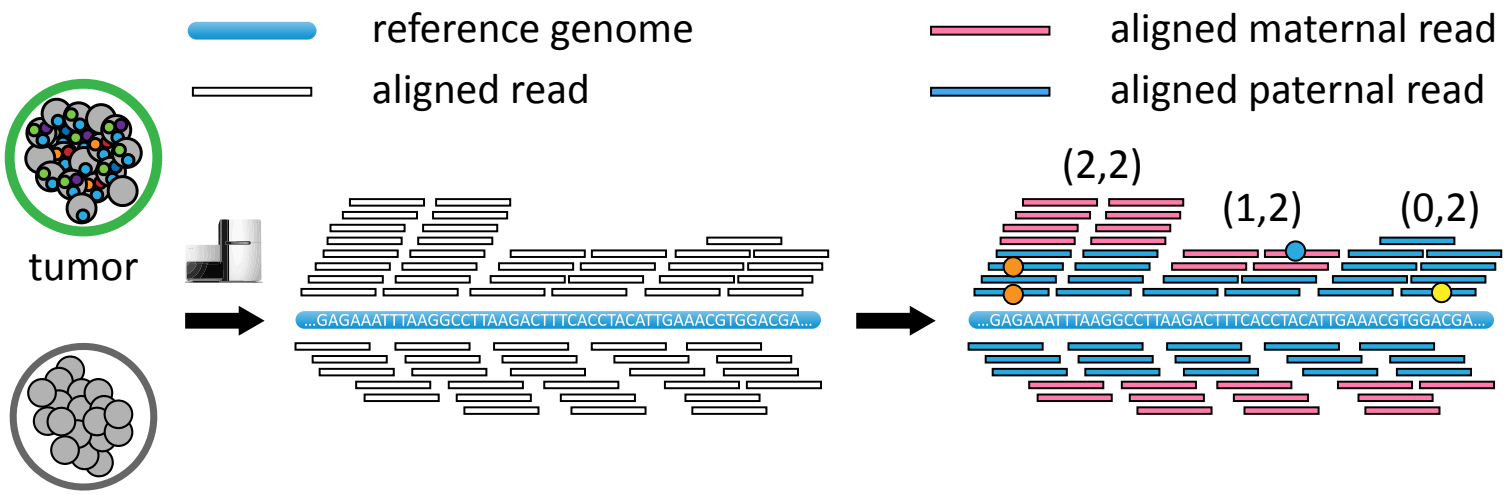
CS 598MEB
Computational Cancer Biology
Lecture 8

Mohammed El-Kebir

February 7, 2019

Cancer Phylogenetics Pipeline



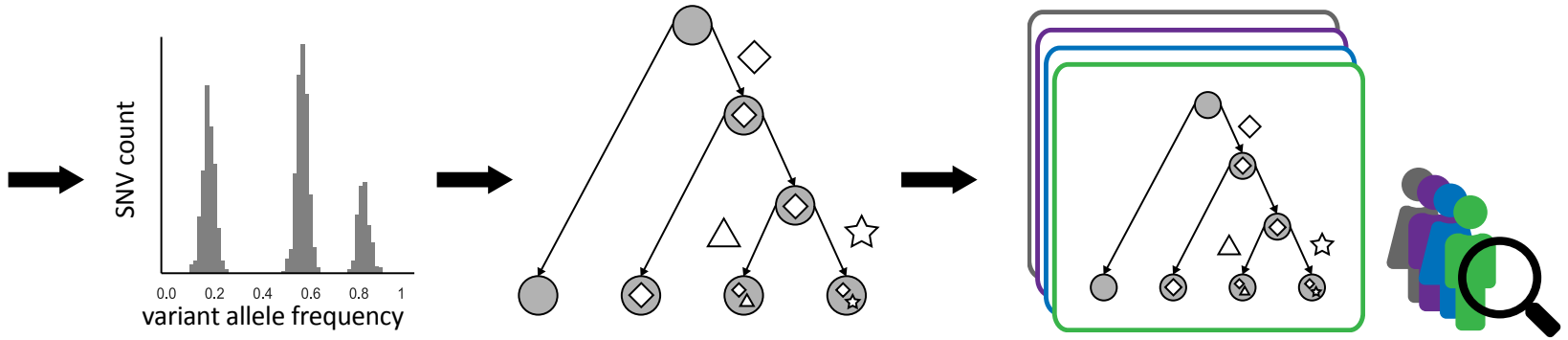
(i) Alignment

(ii) Somatic variant calling

single nucleotide variant (SNV)

SNV cluster

(x, y) copy number state (# maternal, # paternal copies)

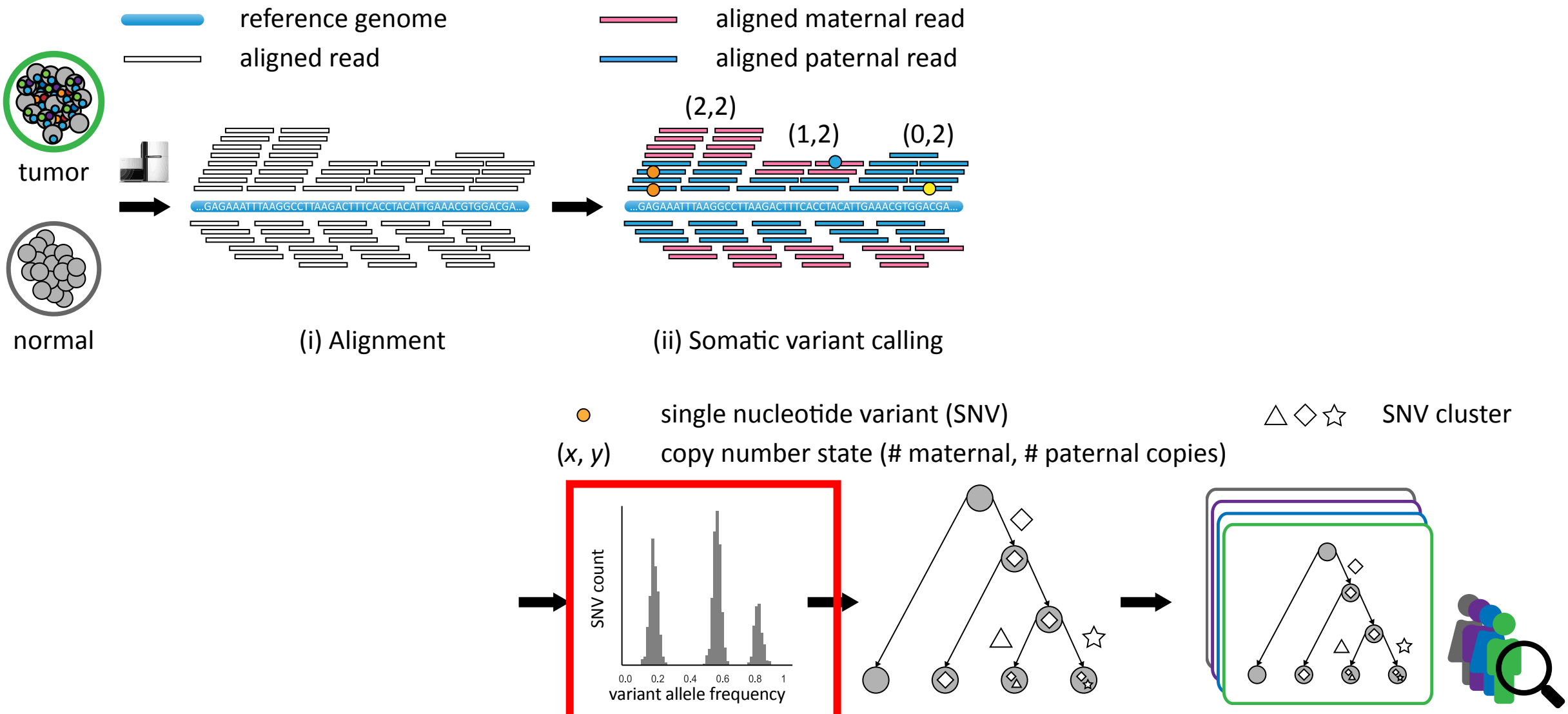


(iii) SNV clustering

(iv) Phylogeny inference

(v) Downstream analysis

Cancer Phylogenetics Pipeline



Outline

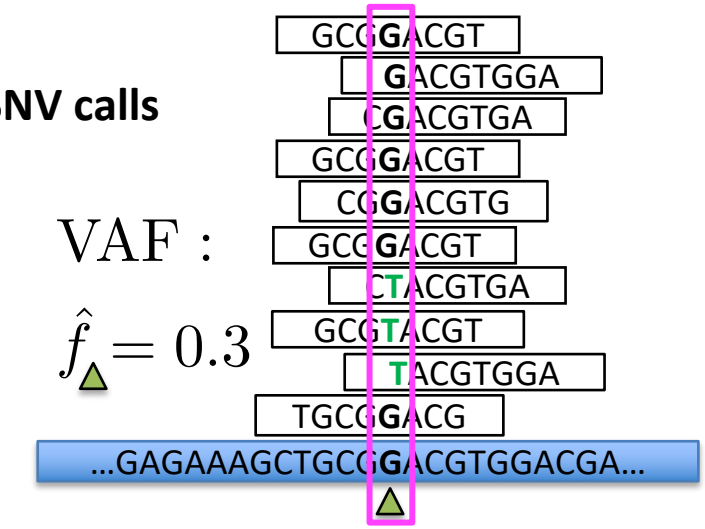
- Summary Statistics of SNV Prevalence
  - SNVF
  - CCF
  - DCF
- Interplay between copy number state and summary statistics
- Simultaneous clustering and inference SNV prevalence statistics

Reading:

- M. El-Kebir, S. Zaccaria and B.J. Raphael. Descendant Cell Fraction: Copy-aware Inference of Clonal Composition and Evolution in Cancer

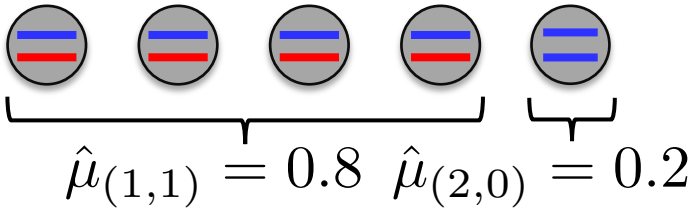
It is hard to infer a phylogeny from bulk DNA-seq data…

SNV calls

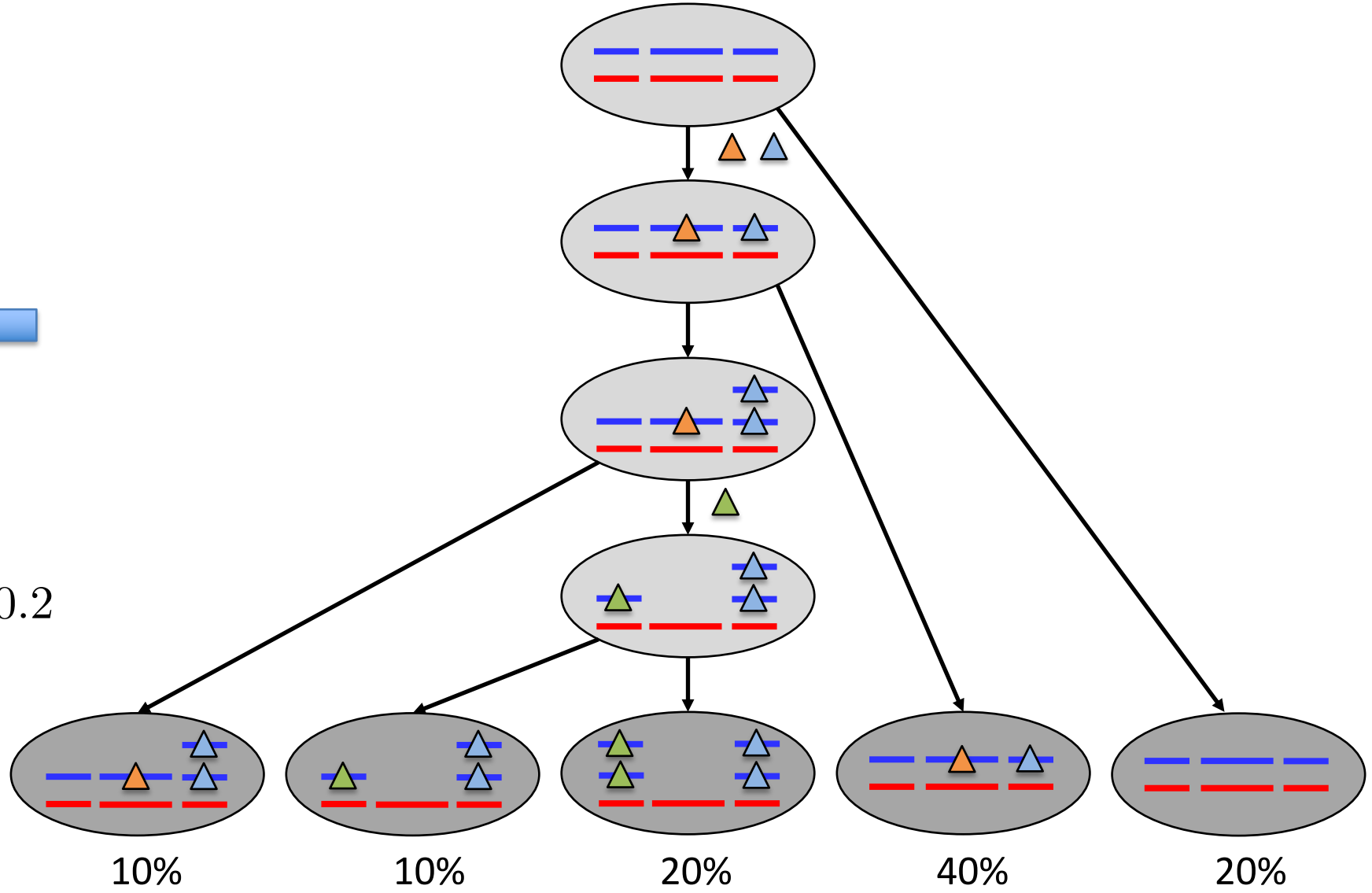


$\text{VAF}: \hat{f}_{\triangle} = 0.3$

CNA calls

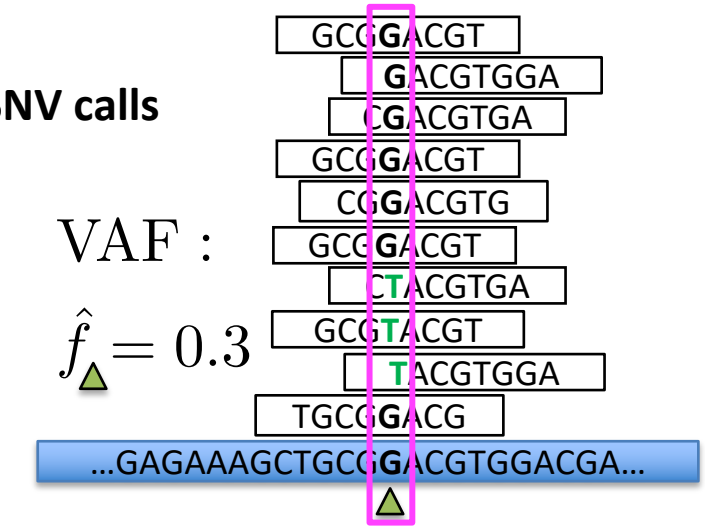


$\hat{\mu}_{(1,1)} = 0.8 \quad \hat{\mu}_{(2,0)} = 0.2$



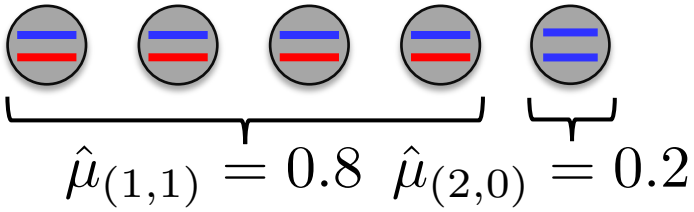
Can we infer SNVs introduced on the same branch?

SNV calls

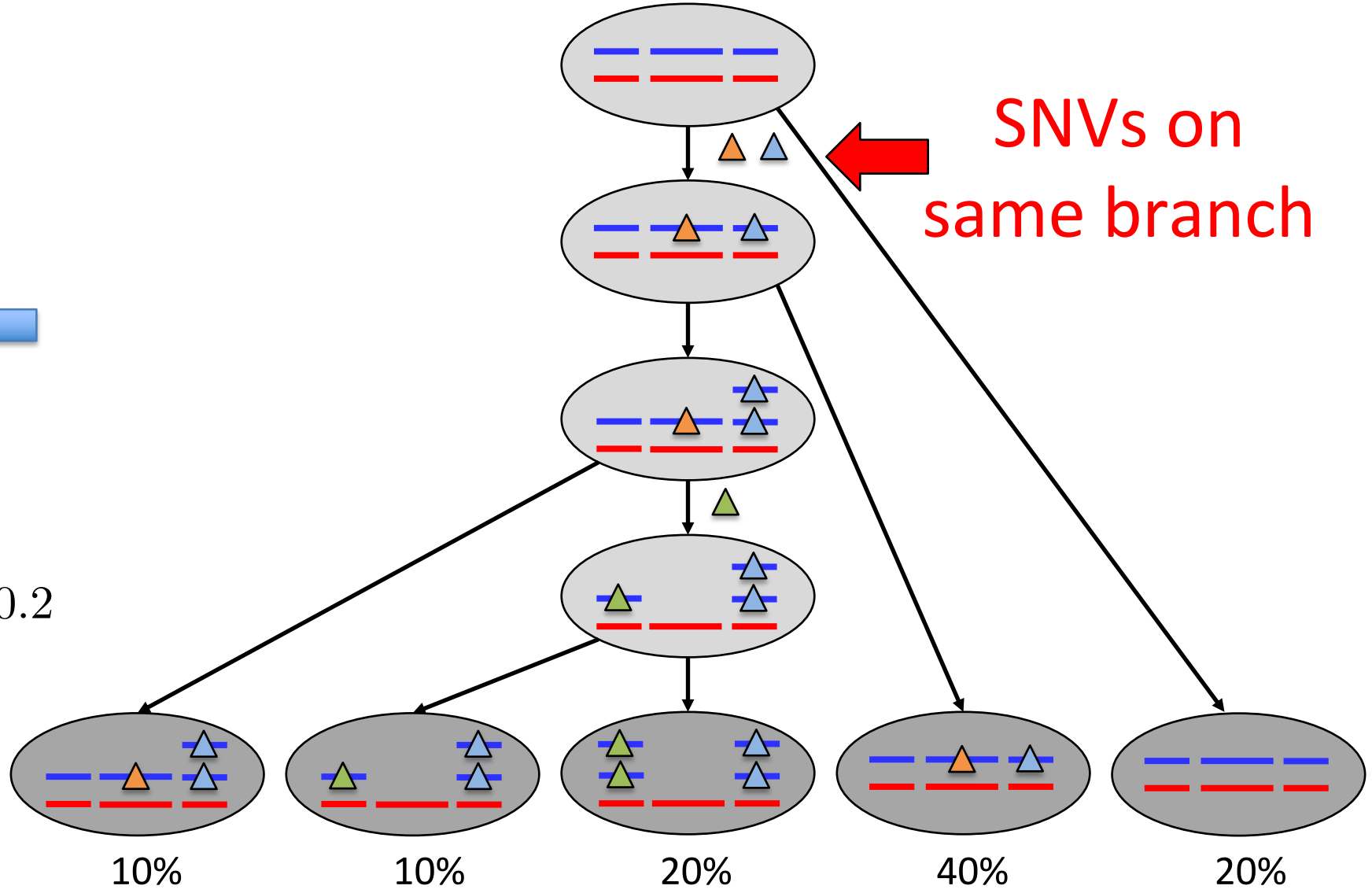


$\text{VAF}: \hat{f}_{\triangle} = 0.3$

CNA calls

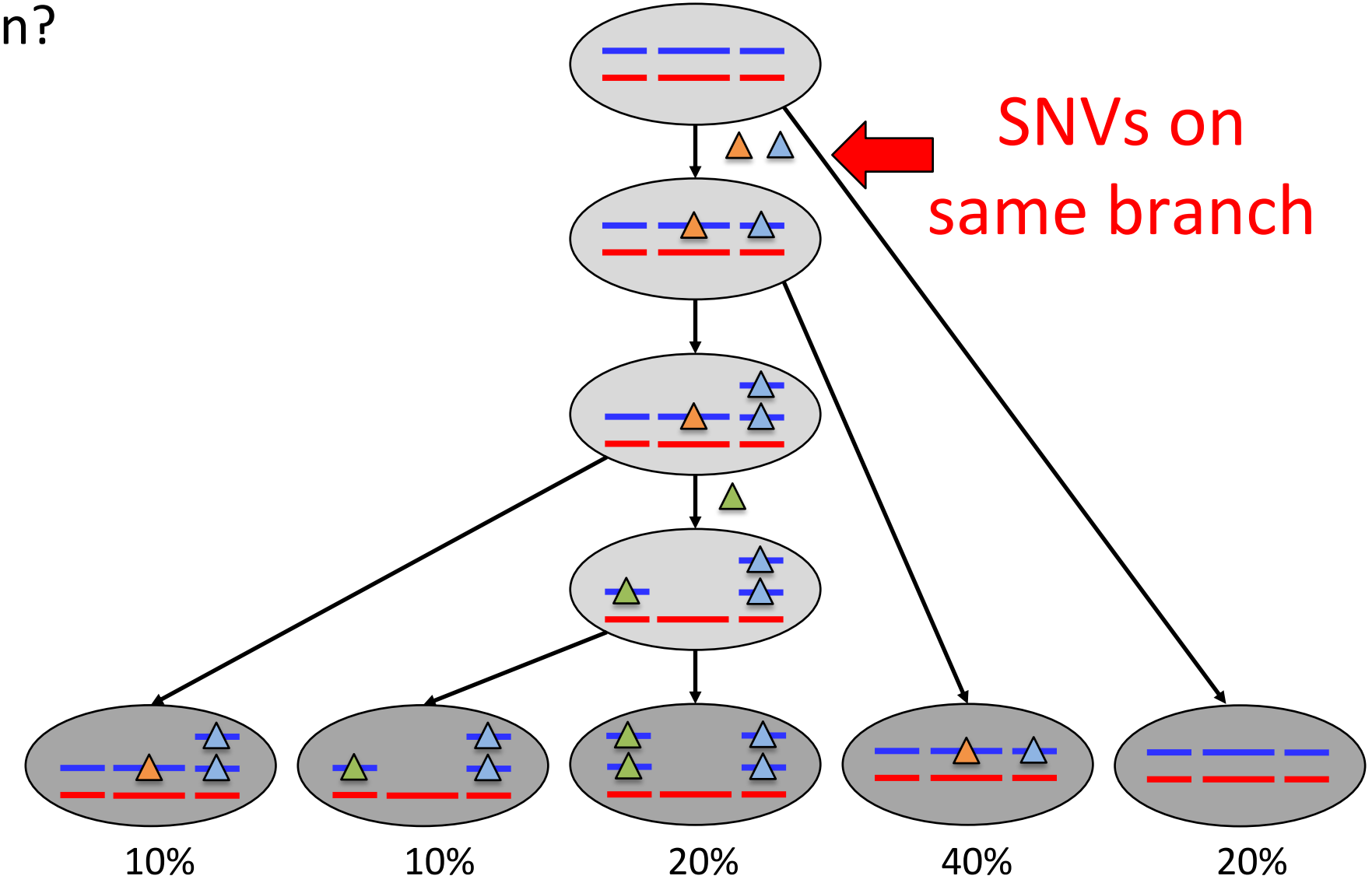


$\hat{\mu}_{(1,1)} = 0.8 \quad \hat{\mu}_{(2,0)} = 0.2$

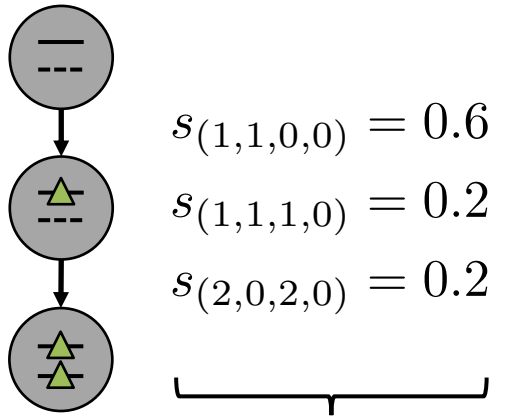


Can we infer SNVs introduced on the same branch?

- What does this mean?



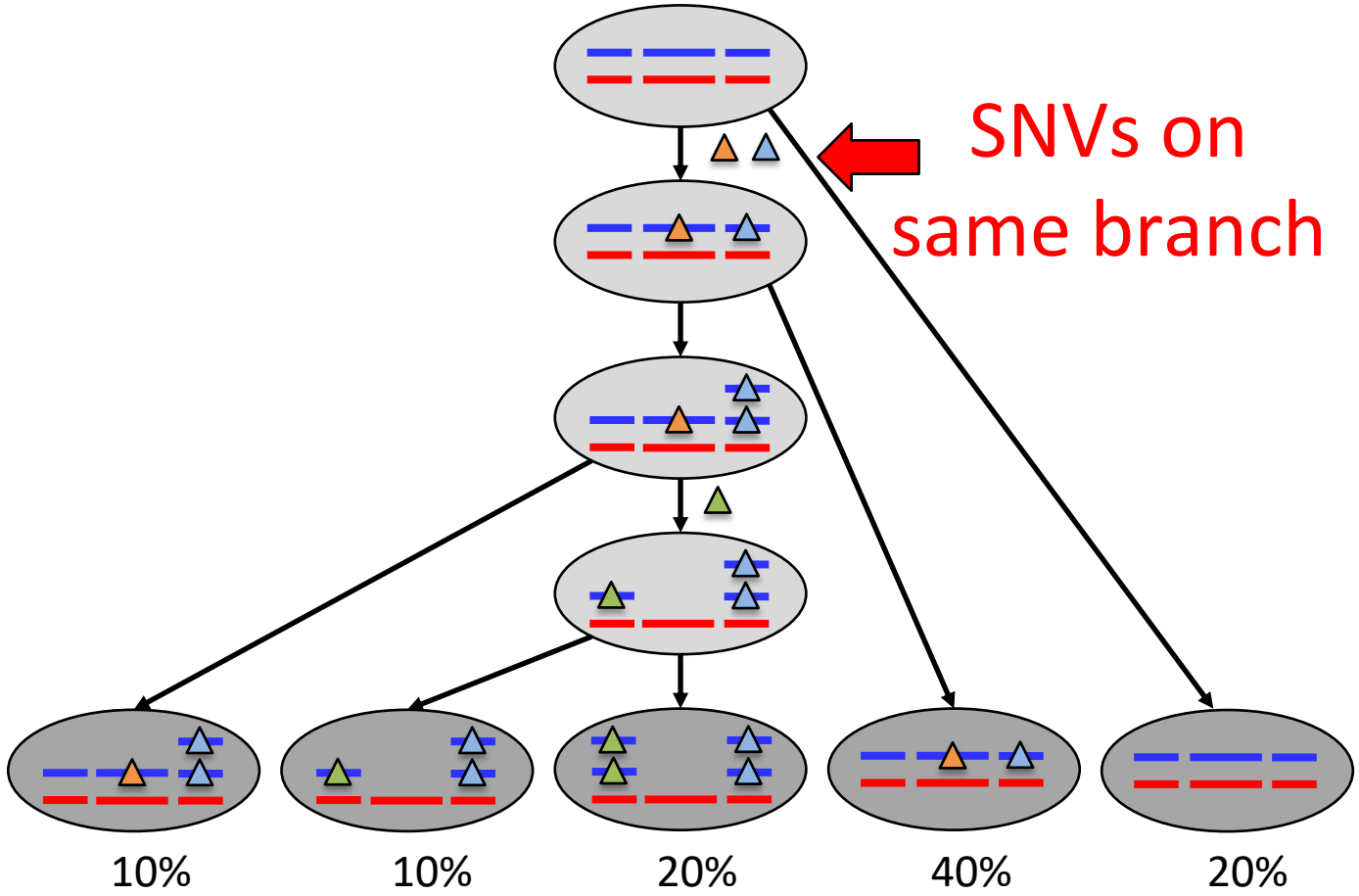
State Trees and State Proportions



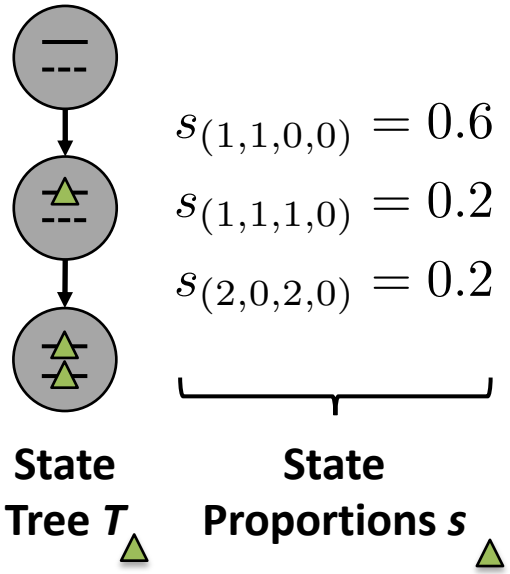
$$s_{(1,1,0,0)} = 0.6$$
$$s_{(1,1,1,0)} = 0.2$$
$$s_{(2,0,2,0)} = 0.2$$

State Tree *T*

State Proportions *s*



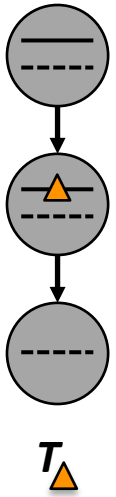
State Trees and State Proportions



State Tree T

$s_{(1,1,0,0)} = 0.6$
$s_{(1,1,1,0)} = 0.2$
$s_{(2,0,2,0)} = 0.2$

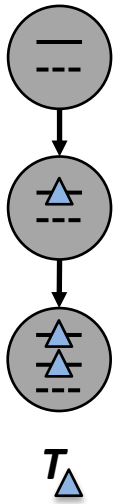
State Proportions s



T

$s_{(1,1,0,0)} = 0.2$
$s_{(1,1,1,0)} = 0.5$
$s_{(1,0,0,0)} = 0.3$

s

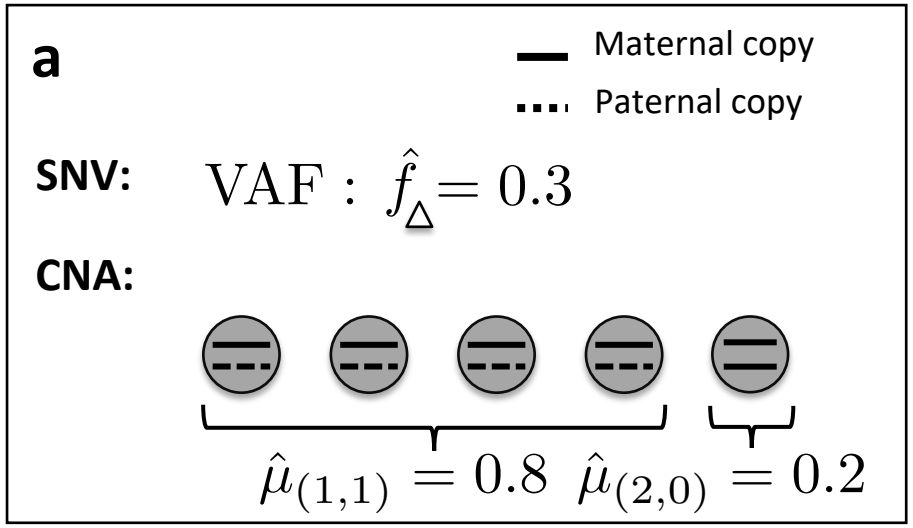


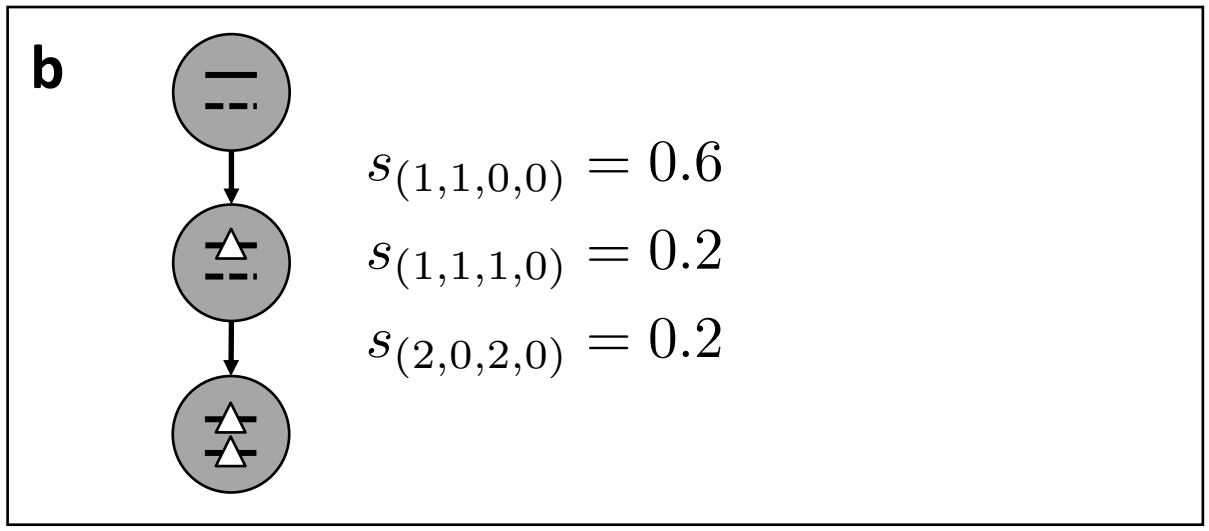
T

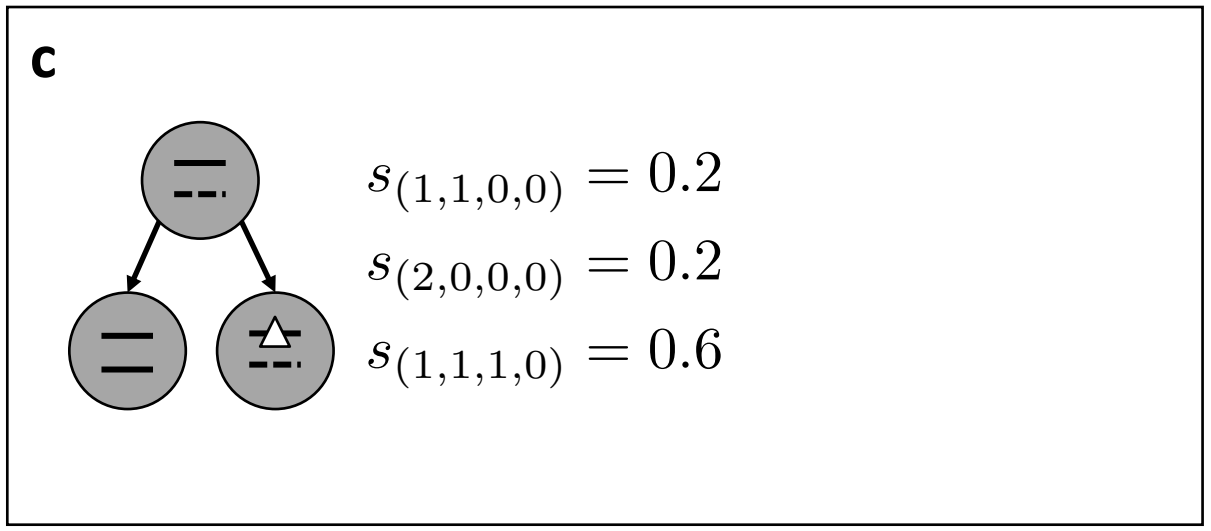
$s_{(1,1,0,0)} = 0.2$
$s_{(1,1,1,0)} = 0.4$
$s_{(2,1,2,0)} = 0.4$

s

Summarizing SNV Prevalence





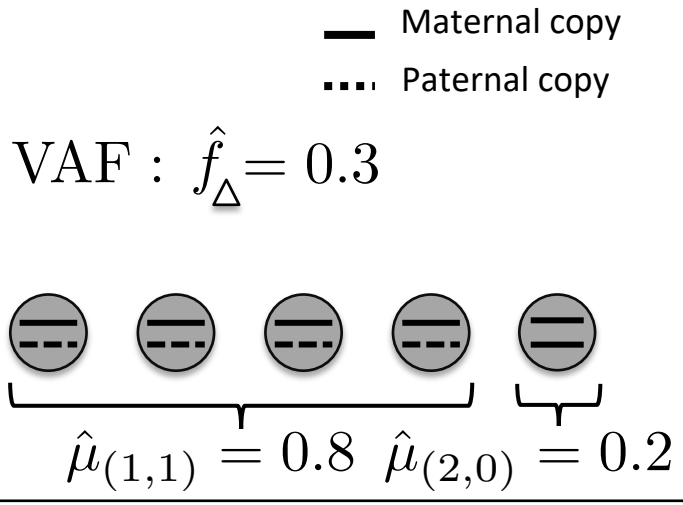


Mutation Loss is Ubiquitous in Cancer

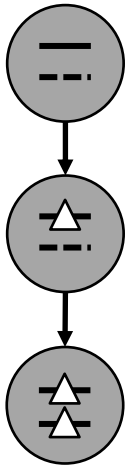
- In a recent ovarian cancer study of 10 patients the proportion of SNVs affected by copy number loss ranged from 0.28 to 0.89 [McPherson et al., Nat Genetics 2016].
- Survival analysis performed in a large-scale study of 100 non–small-cell lung cancer patients required the manual correction of CCFs to account for loss [Jamal-Hanjani et al., NEJM 2017].
- Not correctly accounting for mutation loss biases downstream analyses and may lead to incorrect conclusions regarding:
  - Timing of driver mutations [McGranahan et al., Science Trans Med 2015];
  - Patterns of metastatic spread [McPherson et al., Nat Genetics 2016] [Gundem et al., Nature 2015];
  - Evolutionary history of a tumor [Nik-Zainal et al., Cell 2012].

CCF and DCF are nonidentifiable

a



b

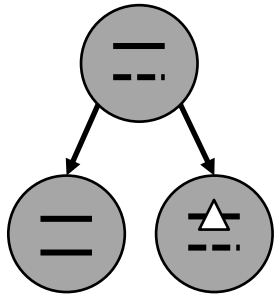


$s_{(1,1,0,0)} = 0.6$ SNVF $f_\Delta = 0.3$

$s_{(1,1,1,0)} = 0.2$ CCF $c_\Delta = 0.4$

$s_{(2,0,2,0)} = 0.2$ DCF $d_\Delta = 0.4$

c



$s_{(1,1,0,0)} = 0.2$ SNVF $f_\Delta = 0.3$

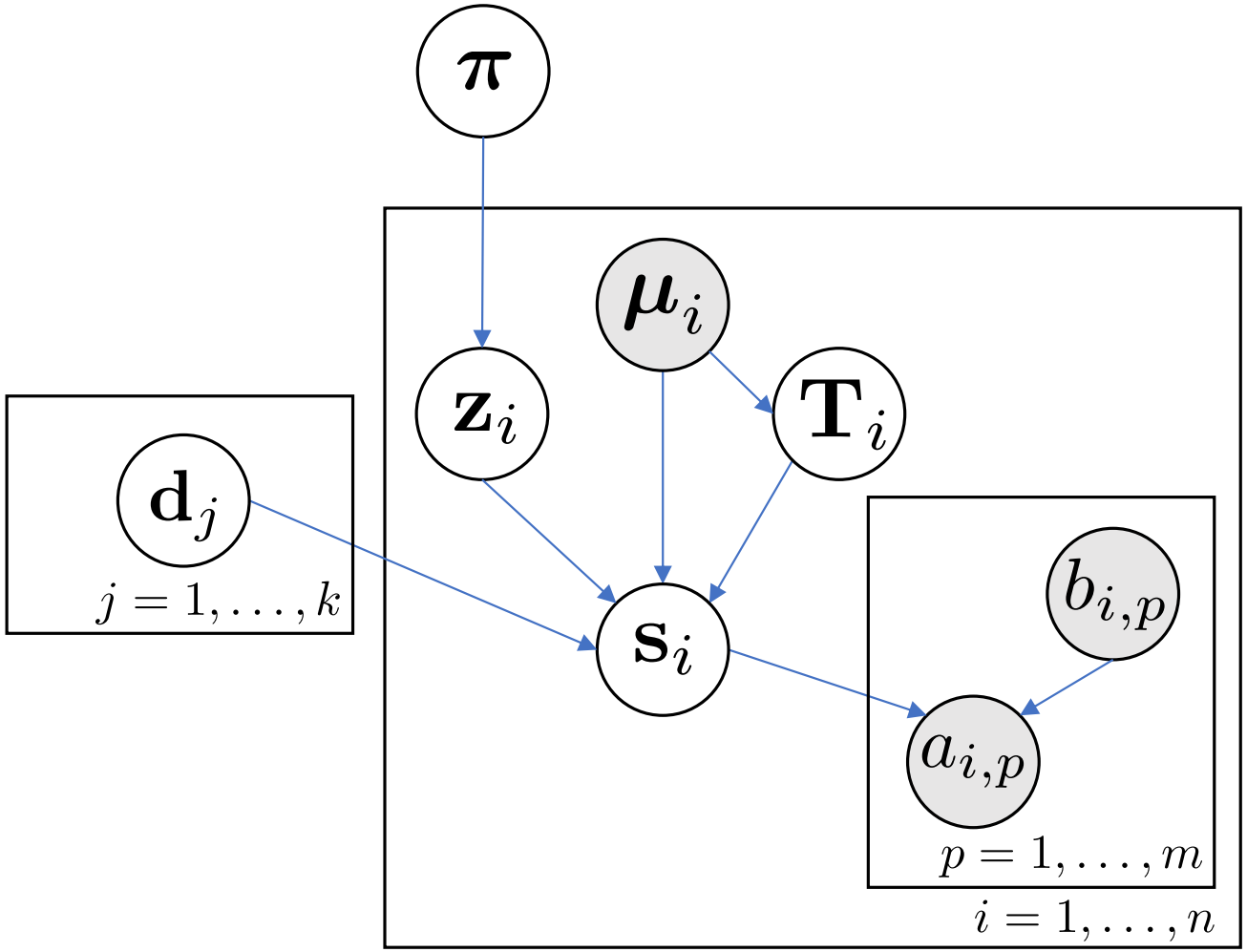
$s_{(2,0,0,0)} = 0.2$ CCF $c_\Delta = 0.6$

$s_{(1,1,1,0)} = 0.6$ DCF $d_\Delta = 0.6$

Summary Statistics for SNV Prevalence and Evolution

Statistic	Summarizes	Confounded by	Identifiable
SNVF ***f***	***s***	copy-number gain & loss	yes (VAF)
CCF ***c***	***s***	copy-number loss	no
DCF ***d***	***s*** and ***T***	-	no

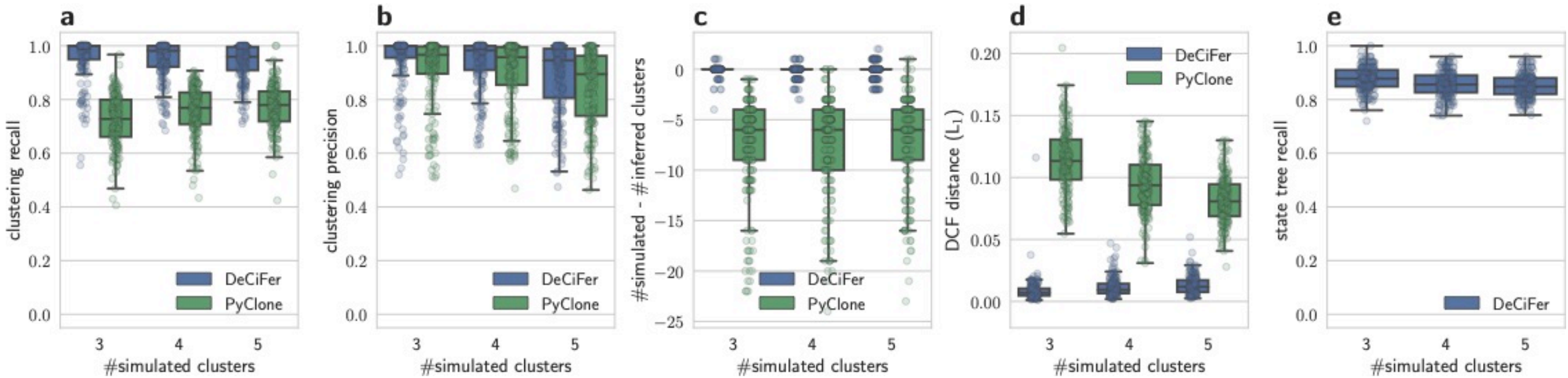
Generative Model



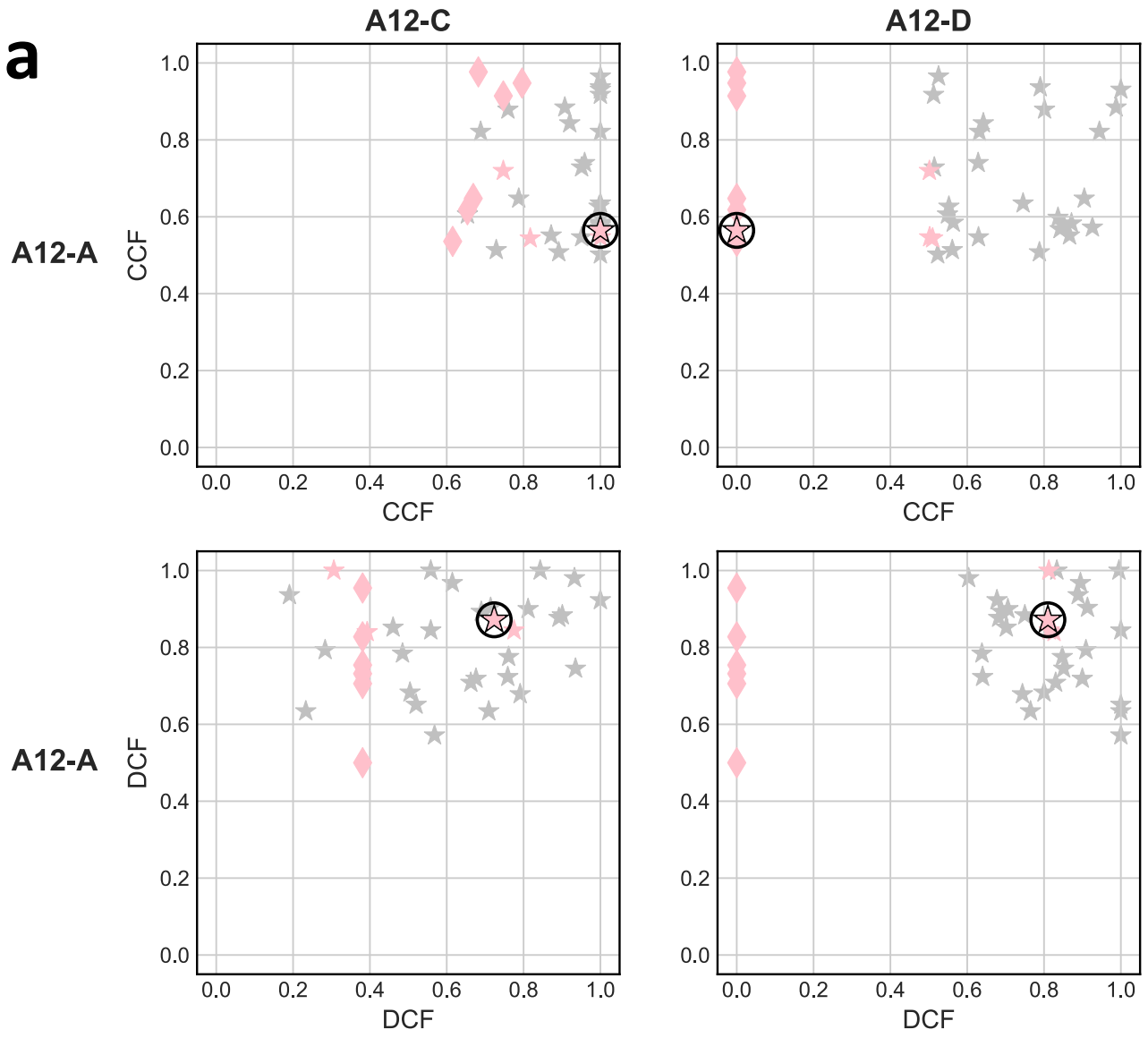
$$\mathbf{z}_i \mid \boldsymbol{\pi} \sim \mathrm{Cat}(\boldsymbol{\pi})$$

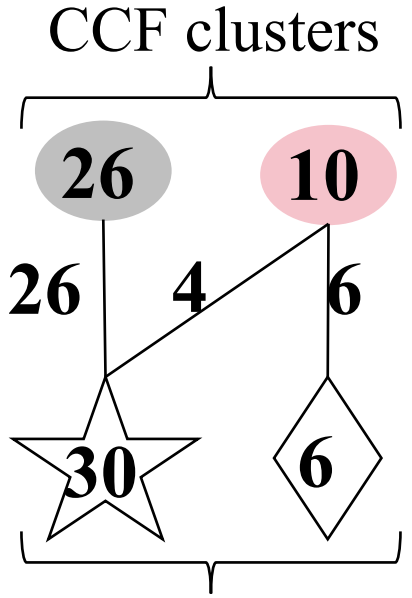
$$a_{i,p} \mid b_{i,p}, \mathbf{s}_{i,p} \sim \mathrm{Binom}(b_{i,p}, f(\mathbf{s}_{i,p}))$$

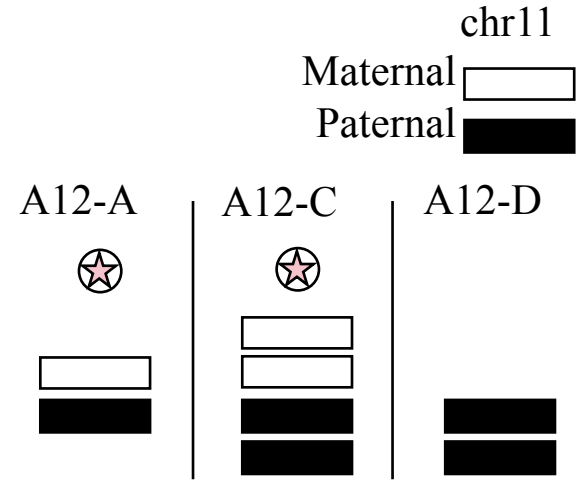
Simulations



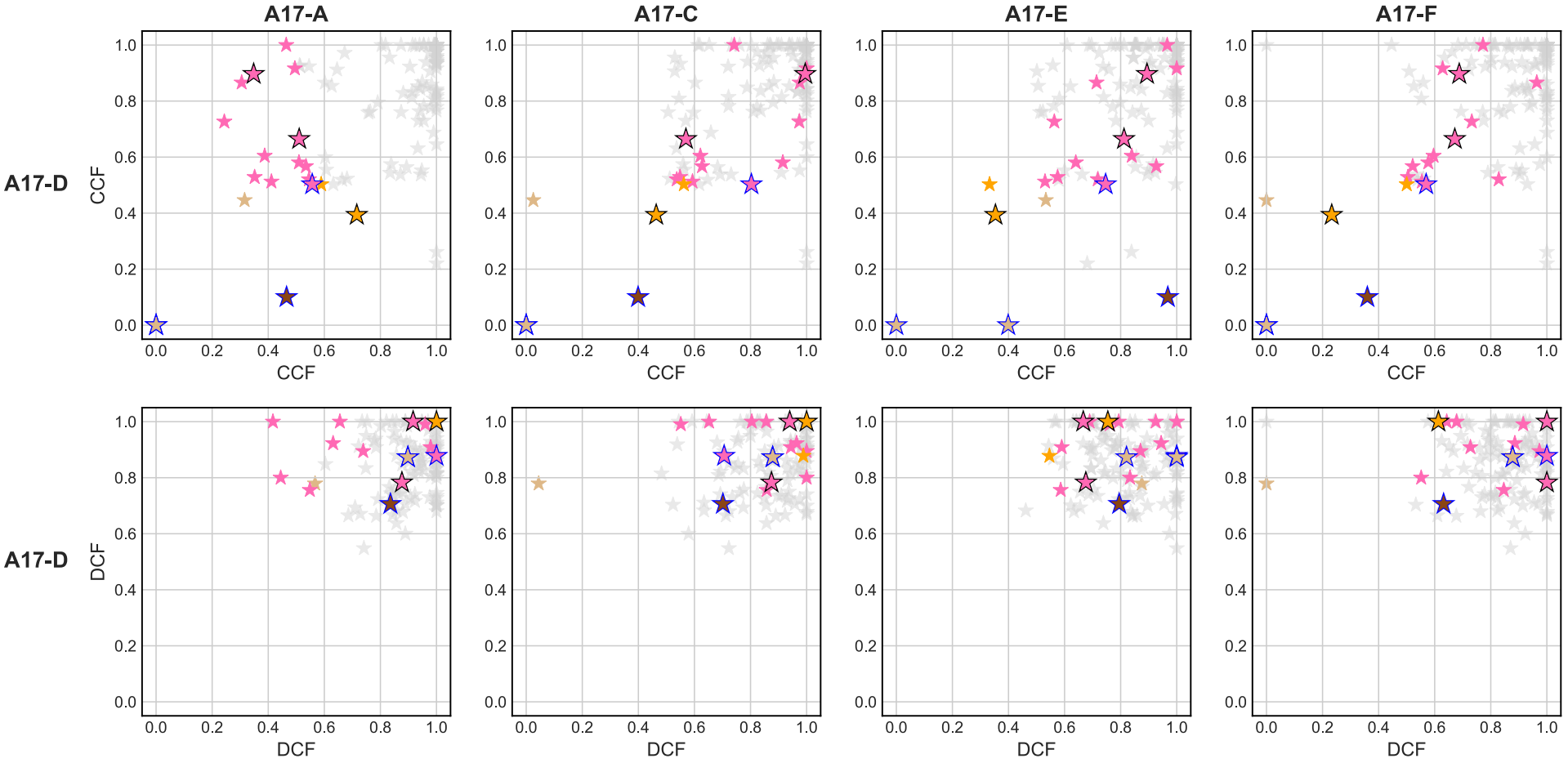
Prostate Cancer A12

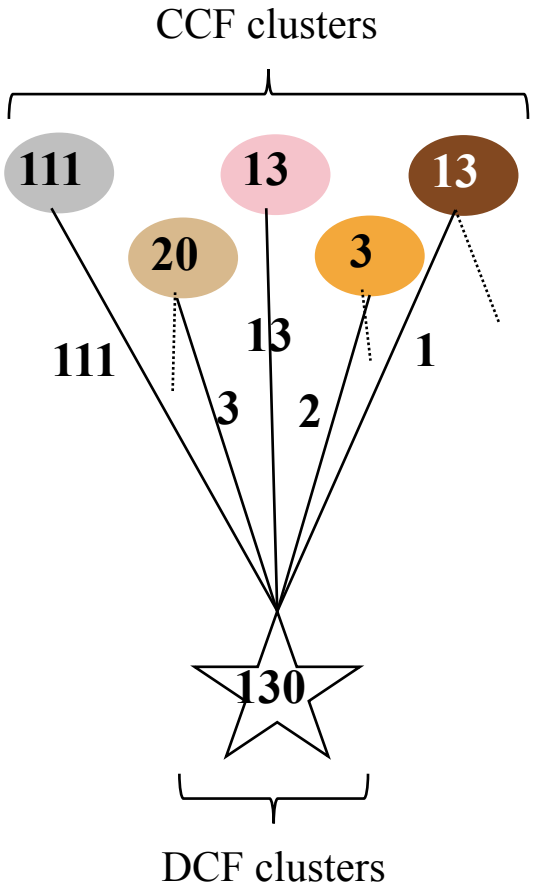




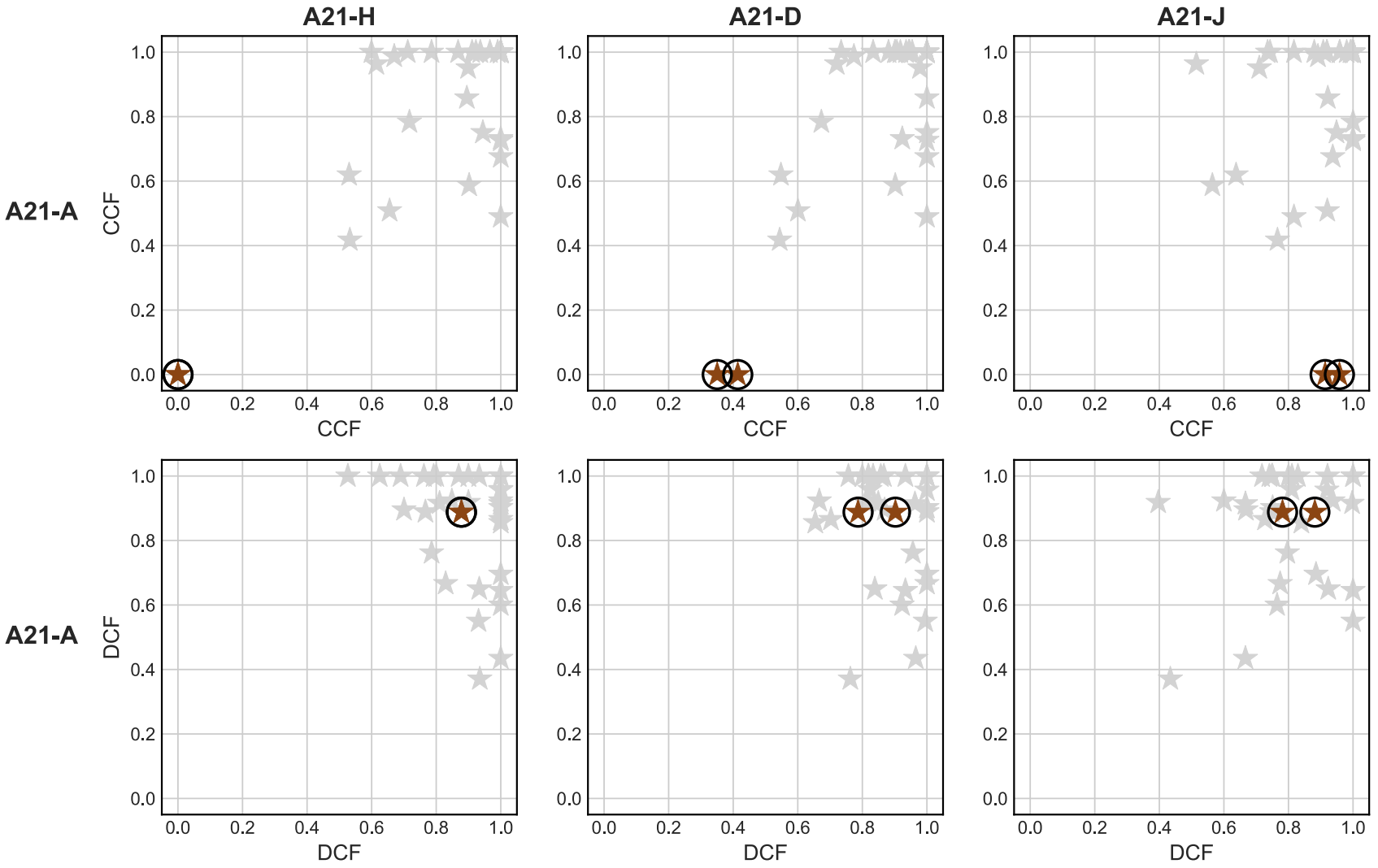


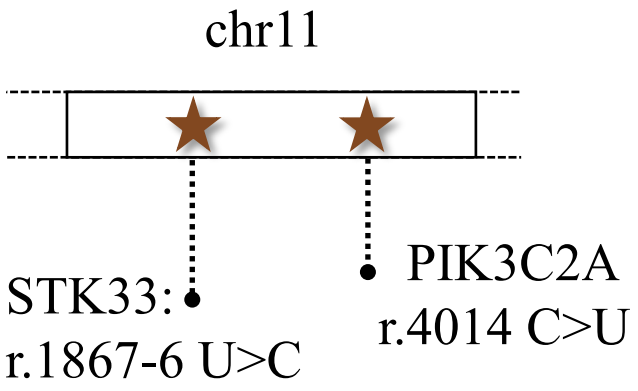
Prostate Cancer A17

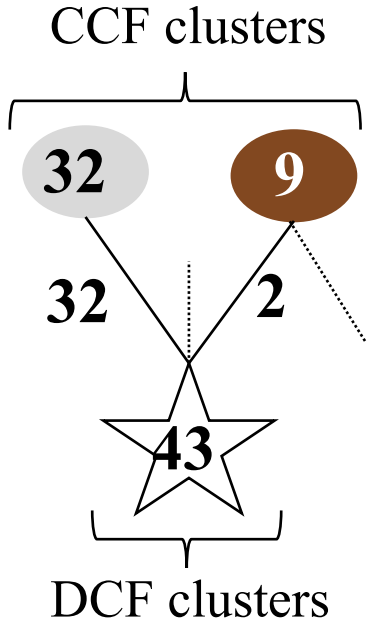




Prostate Cancer A21







Prostate Cancer A21