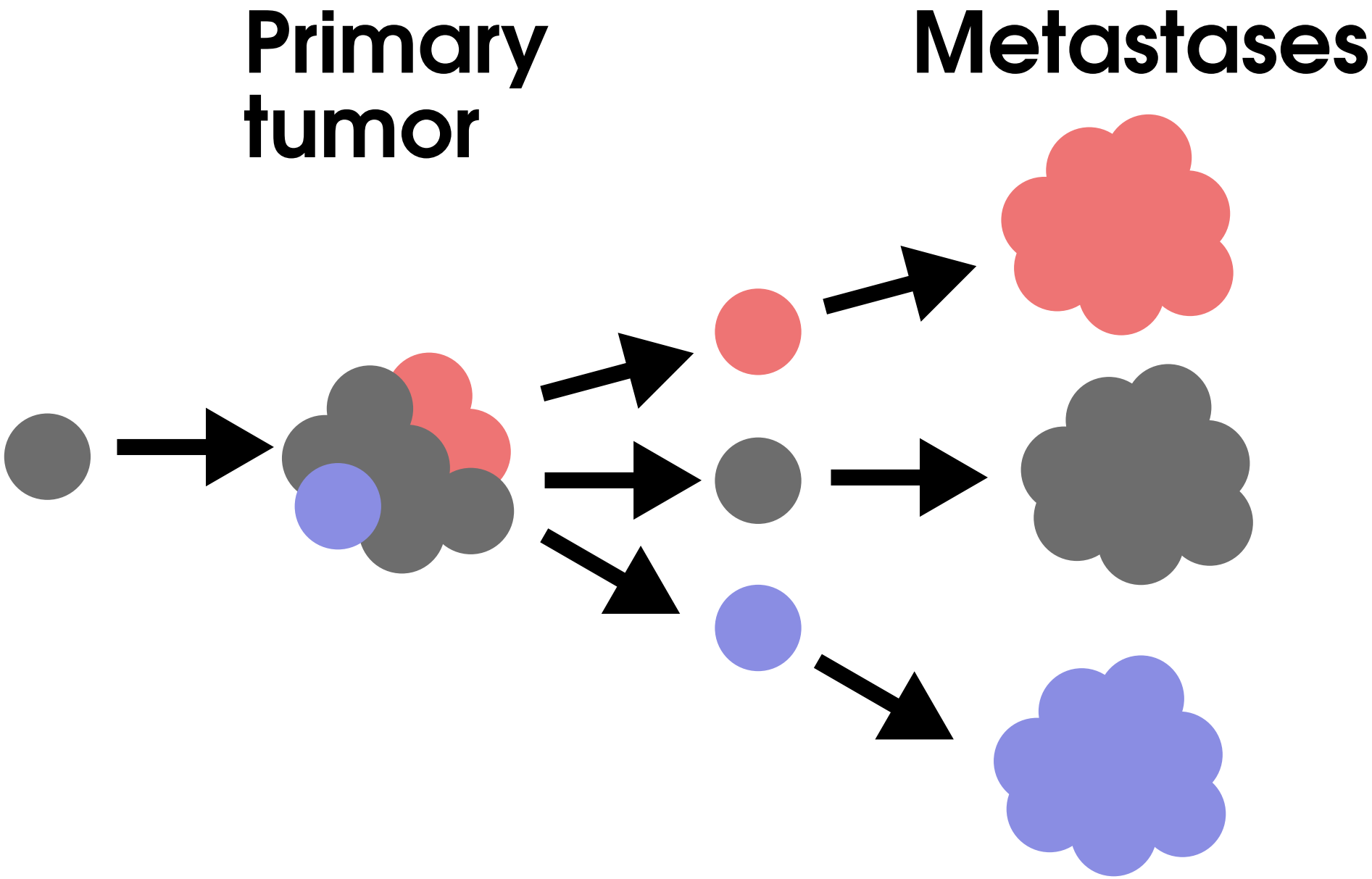
Inferring within patient cancer cell population history from single-cell sequencing data

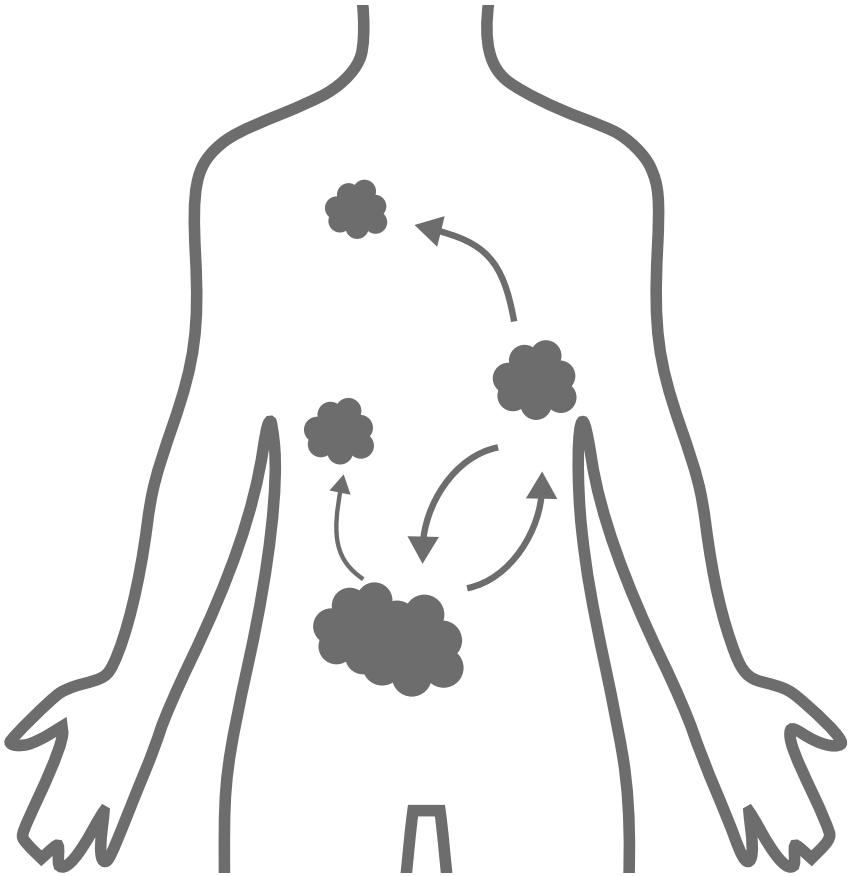
Jiří C. Moravec[1]
R. Bouckaert[2], S. D. Diermeier[3]
A. J. Drummond[2], A. Gavryushkin[1]

[1] Department of Computer Science, University of Otago
[2] Centre for Computational Evolution, University of Auckland
[3] Department of Biochemistry, University of Otago

Motivation



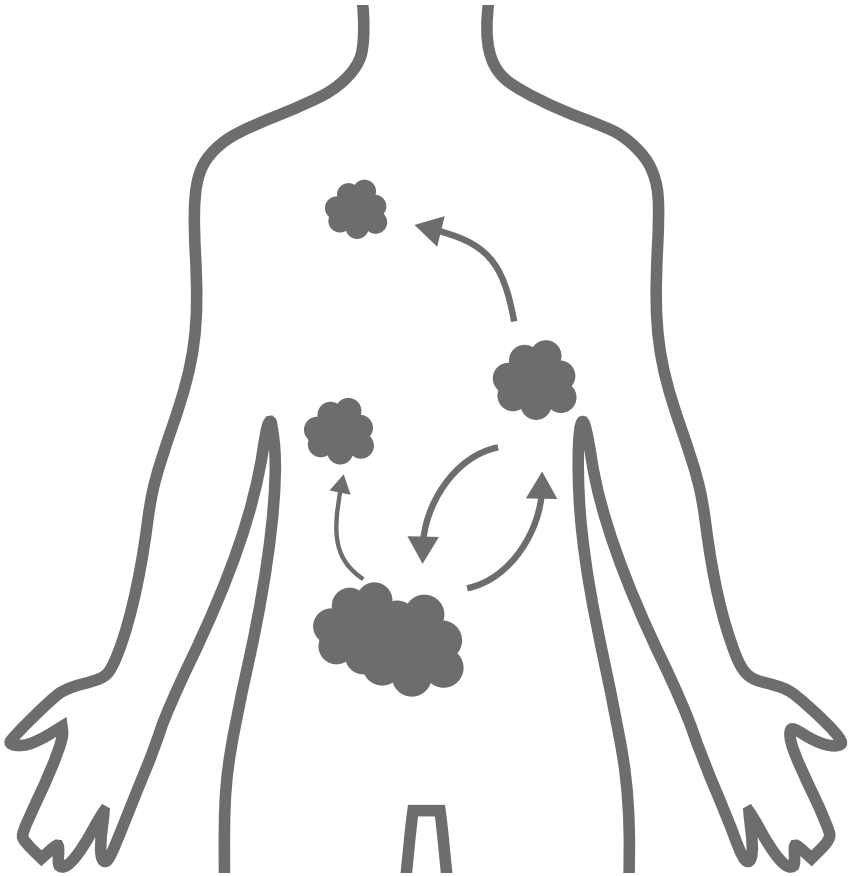
Motivation



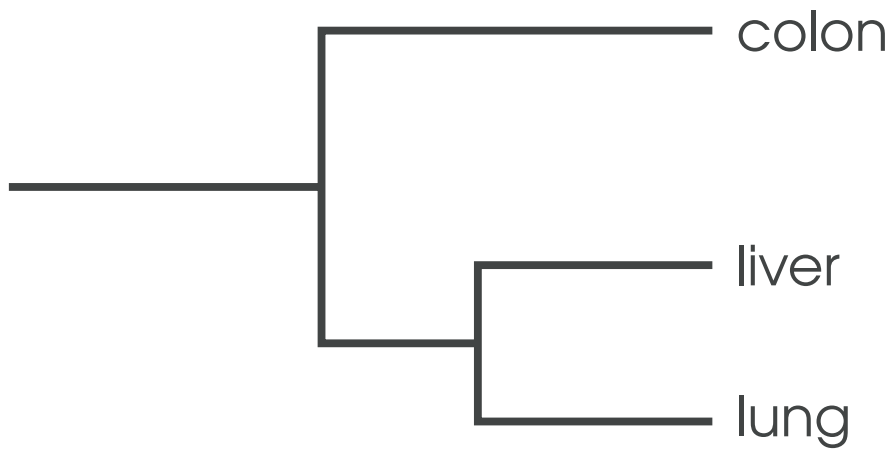
Complex population history of cancer

- multiclonal origin
- self-seeding
- multiple-seeding
- wound theory

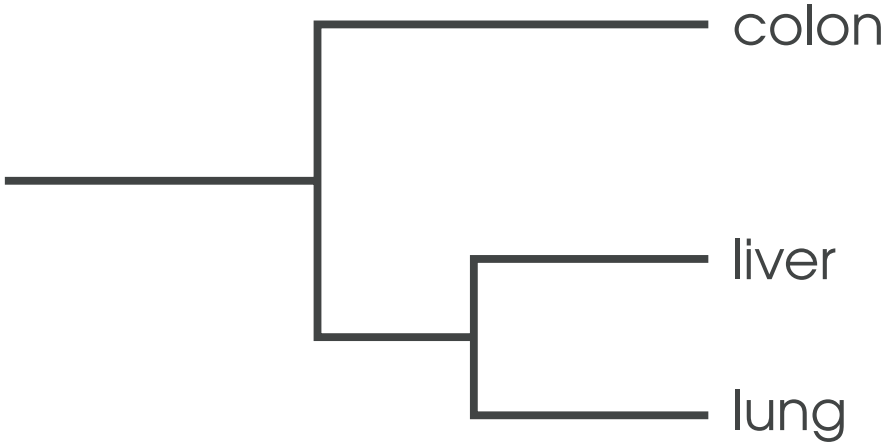
Motivation



The real population history is a tree!



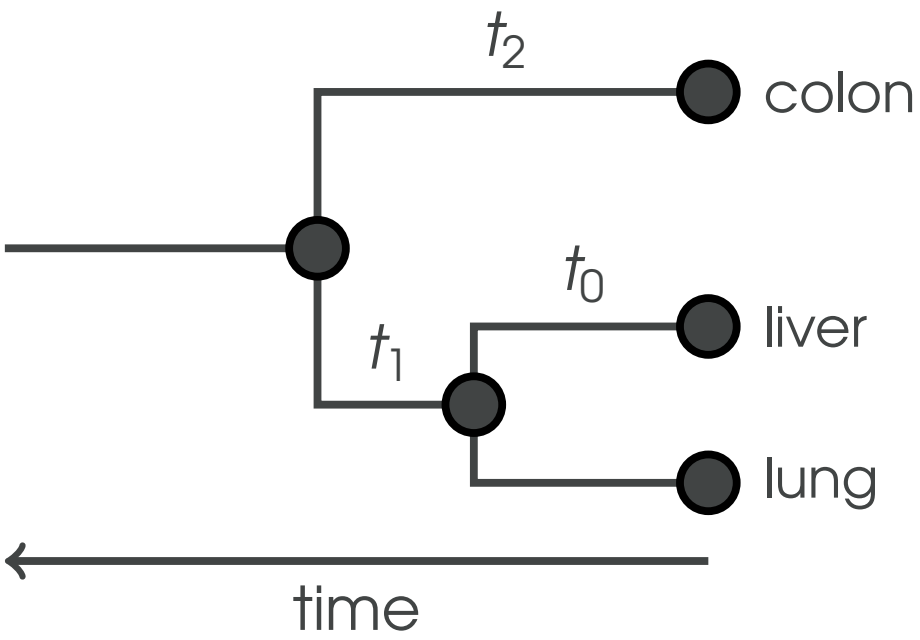
Motivation



Sometimes tree is not just tree

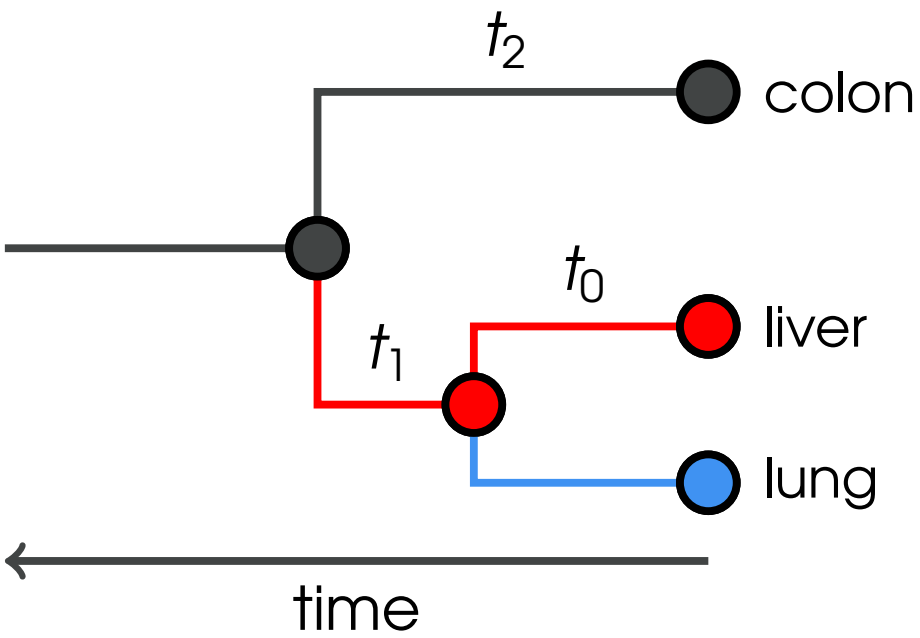
- evolutionary model
- branching process
- mutations
- demographic
- migration

Coalescent theory



- describe merging of lineages in time
- time then used to model mutations on sequence (with HKY, GTR ...)
- basis for other models:
  - migration models (e.g., mugration)
  - population change (e.g., BayesianSkyline)

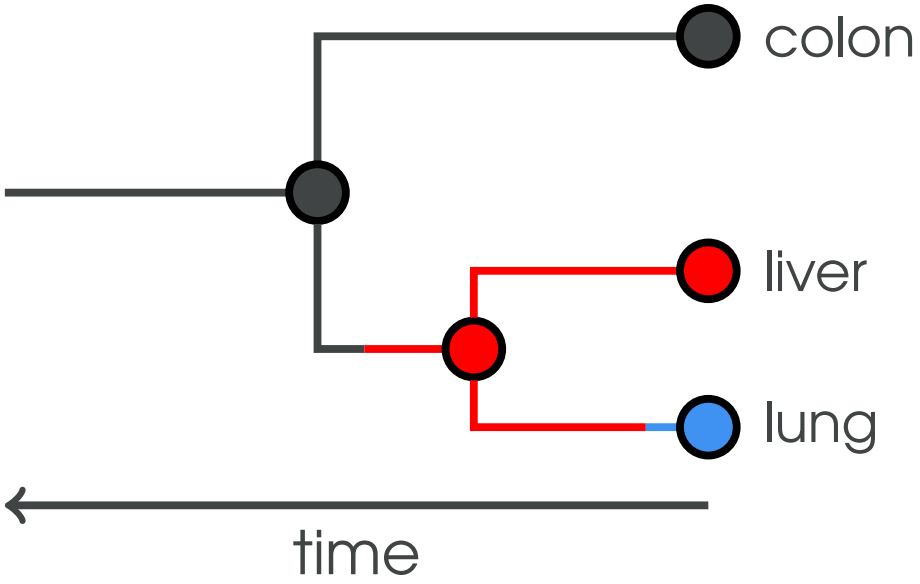
Coalescent theory



No population structure

- migration models are duck-taped on standard coalescent
- population structure can cause spurious signals
- can cause incorrectly estimated topology or migration patterns

Structured coalescent



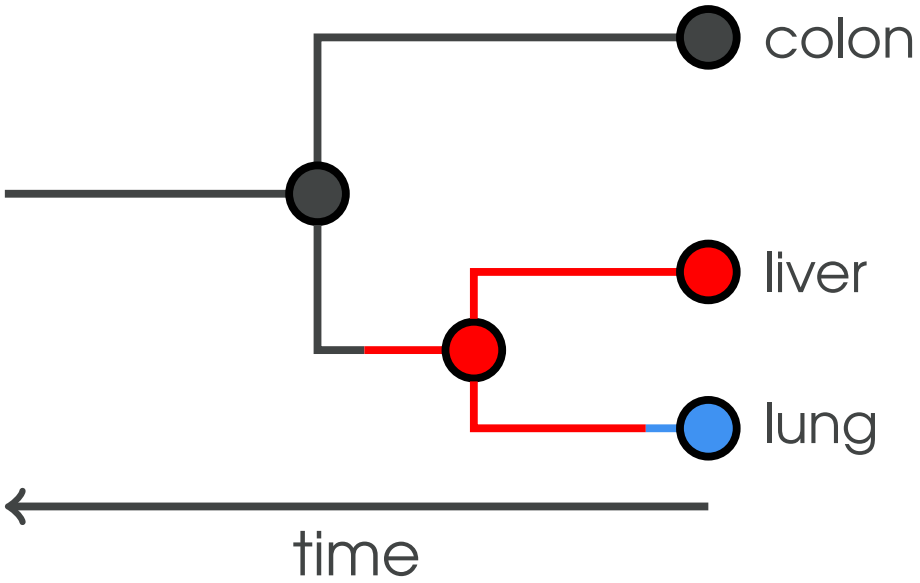
- Takahata (1988) [1], Hudson (1990)[2], Notohara (1990)[3]
- change of location can happen any time
- computationally intensive
- only constant population size

[1] Takahata, N. (1988). The coalescent in two partially isolated diffusion populations. *Genetical Research*, 52(3):213–222

[2] Hudson, R. R. (1990). Gene genealogies and the coalescent process. *Oxford surveys in evolutionary biology*, 7(1):44

[3] Notohara, M. (1990). The coalescent and the genealogical process in geographically structured population. *Journal of Mathematical Biology*, 29(1):59–75

Structured coalescent



Computationally tractable approaches to s.c.

- SISCO (Volz, 2012)[4]
- MultiTypeTree (Vaughan, 2014)[5]
- BASTA (De Maio, 2015)[6]
- MASCO (Müller, 2017)[7]

[4] Volz, E. M. (2012). Complex population dynamics and the coalescent under neutrality. *Genetics*, 190(1):187–201

[5] Vaughan, T. G., Kühnert, D., Popinga, A., Welch, D., and Drummond, A. J. (2014). Efficient Bayesian inference under the structured coalescent. *Bioinformatics*, 30(16):2272–2279

[6] De Maio, N., Wu, C.-H., O'Reilly, K. M., and Wilson, D. (2015). New routes to phylogeography: A bayesian structured coalescent approximation. *PLOS Genetics*, 11(8):1–22

[7] Müller, N. F., Rasmussen, D. A., and Stadler, T. (2017). The Structured Coalescent and Its Approximations. *Molecular Biology and Evolution*, 34(11):2970–2981

Our approach

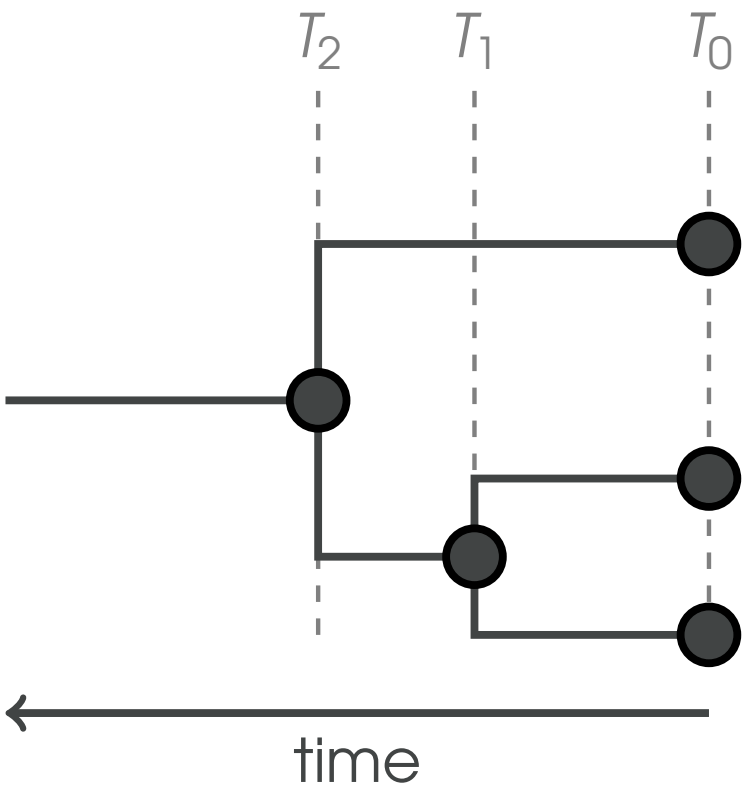
Reconstruct population history of cancer cells:

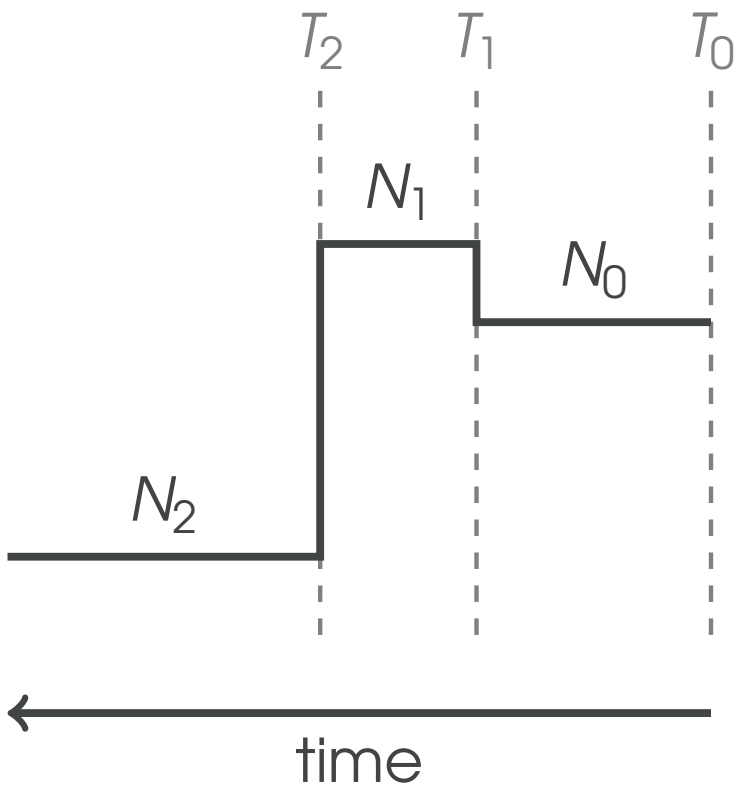
- Structured Coalescent (MASCOT[8] in Beast2)
- with varied population size (Bayesian Skyline[9])
- with model utililzing both DNA and Methylation data

[8] Müller, N. F., Rasmussen, D., and Stadler, T. (2018). MASCOT: parameter and state inference under the marginal structured coalescent approximation. *Bioinformatics*, 34(22):3843–3848

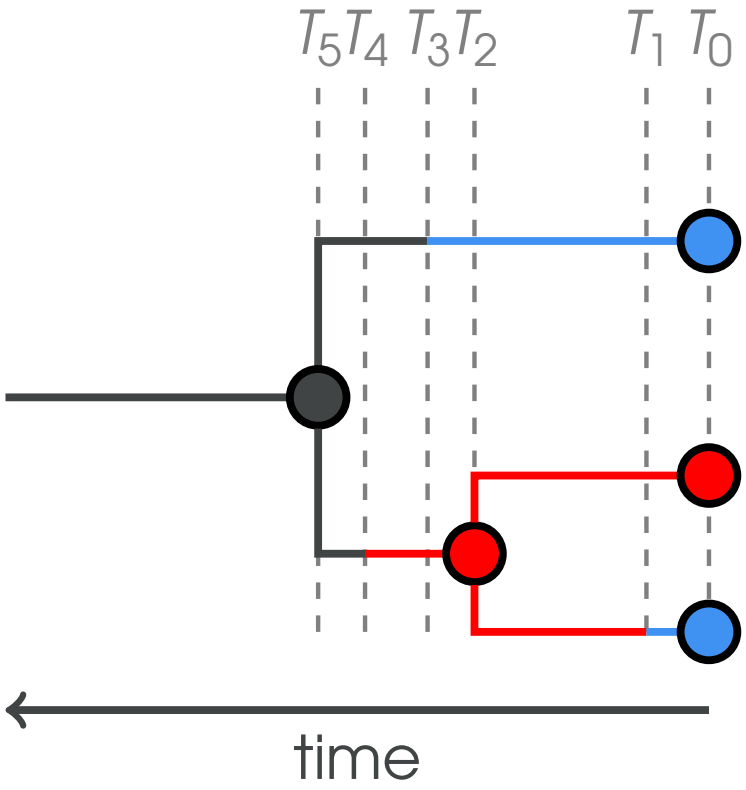
[9] Drummond, A. J., Rambaut, A., Shapiro, B., and Pybus, O. G. (2005). Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. *Molecular Biology and Evolution*, 22(5):1185–1192

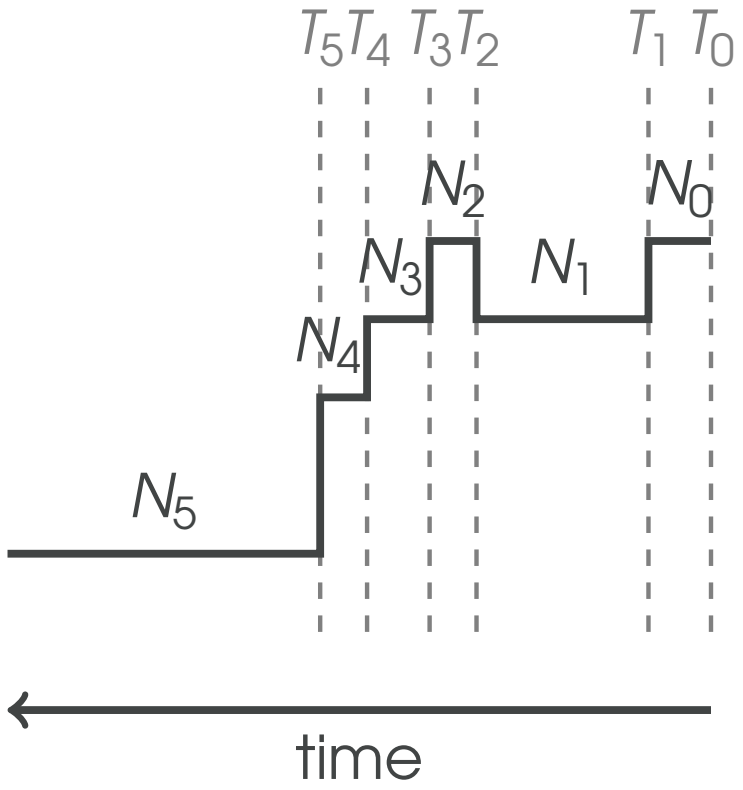
Bayesian Skyline





Bayesian Skyline in s.c.





MethylationHKY

- Combine methylation and DNA information in single matrix
  - DNA provide basic phylogenetic structure
  - methylations then can distinguish closely related cells
- enforce particular pattern of methylation and demethylation

$$
\begin{array}{c} \\ A \\ C \\ G \\ T \\ C' \\ G' \end{array}
\begin{array}{c}
\begin{array}{cccccc} A & C & G & T & C' & G' \end{array} \\
\begin{bmatrix}
- & 1 & \kappa & 1 & 0 & 0 \\
1 & - & 1 & \kappa & \alpha & 0 \\
\kappa & 1 & - & 1 & 0 & \alpha \\
1 & \kappa & 1 & - & 0 & 0 \\
1 & \beta & 1 & \kappa+\gamma & - & 0 \\
\kappa+\gamma & 1 & \beta & 1 & 0 & -
\end{bmatrix}
\end{array}
$$

- 1 – transversion rate
- κ – transition rate
- α – methylation rate from C to MethC
- β – demethylation rate from MethC to C
- γ – demethylation rate from MethC to T

Data

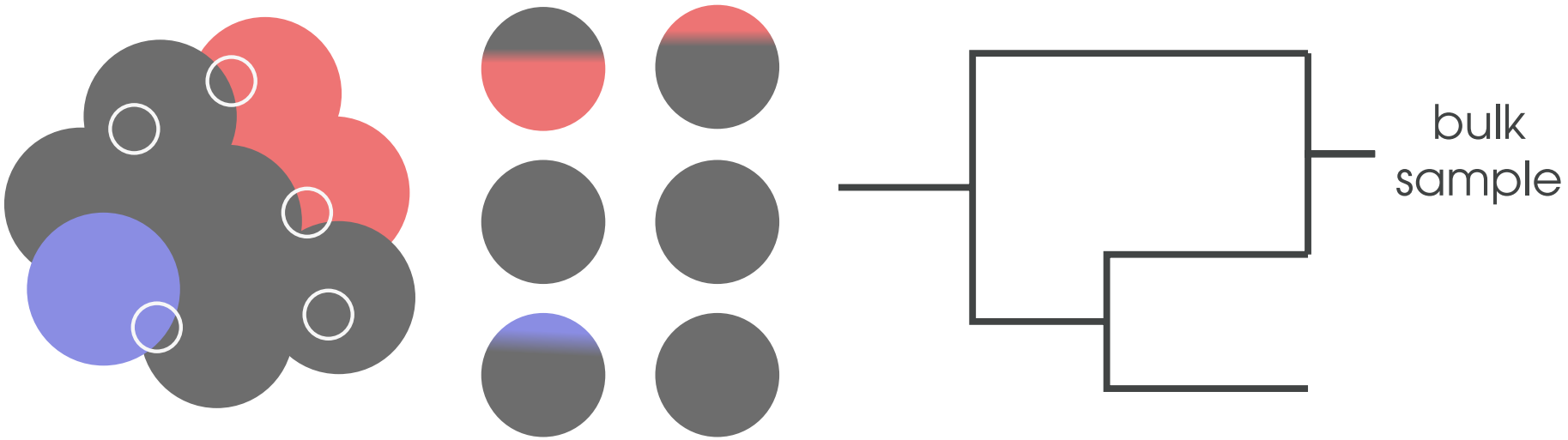
Single cell sequencing

- Advantages:
  - single-origin DNA
  - more data, more structure
  - does not mess up signal
- Disadvantages:
  - harder and more expensive to get
  - sometimes too much data

Bulk sequencing

- Disadvantages:
  - DNA of potentially multiple origin
  - messes up phylogeny
- Advantages:
  - easier to get
  - "cleaner"
  - less data (computationally feasible)

Data



Data

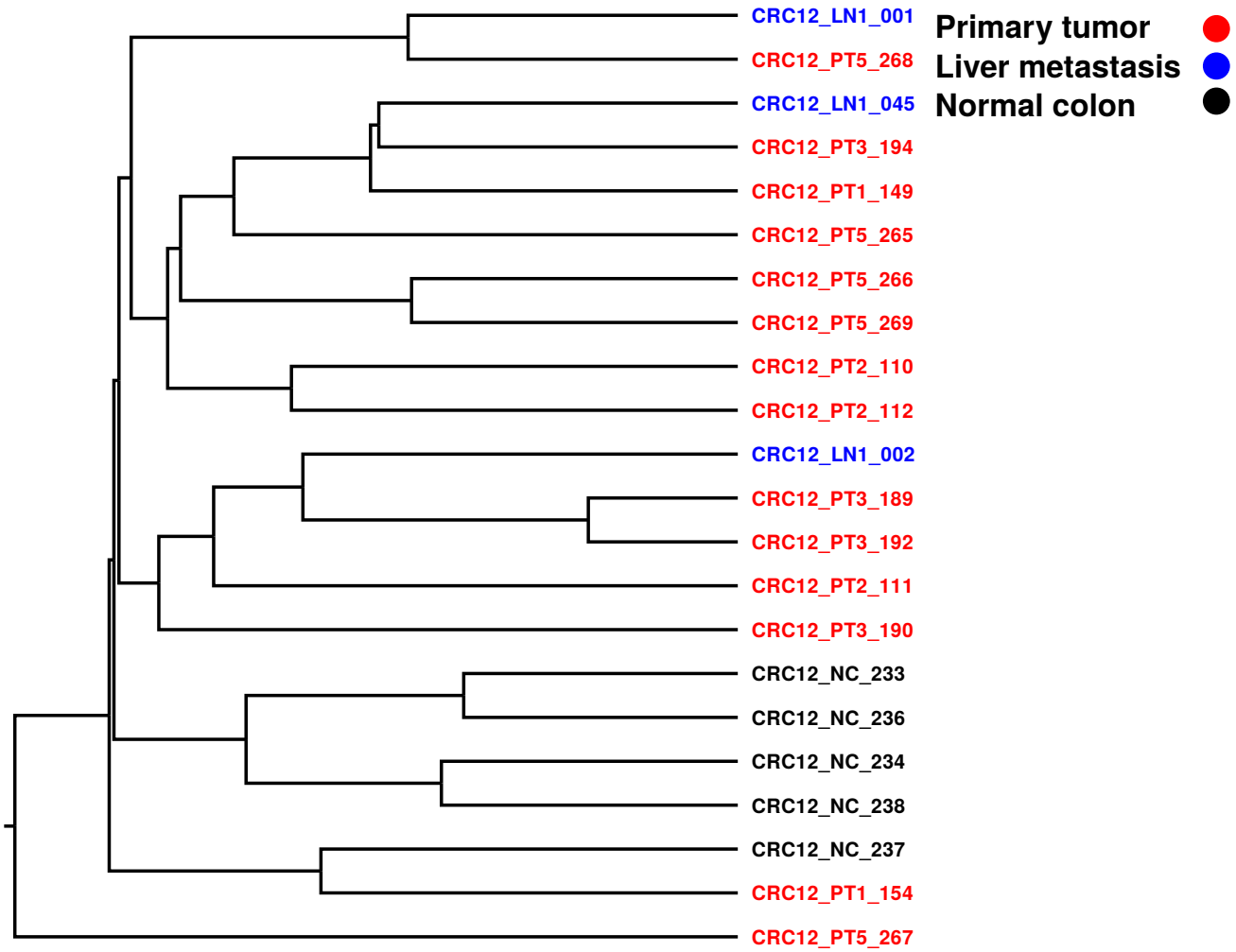
Single cell sequencing

- Bian et al. 2018
- DNA and methylation
- WGS
- multiple patients and multiple regional samples
- not serially sampled

Bulk sequencing

- from Auckland collaborators
- DNA and methylation (but separate)
- SNP
- single patient, 16 samples
- not serially sampled

Work in Progress



Work in Progress

Ideal data:

- Single cell sequencing
- DNA and methylation (with shared positions)
- multiple regional samples
- serially sampled

Summary

Reconstructing demographic history of cancer

- Structured coalescent with Mascot
- Extend SC with Bayesian Skyline
- use MethylationHKY model to utilize both DNA and methylation data

Acknowledgements:

Collaborators:
Cristin Print
Tamsin Robb
Cherie Blenkiron

RUTHERFORD FOUNDATION

ROYAL SOCIETY TE APĀRANGI

SAPERE AUDE

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

References:

Bian, S., Hou, Y., Zhou, X., Li, X., Yong, J., Wang, Y., Wang, W., Yan, J., Hu, B., Guo, H., Wang, J., Gao, S., Mao, Y., Dong, J., Zhu, P., Xiu, D., Yan, L., Wen, L., Qiao, J., Tang, F., and Fu, W. (2018). Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science*, 362(6418):1060–1063.

De Maio, N., Wu, C.-H., O'Reilly, K. M., and Wilson, D. (2015). New routes to phylogeography: A bayesian structured coalescent approximation. *PLOS Genetics*, 11(8):1–22.

Drummond, A. J., Rambaut, A., Shapiro, B., and Pybus, O. G. (2005). Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. *Molecular Biology and Evolution*, 22(5):1185–1192.

Hudson, R. R. (1990). Gene genealogies and the coalescent process. *Oxford surveys in evolutionary biology*, 7(1):44.

Müller, N. F., Rasmussen, D., and Stadler, T. (2018). MASCOT: parameter and state inference under the marginal structured coalescent approximation. *Bioinformatics*, 34(22):3843–3848.

Müller, N. F., Rasmussen, D. A., and Stadler, T. (2017). The Structured Coalescent and Its Approximations. *Molecular Biology and Evolution*, 34(11):2970–2981.

Notohara, M. (1990). The coalescent and the genealogical process in geographically structured population. *Journal of Mathematical Biology*, 29(1):59–75.

Takahata, N. (1988). The coalescent in two partially isolated diffusion populations. *Genetical Research*, 52(3):213–222.

Vaughan, T. G., Kühnert, D., Popinga, A., Welch, D., and Drummond, A. J. (2014). Efficient Bayesian inference under the structured coalescent. *Bioinformatics*, 30(16):2272–2279.

Volz, E. M. (2012). Complex population dynamics and the coalescent under neutrality. *Genetics*, 190(1):187–201.