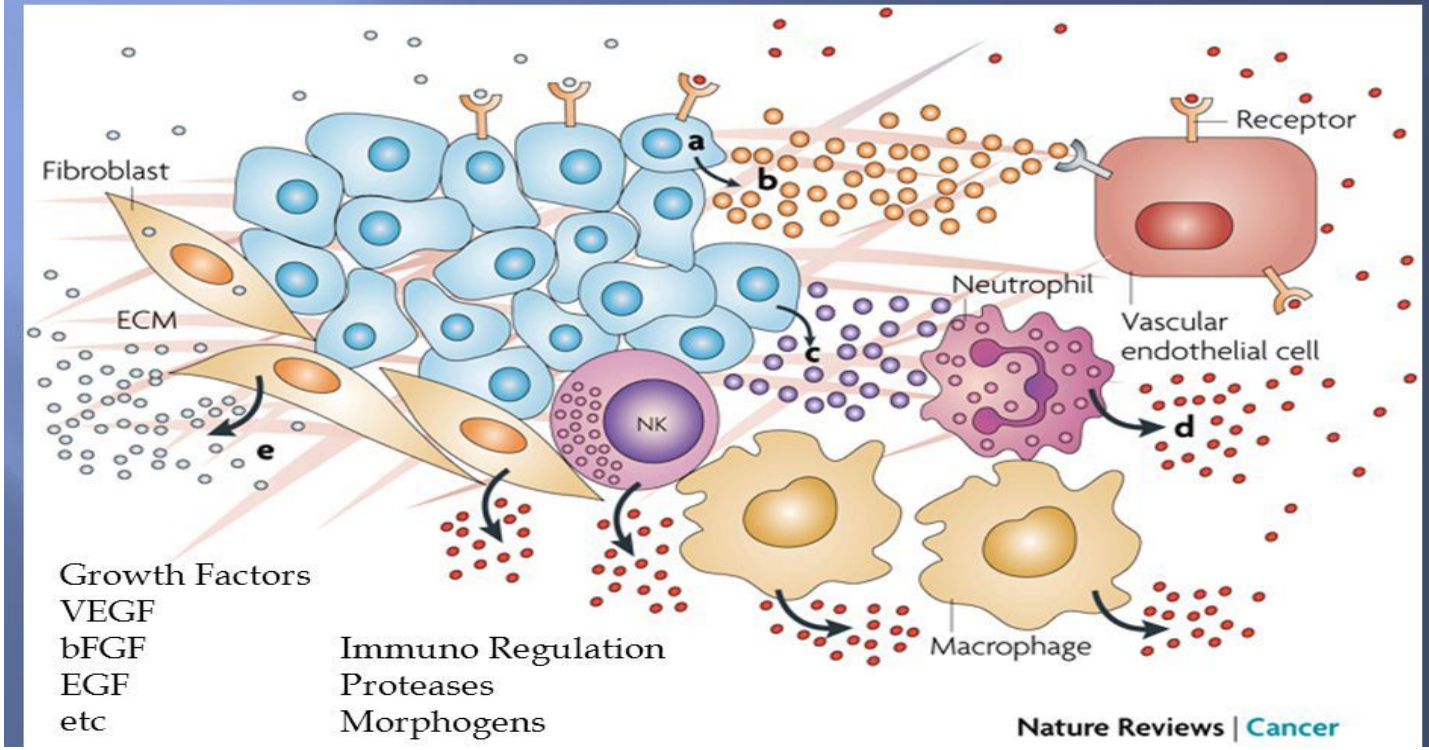
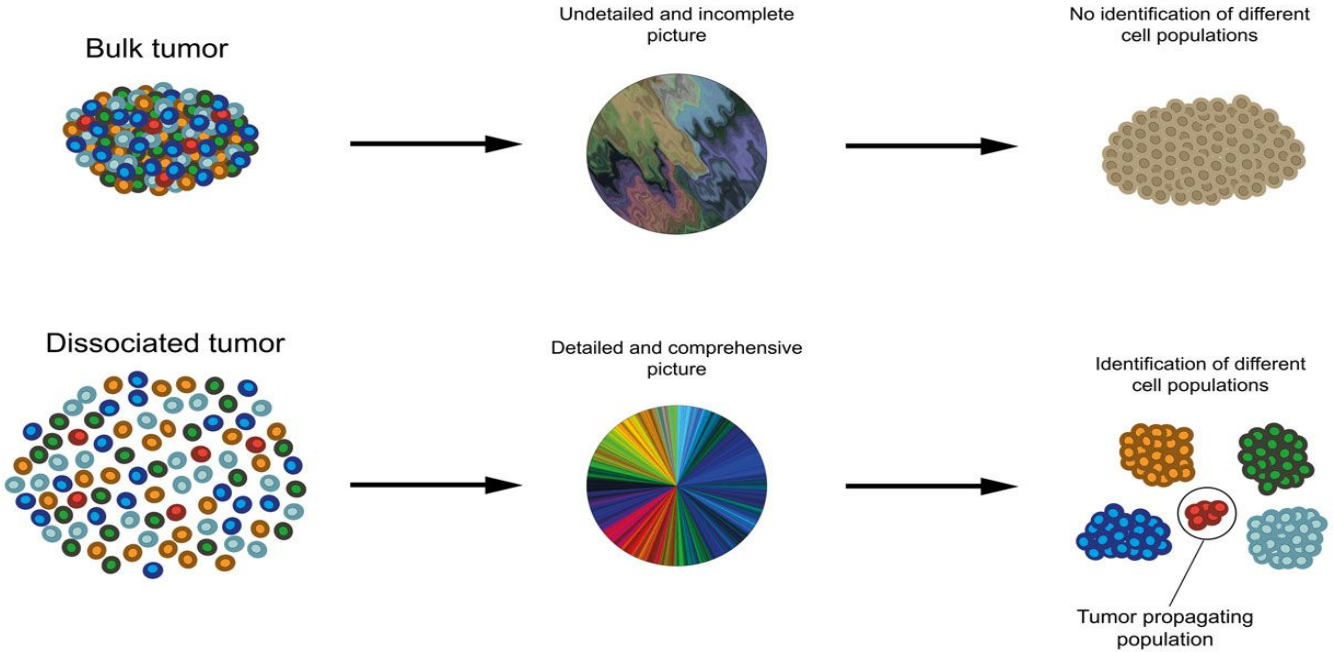
Tumor Phylogeny Simulator

Jesse Eaton
Murtaza Saif
Shefali Umrania

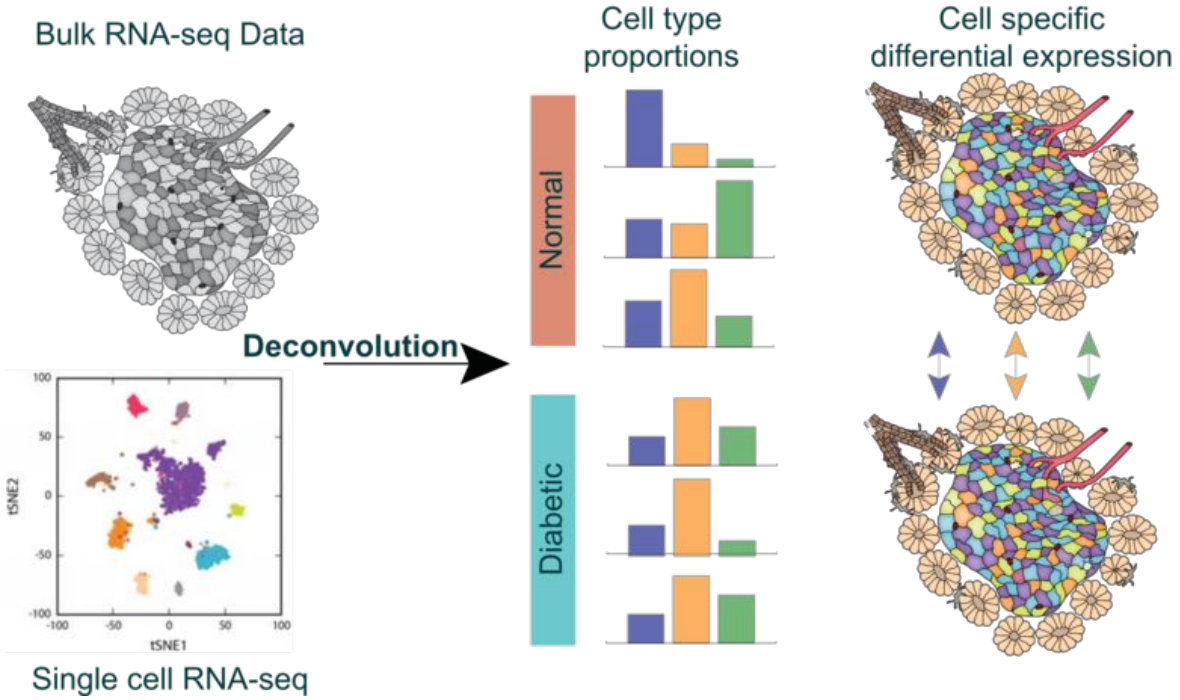
Tumor Microenvironment:
Multiple Cell Types , Factors, Receptors



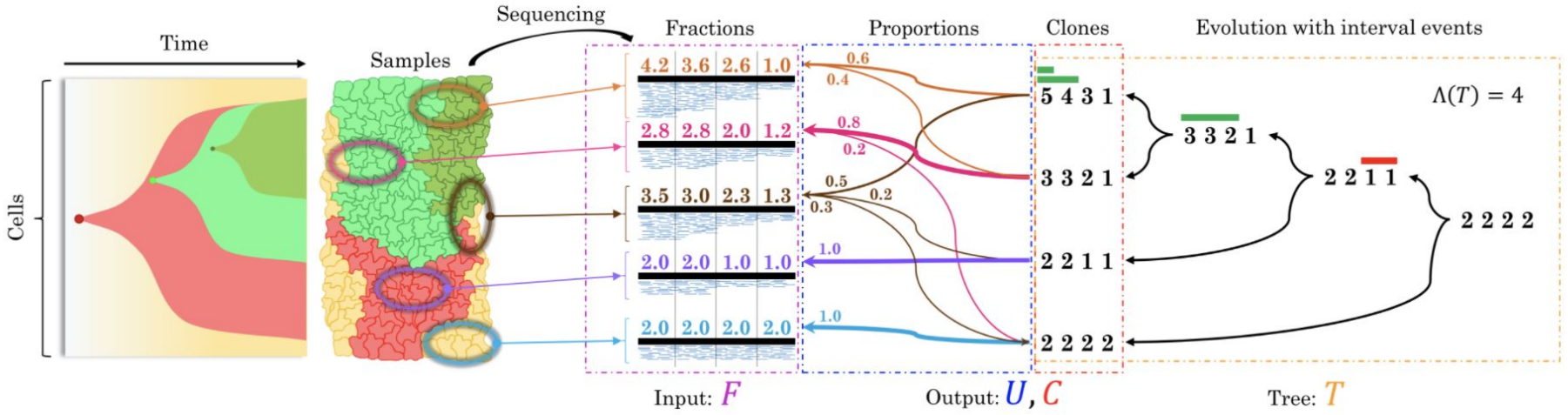
Different sequencing methods



Deconvolution



CNT-MD: Copy Number Tree Mixture Deconvolution



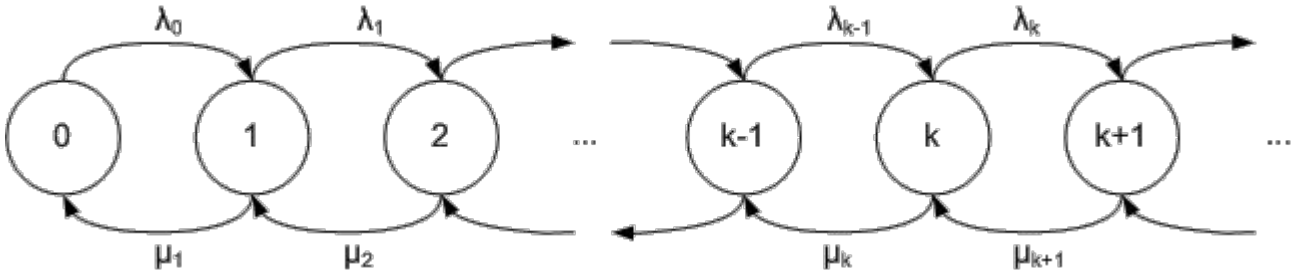
Simone Zaccaria, Mohammed El-Kebir, Gunnar W. Klau, Benjamin J. Raphael: The Copy-Number Tree Mixture Deconvolution Problem and Applications to Multi-sample Bulk Sequencing Tumor Data. RECOMB 2017: 318-335

Problem: Obtaining large amounts of biological mutation sequence data to study tumor phylogenetic evolution

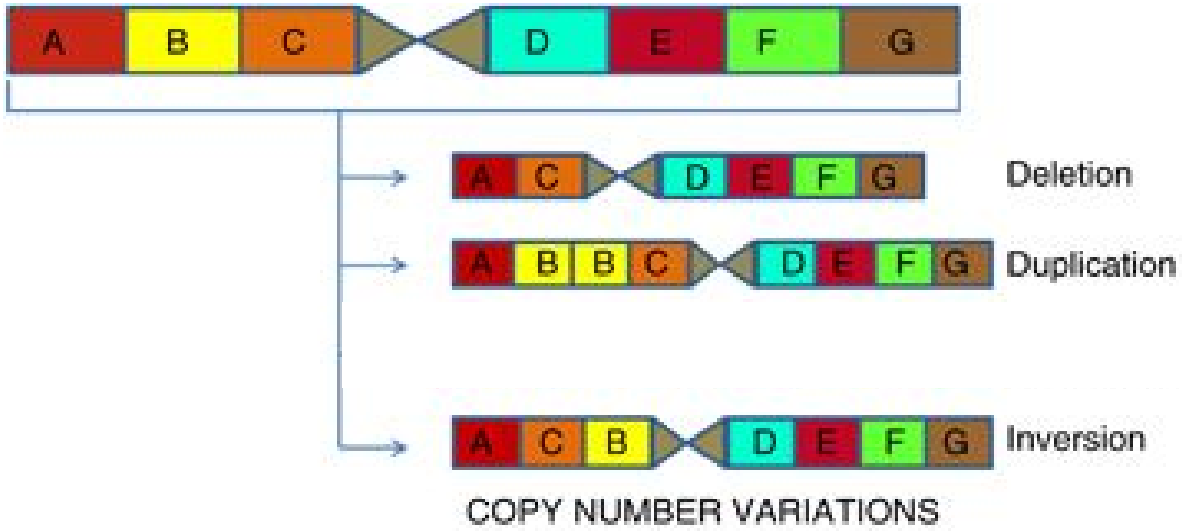
Solution: simulate synthetic mutation datasets

Method:

Create a phylogenetic simulator using a Continuous Time Markov Model to determine WHEN and WHICH mutations occur at each time step.



Simulated Mutations



F = U * C

Fractional Copy number (m x s)	Usage (m x n)	Copy Number (n x s)

m: Number of tumor samples
s: Number of segments
n: Number of cell types

Variables for Dataset Creation

Symbol	Description	Range
ℓ	mean length of a single mutation event	$\ell_{toy} \in \{100, 1000\}$, $\ell_{real} \in \{50000, 100000\}$
m	number of mutations	$m \in \{10, 50, 300\}$
s	number of samples	$s \in \{1, 3, 5, 10\}$
v	variance in probability of generating a perfectly mixed sample	$v \in \{0.002, 0.01, 0.05\}$
α	parameter to Beta distribution	$(\alpha, \beta) \in \{(1, 1), (5, 1)\}$
β	parameter to Beta distribution	

Table 1: Parameters for large generated data set. Draw from beta distribution determines the percent of cells that will mutate at each mutation. Smaller α (and larger β increases population of normal children while smaller β (and larger α) increase the population of mutated children.

Results

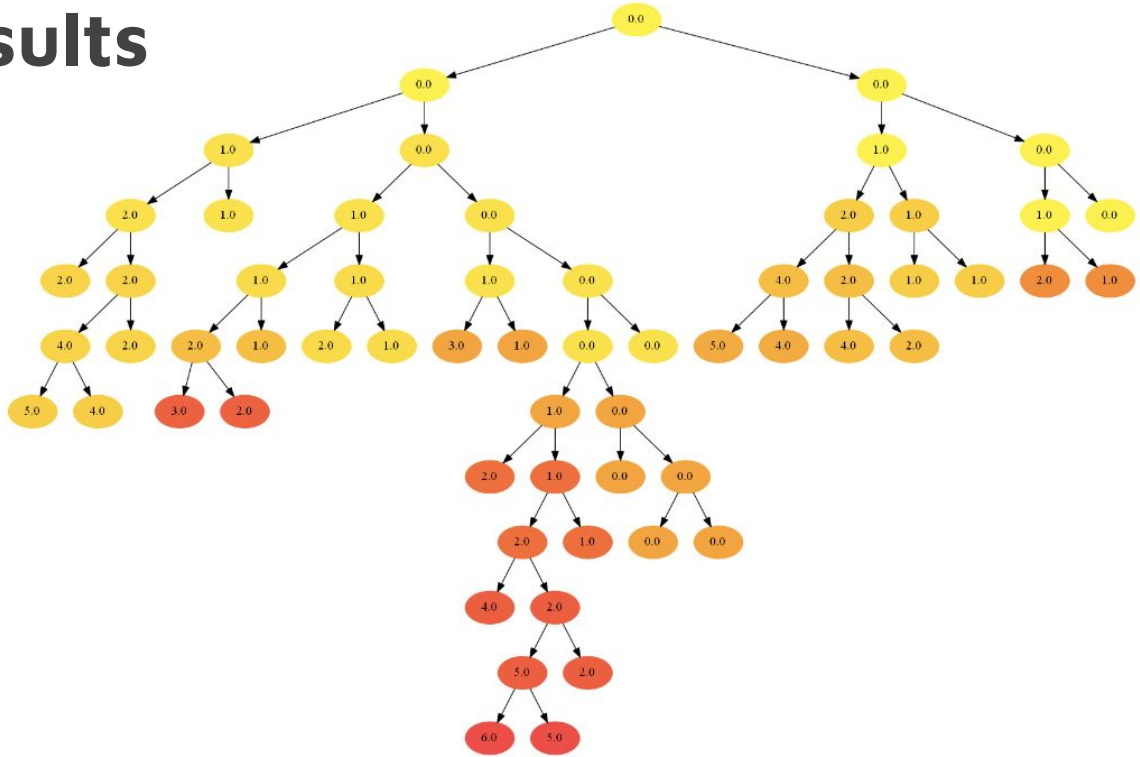
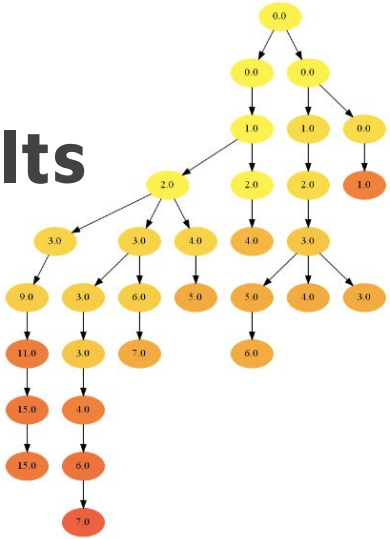
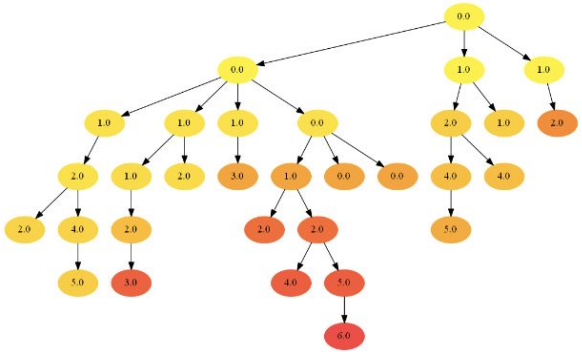


Figure 1: Phylogenetic tree generated by CTMM with no bias on percent population of mutated children. Values of nodes are L1 distances between their copy number profile and normal (root). Yellow color indicates earlier and red indicates later mutations. Single chromosome of length 10,000 with mean mutation size 1,000.

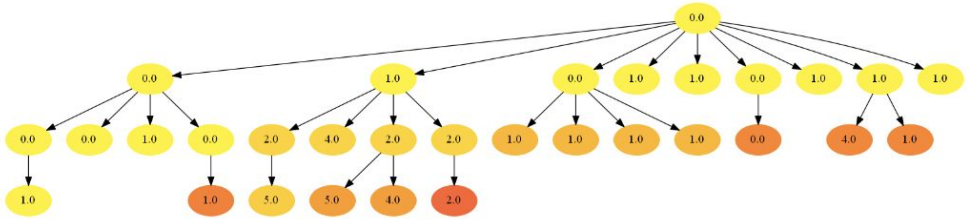
Results



(a) $Beta(\alpha = 5, \beta = 1)$



(b) $Beta(\alpha = 1, \beta = 1) = Unif(0, 1)$



(c) $Beta(\alpha = 1, \beta = 5)$

Figure 2: Phylogenetic trees for varying probability of large population size for mutated offspring. Percent of mutant offspring $p \sim Beta(\alpha, \beta)$. Increased $\frac{\alpha}{\beta}$ leads to larger mutant population and therefore increased chance a mutant cell type will mutate again. Values of nodes are L1 distances between their copy number profile and normal (root). Yellow color indicates earlier and red indicates later mutations. Single chromosome of length 10,000 with mean mutation size 1,000.

Results

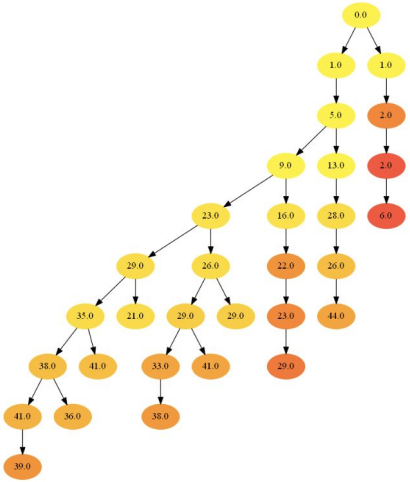
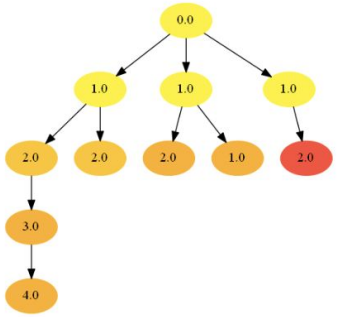
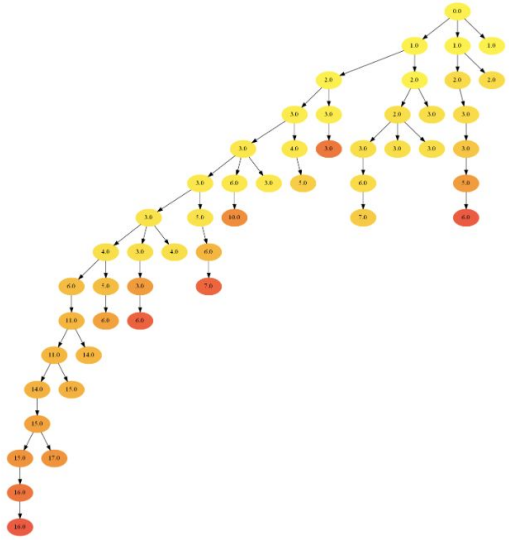


Figure 3: Increasing mean mutation length increases the L1 distance for more mutated nodes. Mean mutation length $\ell = 5000$ and percent of mutant offspring $p \sim Beta(\alpha = 5, \beta = 1)$. Values of nodes are L1 distances between their copy number profile and normal (root). Yellow color indicates earlier and red indicates later mutations. Single chromosome of length 10,000.



(a) Number of mutations $m = 10$



(b) Number of mutations $m = 50$

Figure 4: Phylogenetic topology for varying number of mutations. Values of nodes are L1 distances between their copy number profile and normal (root). Yellow color indicates earlier and red indicates later mutations. Single chromosome of length 10,000 with mean mutation size 1,000.

Summary

- Developed a method for simulating tumor phylogenies
- At a single mutation resolution (inversion, amplification, deletion)
- Using Continuous Time Markov Models
- Various input parameters allow flexibility to test deconvolution tools
- Created 288 data sets using pre-defined parameters
- In the future: increase types of simulated mutations

Thank you

Questions?