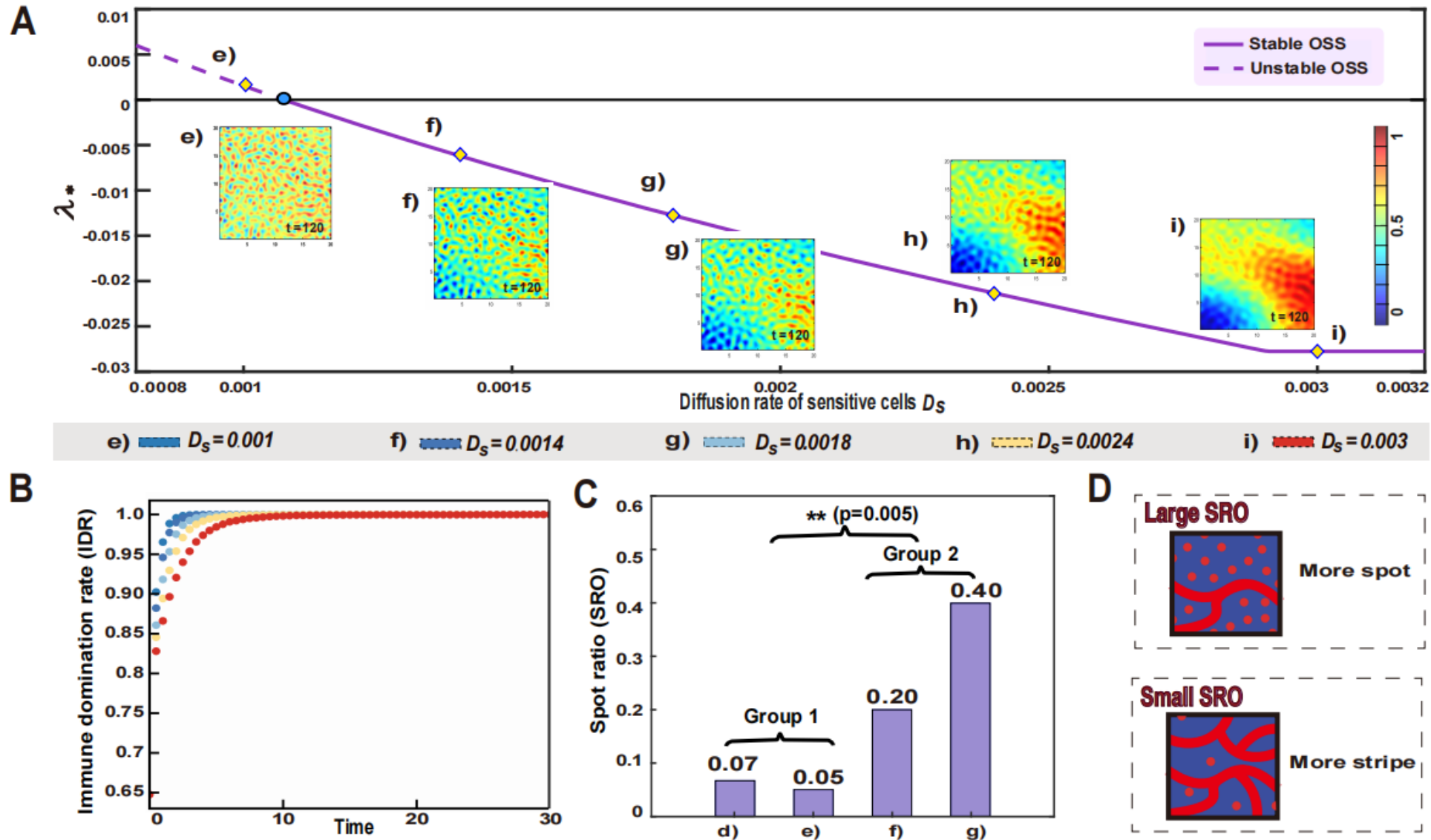
Fig. Structural Transition from Striped to Spotted Patterns near the Turing Instability Region

Application to spatial transcriptomic data of lung cancer

Multi-omic profiling highlights factors associated with resistance to immuno-chemotherapy in non-small cell lung cancer
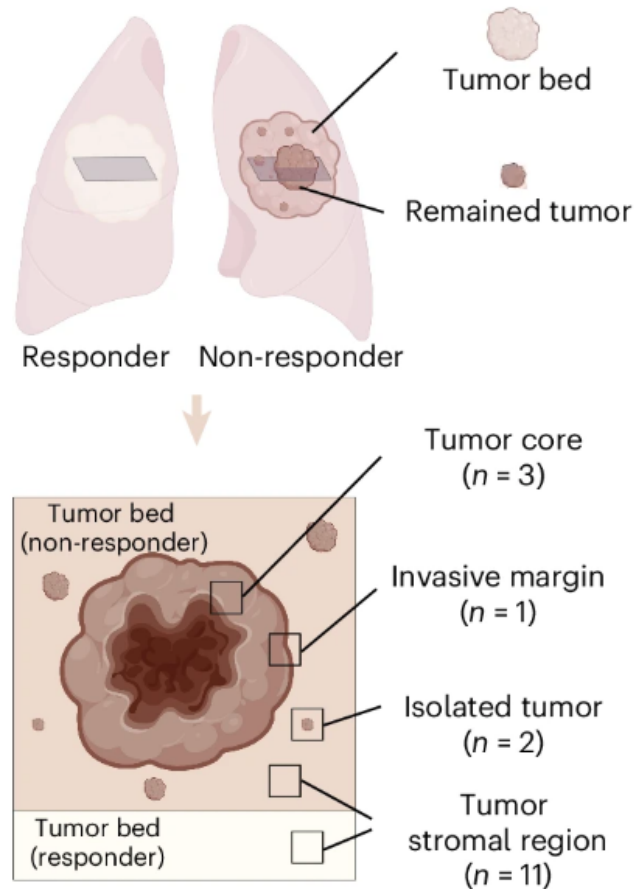
Yilv Yan, et al. *Nature Genetics* 57: 126-139, 2025.
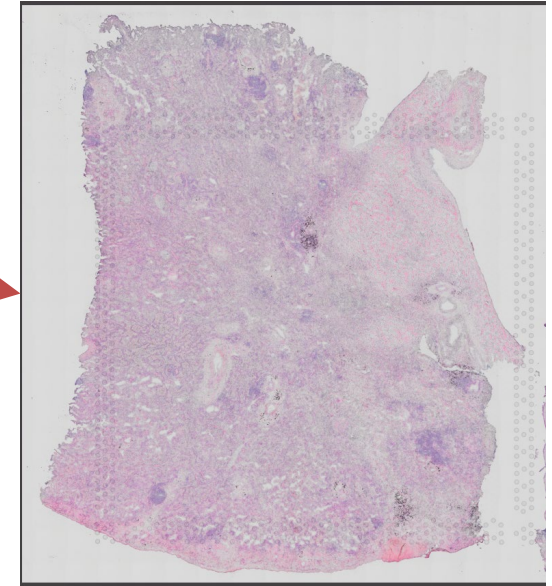
Clinical information for NSCLC patients
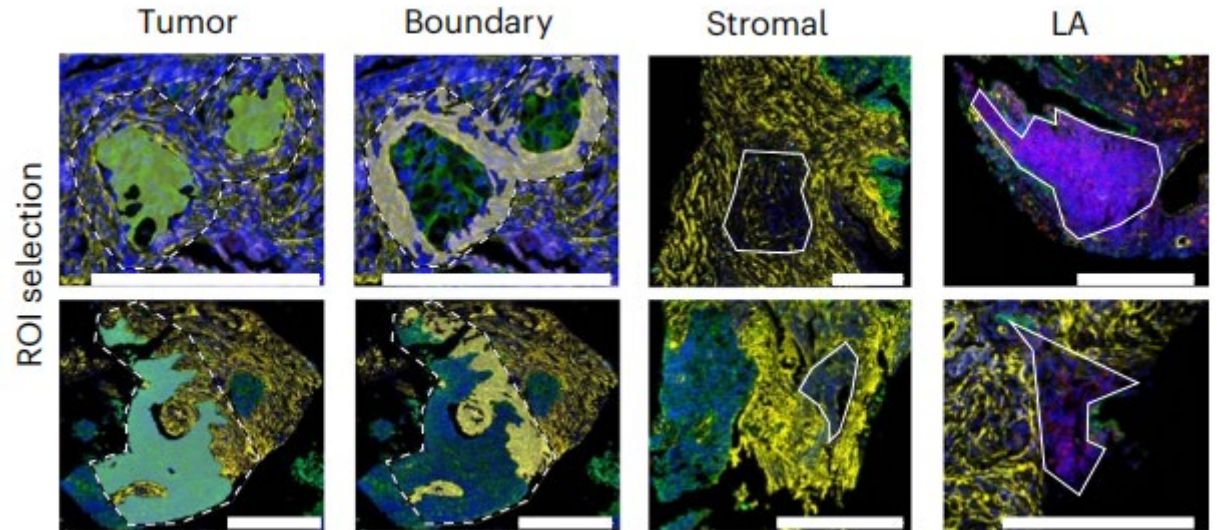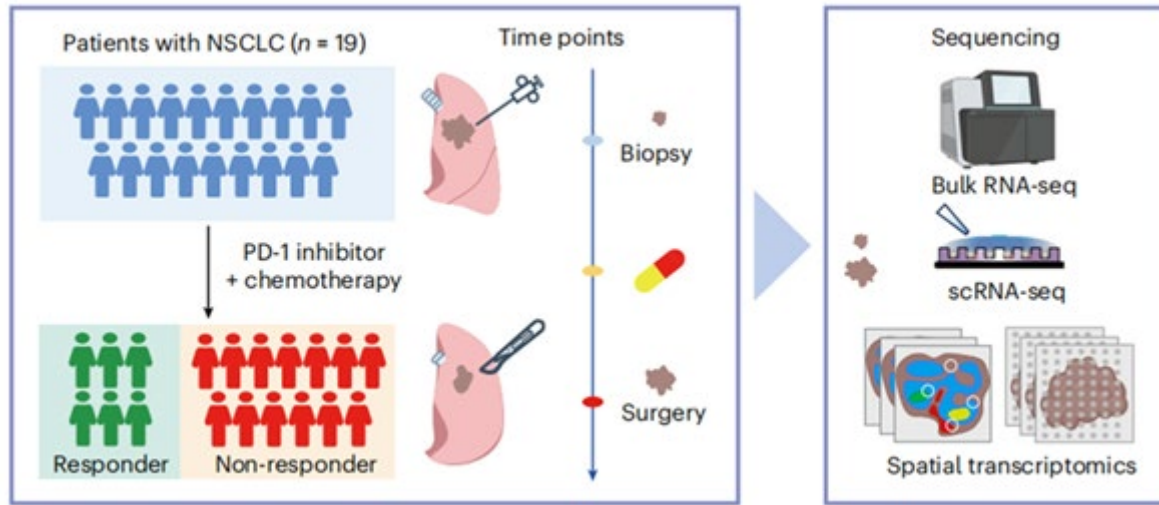
Supplementary Table 1. Clinical information of tumor samples

Patient ID	Sex	Age	Clinical stage	PD-L1 TPS	Neoadjuvant therapy	Pathology	Pathologic response	Residual tumor	RECIST	Matched scRNA-seq	scRNA-seq platform	ST barcode	cTNM	ypTNM
P01	female	61	IIIA	2%	Camrelizumab + Carboplatin+Pemetrexed	LUAD	NMPR	90%	SD	Yes	BD Biosciences Singleron	2B	IIIA	ypT2aN1M0
P02	male	55	IIIA	NA	Sintilimab + Carboplatin+Pemetrexed	LUSC	NMPR	90%	SD	No	BD Biosciences	1C	IIIA	ypT1bN2M0
P03	male	57	IIIA	NA	Toripalimab + Carboplatin+Gemcitabine	LUSC	NMPR	90%	SD	No	BD Biosciences	1B	IIIA	ypT2aN1M0
P04	female	45	IIIA	NA	Toripalimab + Carboplatin+Pemetrexed	LUAD	pCR	0%	PR	No	BD Biosciences	1A	IIIA	ypT0N0M0
P05	male	71	IB	NA	Sintilimab + Carboplatin+Pemetrexed	LUAD	pCR	0%	CR	Yes	Singleron	4C	IB	ypT0N0M0
P06	male	46	IIIA	< 1%	Toripalimab + Carboplatin+Pemetrexed	LUAD	MPR	5%	PR	Yes	Singleron	3D	IIIB	ypT2aN2M0
P07	male	62	IIIA	NA	Camrelizumab + Carboplatin+Gemcitabine	LUSC	NMPR	90%	PR	Yes	Singleron	3C	IIIA	ypT2aN0M0
P08	male	67	IIIA	NA	Camrelizumab + Carboplatin+Gemcitabine	LUSC	NMPR	70%	SD	Yes	Singleron	3A	IIIB	ypT4N2M0
P09	male	60	IIIA	90%	Pembrolizumab + Carboplatin+Pemetrexed	LUAD	pCR	0%	PR	Yes	Singleron	4D	IIIA	ypT0N0M0

16 samples provided all the required data and information.

Yilv Yan, et al. *Nature Genetics* 57: 126-139, 2025.
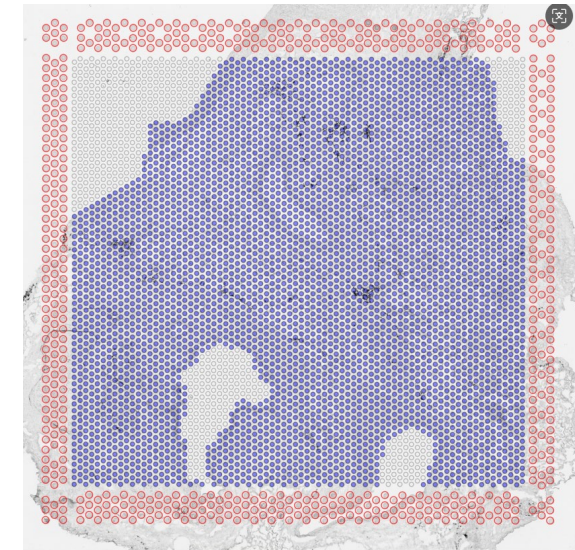
Deconvolution - Reference dataset

scRNA-seq：reference dataset

Subcategory

```
> ct_stat[1:30]              # 前 30 个最多的

        Plasma        E0_AT2      MB_NR4A1     Mono_VEGFA          Treg
         18166         13506         10889           9235          7950
      CD4_TCF7   Macro_CHI3L1 Macro_SELENOP       MB_TXNIP      CD4_NR4A2
          7567          7141          6735           6602          6562
      E1_Malig      CD8_GZMK    CD8_HAVCR2     Macro_CXCL9            AM
          6547          6015          5881           5786          5725
    Macro_CCL18    Macro_CXCL3      E2_Malig       E3_Malig        NK_GNLY
          5456          5191          4484           4019          3743
      E5_Malig    CD4_CXCL13     E6_Mucous       E9_Malig         Venule
          3649          3574          3451           3451          3426
      E7_Malig      E4_Malig          Mast     Macro_SPP1       E10_Malig
          3394          3262          3218           2983          2937
```

Classes： CD4+T, CD8+T, tumor epithelial cells
malignant tumor cells, other

```
table(meta$celltype)

     CD4 T        CD8 T        Other    Tumor_Epi  Tumor_Malig
     17703        18250        133180       23229        39718
```

Yilv Yan, et al. *Nature Genetics* 57: 126-139, 2025.

CARD Deconvolution method
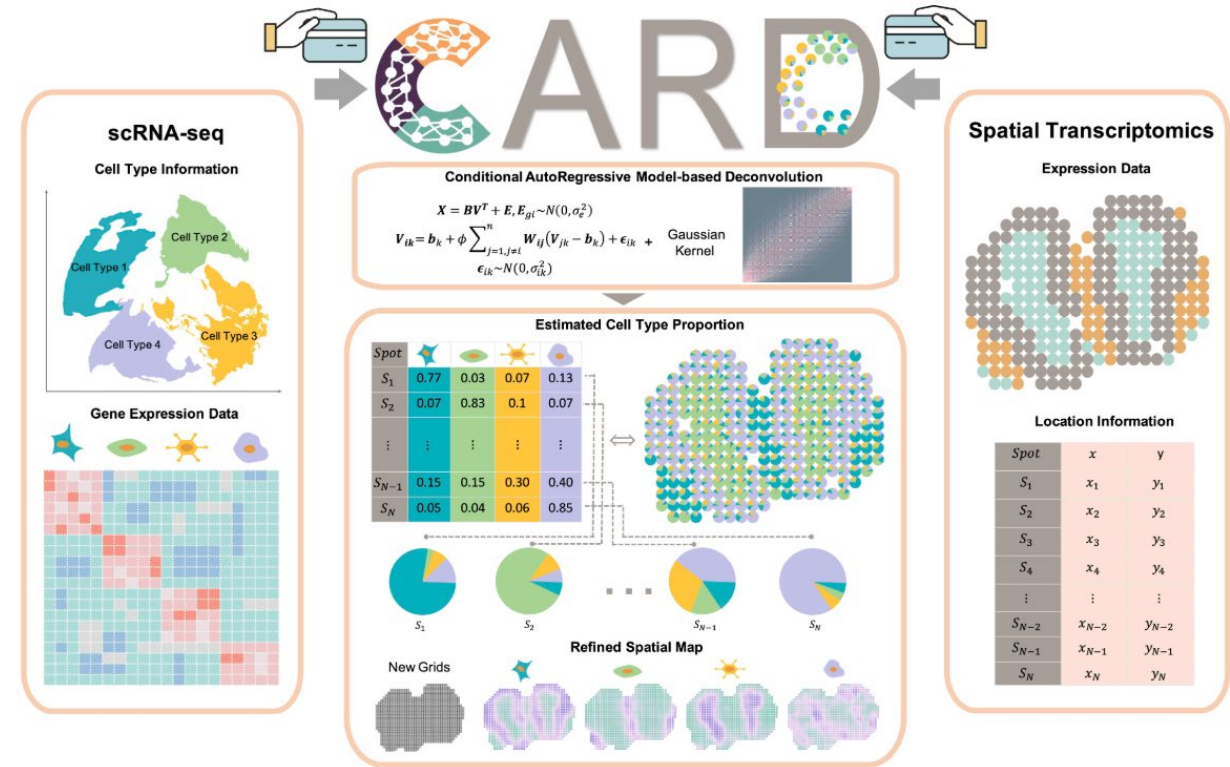
Y Ma, X Zhou, Spatially informed cell-type deconvolution for spatial transcriptomics. Nat. biotechnology 40, 1349–1359 (2022).
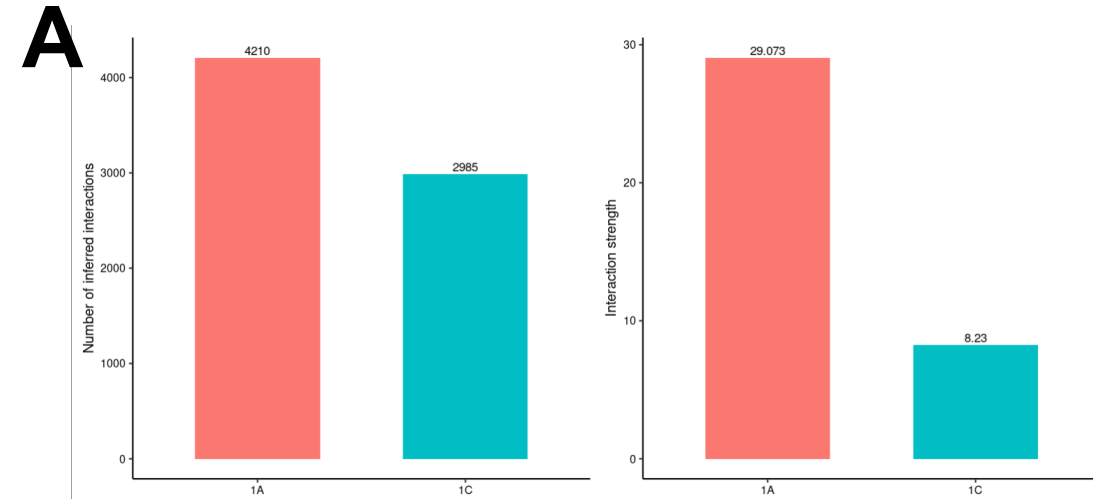
Spot-like Immune Patterns link to Favorable Prognosis

A

Patient ID: P04
ST barcode: 1A
Response: pCR
Recist: PR
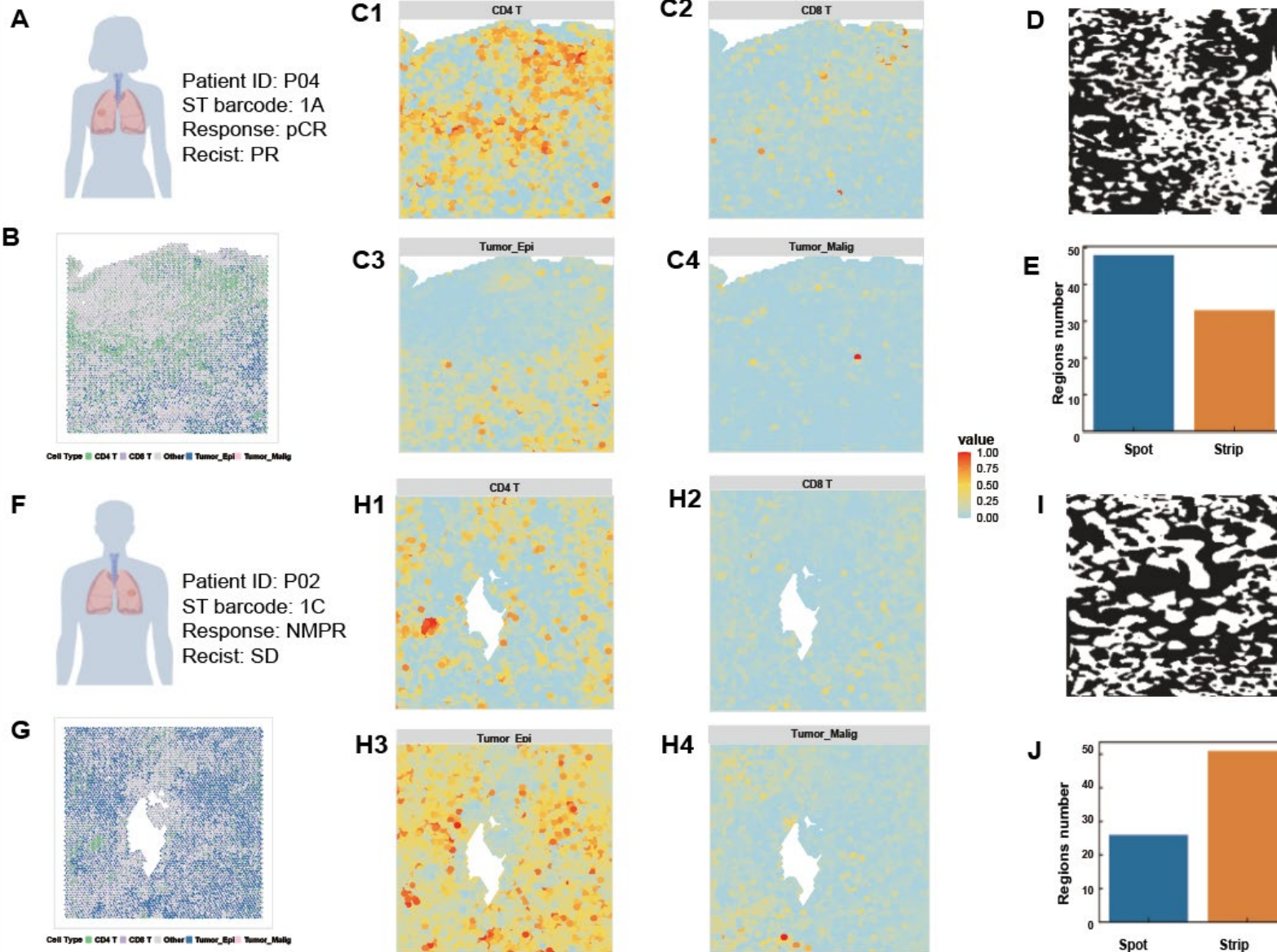
Spatial patterns correlate with clinical outcomes

PR: partial response;

SD: stable disease.

F

Patient ID: P02
ST barcode: 1C
Response: NMPR
Recist: SD

Comparison of Cell-Cell Communication Between Responder (1A) and Non-Responder (1C)

Jin at al, Nature Communication, 2021, Nature protocols, 2024
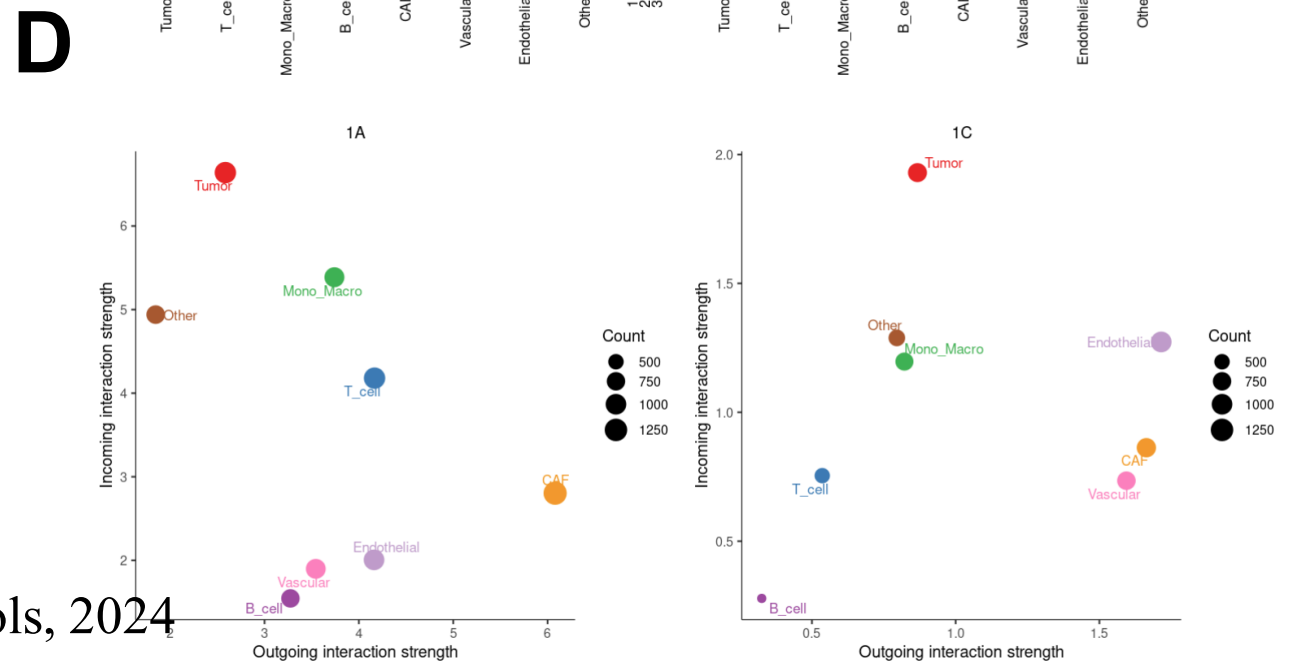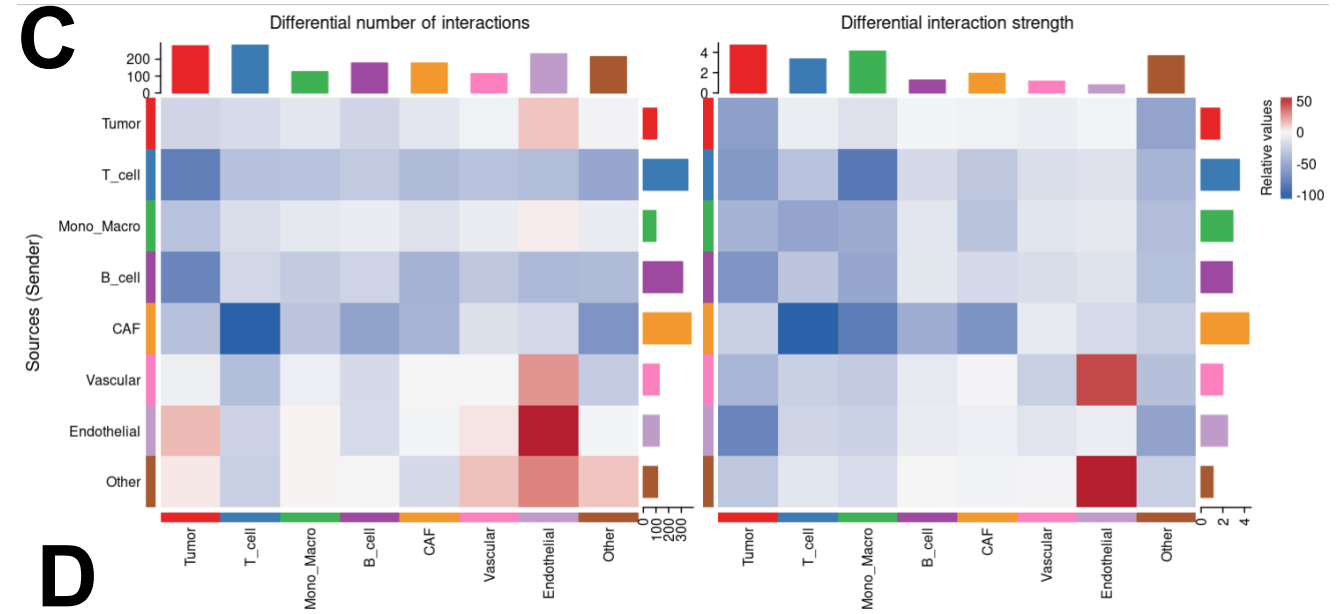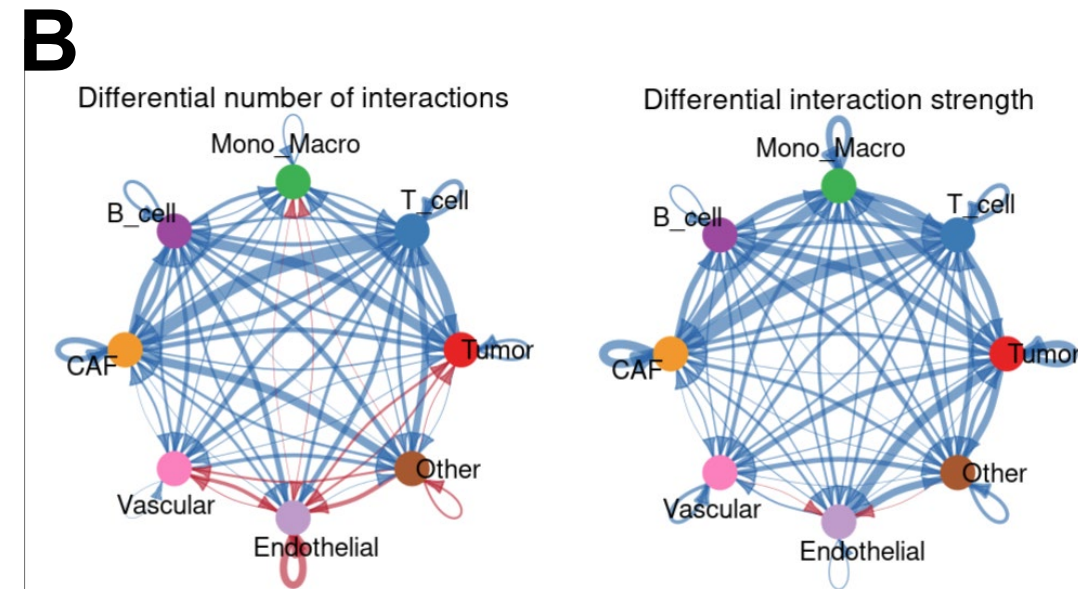
Distinct immune-related signaling networks

MHC-I signaling pathway network
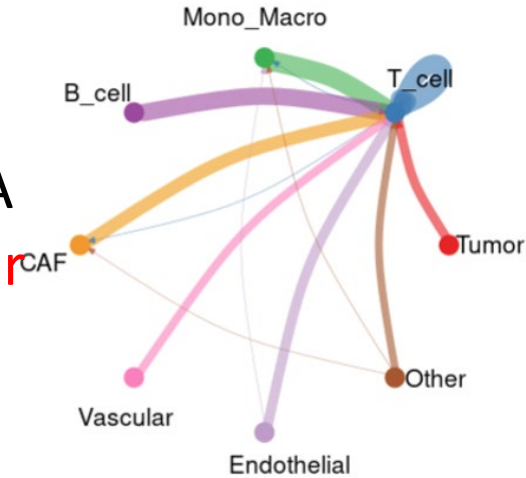
CCL signaling pathway network
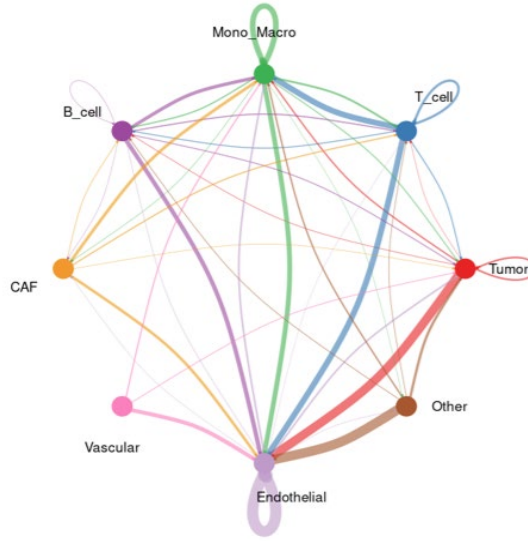
CXCL signaling pathway network
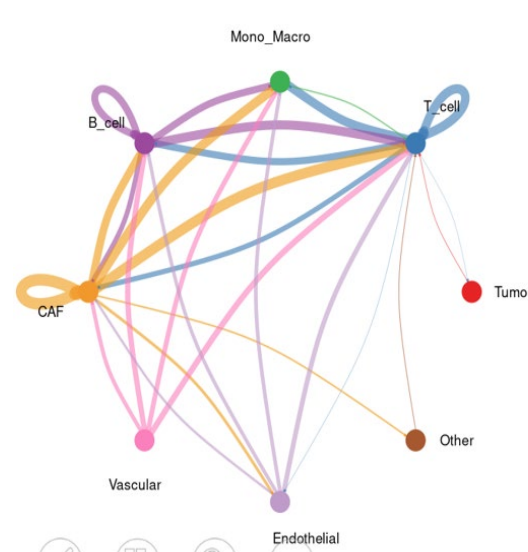
VEGF signaling pathway network

Patient 1A
Responder
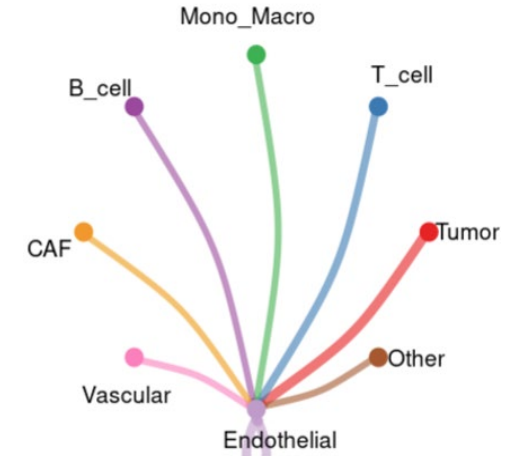
Patient 1C
Nonresponder
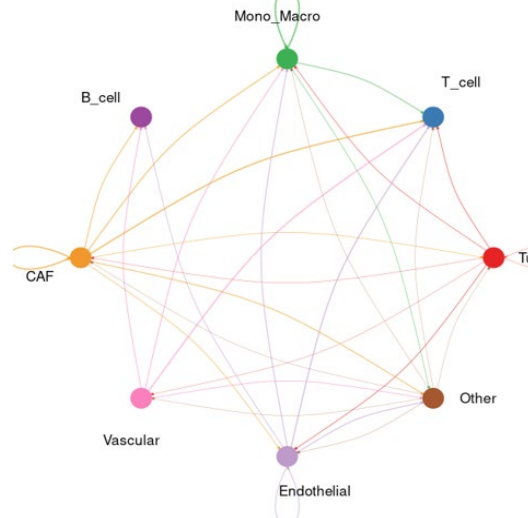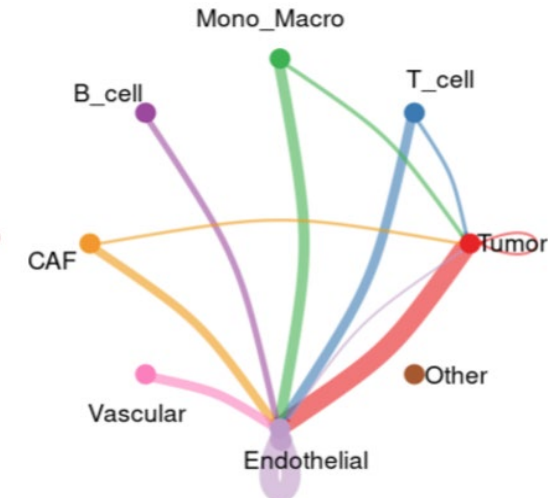
Summary II

➢ The reaction-diffusion mechanistic model identifies transient Turing instability patterns in tumor-immune dynamics.

➢ Immune dominance rate and duration directly correlate with topological pattern (spot-like vs. stripe-like structures).

➢ Spatial transcriptomics confirms that spot-like immune infiltration associates with improved patient outcomes.

➢ Cell-cell communication differs between responders and nonresponders.

Work III

➢ **Multiscale Modeling and Multidimensional Data Reveal Epigenetic Instability's Role in Cancer Drug Resistance**

# Continuous treatment leads to drug resistance
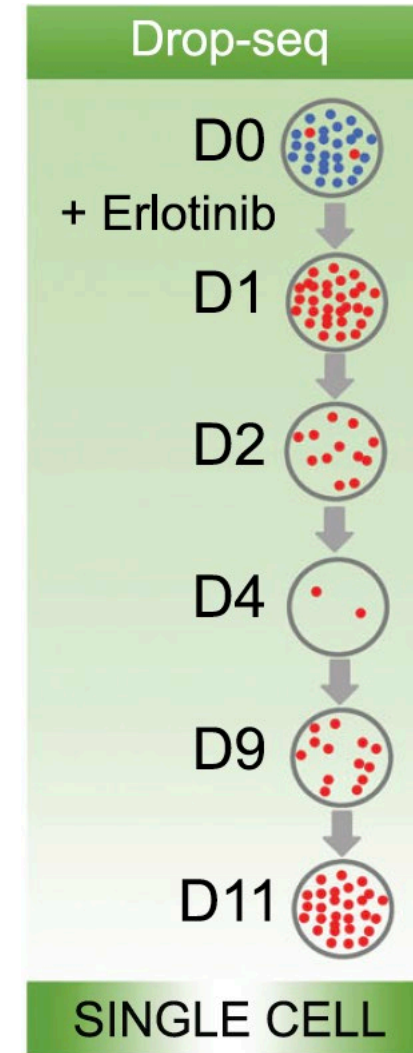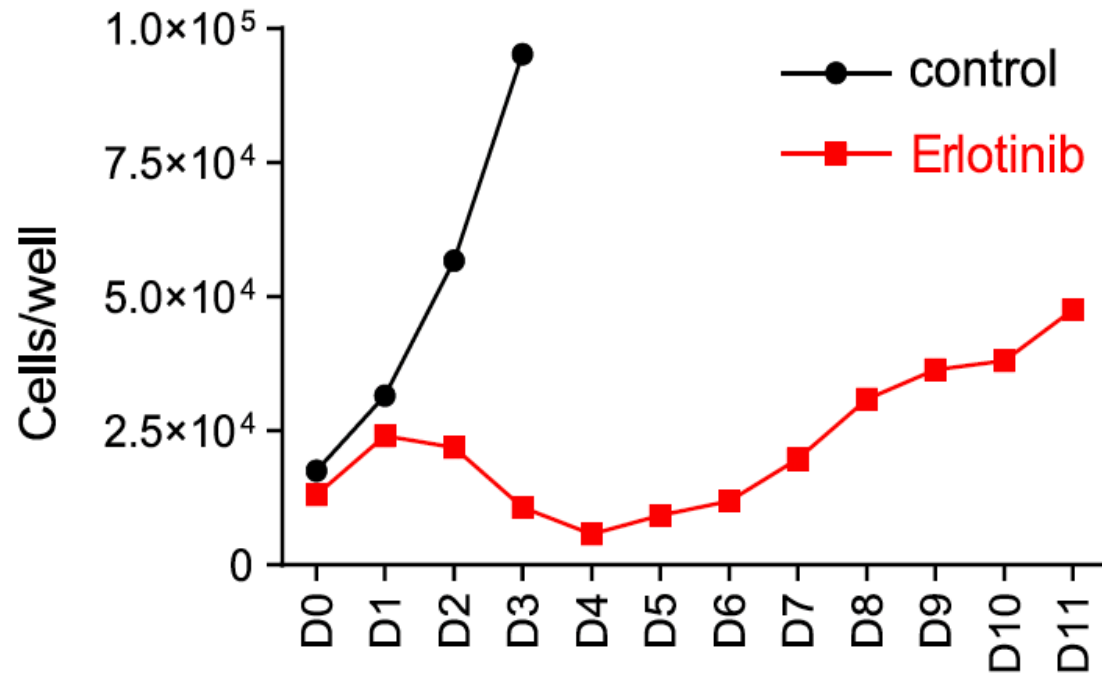
non-small-cell lung carcinoma (NSCLC) cell line PC9

A. F. Aissa et al., Single-Cell Transcriptional Changes Associated with Drug Tolerance and Response to Combination Therapies in Cancer, Nature Communications 12, 1 (2021).

Schematic for the set of consecutive samples for single-cell RNA-seq.
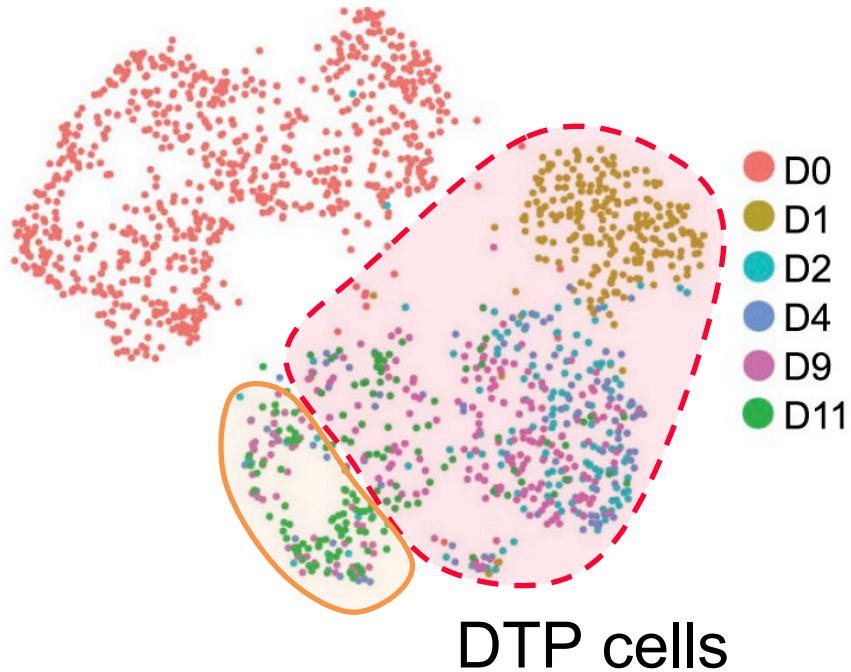
scRNA-seq reveals drug-tolerant cell subpopulations

drug-tolerant
expanded persister
(DTEP)

drug-tolerant
persister (DTP)
cells
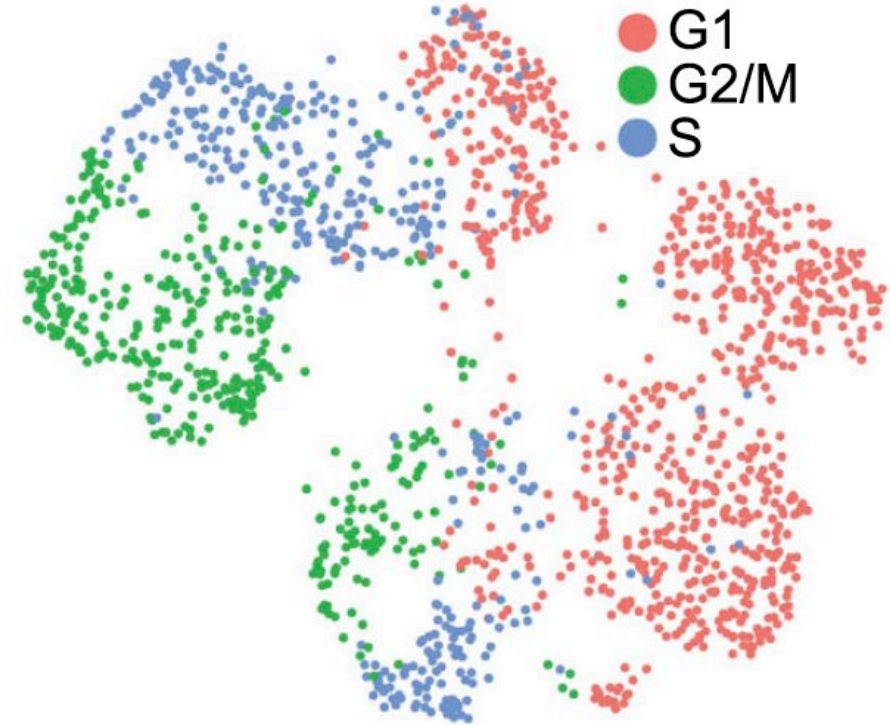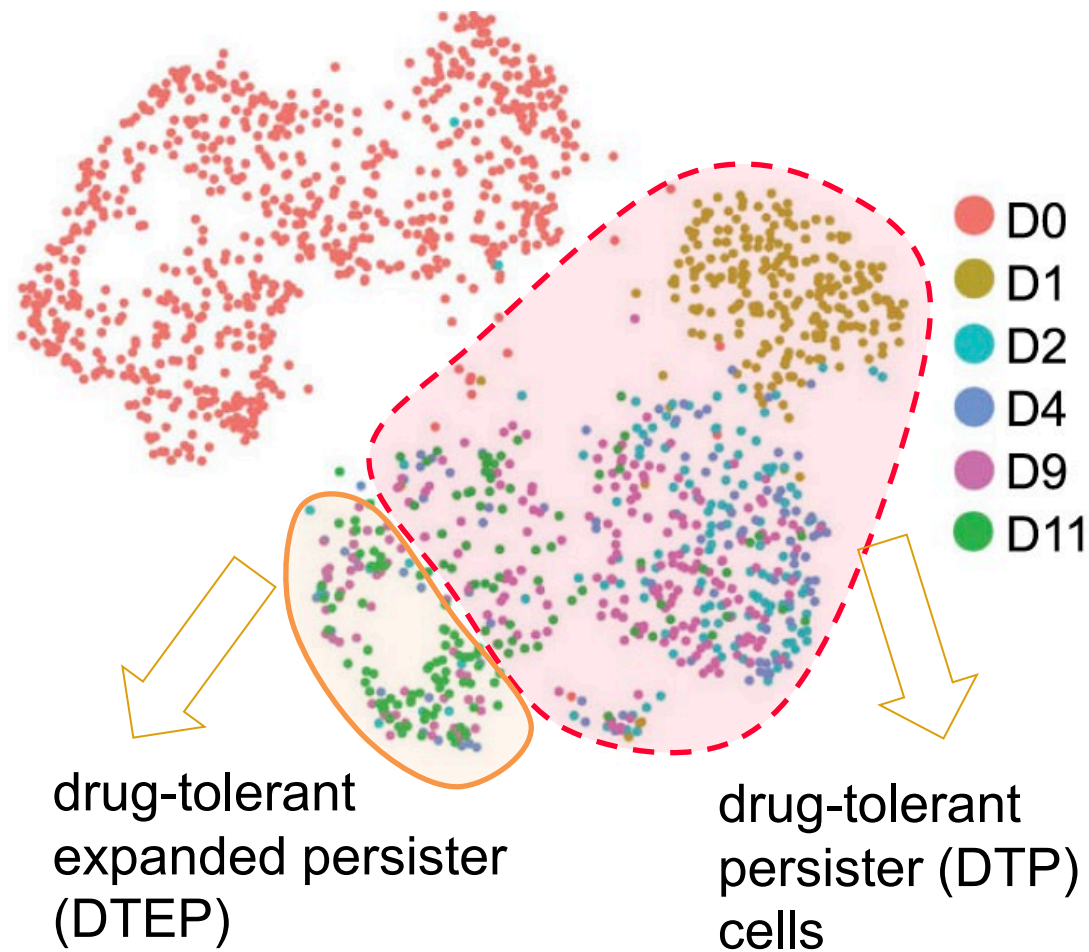
DTPs: a quiescent state
DTEPs: proliferate slowly
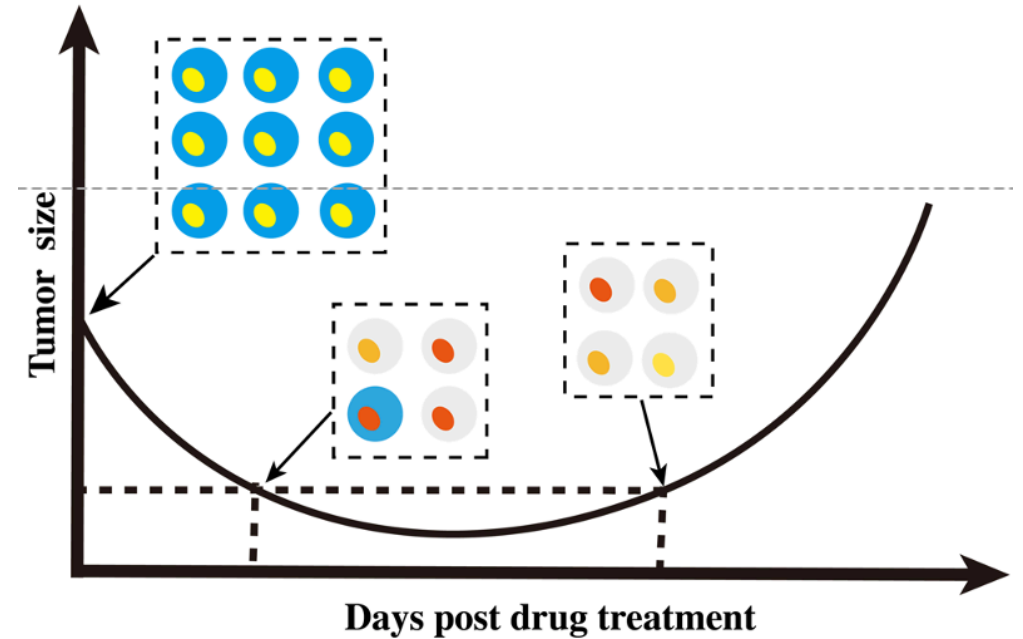
The Proposed Multiscale Model Framework
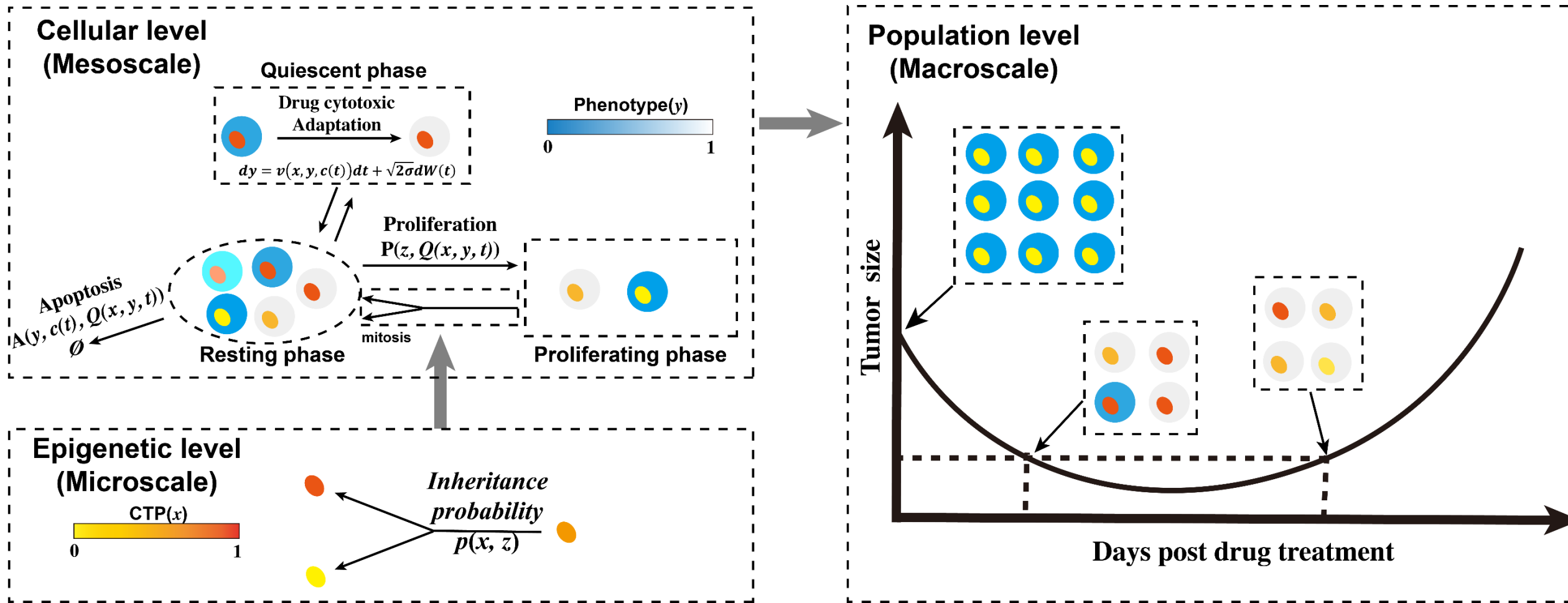
We used population balance analysis (PBA) and epigenetic inheritance dynamics during cell proliferation to describe the evolution of tumor cell epigenetics under exposure to cytotoxic drug concentration *c(t)*.

Phenotypic dynamics (at the macroscale)

$$\frac{\partial Q(x,y,t)}{\partial t} + \overbrace{\frac{\partial}{\partial y}\{J(x,y,c(t))Q(x,y,t)\}}^{\text{the flux of phenotype y}} = \overbrace{P(x,\hat{Q},Q)}^{\text{cell proliferation}} - \overbrace{A(y,c,Q)}^{\text{cell death}}$$

$Q(x, y, t)$: the number of cells at time point t in the resting phase, with epigenetic state CTP(x) and phenotype y

The flux of phenotype:
$$J(x,y,c(t)) = \overbrace{v(x,y,c(t);\ \hat{v})}^{\text{stress−induced adpatation to cell survival}} - \sigma\overbrace{\frac{\partial}{\partial y}\log Q(x,y,t)}^{\text{fluctuations in y}}$$

Epigenetic dynamics (at the microscale)

$$P(x, \hat{Q}, Q) = 2 \int_0^1 \beta\left(z, \hat{Q}(t)\right) Q(z, y, t) p(x, z) dz - \beta\left(x, \hat{Q}(t)\right) Q(x, y, t)$$

cell mitosis   entering proliferation phase

$$where\ \hat{Q}(t) = \int_0^1 \int_0^1 Q(x, y, t) dy dx\ ,\ \int_0^1 p(x, z) dx = 1\ , \forall z \in [0,1]$$

proliferation rate regulated by CTP   logistic growth

$$\beta(x, \hat{Q}) = \overbrace{\beta_1(x)} \overbrace{\left(1 - \frac{\hat{Q}}{K}\right)}\ ,\ \beta_1(x) = \beta_0 + \beta_{10} \frac{a_1 x + (a_2 x)^6}{1 + (a_3 x)^6}$$

First, we denoted the probability density of a single cell with CTP (x) and phenotype (y) as:

$$f(x, y, t) = Q(x, y, t)/\hat{Q}(t), \text{where } \hat{Q}(t) > 0, \forall t > 0$$

Using the chain law, we can easily obtain

$$\frac{\partial f(x, y, t)}{\partial t} = \frac{1}{\hat{Q}(t)} \left(\frac{\partial Q(x, y, t)}{\partial t} - f(x, y, t) \frac{d\hat{Q}(t)}{dt} \right)$$

$$= -\frac{\partial}{\partial y} \{ v(x, y, c(t); \hat{v}) f(x, y, t) \} + \sigma \frac{\partial^2}{\partial y^2} f(x, y, t) - \beta \left(x, \hat{Q}(t) \right) f(x, y, t)$$

$$+ 2 \int_0^1 \beta \left(x, \hat{Q}(t) \right) f(z, y, t) p(x, z) dz - \gamma(y, c, E) f(x, y, t)$$

$$- f(x, y, t) \int_0^1 \int_0^1 \left(\beta \left(z, \hat{Q}(t) \right) - \gamma(y, c) \right) f(x, y, t) dx dy$$

Single-cell epigenetic dynamics of acquired drug resistance

Next, we proposed a conditional probability for phenotype (y) denoted as h(y, t; x) and a marginal probability of CTP (x) denoted as g(x, t), let f(x, y, t) = h(y, t; x)g(x, t). Using the chain law, we easily obtain

Phenotype dynamics

$$\frac{\partial f(x,y,t)}{\partial t} = g(x,t)\frac{\partial h(y,t;x)}{\partial t} + h(y,t;x)\frac{\partial g(x,t)}{\partial t}$$

$$\frac{\partial h(y,t;x)}{\partial t} = -\frac{\partial}{\partial y}\{v(x,y,c(t))h(y,t;x)\} + \sigma\frac{\partial^2}{\partial y^2}h(y,t;x)$$

$$- h(y,t;x)(\gamma(y,c) - \int_0^1\int_0^1 \gamma(y,c)h(y,t;x)g(x,t)dxdy).$$

Epigenetic dynamics

（17）

$$\frac{\partial g(x,t)}{\partial t} = 2\int_0^1 \beta_1(z)g(z,t)p(x,z)dz - g(x,t)(\beta_1(x) + \int_0^1 \beta_1(z)g(z,t)dz).$$

（18）

The proposed multiscale model provides continuity and interpretability from macroscopic population dynamics to microscopic epigenetic dynamics.
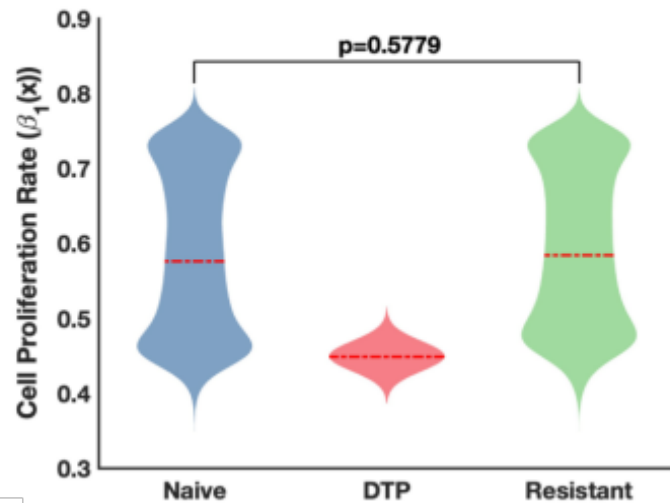
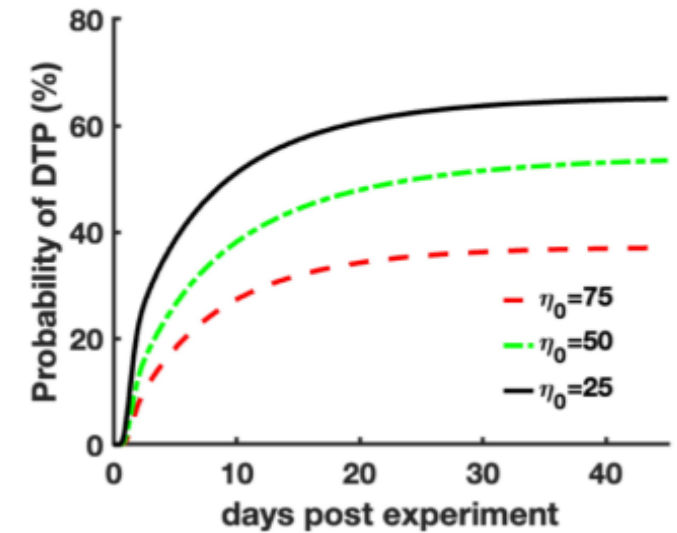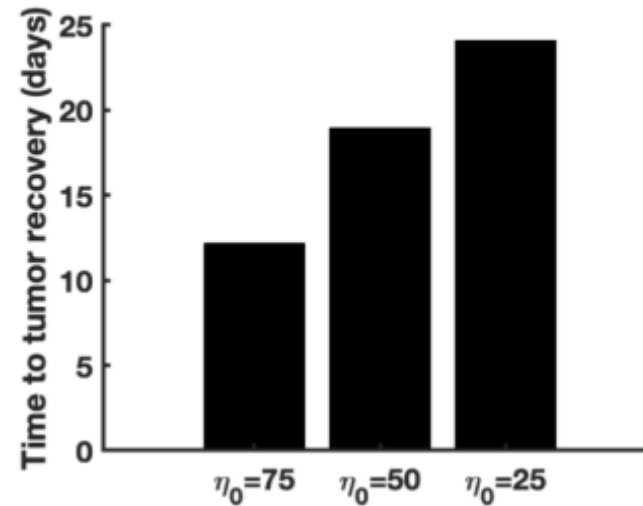# Simulations and Comparison with experimental data
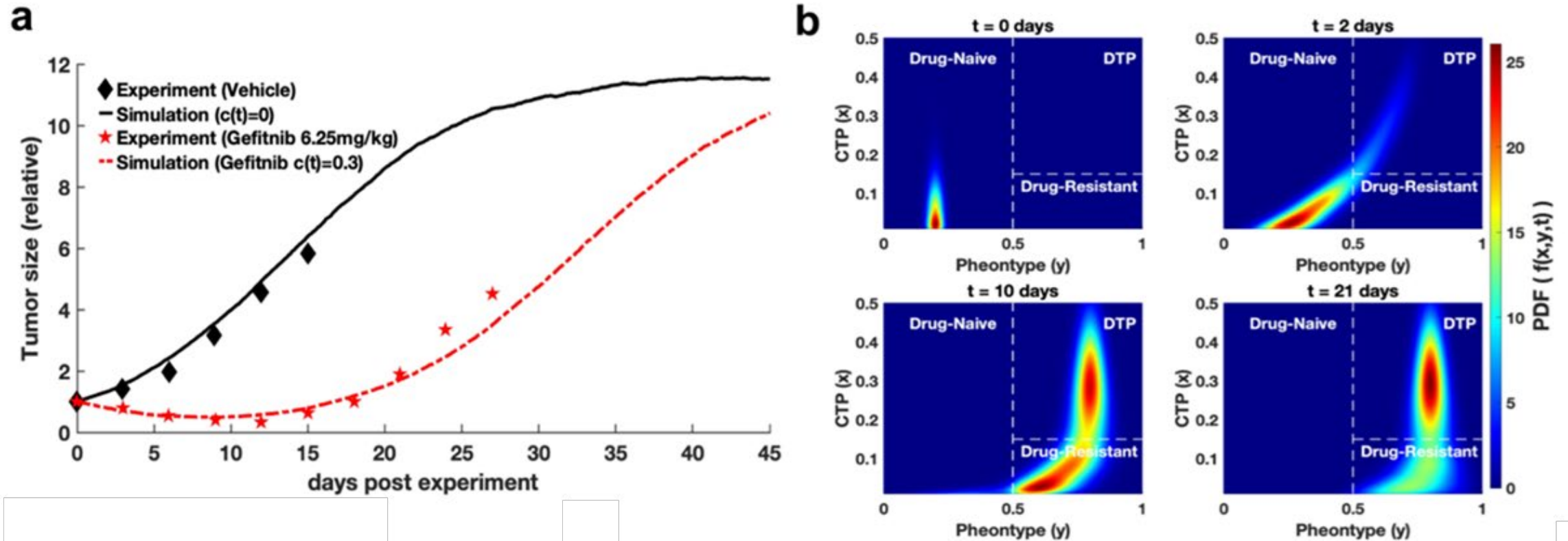
The model reconstructs the dynamic process of acquired drug resistance at the cell population level and defines three types of cells at the epigenetic level.
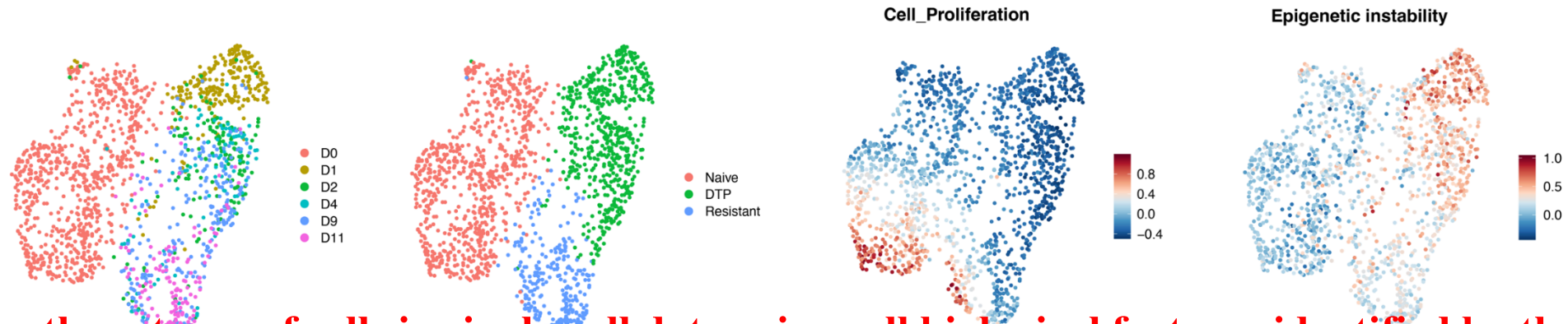
Features:
➤ **The proliferation capacity of DTPs is lower than that of the other two types of cells;**
➤ **High epigenetic instability is the key factor in observing the emergence of tolerant cell subpopulations，**

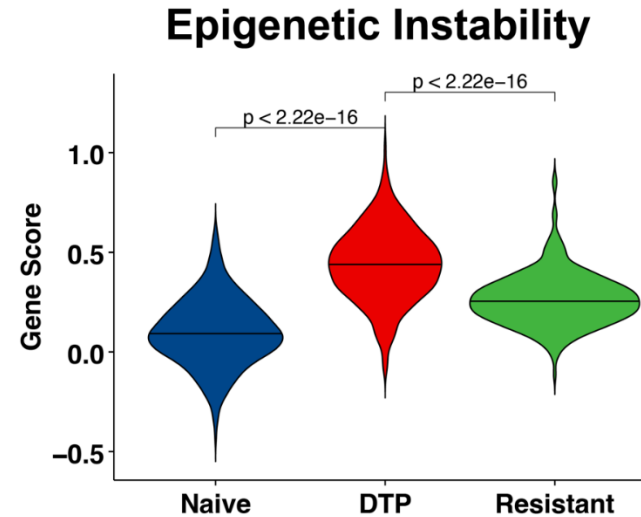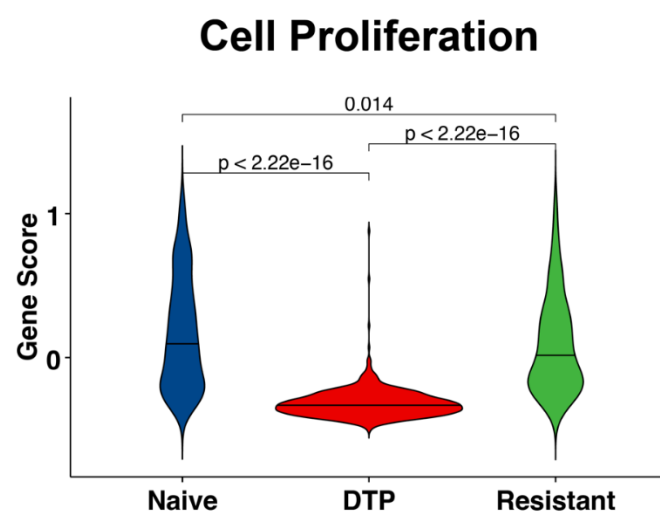# Model-supervised single-cell clustering analysis

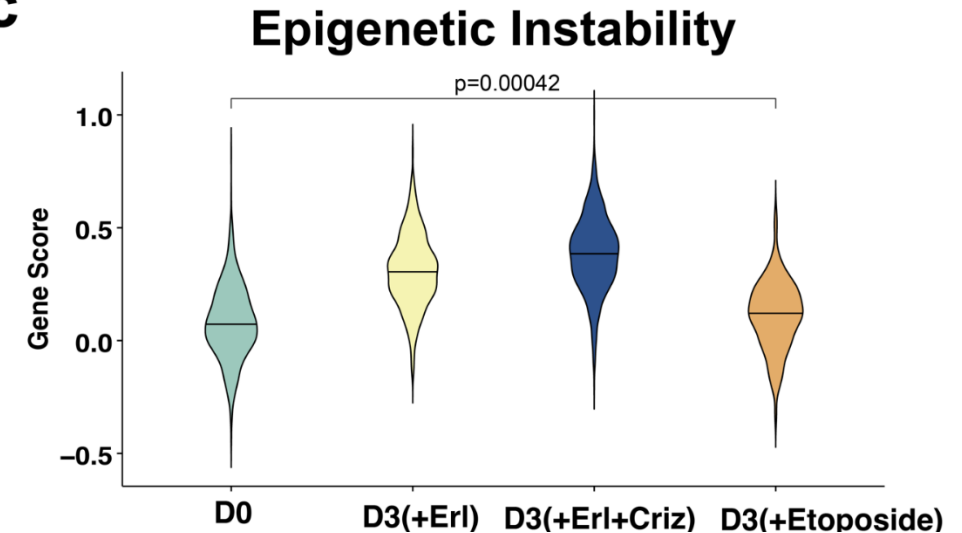> **Cell proliferation and epigenetic instability are identified as two key dynamic features**

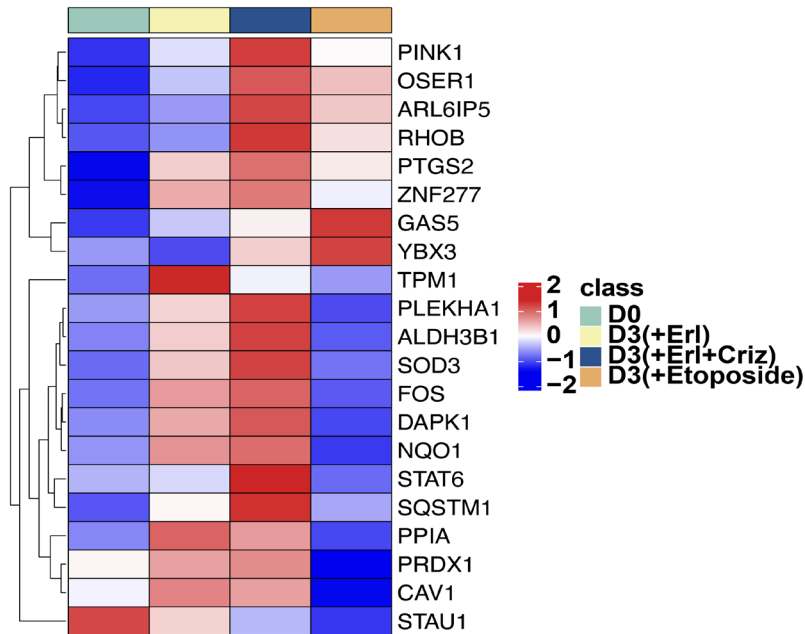**Clustering three types of cells in single-cell data using cell biological features identified by the model**
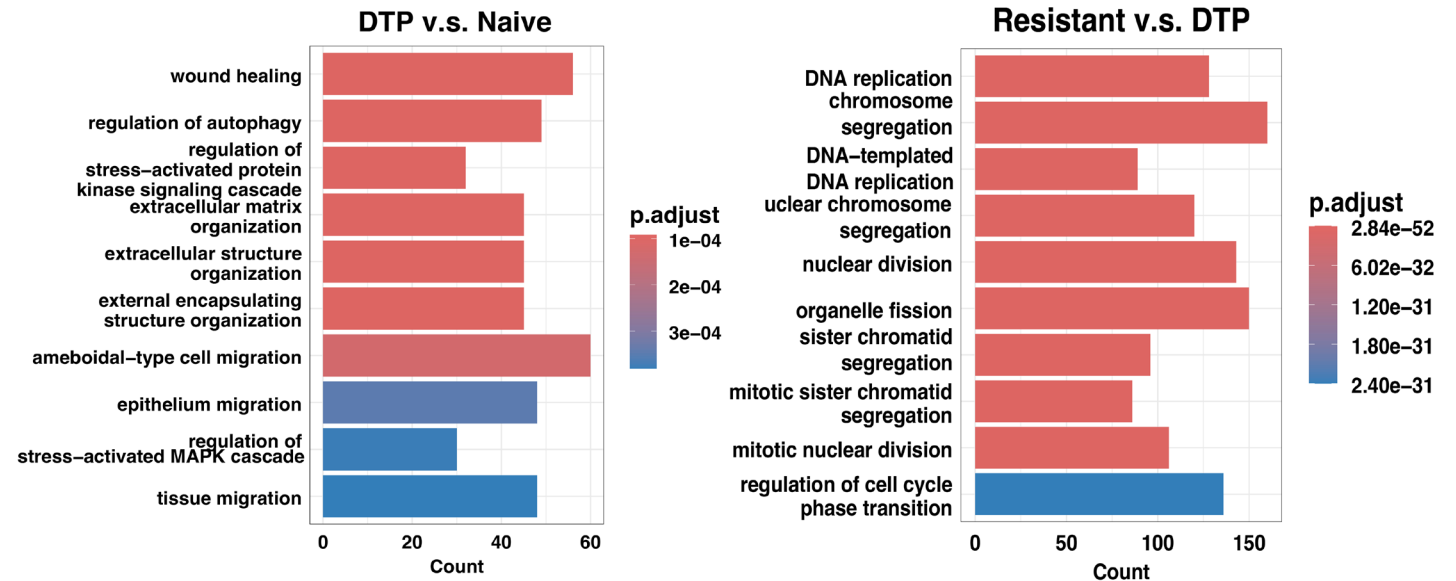
Identifying drug tolerance biomarkers via scRNA-seq

Gene sets associated with epigenetic instability

GO enrichment analysis

The results identify three drug tolerance marker genes (TPM1, CAV1, and PPIA) and two experimentally validated biomarker functions of drug-tolerant persister (DTP) cells: cell autophagy and cell migration.

The GO enrichment analysis of differentially expressed genes between drug-resistant and DTP cells revealed that the predominant difference between these two cell types was the regulation of cell proliferation, such as DNA replication and mitotic nuclear division (Fig 4e).
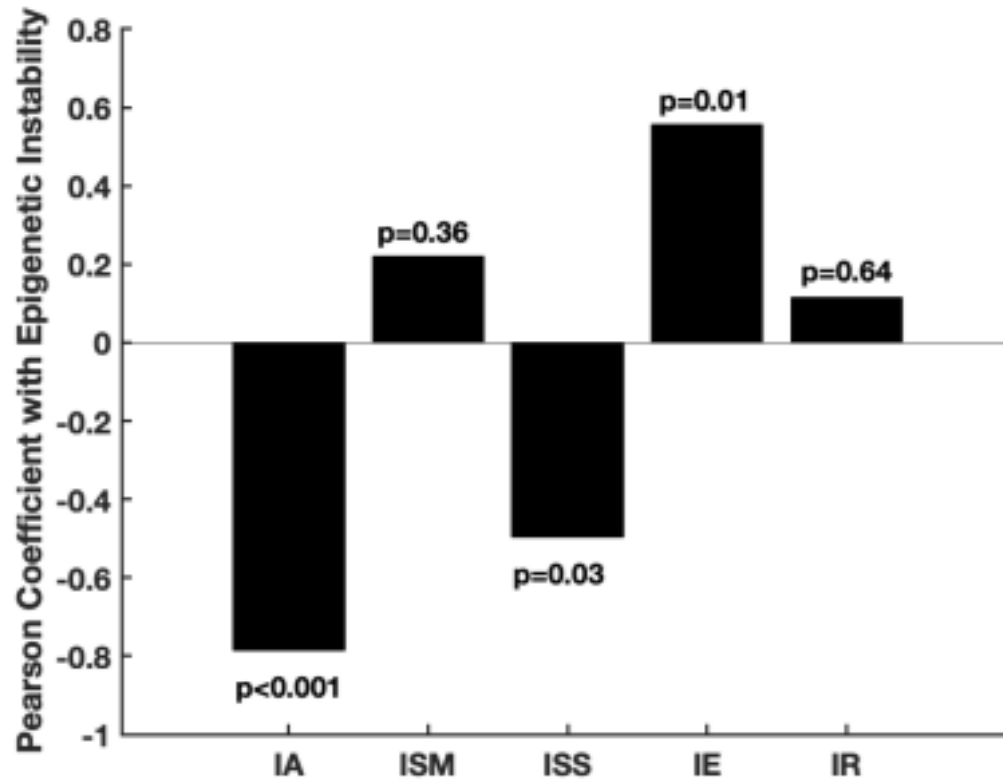
Five subtypes of the immune microenvironment
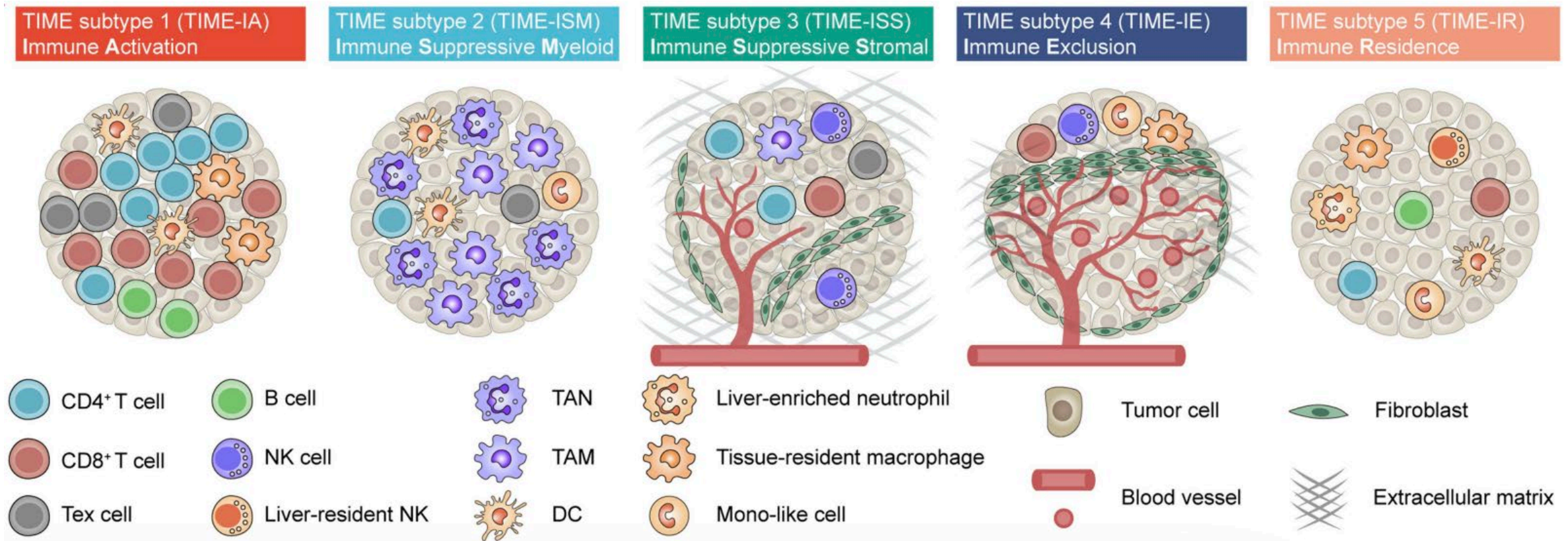
scRNA-seq data of liver cancer

Xue et al., Nature 612, 141–147 (2022)

IA：（Immune Activation），ISM：（Immune Suppressive Myeloid），
ISS：（Immune Suppressive Stromal），IE：（Immune Exclusion），IR：（Immune Residence）

Relationship between drug tolerance and immune microenvironment

IA：　(Immune Activation)，

ISM：　(Immune Suppressive Myeloid)，

ISS：　(Immune Suppressive Stromal)，

IE：　(Immune Exclusion)，

IR：　(Immune Residence)

The results suggest that drug treatment tolerance may be positively correlated with an immune-excluded tumor microenvironment subtype.

Fig. F. Pearson relationship analysis between epigenetic instability and five subtypes of tumor immune microenvironment in scRNA-seq data. P-values are indicated on the top of each bar. IA：　Immune Activation, ISM: Immune Suppressive Myeloid, ISS: Immune Suppressive Stromal, IE: Immune Exclusion, IR: Immune Residence.

Summary of predictive biomarkers for acquired drug resistance

This figure was created in BioRender

Summary III

➢ Developed a multiscale computational framework to explore the dynamic features of the evolution of acquired drug resistance.

➢ Tested preliminary model predictions using scRNA-seq data.

➢ Predicting novel biomarks.

Conclusions

Integrating AI approaches and mathematical models offers new avenues to realize the systems learning from multidimensional data.

➢ **Bridging heterogeneous tumor growth pattern with diverse cellular subclones (Single-cell data+ multiscale modeling)**

➢ **Decoding mechanisms of therapy resistance and** identifying novel biomarks (**Single-cell data + population data+ multiscale modeling)**

➢ **Predicting clinical outcomes based on spatial pattern (Spatial RNA-seq data + PDE model)**

Acknowledgements

Ke Qi

Shun Wang

Tengfei Wang

Zhihao Yao
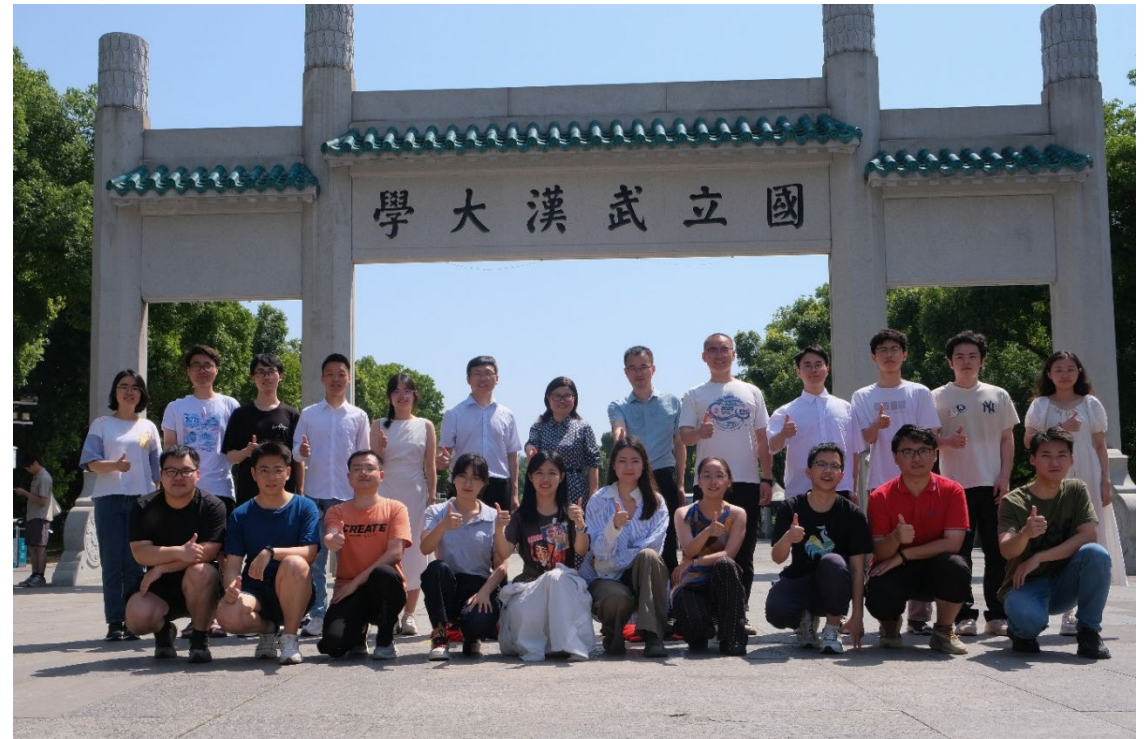
Thanks for your attention!