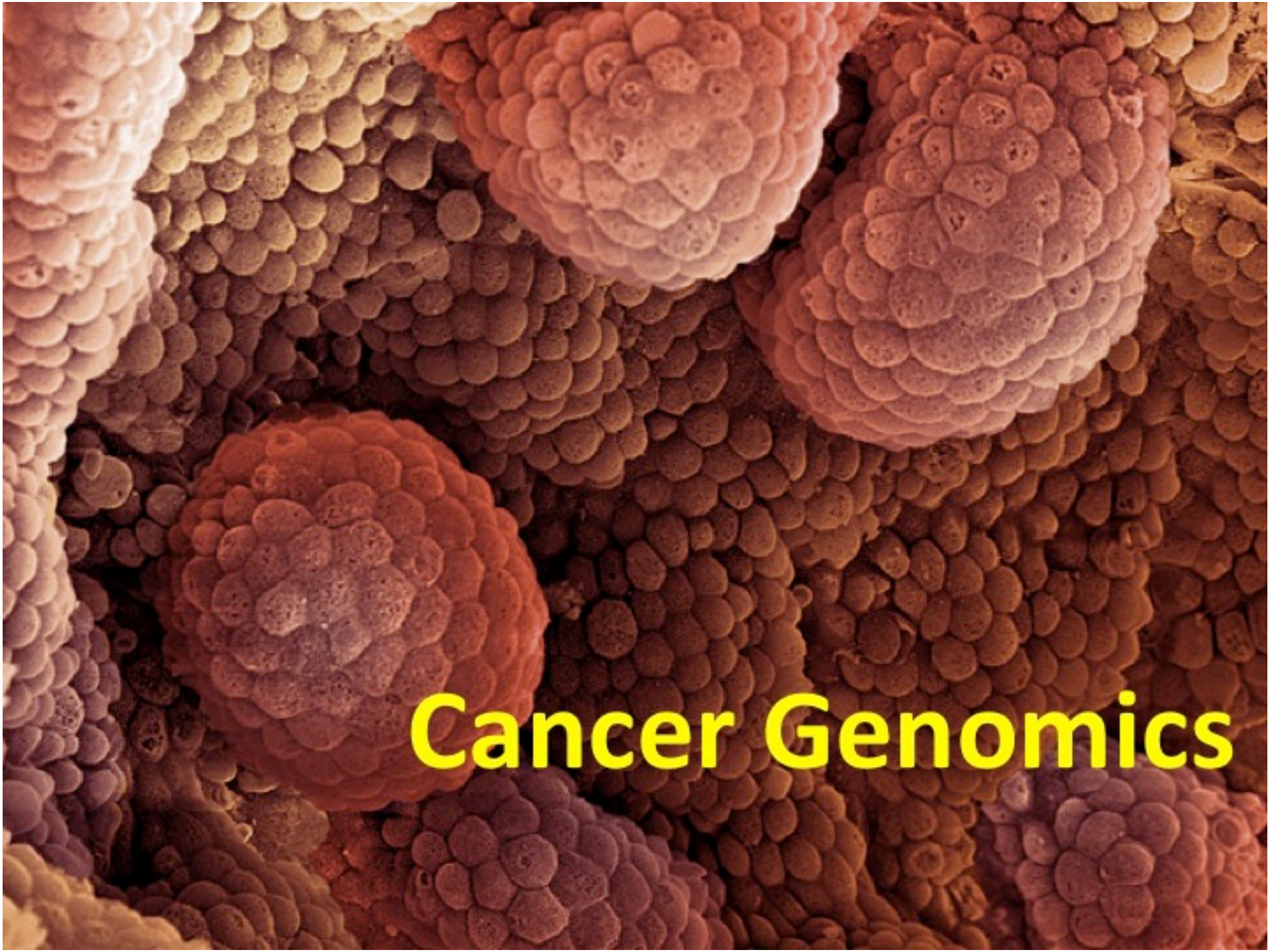
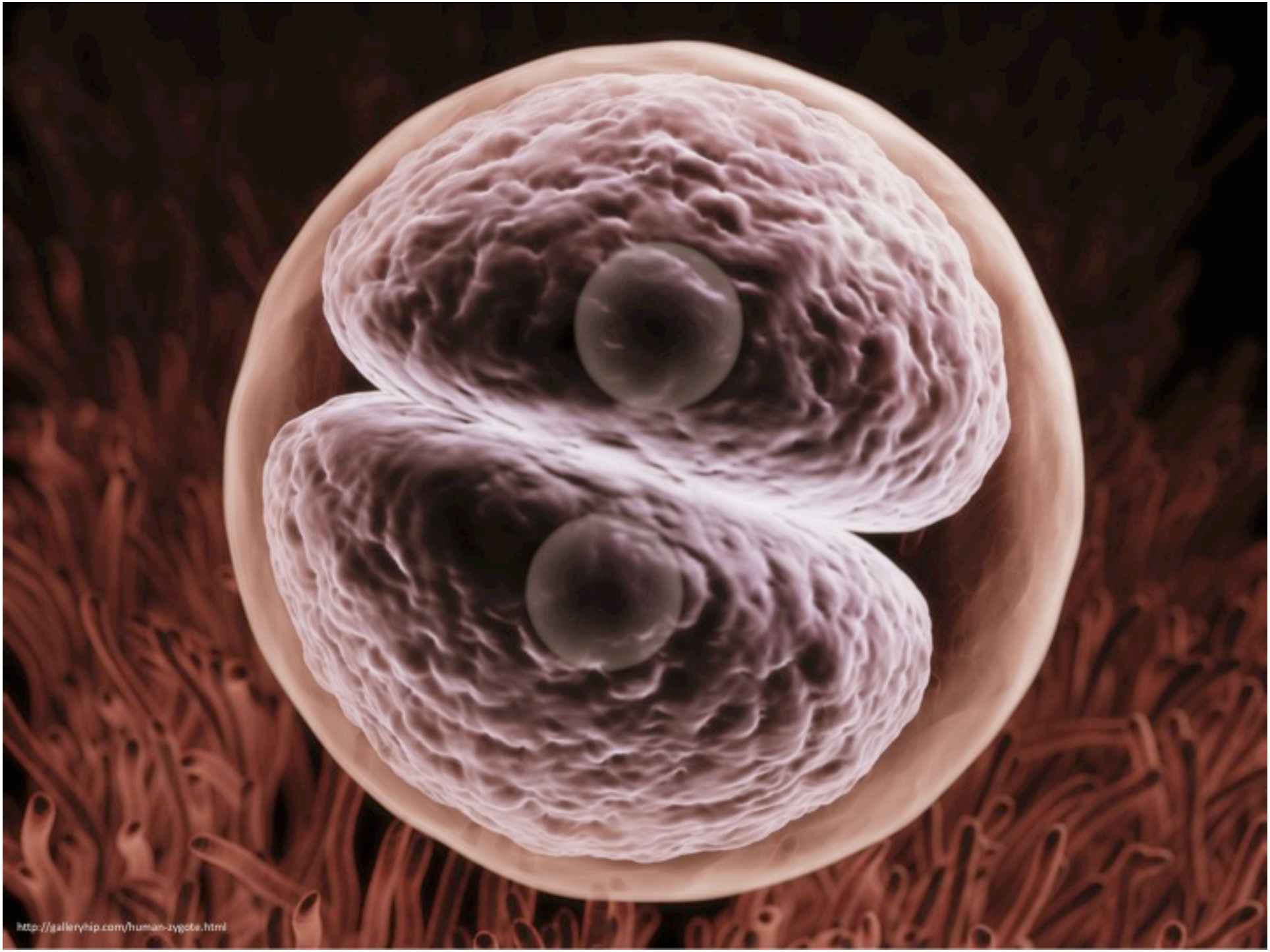
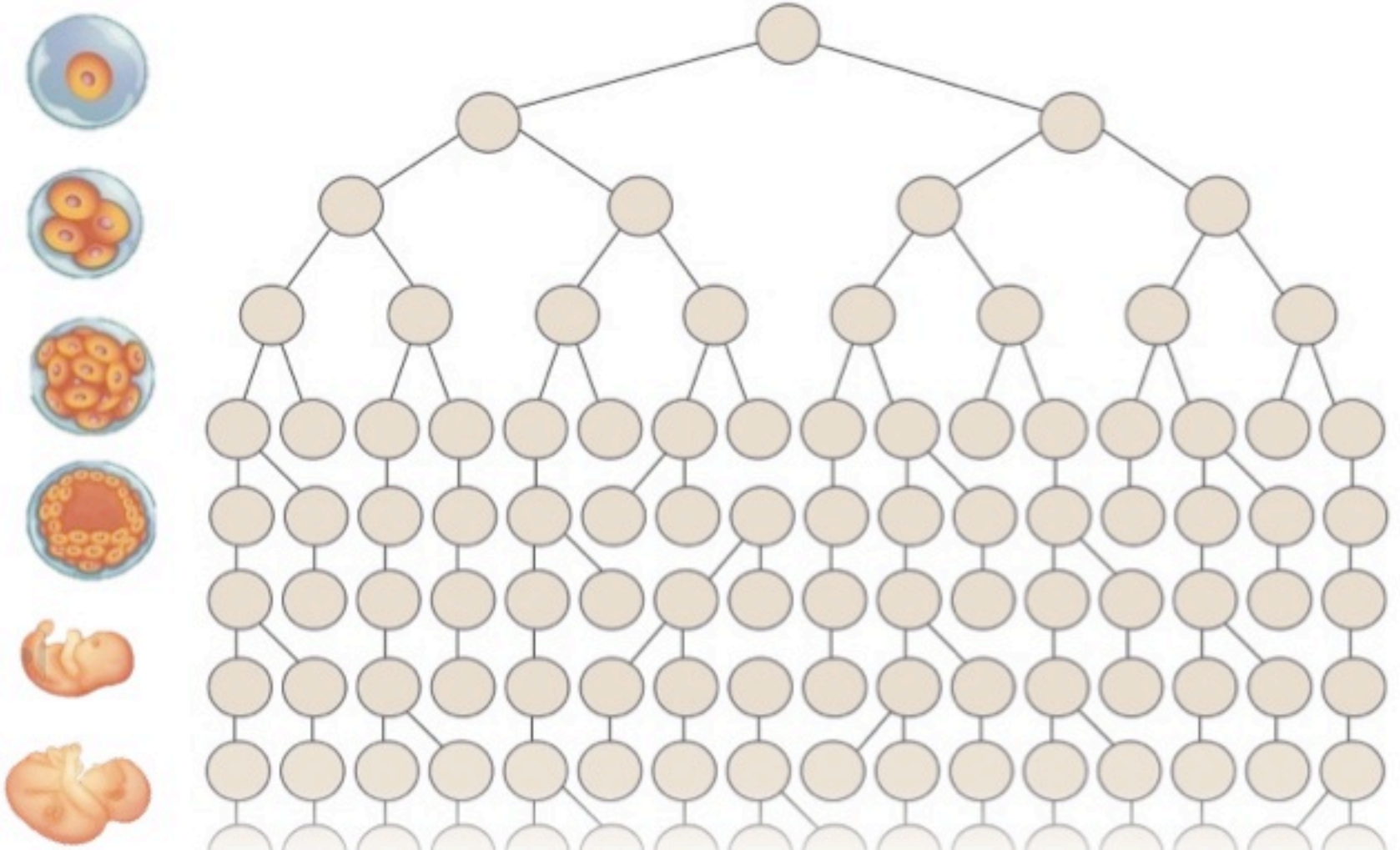
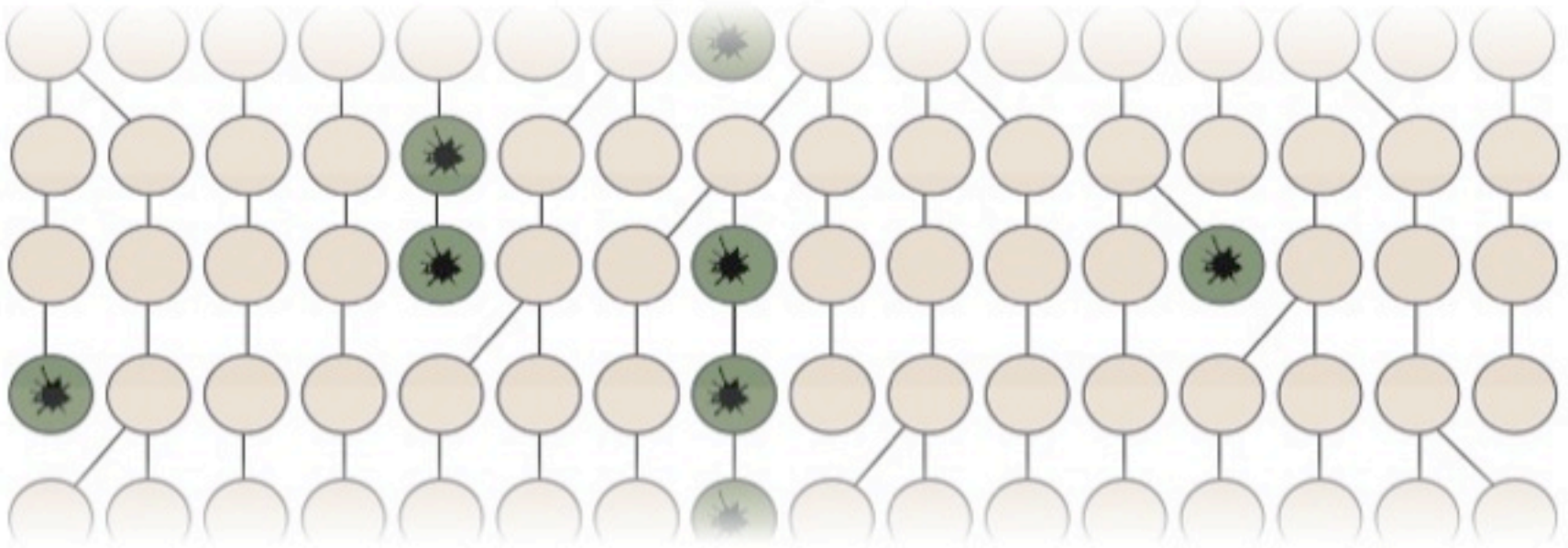

Cancer Genomics

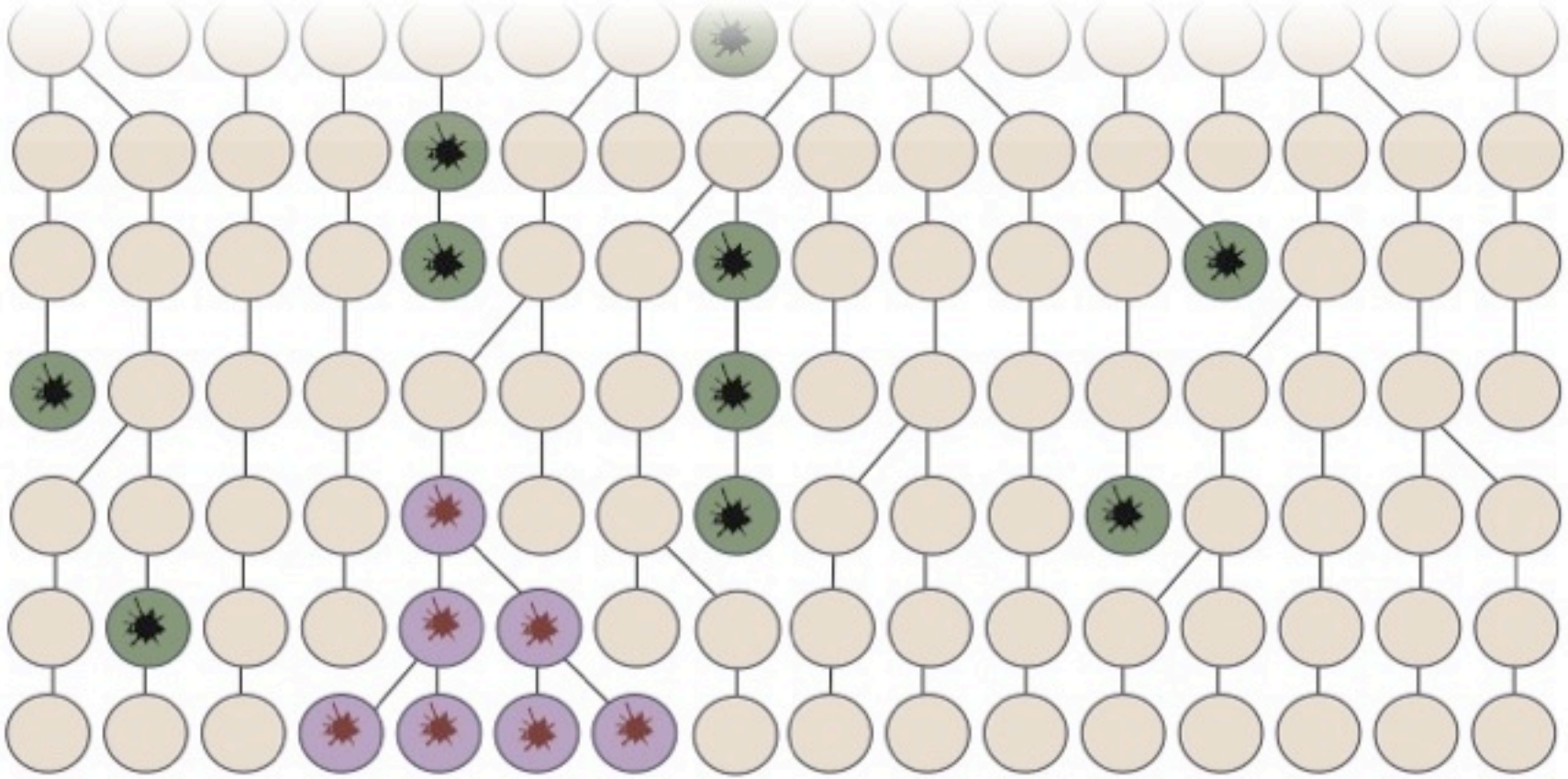

http://galleryhip.com/human-zygote.html



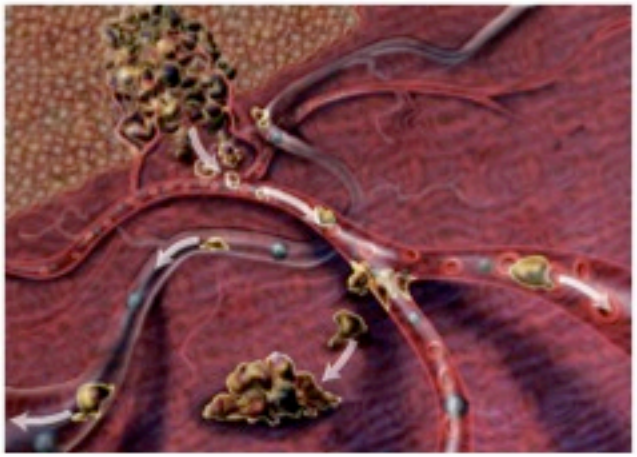
Source: Dorna Kashef-Haghighi

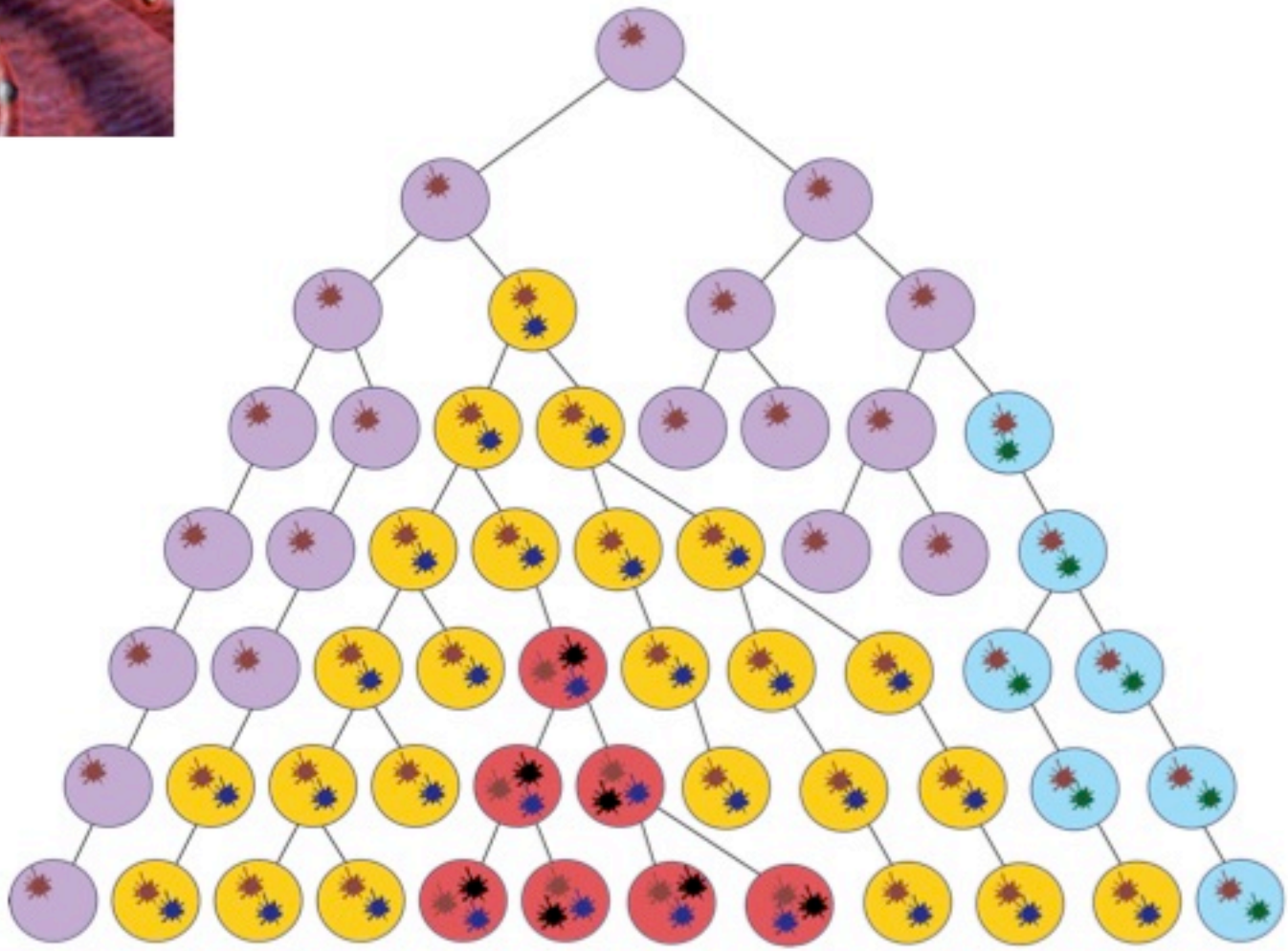


Source: Dorna Kashef-Haghighi

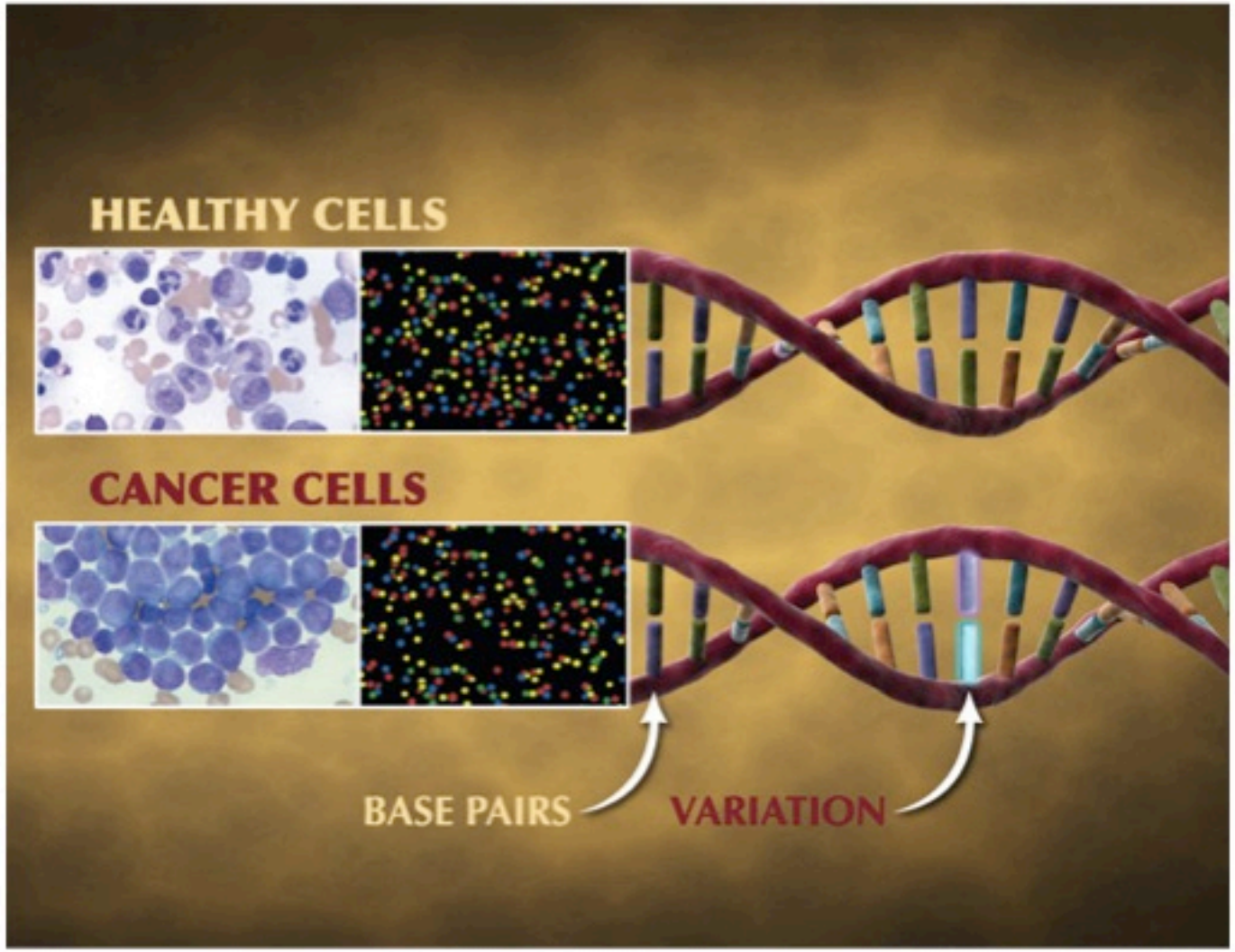


Source: Dorna Kashef-Haghighi





Source: Dorna Kashef-Haghighi



Cancer is a disease of the genome.

Variant Calling Process

Pipeline

Alignment → PCR duplication removal → Base quality Recalibration → Local Realignment → GATK Unified-Genotyper → Filters → Rescue → Final Variants Set

Challenges

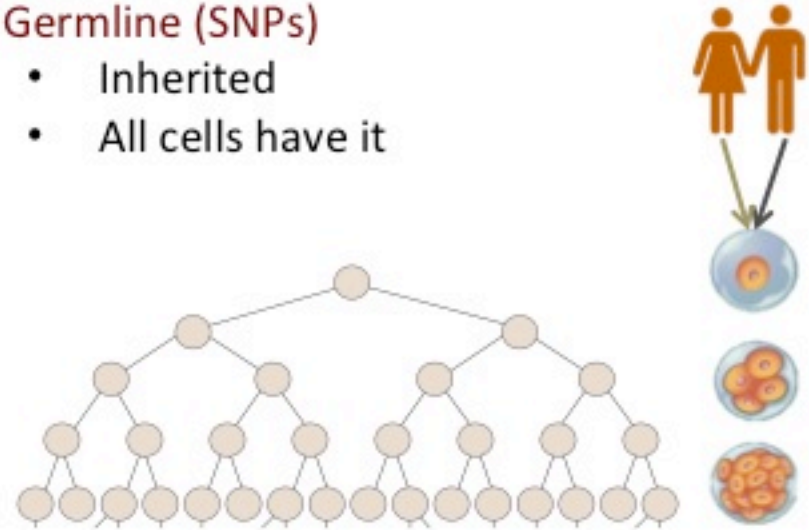
- Sequencing Error
- Mapping Error
- Sequencing Coverage
- Normal Contamination

Source: Dorna Kashef-Haghighi

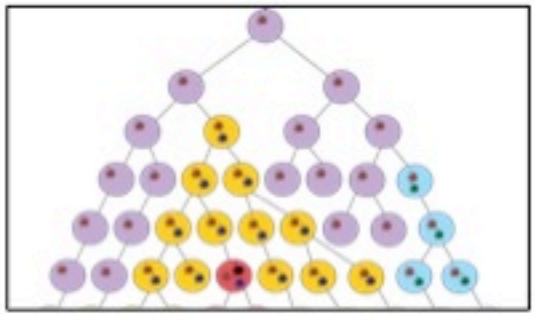
Single Nucleotide Variants (SNVs)

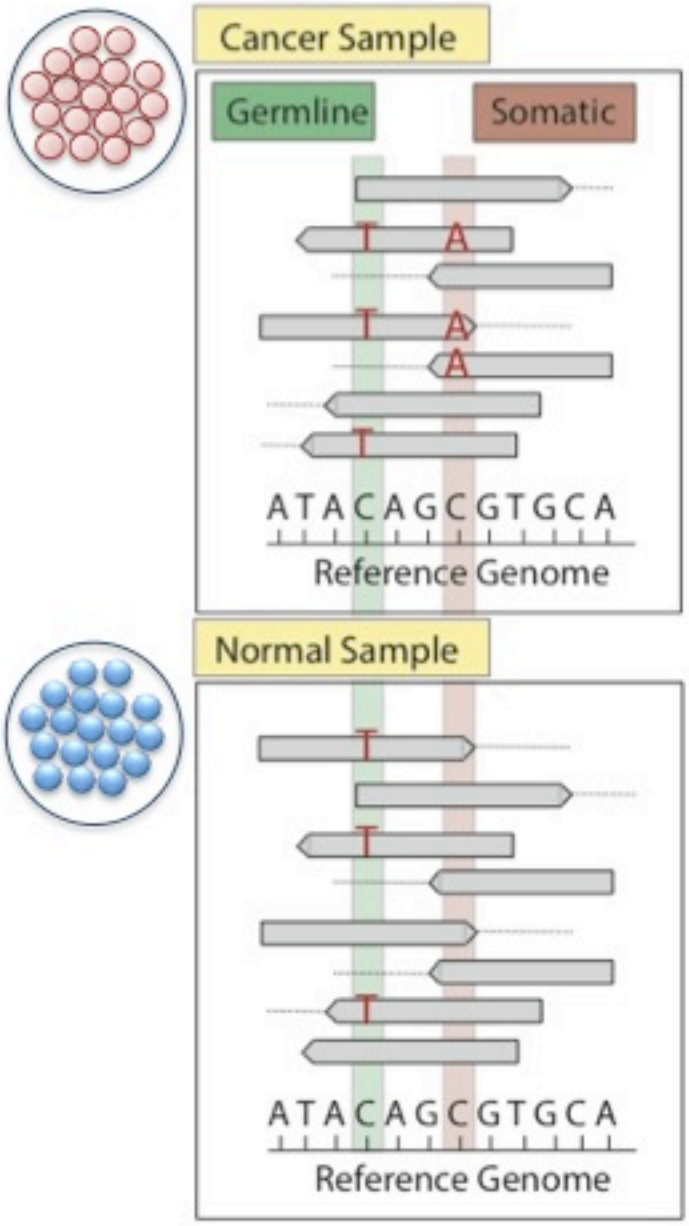
Types of SNVs in a cancer sample:

1. Germline (SNPs)
   - Inherited
   - All cells have it



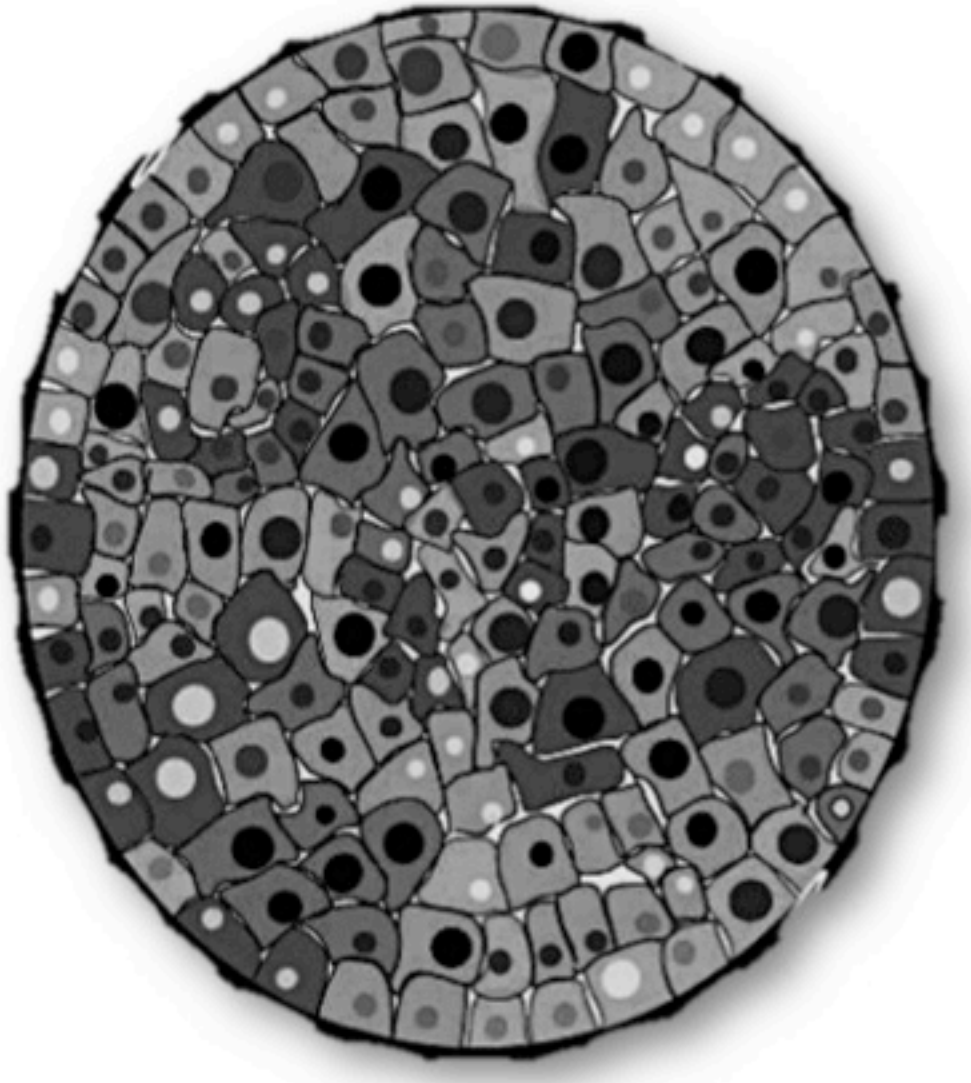
2. Somatic (SSNVs)
   - Acquired during cancer progression
   - Not present in normal cells





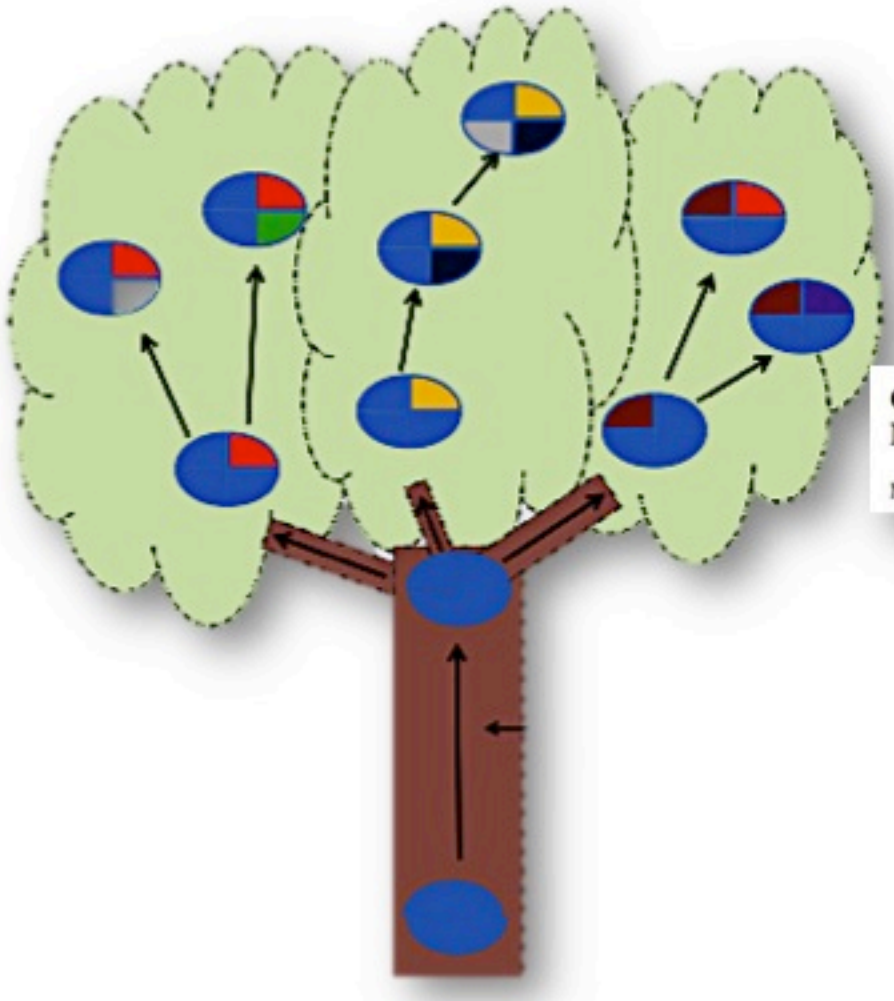
Source: Dorna Kashef-Haghighi

Intra-Tumor Heterogeneity



Intra-Tumor Heterogeneity

Branched Tree Evolution Model



Glioblastoma Multiforme: A Look Inside Its Heterogeneous Nature

Maria-del-Mar Inda [1,7,*], Rudy Bonavia [1,7] and Joan Seoane [1,2]

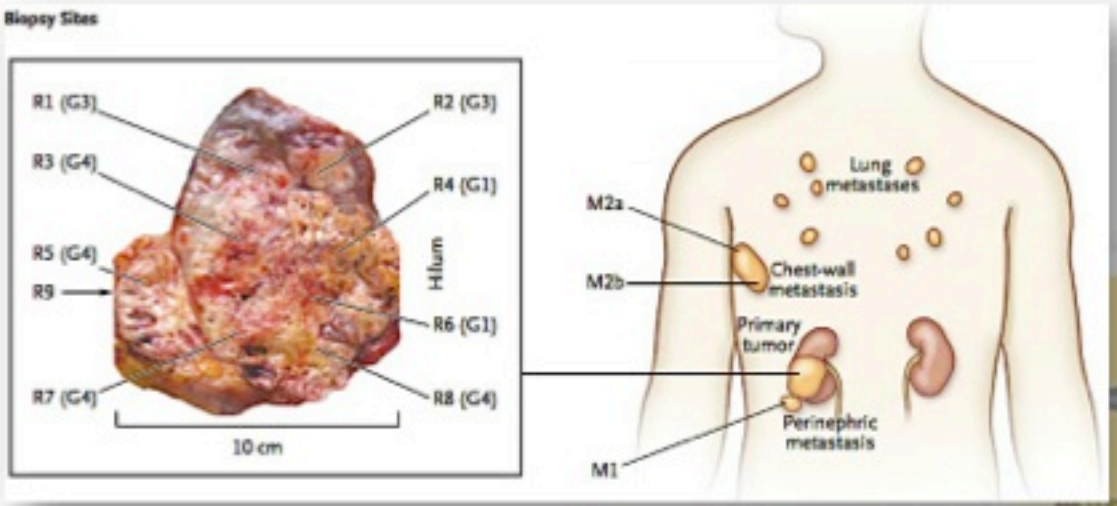
Intratumor Heterogeneity:
Seeing the Wood for the Trees

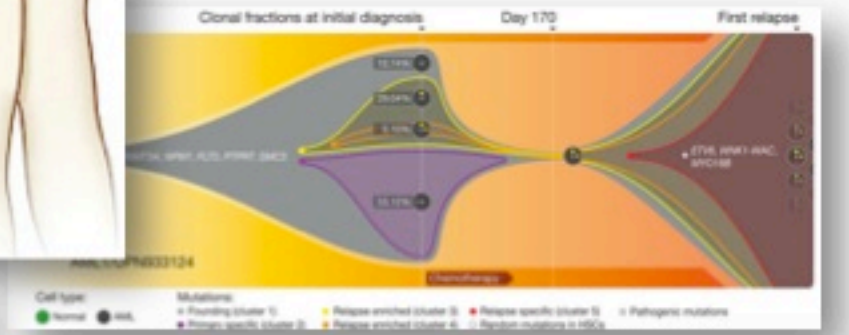
Timothy A. Yap,[1*] Marco Gerlinger,[2,3*] P. Andrew Futreal,[4] Lajos Pusztai,[5] Charles Swanton[2,6†]

Clonal evolution in cancer

Mel Greaves & Carlo C. Maley

Multi-Sample Sequencing





Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing

Marco Gerlinger, M.D., Andrew J. Rowan, B.Sc., Stuart Horswell, M.Math., James Larkin, M.D., Ph.D., David Endesfelder, Dip.Math., Eva Gronroos, Ph.D., Pierre Martinez, Ph.D., Nicholas Matthews, B.Sc., Aengus Stewart, M.Sc., Patrick Tarpey, Ph.D., Ignacio Varela, Ph.D., Benjamin Phillimore, B.Sc., Sharmin Begum, M.Sc., Neil Q. McDonald, Ph.D., Adam Butler, B.Sc., David Jones, M.Sc., Keiran Raine, M.Sc., Calli Latimer, B.Sc., Claudio R. Santos, Ph.D., Mahrokh Nohadani, H.N.C., Aron C. Eklund, Ph.D., Bradley Spencer-Dene, Ph.D., Graham Clark, B.Sc., Lisa Pickering, M.D., Ph.D., Gordon Stamp, M.D., Martin Gore, M.D., Ph.D., Zoltan Szallasi, M.D., Julian Downward, Ph.D., P. Andrew Futreal, Ph.D., and Charles Swanton, M.D., Ph.D.

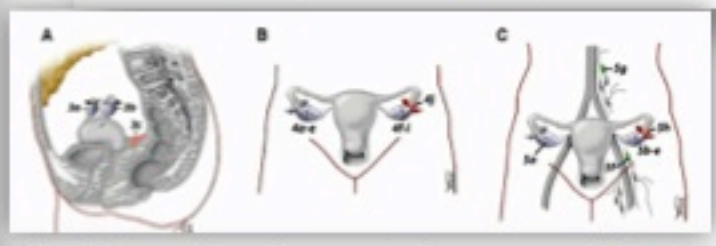
N Engl J Med 2012; 366:883-892 | March 8, 2012 | DOI: 10.1056/NEJMoa1113205

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. Welch, Julie K. Ritchey, Margaret A. Young, Tamara Lamprecht, Michael D. McLellan, Joshua F. McMichael, John W. Wallis, Charles Lu, Dong Shen, Christopher C. Harris, David J. Dooling, Robert S. Fulton, Lucinda L. Fulton, Ken Chen, Heather Schmidt, Joelle Kalicki-Veizer, Vincent J. Magrini, Lisa Cook, Sean D. McGrath, Tammi L. Vickery + *et al.*

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

Marco Gerlinger, Stuart Horswell, James Larkin, Andrew J Rowan, Max P Salm, Ignacio Varela, Rosalie Fisher, Nicholas McGranahan, Nicholas Matthews, Claudio R Santos, Pierre Martinez, Benjamin Phillimore, Sharmin Begum, Adam Rabinowitz, Bradley Spencer-Dene, Sakshi Gulati, Paul A Bates, Gordon Stamp, Lisa Pickering, Martin Gore, Steven Hazell, P Andrew Futreal, Aengus Stewart & Charles Swanton



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

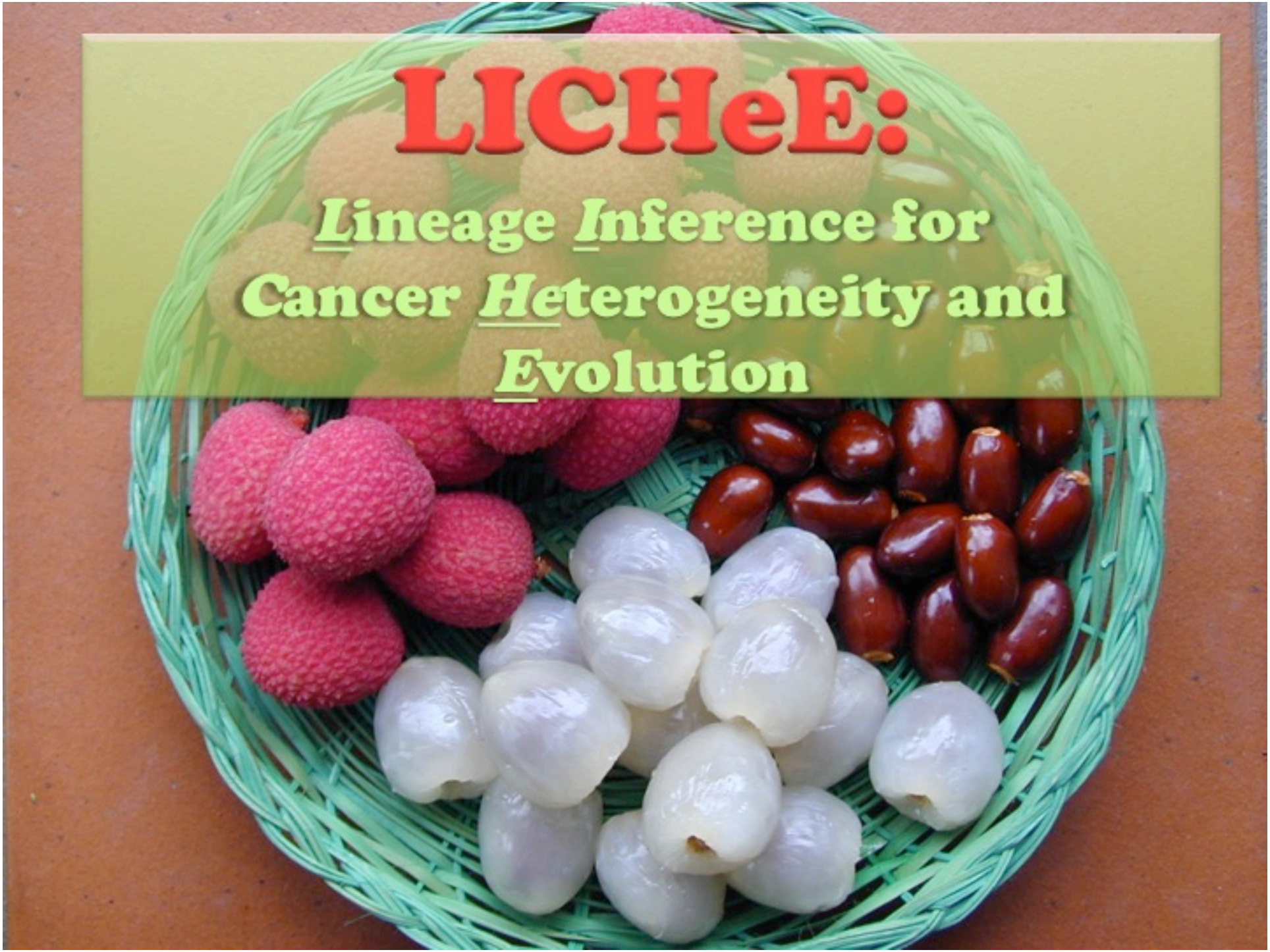
Bashashati A[1], Ha G, Tone A, Ding J, Prentice LM, Roth A, Rosner J, Shumansky K, Kalloger S, Senz J, Yang W, McConechy M, Melnyk N, Anglesio M, Luk MT, Tse K, Zang T, Moore R, Zhao Y, Marra MA, Gilks B, Yip S, Huntsman DG, McAlpine JN, Shah SP.

Fast and Scalable Inference of Multi-Sample Cancer Lineages

V. Popic[1], R. Salari[1], I. Hajirasouliha[1],
D. Kashef-Haghighi[1], R. West[2],
S. Batzoglou[1]

[1]Department of Computer Science, Stanford University
[2] Department of Pathology, Stanford University School of Medicine

No conflicts of interest to declare


LICHeE:
Lineage Inference for
Cancer Heterogeneity and
Evolution

LICHeE: Method Overview

Given: SSNV multi-sample variant allele frequencies (VAFs)

Algorithm steps:

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

Perfect Phylogeny Model: Constraints

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 SSNV 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 SSNV 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

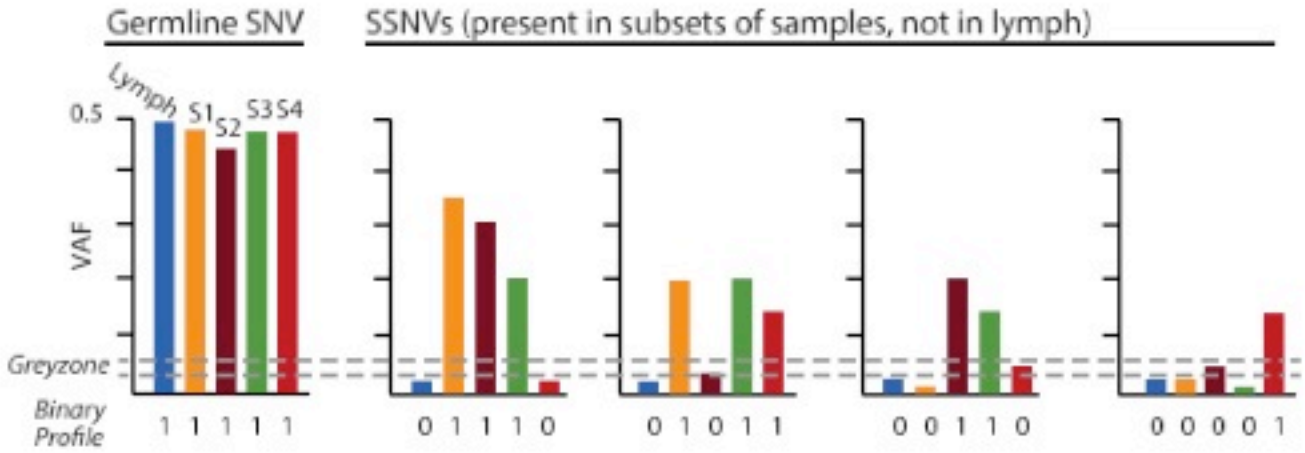
Goal: find all lineage trees that satisfy the above three constraints

1. **Grouping and clustering SSNVs**

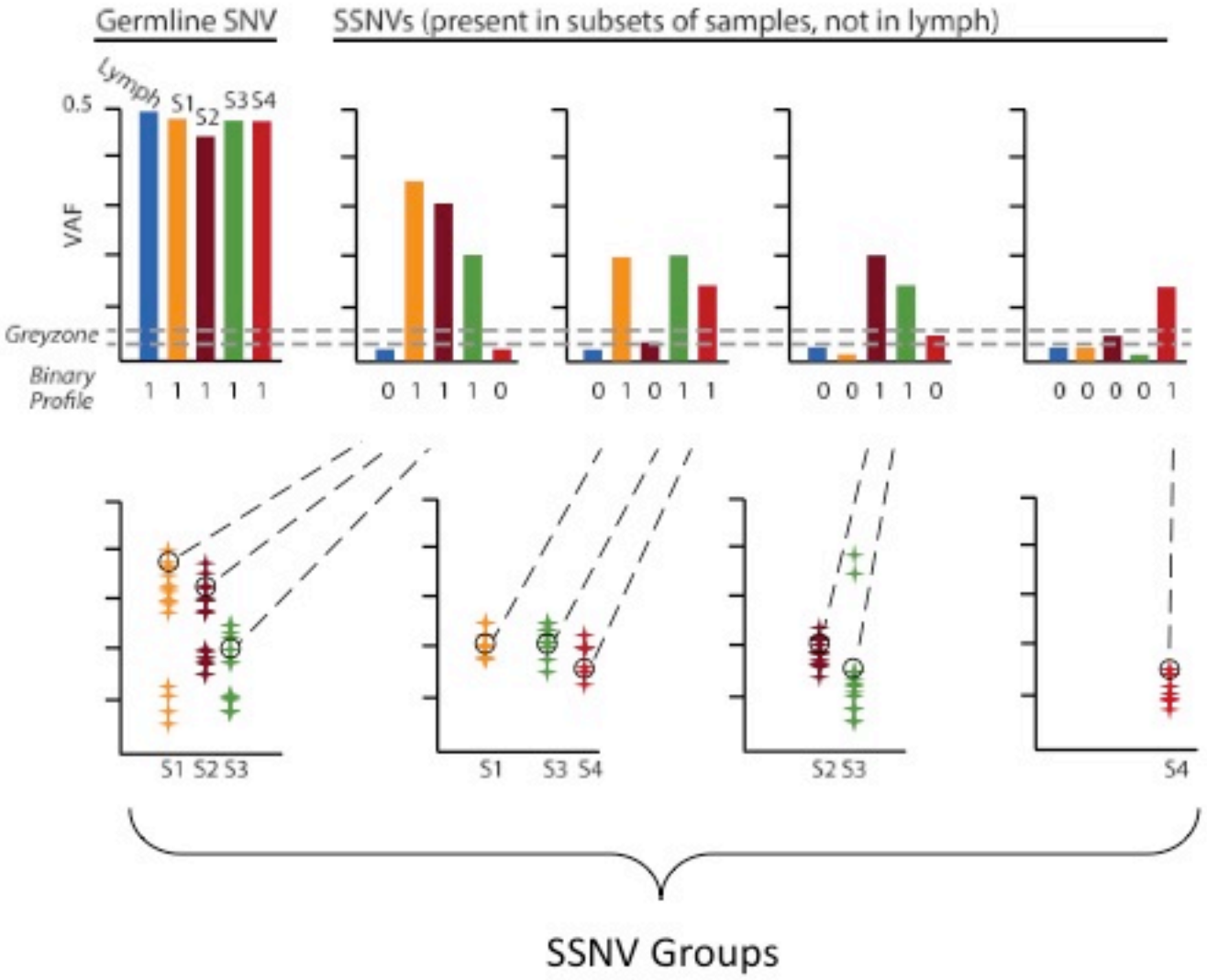
   → presence patterns across samples

   → VAF similarity

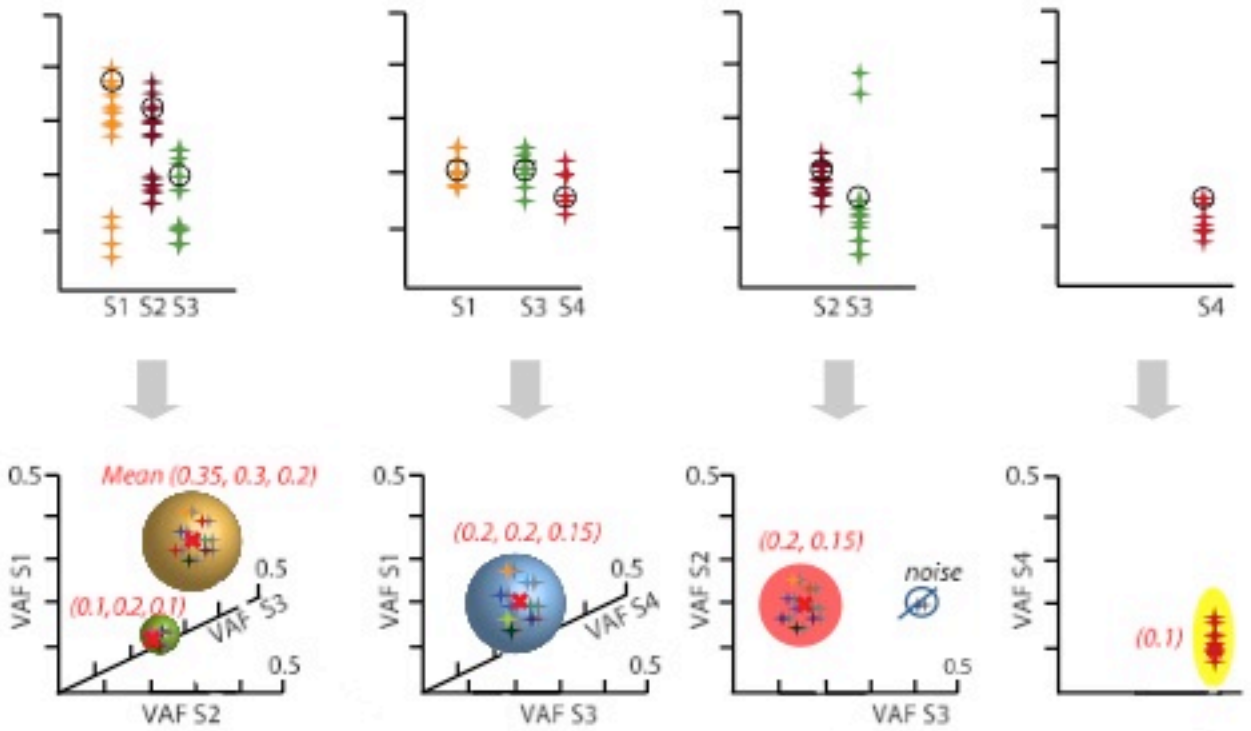
Presence Patterns Across Samples



Presence Patterns Across Samples



VAF-Based Clustering

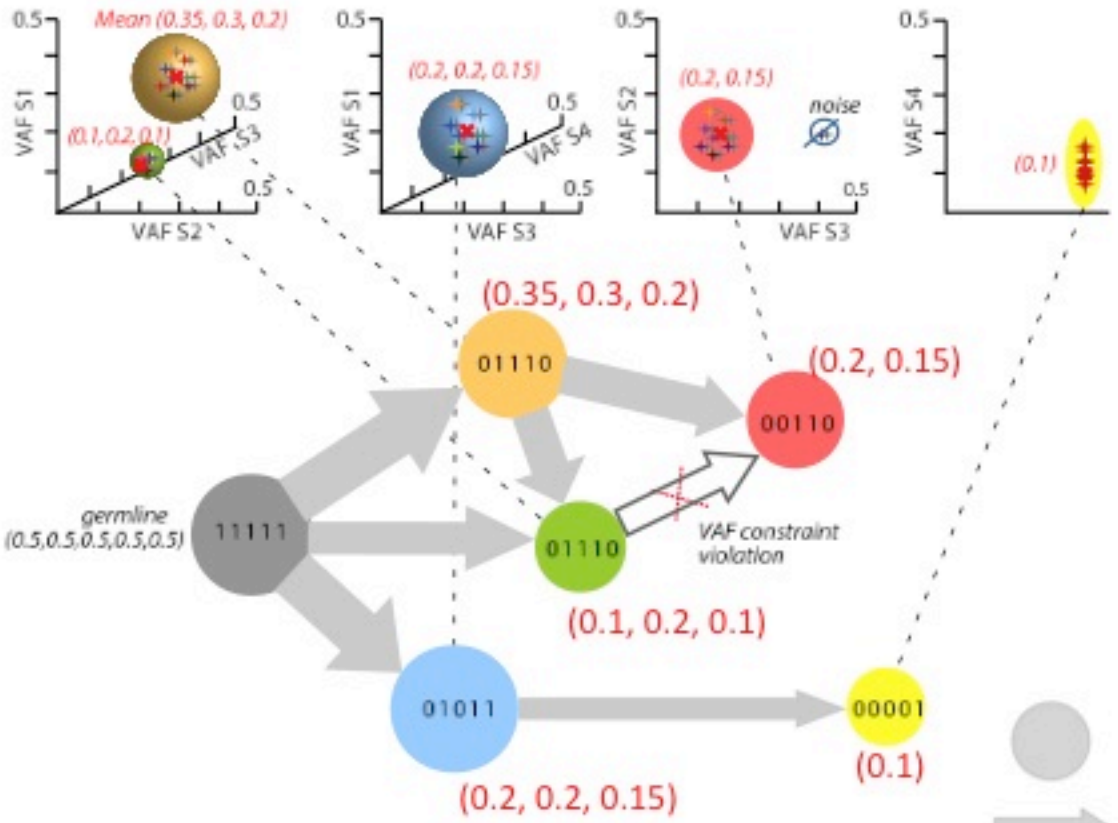


2. Evolutionary Constraint Network Construction

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

→ valid lineage trees are embedded in this network

Evolutionary Constraint Network



(1) $u.\overline{\mathrm{VAF}}_i \geq v.\overline{\mathrm{VAF}}_i - \epsilon_{uv}$ and
(2) if $u.\overline{\mathrm{VAF}}_i = 0, v.\overline{\mathrm{VAF}}_i = 0$,

3. Lineage Tree Search and Ranking

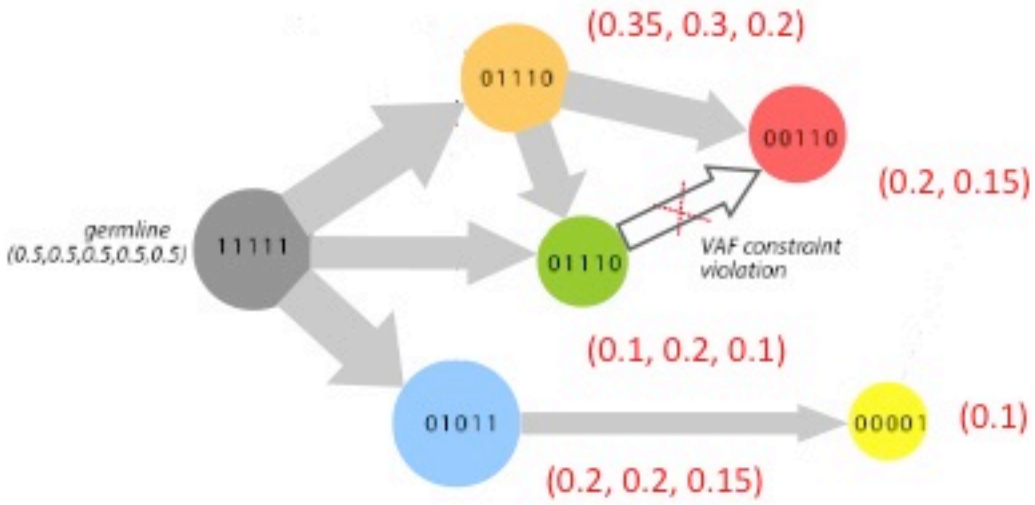
→ search for spanning trees satisfying VAF constraints within an error margin

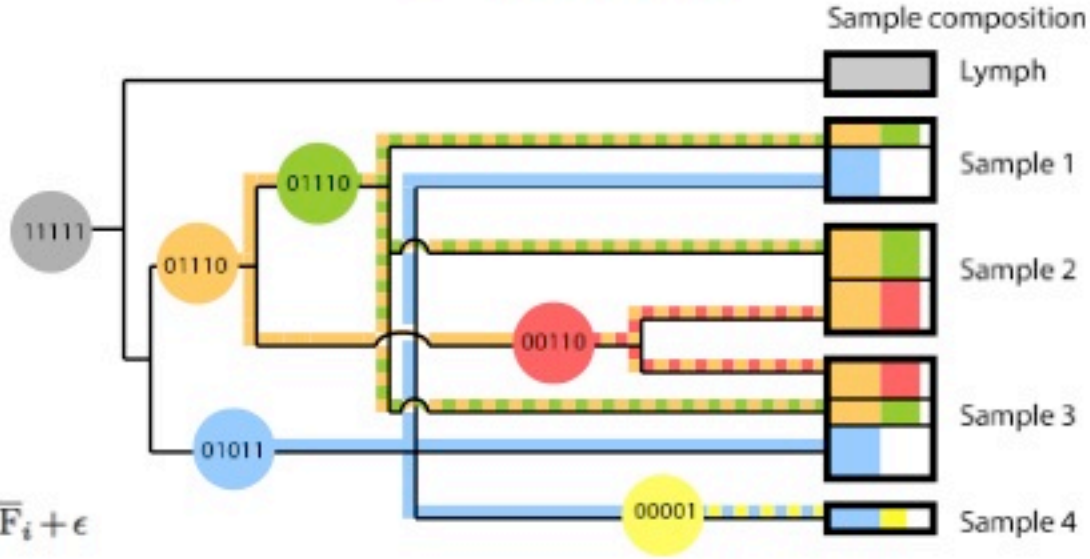
→ top tree minimizes the squared deviation from the cluster centroids

Lineage Tree Search

Algorithm 1 Finding All Lineage Trees

```
1: Initialization: f ← empty list, L ← null // stores the last tree output
2: procedure LINEAGE TREE SEARCH(N) // N is a constraint network rooted at r
3:     Tree t ← new empty Tree
4:     t.ADDNODE(r)
5:     add all edges (r → v) ∈ N to f
6:     GROW(t)
7: procedure GROW(t, N)
8:     if t contains all the nodes in N then
9:         L ← t
10:        output L
11:    else
12:        s ← empty list
13:        b ← false
14:        while (not b and f not empty) do
15:            // e defined as (e.From → e.To)
16:            Edge e ← f.REMOVELAST()
17:            Node v ← e.To
18:            t.ADDNODE(v)
19:            t.ADDEDGE(e.From → v)
20:            // ret. true if Eqn. (5) is satisfied for node e.From
21:            if t.CHECKCONSTRAINT(e.From) then
22:                add all edges (v → w), w ∉ t to f
23:                remove all edges (w → v), w ∈ t from f
24:                GROW(t)
25:                if number of returned trees > max_trees return
26:                remove all edges (v → w), w ∉ t from f
27:                add all edges (w → v), w ∈ t to f
28:            t.REMOVEEDGE(e.From → e.To)
29:            N.REMOVEEDGE(e.From → e.To)
30:            s.ADD(e)
31:            if ∃ an edge (w → v) s.t. w not a descendent of v in L then
32:                b ← false
33:            else b ← true
34:        for all edges e starting from the end of s do
35:            remove e from s, add e to f, add e to N
```

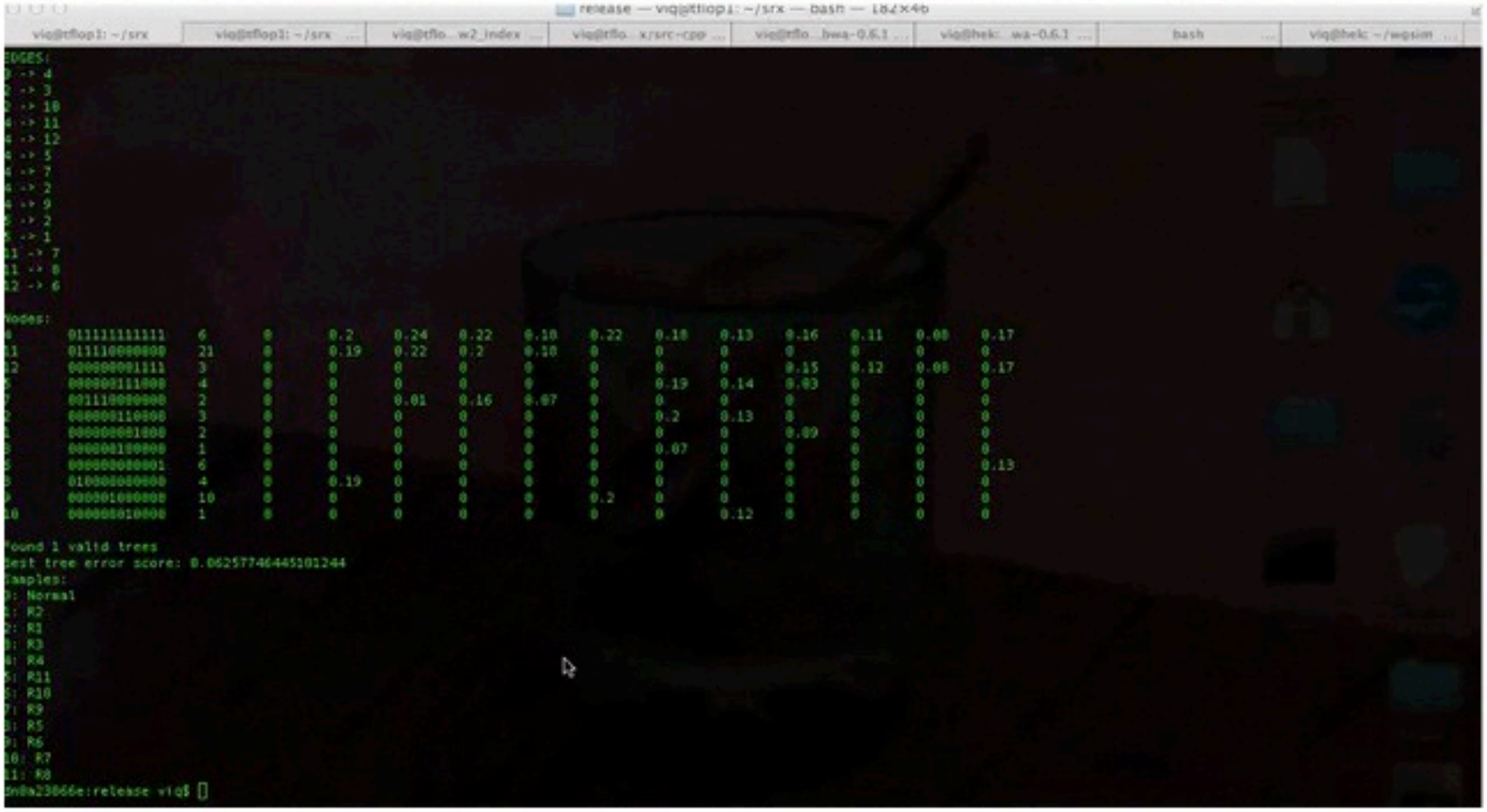




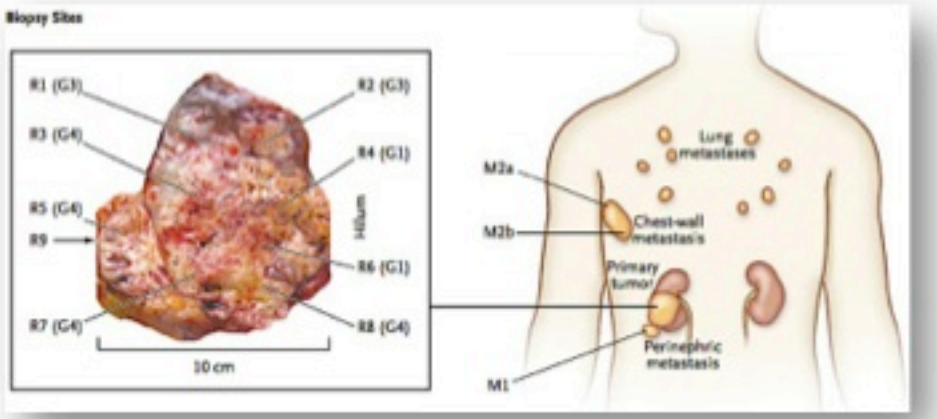
$$\forall i \in \text{samples} \;:\; \sum_{v \text{ s.t. } (u \to v) \in T} v.\overline{\text{VAF}}_i \leq u.\overline{\text{VAF}}_i + \epsilon$$

RESULTS

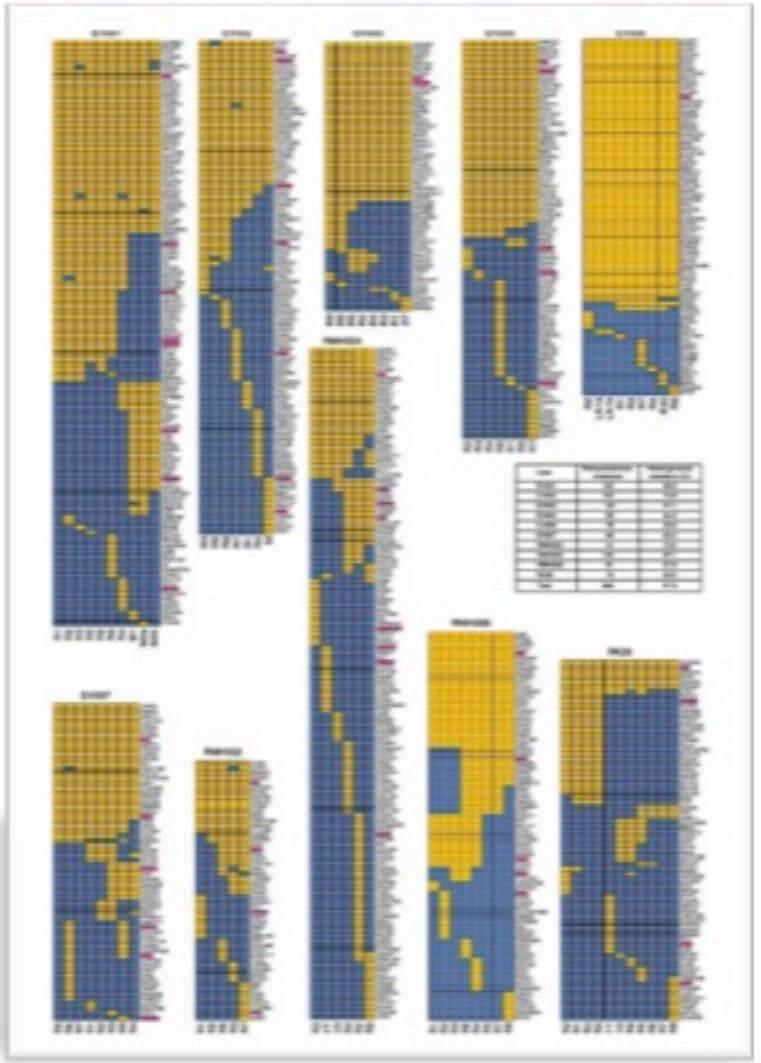
LICHeE Runtime DEMO Movie



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

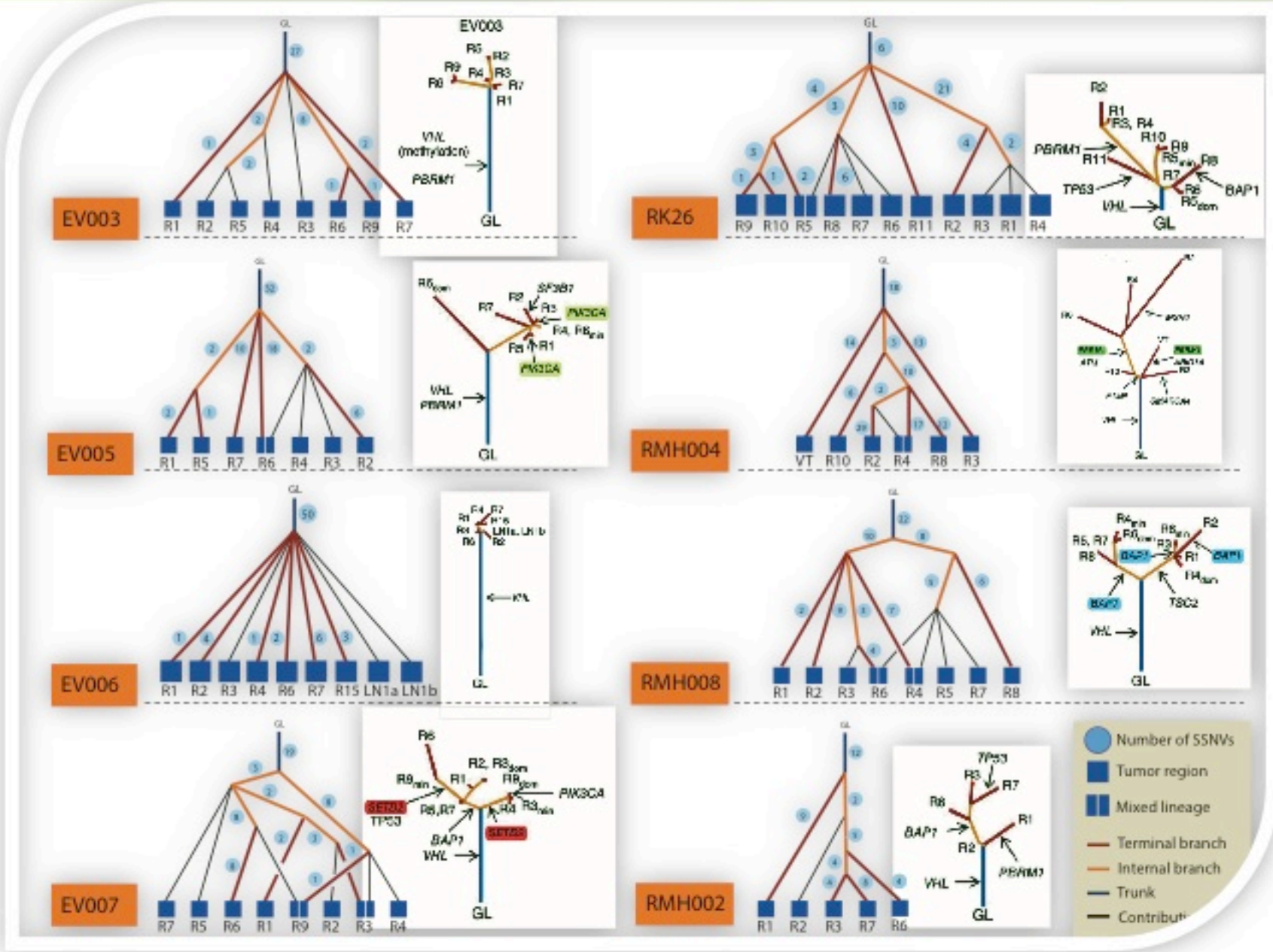


8 patients, 587 SNVs

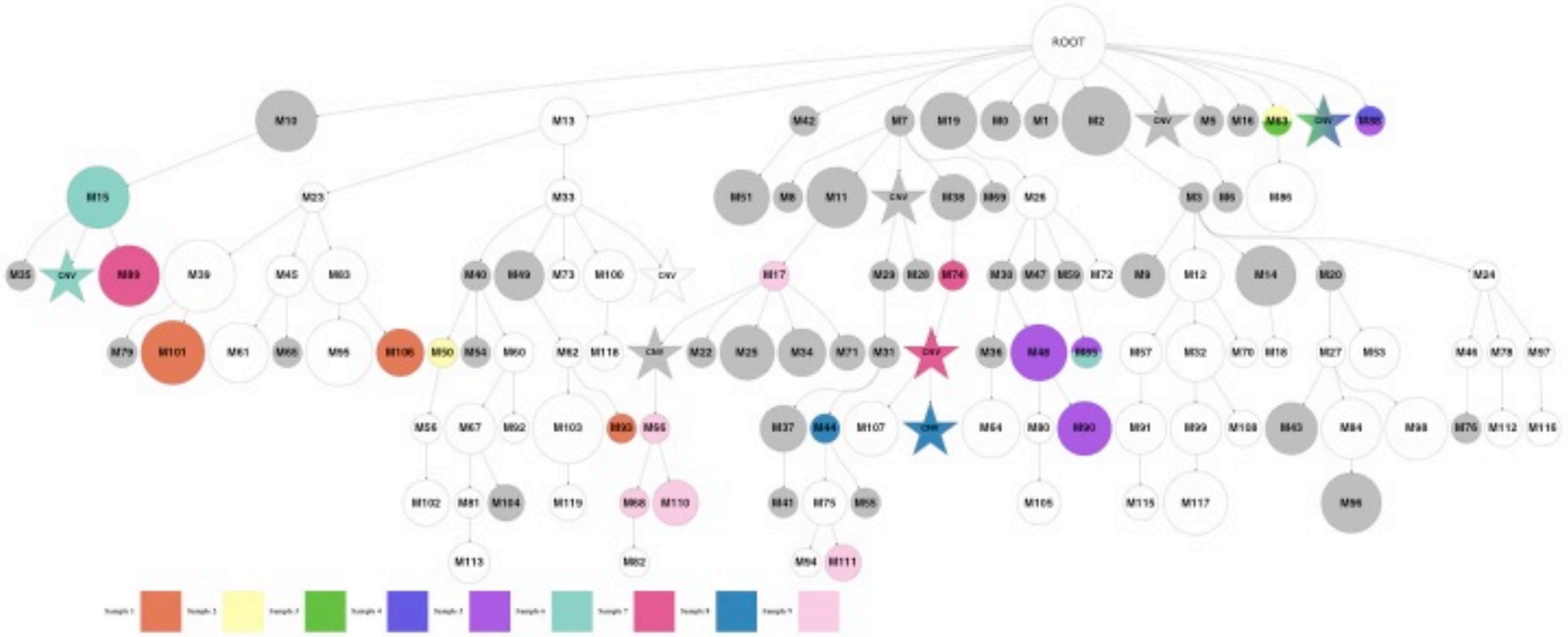


Gerlinger, M., et al. (2014). "Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing." Nature genetics **46**(3): 225-233.

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



Simulations



Method Overview

1. Call somatic SNVs in samples

2. Group SNVs using sample presence patterns

3. Cluster groups based on VAFs

4. Construct the evolutionary constraint network: captures all phylogenetically valid precedence relationships among cluster pairs

5. Search for valid lineage trees (applying VAF constraints) and rank the trees