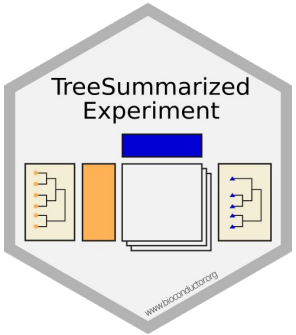
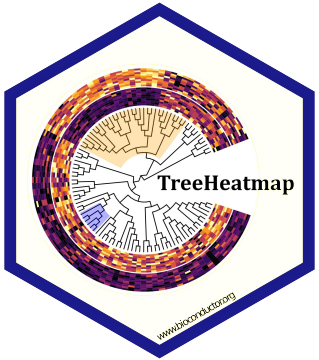
Universität Zürich[UZH]

Swiss Institute of Bioinformatics

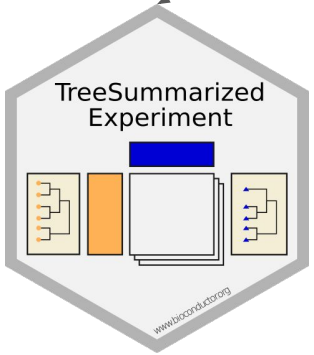
Tree-based Signal Aggregation

Ruizhu Huang

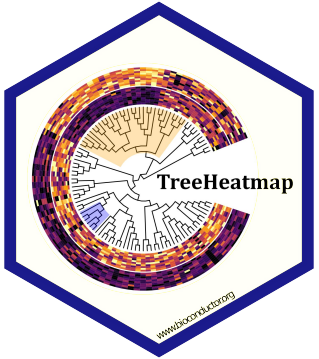




Introduction	Data container	Visualization	Algorithm

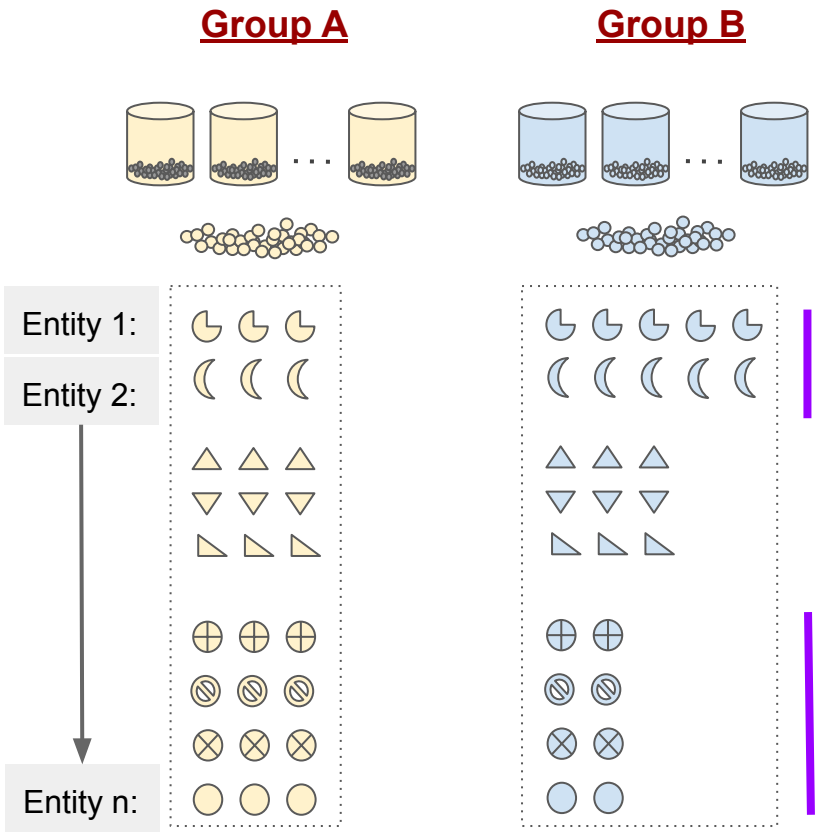


TreeSummarizedExperiment



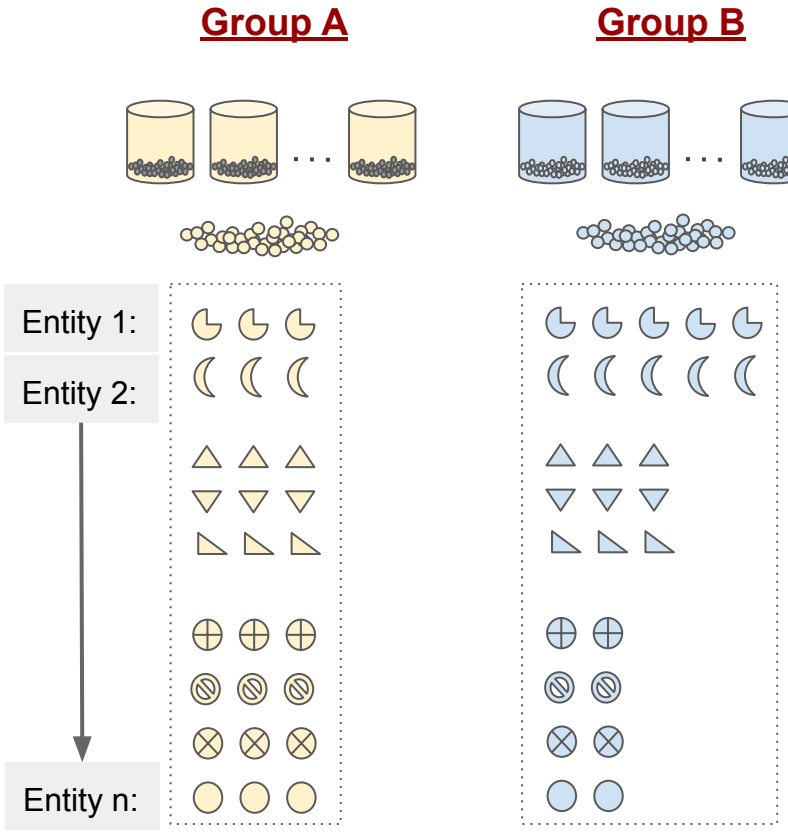
TreeHeatmap

treeclimbR



Samples from different conditions (groups)

Are some entities differentially abundant between groups?



Microbial data

Soil ⟷ Ocean

Oral ⟷ Skin

Day 1 ⟷ Day 2

Microbial species

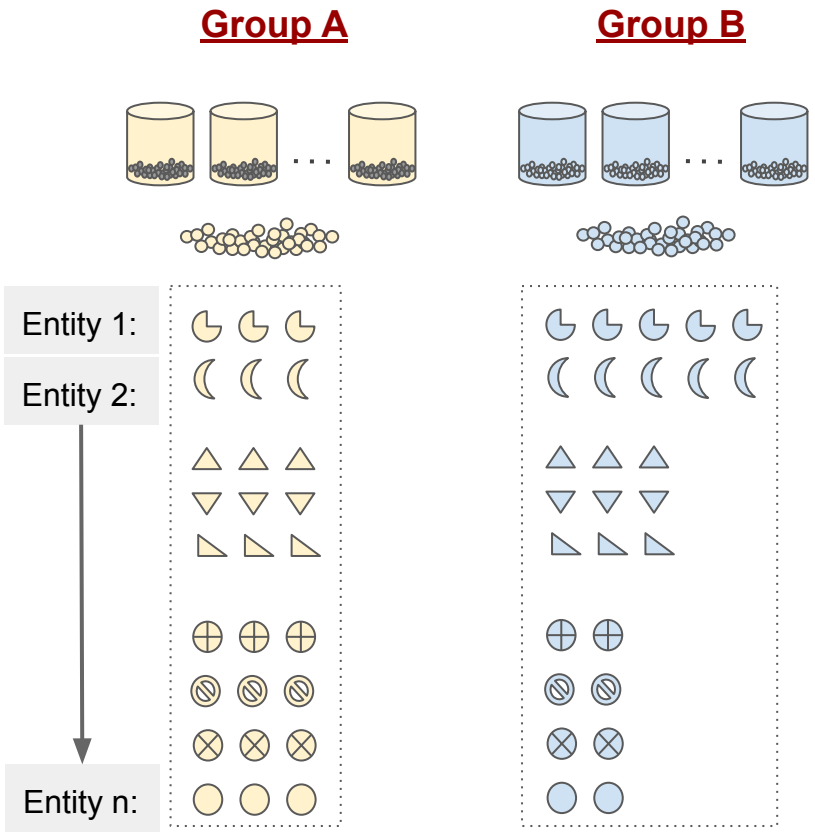
Differential abundance?

Single cell data

control ⟷ stimulated

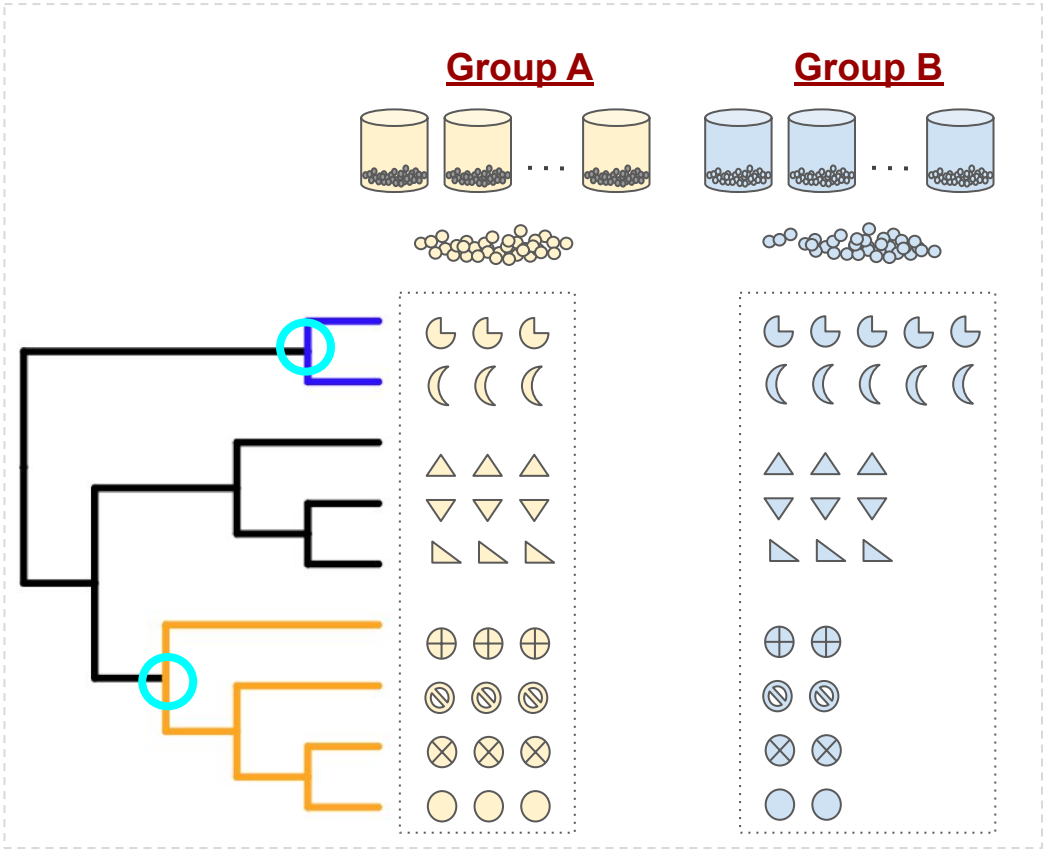
Cell types or genes

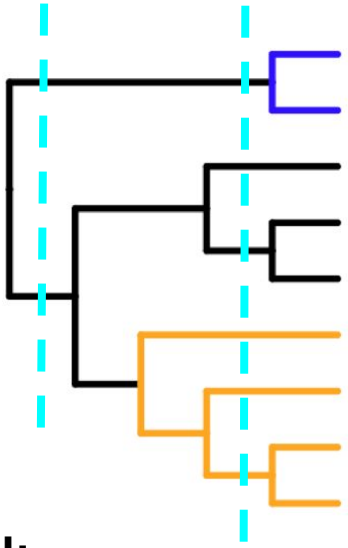
Differential abundance or states?



Issues:

1. Result is long
2. Difference is small

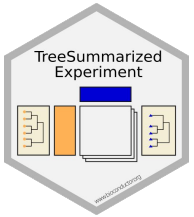


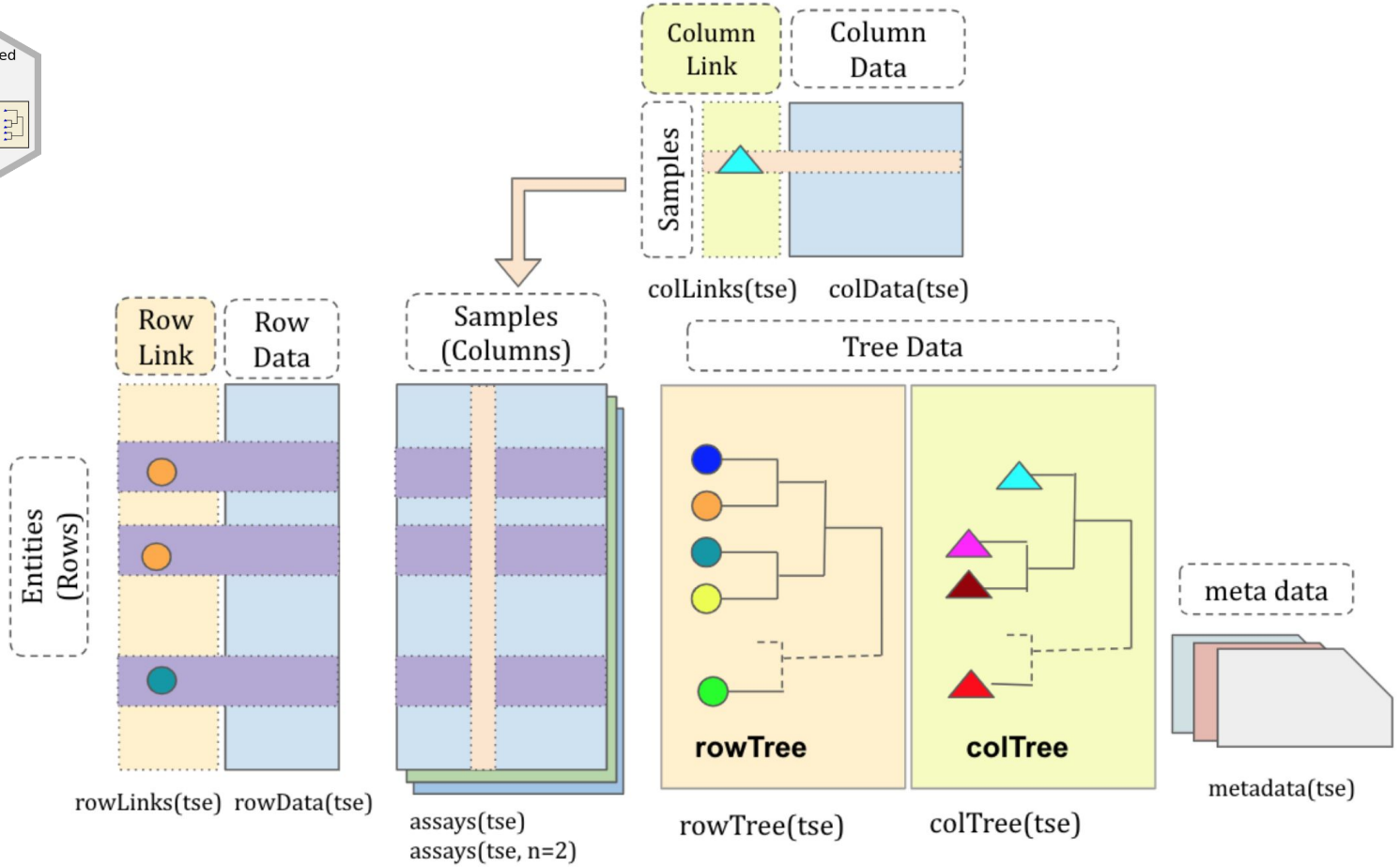


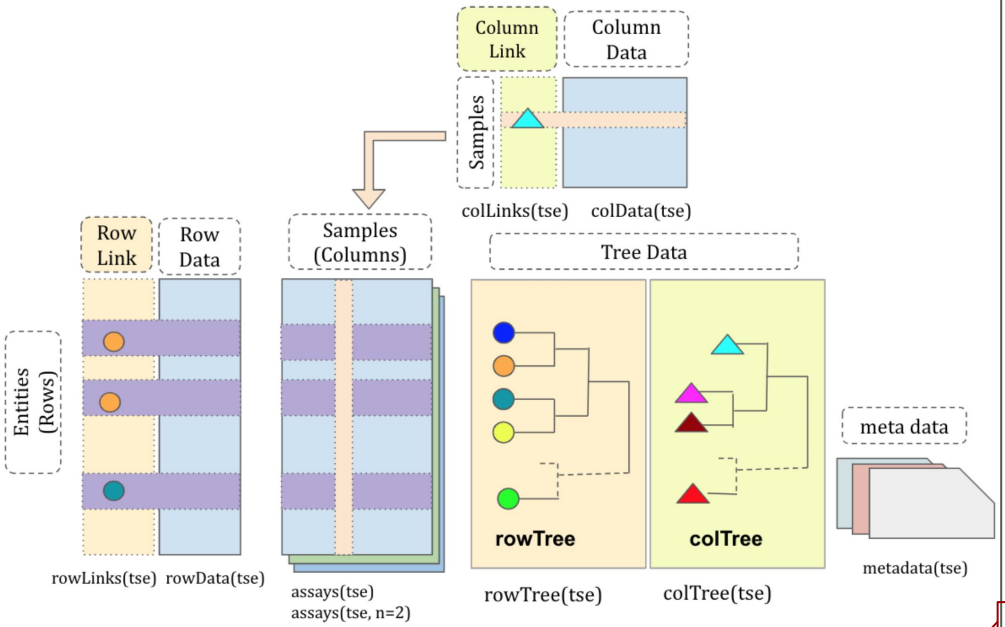
Goal:

Which level on the tree to interpret the difference?

An arbitrary level X


TreeSummarized
Experiment

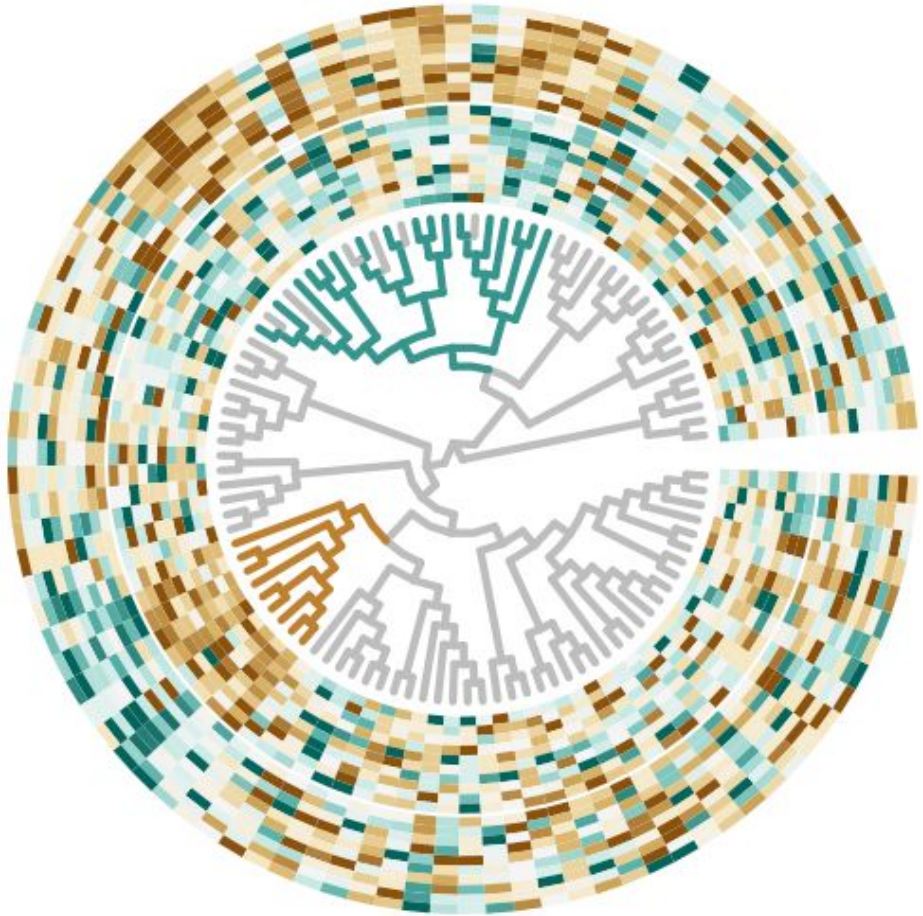

Column
Link
Column
Data
Samples
colLinks(tse)
colData(tse)
Row
Link
Row
Data
Samples
(Columns)
Tree Data
Entities
(Rows)
rowTree
colTree
meta data
rowLinks(tse)
rowData(tse)
assays(tse)
assays(tse, n=2)
rowTree(tse)
colTree(tse)
metadata(tse)

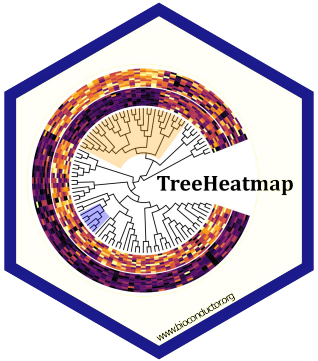


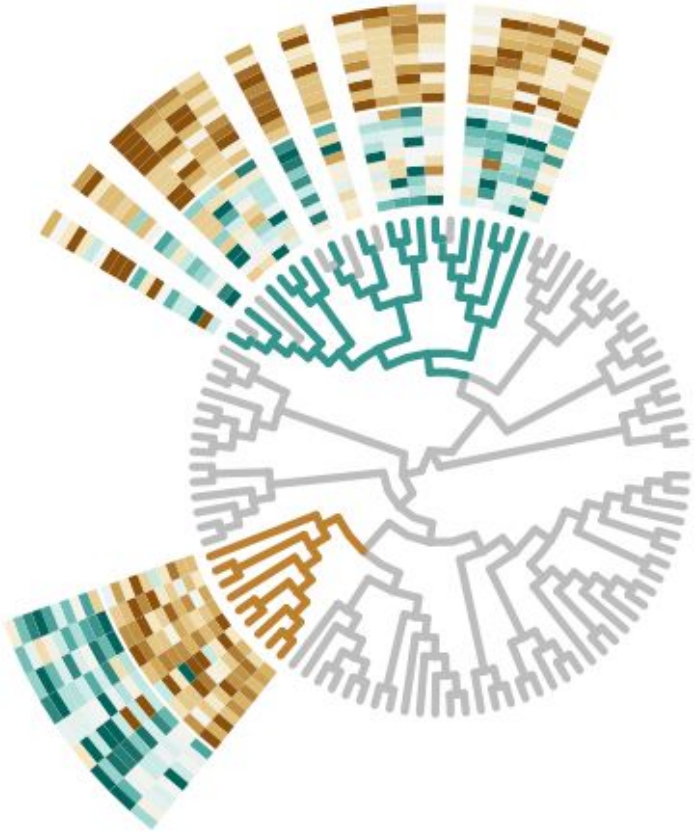

```
> rowLinks(lse)
LinkDataFrame with 100 rows and 4 columns
        nodeLab nodeLab_alias    nodeNum     isLeaf
    <character>   <character> <integer> <logical>
1           t73       alias_1         1      TRUE
2           t67       alias_2         2      TRUE
```

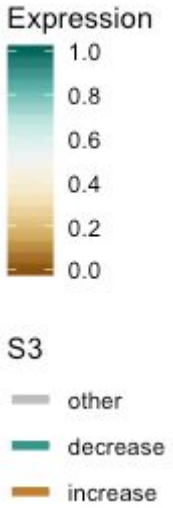
```
> lse <- TreeSummarizedExperiment(assays = count,
+                                 colData = sample_data,
+                                 rowData = entity_data,
+                                 rowTree = tree,
+                                 rowNodeLab = lab)
> lse
class: TreeSummarizedExperiment
dim: 100 20
metadata(0):
assays(1): ''
rownames(100): t73 t67 ... t35 t49
rowData names(0):
colnames(20): C1_1 C1_2 ... C2_9 C2_10
colData names(1): group
reducedDimNames(0):
spikeNames(0):
altExpNames(0):
rowLinks: a LinkDataFrame (100 rows)
rowTree: a phylo (100 leaves)
colLinks: NULL
colTree: NULL
```

100 entities,
20 samples (10 in group C1, 10 in group C2)
Tree: 100 leaves (entity ← → leaf)


Expression
1.0
0.8
0.6
0.4
0.2
0.0
S3
other
decrease
increase

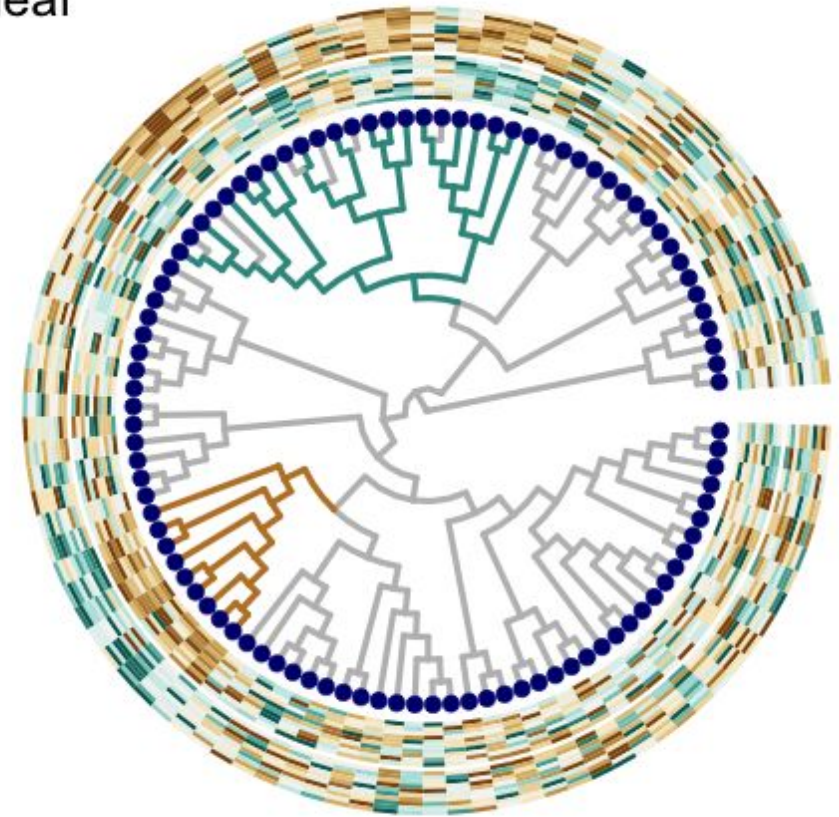






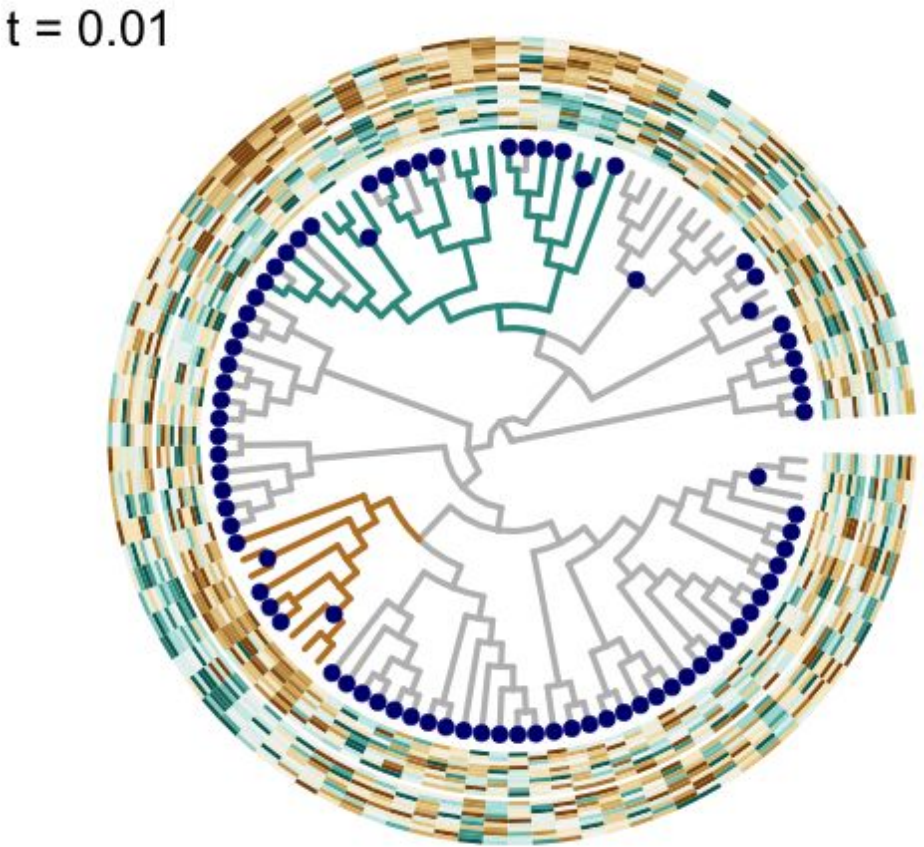
Based on : ggtree + ggplot2

t = leaf

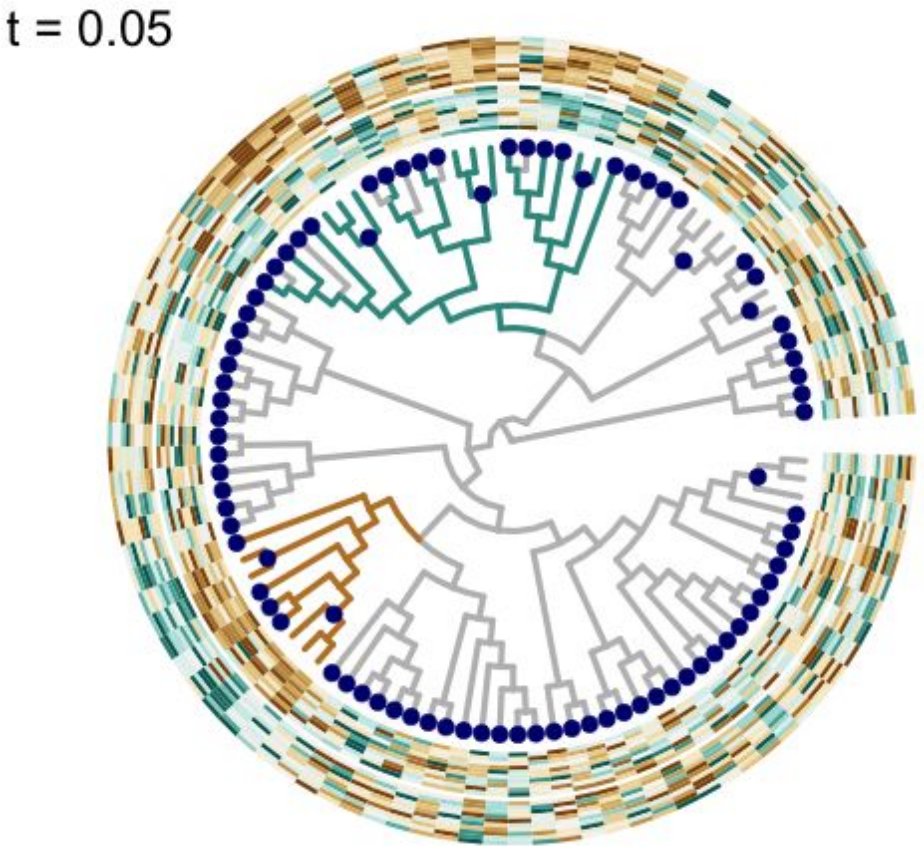


1. **Find candidate levels**
2. Pick the best candidate

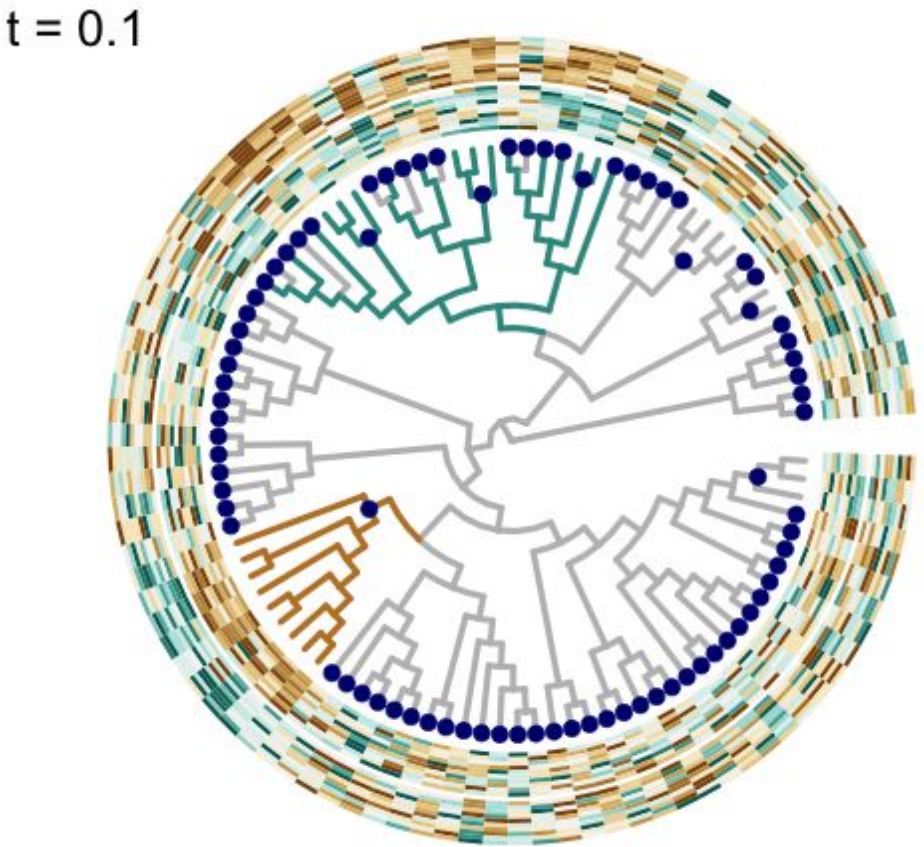
1. **Find candidate levels**
2. Pick the best candidate



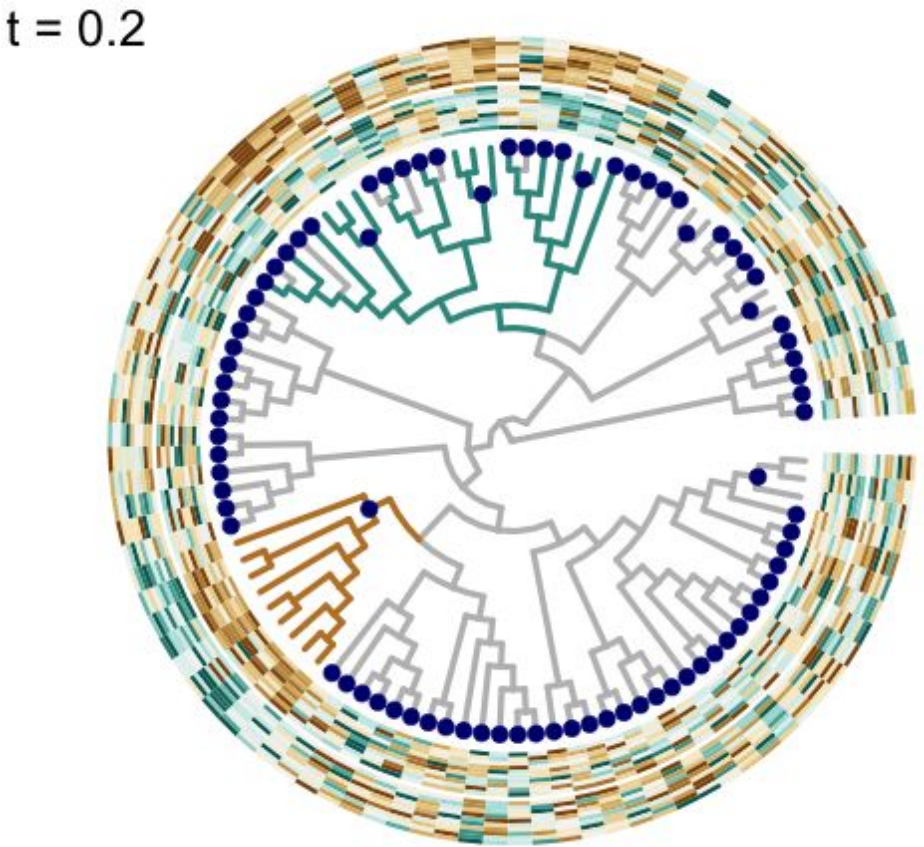
1. **Find candidate levels**
2. Pick the best candidate



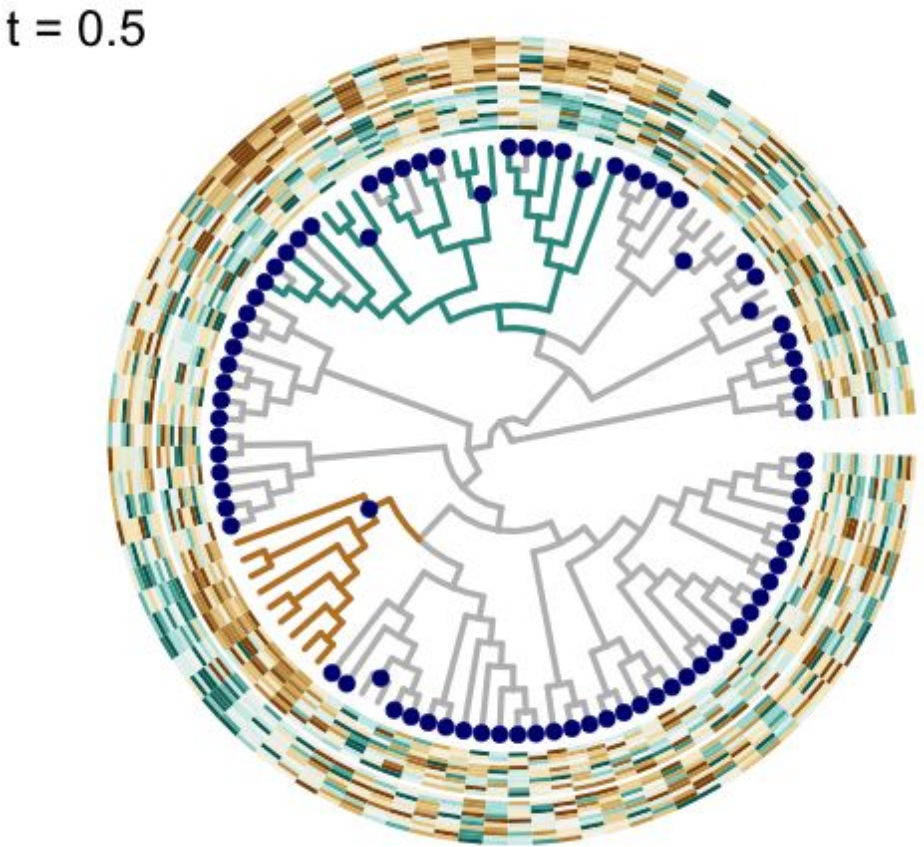
1. **Find candidate levels**
2. Pick the best candidate



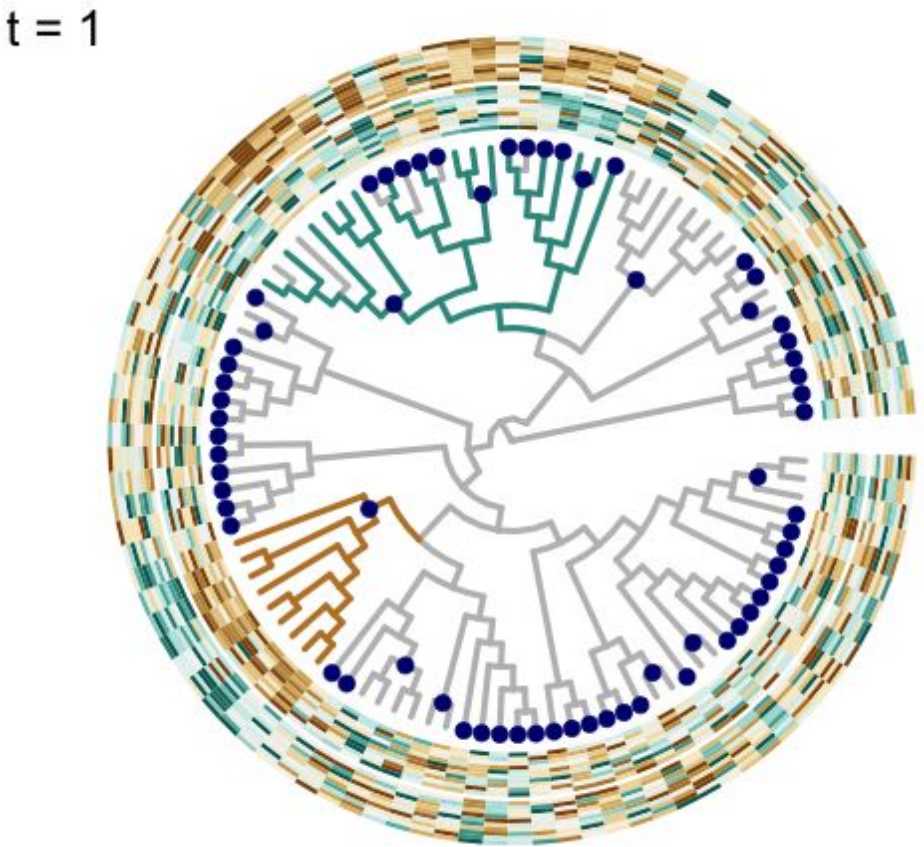
1. **Find candidate levels**
2. Pick the best candidate



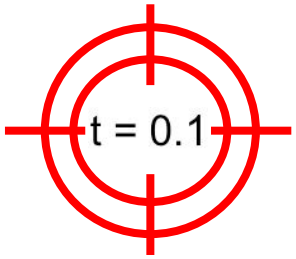
1. **Find candidate levels**
2. Pick the best candidate



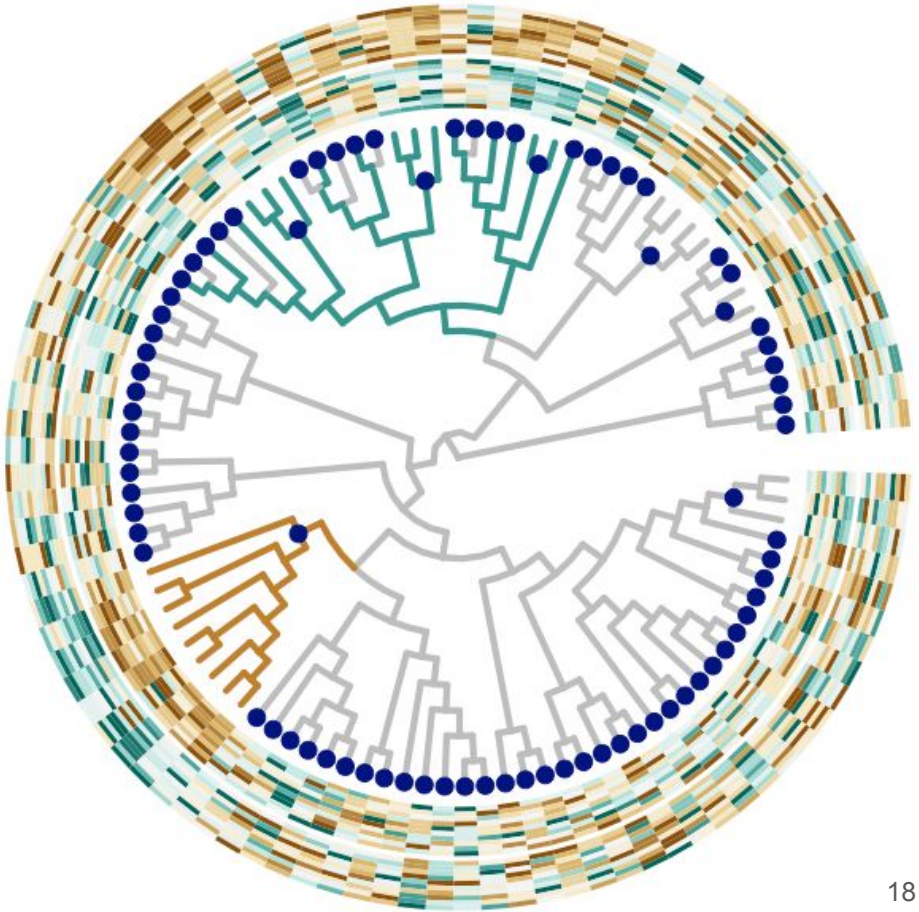
1. **Find candidate levels**
2. Pick the best candidate

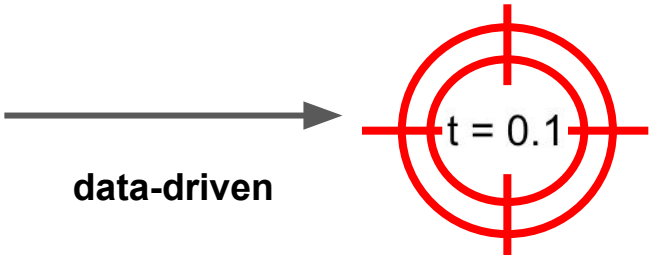


data-driven

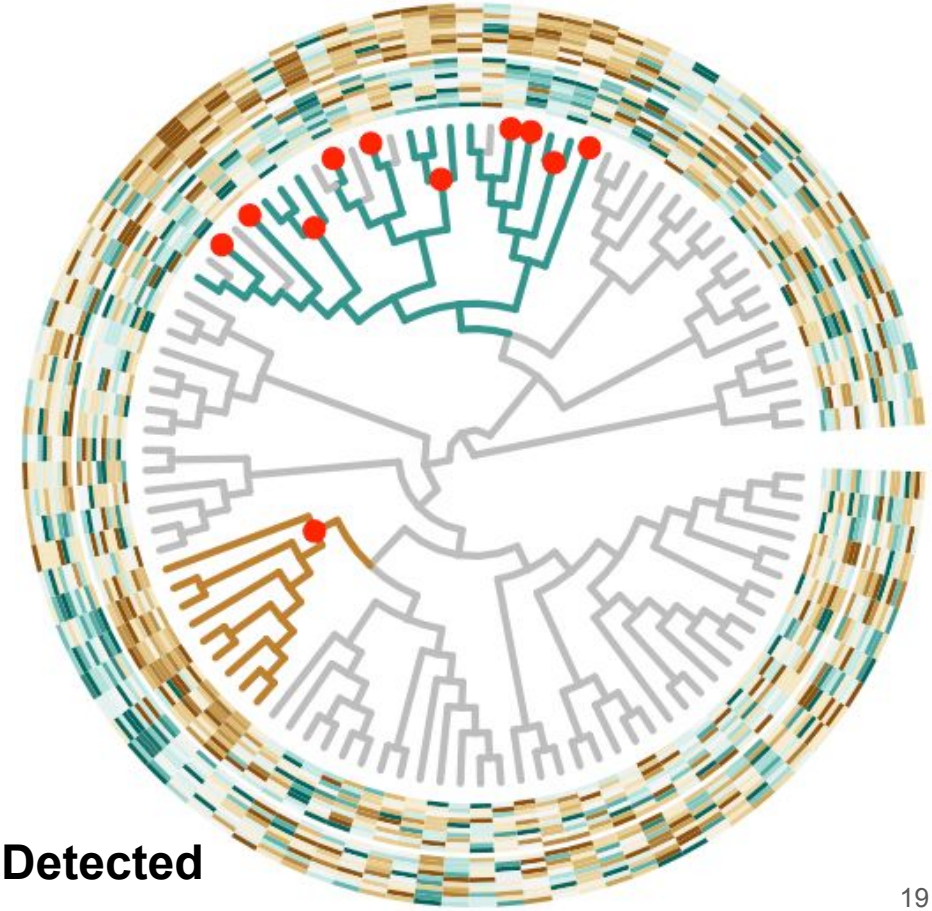


1. Find candidate levels
2. **Pick the best candidate**





1. Find candidate levels
2. **Pick the best candidate**
3. **Perform the multiple hypothesis correction**



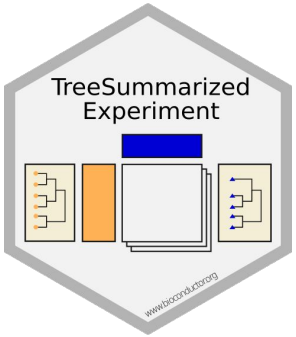
Acknowledgement

Supervisors:

Mark Robinson
Charlotte Soneson

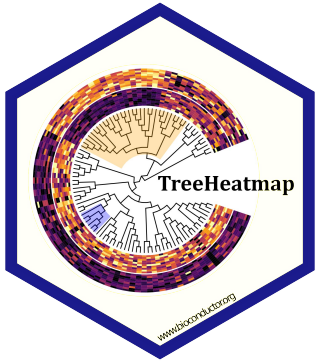
Collaborator in TreeHeatmap

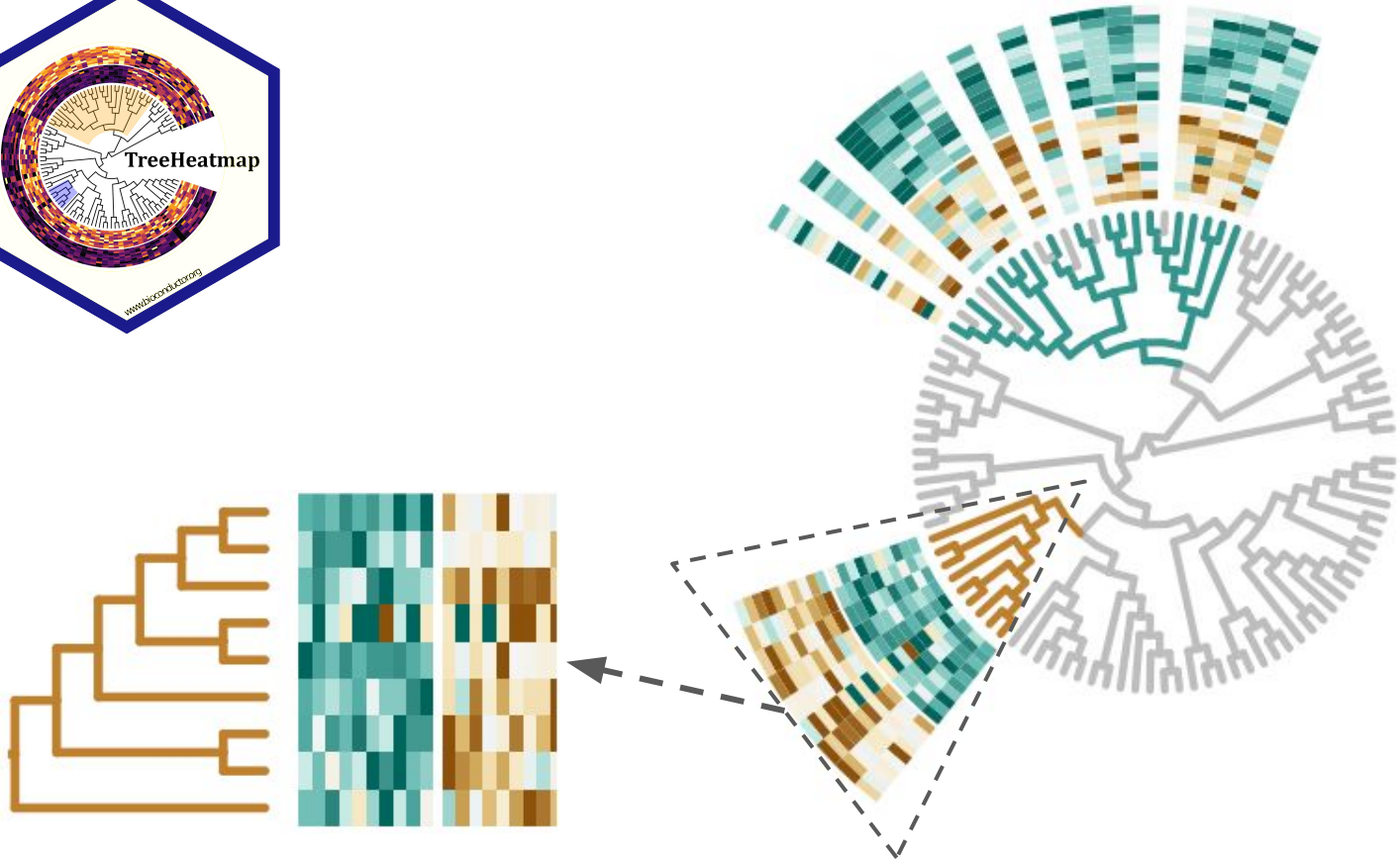
Guangchuang YU



Acknowledgements

- Mark D Robinson
- Héctor Corrada Bravo
- Jayaram Kancherla
- Domenick Braccia
- Hervé Pagès
- Guangchuang YU
- Charlotte Soneson
- Levi Waldron
- Vince Carey
- Kasper D Hansen
- Kévin Rue-Albrecht
- Martin Morgan
- Federico Marini
- Davide Risso
- Stephanie Hicks
- Daniel van Twisk
- Marcel Ramos


TreeHeatmap
www.bioconductor.org


Expression
1.0
0.8
0.6
0.4
0.2
0.0
S3
other
decrease
increase