

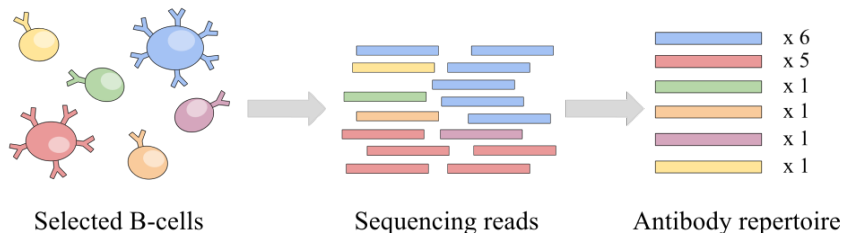
# Improving accuracy of antibody repertoire construction from barcoded Rep-seq data

Student: Maria Chernigovskaya  
Scientific advisor: Sergey Bankevich



## Antibody repertoire problem

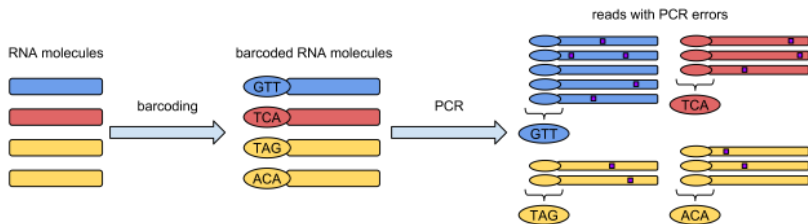
- ▶ Antibody repertoire is a set of antibodies circulating in organism with their abundances.
- ▶ Fundamental problem in immunoinformatics field.



- ▶ Computational problem: clustering with many closely located centers.

# Unique Molecular Identifiers

Help to identify all reads derived from identical antibodies.



State-of-the-art tools:

- ▶ barcodedIgReC (Shlemov et al., 2016)
- ▶ MiGEC (Shugay et al., 2014)
- ▶ pRESTO (Vander et al., 2014)

## Goal of the project

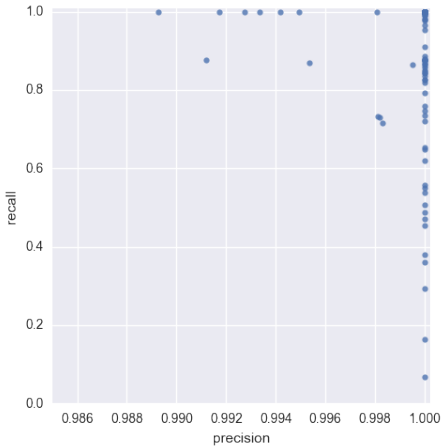
Improve the accuracy of the antibody repertoire construction:  
detect barcodedIgReC bottlenecks.

### Results:

- ▶ An effective way to improve consensus calculation procedure found.
- ▶ Bug in PCR simulation found.

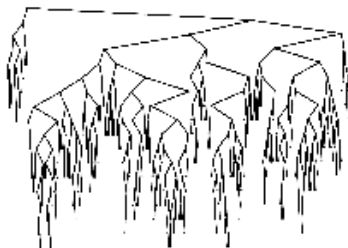
# Clusters with correct consensus

Some clusters have low recall ( = true cluster split into pieces).



## Analysis of the problem

- ▶ PCR is a branching process.
- ▶ Sometimes errors on early stages make the branching tree unbalanced.

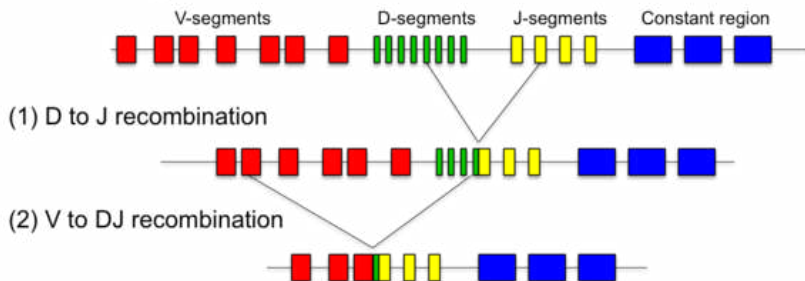


- ▶ Reads from big branch can form separate cluster with wrong consensus or just spoil the consensus.

## Solution

- ▶ When unsure at certain position, pick a nucleotide from the germline.
- ▶ Correction could be wrong e.g., when an antibody has an SHM at that position.

Germline configuration:



## Numeric results

	total clusters	clusters with correct consensus	percent of clusters with correct consensus
Before	3706	305	8.2%
After	2219	311	14%



Thank you for your attention!