

Клональные
деревья не
ТОЛЬКО
В-клеток

И почему
филогения здесь
ни при чём

1. Cancer clonal theory

- All cells within a tumor have descended from a single founder cell
- Inference of the clonal evolutionary history of somatic mutations provides useful insight in the tumor's development.

1. Cancer clonal theory

- ◉ Single sample: all cells come from a solid tumor at one timepoint.
- ◉ Multiple samples:
 - multiple spatially distinct regions from the same tumor
 - multiple time points

1. Cancer clonal theory

- ◉ Single sample: all cells come from a solid tumor at one timepoint.
- ◉ Multiple samples:
 - multiple spatially distinct regions from the same tumor
 - multiple time points

Single sample methods

Variant Allele Frequencies of SNPs

- fraction of tumor cells that contain each mutation

Probabilistic models

- require additional assumptions about the evolutionary process

1. Cancer clonal theory

- ◉ Single sample: all cells come from a solid tumor at one timepoint.
- ◉ Multiple samples:
 - multiple spatially distinct regions from the same tumor
 - multiple time points

Multiple samples methods

Two different problems, we need to:

- Infer the clones fractioning
- Reconstruct global ancestral relationships

VARIANT ALLELE FREQUENCY FACTORIZATION PROBLEM

Given a sample of clones such that

- For any pairs of clones j and k either SNP sets $I(j)$ and $I(k)$ are disjoint, or one contains the other
- No mutation appears more than once

VARIANT ALLELE FREQUENCY FACTORIZATION PROBLEM

Two different problems, we need to:

- ? Infer the clones fractioning
- ✓ Reconstruct global ancestral relationships

2. B-cells

Again we have two problems:

- ◉ Decomposition on clonal lineages
- ◉ Reconstruction of the ancestral relationships within one lineage

Simple
methods

Clusterization + some
phylogenetics
algorithm

Simple
methods

Clusterization + some
phylogenetics
algorithm



Simple methods

Clusterization + some phylogenetics algorithm

Boring

- We do not use ancestral information during clusterization phase
- Phylogenetics algorithms assume that all of the cell are leaves of the tree

2. B-cells

Again we have two problems:

- ◉ Decomposition on clonal lineages
- ◉ Reconstruction of the ancestral relationships within one lineage

2. B-cells

Again we have two problems:


- ◉ Decomposition on clonal lineages
- ◉ Reconstruction of the ancestral relationships within one lineage

2. B-cells

Again we have two problems:

- Decomposition on clonal lineages
- **Reconstruction of the ancestral relationships within one lineage**

Only phylogenetics!




2. B-cells

Again we have two problems:

- Decomposition on clonal lineages
- **Reconstruction of the ancestral relationships within one lineage**

Only phylogenetics! ...



2. B-cells

Again we have two problems:

- ◉ Decomposition on clonal lineages
- ◉ Reconstruction of the ancestral relationships within one lineage

2. B-cells

Again we have two problems:

- ◉ Decomposition on clonal lineages
 - seeded
 - unseeded

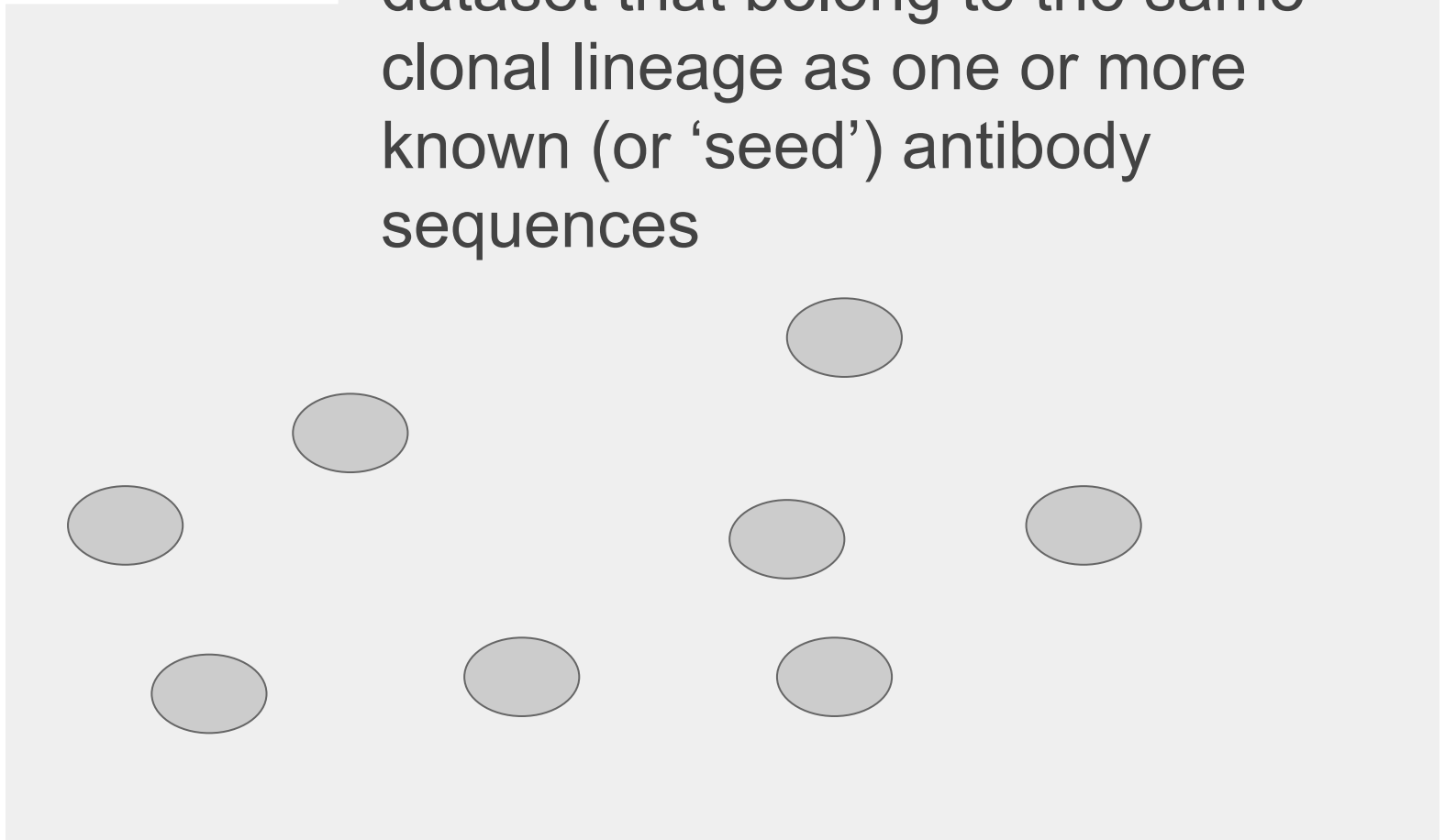
2. B-cells

Again we have two problems:

- ◉ Decomposition on clonal lineages
 - seeded
 - unseeded

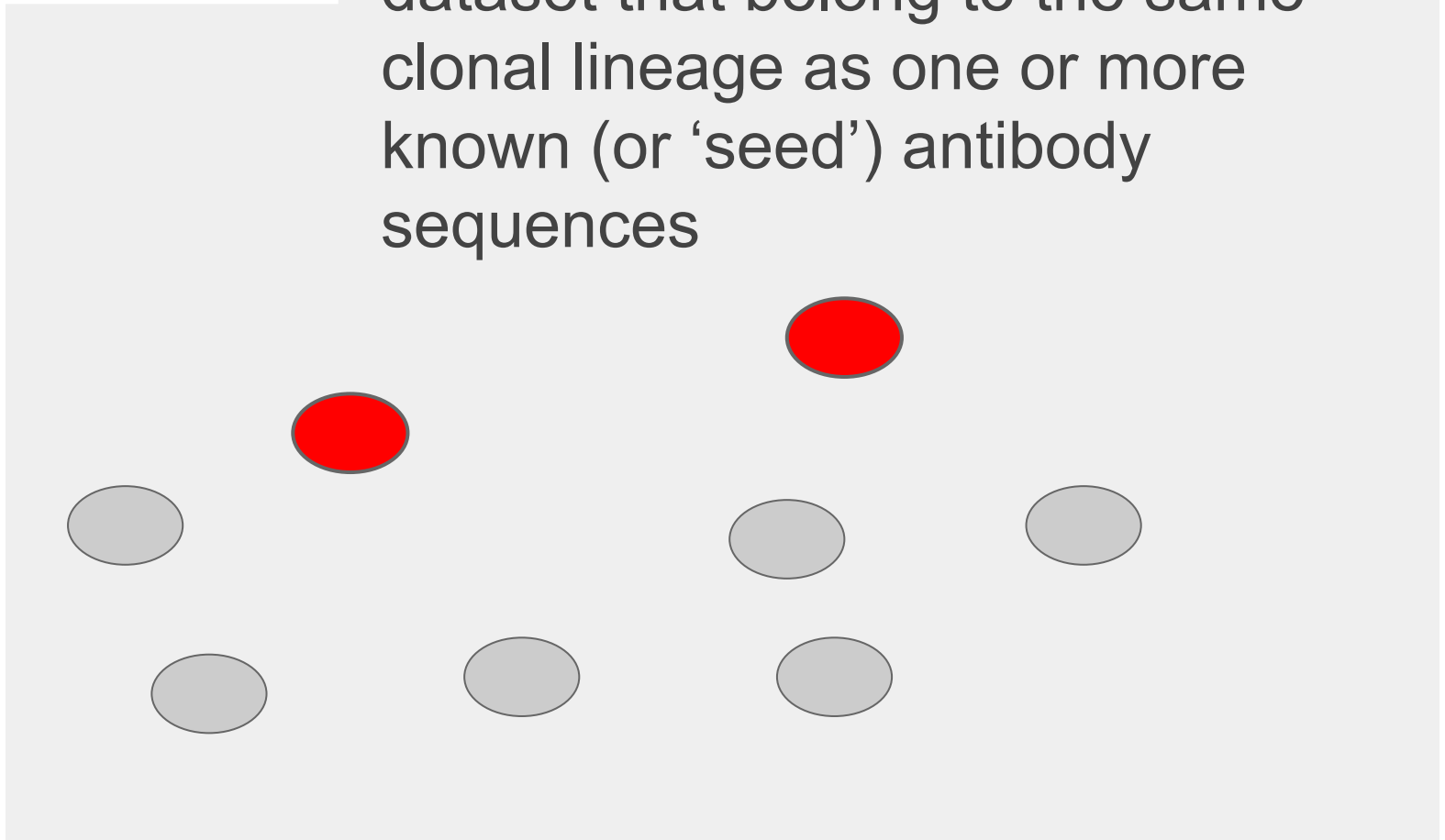
Seeded clonal lineage assignment

Identifying sequences from a dataset that belong to the same clonal lineage as one or more known (or 'seed') antibody sequences



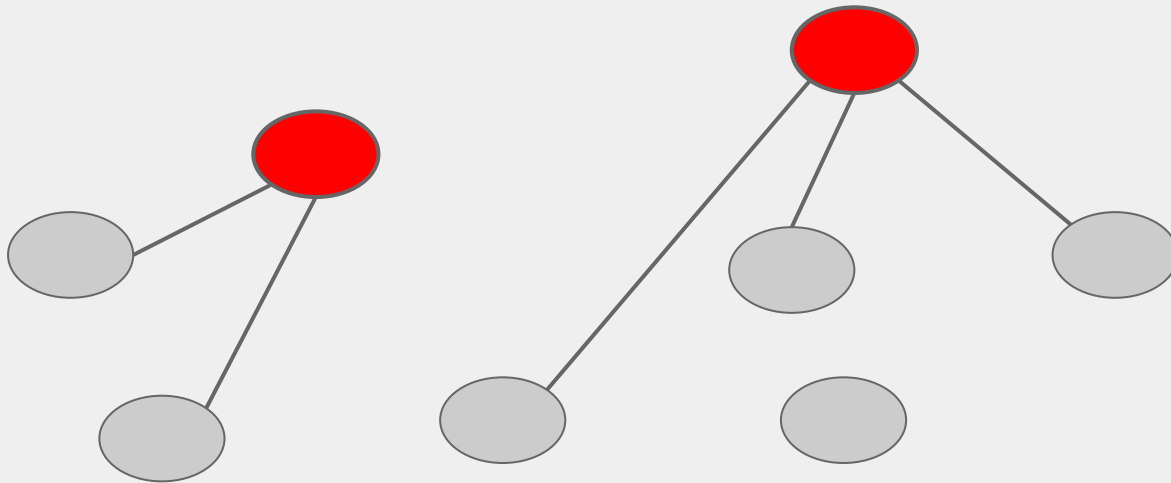
Seeded clonal lineage assignment

Identifying sequences from a dataset that belong to the same clonal lineage as one or more known (or 'seed') antibody sequences



Seeded clonal lineage assignment

Identifying sequences from a dataset that belong to the same clonal lineage as one or more known (or 'seed') antibody sequences



Seeded clonal lineage assignment

Identifying sequences from a dataset that belong to the same clonal lineage as one or more known (or 'seed') antibody sequences

Outputs really few lineages, poorly scalable.

2. B-cells

Again we have two problems:

- ◉ Decomposition on clonal lineages
 - seeded
 - unseeded

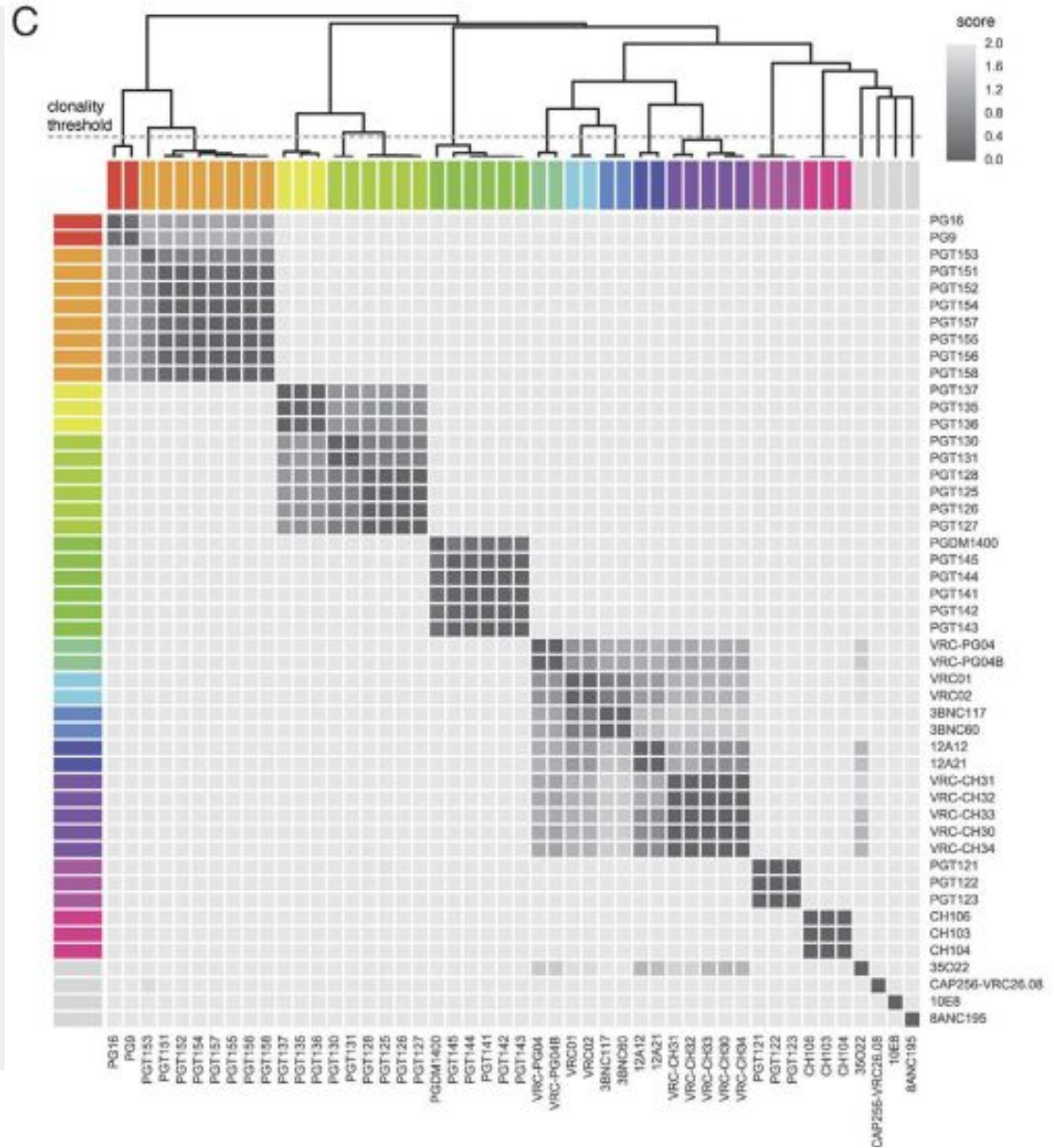
Clonify

Unseeded clonal lineage assignment

- Calculate pairwise affinity matrix
 - normalized CDR3 Levenshtein distance
 - V and J gene use
 - shared SHMs count
- Run hierarchical clustering
- profit

Unseeded clonal lineage assignment

Clonify



Unseeded clonal lineage assignment

PARTIS

- HMM for VDJ-recombination
 - has one state for each position in every V, D, and J gene, and
 - a state for each of the joining N-regions for heavy chain sequences
- Agglomerative clustering: merge x and y such that they maximize likelihood of them coming from the same rearrangement
- Likelihood is given by the forward algorithm

Questions?

Questions?