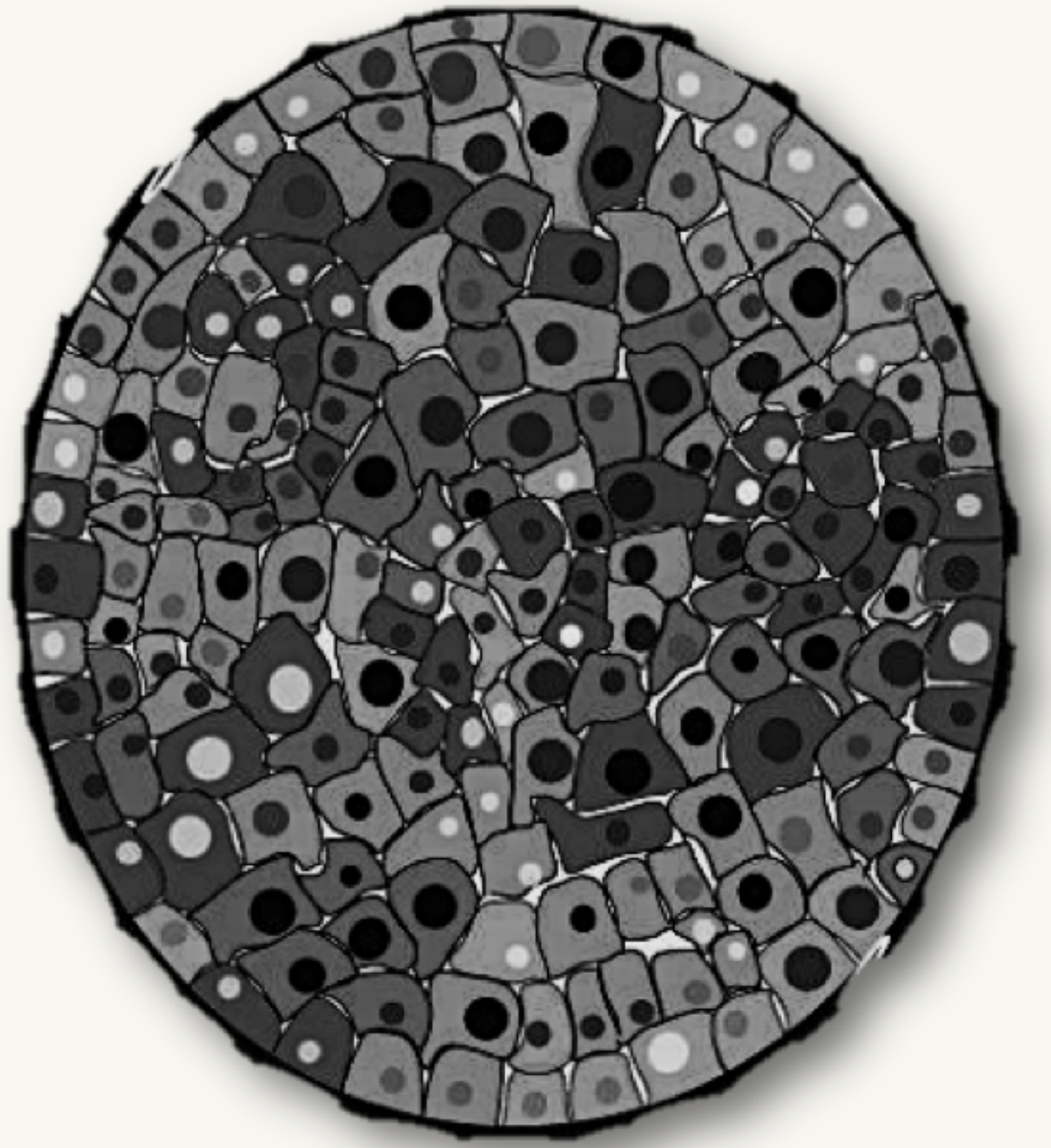
Cancer Phylogeny Inference Using Multi-Sample Somatic Variants

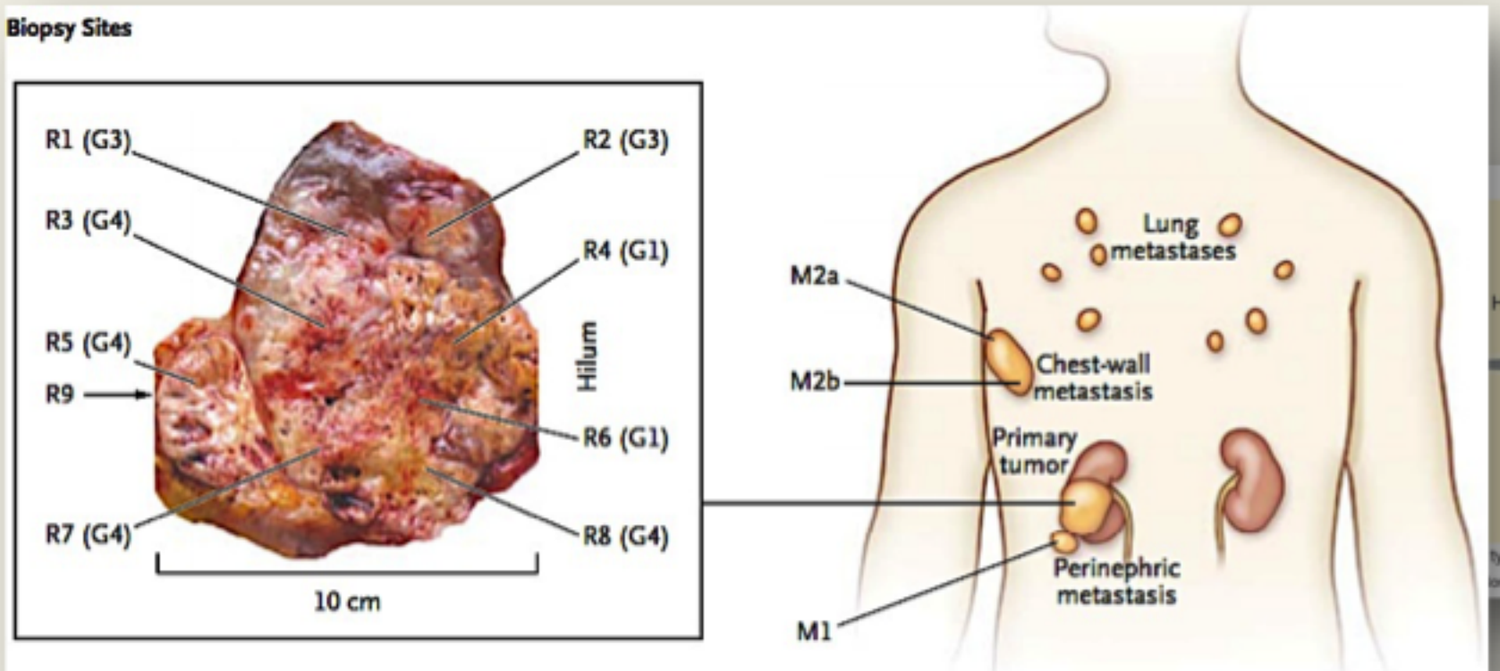
Victoria Popic
PhD Student in *Batzoglou Lab*
Stanford University

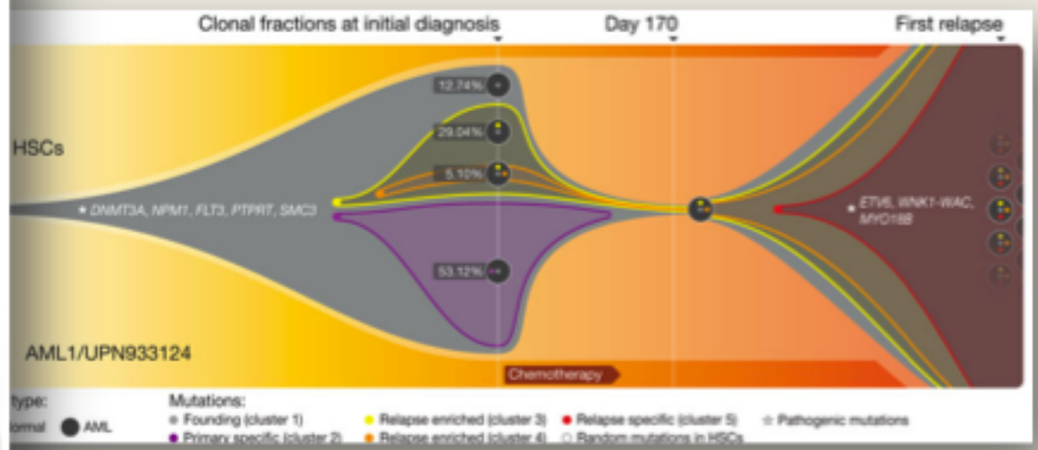
Intra-Tumor Heterogeneity



Intra-Tumor Heterogeneity

Multi-Sample Sequencing Studies





Intratumor Heterogeneity and Branched Evolution Revealed
Multiregion Sequencing

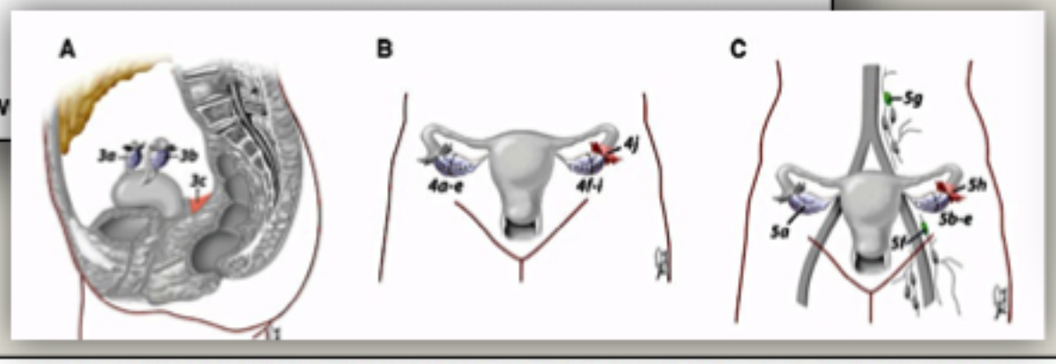
Marco Gerlinger, M.D., Andrew J. Rowan, B.Sc., Stuart Horswell, M.Math., James Larkin, M.D., Ph.D., David

Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing

Li Ding, Timothy J. Ley, David E. Larson, Christopher A. Miller, Daniel C. Koboldt, John S.

Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing

Marco Gerlinger, Stuart Horsw



Genome evolution during progression to breast cancer

Daniel E. Newburger[1,6], Dorna Kashef-Haghighi[2,6], Ziming Weng[3,6], Raheleh Salari[2], Robert T. Sweeney[3], Alayne L. Brunner[3], Shirley X. Zhu[3], Xiangqian Guo[3], Sushama Varma[3], Megan L. Troxell[4], Robert B. West[3,7], Serafim Batzoglou[2,7] and Arend Sidow[3,5,7]

Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling.

Bashashati A[1], Ha G, Tone A, Ding J, Prentice LM, Roth A, Rosner J, Shumansky K, Kalloger S, Senz J, Yang W, McConechy M, Melnyk N, Anglesio M, Luk MT, Tse K, Zeng T, Moore R, Zhao Y, Marra MA, Gilks B, Yip S, Huntsman DG, McAlpine JN, Shah SP.

LICHeE*: Fast and Scalable Inference of Multi-Sample Cancer Lineages

Popic V, Salari R, Hajirasouliha I, Kashef-Haghighi D, West RB, Batzoglou S.
ASHG, 2014
Genome Biology, 2015

Raheleh Salari

Rob West

Arend Sidow

* Initially called "SMuTH" ('13-'14)

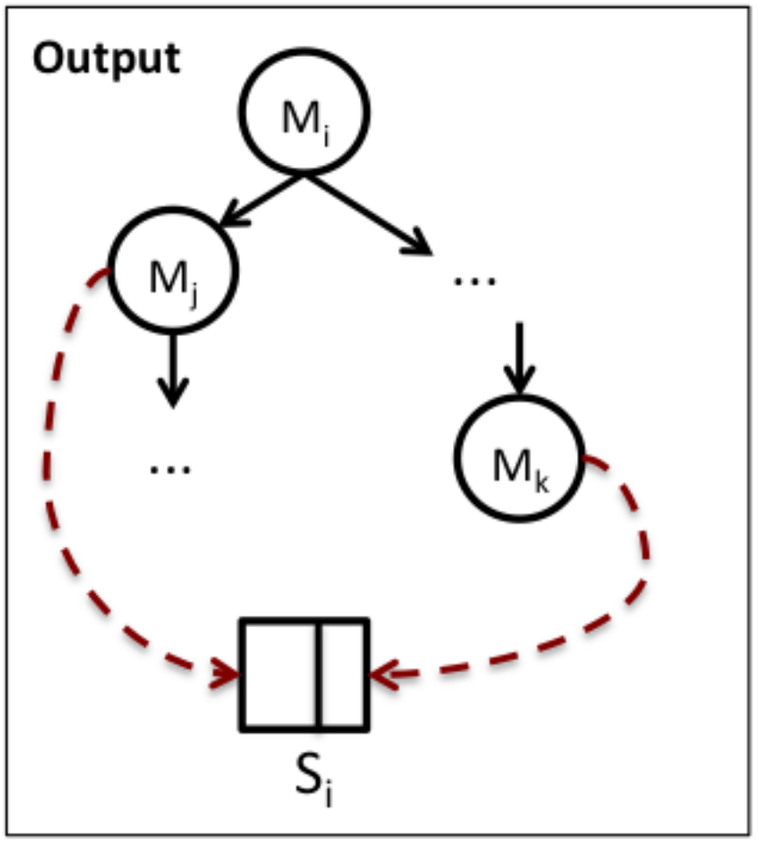
Input Data

Variant allele frequencies (VAFs) per sample

Single Nucleotide Variants (SNVs)

	#chr	position	description	Normal	S_1	S_2	S_3	...	S_M
M_1	1	184306474	A/G HMCN1	0.0	0.1	0.2	0.25		0.15
M_2	1	18534005	C/A IGSF21	0.0	0.1	0.25	0.2		0.1
M_3	1	110456920	G/A UBL4B	0.0	0.4	0.4	0.45		0.45
...									
M_N	10	26503064	C/G MYO3A	0.0	0.4	0.0	0.0		0.24

Note: In general, the method can handle any type of variant given its cell prevalence (CP) values in each sample



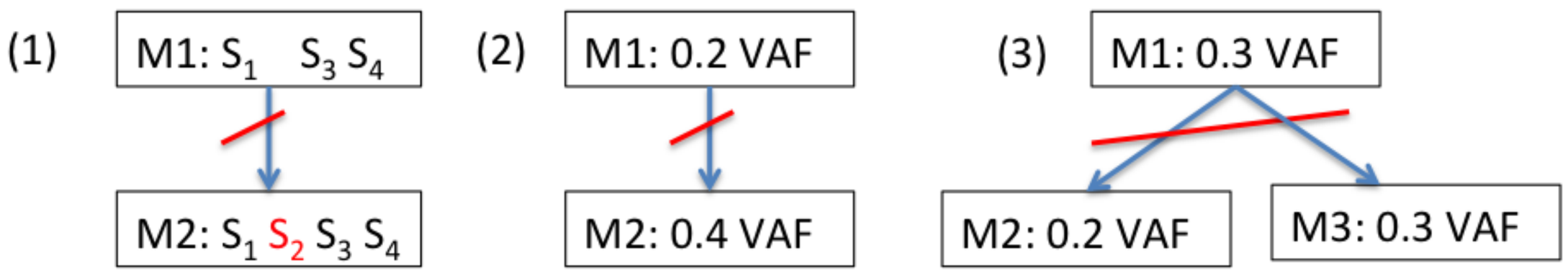
Perfect Phylogeny Model Assumption

Mutations do not recur independently in different cells
⇒ cells sharing the same mutation must have inherited it from a common ancestral cell

Perfect Phylogeny Model: Constraints

Three SNV Ordering Constraints:

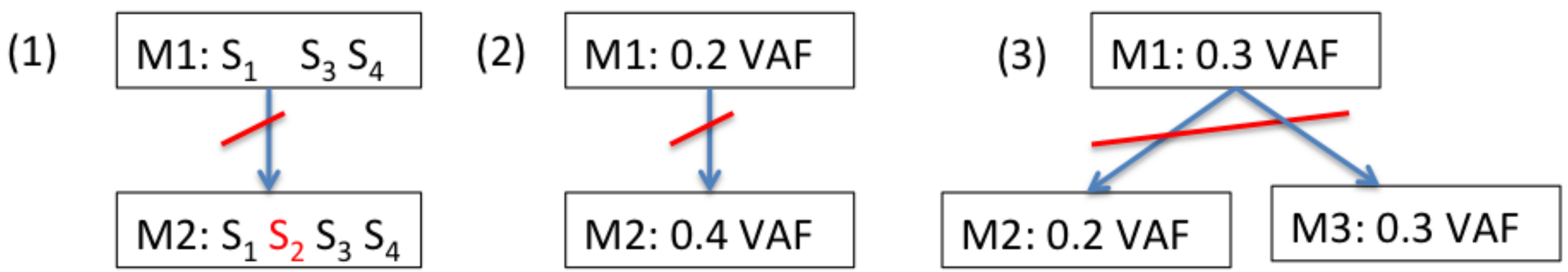
1. a mutation present in a given set of samples cannot be a successor of a mutation present in a smaller subset of these samples
2. a mutation cannot have a VAF higher than that of its predecessor mutation (except due to CNVs)
3. the sum of the VAFs of mutations disjointly present in distinct subclones cannot exceed the VAF of a common predecessor mutation present in these subclones



Perfect Phylogeny Model: Constraints

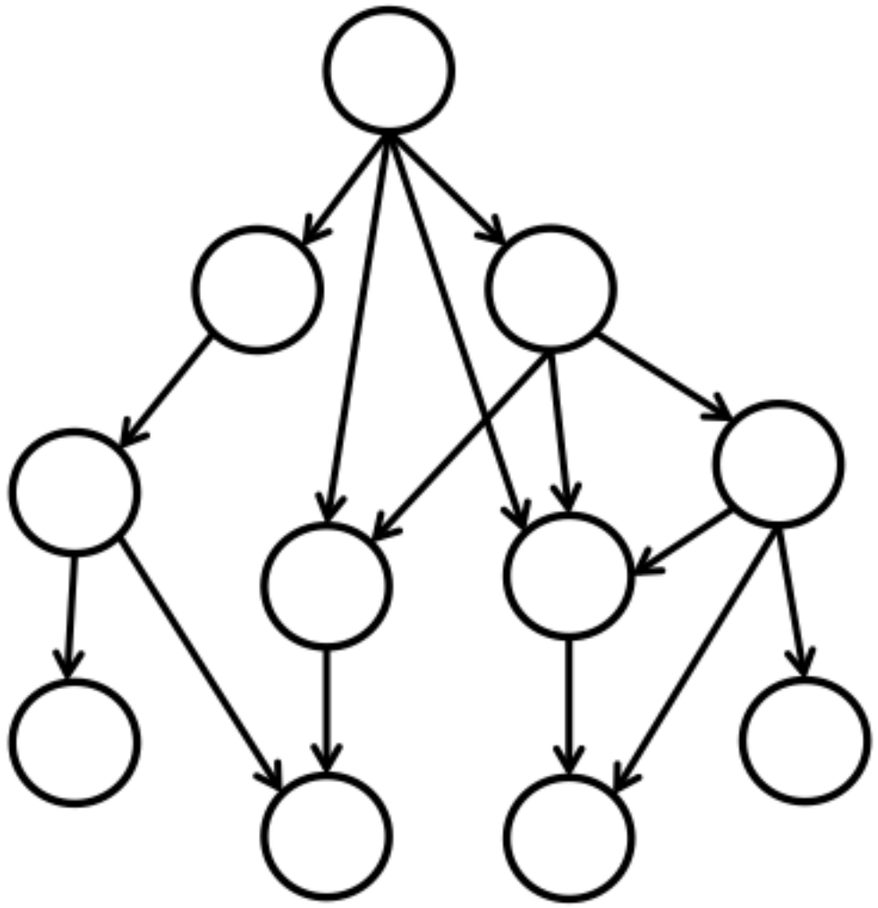
Three SNV Ordering Constraints:

1. a mutation present in a given set of samples cannot be a successor of a mutation present in a smaller subset of these samples
2. a mutation cannot have a VAF higher than that of its predecessor mutation (except due to CNVs)
3. the sum of the VAFs of mutations disjointly present in distinct subclones cannot exceed the VAF of a common predecessor mutation present in these subclones



Goal: find all lineage trees that satisfy the above three constraints

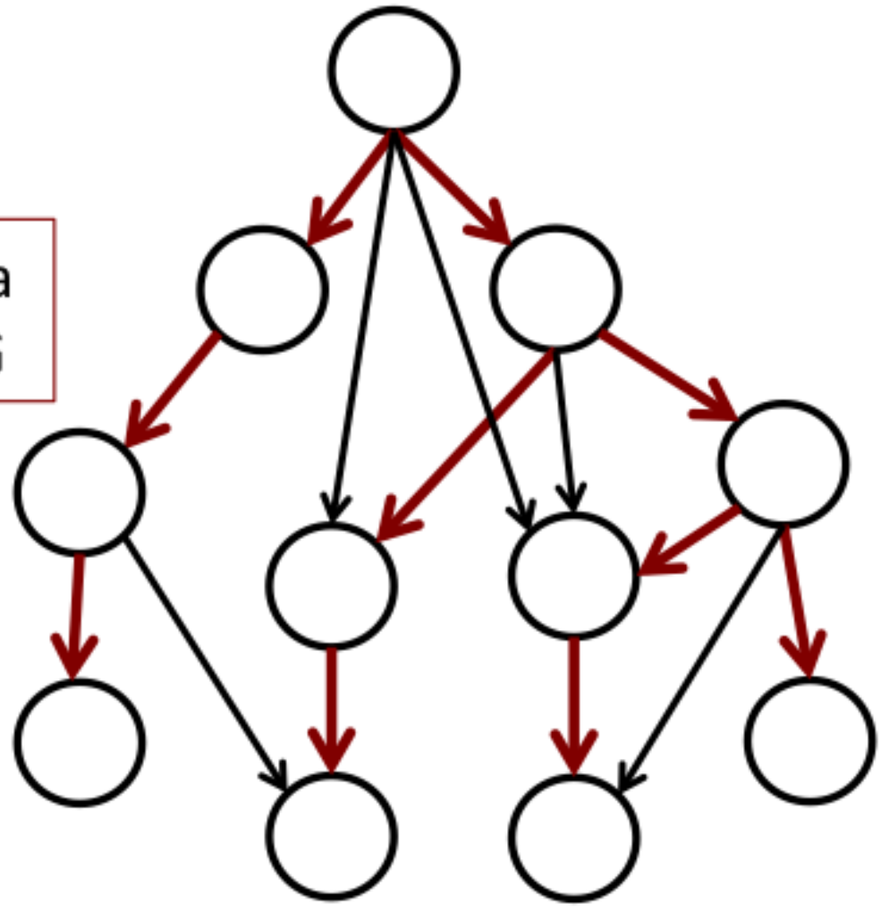
LICHeE's Problem Formulation



DAG encoding all pairwise valid precedence relationships – *evolutionary constraint network*

LICHeE's Problem Formulation

True lineage tree will be a spanning tree of this DAG



DAG encoding all pairwise valid precedence relationships – *evolutionary constraint network*

➔ search for all lineage trees that satisfy constraint (3)

LICHeE: Method Overview

Given: SSNV multi-sample VAFs

Algorithm steps:

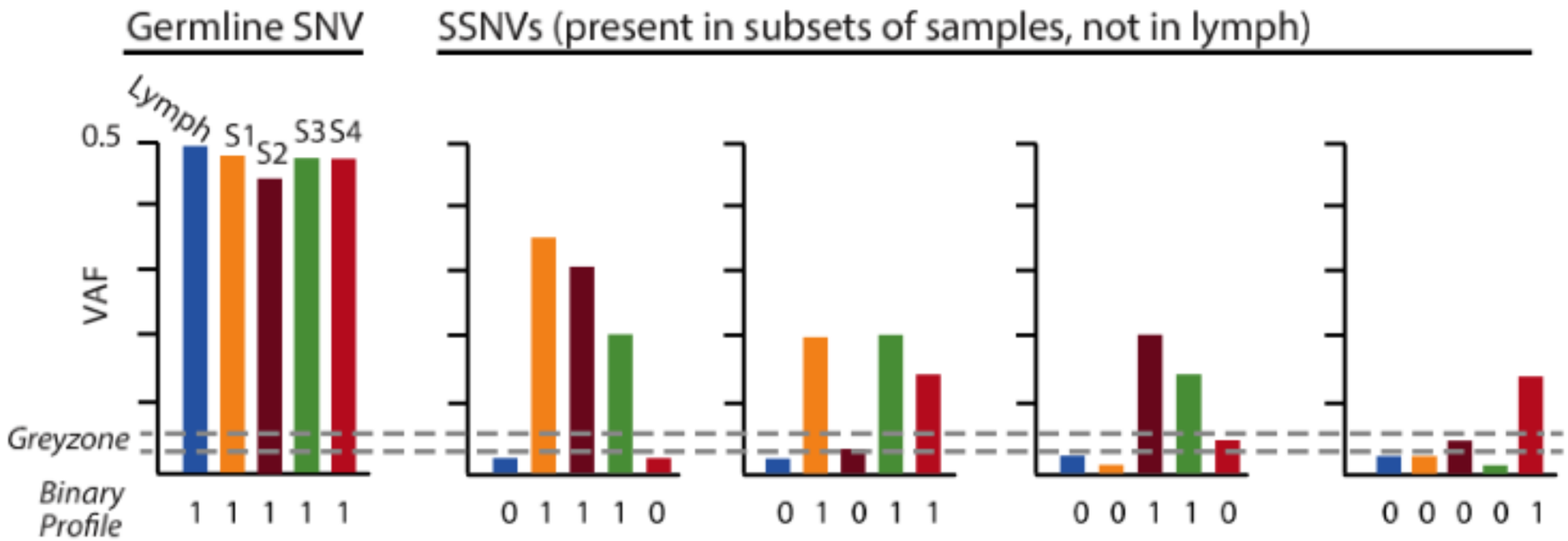
1. Grouping and clustering SSNVs
2. Evolutionary Constraint Network Construction
3. Lineage Tree Search and Ranking

1. Grouping and clustering SSNVs

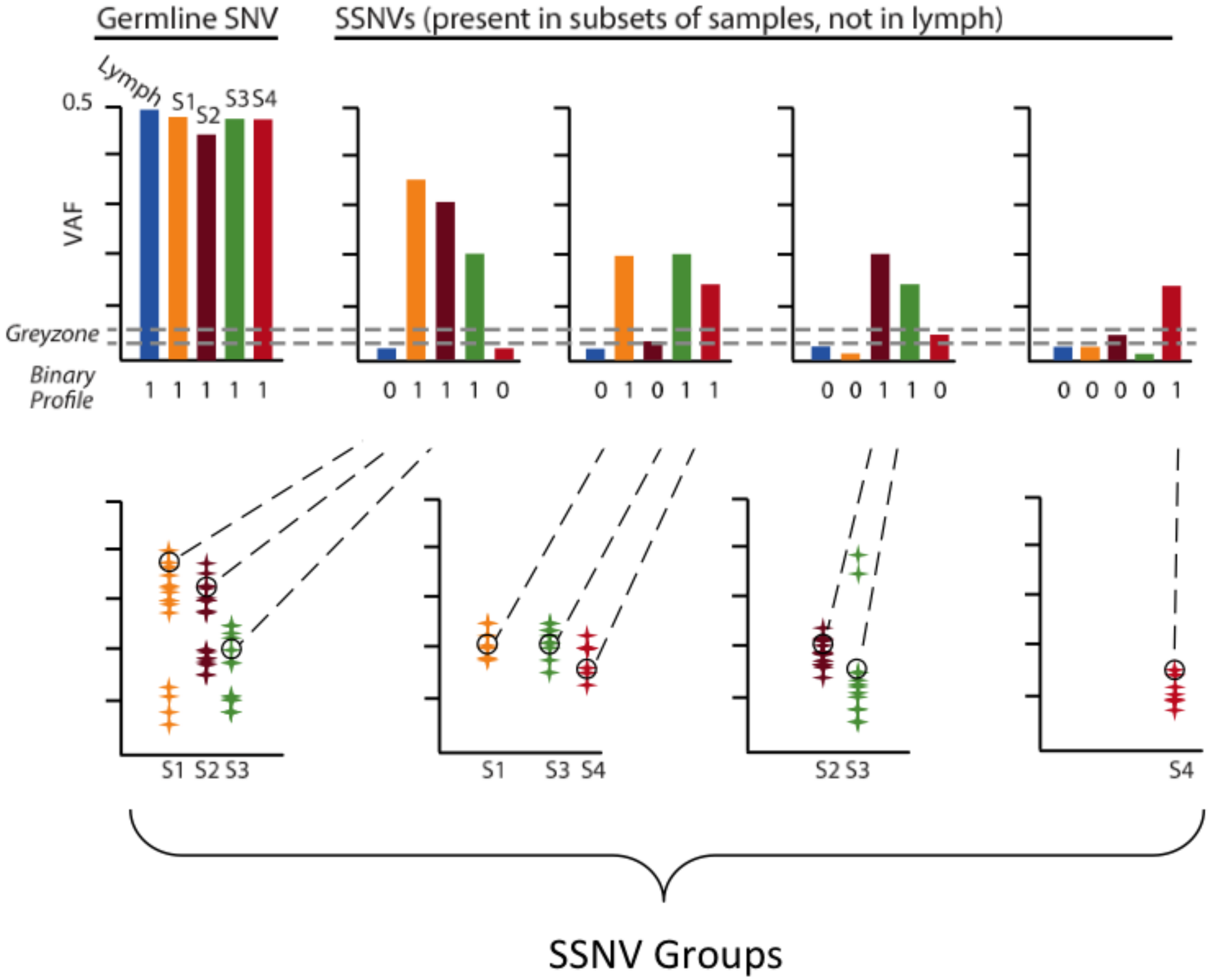
→ presence patterns across samples

→ VAF similarity

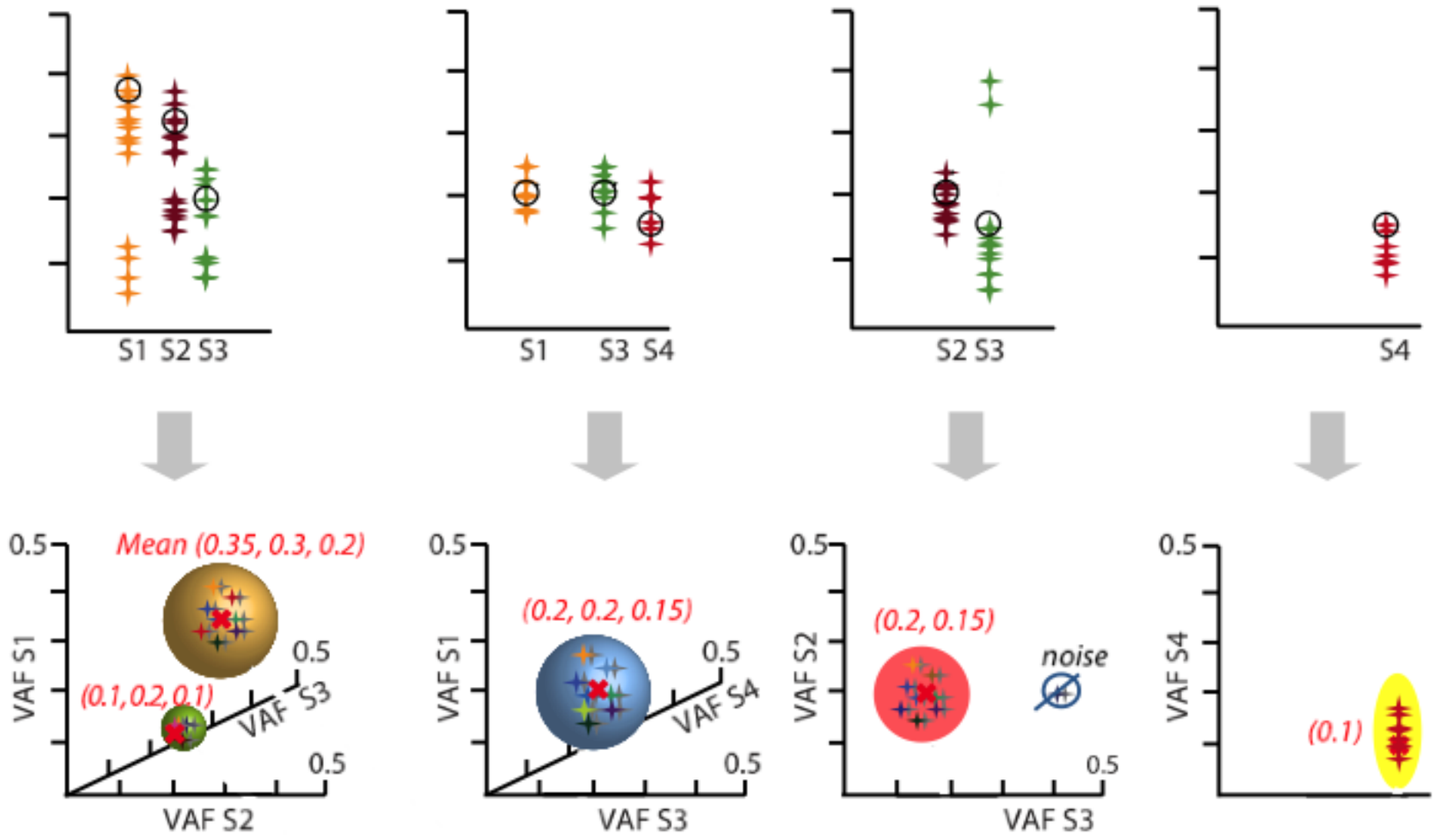
Presence Patterns Across Samples



Presence Patterns Across Samples



VAF-Based Clustering

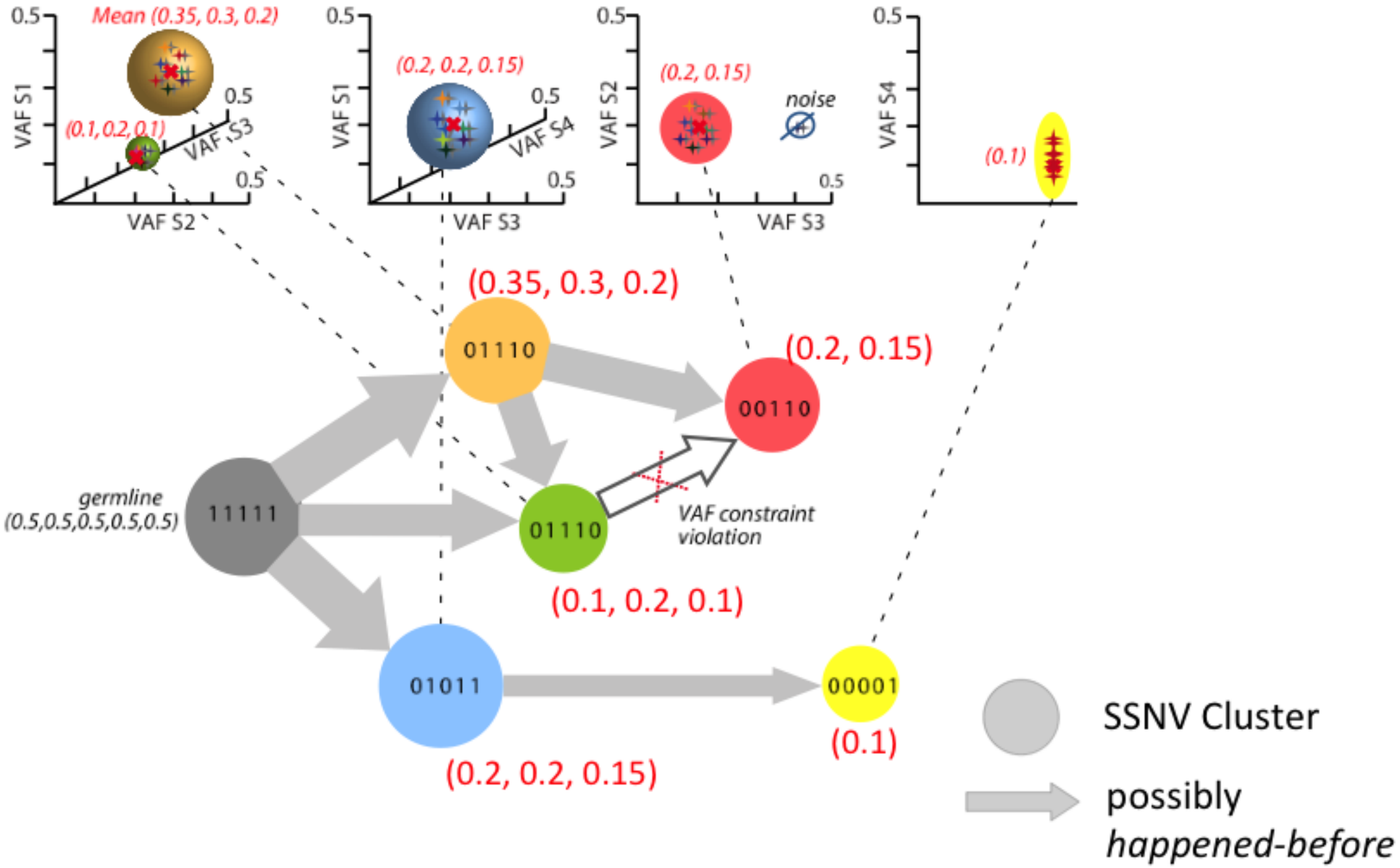


2. Evolutionary Constraint Network Construction

→ encodes whether a given cluster of SSNVs could have preceded another

→ valid lineage trees are embedded in this network

Evolutionary Constraint Network



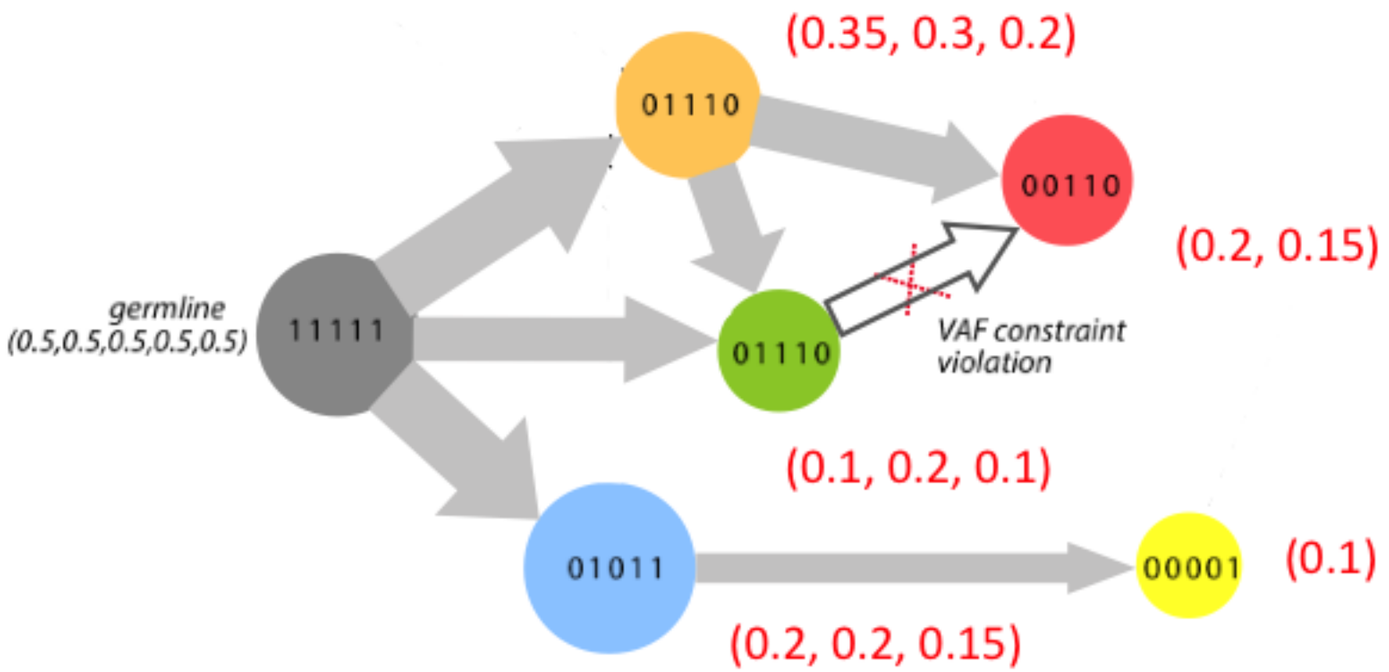
3. Lineage Tree Search and Ranking

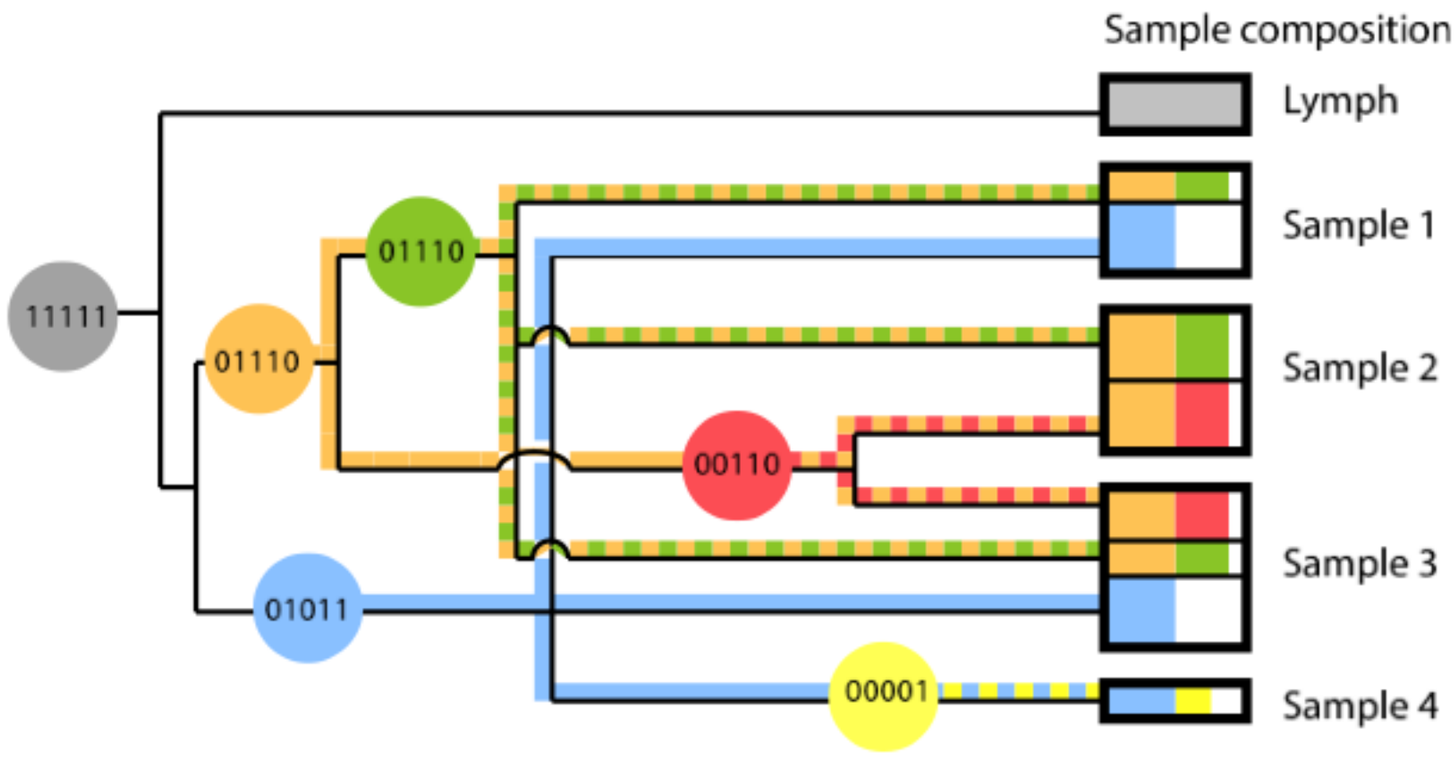
→ search for spanning trees satisfying VAF constraints

within an error margin (extension of Gabow and Myers'78)

→ top tree minimizes the squared deviation from the cluster centroids

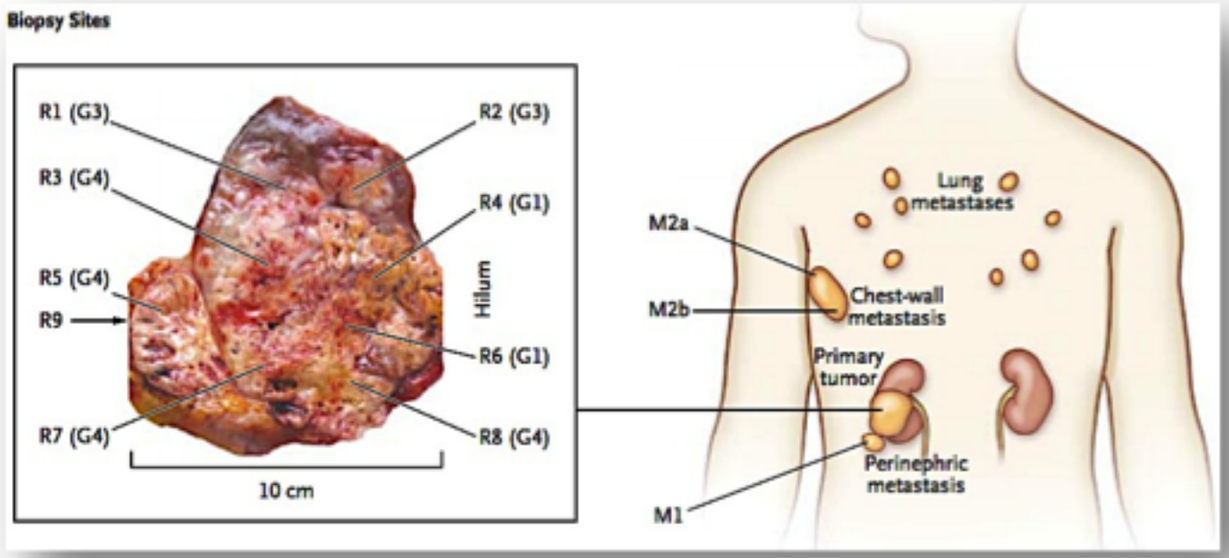
Lineage Tree Search



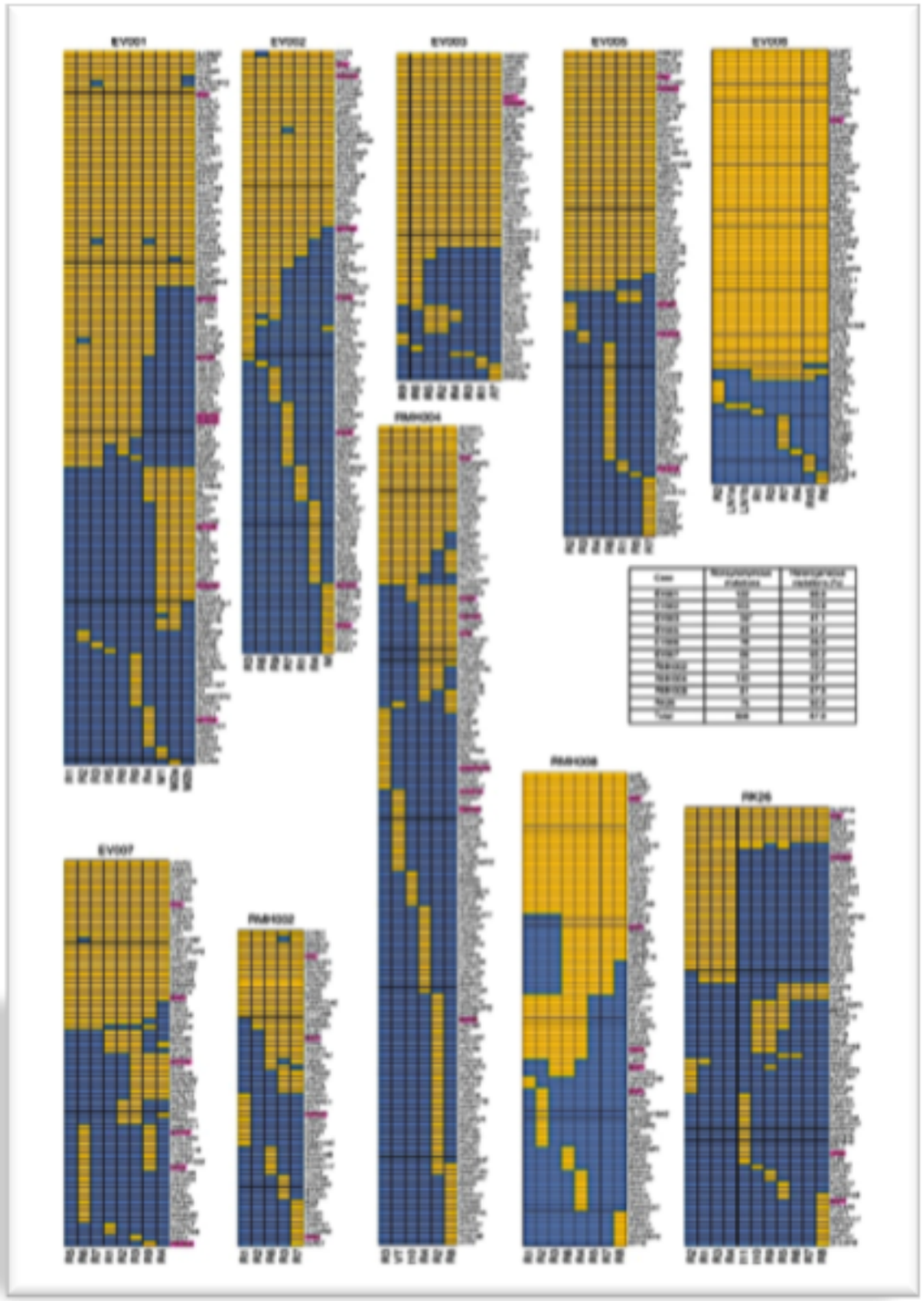


RESULTS

ccRCC Study by Gerlinger *et. al* (2014)

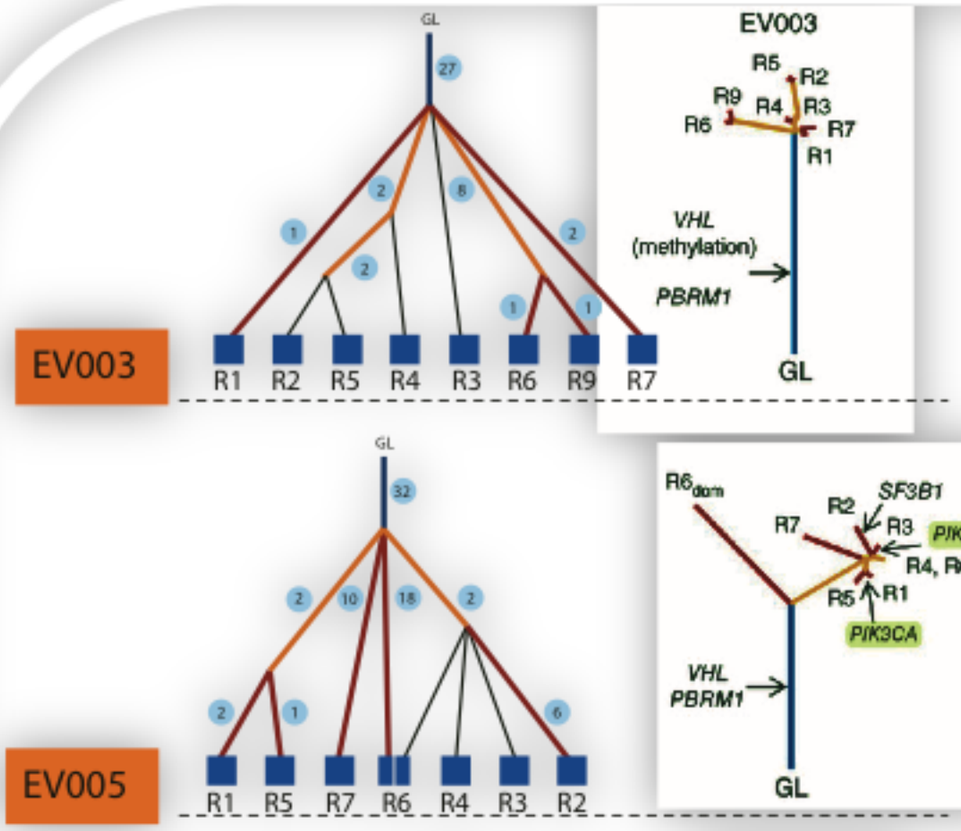


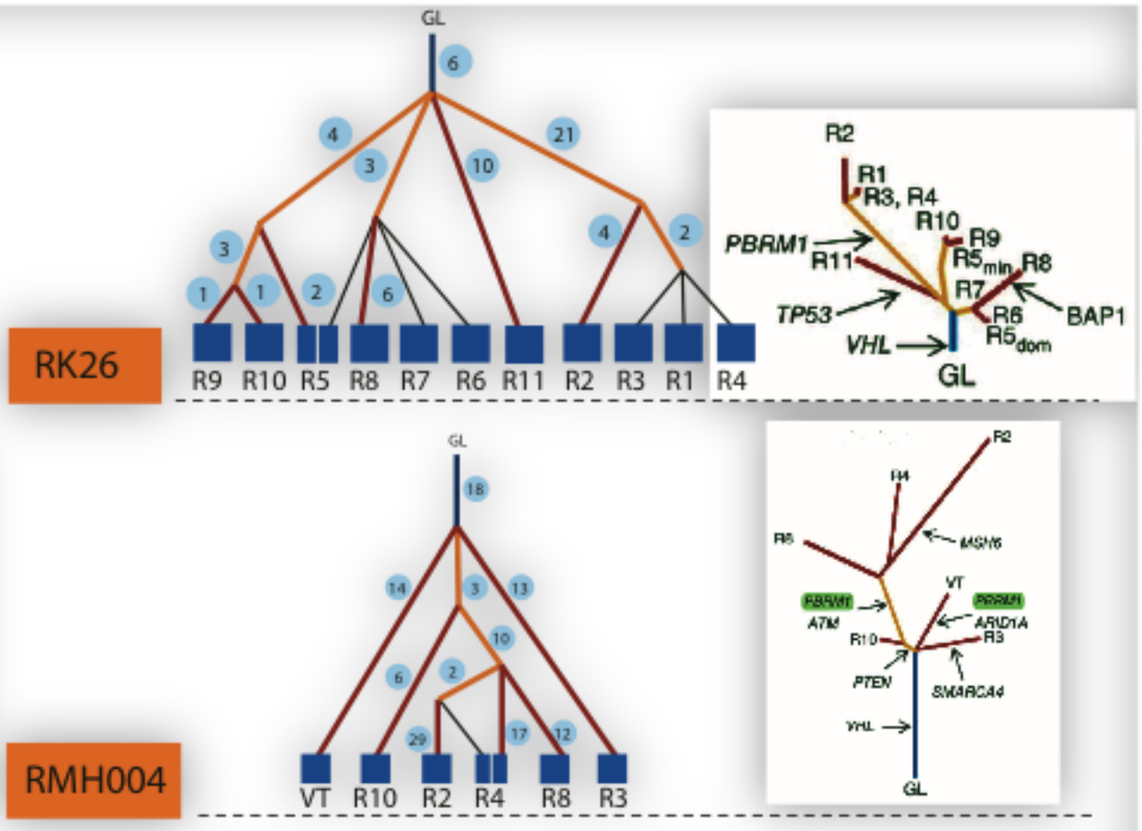
8 patients, 587 SNVs

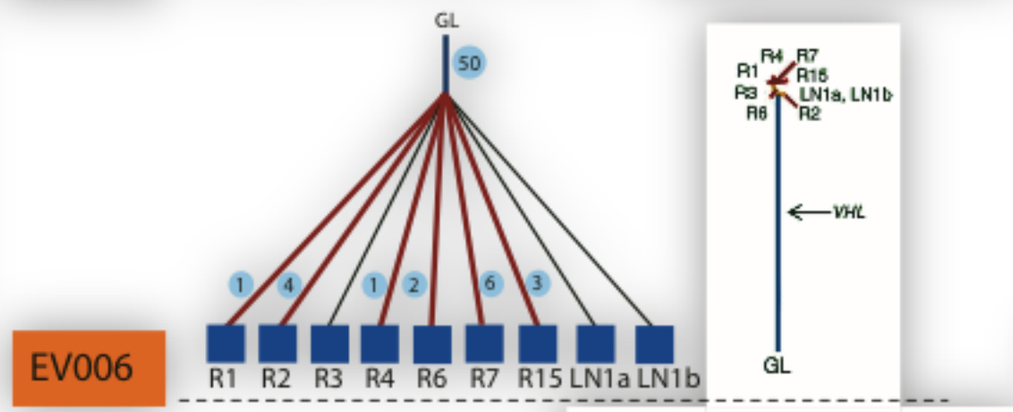


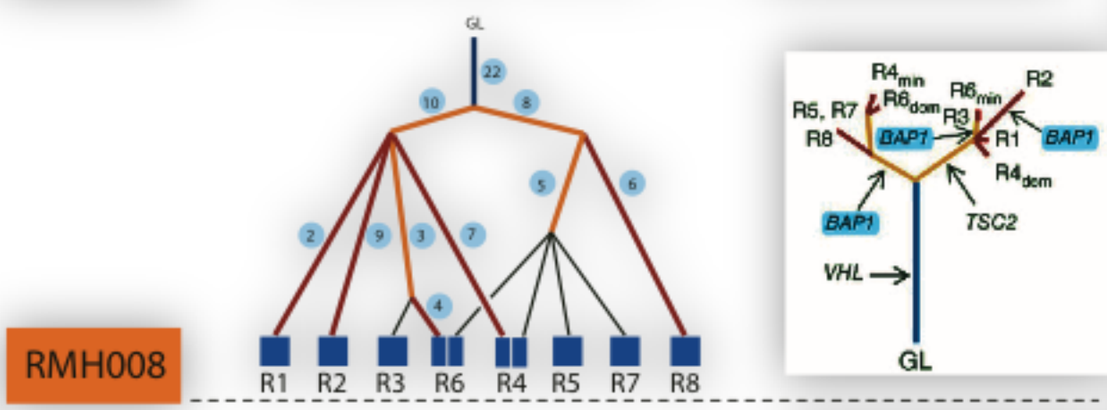
Gerlinger, M., et al. (2014). "Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing." Nature genetics **46**(3): 225-233.

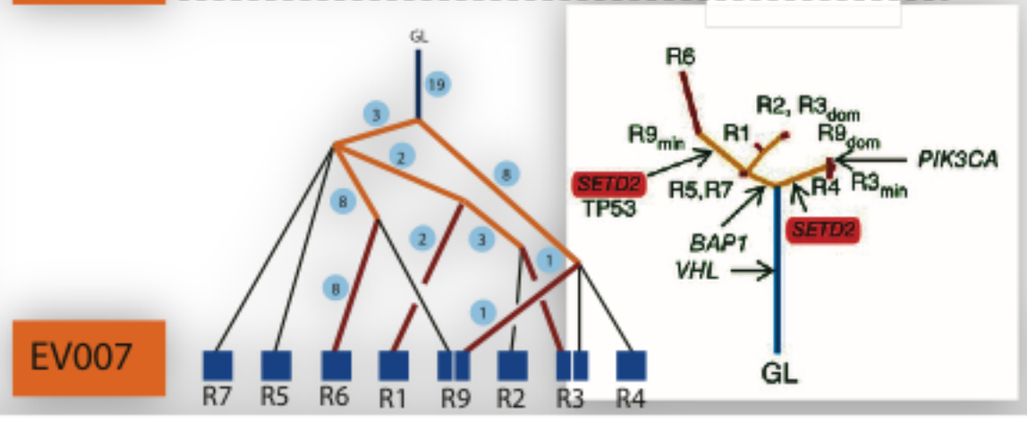
ccRCC Study by Gerlinger *et. al* (2014)

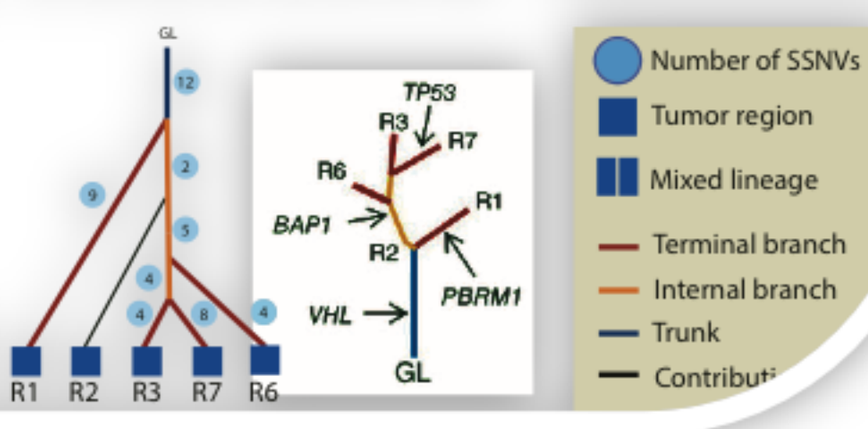




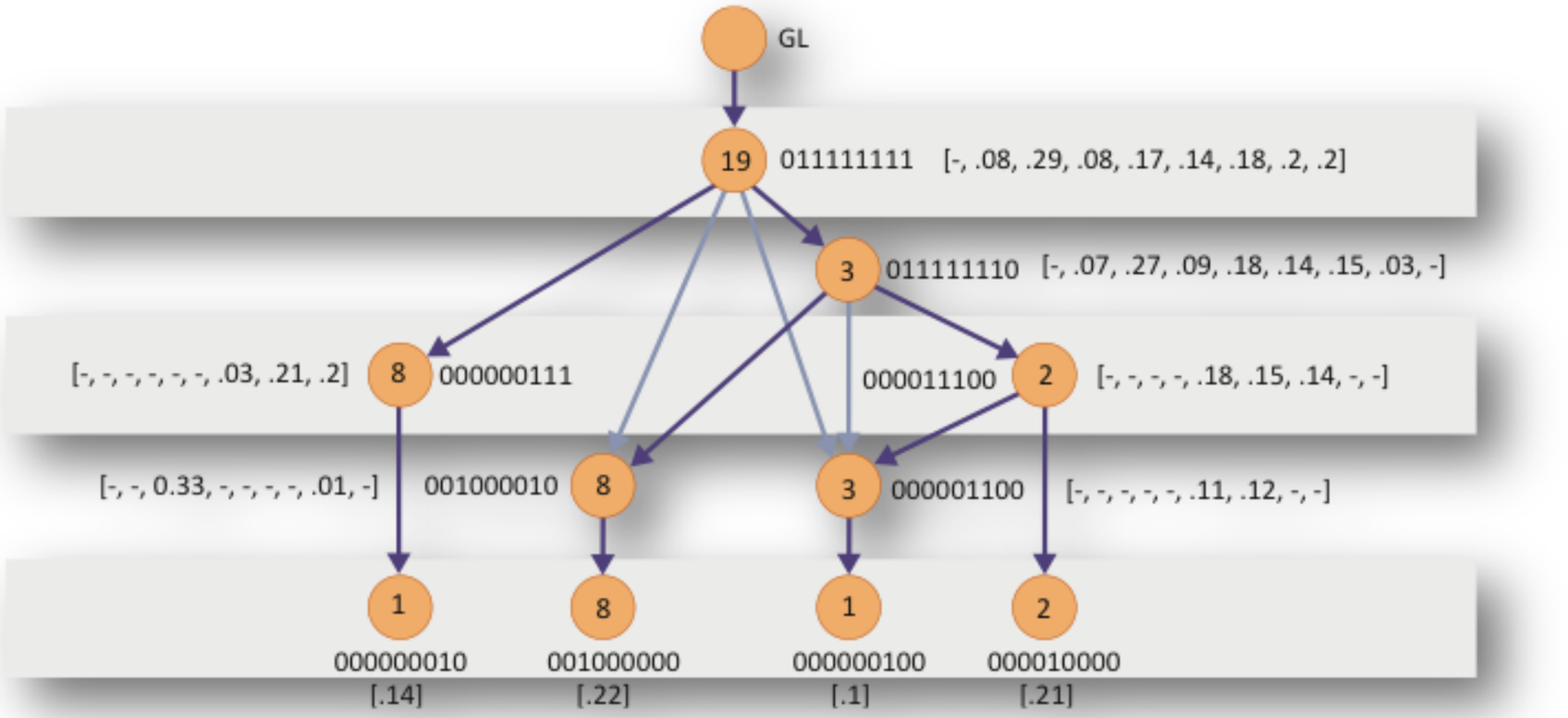


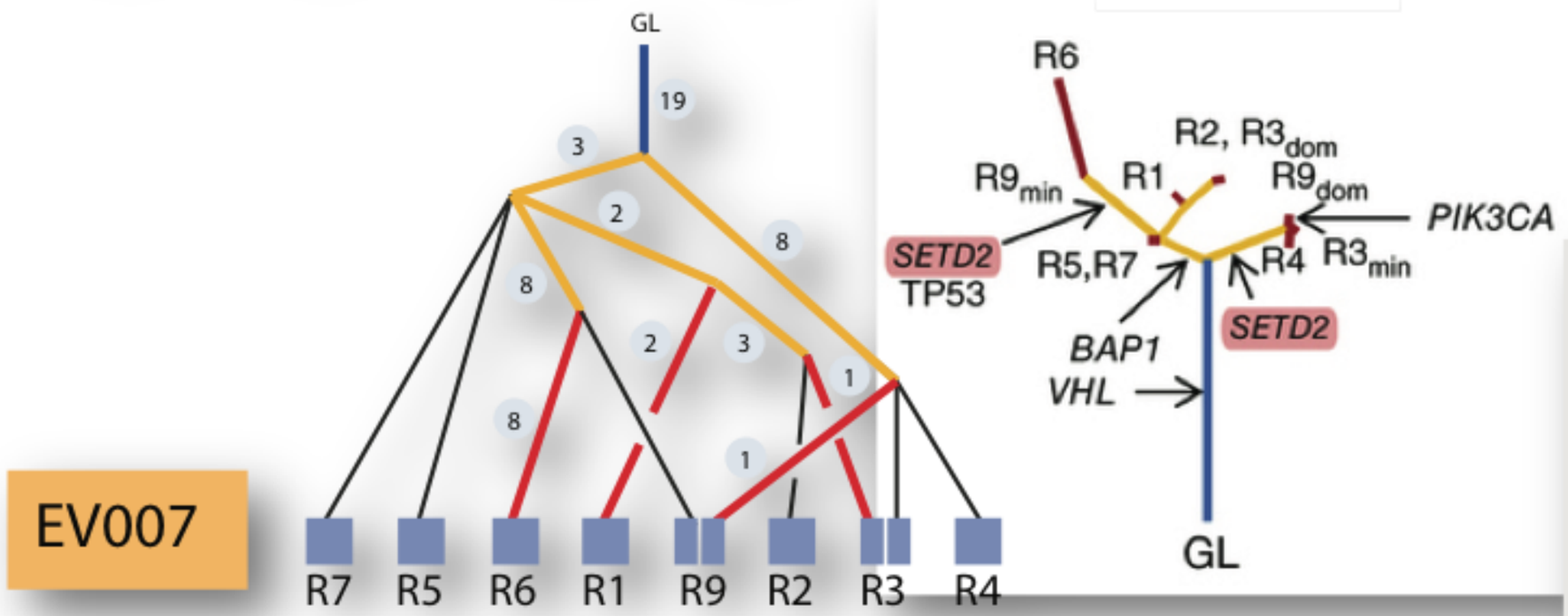






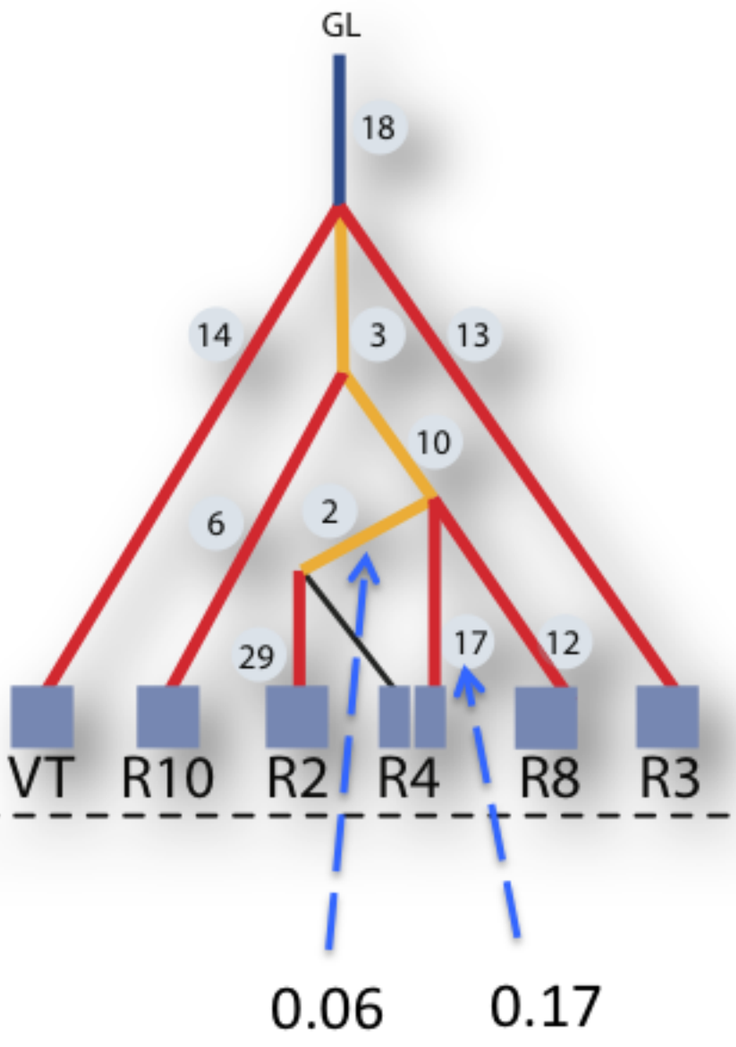
ccRCC Study by Gerlinger *et. al* (2014)

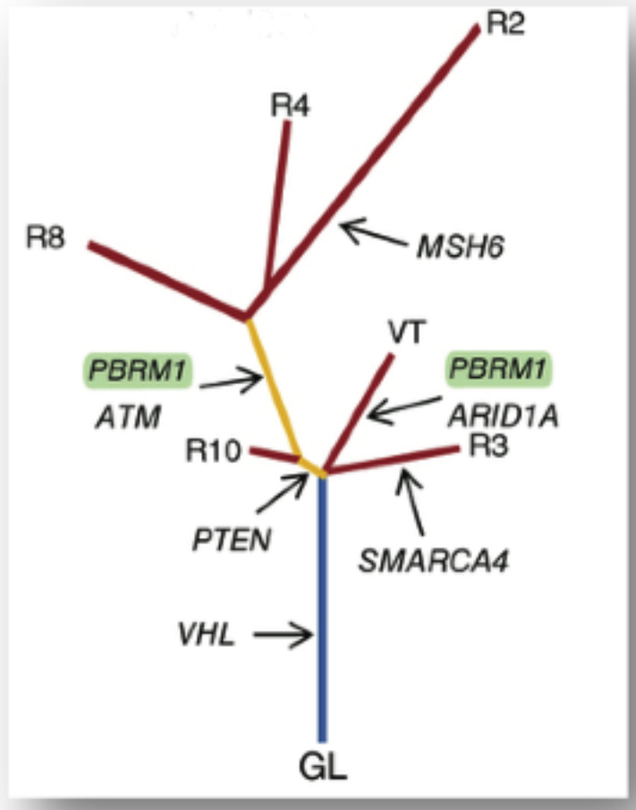




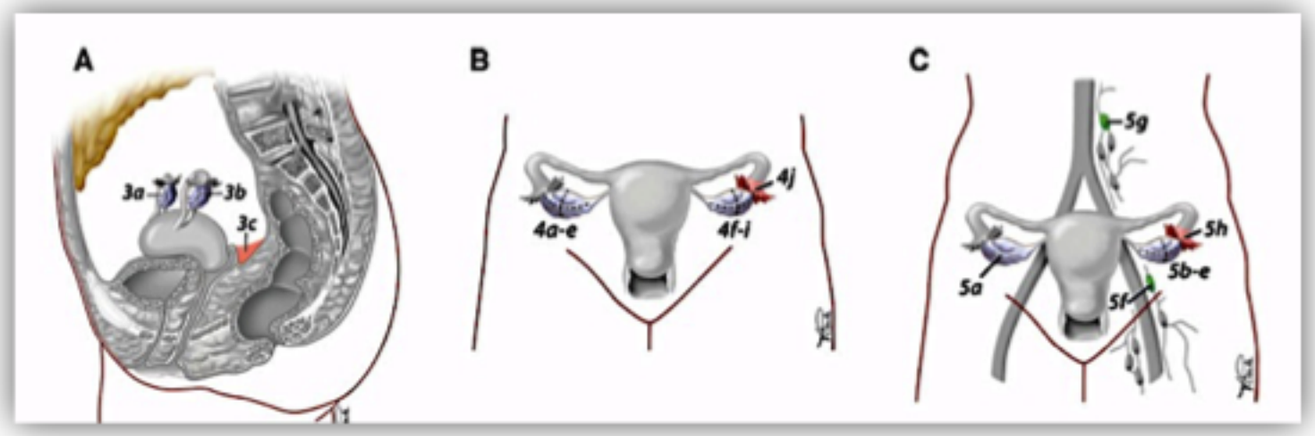
ccRCC Study by Gerlinger *et. al* (2014)

RMH004

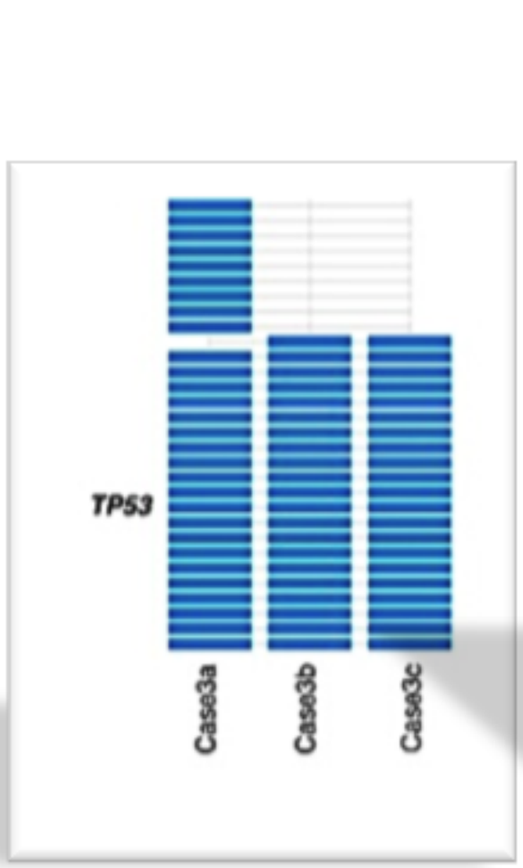


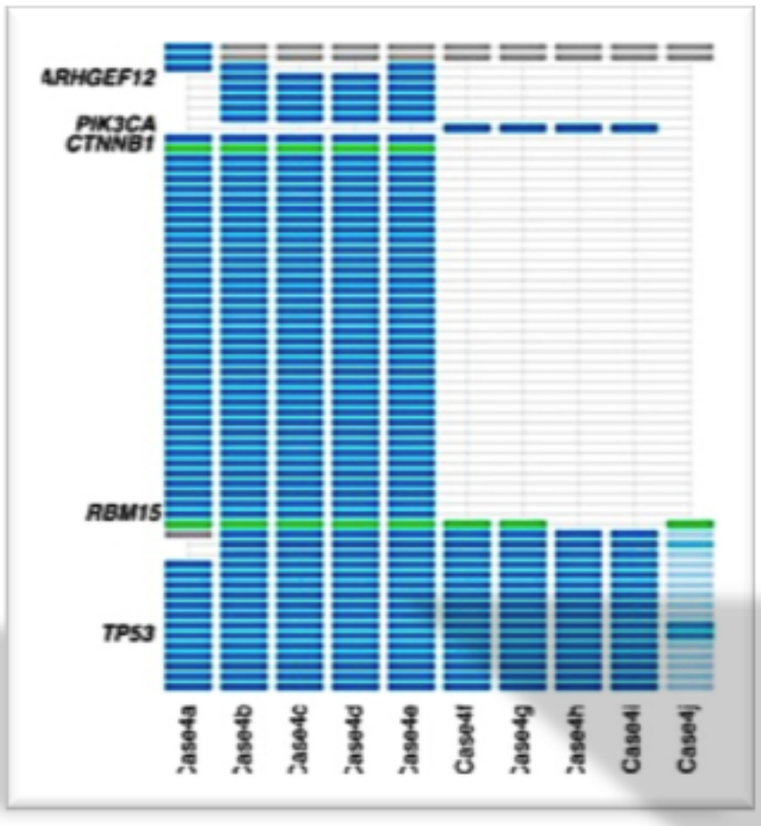


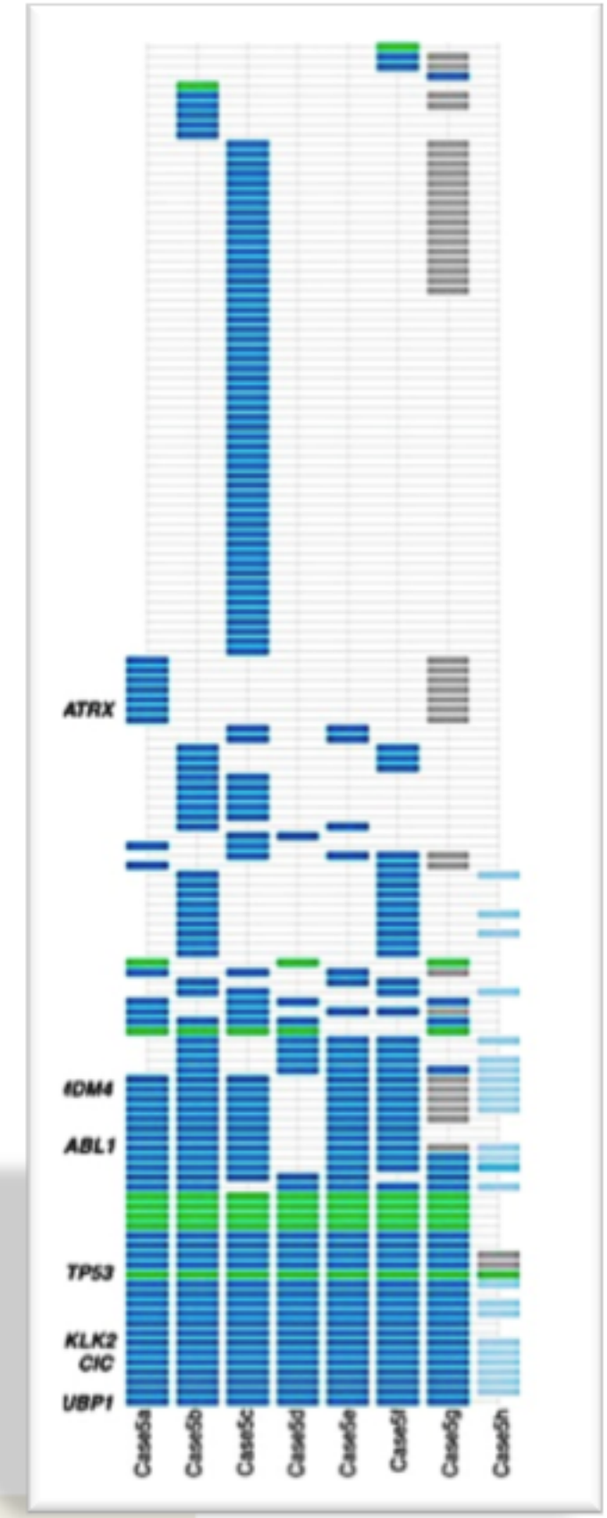
HGSC Study by Bashashati *et. al* (2013)



19 tumors, 6 patients, 340 SNVs

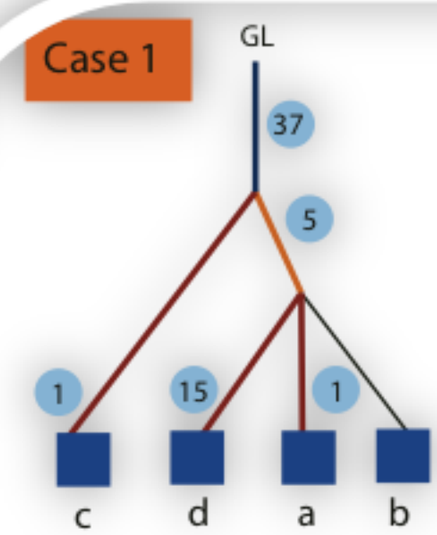


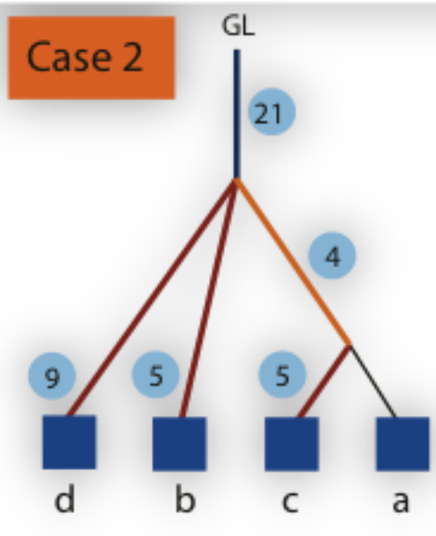


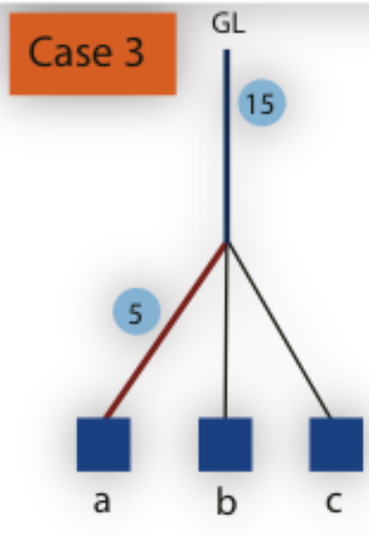


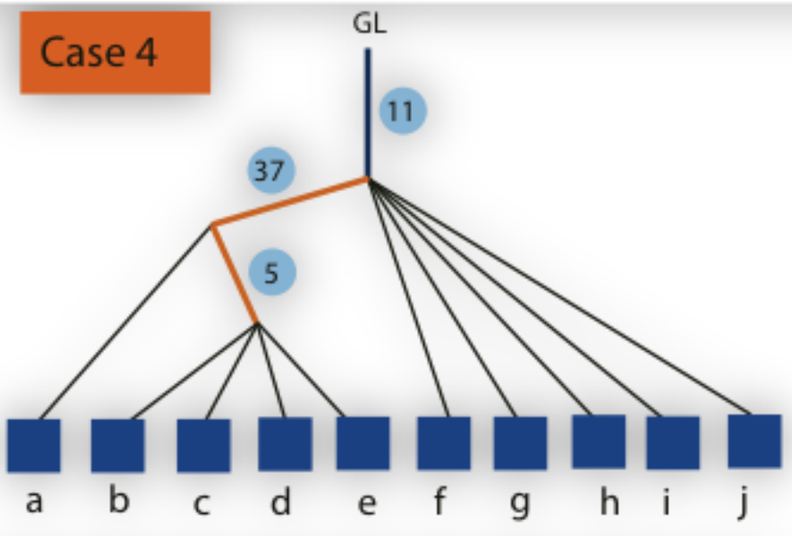
Bashashati, A., et al. (2013). "Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling." The Journal of pathology **231**(1): 21-34.

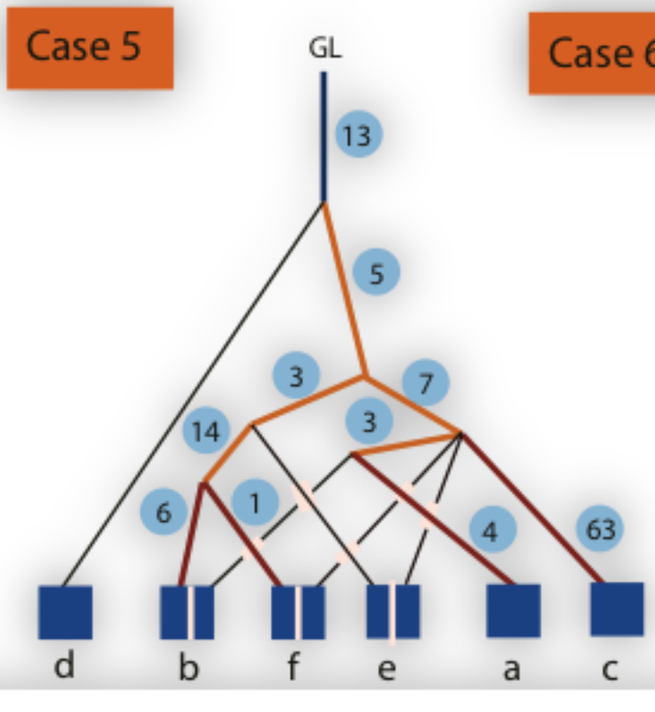
HGSC Study by Bashashati *et. al* (2013)

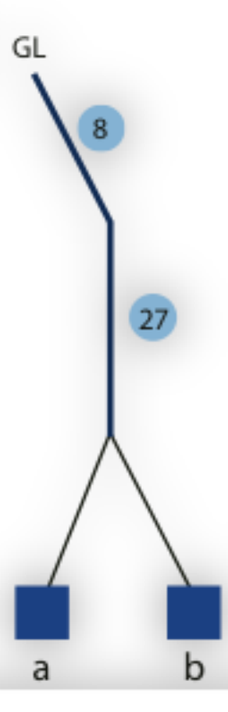


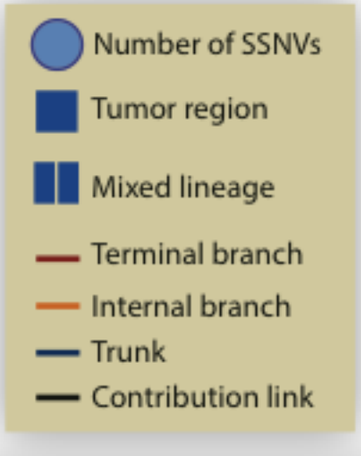




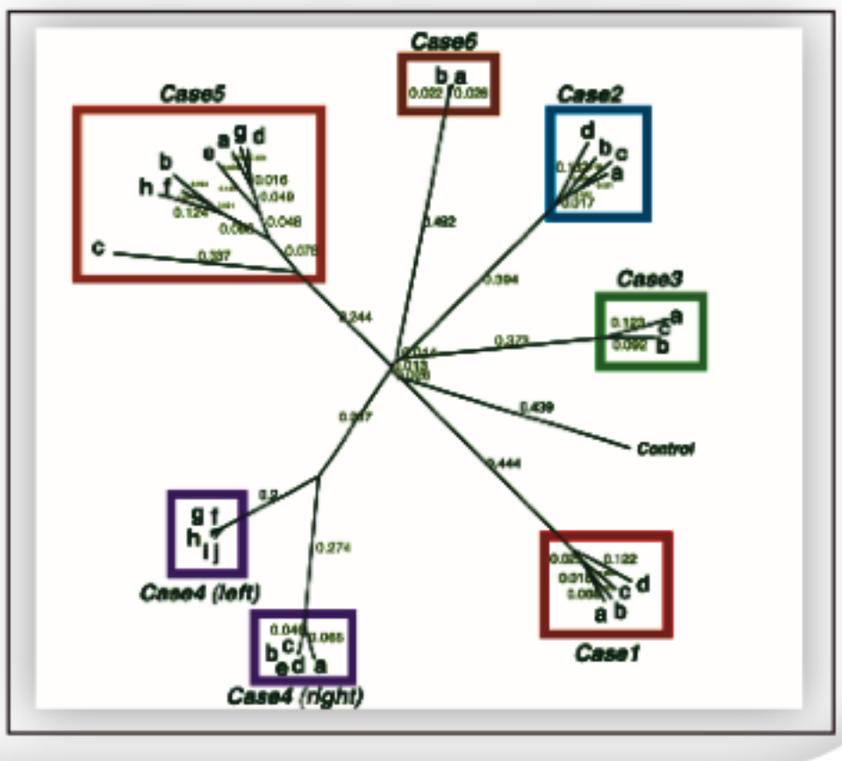




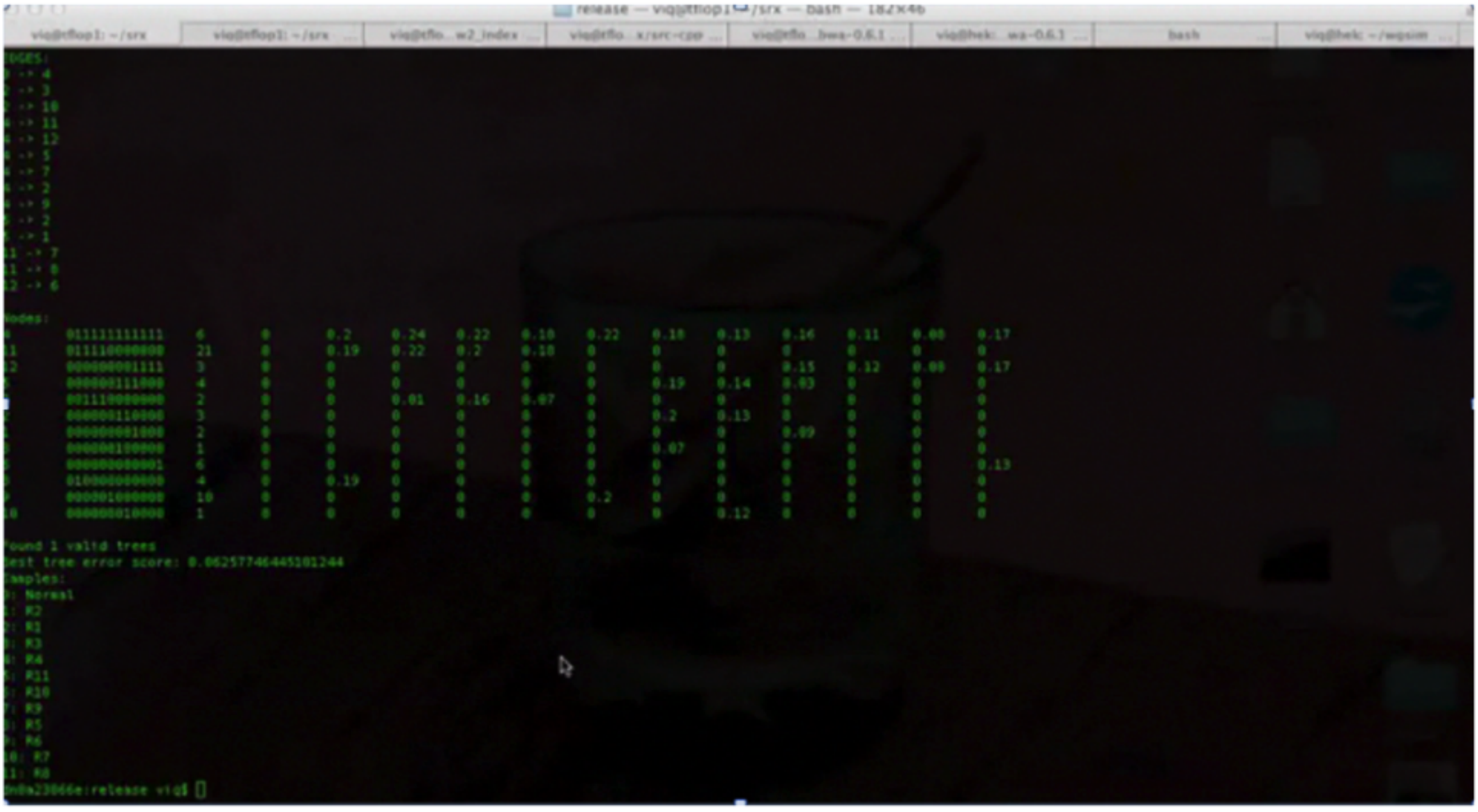




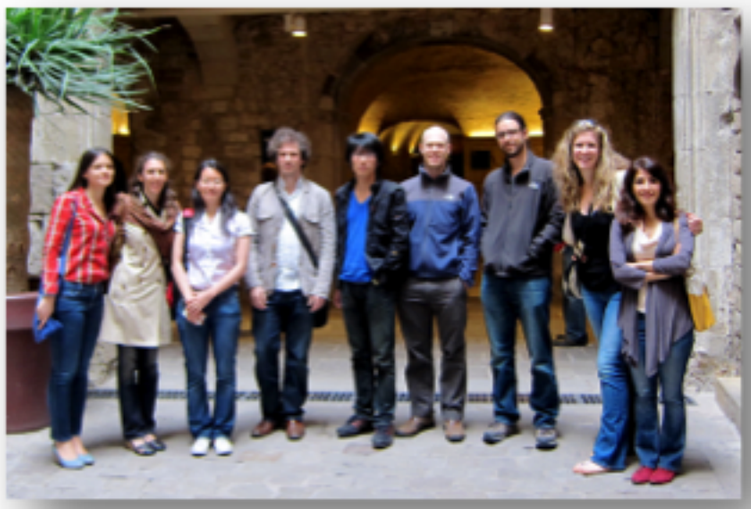
b.



LICHeE Runtime DEMO Movie



Serafim Batzoglou Lab
@ Stanford University



Acknowledgements

Raheleh Salari

Iman Hajirasouliha

Dorna Kashef-Haghighi

Daniel Newburger

Robert West

Arend Sidow

Serafim Batzoglou

Everyone in the audience

Thank You