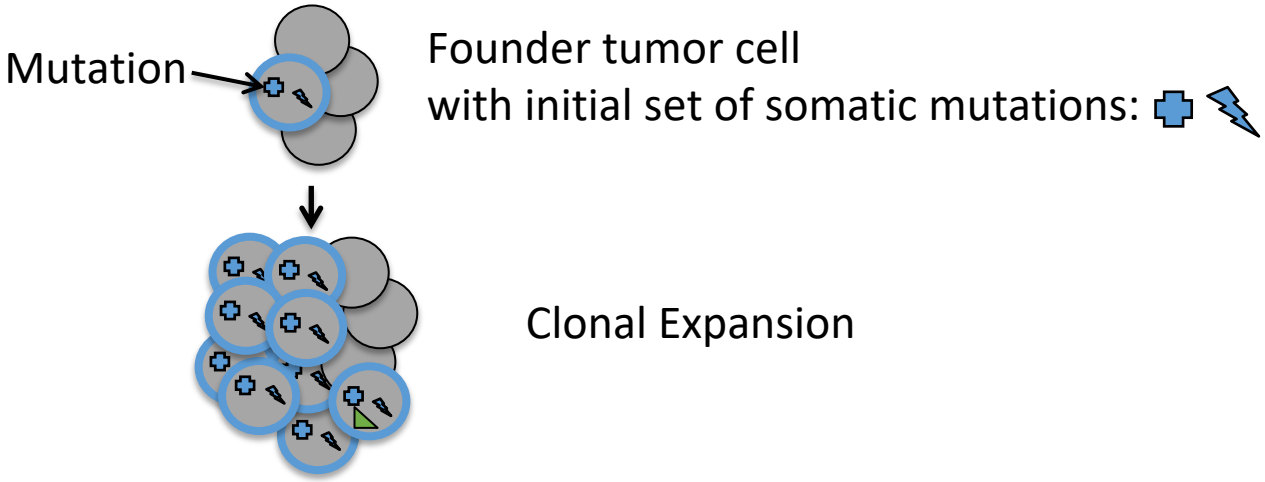
Descendant Cell Fraction: Accounting for Mutation Loss in Cancer Evolution

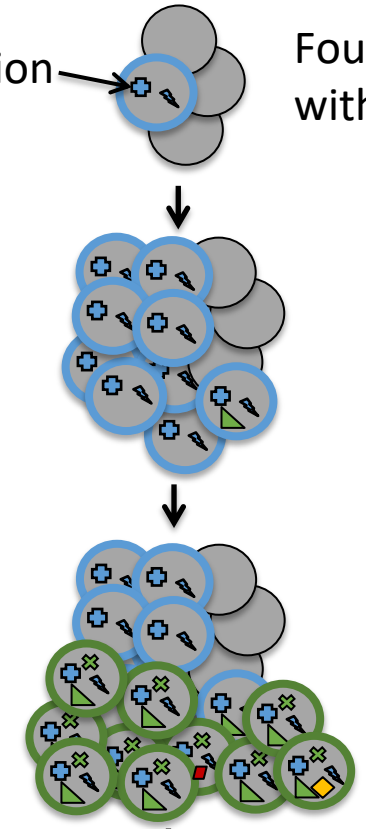
Mohammed El-Kebir[1], Simone Zaccaria[2], Gryte Satas[2,3] and Ben Raphael[2]
[1]University of Illinois at Urbana Champaign, Department of Computer Science
[2]Princeton University, Department of Computer Science
[3]Brown University, Department of Computer Science

ISMB/HitSeq 2019

Cancer is an evolutionary process



Cancer is an evolutionary process

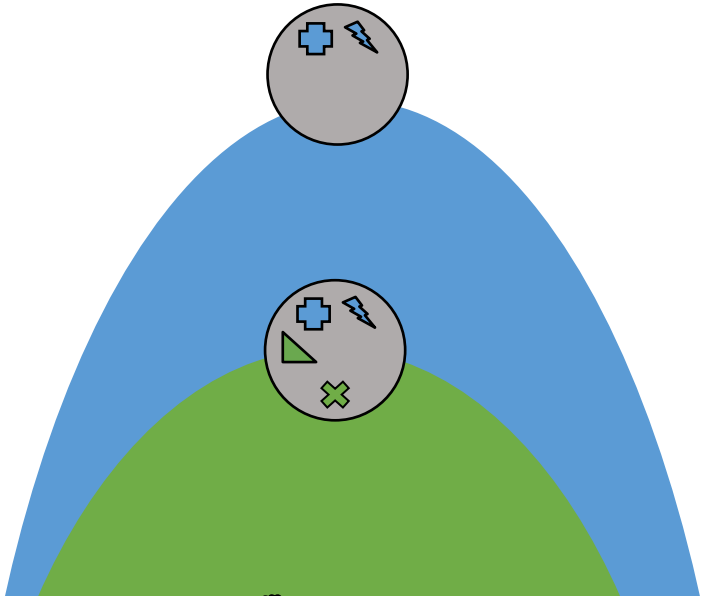


Founder tumor cell
with initial set of somatic mutations:

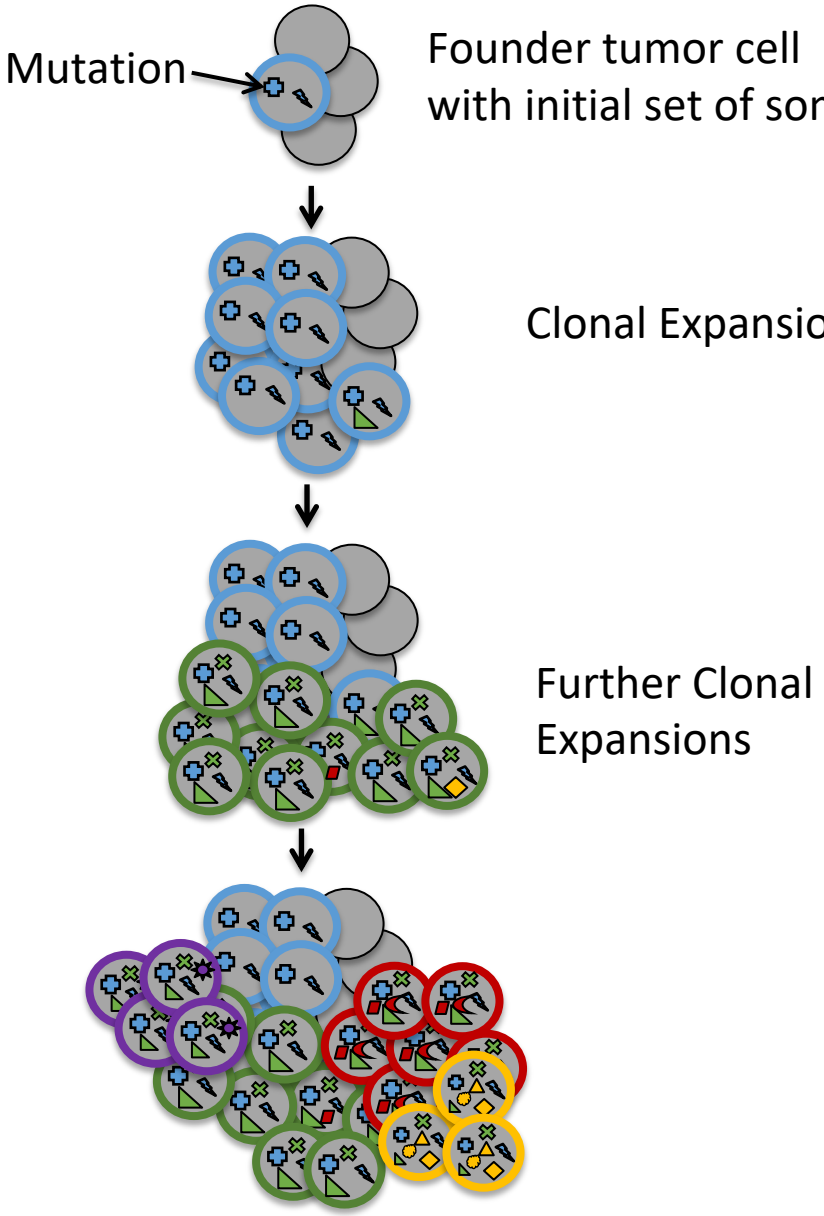
Clonal Expansion

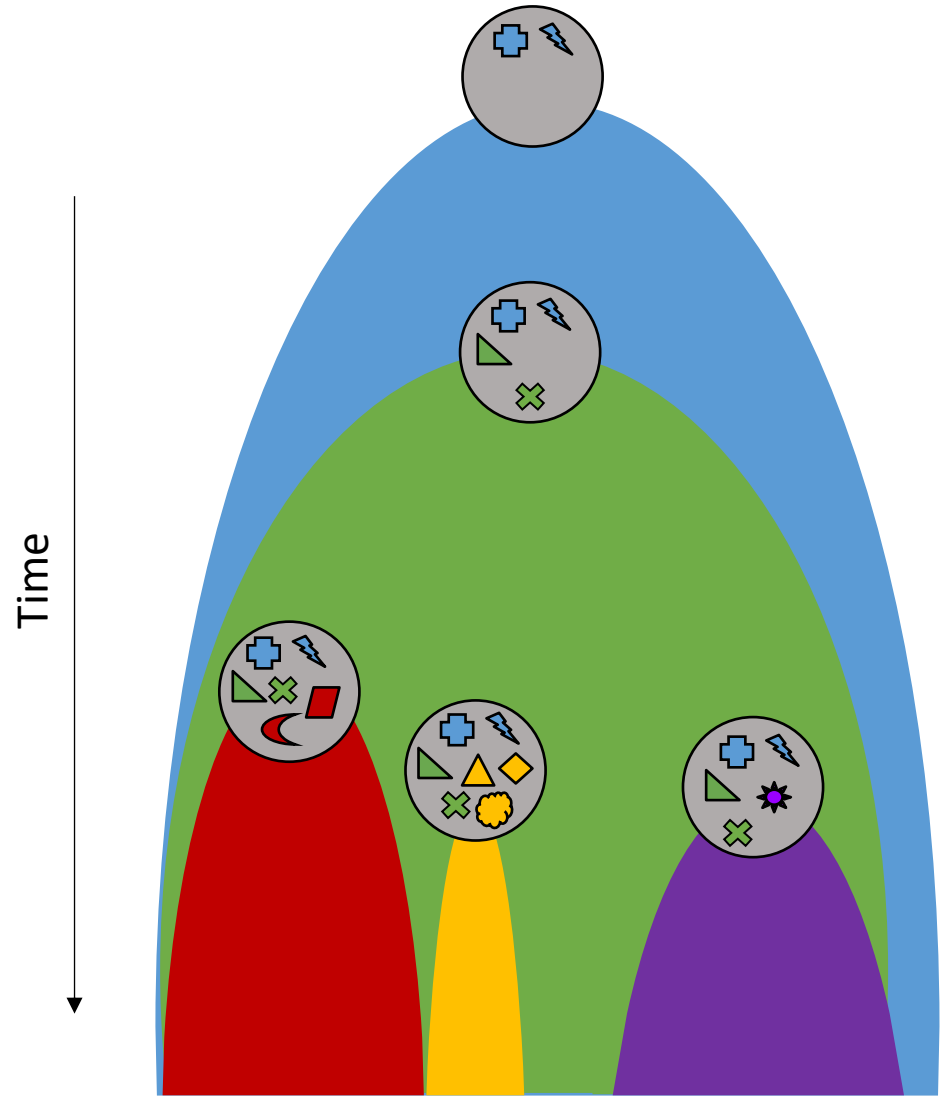
Further Clonal
Expansions

Time



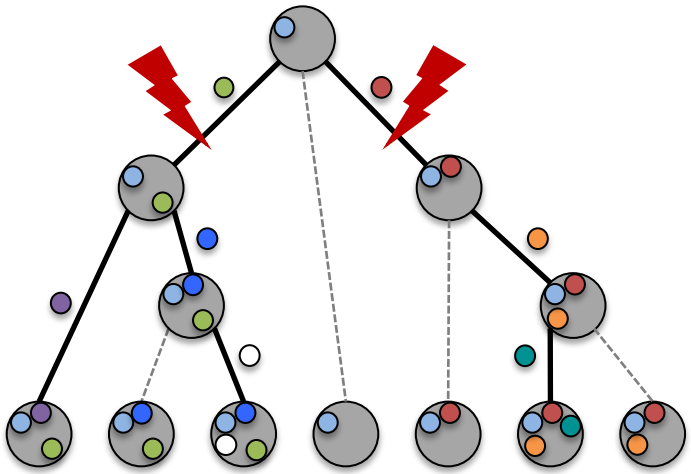
Cancer is an evolutionary process



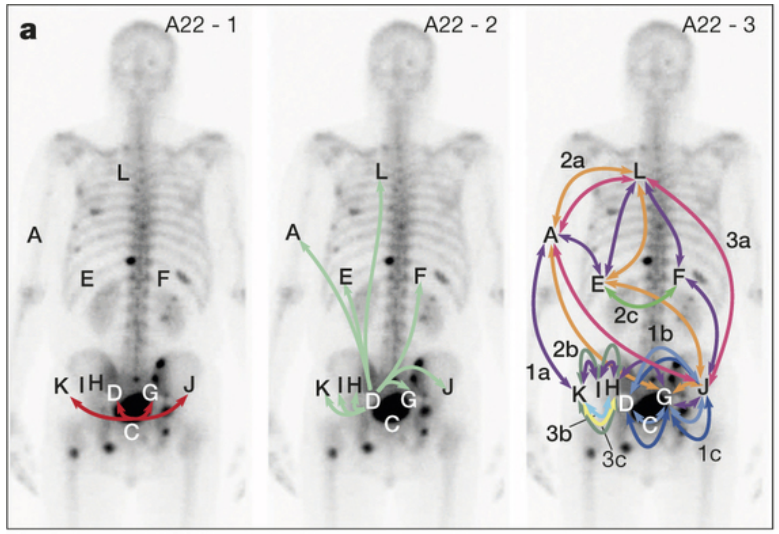


Tumor Evolution is Key to Understanding and Treating Cancer

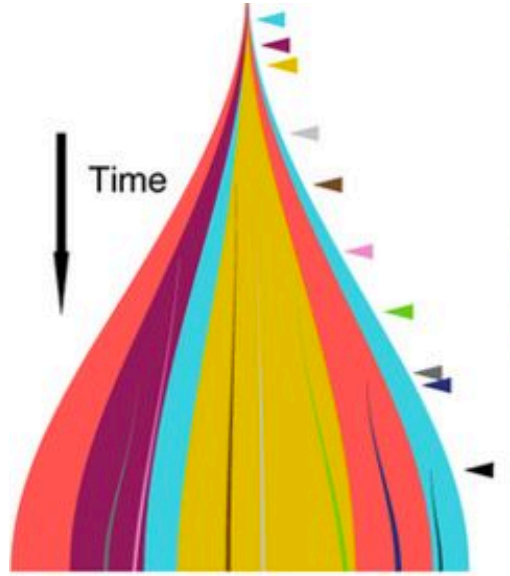
Identify targets for treatment



Reconstruct metastatic development

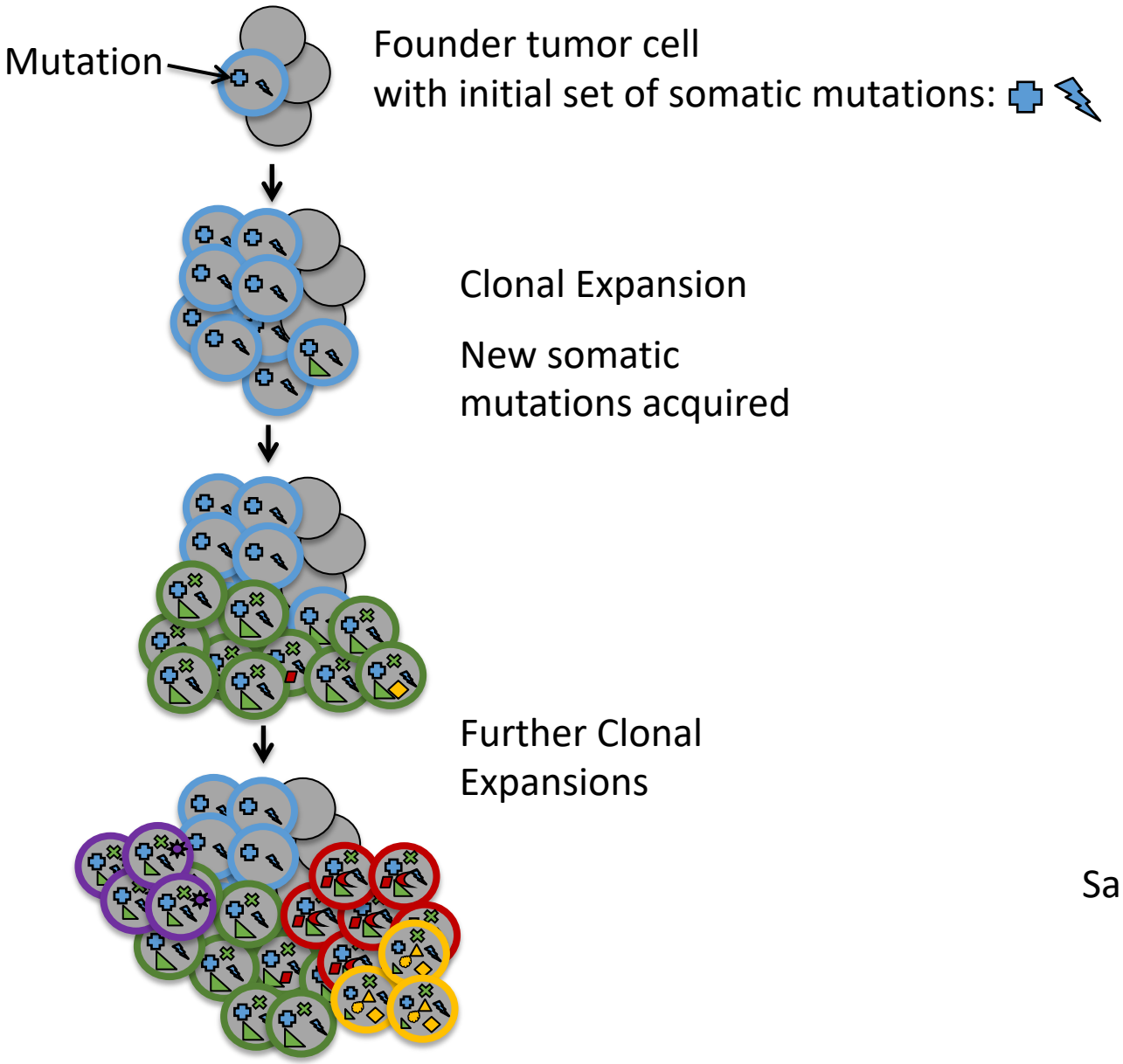


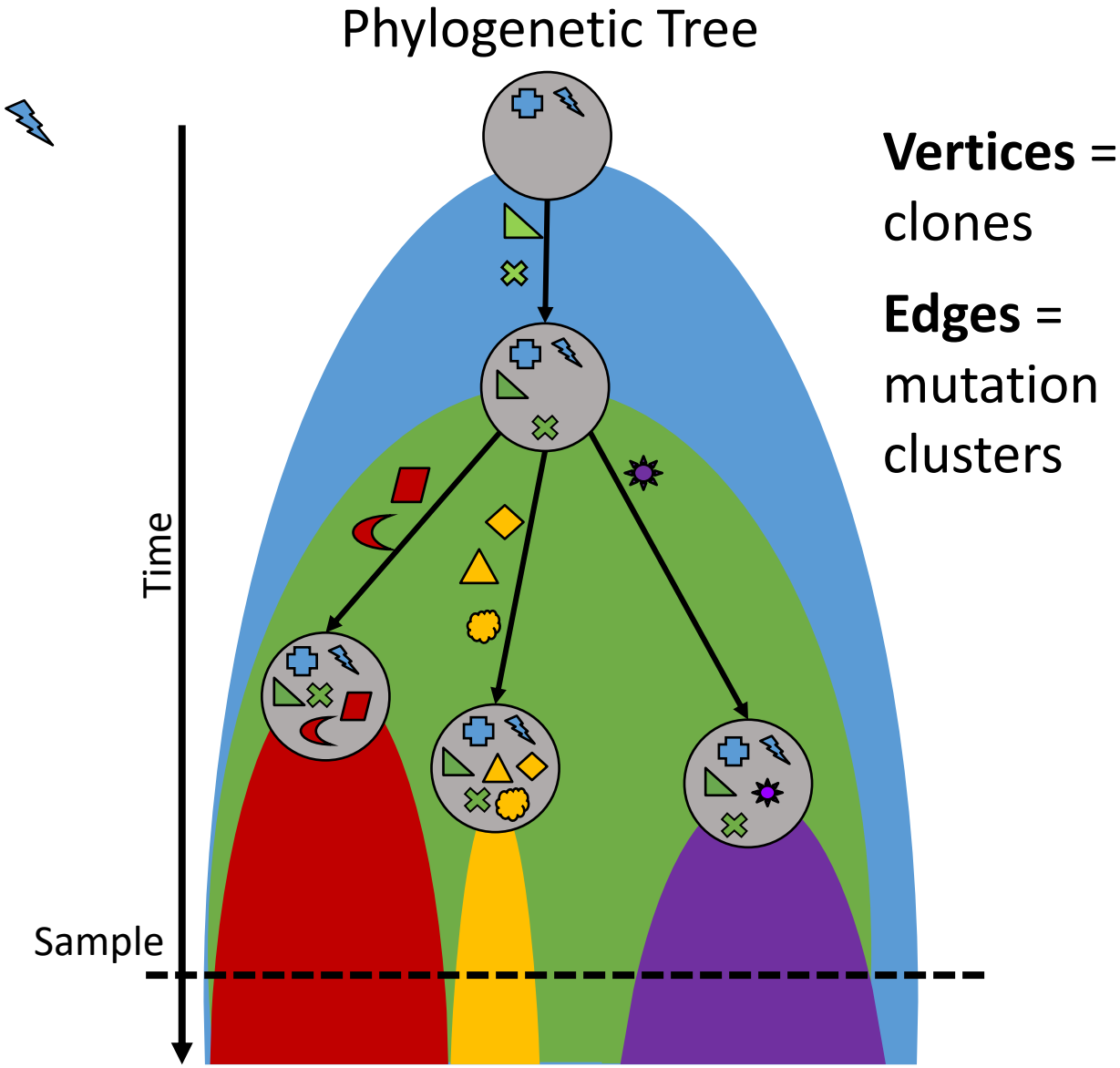
Understanding population dynamics and selection in tumors



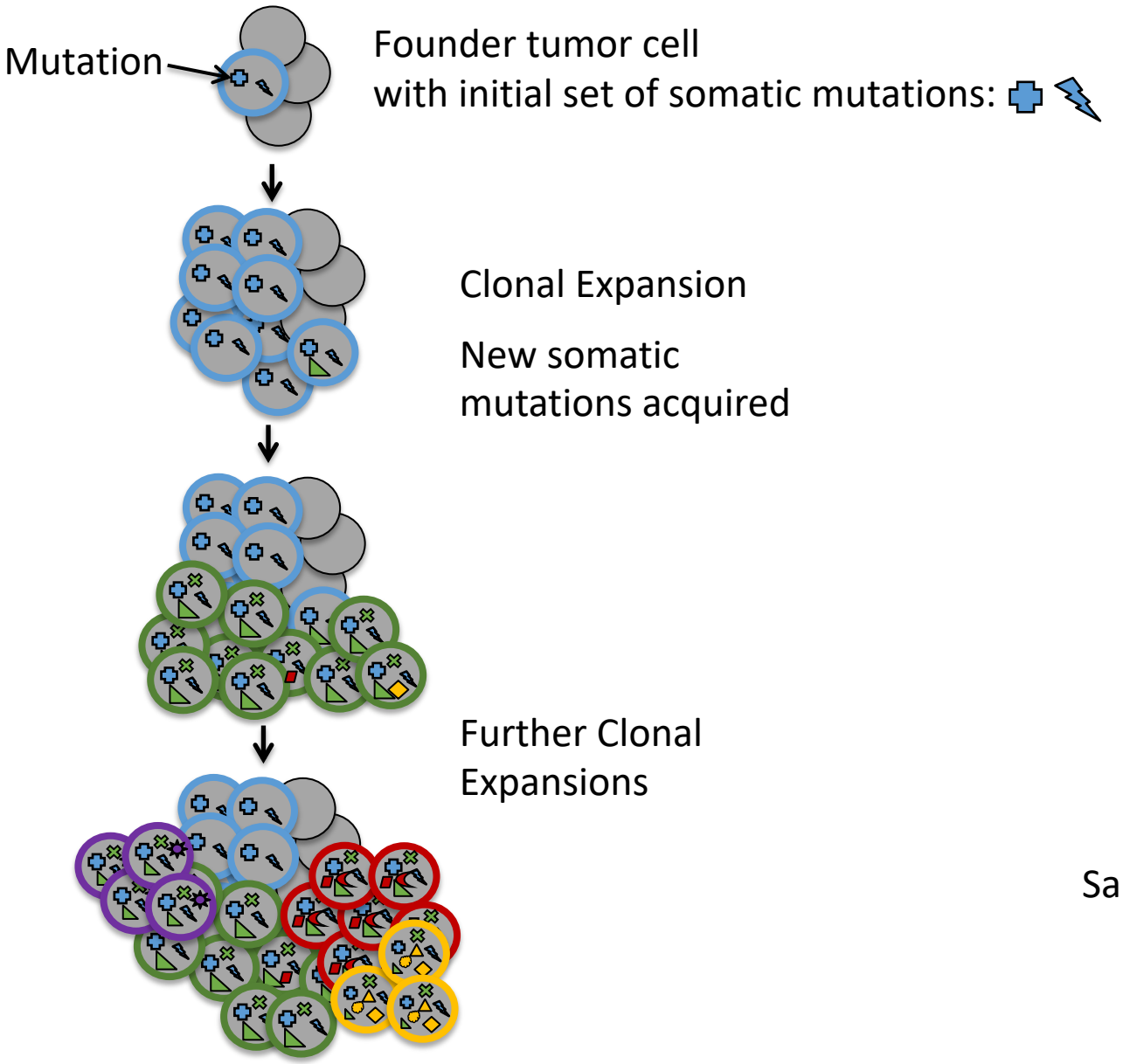
Figures adapted from Gundem et al. (Nature 2015), Sottoriva et al. (Nat Gen 2015)

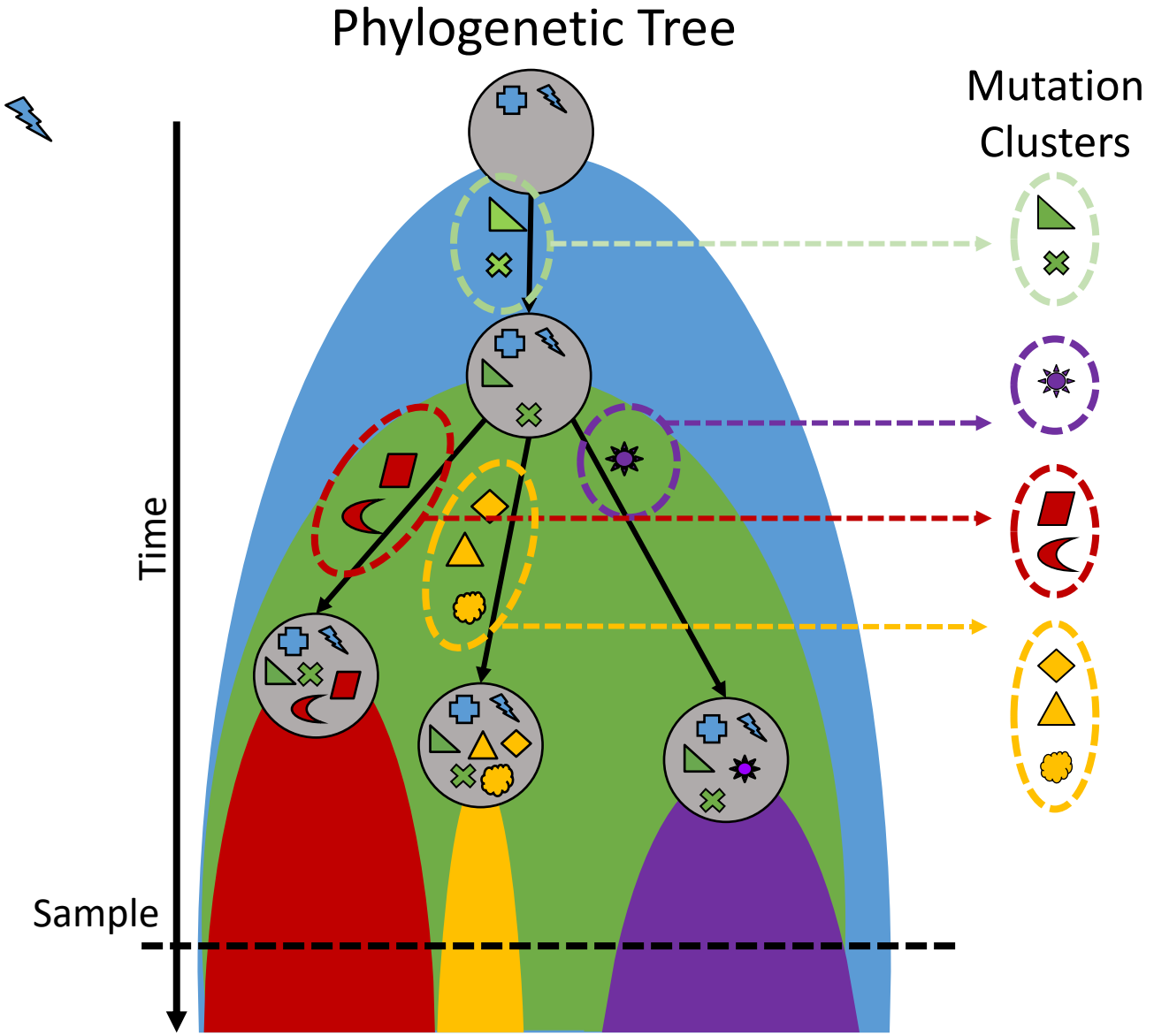
Cancer evolution is described by phylogenetic tree



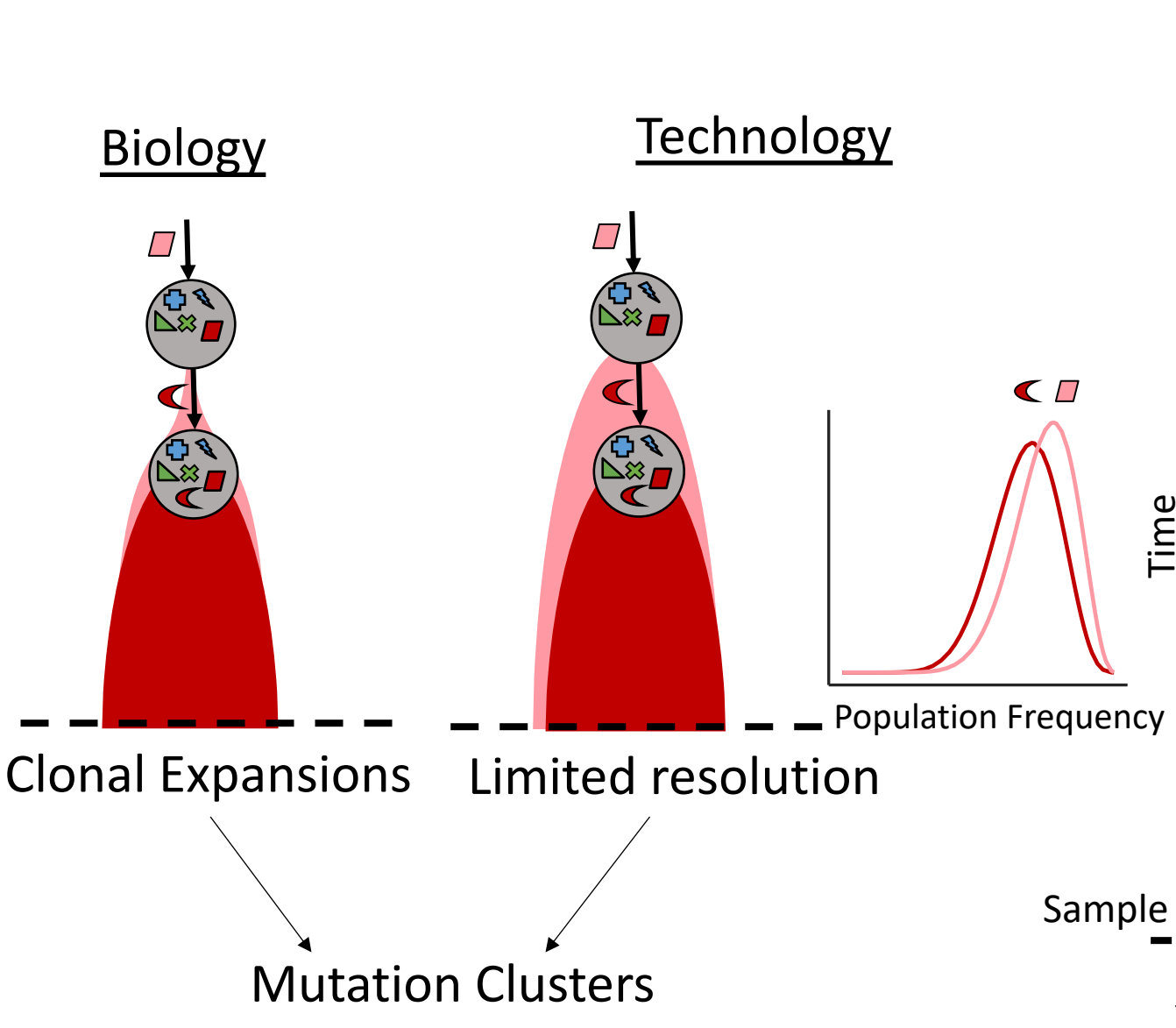


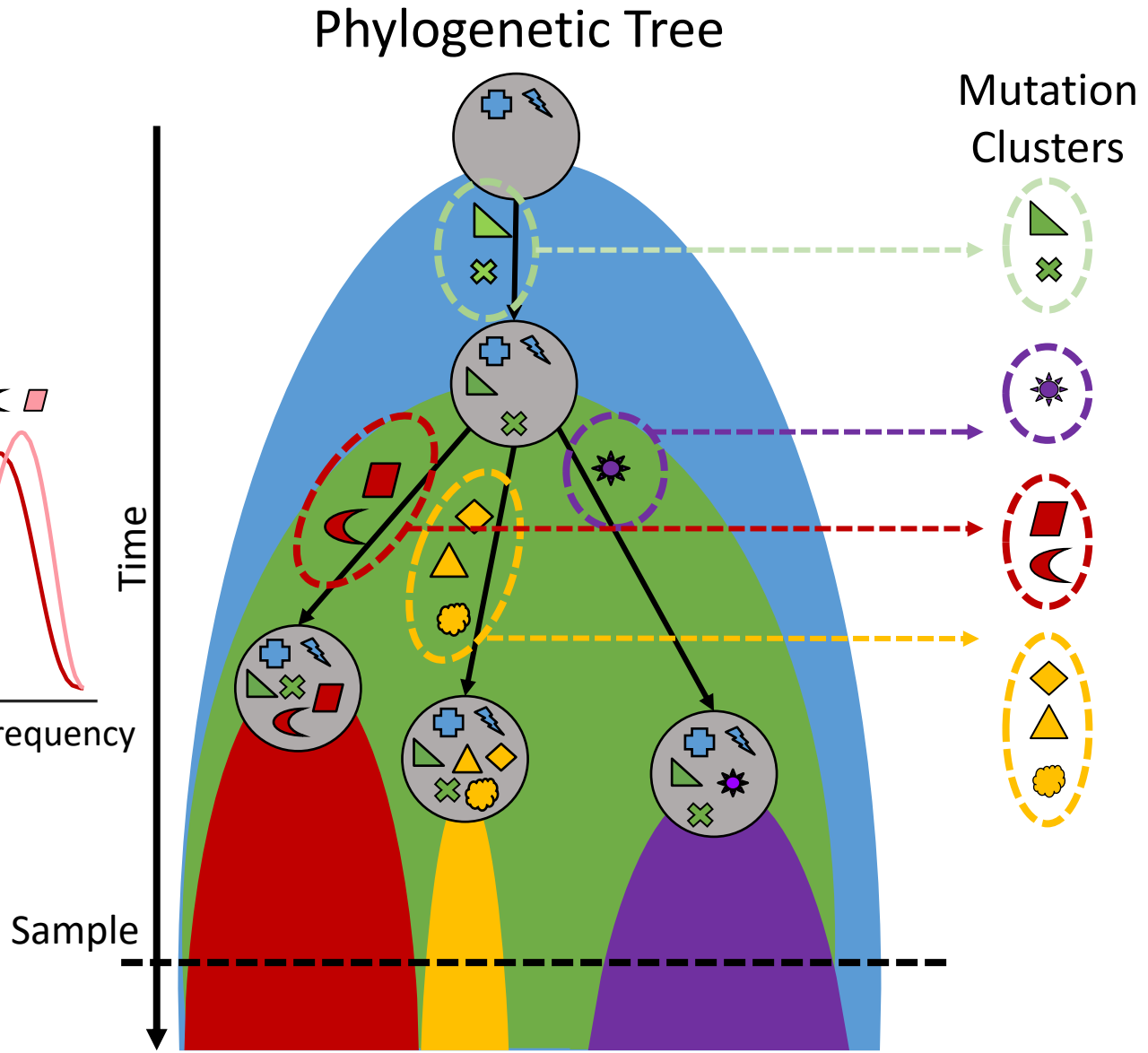
Cancer evolution is described by phylogenetic tree



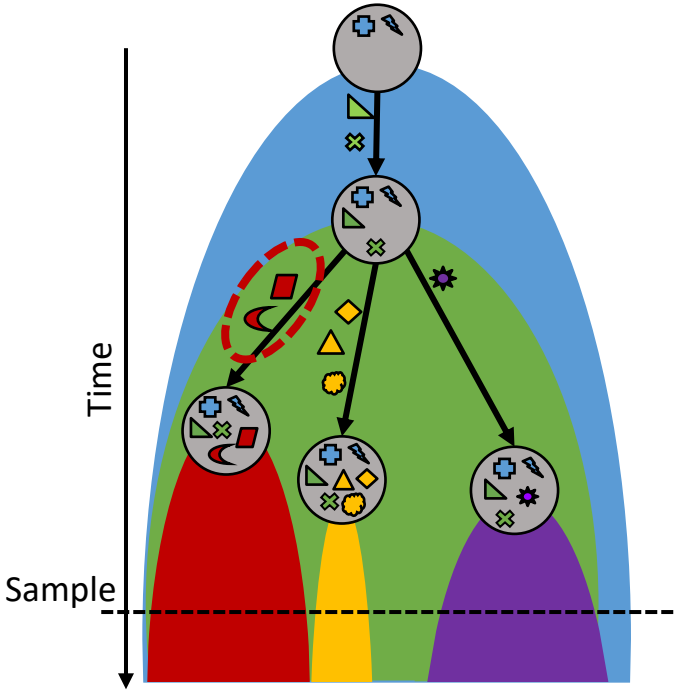


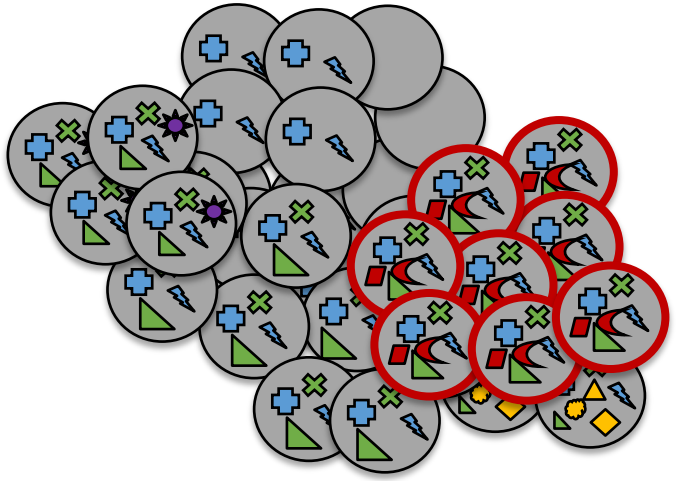
Clonal expansions → mutation clusters





Clustering mutations by VAF

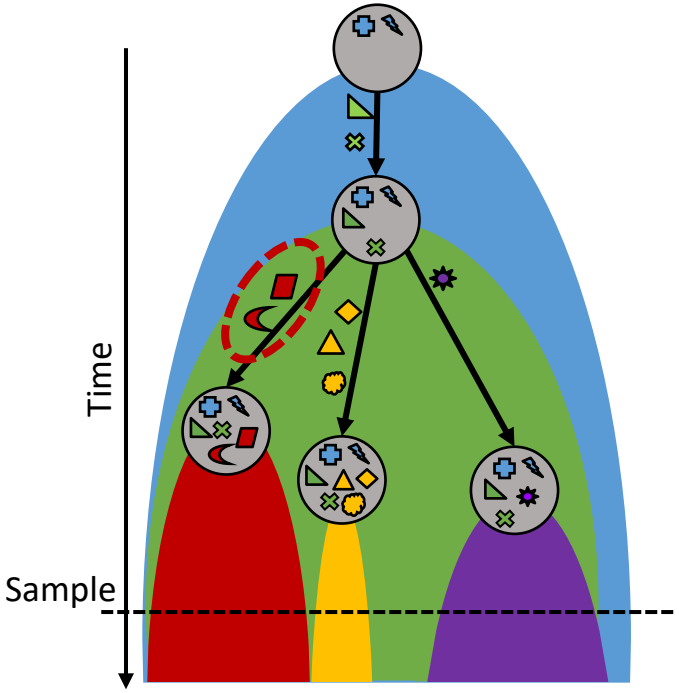


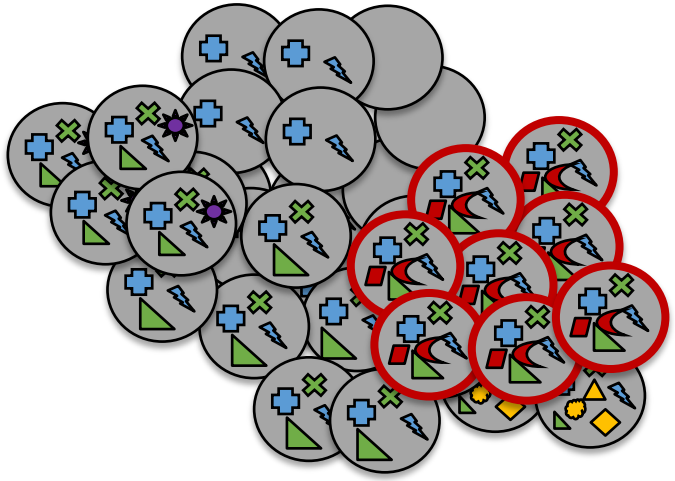


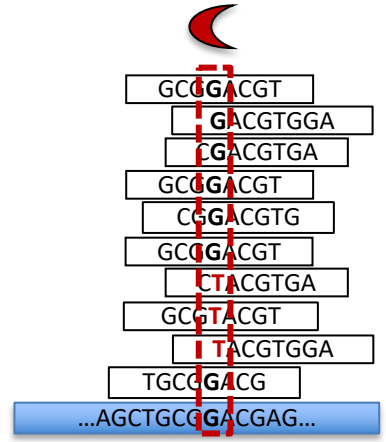
Cancer Cell Fraction (CCF): proportion of cancer cells with mutation

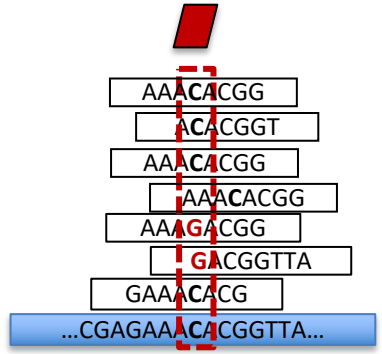
Same branch → Same CCF

Clustering mutations by VAF









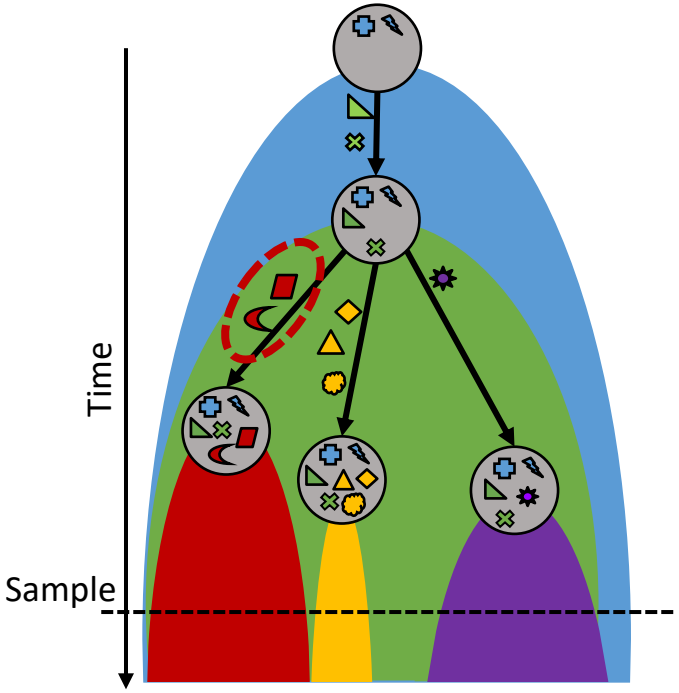
$VAF_{\text{C}} = 0.3$ $VAF_{\text{I}} \approx 0.3$

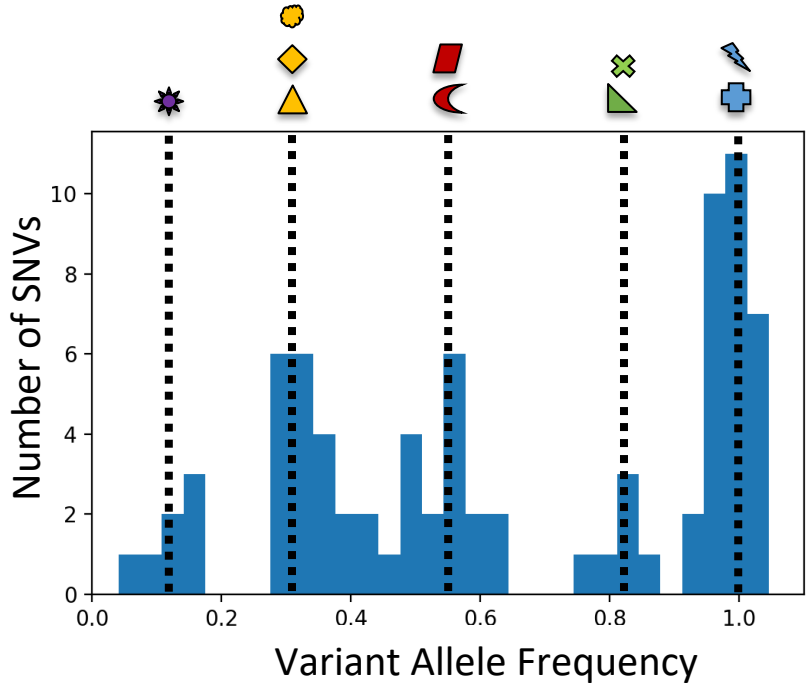
Cancer Cell Fraction (CCF): proportion of cancer cells with mutation

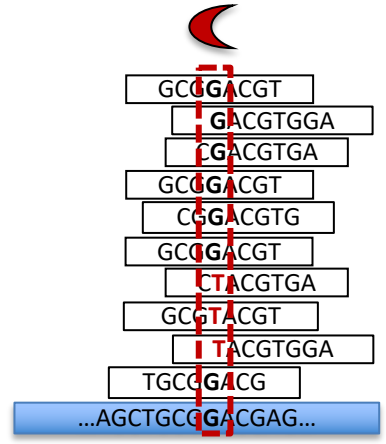
Variant Allele Frequency (VAF): proportion of reads with mutation

Same branch → Same CCF → Same VAF

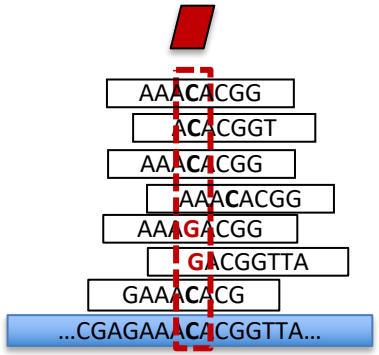
Clustering mutations by VAF







$VAF_{\color{red}{\text{C}}} = 0.3$



$VAF_{\color{red}{\text{I}}} \approx 0.3$

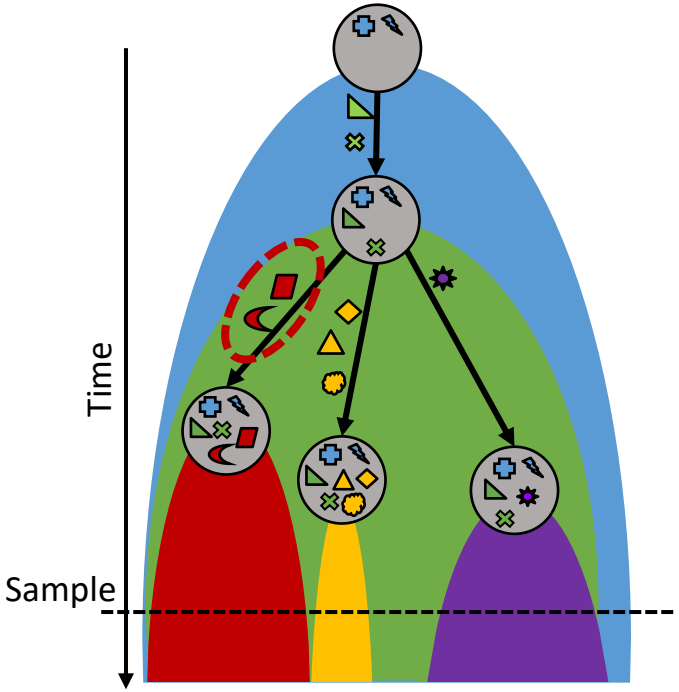
Variant Allele Frequency (VAF):
proportion of reads with mutation

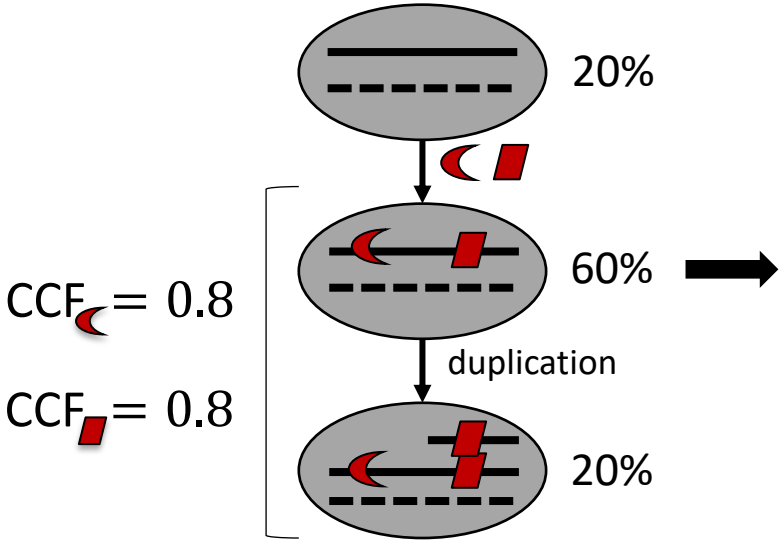
Same branch → Same CCF → Same VAF

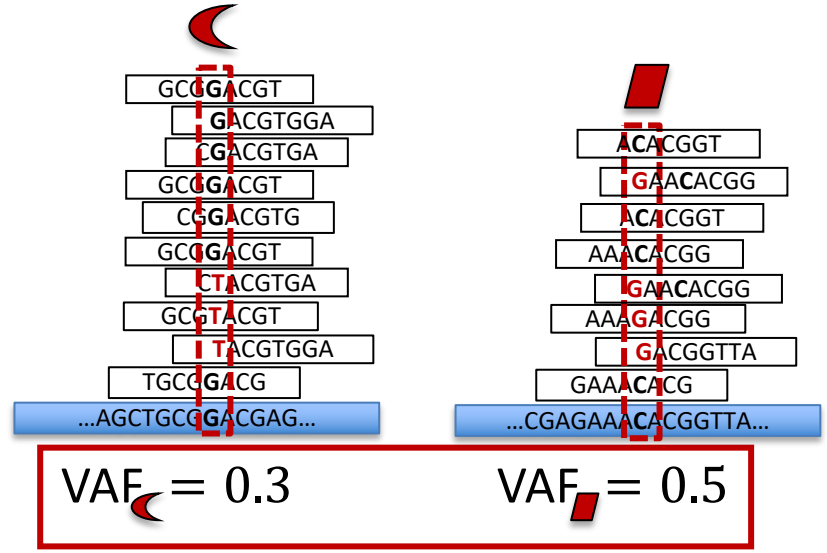
Cluster

SciClone (Miller et al., *PLOS CB* 2014), **Clomial** (Zare et al., *PLOS CB*, 2014), ...

Copy-number aberrations alter VAF:CCF correspondence







Variant Allele Frequency (VAF):
proportion of reads with mutation

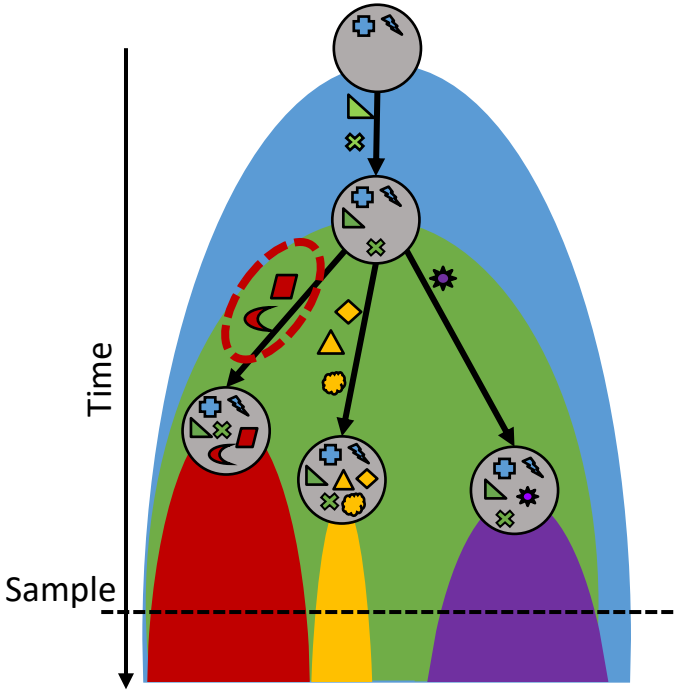
Copy-number aberrations

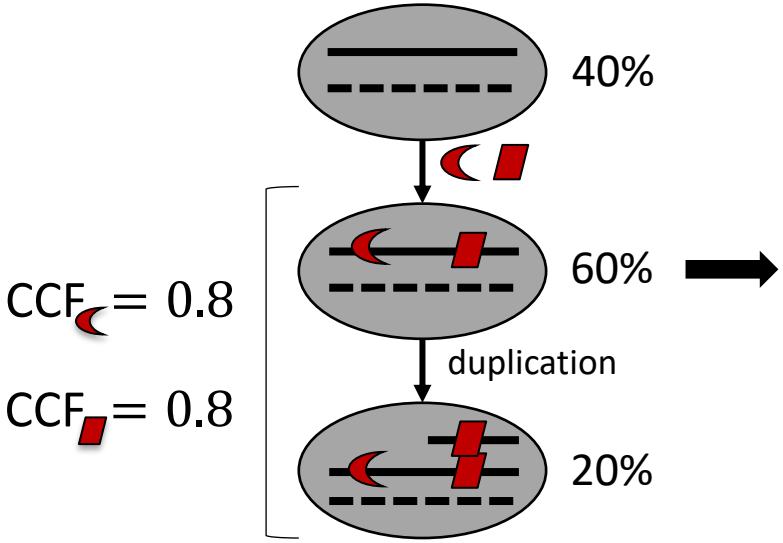
Same branch → Same CCF → Same VAF

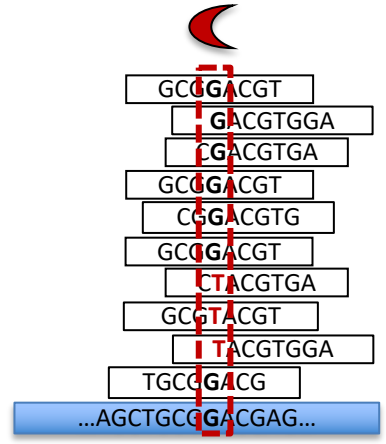
Cluster

SciClone (Miller et al., *PLOS CB* 2014), **Clomial** (Zare et al., *PLOS CB*, 2014), ...

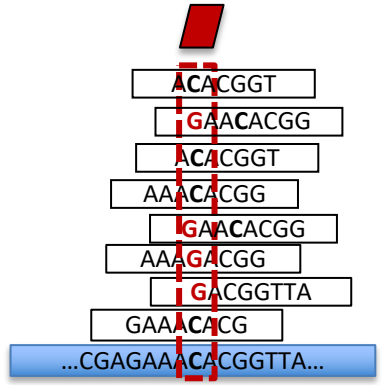
Clustering mutations by CCF







VAF = 0.3



VAF = 0.5

Same branch → Same CCF

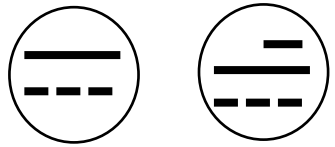
1

2 Cluster

Variant Allele Frequency (VAF)

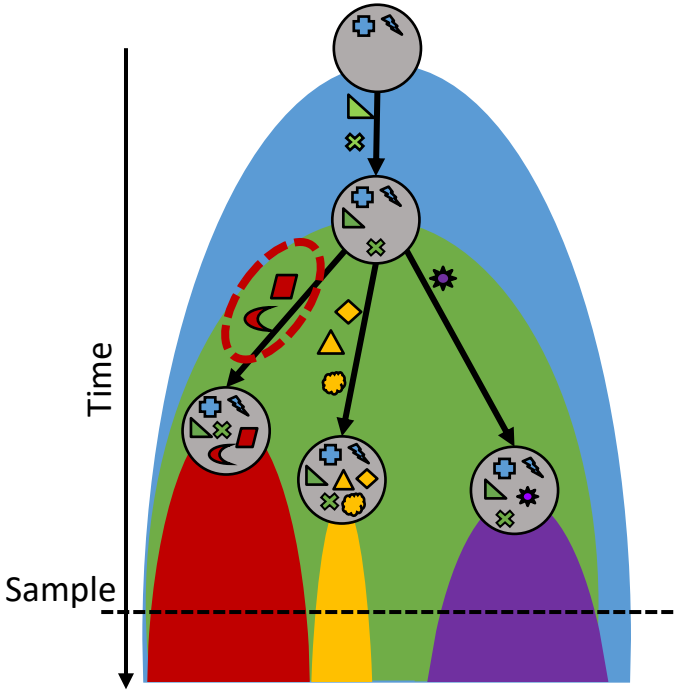
80% 20%

CNA Data



PyClone (Roth et al., *Nat Meth* 2014),
DPClust (Dentro et al., *Biorxiv* 2018), etc.

CCFs are not identifiable from VAFs & CNAs



Same branch → Same CCF

Cluster (2) ← Same CCF

Non-identifiability

$VAF_\triangle = 0.3$

80% 20%

20% 60% 20%

$CCF_\triangle = 0.6$

60% 20% 20%

$CCF_\triangle = 0.4$

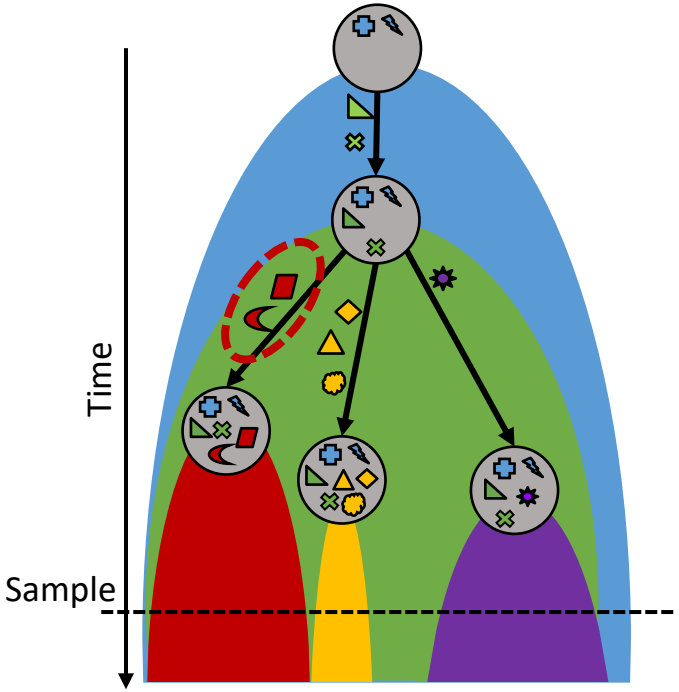
Same CCF ← (1) Variant Allele Frequency (VAF)

80% 20%

CNA Data

PyClone (Roth et al., *Nat Meth* 2014),
DPClust (Dentro et al., *Biorxiv* 2018), etc.

CCFs are not identifiable from VAFs & CNAs



Same branch → Same CCF

Cluster (2) ← Same CCF (1)

Non-identifiability

$VAF_\triangle = 0.3$

80% 20%

20% 60% 20%

$CCF_\triangle = 0.6$

60% 20% 20%

$CCF_\triangle = 0.4$

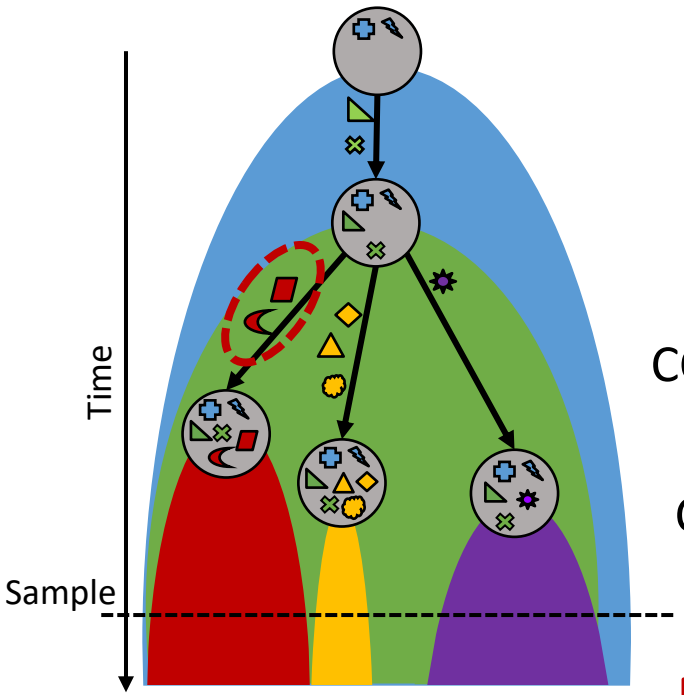
Variant Allele Frequency (VAF)

80% 20%

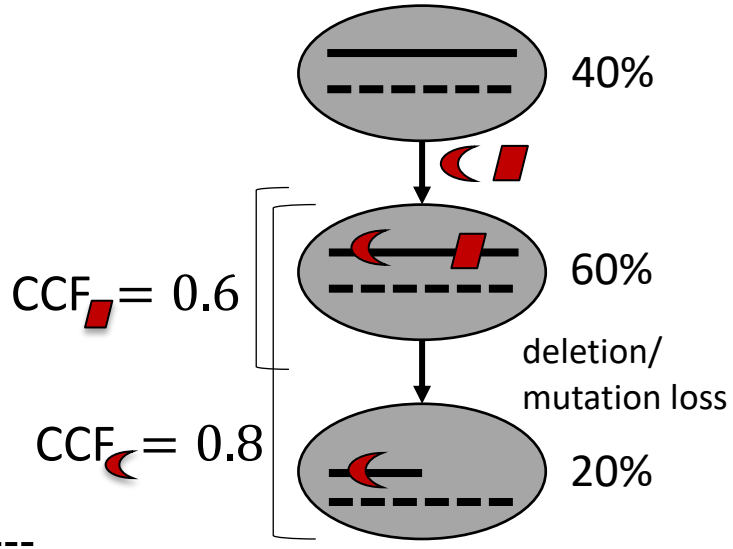
CNA Data

PyClone (Roth et al., *Nat Meth* 2014),
DPClust (Dentro et al., *Biorxiv* 2018), etc.

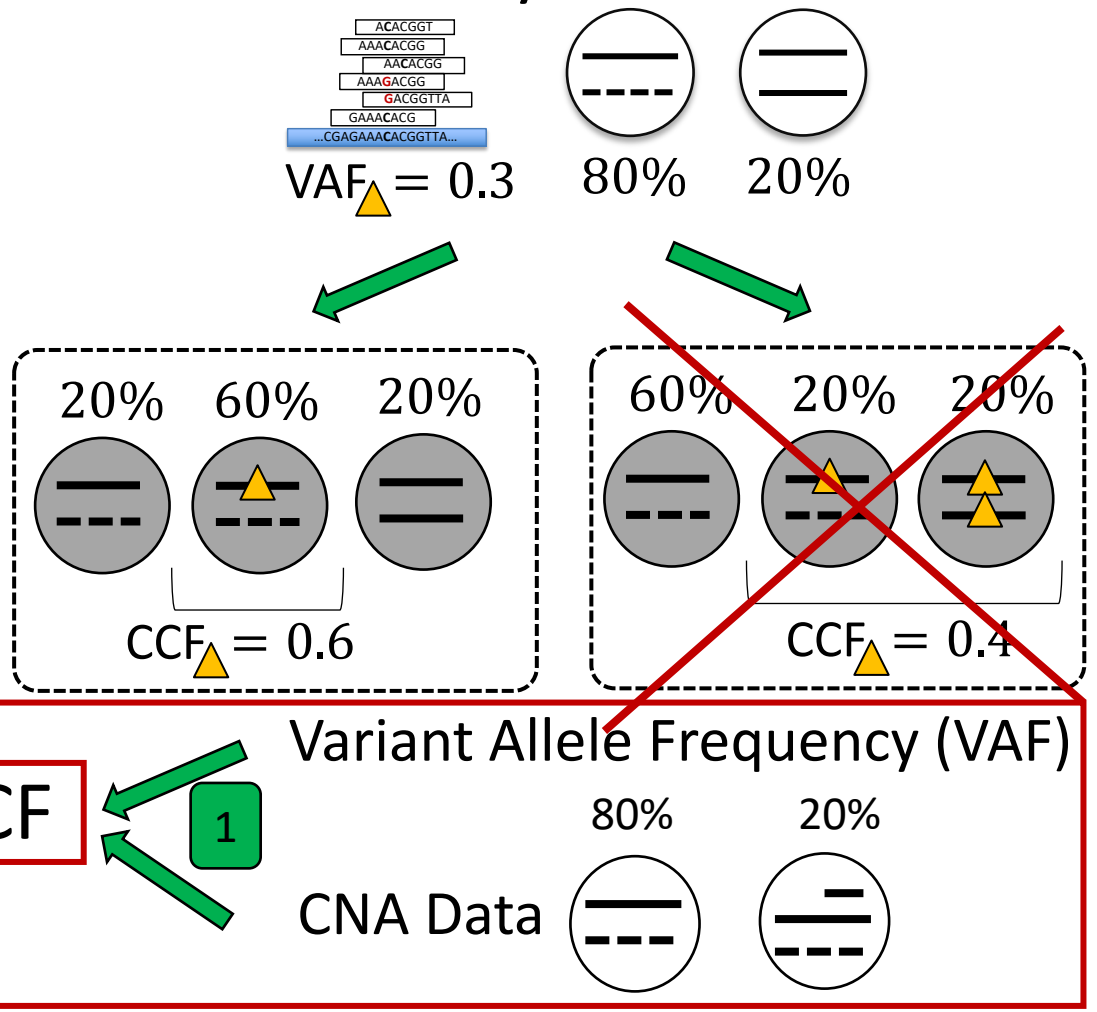
Mutation losses alter branch:CCF relationship



Mutation Losses

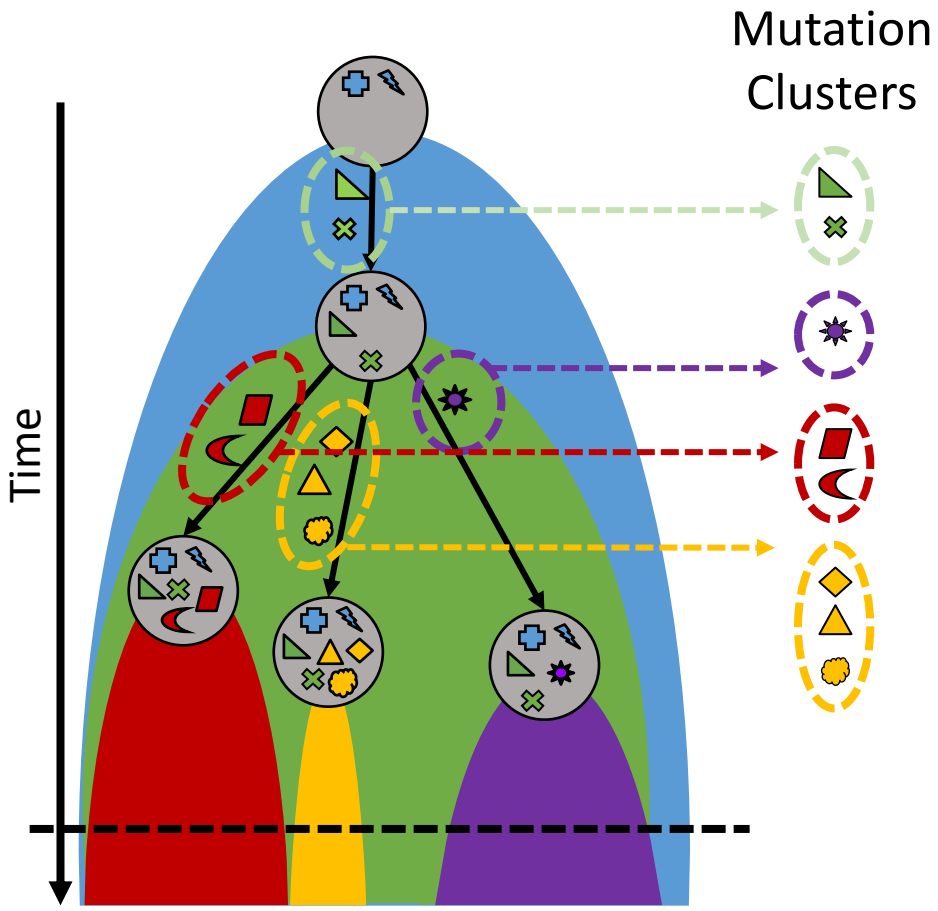


Non-identifiability



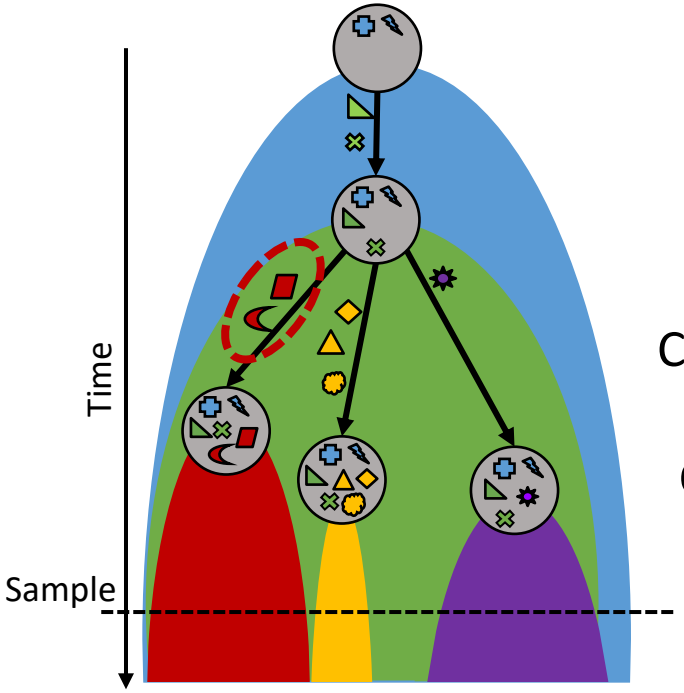
PyClone (Roth et al., *Nat Meth* 2014),
DPClust (Dentro et al., *Biorxiv* 2018), etc.

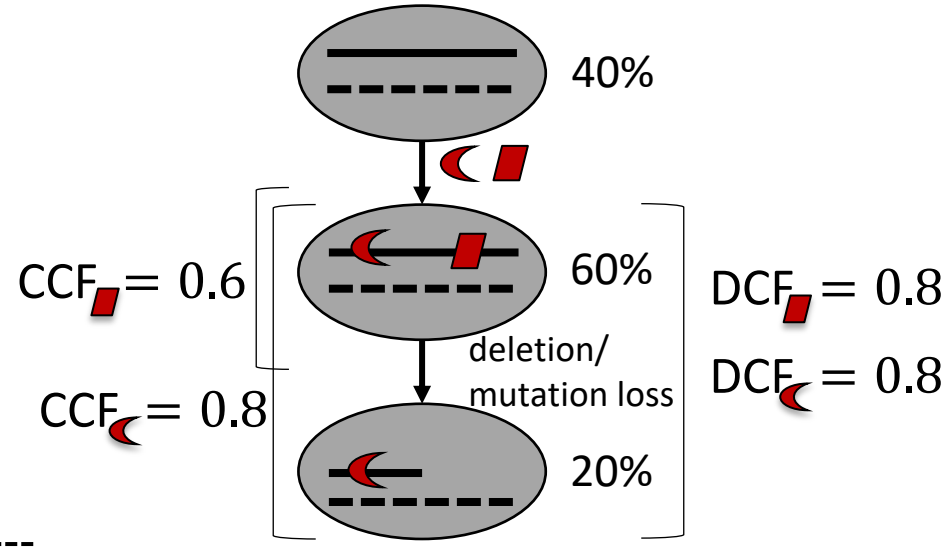
Summary: Current approaches for mutation clustering



Quantity:	VAF	CCF
Identifiable	YES	NO
Accounts for copy-number aberrations	NO	YES
Accounts for mutation losses	NO	NO

New Approach: Descendent Cell Fraction (DCF) accounts for mutation losses





$CCF_{\blacksquare} = 0.6$

$CCF_{\leftarrow} = 0.8$

$DCF_{\blacksquare} = 0.8$

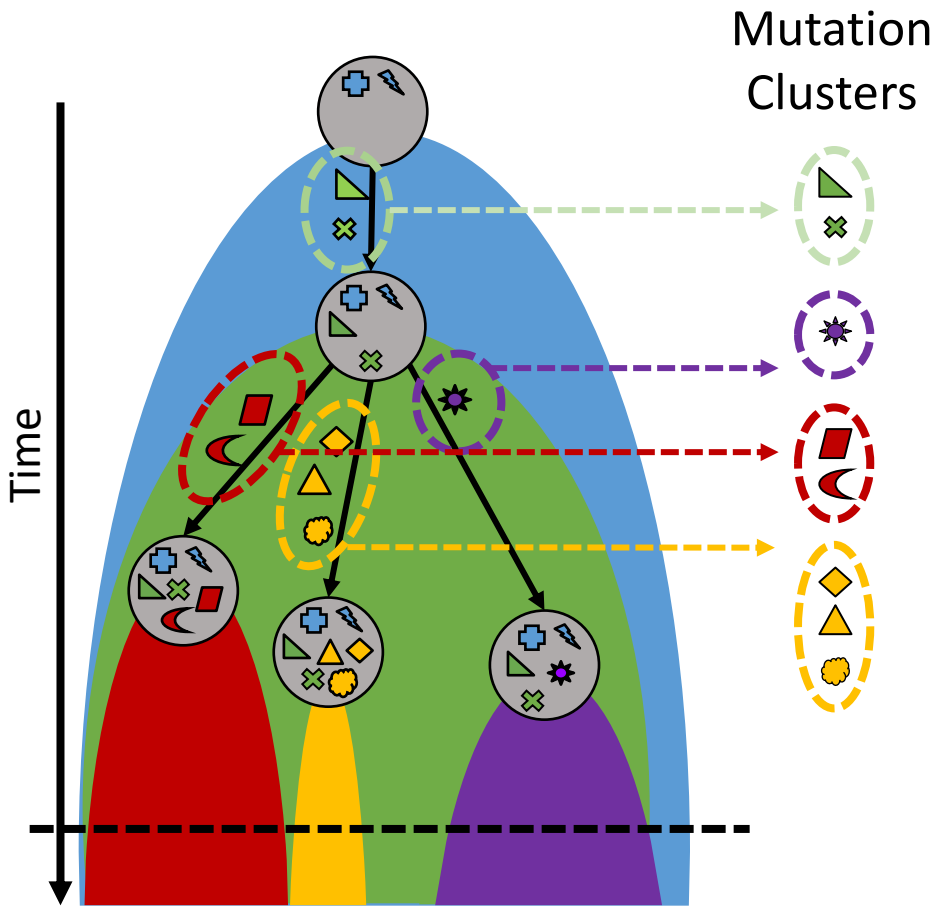
$DCF_{\leftarrow} = 0.8$

Cancer Cell Fraction (CCF):
proportion of extant cells *containing SNV*

Descendent Cell Fraction (DCF):
proportion of extant cells *descending from cell* that first introduced SNV

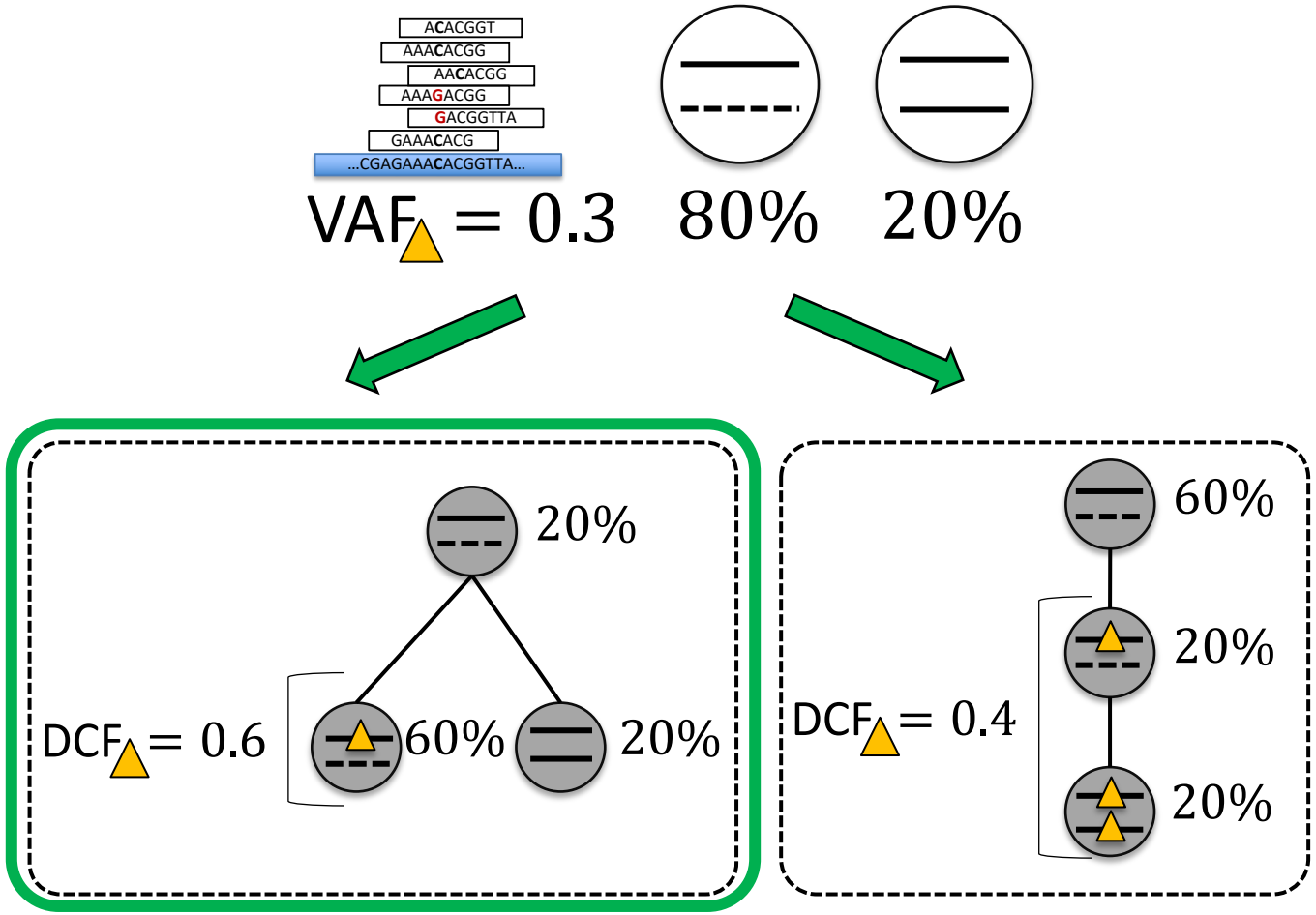
Same branch → Same DCF

Summary: Mutation Clustering Approaches

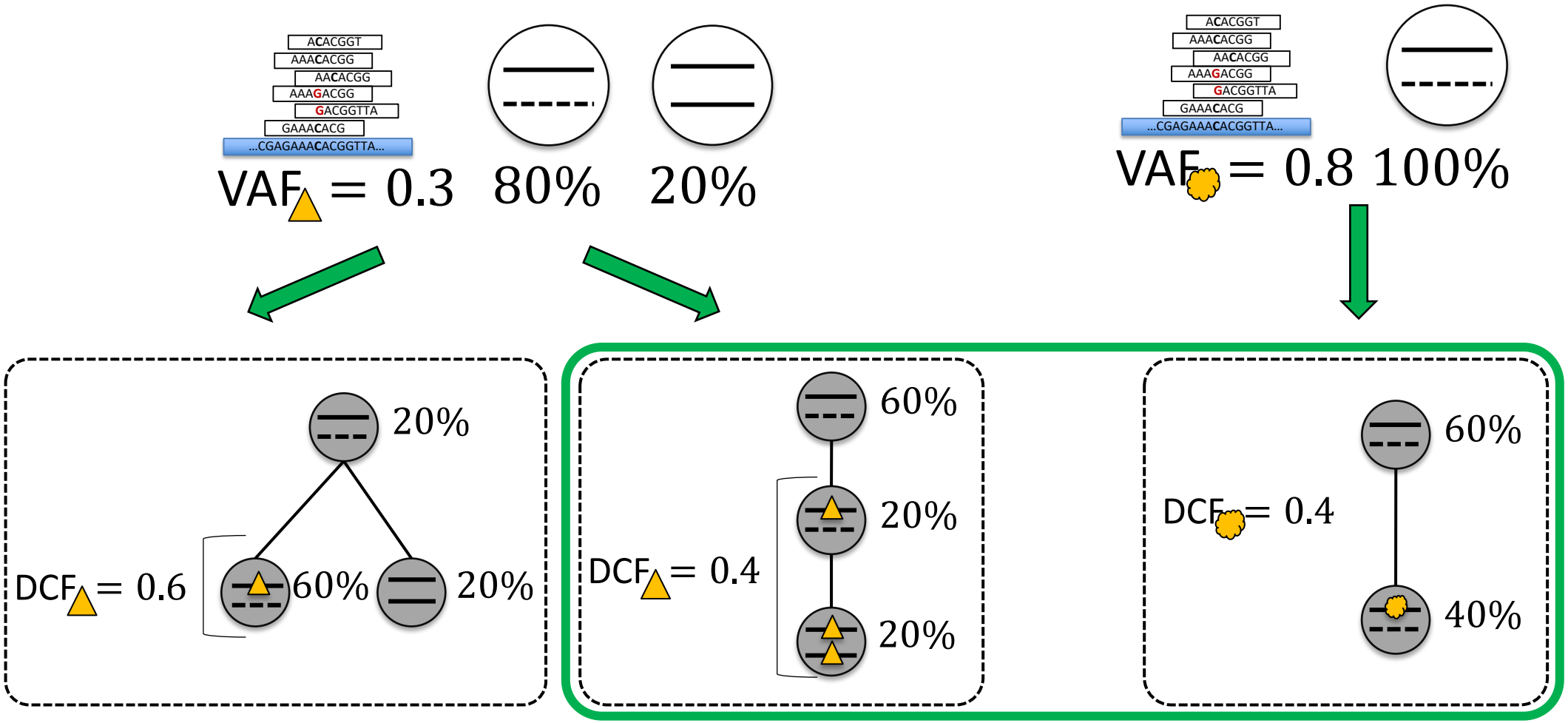


Quantity:	VAF	CCF	DCF
Identifiable	YES	NO	NO
Accounts for copy-number aberrations	NO	YES	YES
Accounts for mutation losses	NO	NO	YES

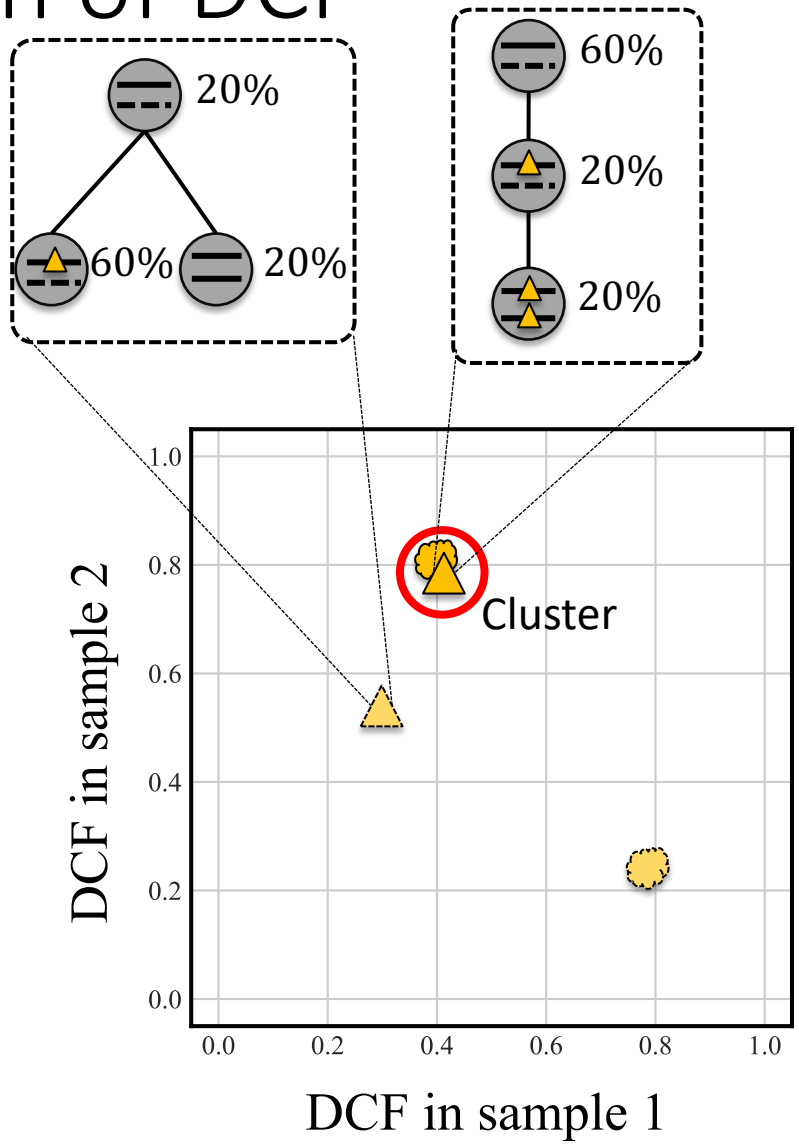
DeCiFer: *Simultaneous* clustering of mutations and calculation of DCF

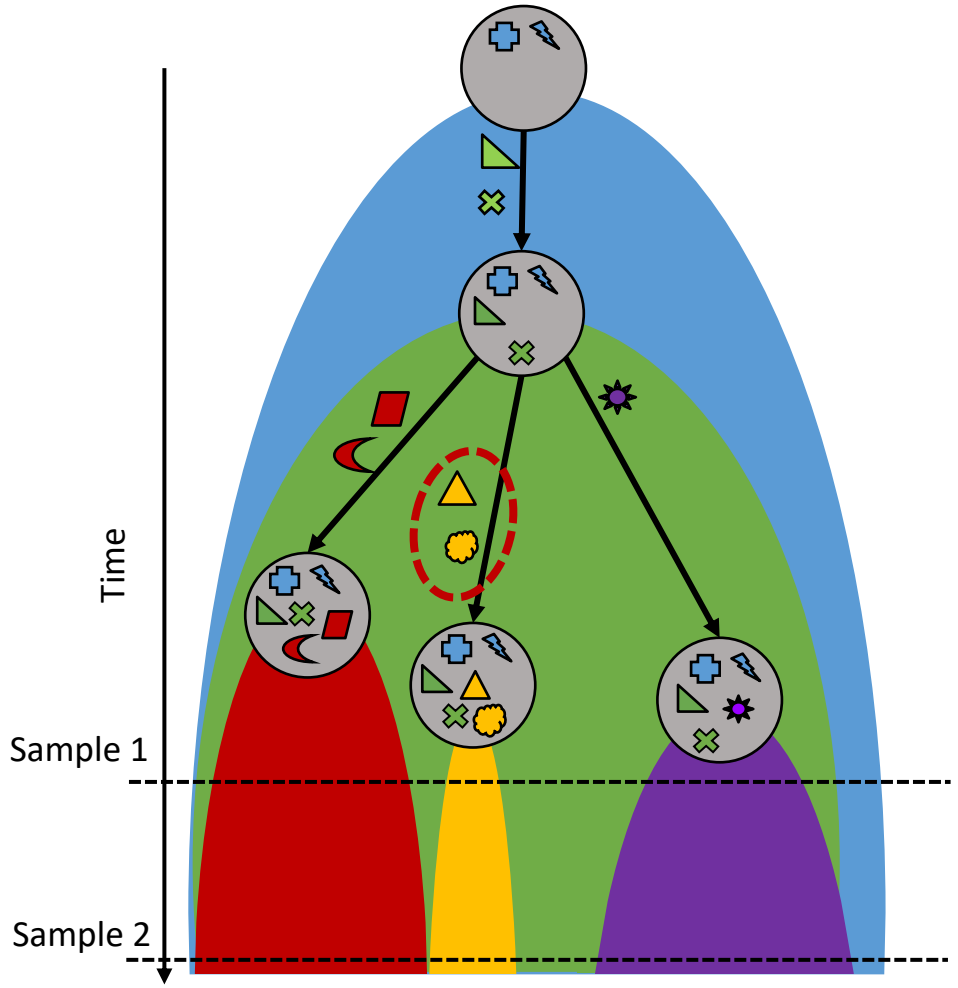


DeCiFer: *Simultaneous* clustering of mutations and calculation of DCF



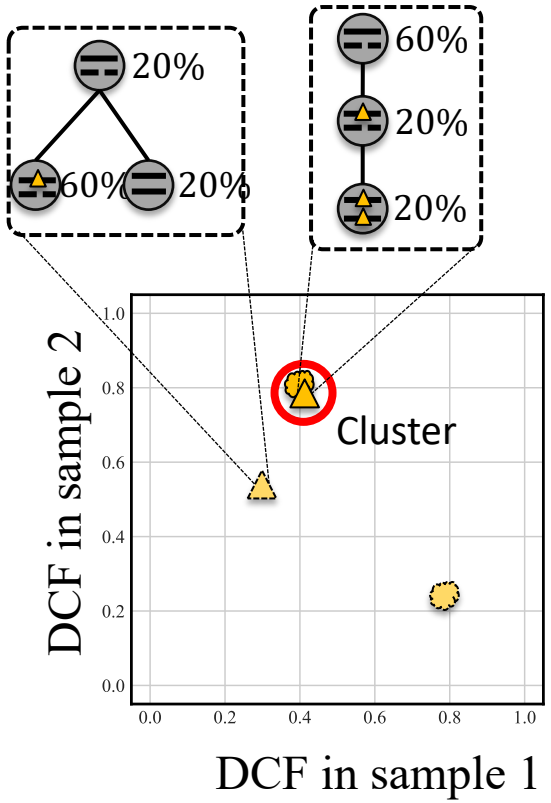
DeCiFer: *Simultaneous* clustering of mutations and calculation of DCF

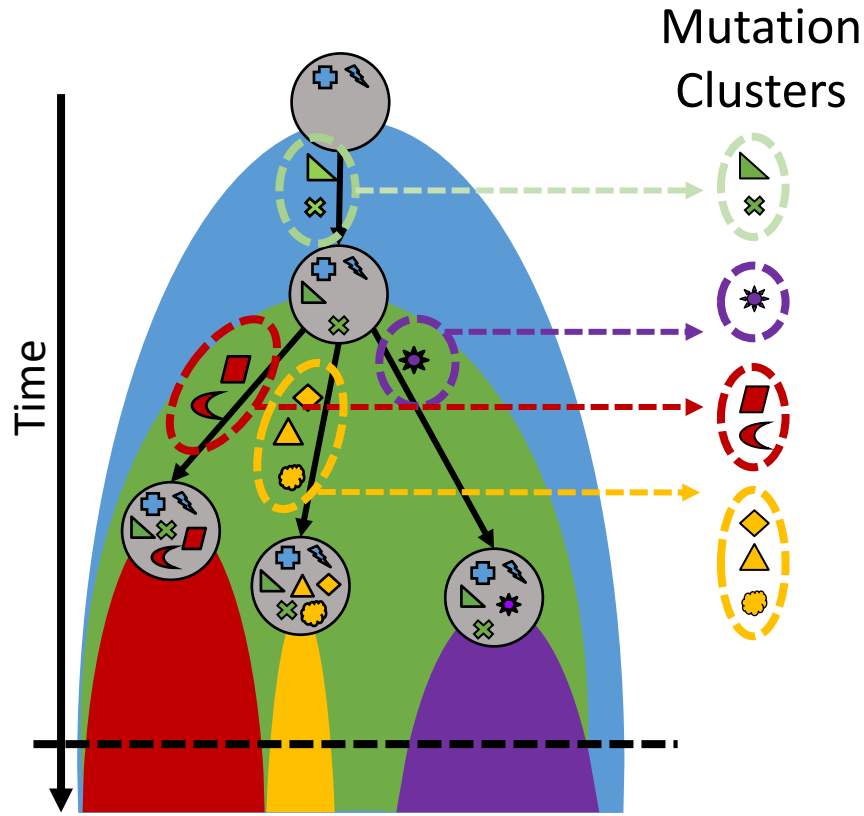




DeCiFer: DCF Clustering Algorithm

EM-based algorithm using Linear Programming performs simultaneous clustering and DCF assignment



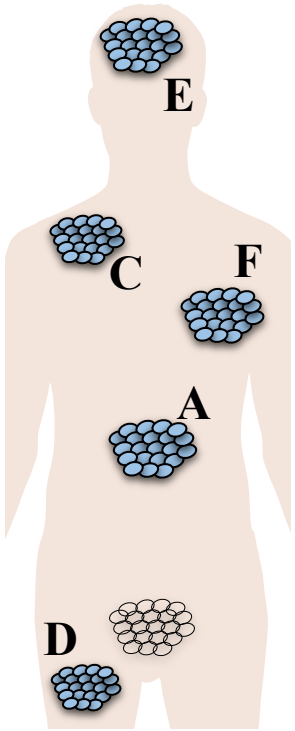


DCF accounts for mutation losses

Simultaneous clustering across multiple samples handles non-identifiability of DCF values

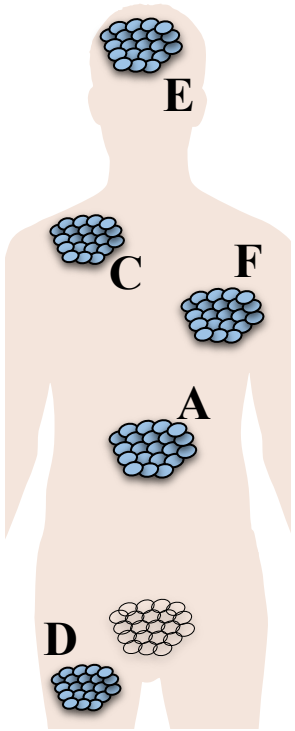
Prostate Cancer Patient A17

Metastatic prostate cancer patient A17:
198 SNVs from 5 bulk tumor samples
(Gundem et al., *Nature* 2015)



Prostate Cancer Patient A17

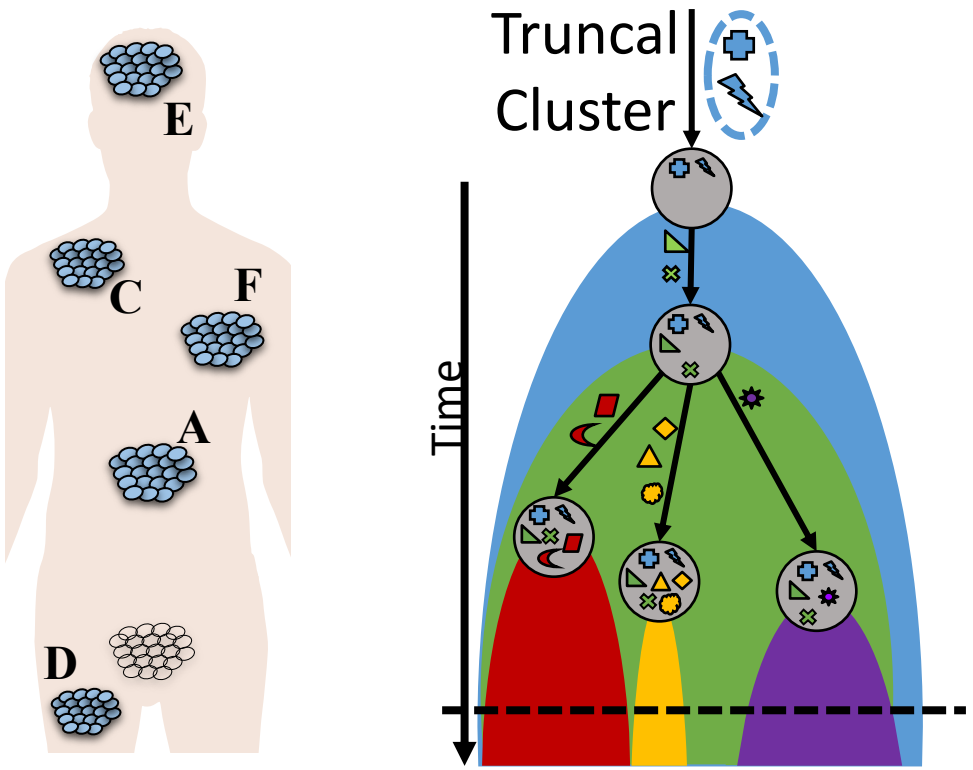
Metastatic prostate cancer patient A17:
198 SNVs from 5 bulk tumor samples
(Gundem et al., *Nature* 2015)



Mutation clusters:
Published: 13 clusters
DeCiFer: 5 clusters

Prostate Cancer Patient A17

Metastatic prostate cancer patient A17:
198 SNVs from 5 bulk tumor samples
(Gundem et al., *Nature* 2015)

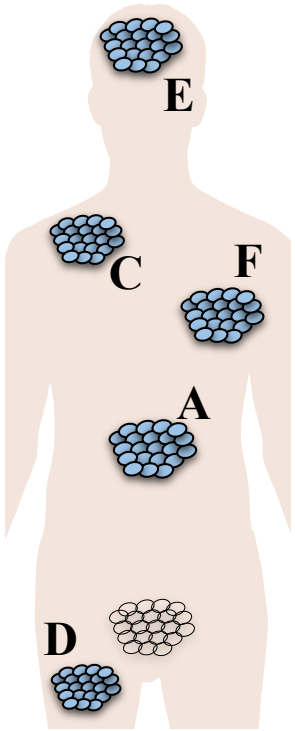


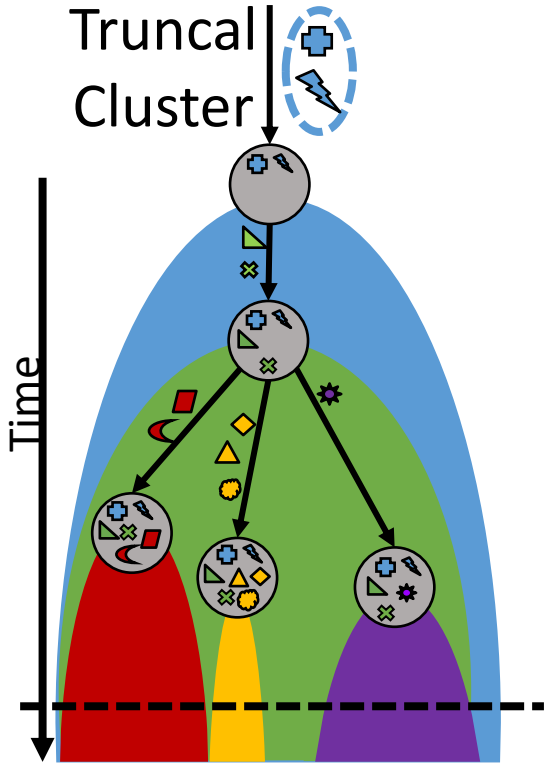
Truncal Cluster:
Published: 111 mutations
DeCiFer: 129 mutations

Mutation clusters:
Published: 13 clusters
DeCiFer: 5 clusters

Prostate Cancer Patient A17

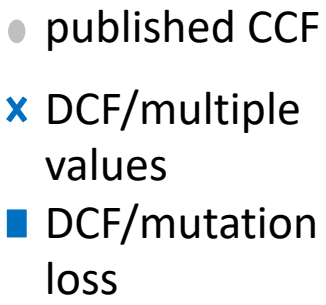
Metastatic prostate cancer patient A17:
198 SNVs from 5 bulk tumor samples
(Gundem et al., *Nature* 2015)

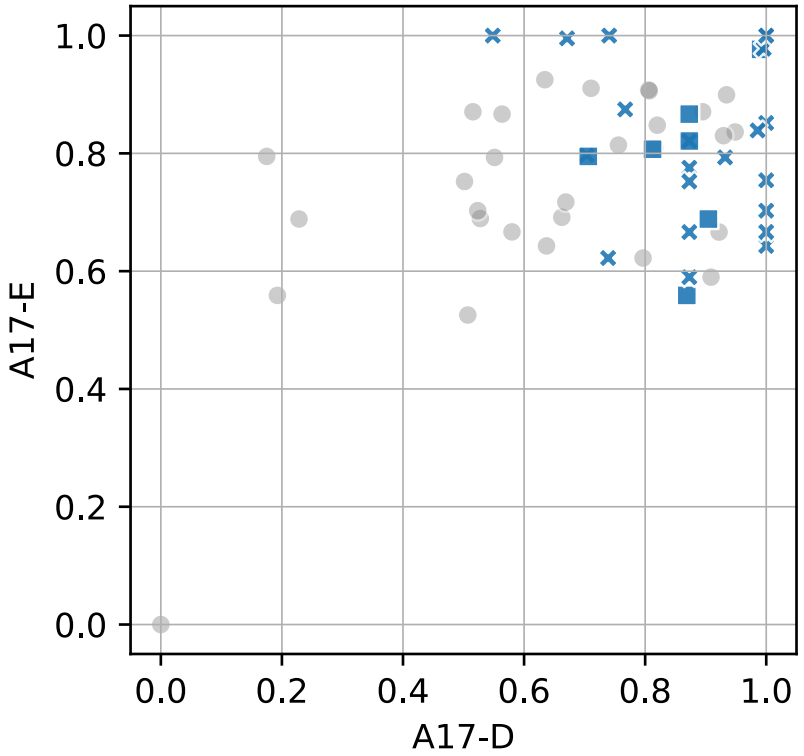




Truncal Cluster:
Published: 111 mutations
DeCiFer: 129 mutations

111 published | 12 multiple values | 6 mutation loss

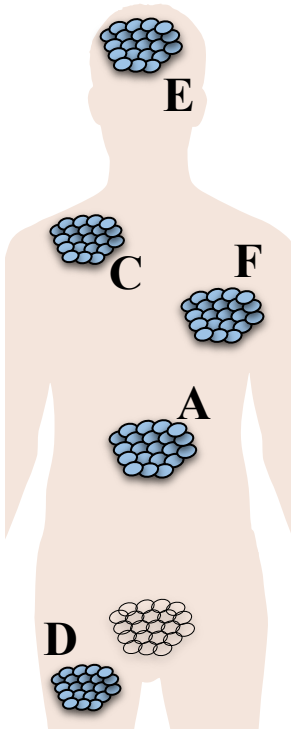


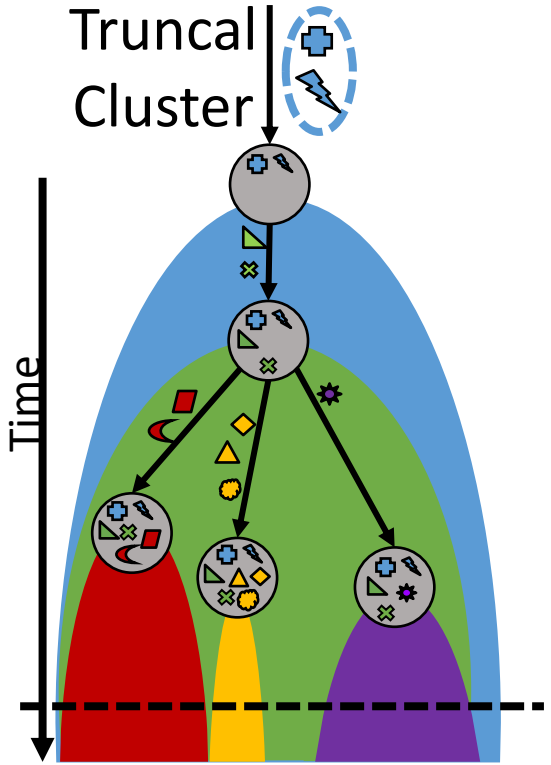


Mutation clusters:
Published: 13 clusters
DeCiFer: 5 clusters

Prostate Cancer Patient A17

Metastatic prostate cancer patient A17:
198 SNVs from 5 bulk tumor samples
(Gundem et al., *Nature* 2015)



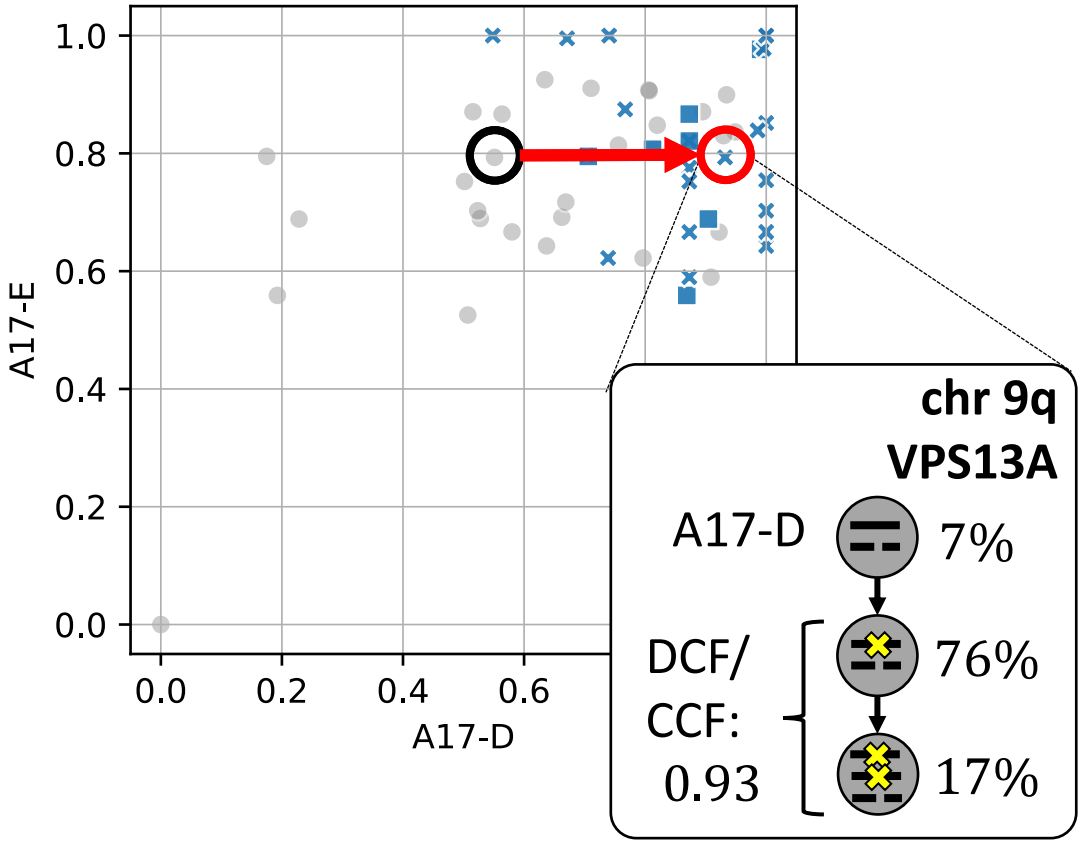


Mutation clusters:
Published: 13 clusters
DeCiFer: 5 clusters

Truncal Cluster:
Published: 111 mutations
DeCiFer: 129 mutations

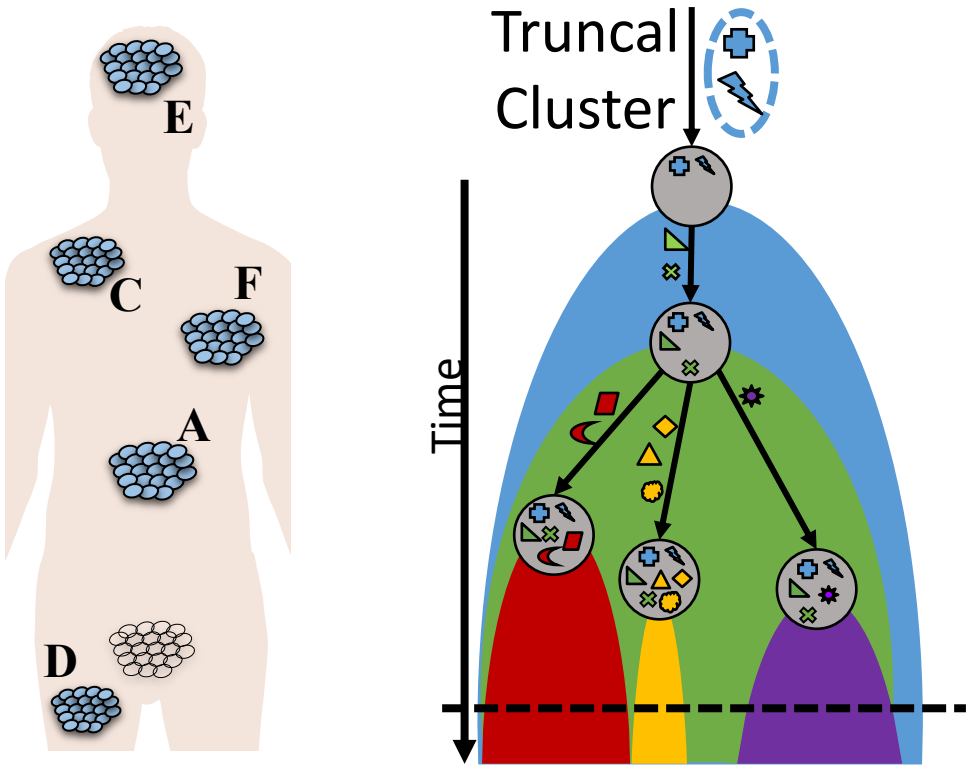
111 published | 12 multiple values | 6 mutation loss

published CCF
DCF/multiple values
DCF/mutation loss



Prostate Cancer Patient A17

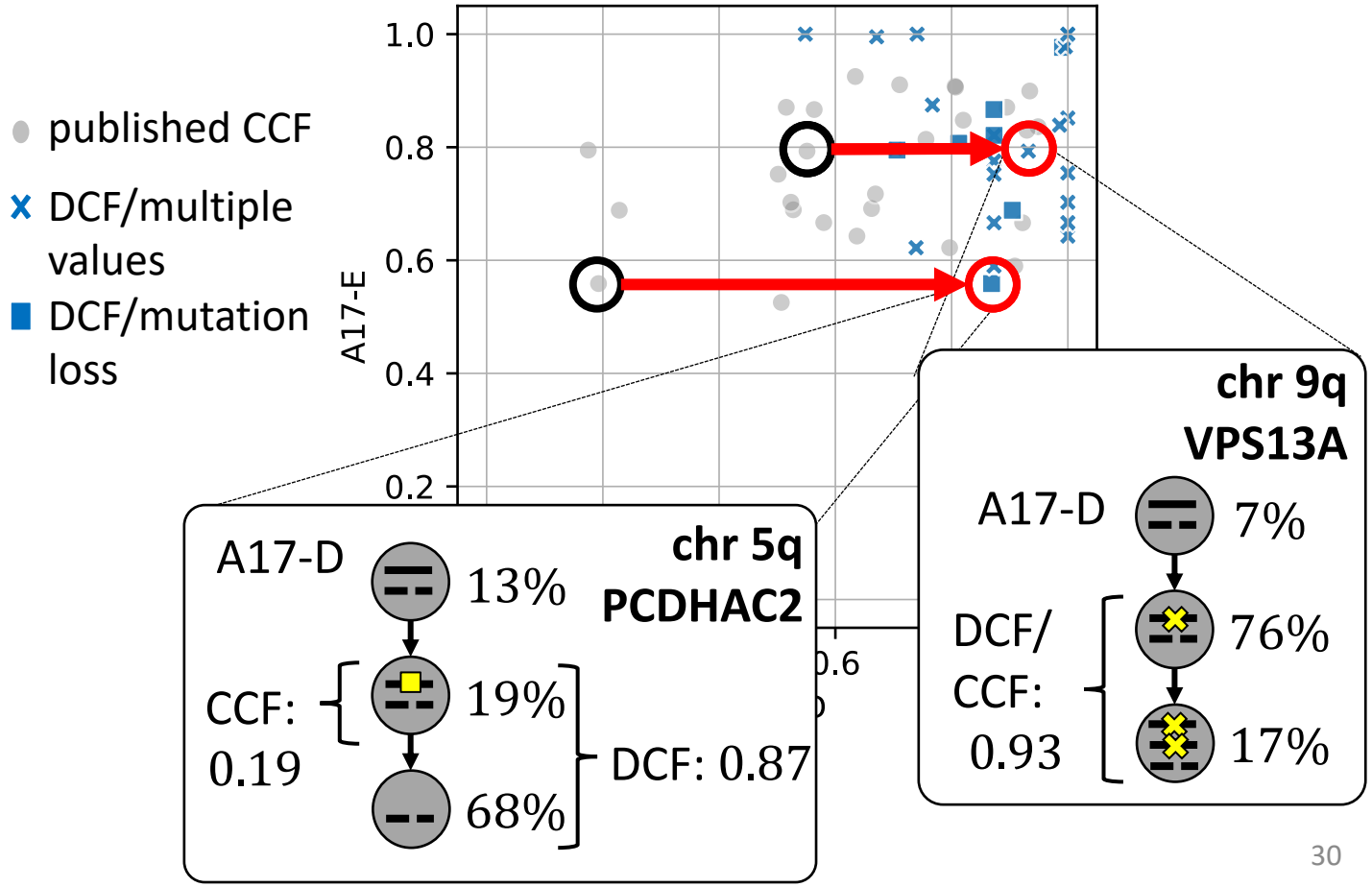
Metastatic prostate cancer patient A17:
198 SNVs from 5 bulk tumor samples
(Gundem et al., *Nature* 2015)



Mutation clusters:
Published: 13 clusters
DeCiFer: 5 clusters

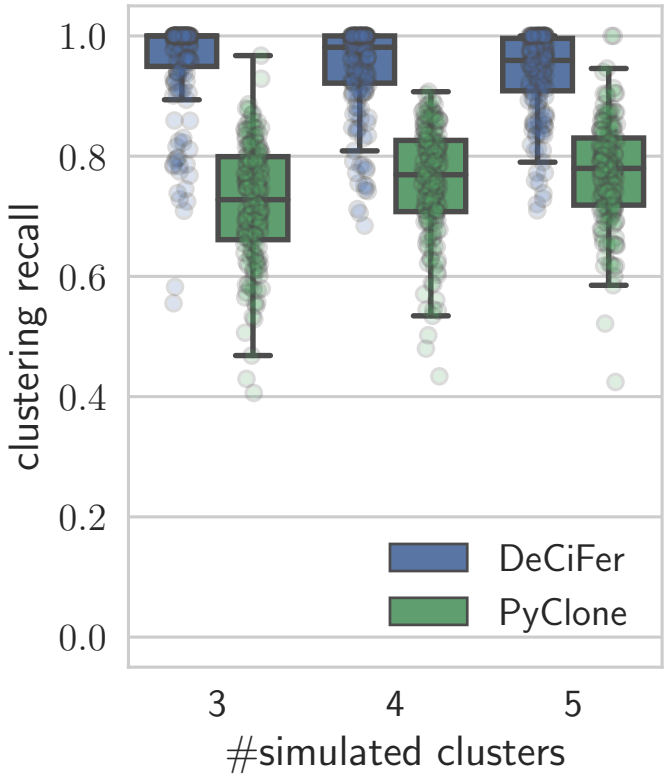
Truncal Cluster:
Published: 111 mutations
DeCiFer: 129 mutations

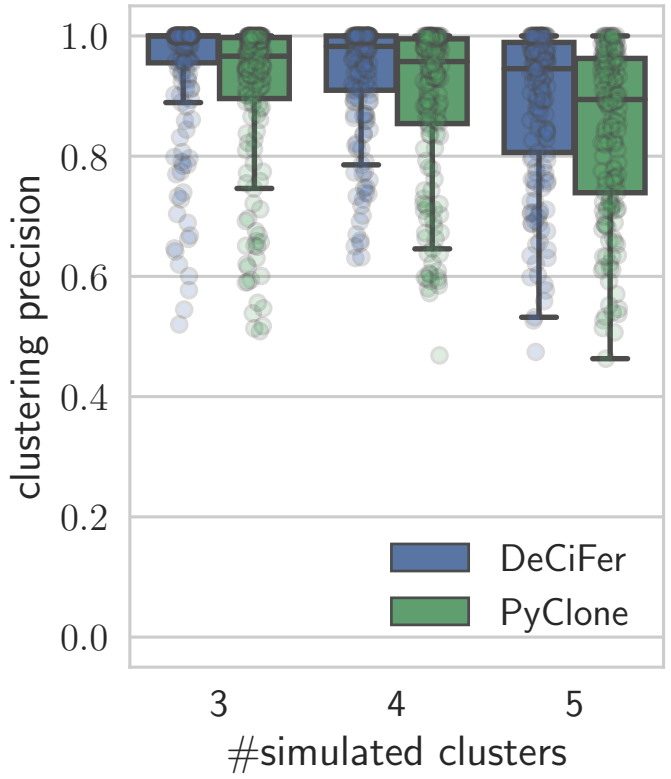
111 published | 12 multiple values | 6 mutation loss

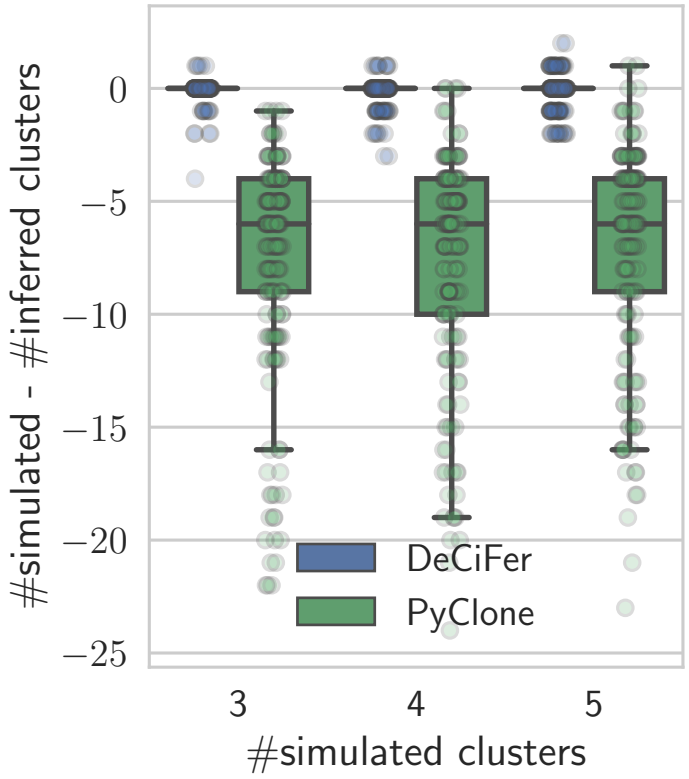


Simulations

180 instances for each #clusters $k \in \{3, 4, 5\}$ with copy number loss (#SNVs $\in \{50, 100, 500\}$, #samples $\in \{2, 3, 4\}$)







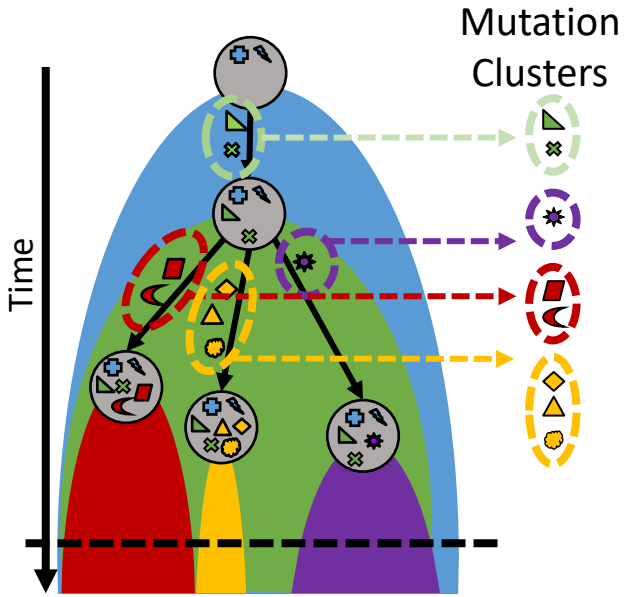
1. DeCifer improves recall

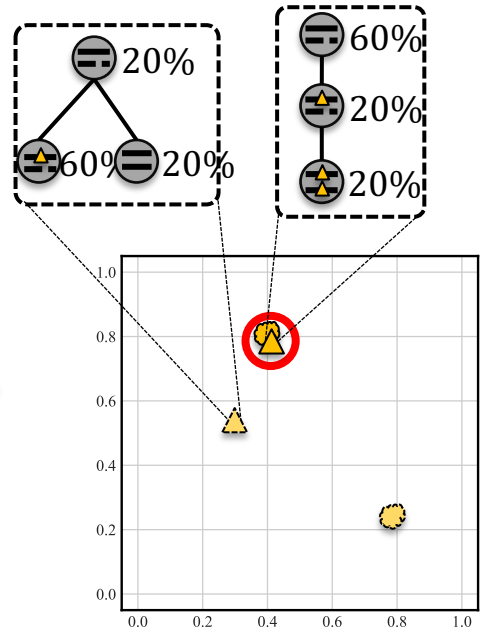
2. w/o loss of precision

3. more accurately infers # of clusters

Conclusions

- Mutation clusters indicate SNVs present on same phylogenetic branch
- Clustering mutations by VAF or CCF does not account for mutation losses
- **Descendent Cell Fraction (DCF**) accounts for mutation losses
- **DeCiFer** simultaneously clusters and calculates DCF values to better handle non-identifiability
- **DeCiFer** leads to more accurate clustering on simulated data, and simpler tumor composition on prostate cancer data





Acknowledgements

Mohammed El-Kebir

Simone Zaccaria

Ben Raphael

Raphael Group