

Leveraging gating hierarchies to evaluate automated annotation of cells in single-cell data

David Novak^{1,2}, Sofie van Gassen^{1,2}, Yvan Saeys^{1,2}

¹ Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Gent, Belgium
² Data Mining and Modeling for Biomedicine, VIB Center for Inflammation Research, Gent, Belgium

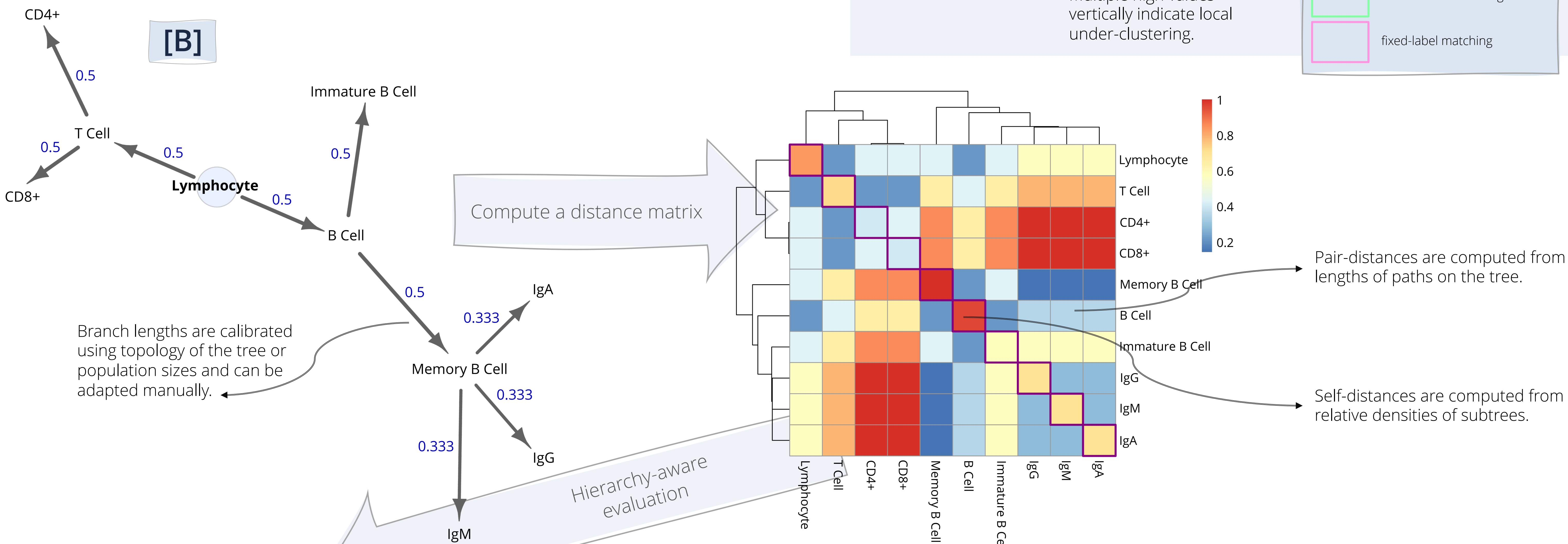
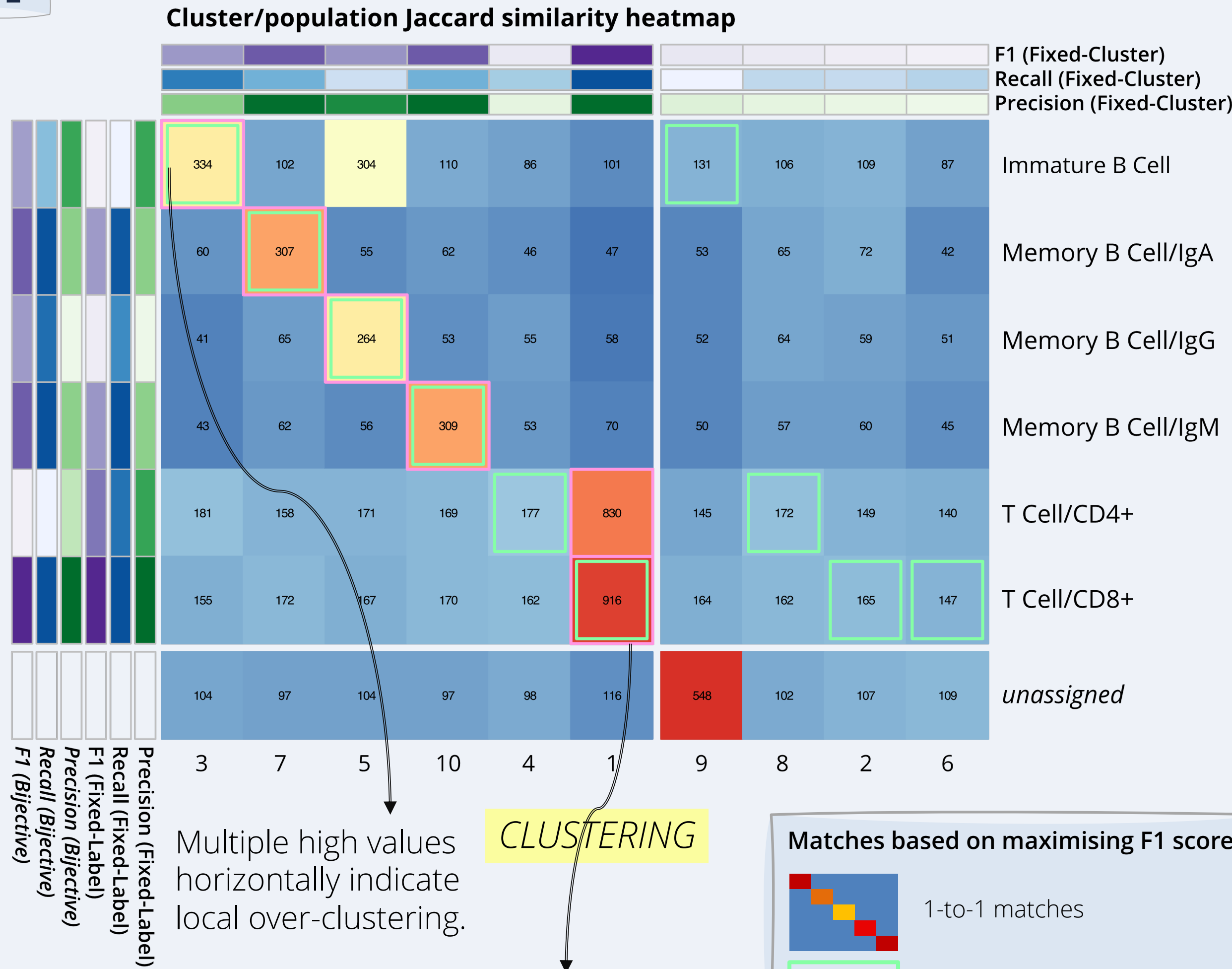
github.com/davnovak/hloss

Clustering of cytometry, scRNA-seq or CITE-seq data is a powerful tool for identifying cell populations of interest by phenotype and function.

Evaluation of clustering methods on manually annotated subsets of single-cell data is helpful in validating each approach.

With **SingleBench**, we address the issues of **determining optimal clustering resolution**^[A] and **incorporating cell-population taxonomies as priors**^[B] to enable more thorough evaluation of performance, tailored to your dataset.

[A]



Rand index (RI) works with all pairs of points from the vector of predicted cluster numbers and vector of true cell labels and counts (mis)matches between the vectors.

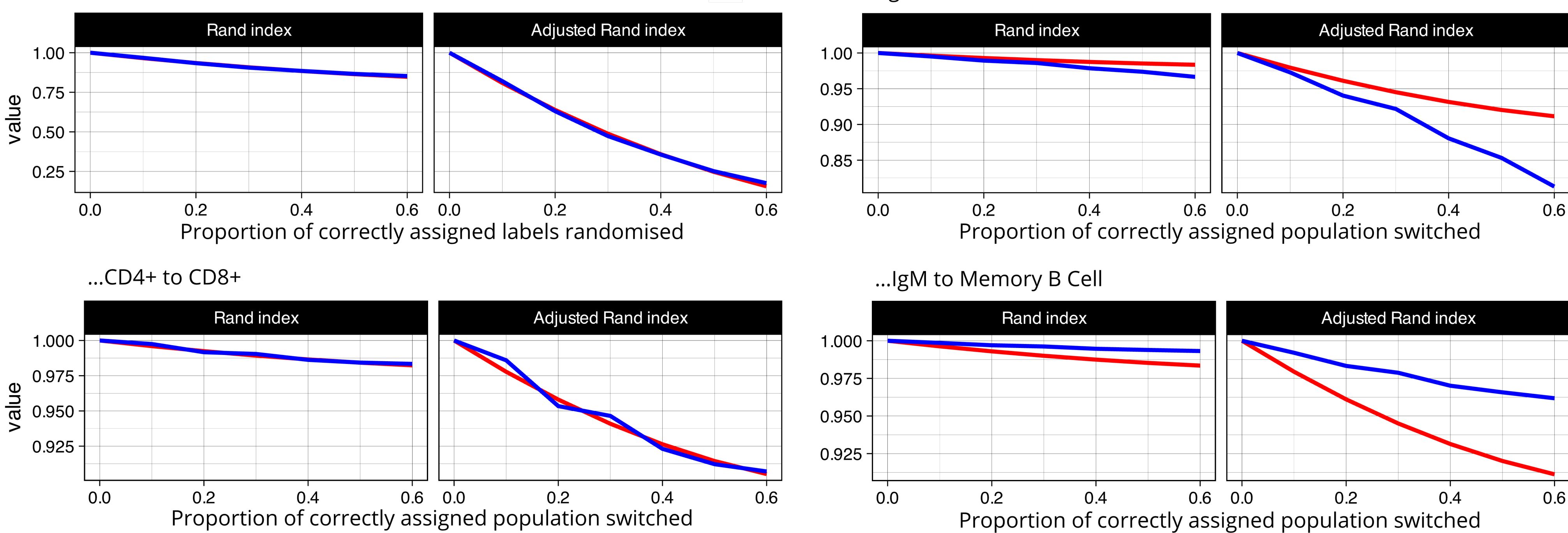
Hierarchy-aware Rand index uses entries in the distance matrix for the true labels to weight matches (and one-minus-distance to weight mismatches).

$$RI = \frac{a + b}{a + b + c + d}$$

The **Adjusted Rand index (ARI)** is corrected for chance and more commonly used.

| Term | Predicted clusters | True labels | Effect of weighting |
|------|--------------------|-------------|---|
| a | Same | Same | Clustering points together correctly is rewarded more in denser regions of the tree. |
| b | Different | Different | Clustering points apart correctly is rewarded less if they are far apart in the tree. |
| c | Same | Different | Clustering points together incorrectly is penalised more if they are far apart in the tree. |
| d | Different | Same | Clustering points apart incorrectly is penalised less in denser regions of the tree. |

Disruption: randomisation through replacing an increasing proportion of correct cluster assignments with random ones



Future work

We are testing hierarchy-aware ARI and F1 scores on real-life datasets with taxonomic labelling.

We are testing various calibration strategies for determining branch lengths of taxonomic trees correctly.