Single-molecule genomics data delineate patient-specific tumor profiles and cancer stem cell organization

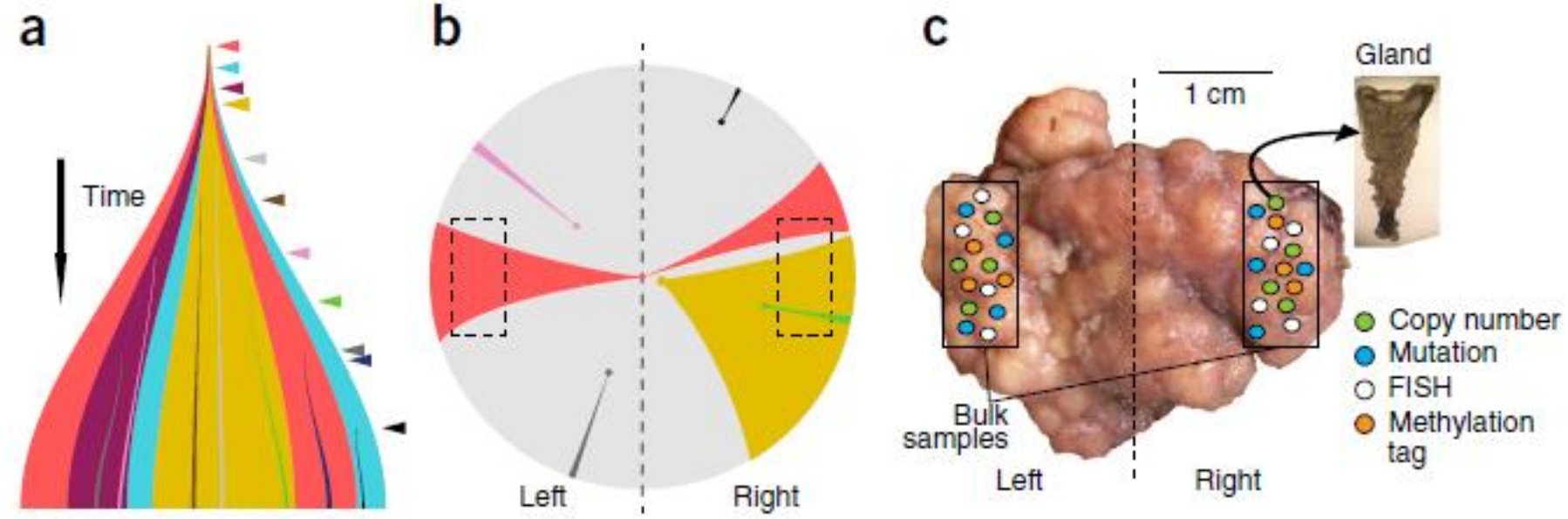
Yifan Yang

2015-3-2

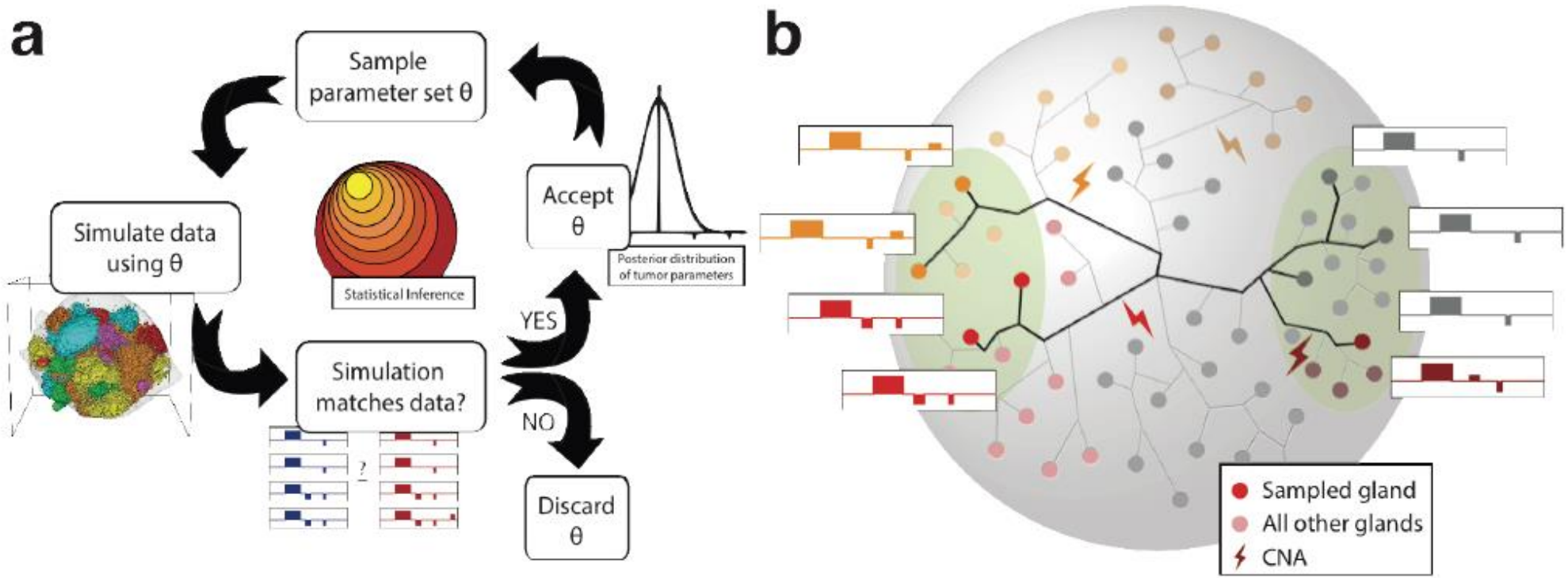
- https://www.synapse.org/#!Challenges:DREAM

A big bang model of colon tumor growth

- “It’s like going back in time. The history of each tumor is written in its genomes. To prevent tumors, you want to see what happened early on and how to stop their first cell divisions.”



Aim: How to model tumor growth considering the heterogeneity



Data:

Table 1. Patient clinical information

Patient	Age	Tumor size (cm)	Stage	MSI+
CT	53	4.5	3	N
CU	50	4.5	1	Y
CX	44	9	3	N
HA	61	7.5	3	N
Z	83	6	3	NA

Method: Three building blocks

- The patient molecular data (the history of tumor evolution)
  - somatic point mutation
  - microsatellites
  - neutral methylation
- The mathematical/computational model
  - simulate tumor growth in a spatial fashion
  - mutation rate, apoptosis, etc
- The statistical inference
  - probability distribution of parameters.

Block 1: The patient molecular data

- Choose a clock: Neutral somatic mutation
- This measurement is used to indicate the historical mutation/changing during tumor growth, also related to tumor status.
- DNA methylation of IRX2
- 8 GpC islands : 2^8 = 256 patterns (binary representation)
- Low methylation: non-dividing tissues (differentiated)
- High methylation: dividing tissues(proliferative)

Block 2: simulate tumor growth in a spatial fashion

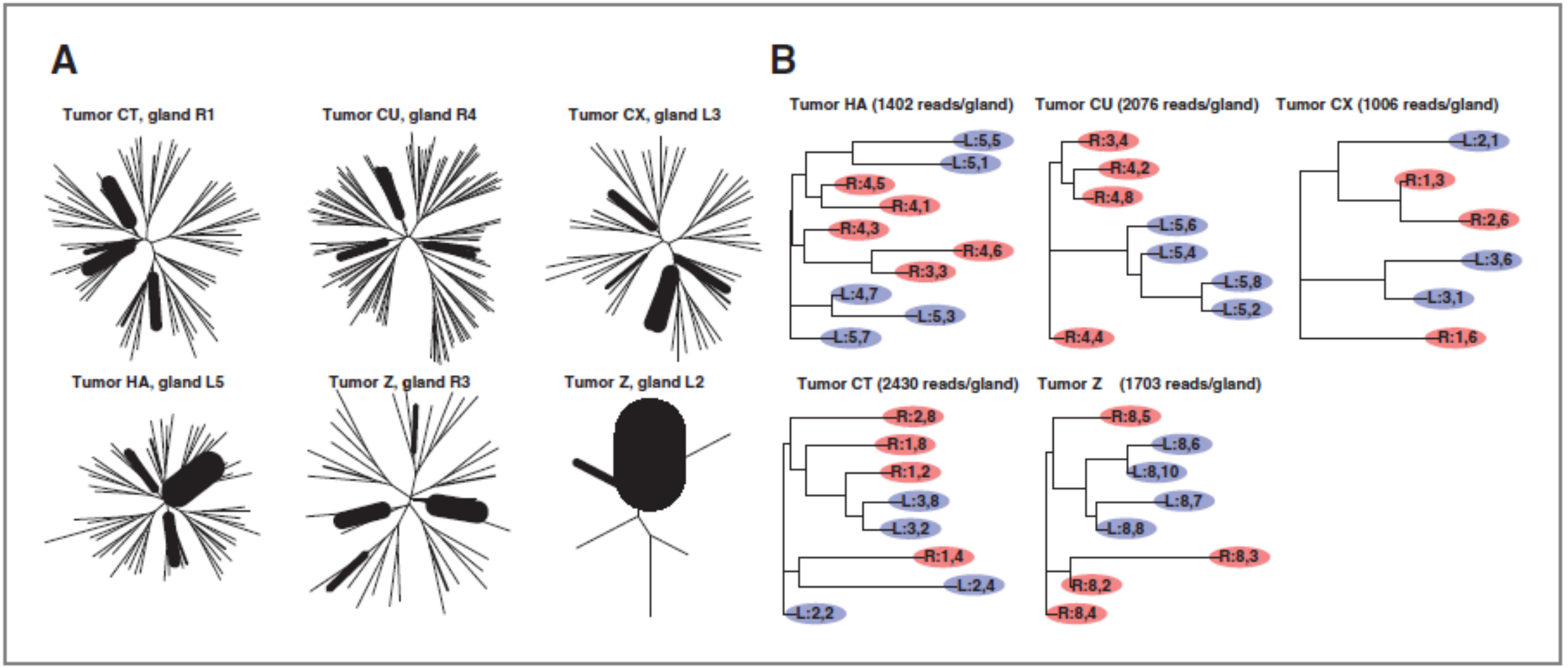
- Dynamic models (numerous)
- Build a phylogenetic tree (spatial property)
- Go to Pseudocode 1 (crecspace)
- Go to Pseudocode 2 (crecmeth)

Block 3: statistical inference

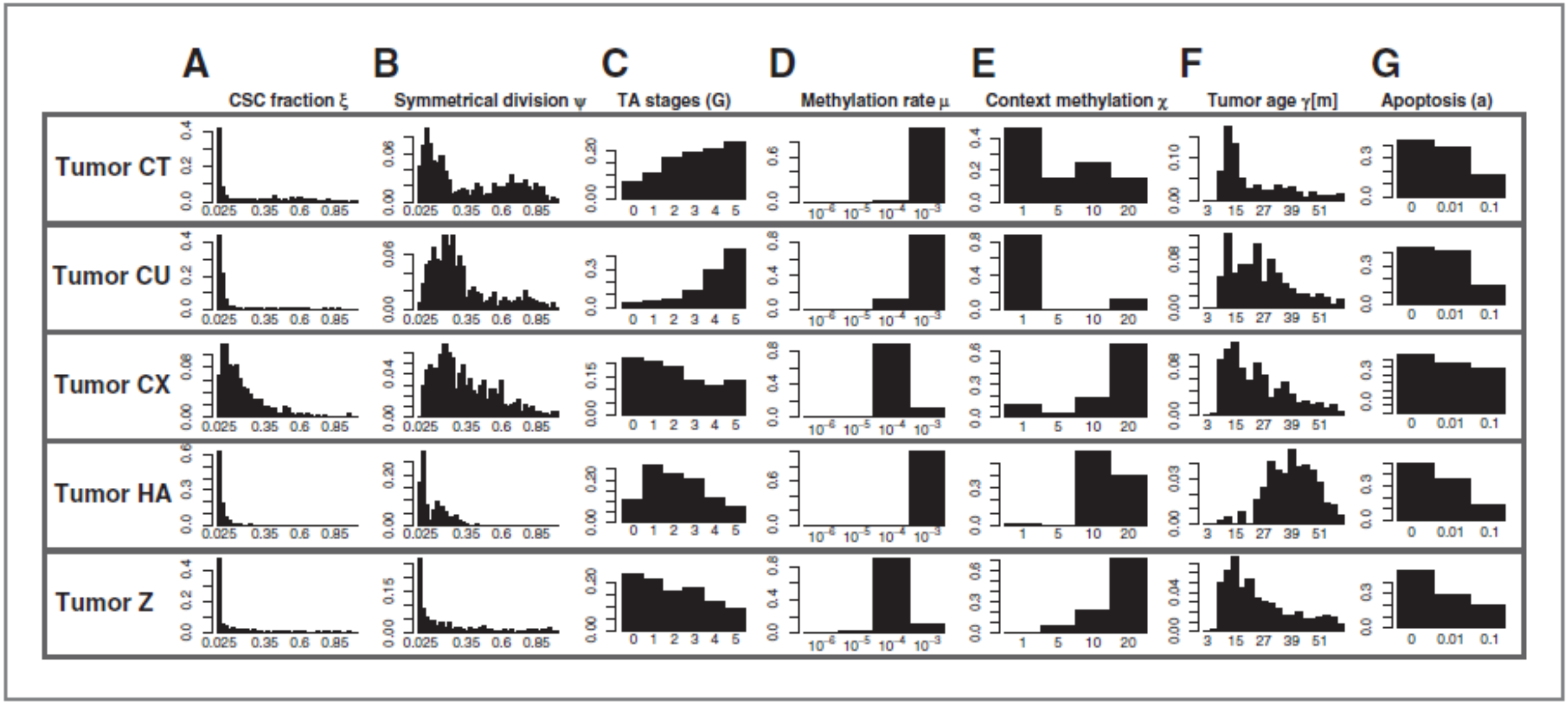
- Rejection sampling

1. Sample the parameter θ from the prior distribution $P(\theta)$.
2. Simulate the data D' from the computational model with input θ.
3. If $\rho(S[\mathrm{D}], S[\mathrm{D}']) < \varepsilon$ accept θ.
4. Go to 1.

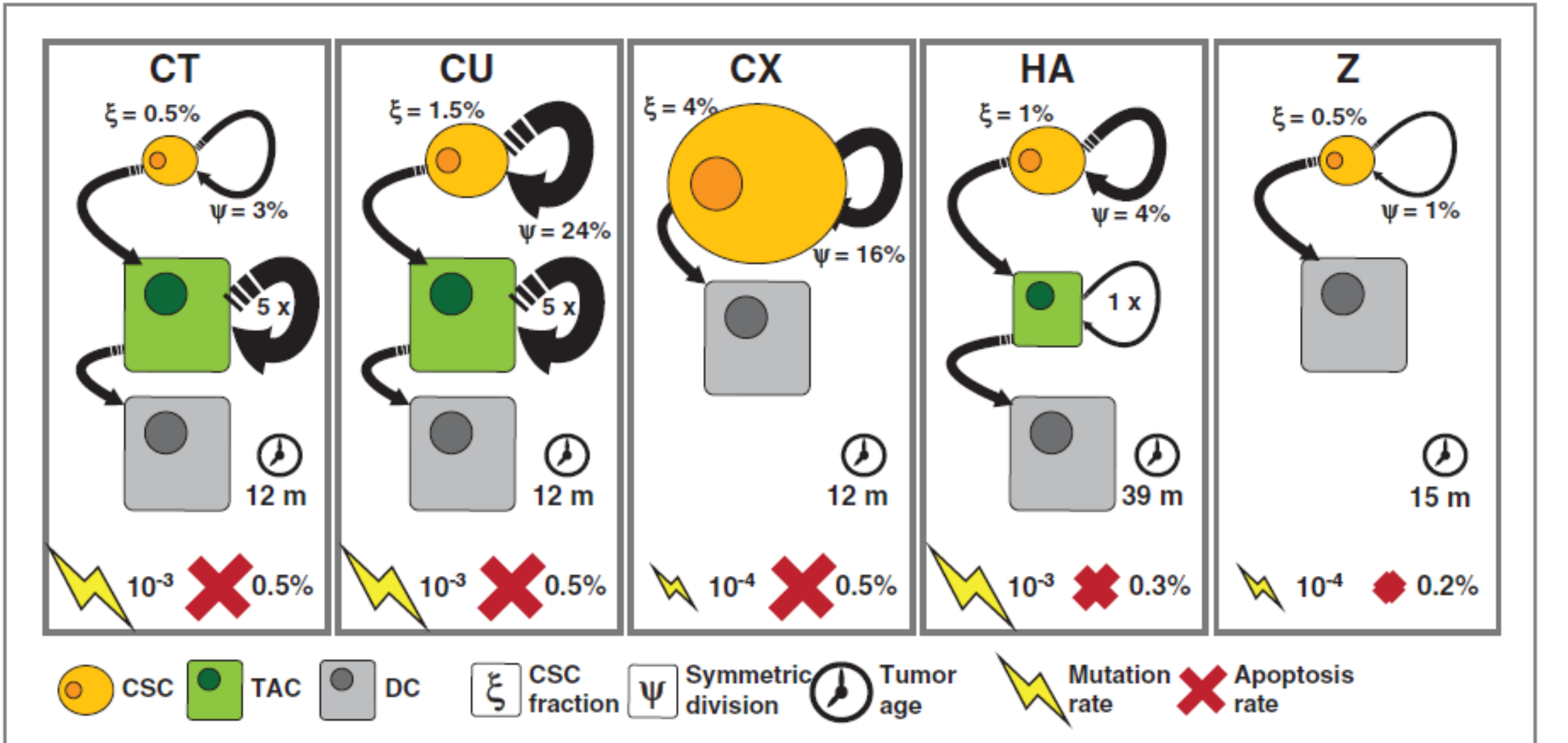
Results 1: Cancer glands are heterogeneous



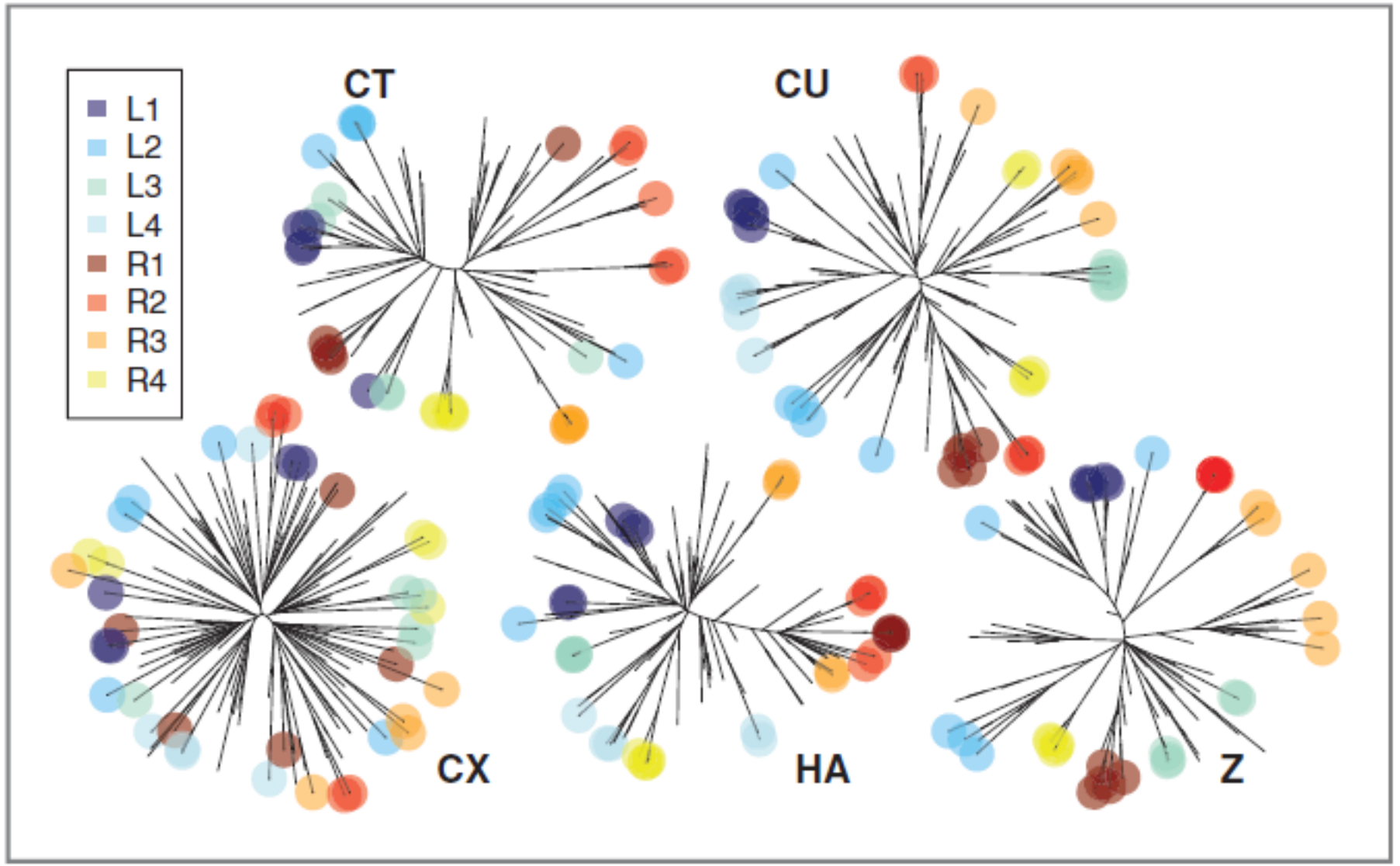
Results 2: parameters summary for each patient



Results 3:Cartoon representation of the tumor growth characteristics for each patient



Results 4: reconstructing tumor ancestral tree



Discussion:

- Deep sequencing technologies have enable the generation of high-throughput molecular data, which coupled with careful tumor sampling schemes, can provide new insight into tumorigenesis.
- Using our model, we simulated the entire history of a tumor using the inferred parameters to generate a representation of a phylogenetic tree.

- How to decide the number of subclones in a tumor?
- How to utilize molecular data (multi-genes? RNA-seq data? Meta-data?) to retrieve the tumor growth history?
- How to simulate tumor growth considering single-cell heterogeneity?