# PhySigs: Phylogenetic Inference of Mutational Signature Dynamics

Sarah Christensen<sup>1</sup>, Mark D.M. Leiserson<sup>2</sup>, and Mohammed El-Kebir<sup>1</sup>

<sup>1</sup>Dept. of CS, University of Illinois at Urbana-Champaign <sup>2</sup>Dept. of CS, University of Maryland College Park



PSB 2020



#### Signature Exposures at the Tumor Level



#### Intra-tumor Heterogeneity



[Schwartz and Schäffer, Nat Rev Genetics, 2017]

### Heterogeneity in Signatures in Tumor Clones?

A clone can be distinguished from its parent by the set of newly introduced mutations.

Idea is to look at differences in exposures for these *newly introduced* mutations across the tree.



#### Previous: Single Exposure Inference

sig. 1 sig. 2 ··· sig. 30



#### Previous: Single Exposure Inference

sig. 30



every clone

Alexandrov et al., 2013; Rosenthal et al., 2016 Huang et al., 2017;Blokzijl et al,. 2018

#### Previous: Independent Exposure Inference



Alexandrov et al., 2013; Rosenthal et al., 2016 Huang et al., 2017;Blokzijl et al, 2018 McPherson et al., 2016; Jamal-Hanjani et al., 2017

#### This Work: Tree Constrained Exposure Inference



Alexandrov et al., 2013; Rosenthal et al., 2016 Huang et al., 2017;Blokzijl et al,. 2018 McPherson et al., 2016; Jamal-Hanjani et al., 2017

# PhySigs

Problem Statement and Methodology

#### Tree-constrained Exposure Problem

**Problem 3 (Tree-constrained Exposure (TE)).** Given feature matrix P, corresponding count matrix C, signature matrix S, phylogenetic tree T and integer  $k \ge 1$ , find relative exposure matrix D such that  $||P - SDC||_F$  is minimum and D is composed of k sets of identical columns, each corresponding to a connected subtree of T.

*P*: Mutation Count Matrix





### PhySigs Algorithm

Step 1: Solve TE Problem for each possible number *k* of clusters



### PhySigs Algorithm

Step 1: Solve TE Problem for each possible number *k* of clusters





*One Shift (k=2)* 



Two Shifts (k=3)



### PhySigs Algorithm

Step 1: Solve TE Problem for each possible number *k* of clusters

Step 2: Choose best number *k* of clusters using Bayesian Information Criterion (BIC)







Lowest BIC Selected

Two Shifts (k=3)



## Results

On Simulated and Biological Data

## PhySigs on Lung Cancer Cohort

- 91 patient tumors
- Number of clones per patient ranges from 2 to 15 (median of 5).
- Number of equally likely trees per patient ranges from 1 to 17 (median of 1).



Data from [Jamal-Hanjani et al., 2017]

#### PhySigs Explains Data without Overfitting



Data from [Jamal-Hanjani et al., 2017]

#### PhySigs Finds Explainable Shift for Patient CRUK0064



Data from [Jamal-Hanjani et al., 2017]

#### PhySigs Constrains Solutions for Patient CRUK0025



#### PhySigs Constrains Solutions for Patient CRUK0025



#### Conclusions and Discussion

Key concept: mutational signature exposure may not be constant across clones due to intra-tumor heterogeneity

PhySigs identifies shifts along edges of a tumor's evolutionary tree

• May want to consider additional patterns for shifts

PhySigs works by reducing to single exposure problem and then can be solved with existing algorithms

• Hardness of tree constrained exposure for a fixed k remains open

Availability: https://github.com/elkebir-group/PhySigs

## Acknowledgements

#### El-Kebir lab:

- Chuanyi Zhang
- Jiaqi Wu
- Juho Kim
- Leah Weber
- Nuraini Aguse
- Sarah Christensen
- Yerong Li
- Yuanyuan Qi





#### National Science Foundation (IIS 15-13629 to SC, CCF 18-50502 to MEK)