Phylogenies Derived from Matched Transcriptome in Breast Cancer Brain Metastases

Yifeng Tao[1,2], Haoyun Lei[1,2], Adrian V. Lee[3], Jian Ma[1], Russell Schwartz[1,4]

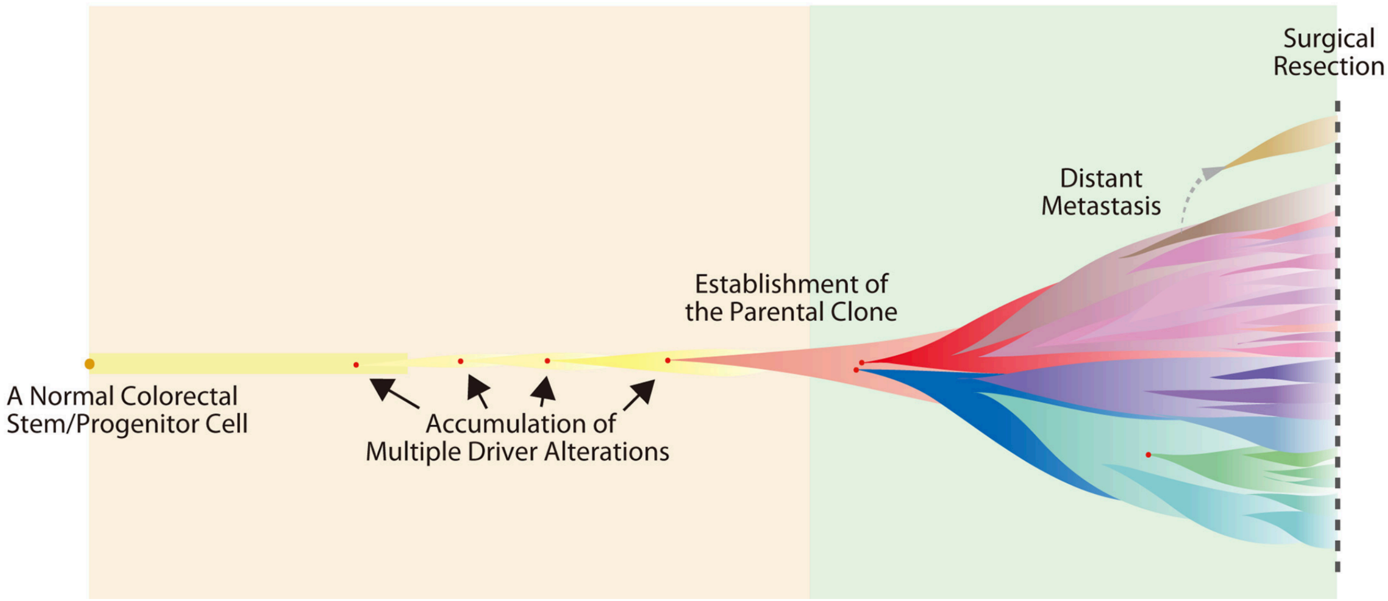
[1]Computational Biology Department, School of Computer Science, Carnegie Mellon University

[2]Joint Carnegie Mellon-University of Pittsburgh Ph.D. Program in Computational Biology

[3]Department of Pharmacology and Chemical Biology, UPMC Hillman Cancer Center, Magee-Womens Research Institute, University of Pittsburgh

[4]Department of Biological Sciences, Carnegie Mellon University

Background: Cancer Progression

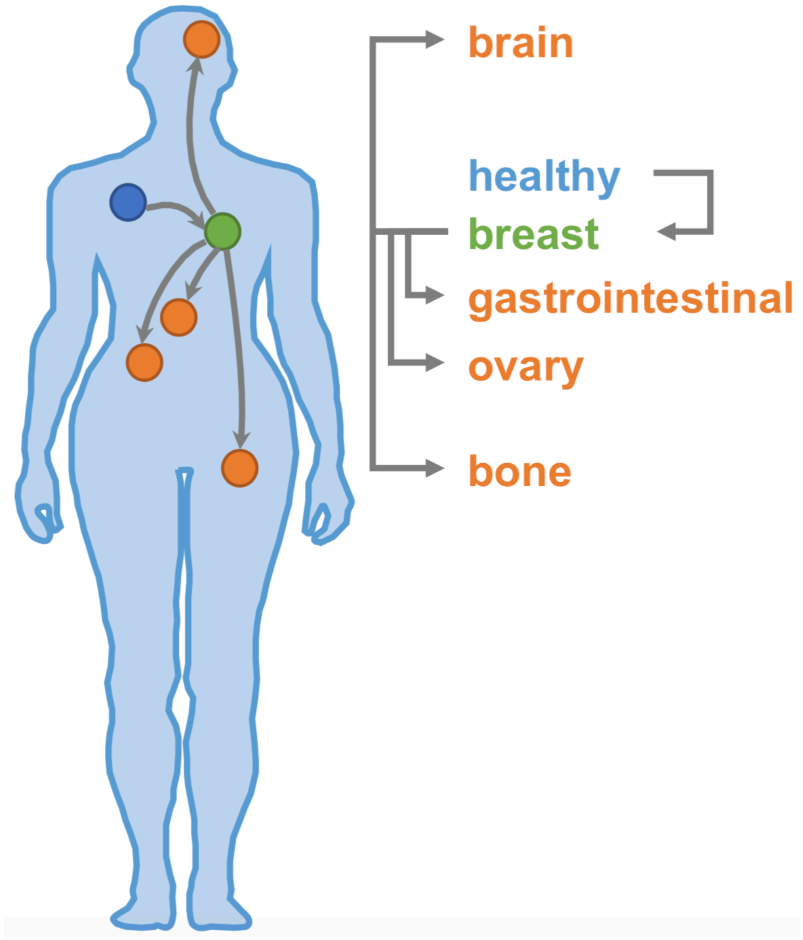


[Uchi, R. et al., *PLOS Genetics*. 2016]

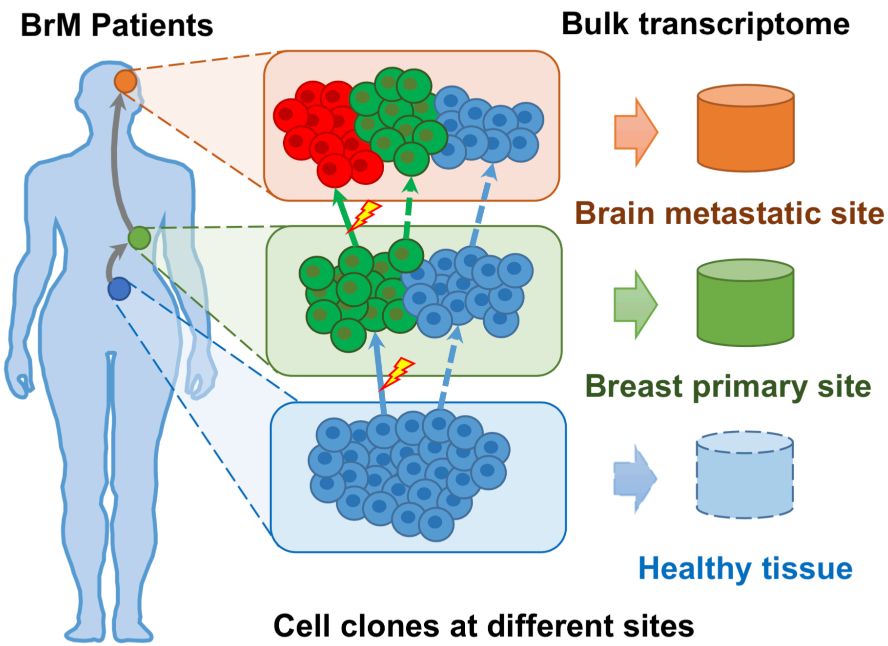
- Cancer: mainly caused by accumulated somatic alterations
- Tumor cells: heterogeneous populations/clones
- Tumor phylogeny: tumor cells follow a clonal evolution
- Metastasis: transfer from primary site to other sites
- Cell communities vs. cell clones

Background: Breast Cancer Metastasis

- Breast cancer: 2nd common cause of death from cancer in women
- Metastatic breast cancer
  - Causes majority of those deaths
  - Limited viable treatment options
  - Early detection is important
- Mechanism of tumor progression/evolution during metastasis?



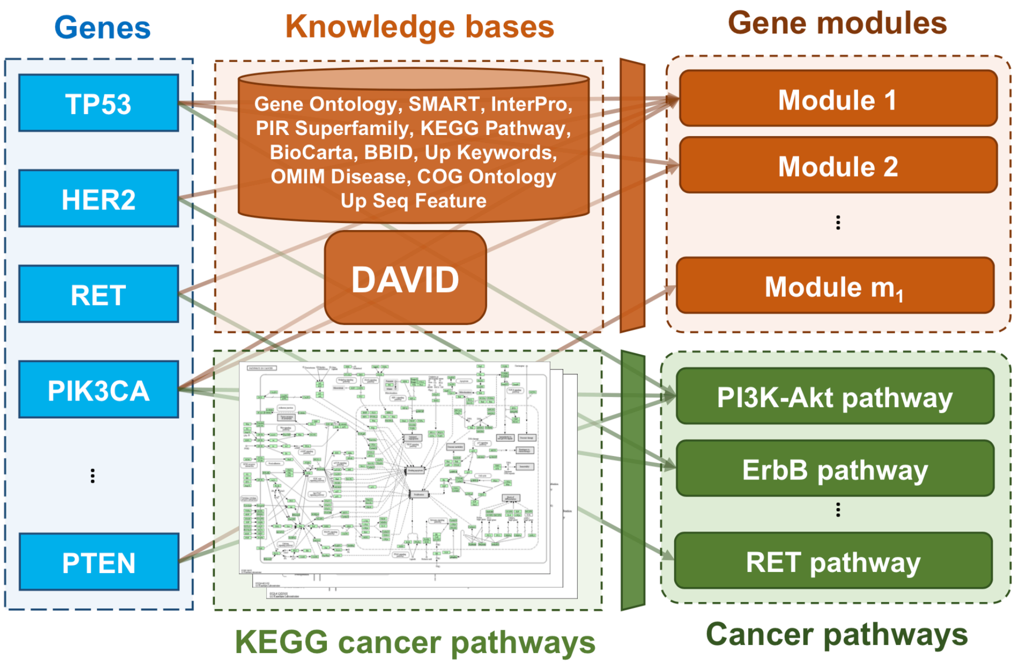
Tumor Evolution Derived from Match Bulk Transcriptome



[Vareslija, D. et al., *Journal of the National Cancer Institute*. 2018]

- Given matched primary and metastatic bulk transcriptome:
  - Q1: How to cope with high-dim, noisy, and uninformative transcriptome?
  - Q2: What model and solver to unmix/deconvolve clones?
  - Q3: How to infer evolutionary trajectory and perturbed pathways/functions?
- Yes! We proposed a three-step pipeline.

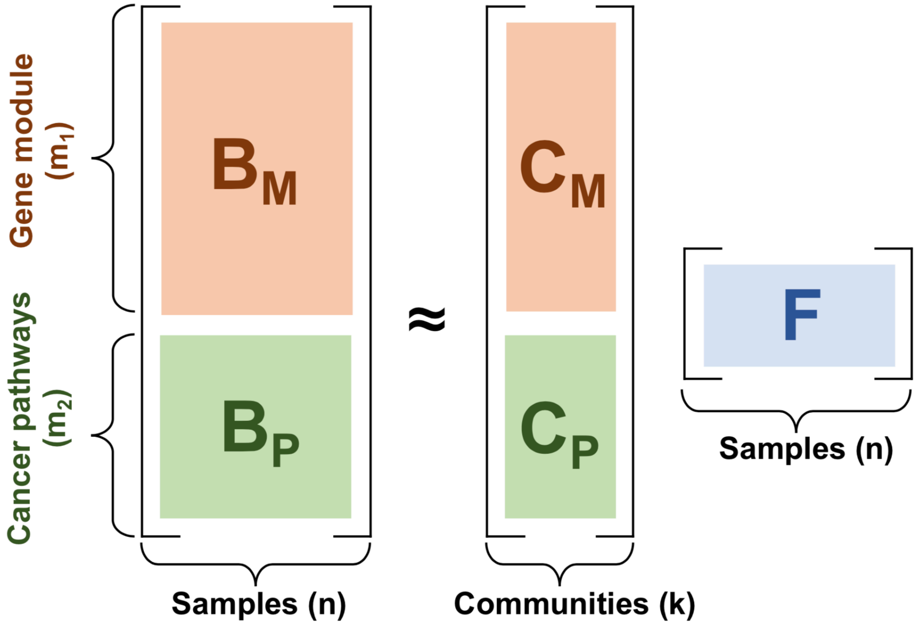
Step 1: Mapping to Gene Modules and Cancer Pathways



[Desmedt, C. et al., *Clinical Cancer Research*. 2008]

- Q1: How to cope with high-dim, noisy, and uninformative RNA?
  - Gene modules
    - Compress high dimensional and noisy data → accurate deconvolution
  - Cancer pathways
    - Markers/probes → interpretation purpose

Step 2: Deconvolution of Cell Communities



$$\min_{\mathbf{C},\mathbf{F}} \quad \|\mathbf{B} - \mathbf{CF}\|_{\mathrm{Fr}}^2 ,$$

$$\text{s.t.} \quad \mathbf{F}_{lj} \geq 0,$$

$$\sum_{l=1}^{k} \mathbf{F}_{lj} = 1$$

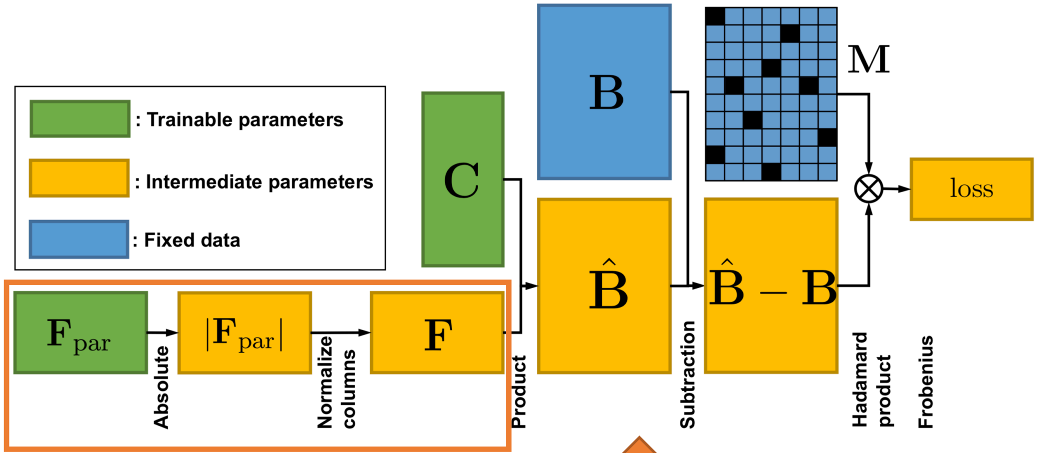
[Lee, D.D. et al., *NIPS*. 2000]

- Q2: What model to unmix/deconvolve clones?
  - Matrix factorization
  - C: expression profiles of communities
  - F: fractions of communities in samples
- However, it is non-convex and not trivial to solve…

Step 2: Deconvolution of Cell Communities

Gradient descent
by backpropagation

[Rumelhart, D.E. et al., *Nature*. 1986]



$$\min_{\mathbf{C},\mathbf{F}} \quad \|\mathbf{B} - \mathbf{CF}\|_{\mathrm{Fr}}^2 ,$$

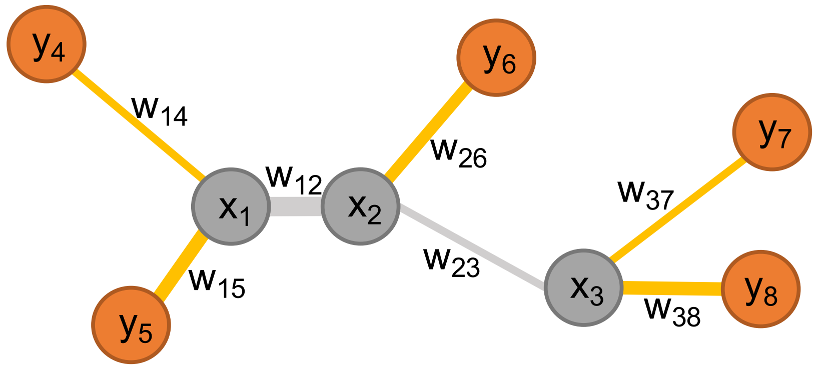
$$\text{s.t.} \quad \boxed{\begin{array}{l} \mathbf{F}_{lj} \geq 0, \\ \sum_{l=1}^{k} \mathbf{F}_{lj} = 1 \end{array}}$$

$$\min_{\mathbf{C},\mathbf{F}_{\mathrm{par}}} \quad \|\mathbf{B} - \mathbf{CF}\|_{\mathrm{Fr}}^2 ,$$

$$\text{s.t.} \quad \boxed{\mathbf{F} = \mathrm{cwn}\left(|\mathbf{F}_{\mathrm{par}}|\right)}$$

- Q2: What model and solver to unmix clones?
  - Neural network deconvolution (NND)
  - # components: trade-off of model complexity vs. sample size
    - Mask matrix for cross-validation in NND

Step 3: Inference of Phylogeny



$$\mathcal{G} = (\mathcal{V}, \mathcal{E})$$

[Nei, M. et al., *Molecular Biology and Evolution*. 1987]

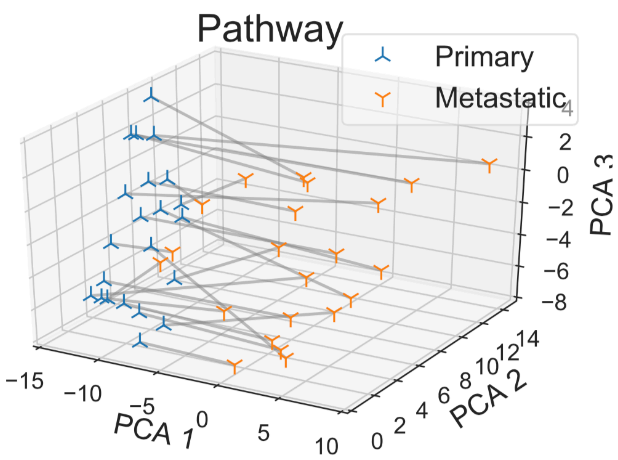
- Q3: How to infer evolutionary trajectory and perturbed pathways?
  - Phylogeny skeleton built using neighbor-joining algorithm
  - Pathway of Steiner nodes inferred by minimizing the elastic potential energy:

$$\min_{\mathbf{x}} \quad U(\mathbf{x}, \mathbf{y};\ \mathcal{W}) = \sum_{\substack{(u,v)\in\mathcal{E} \\ v\leq k-2}} \frac{1}{2} w_{uv}(x_u - x_v)^2 + \sum_{\substack{(u,v)\in\mathcal{E} \\ v\geq k-1}} \frac{1}{2} w_{uv}(x_u - y_v)^2$$

$$\Updownarrow$$

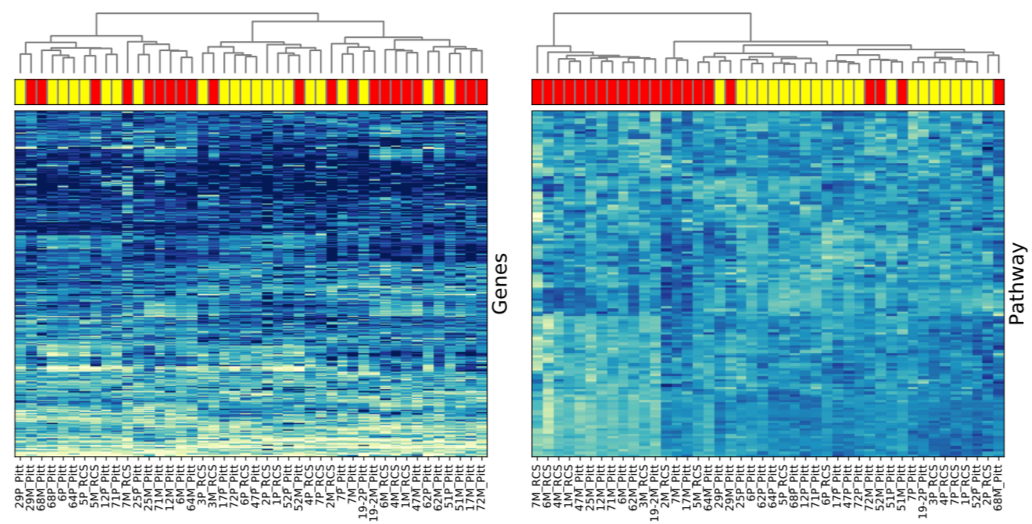
$$\min_{\mathbf{x}} \quad \frac{1}{2}\mathbf{x}^\intercal \mathbf{P}(\mathcal{W})\mathbf{x} + \mathbf{q}(\mathcal{W}, \mathbf{y})^\intercal \mathbf{x}$$

Effective Pathway Representation



Gene expression z_{MSD}=-2.6

Pathway z_{MSD}=-13.4

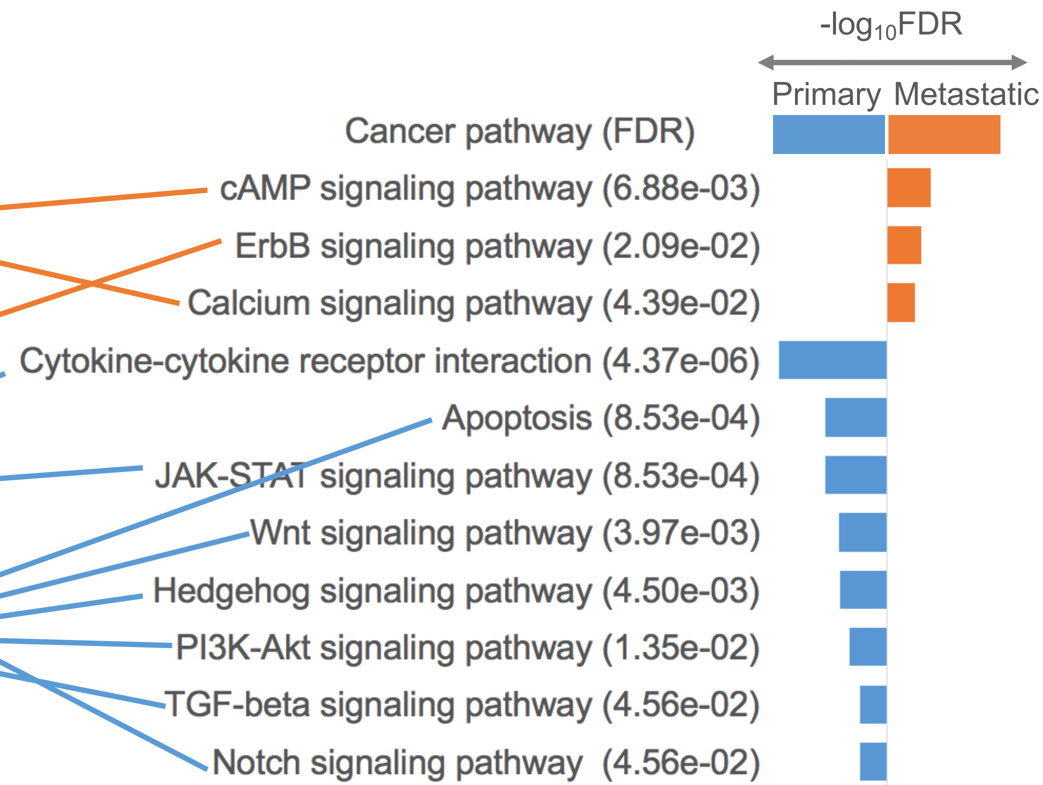


[Park, Y. et al., *Transactions on Computational Biology and Bioinformatics*. 2009]

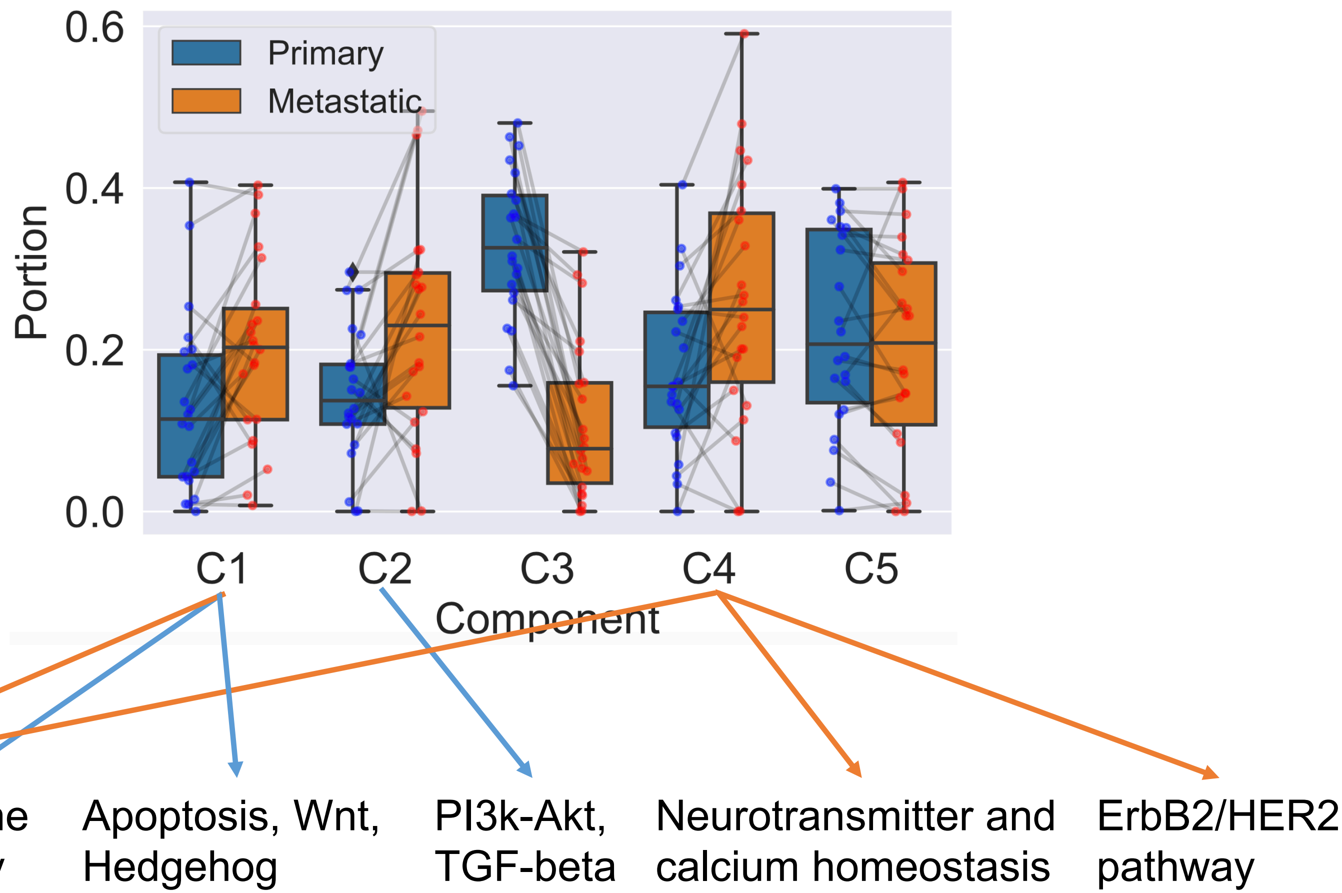
- PC1: recurrent feature between primary and metastatic samples
- PC2+PC3: variability between patients
- Effective in separating primary tumors from metastatic tumors

Differentially Expressed Cancer Pathways

- Neurotransmitter and calcium homeostasis
- ErbB2/HER2 pathway
- Immune activity
- Dysregulation promotes tumor growth

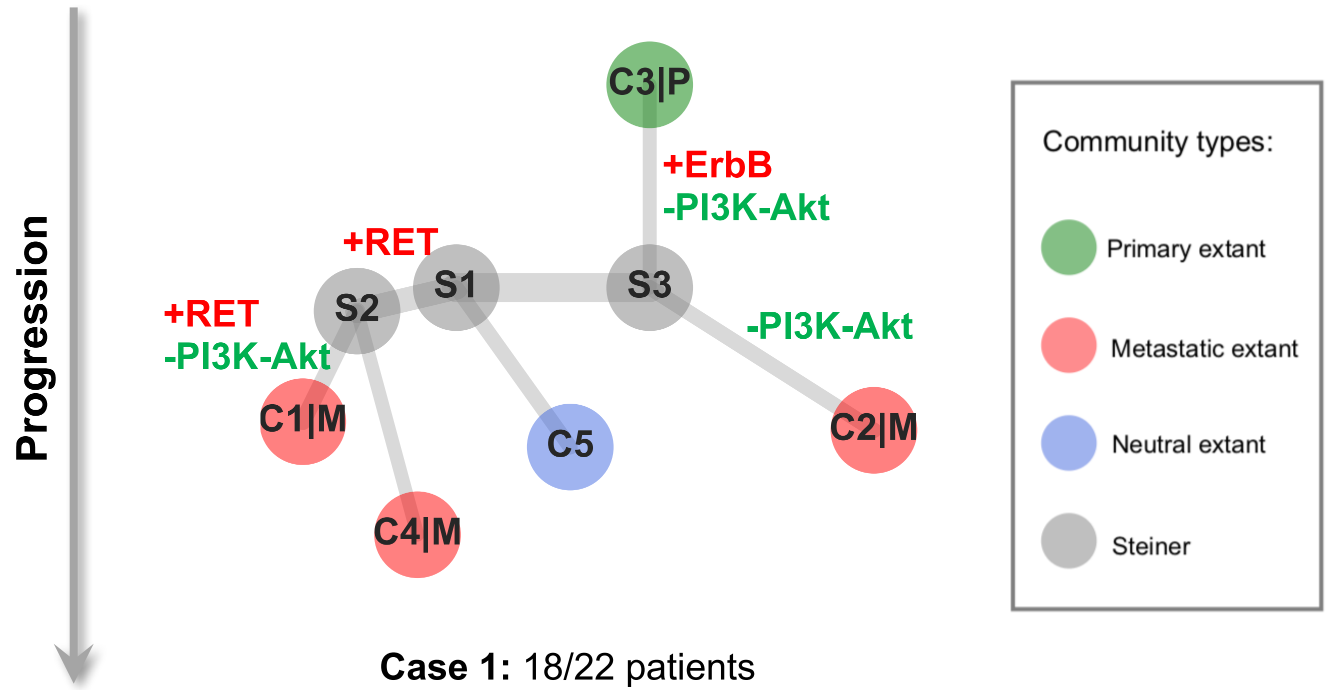


Landscape of Cell Communities



- The deconvolution provides more fine-grained landscape of tumor cell communities

Phylogenies of Cell Communities



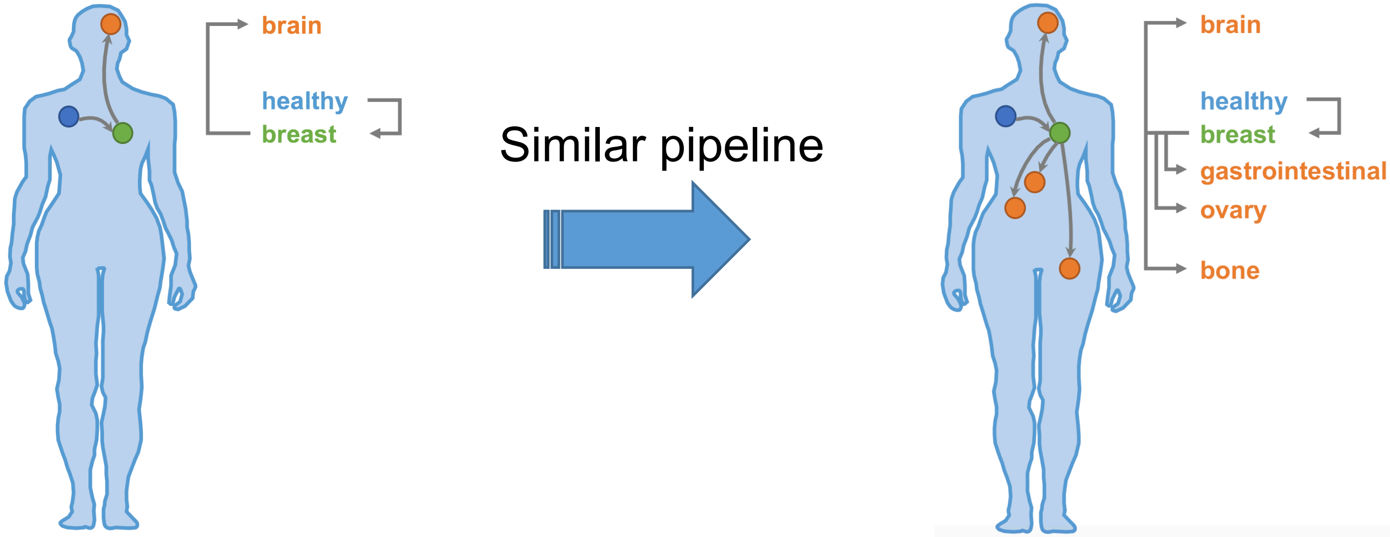
Case 1: 18/22 patients

- Common temporal order of perturbed pathways during metastasis
  - Gained ErbB caused by early events
  - Expansion of minor clonal populations with lost PI3K-Akt and gained RET

Conclusion and Future Work

- Conclusion
  - Pipeline to infer tumor evolution using matched bulk transcriptome
  - Common temporal order of perturbed pathways in breast cancer brain metastases
- Open source code, data and supp:

  https://github.com/CMUSchwartzLab/BrM-Phylo
- Further exploration: multiple metastatic sites



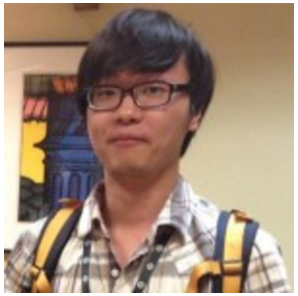
Acknowledgment

○ Authors

Prof. Russell Schwartz

Prof. Jian Ma

Prof. Adrian V. Lee



Haoyun Lei

○ Fundings

NIH National Institutes of Health
Turning Discovery Into Health

BREAST CANCER ALLIANCE

pennsylvania
DEPARTMENT OF HEALTH

Pittsburgh Health Data Alliance
Carnegie Mellon University
University of Pittsburgh
UPMC
Center for Machine Learning and Health
Carnegie Mellon University

Susan G. Komen for the Cure

66 Mario Lemieux Foundation™