PhyloSub

Jiao *et. al. BMC Bioinformatics* 2014, 15:35

Background

- Genetically-diverse subclonal populations of cells in tumors

- Can reconstruct evolutionary history of tumor

- SNV’s (single nucleotide variants)
    - Limitation: frequencies measured independently

Method

- Infinite sites assumption
- Topological constraint rules
- PhyloSub – infers tumor phylogenies from SNV allele frequency
- Partial order plot

Assumptions and Rules of Model

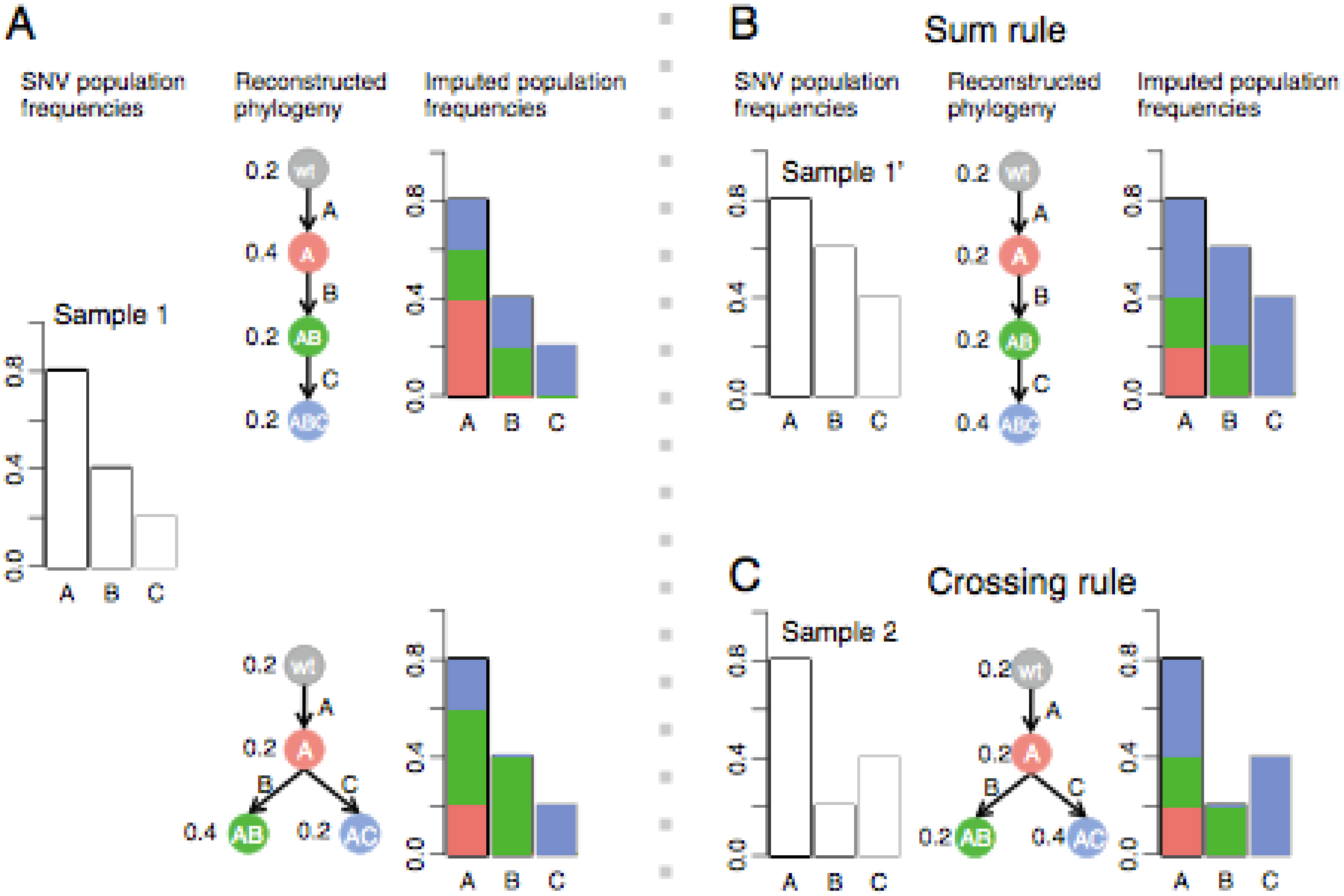
- Clonal evolution theory
- Infinite sites assumption
- Need at least two tumor samples to rule out a linear phylogeny

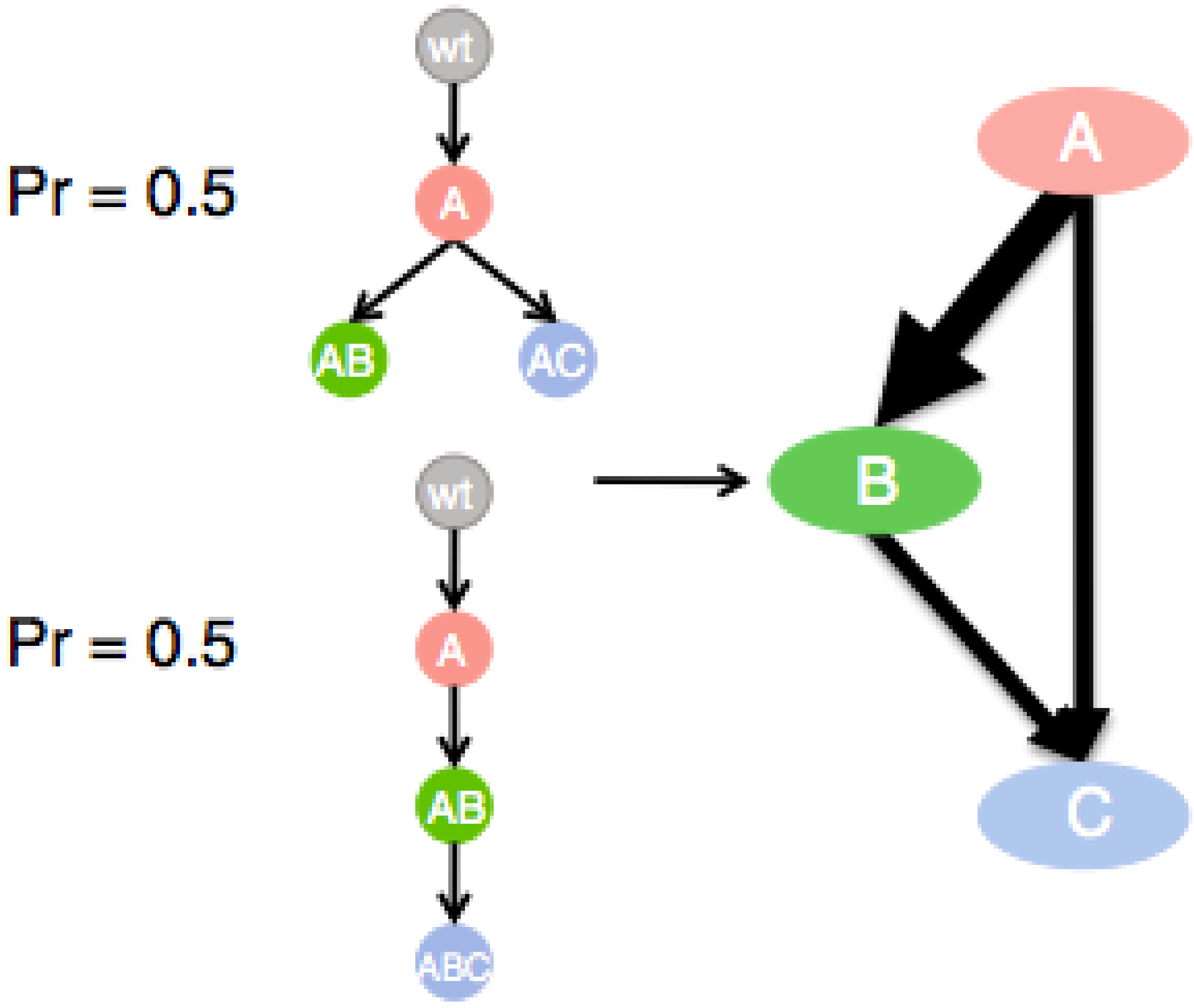
“Ordering Rule”

Given the following frequencies: f_A , f_B and f_C

Rule: If $f_B > f_C$, the phylogeny can be parallel OR B can be an ancestor of C

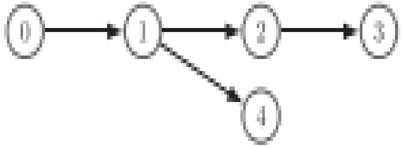
Frequencies of A, B and C give constraint alone


A
SNV population frequencies
Reconstructed phylogeny
Imputed population frequencies
Sample 1
0.2 wt
A
0.4 A
B
0.2 AB
C
0.2 ABC
0.2 wt
A
0.2 A
B
C
0.4 AB
0.2 AC
0.0
0.4
0.8
A B C
B
Sum rule
SNV population frequencies
Reconstructed phylogeny
Imputed population frequencies
Sample 1'
0.2 wt
A
0.2 A
B
0.2 AB
C
0.4 ABC
C
Crossing rule
Sample 2
0.2 wt
A
0.2 A
B
C
0.2 AB
0.4 AC

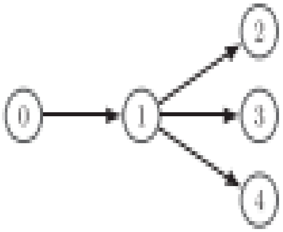

Pr = 0.5
wt
A
AB
AC
Pr = 0.5
wt
A
AB
ABC
A
B
C

Dirichlet Process

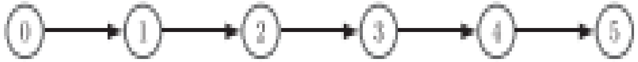
- 3 hyper parameters: α_0 , γ and λ
- α_0 and λ = # of nodes (subclones) in tree
- λ = height of tree
- γ = # of siblings in tree -> width of tree



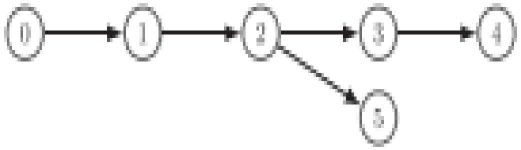
$\alpha_0 = 2, \gamma = 1, \lambda = 0.5$



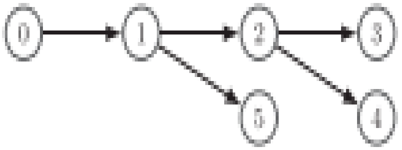
$\alpha_0 = 10, \gamma = 6, \lambda = 0.8$



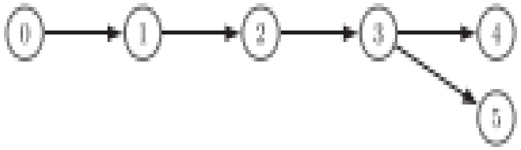
$\alpha_0 = 20, \gamma = 1, \lambda = 0.8$



$\alpha_0 = 20, \gamma = 1, \lambda = 0.5$



$\alpha_0 = 50, \gamma = 1, \lambda = 0.25$



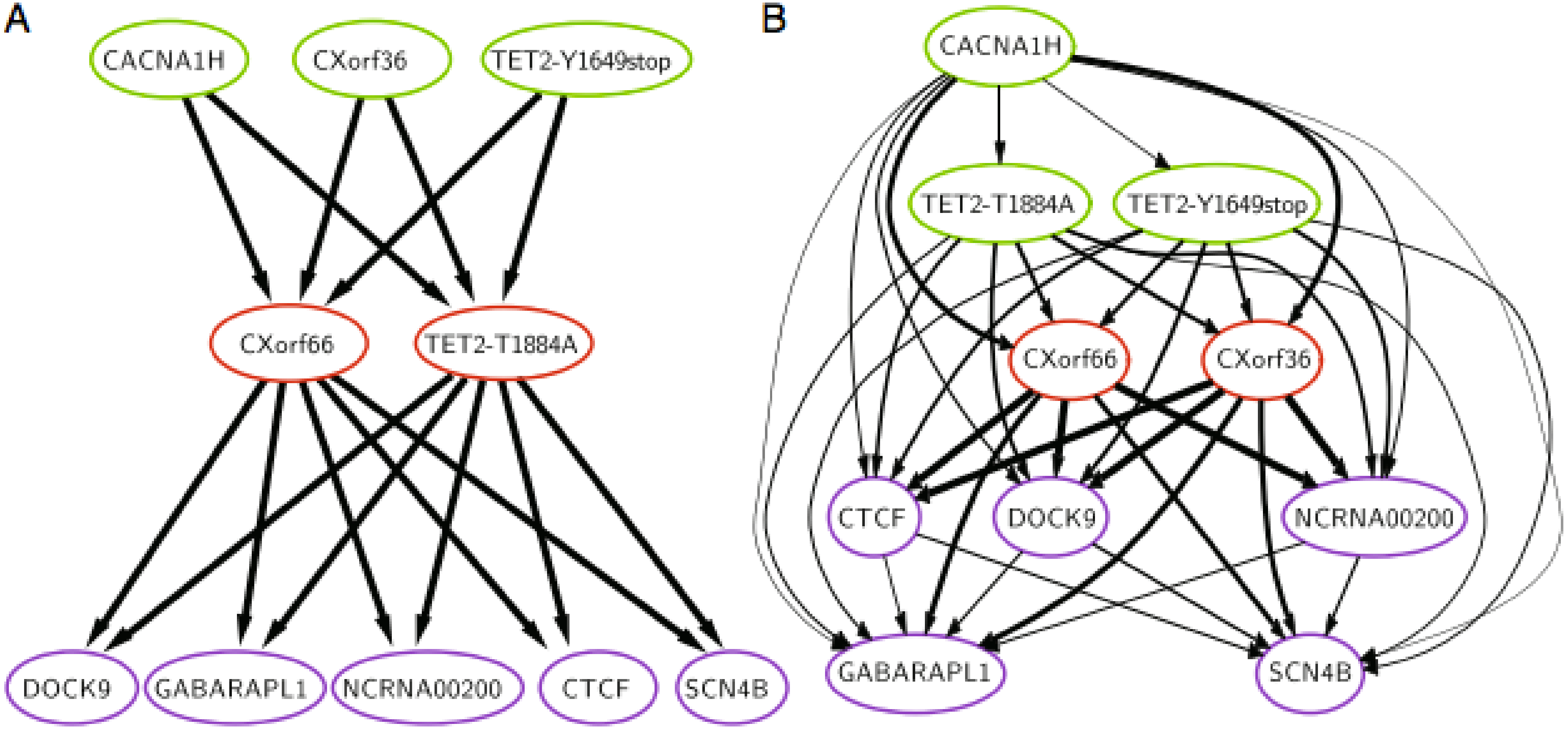
$\alpha_0 = 50, \gamma = 1, \lambda = 0.5$

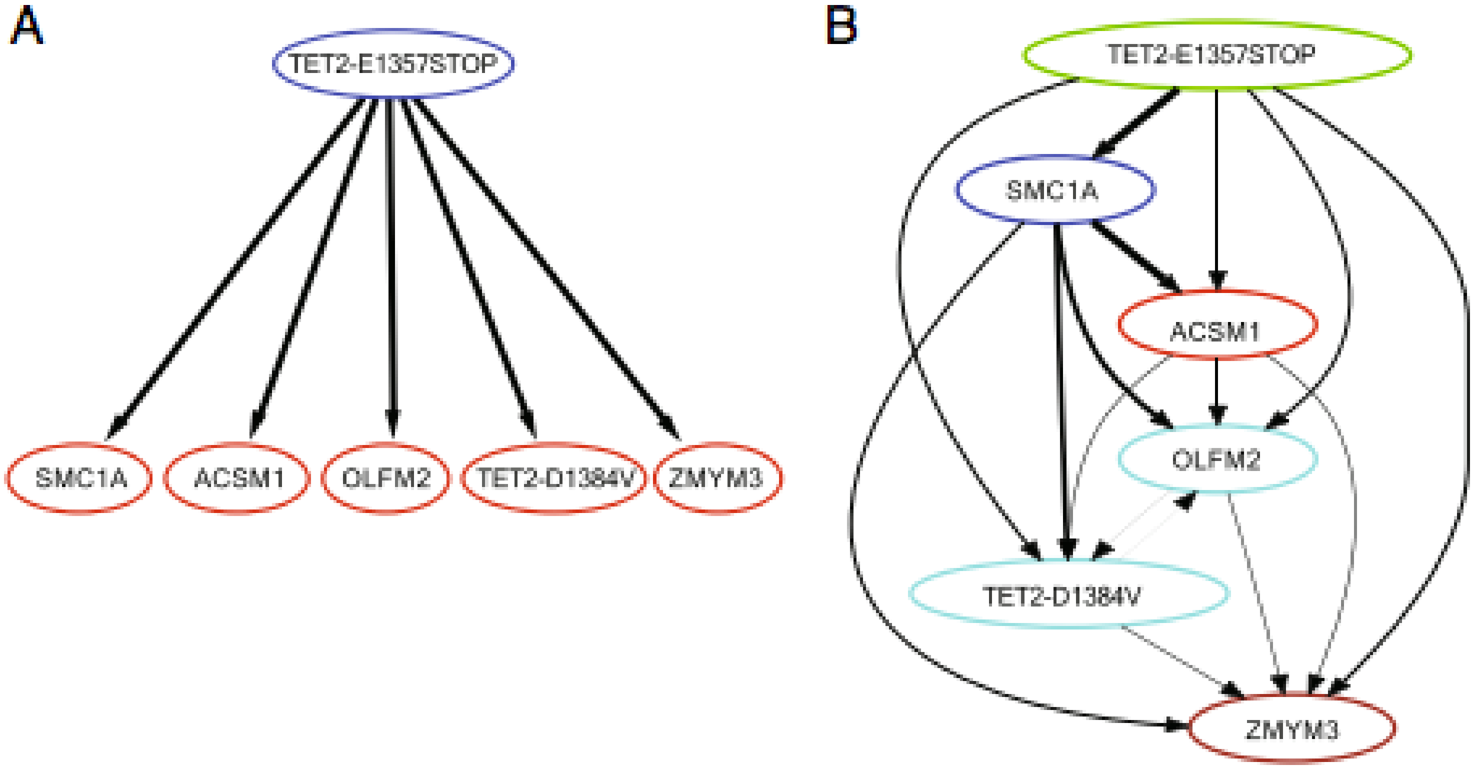
“Partial Order Plot”

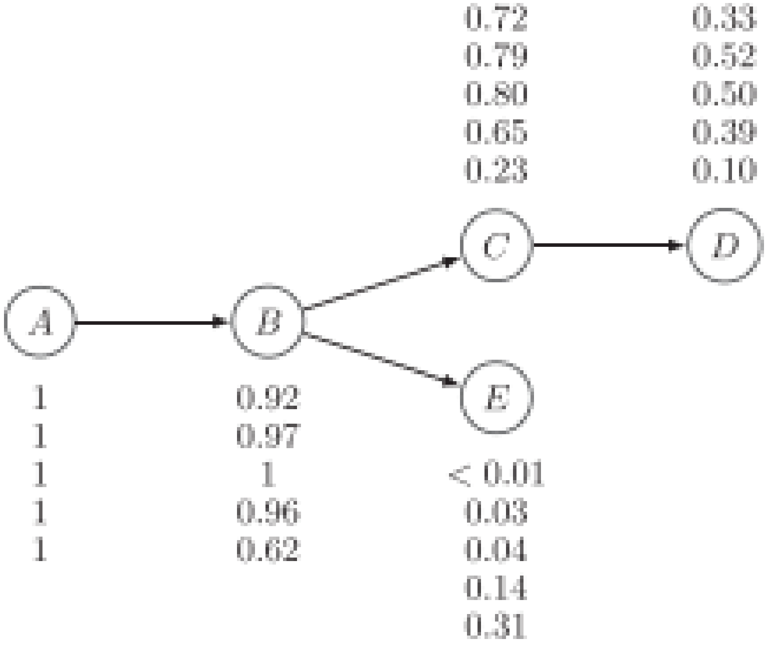
- Data from Jan et. al paper found coexistence of multiple subclonal lineages in HSC from AML samples

- Single assay provided ground truth tree

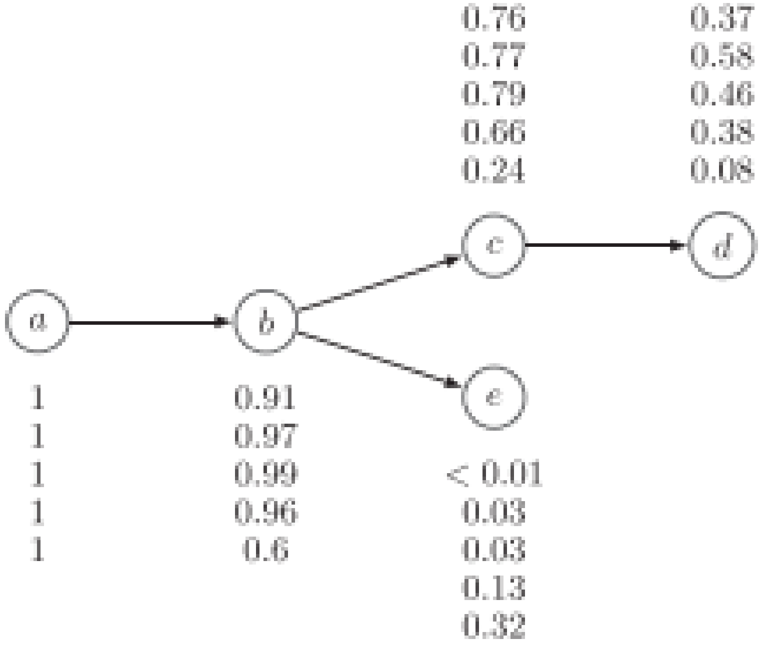
- Samples: SU048 and SU070


A
CACNA1H
CXorf36
TET2-Y1649stop
CXorf66
TET2-T1884A
DOCK9
GABARAPL1
NCRNA00200
CTCF
SCN4B
B
CACNA1H
TET2-T1884A
TET2-Y1649stop
CXorf66
CXorf36
CTCF
DOCK9
NCRNA00200
GABARAPL1
SCN4B

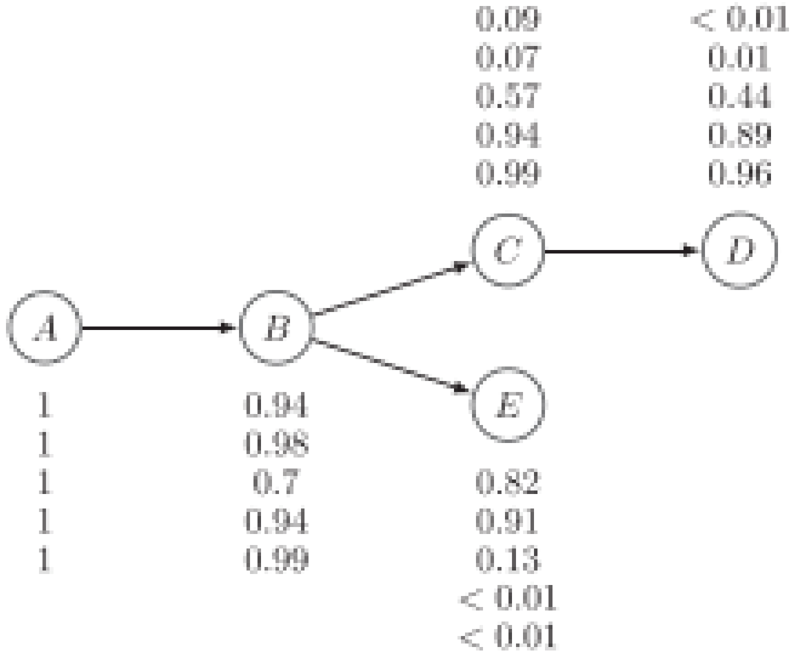

A
TET2-E1357STOP
SMC1A
ACSM1
OLFM2
TET2-D1384V
ZMYM3
B
TET2-E1357STOP
SMC1A
ACSM1
OLFM2
TET2-D1384V
ZMYM3



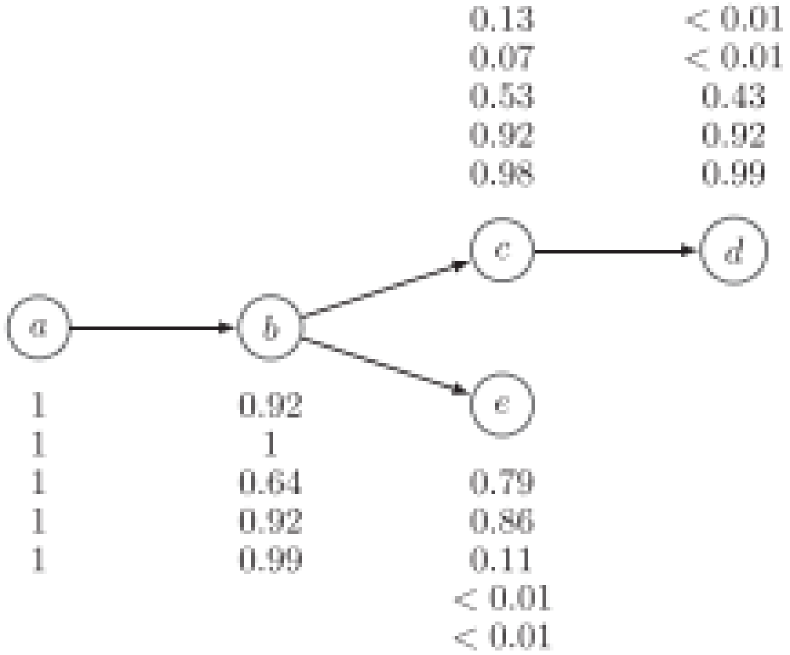
A: –
B: BCL2L13; NAMPTL; GPR158;
SAMHD1;SLC12A1
C: DAZAP1; EXOC6B; GHDC;
OCA2; PLA2G16
D: LRRC16A
E: KLHDC2; COL24A1; NOD1;
HMCN1; MAP2K1



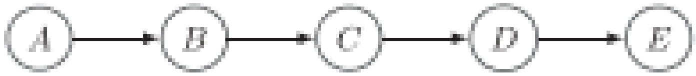
a: –
b: BCL2L13; NAMPTL; GPR158;
SAMHD1; SLC12A1
c: DAZAP1; EXOC6B; GHDC;
OCA2; PLA2G16
d: LRRC16A
e: KLHDC2; COL24A1; NOD1;
HMCN1; MAP2K1



A: –
B: ADAD1; CHTF8; HERC2; IL11RA;
SF3B1; SHROOM1
C: ASXL1; MUSK; SEMA3E
D: NPY; NRG3; FAT3; CHRNB2
E: AMTN; APBB2; ATM; BPIL2;
PLEKHG5; MTUS1; SPTAN1

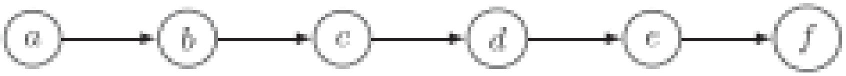


a: –
b: ADAD1; CHTF8; HERC2; IL11RA;
SF3B1; SHROOM1
c: ASXL1; MUSK; SEMA3E
d: NPY; NRG3; FAT3; CHRNB2
e: AMTN; APBB2; ATM; BPIL2;
PLEKHG5; MTUS1; SPTAN1



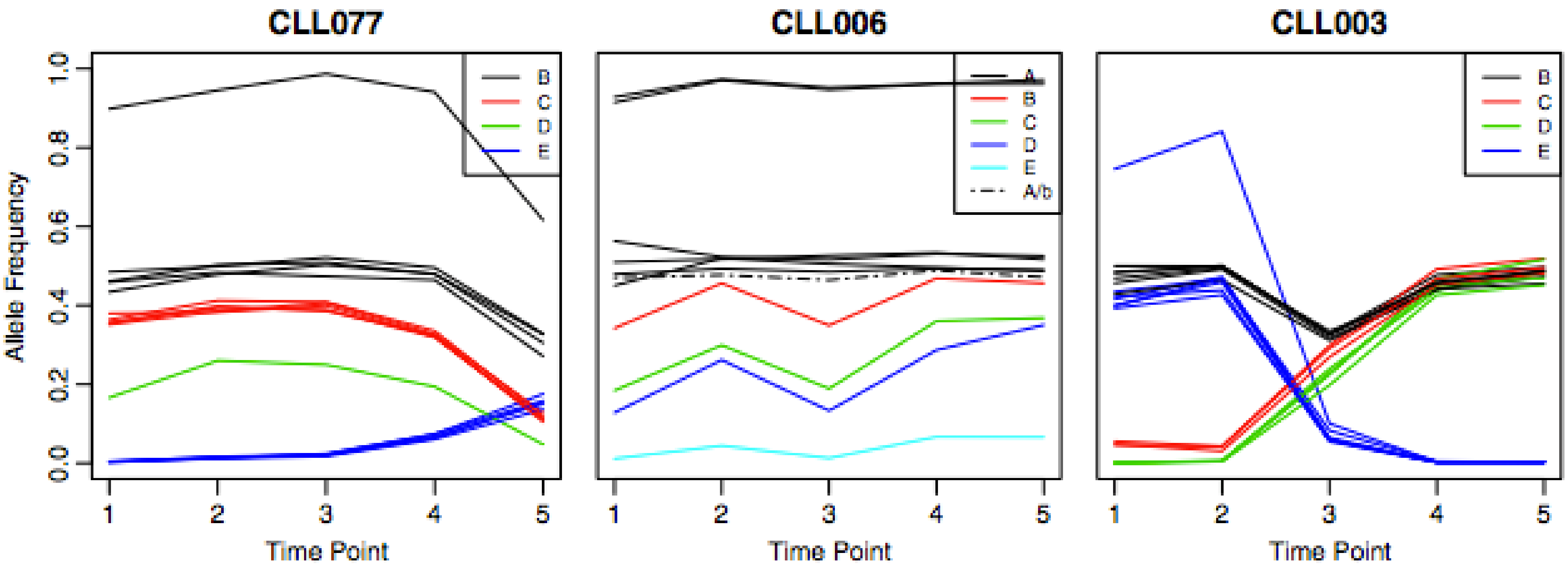
A	B	C	D	E
1	0.68	0.37	0.26	0.02
1	0.91	0.6	0.52	0.09
1	0.7	0.38	0.27	0.03
1	0.94	0.72	0.57	0.13
1	0.91	0.74	0.7	0.13

A: ARHGAP29; EGFR; KIAA0182; KLHL; MED12; PILRB; SIK1
B: U2AF1
C: KIAA0319L
D: IRF4
E: RBPJ



a	b	c	d	e	f
1	0.94	0.71	0.38	0.25	0.02
1	0.98	0.93	0.56	0.53	0.08
1	0.93	0.73	0.37	0.29	0.03
1	0.98	0.94	0.72	0.55	0.14
1	0.95	0.9	0.74	0.69	0.13

a: ARHGAP29; KIAA0182; KLHL; MED12; PILRB; SIK1
b: EGFR
c: U2AF1
d: KIAA0319L
e: IRF4
f: RBPJ


CLL077
CLL006
CLL003
Allele Frequency
0.0
0.2
0.4
0.6
0.8
1.0
Time Point
1
2
3
4
5
B
C
D
E
A
B
C
D
E
A/b
B
C
D
E