ClustIRR

Monthly meeting

June 2023

Recapitulation

Changes

Submission

Graph discussion

Status after last meeting

- Pre-submission phase of ClustIRR
  - Integrate dataset into vignette
  - Submit package

Recapitulation

Changes

Submission

Graph discussion

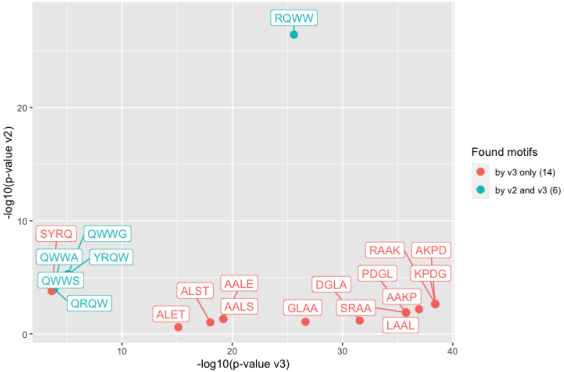
Dataset integration

- Integrated reference dataset of Gliph2
- Demonstrated clustering with inserted ground truth
  - Reference dataset of 10^4 CDR3β sequences
  - Take random sample of $n = 500$ CDR3s
  - Artificially enrich 20 sequences with motif *RQWW*
  - Simulate clonal expansion with two sequences :
    - *CATSRAAKPDGLAALETQYF* and *CATSRAAKPDGLAALSTQYF*
  - that get attached to the sample 15 times each
- Documented local and global clustering

Trim flanks

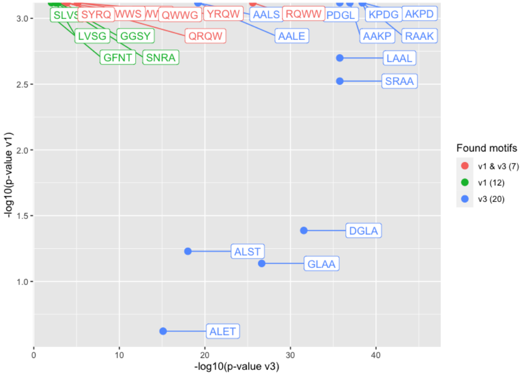
- Sequences could be ”trimmed to death”
- E.g., *CASSCDRTQFV* (length 11) trimmed by $2 * 6 \rightarrow NA$
- Solved by integrating warnings and error message
    - Warnings, if sequences are lost
    - Error message, if all sequences are trimmed

Local clustering - v3 vs v2



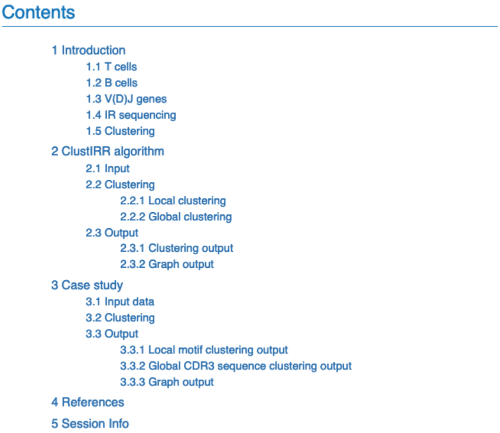
- Motif *RQWW* gets found with high *p* by both versions
  - *QWWA*, *QWWG*, *QWWS*, *QRQW* and *YRQW* motif related
- Clonally expanded sequences also share motifs
  - Six motifs, counting only non-redundant sequences
  - 14 additional motifs, counting also redundant sequences

Local clustering - v3 vs v1



- v1 also does not find clonally expanded sequence motifs
- But v1 finds 5 sequences not found by v2 or v3
  - *GFNT*, *GGSY*, *LVSG*, *SLVS*, and *SNRA*
- Bootstrapping related, high fdr-values both in v2 and v3
  - At least fdr=0.1524478, up to fdr=0.310575

Vignette



Contents

- Added extended intro, detailed algorithm description
- Wrote additional vignette for detailed version comparison

Recapitulation

Changes

Submission

Graph discussion

Submission

- Integrate exemplary graph (optional)
- Submit package

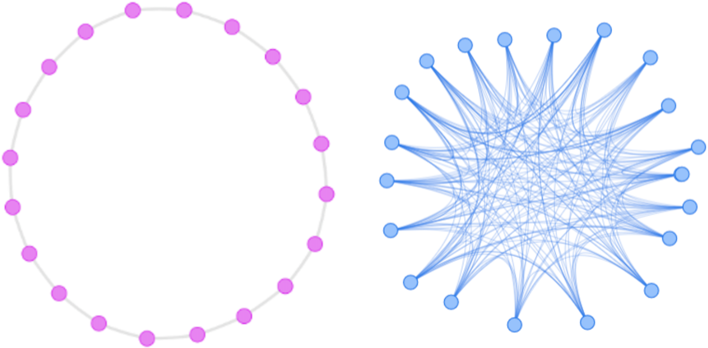
Recapitulation

Changes

Submission

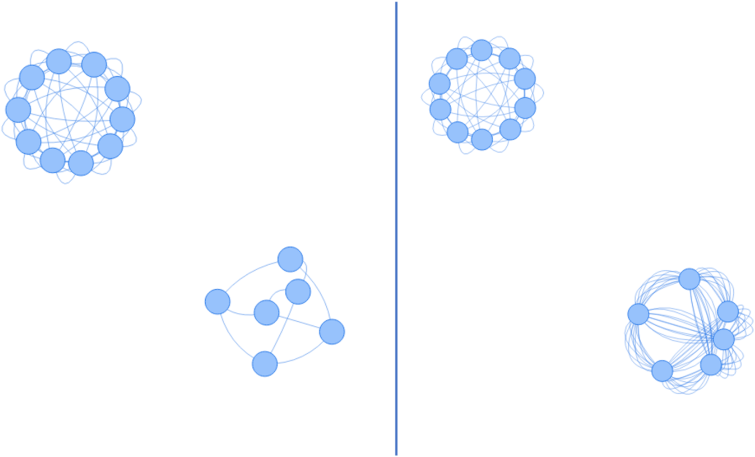
Graph discussion

Motifs - ring vs fully connected



- Edges = shared motif, no information loss with ring (?)
- Could even be reduced to chain without information loss

Motifs - Occluded motifs



- When two CDR3 share > 1 motif, edges get reduced (left)
- Could be solved for example by using count weights

Global similarity - long motif

- CDR3 of length 20 = essentially long motif
- Edges = similarity, enables hubs to exist
- Stop splitting into global and local, use ”levels”?