Robust sub-population discovery using self-pruning decision trees

Assaf Magen
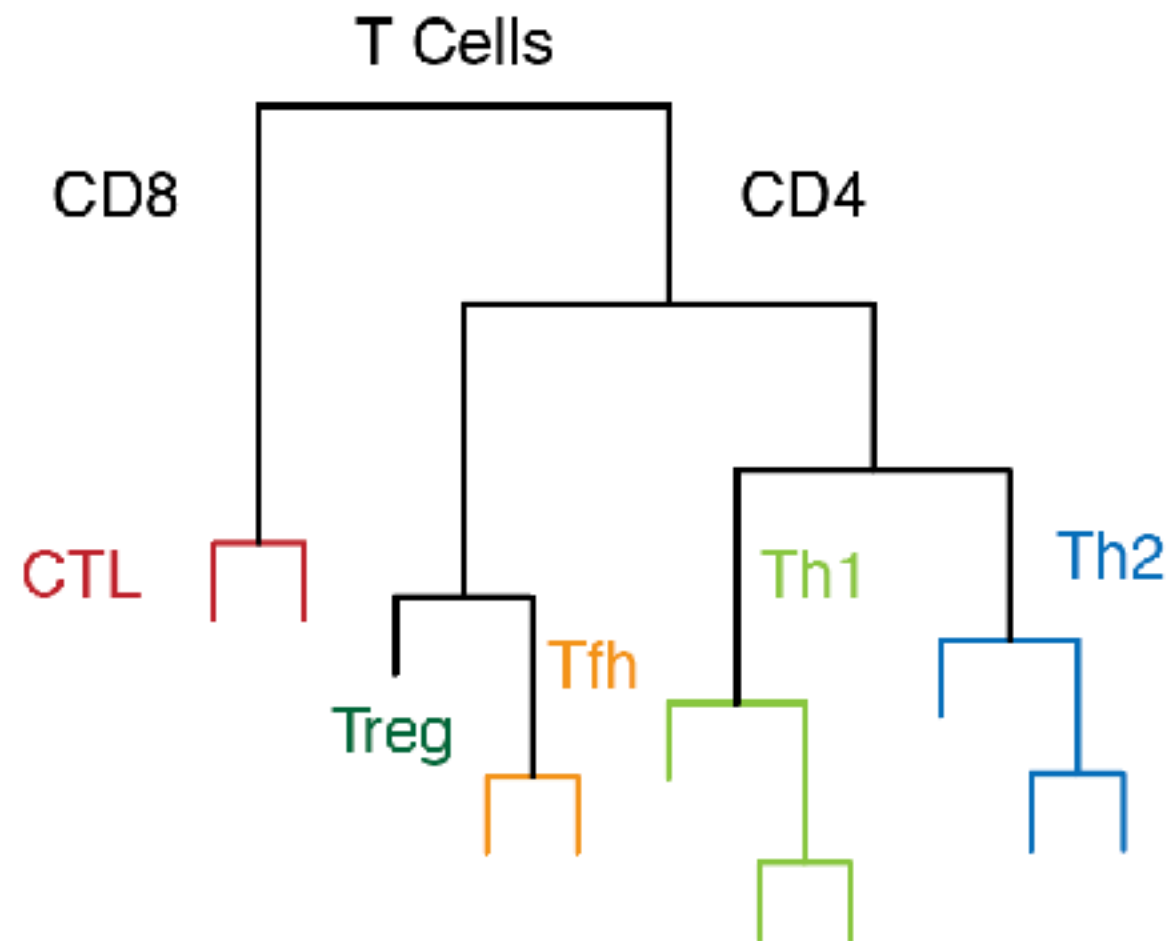
Mamie Wang, Billy Kim

Claire Malley, Amulya Shastry

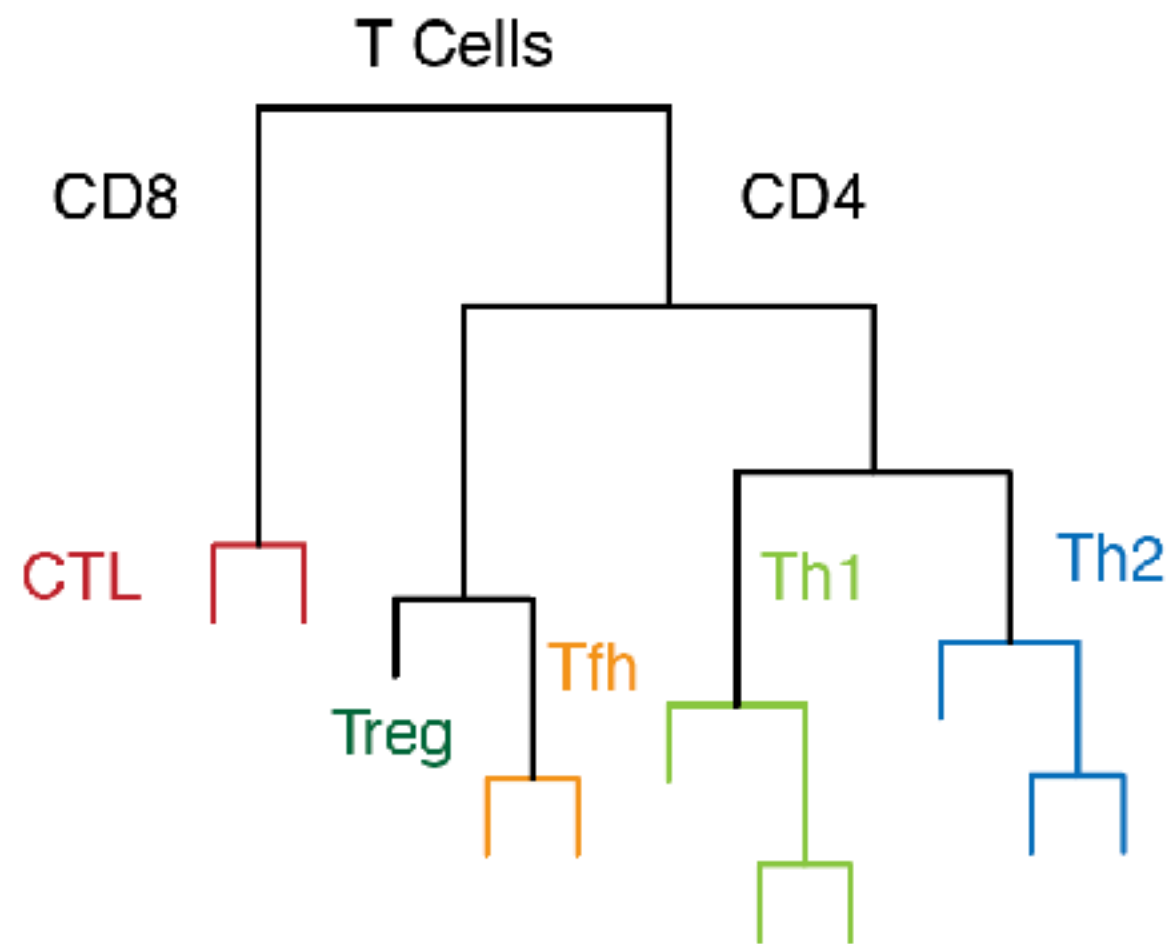Christopher Rhodes, Jonathan Badger

Github: ReSET (Robust Subpopulation dEcision Trees)
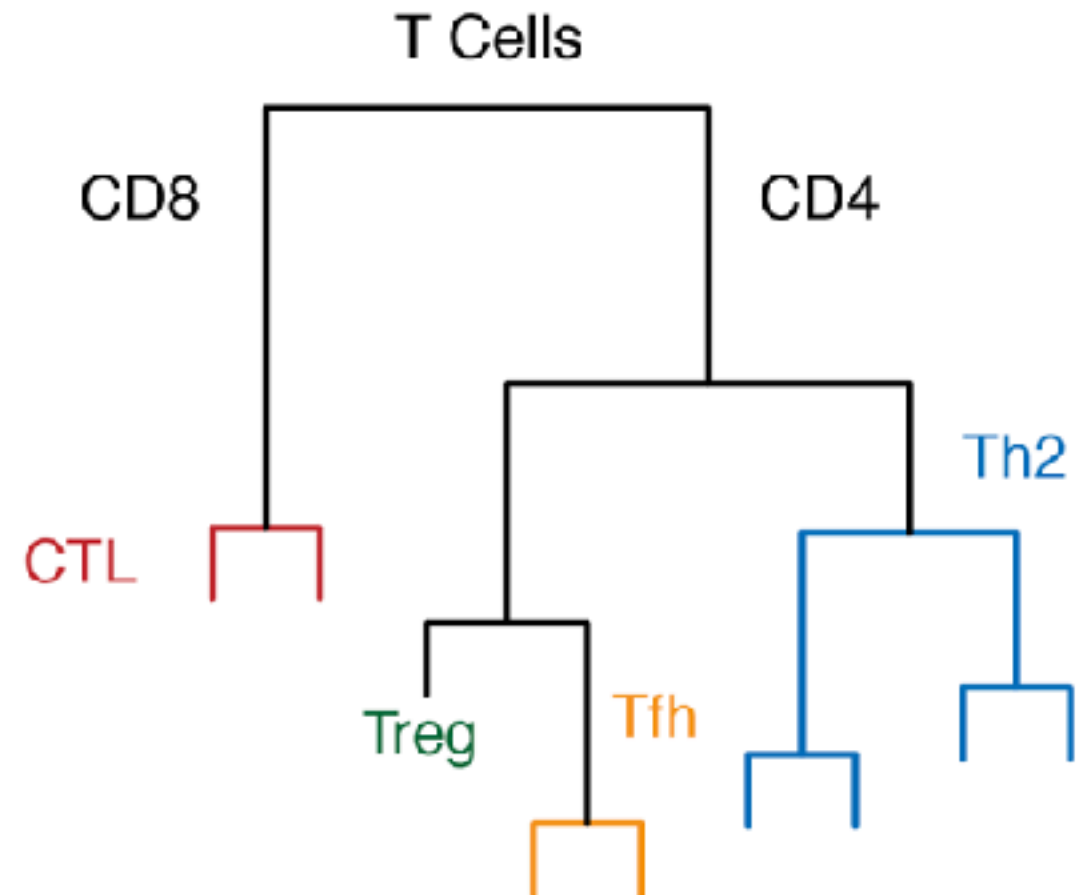
Population structure discovery
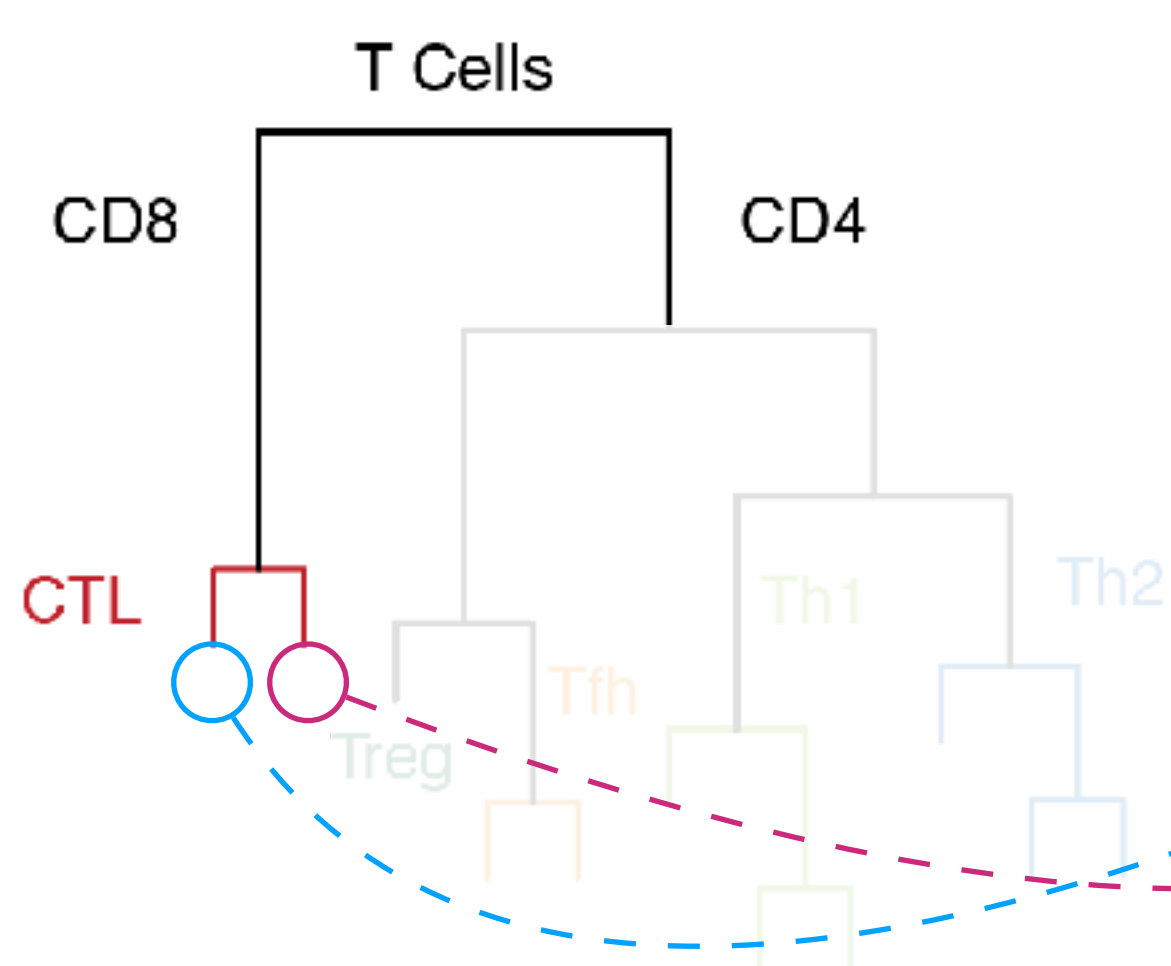
Comparison across replicate experiments

Identify reproducible populations
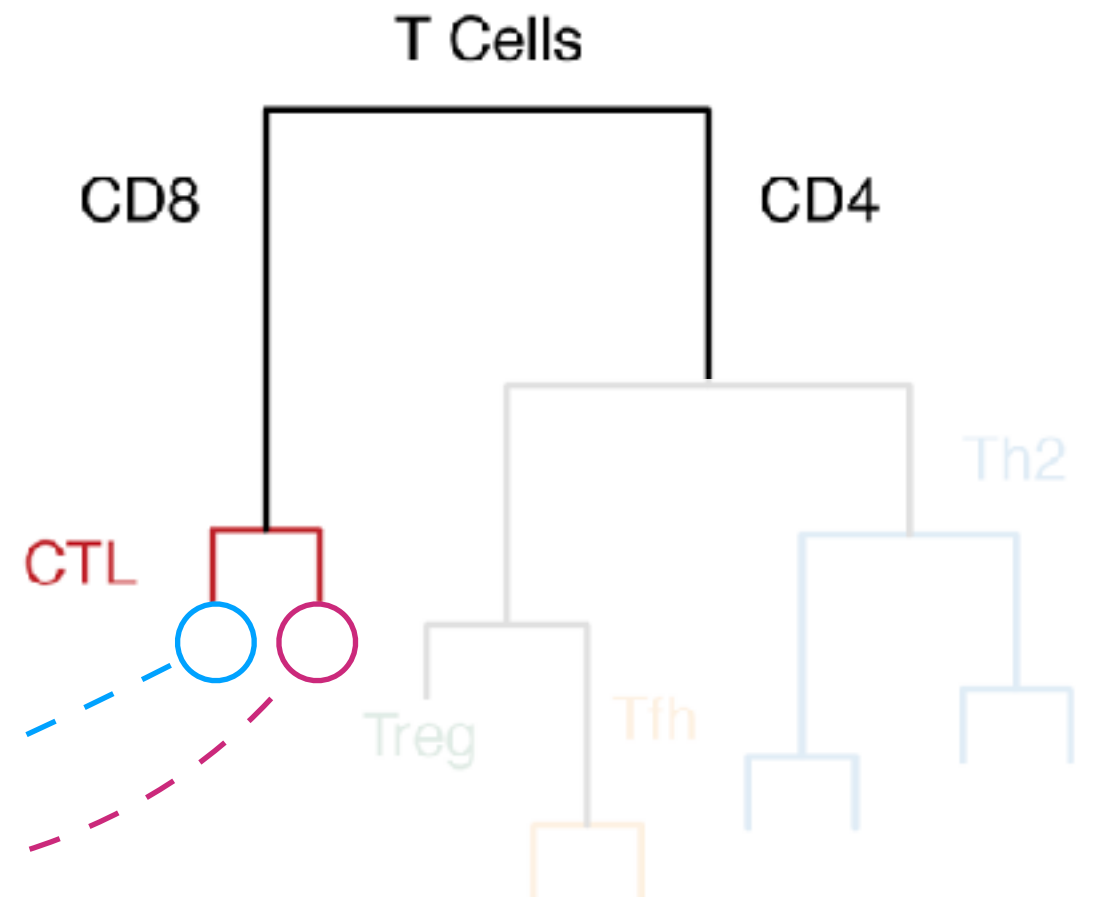
Identify irreproducible populations
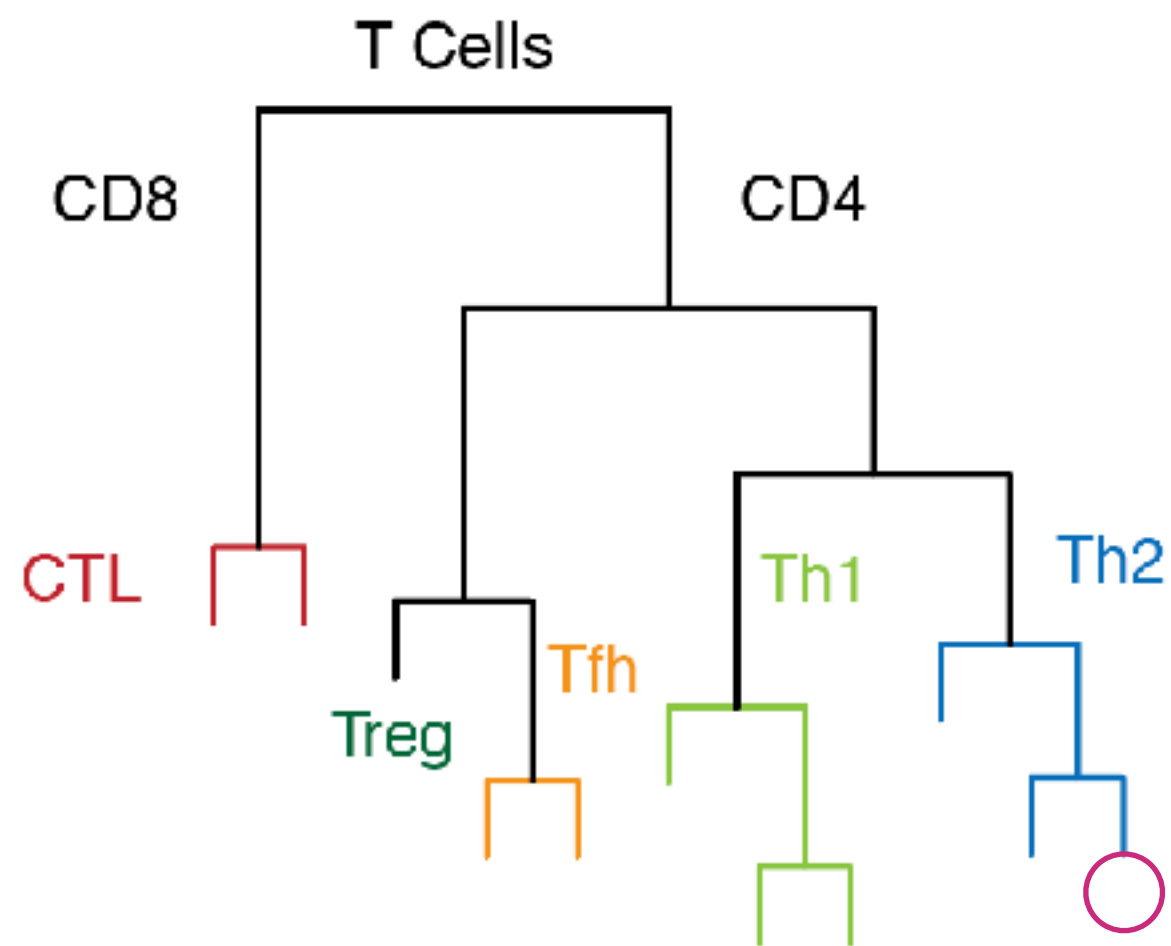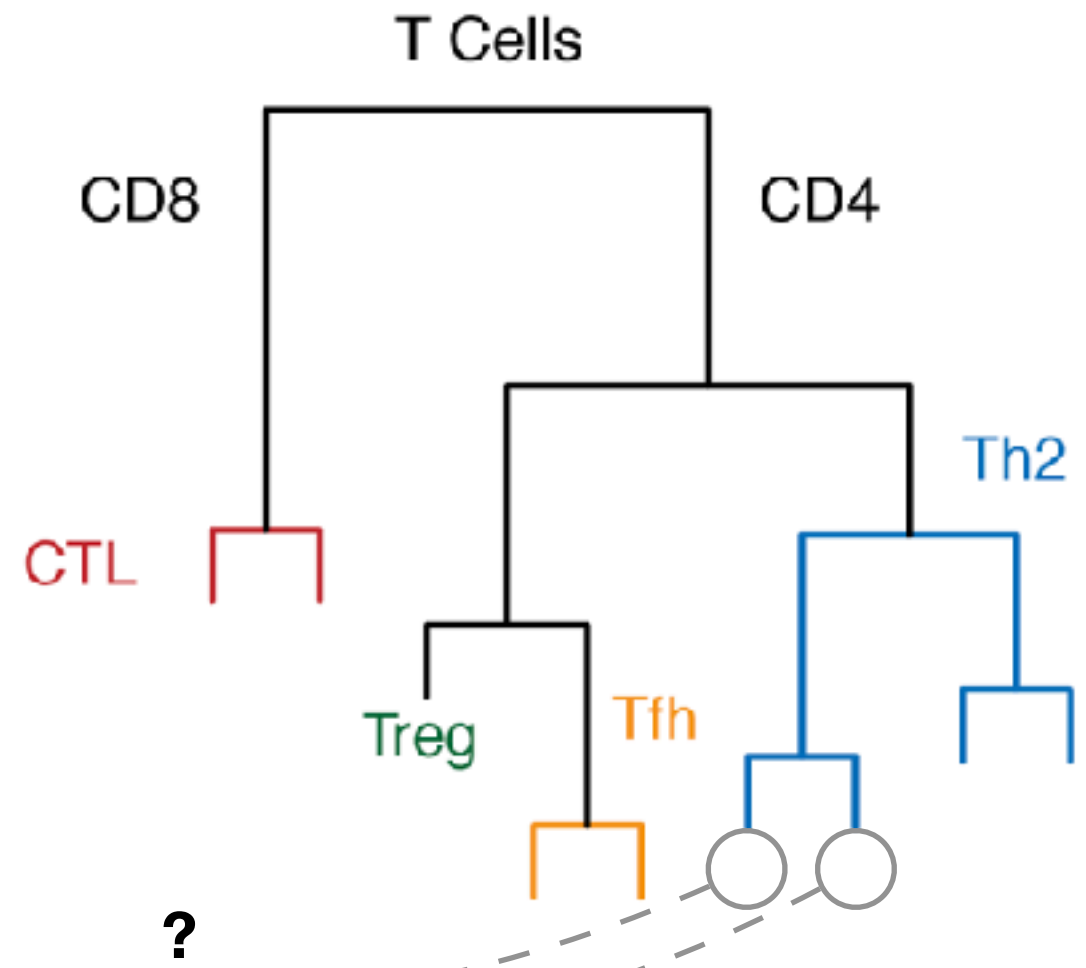
Consider matches of lower granularities

Experiment I

Experiment II

Consider matches of lower granularities

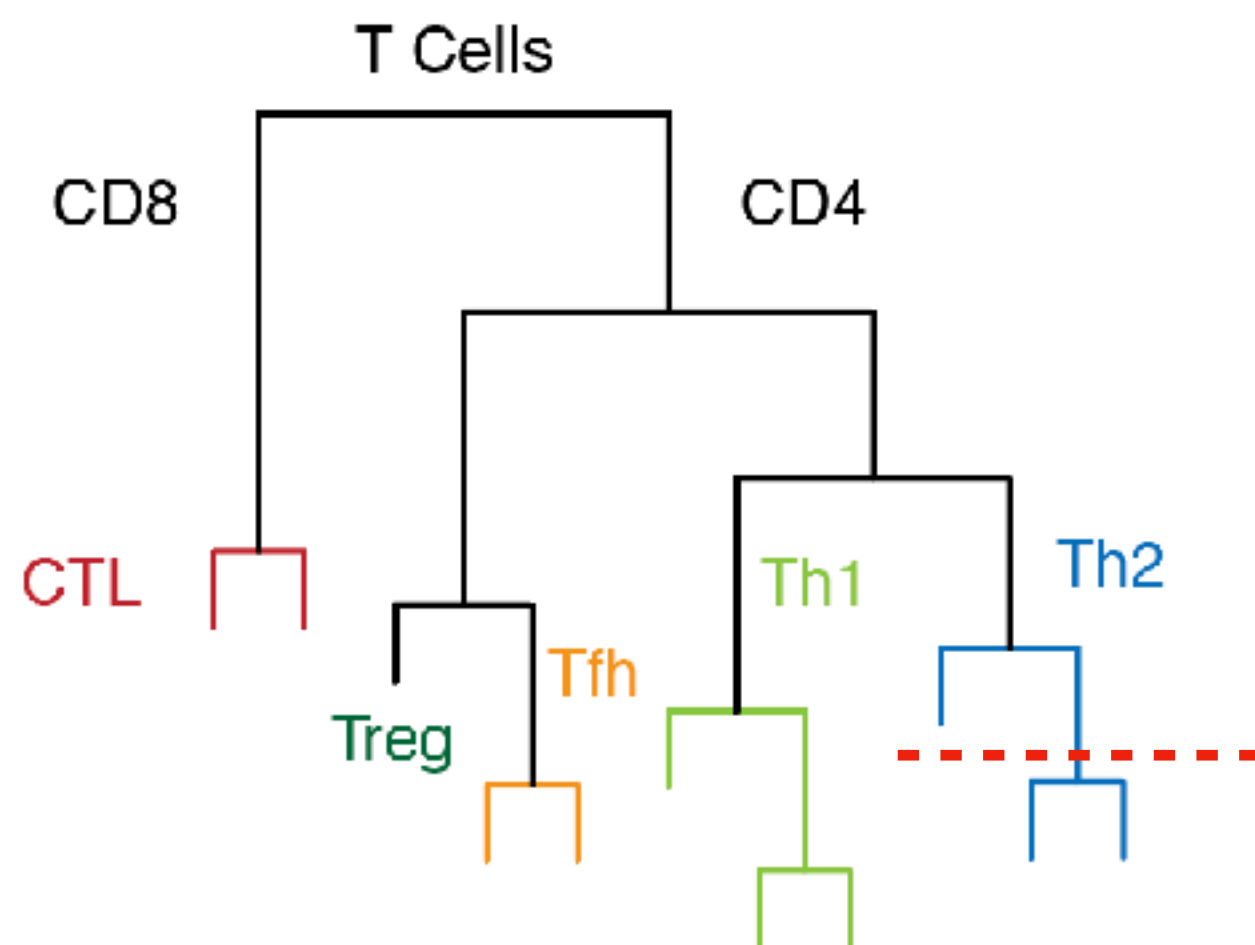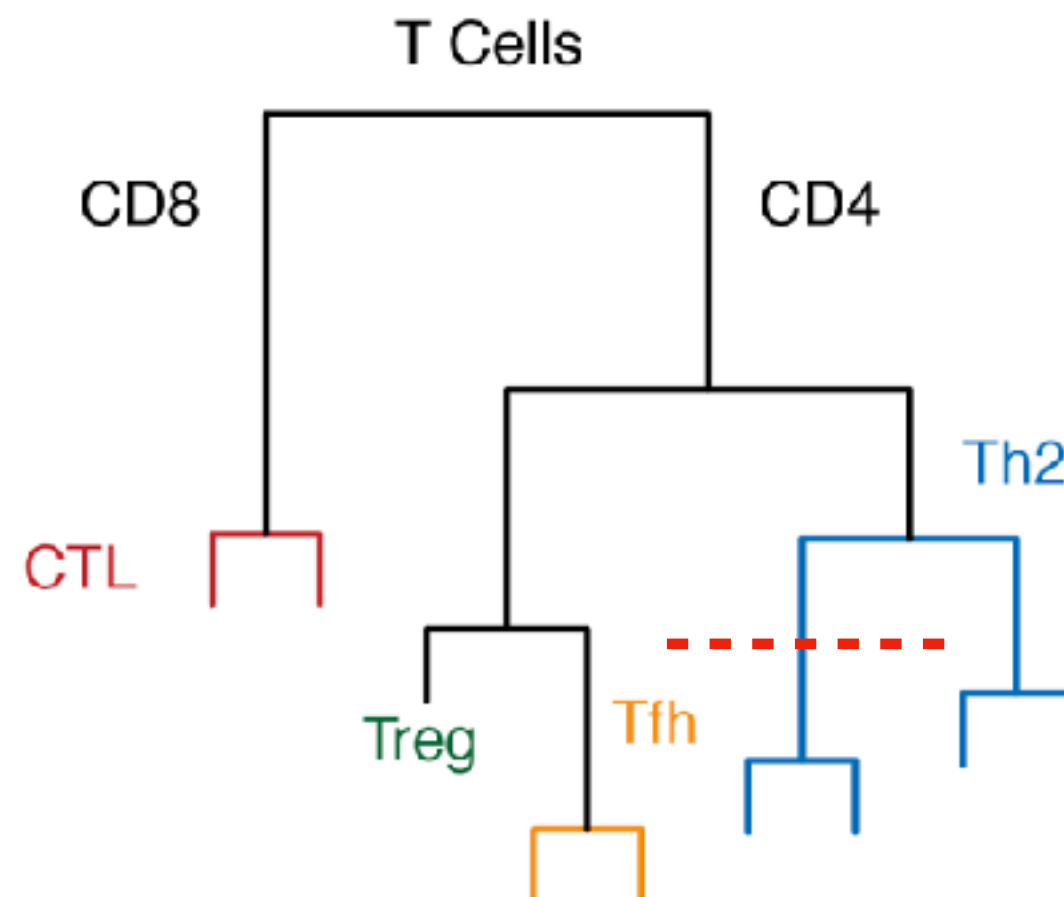# Consider matches of lower granularities

Consider matches of lower granularities

Experiment I
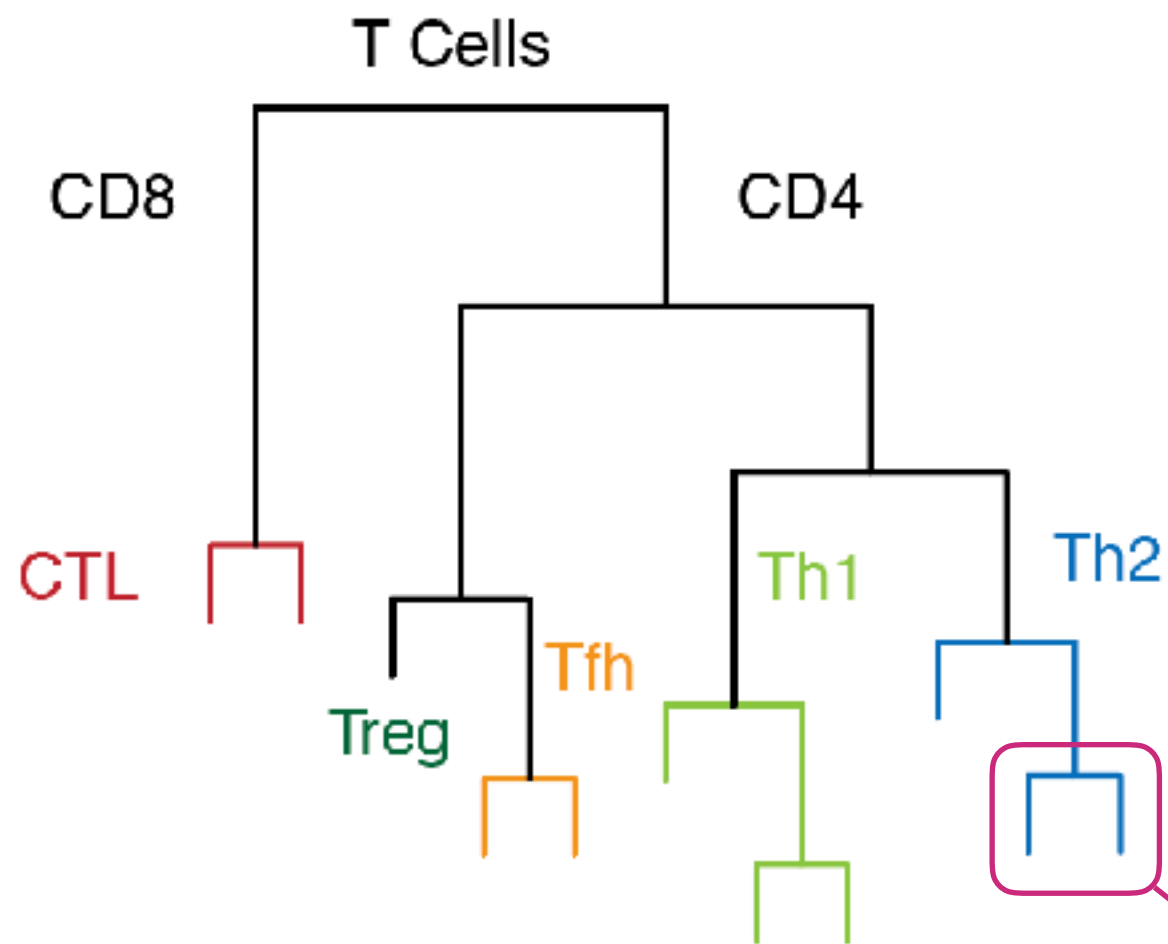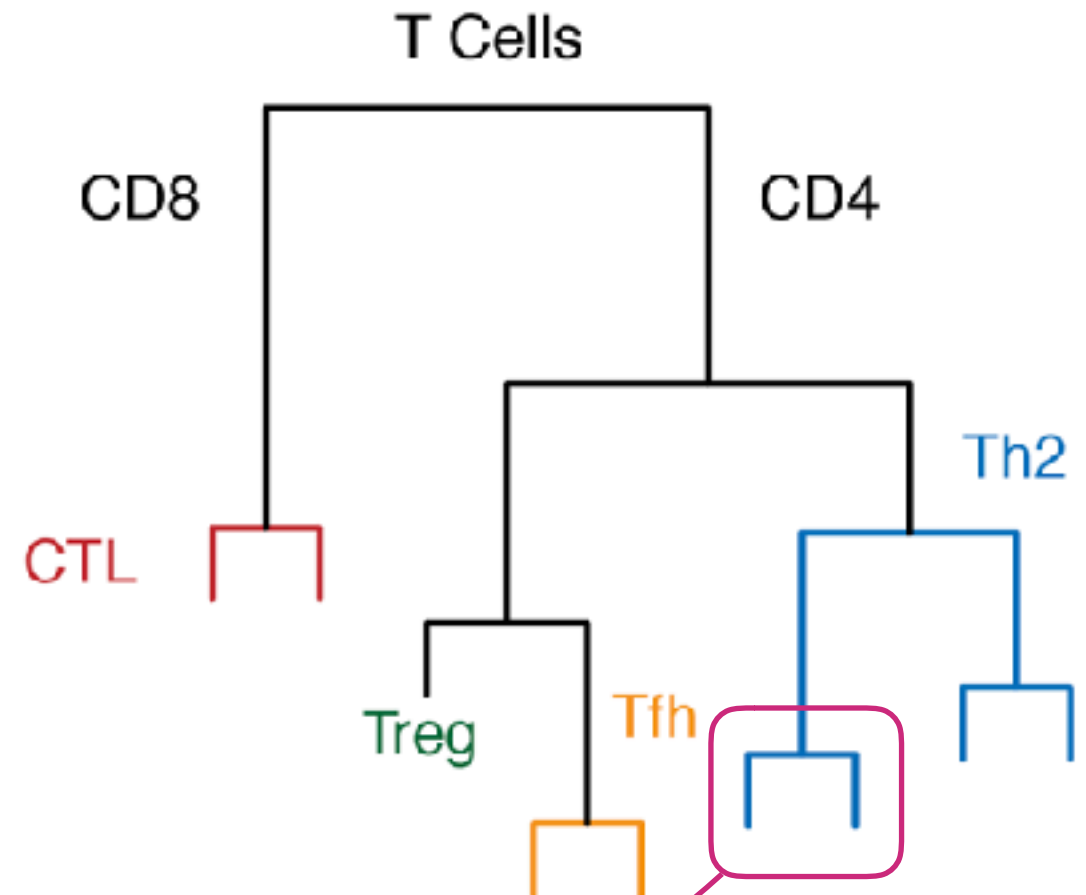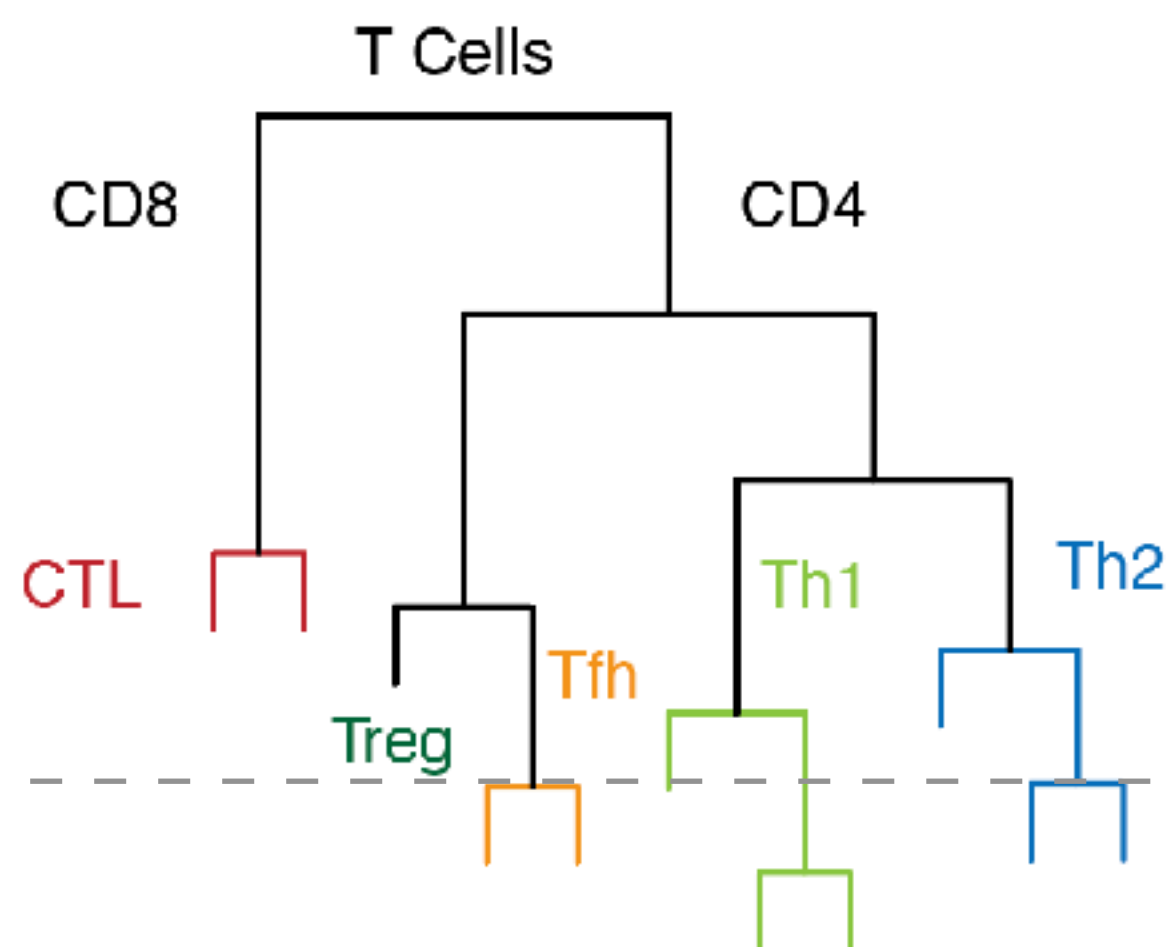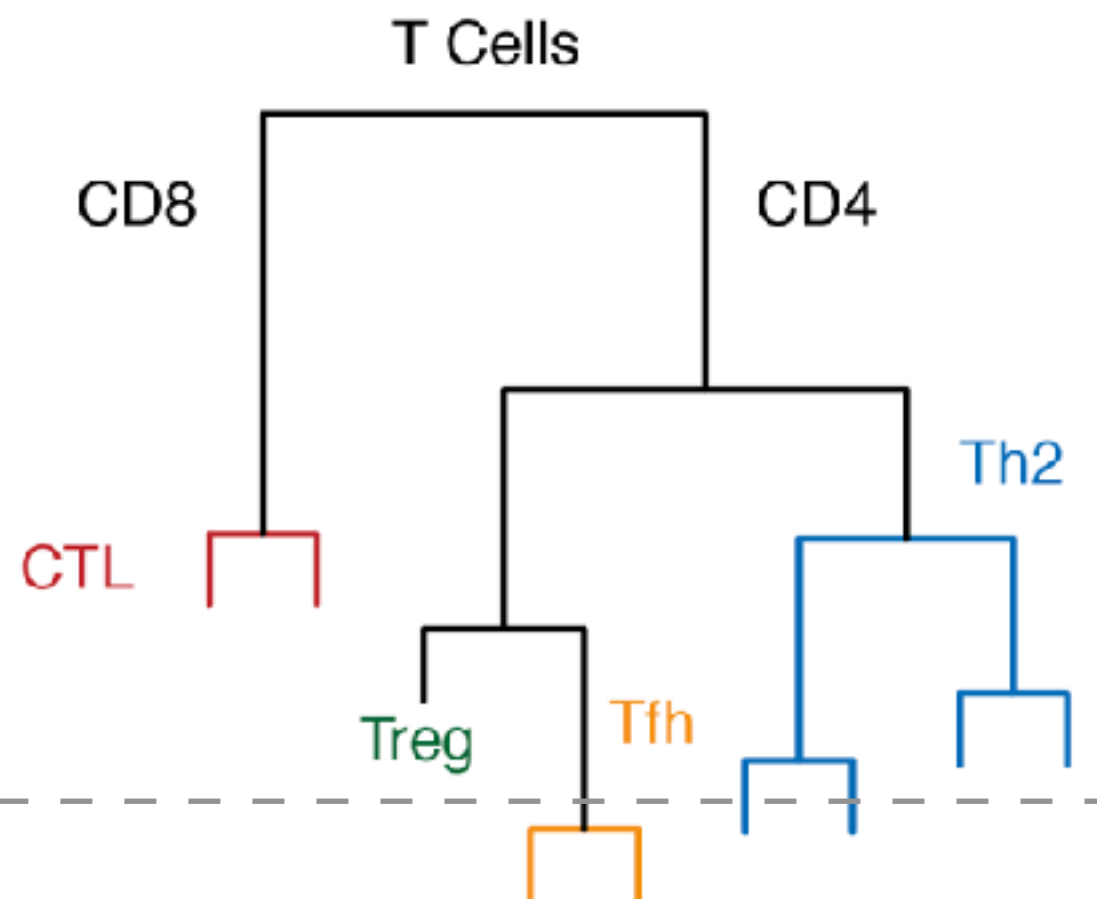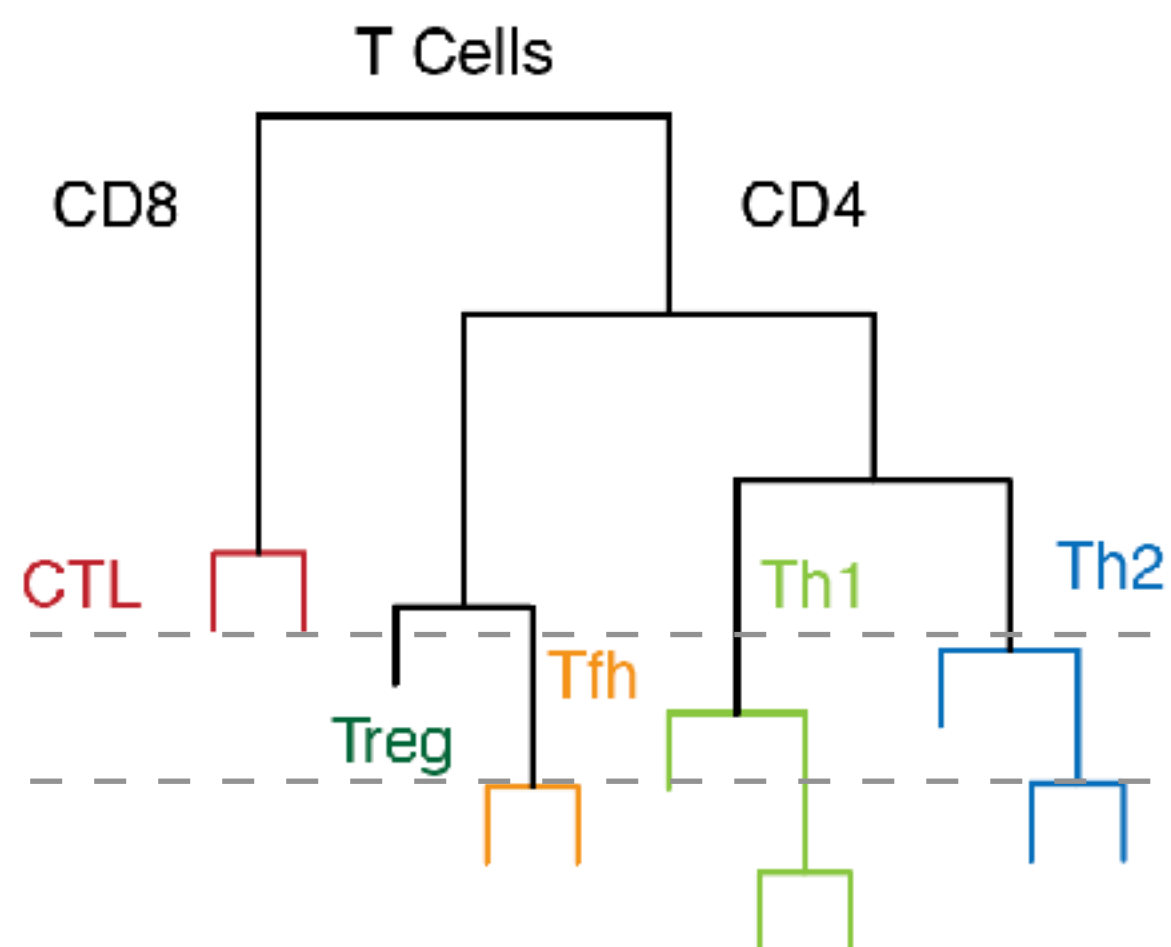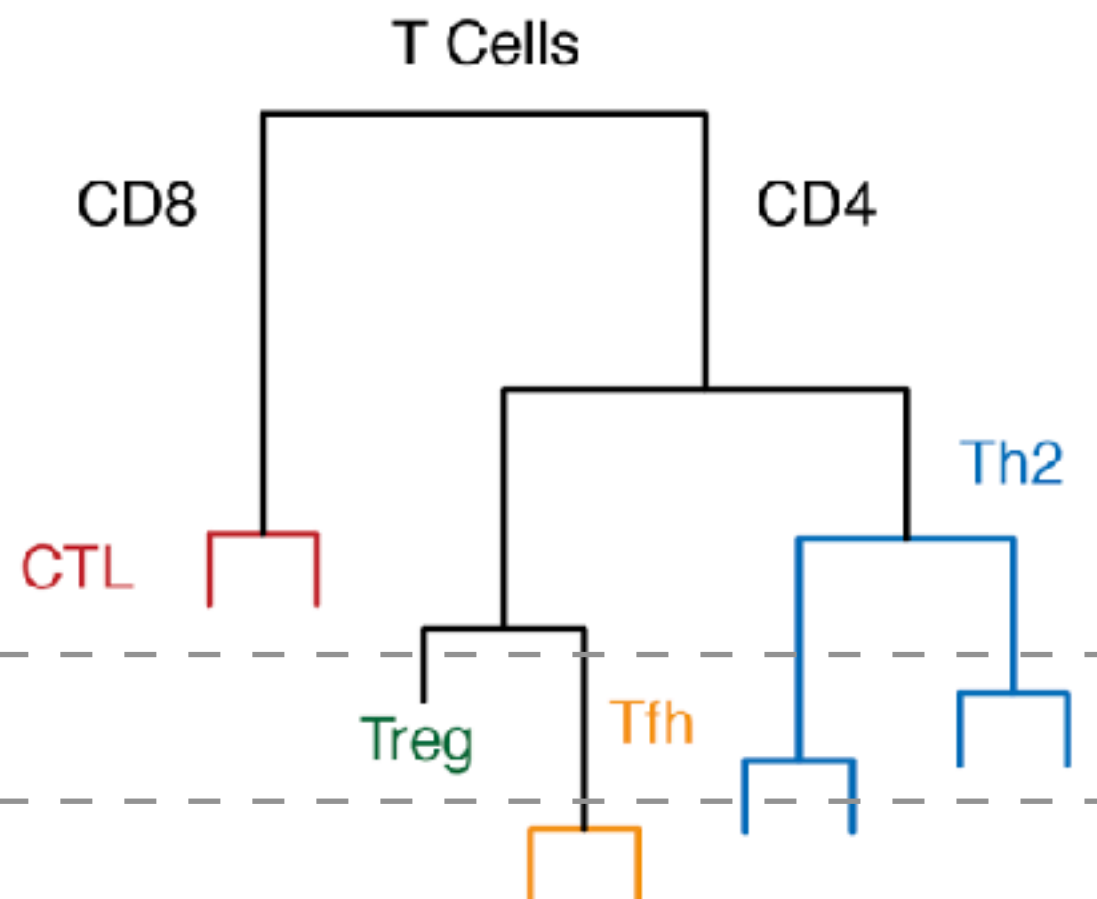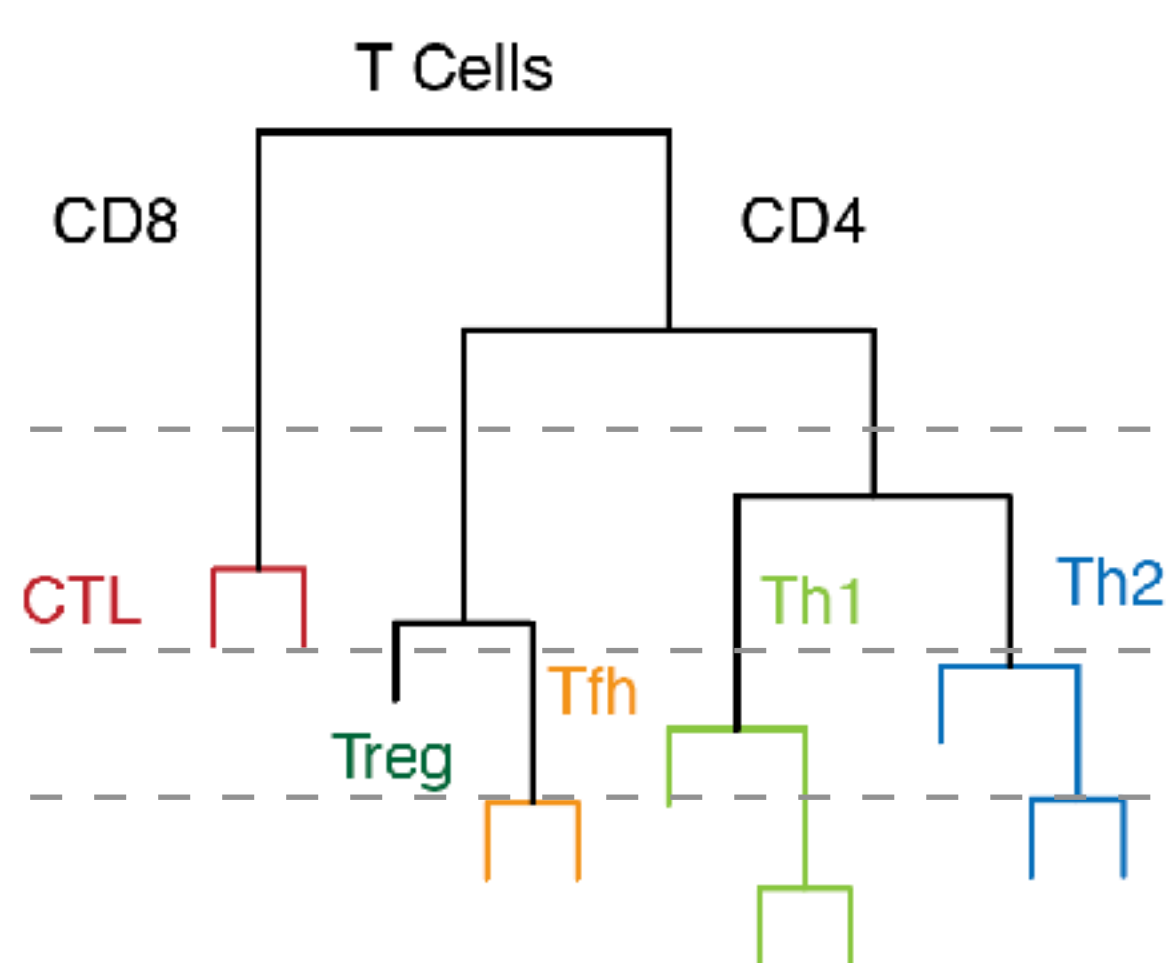
Experiment II

Consider matches of lower granularities

Experiment I
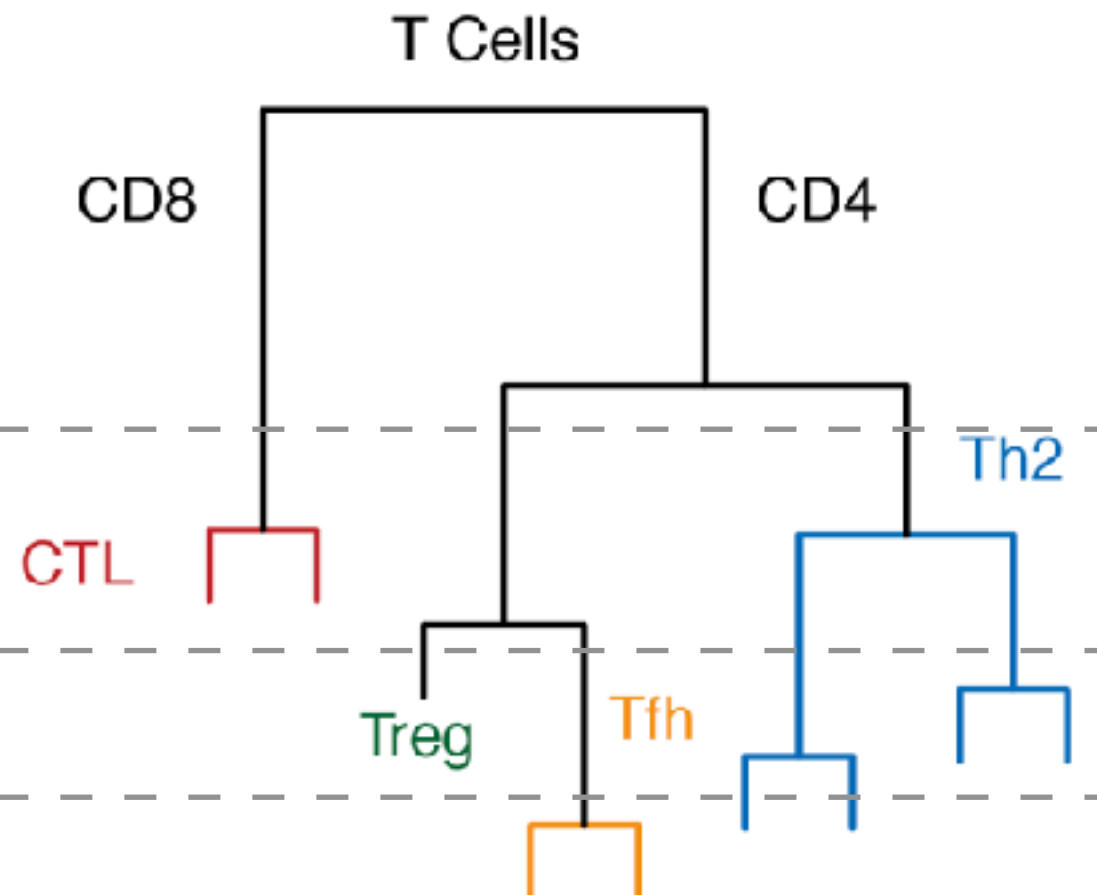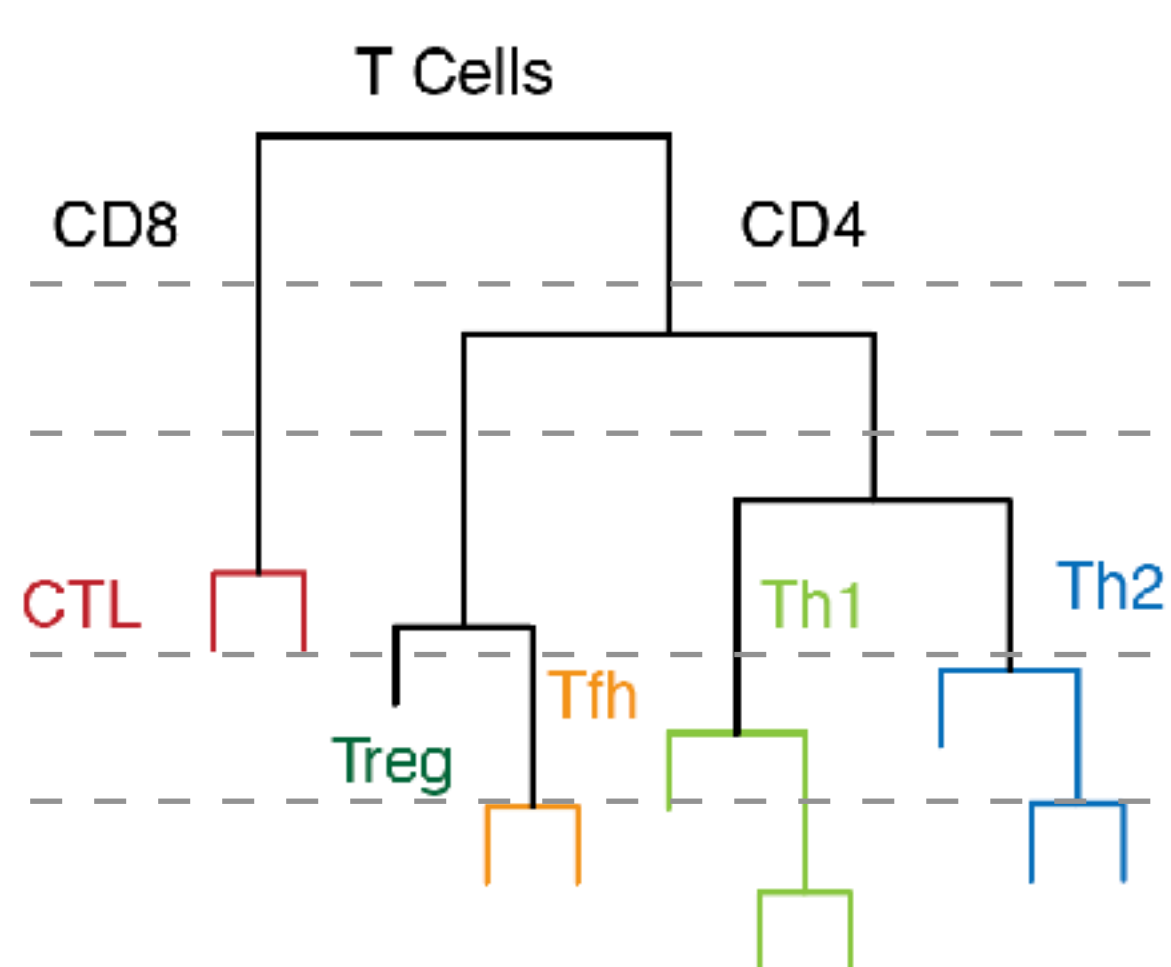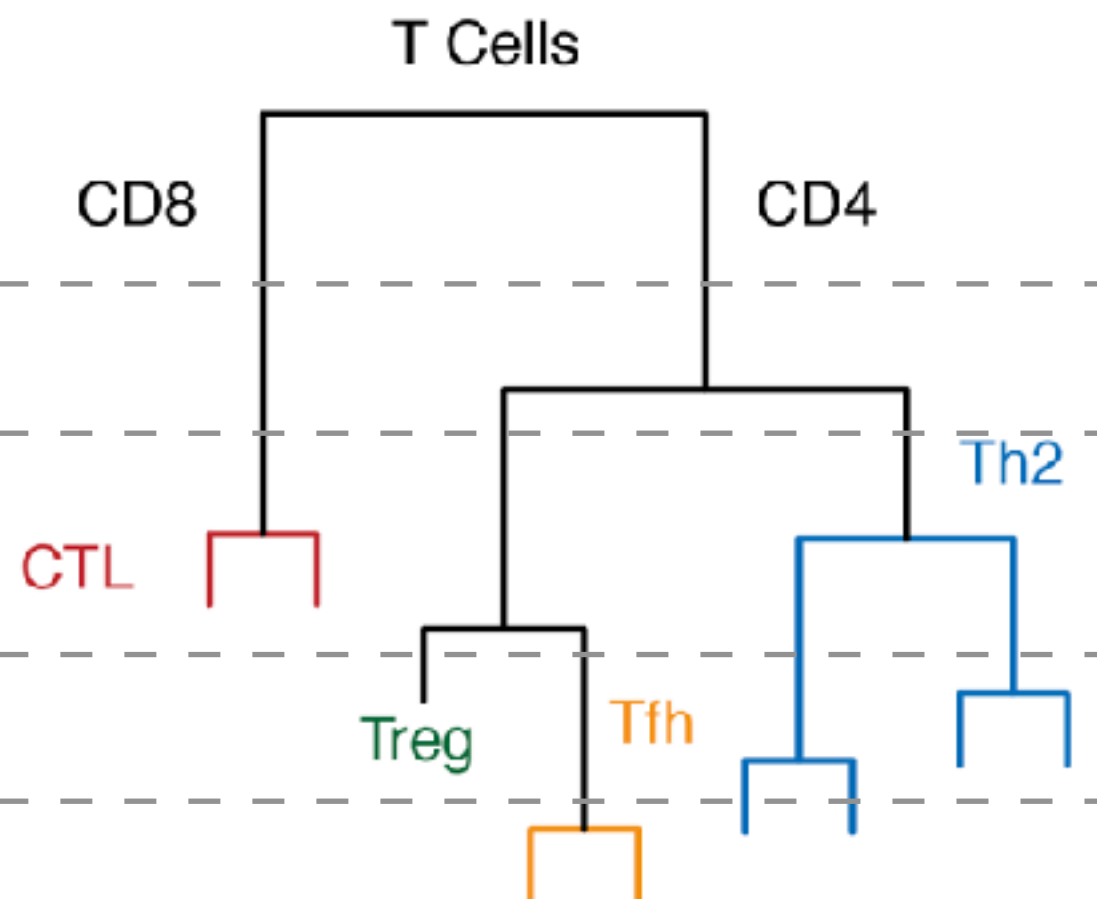

Experiment II

Consider matches of lower granularities

Experiment I

T Cells

CD8   CD4

CTL

Tfh

Treg

Th1

Th2

Experiment II

T Cells

CD8   CD4

Th2

CTL

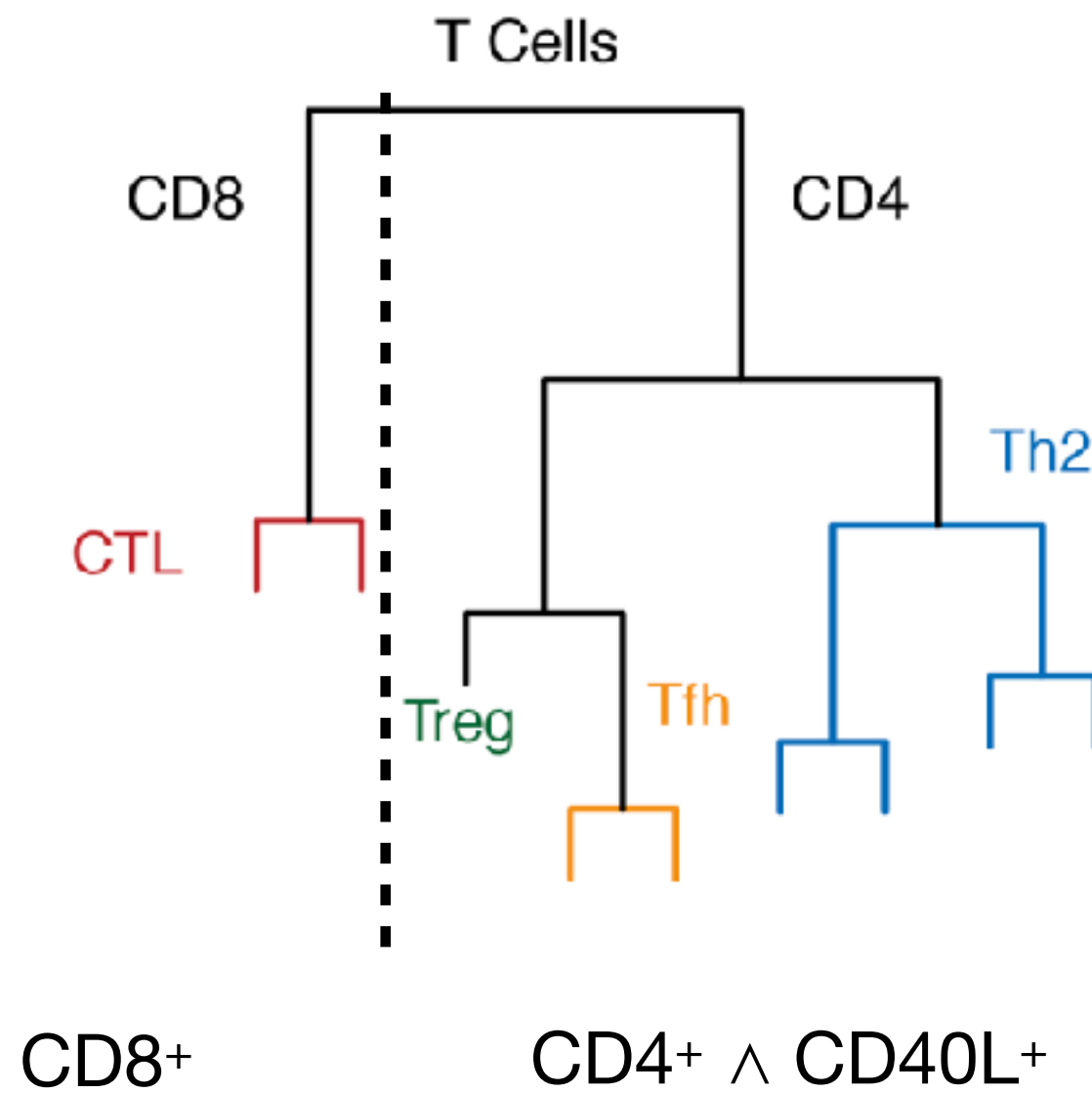Treg   Tfh
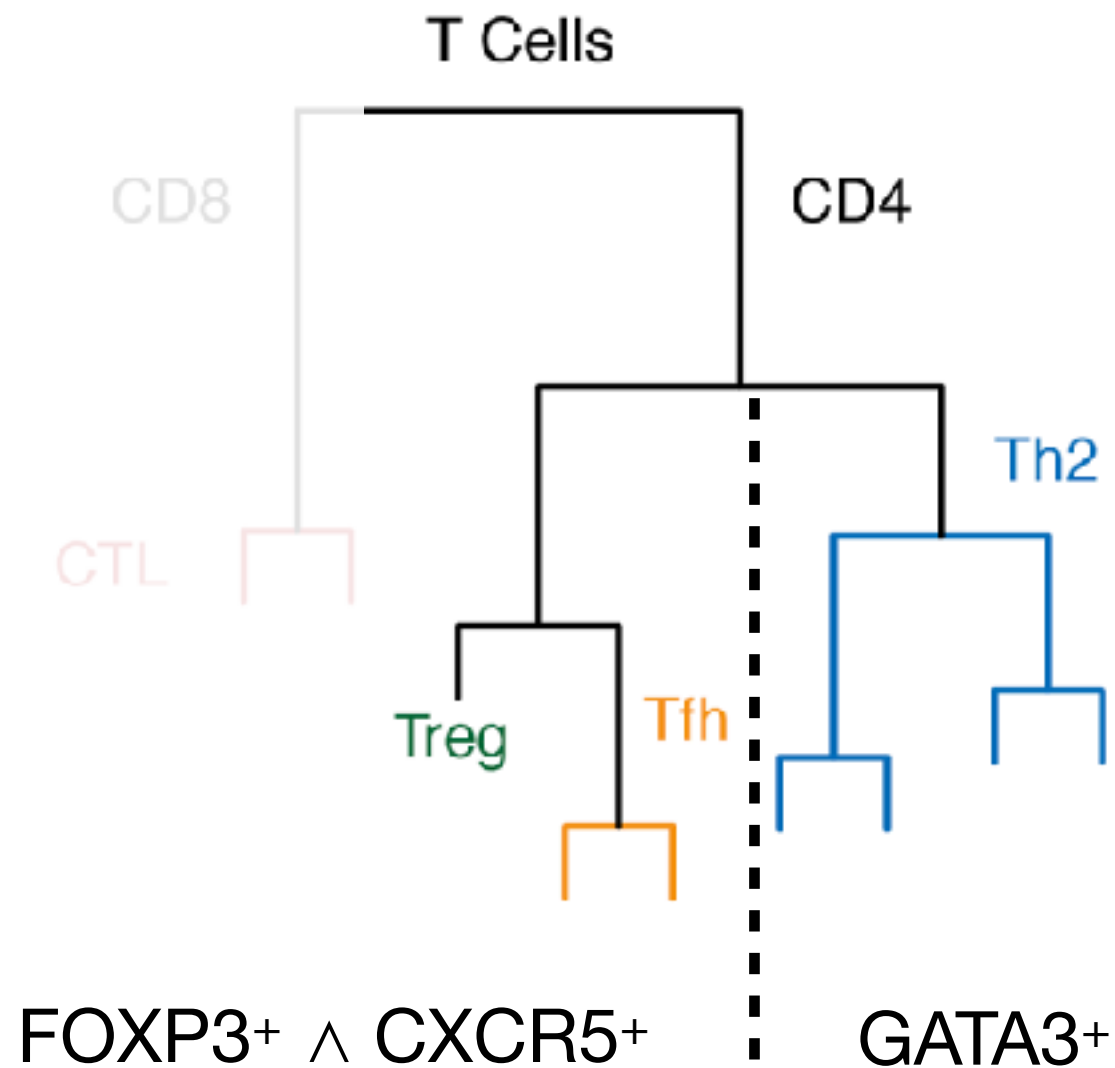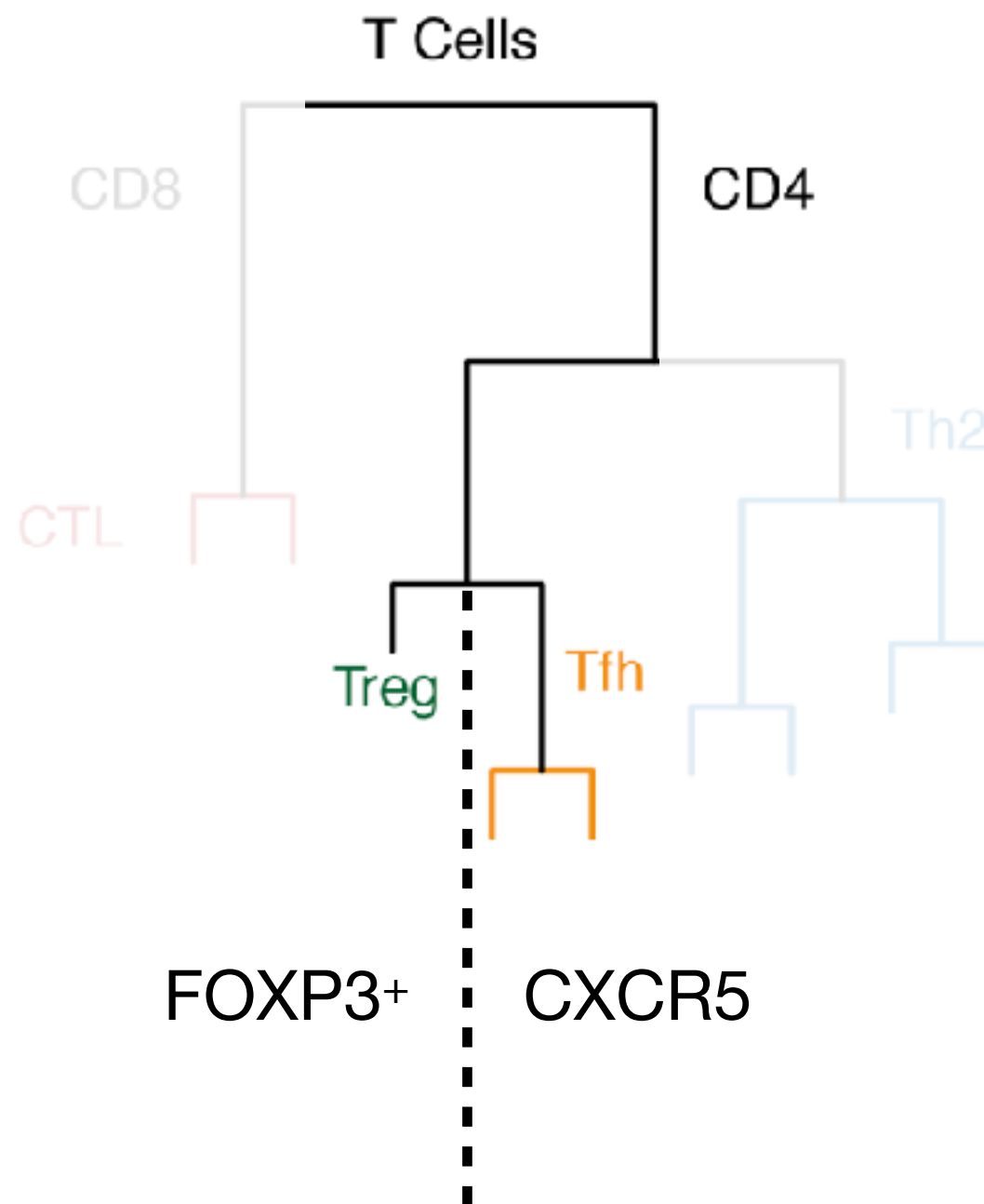
Identify cluster defining genes

Identify cluster defining genes

Identify cluster defining genes

Conclusions

- Clustering hierarchy provides improved functionality

- Optimal clustering granularity can be defined in an unbiased manner

- ML approached may improve marker gene identification

Future Directions

- Improve clustering and hierarchy deduction

- Identify ideal clustering granularity computationally

- Refine cluster marker genes discovery methodology

Acknowledgments

Hackathon lead

Allissa Dillman

Team members

Team lead

Assaf Magen

Cluster reproducibility

Mamie Wang, Billy Kim, Christopher Rhodes

ML strategies for cluster markers

Claire Malley, Amulya Shastry, Jonathan Badger