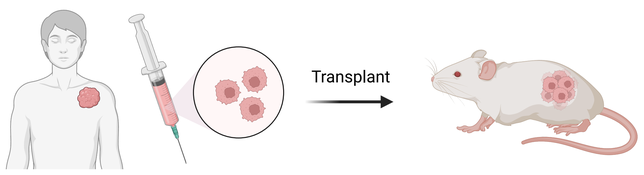
Computational Biology Summer Program

Probabilistic Time Series Modeling of Tumor Growth in Patient-Derived Xenografts

Rohan Subramanian[1], Wesley Tansey[2]
[1]Harvey Mudd College, [2]Memorial Sloan Kettering Cancer Center

Introduction

- Pre-clinical drug screens in limited *in vitro* or *ex vivo* cancer models (cell lines) generalize poorly to *in vivo* systems.
- Patient-derived xenograft models (PDXs) provide a more accurate *in vivo* approach.
- Simulations and predictive models of tumor growth in PDXs can help make targeted predictions of drug responses.



Objectives

- Simulate tumor growth in PDXs using probabilistic models and infer underlying latent disease states.
- Predict tumor growth curves for specific PDX/drug combinations.

Methods

Data and Preprocessing

- 62 drugs and 277 PDX models across 6 cancer types, 4699 total curves [1].
- Normalized by day 0 volume.
- Missing time points imputed using linear interpolation.

Factor Model

- X_{ij} is AUC up to day 35 for PDX i, drug j.
- Learn PDX, drug embeddings W, V.

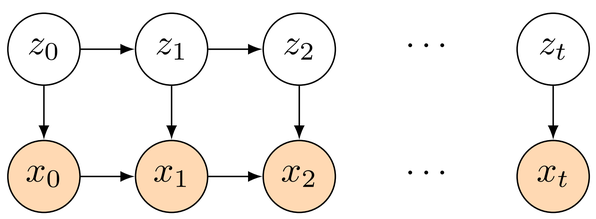
$$\hat{X}_{ij} = W_i \cdot V_j$$

$$X_{ij} \sim \mathcal{N}(\hat{X}_{ij}, \sigma^2)$$

Autoregressive Hidden Markov Model

- x_t is the observed tumor volume and z_t is the discrete latent state on day t [2].

$$x_t | x_{t-1}, z_t \sim \mathcal{N}(A_{z_t} x_{t-1} + b_{z_t}, \sigma^2)$$



Multiview Model

- Transition matrix for PDX i and drug j is computed as the inner product of PDX, drug, current state and next state embeddings.

$$\phi_{ij}(s_{t-1}, s') = \sum_{r=1}^{d} w_{ir} v_{jr} m_{s_{t-1} r} m_{s' r}$$

Results

Learned PDX Embeddings from Factor Model Cluster by Cancer Type

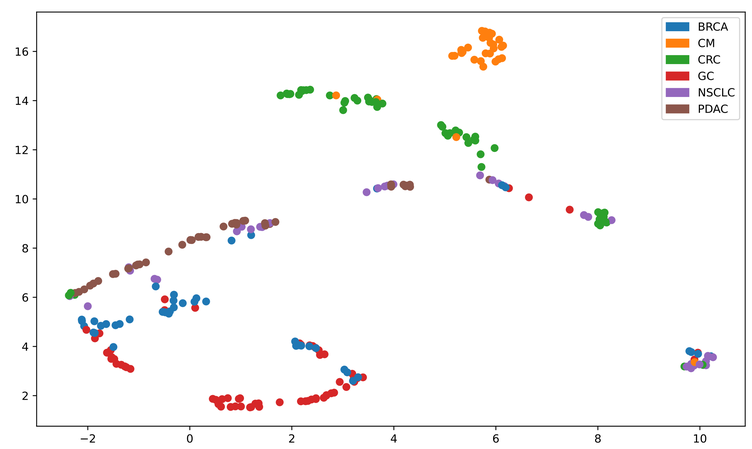


Figure 1. UMAP applied to PDX embeddings.

Prediction of Interpretable Latent States using Multiview AR-HMM

Figure 2. Observed tumor volume (line) with MAP inferred latent state sequence (background color corresponds to 5 discrete states).

Durations of Latent States Are Associated with mRECIST Response Categories

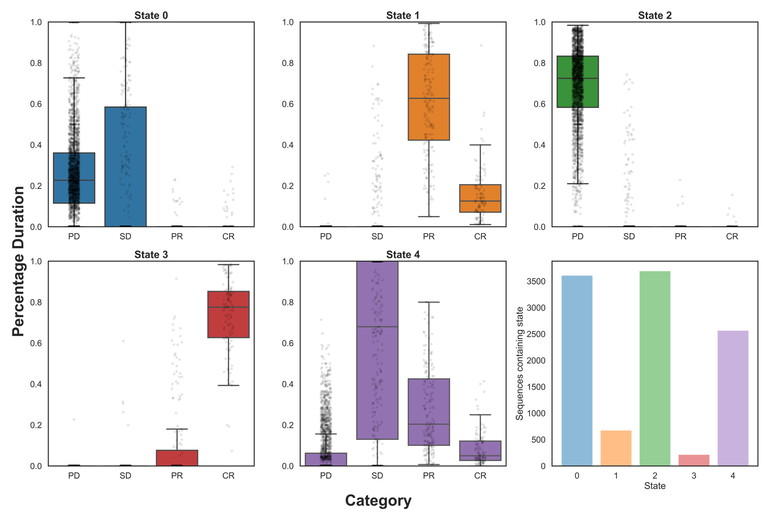


Figure 3. Boxplots show percentage of total time spent in each latent state in PDX curves by response category. Bar chart shows number of sequences predicted to contain each state.

Learned PDX Embeddings from Joint Inference of Embeddings and Time Series

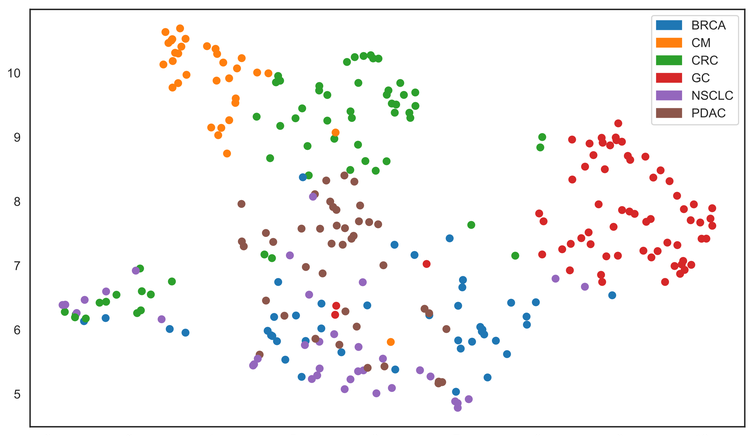


Figure 4. UMAP applied to PDX embeddings.

Simulation of Growth Curves for PDX/Drug Combinations

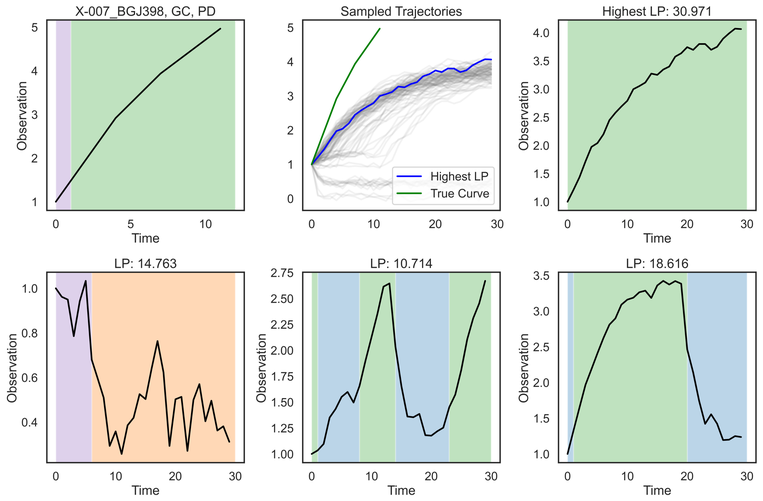


Figure 5. Top row: Observed tumor volume with inferred states, top 100 log probability (LP) sequences from 2000 samples, predicted observation with highest LP with inferred states. Bottom row: draws from the model with lower LP.

Conclusions

- Unsupervised learning of discrete latent states lends insight into disease progression and response.
- Joint learning of time series dynamics and sample-specific embeddings enables simulation and prediction of tumor growth curves.

Limitations and Future Directions

- High stochasticity in sampled sequences makes prediction less feasible.
- Test performance with holdout and robust cross-validation across the whole dataset.
- Enforcing a death state to model survival probability.
- Integration of mutational and transcriptomic data.

References

1. Gao, H., Korn, J., Ferretti, S. *et al.* (2015). High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat Med* **21**, 1318–1325.
2. Linderman, S. W., Miller, A. C., Adams, R. P., Blei, D. M., Paninski, L., & Johnson, M. J. (2016). Recurrent switching linear dynamical systems.

Acknowledgments

Tansey Lab, CBSP (Tri-Institutional PhD Program)
Special Thanks: Jeff Quinn, Yiwei Gong, Tom Dougherty