



High-throughput sequencing of B-cell receptors in lymphocytic leukemia

September 2018

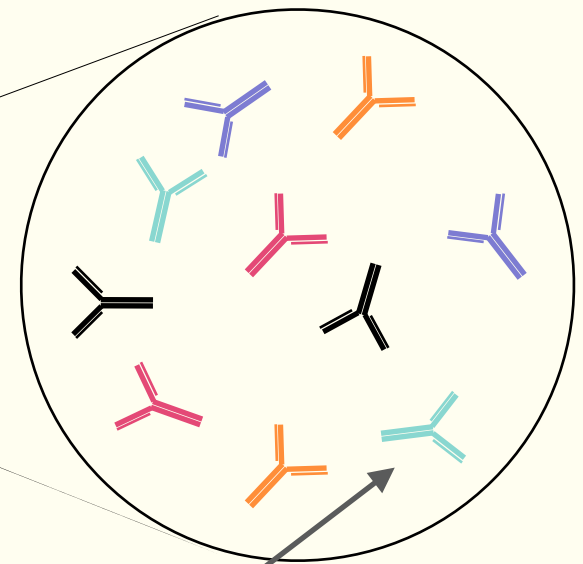
Plan

B-cells and Immunoglobulin proteins

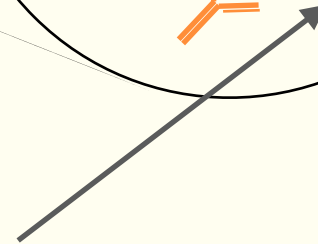
Comparison and evaluation of Rep-seq analysis tools

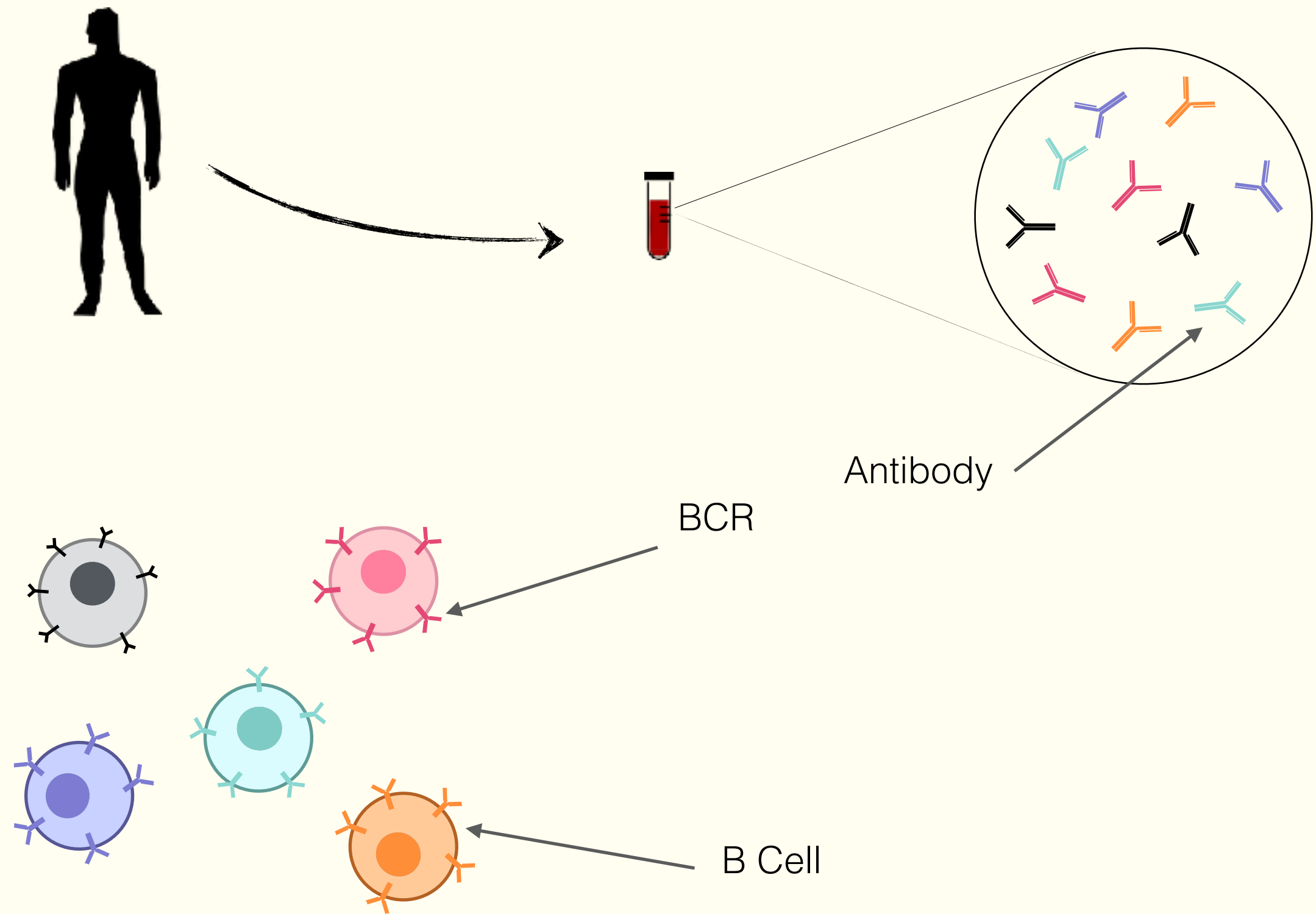
MixClus: an ultra fast and sensitive algorithm for clustering BCR-IG sequences

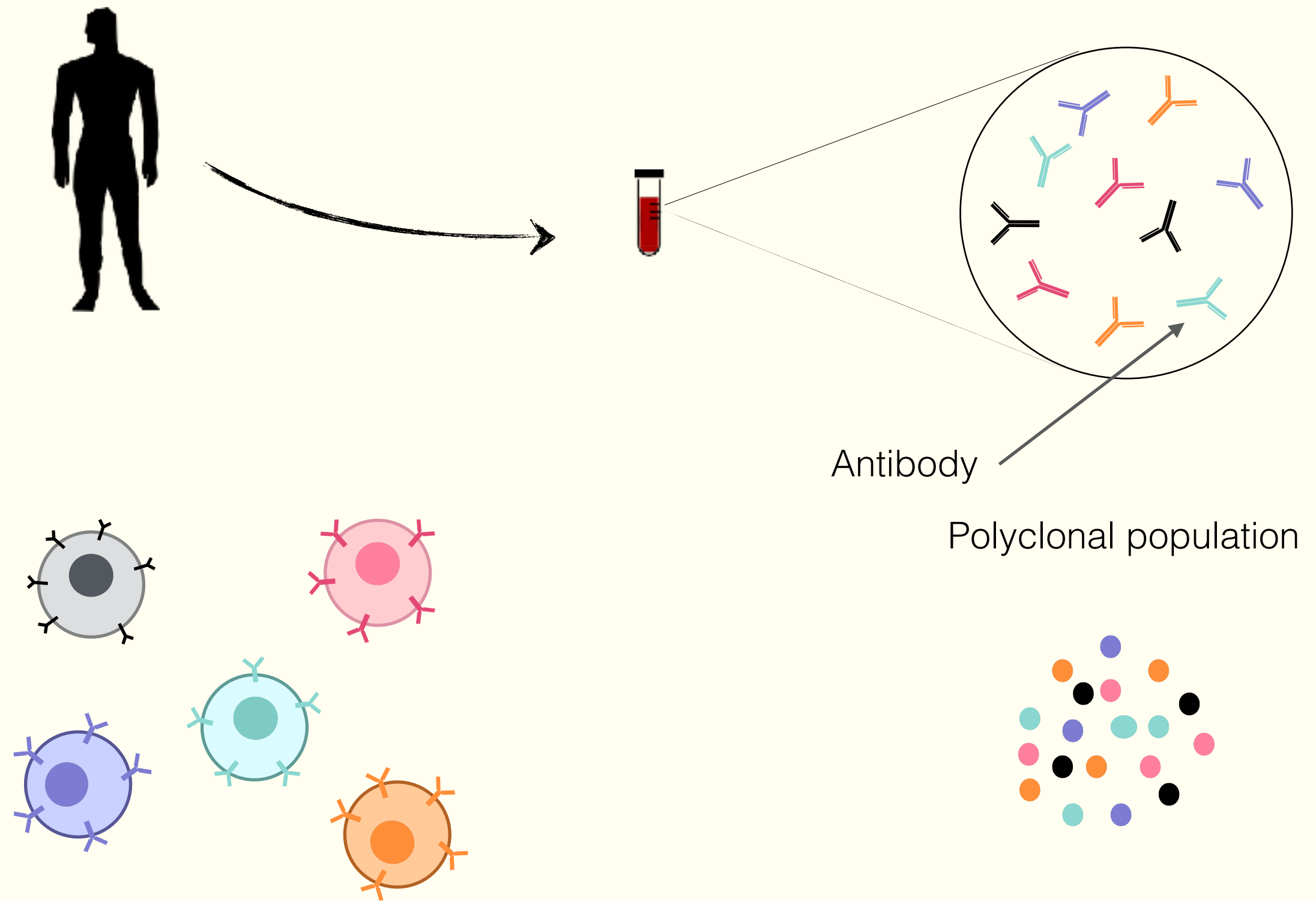
Overall conclusion and perspective

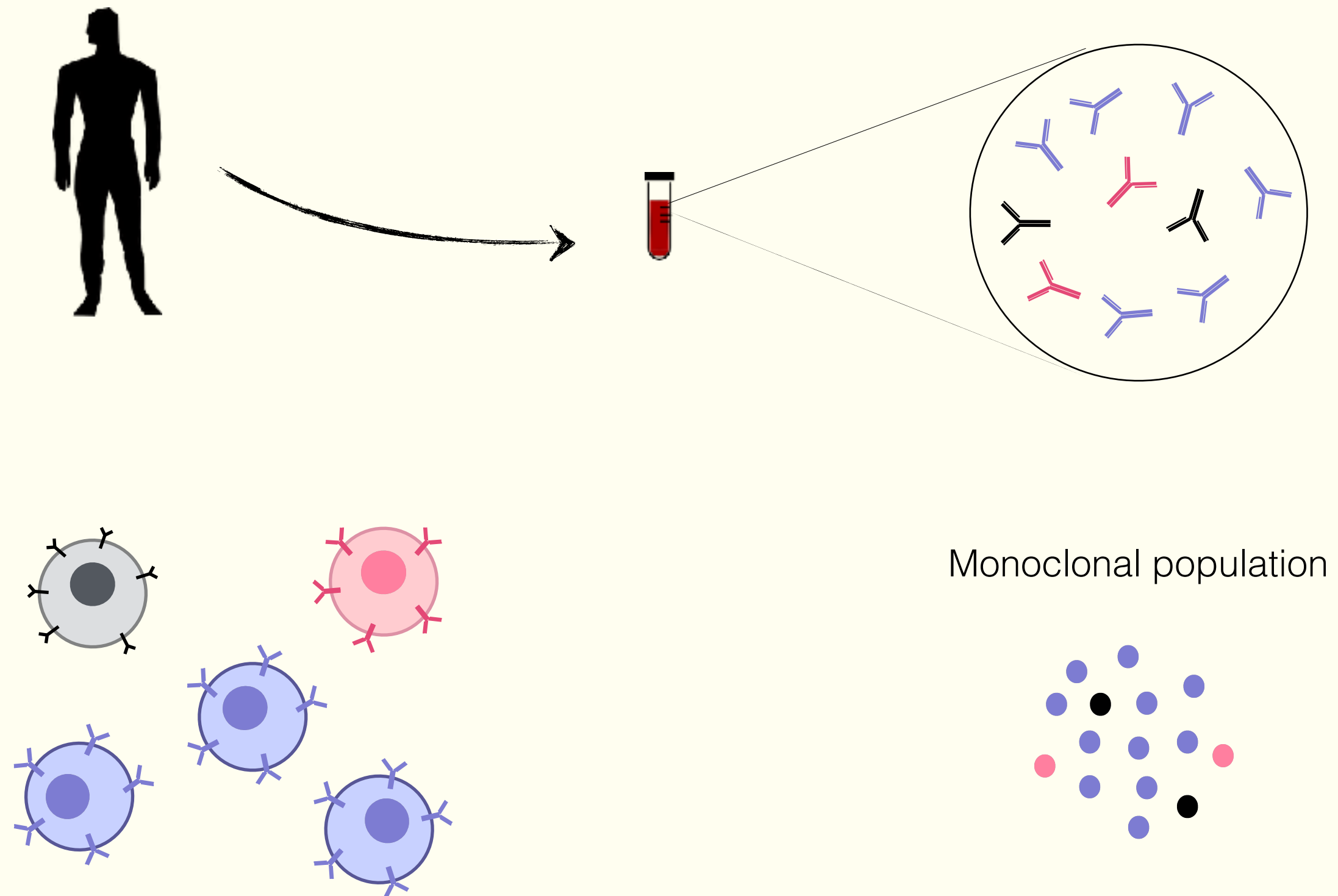


Antibody

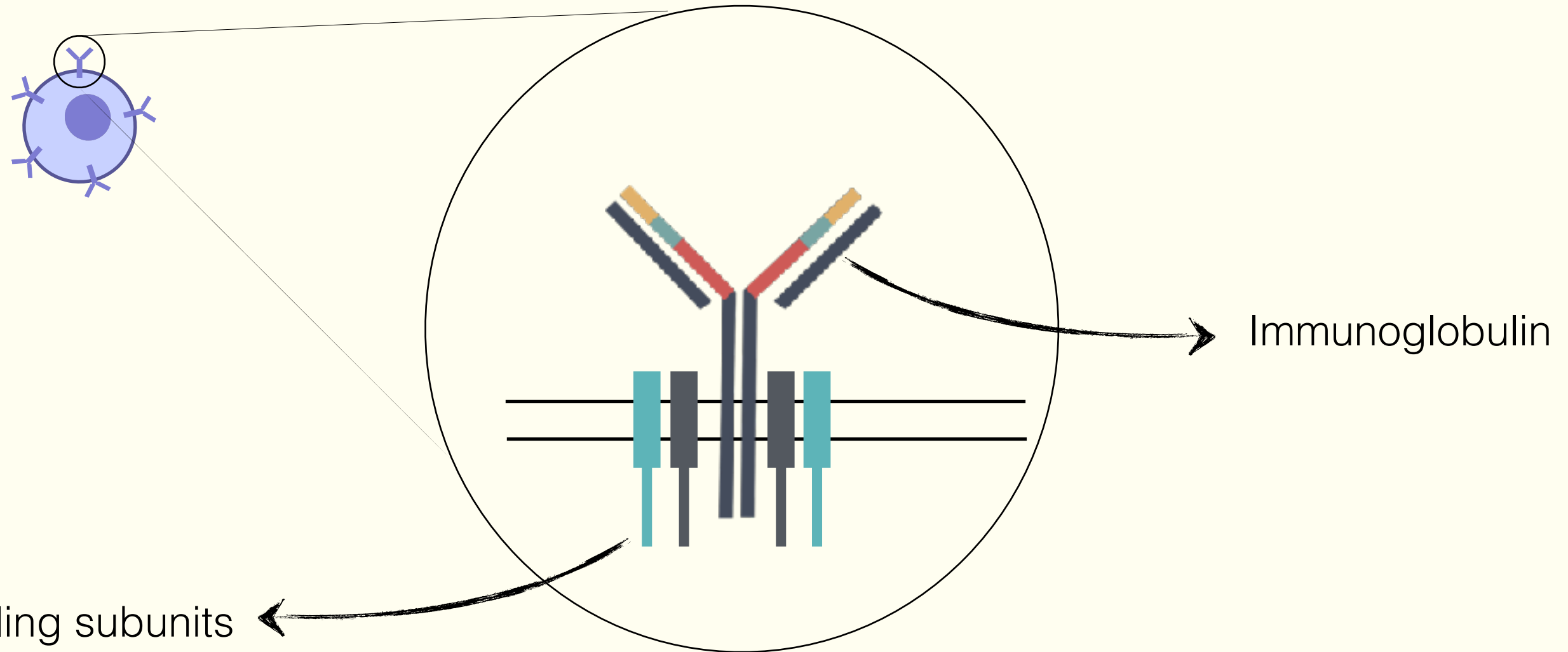




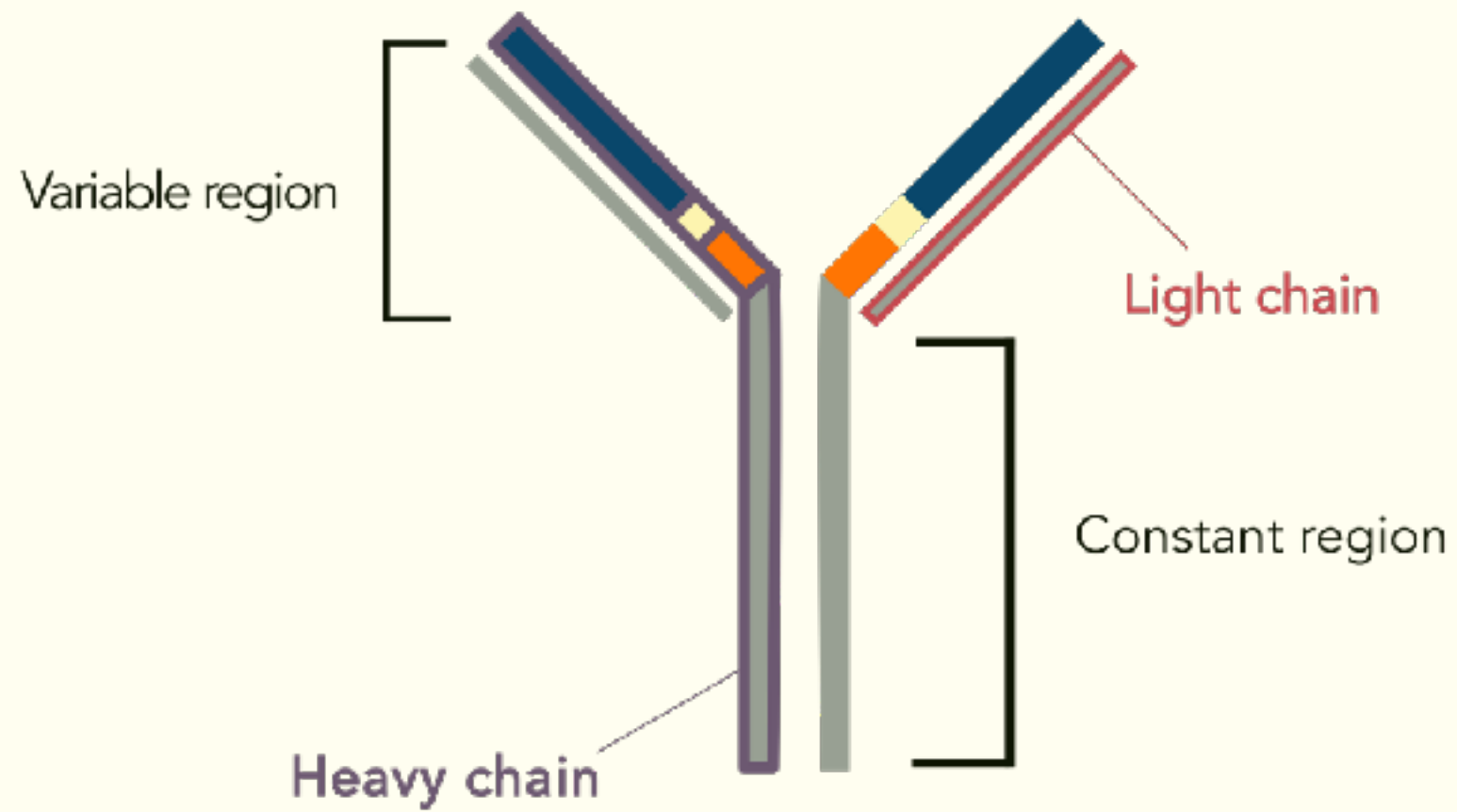




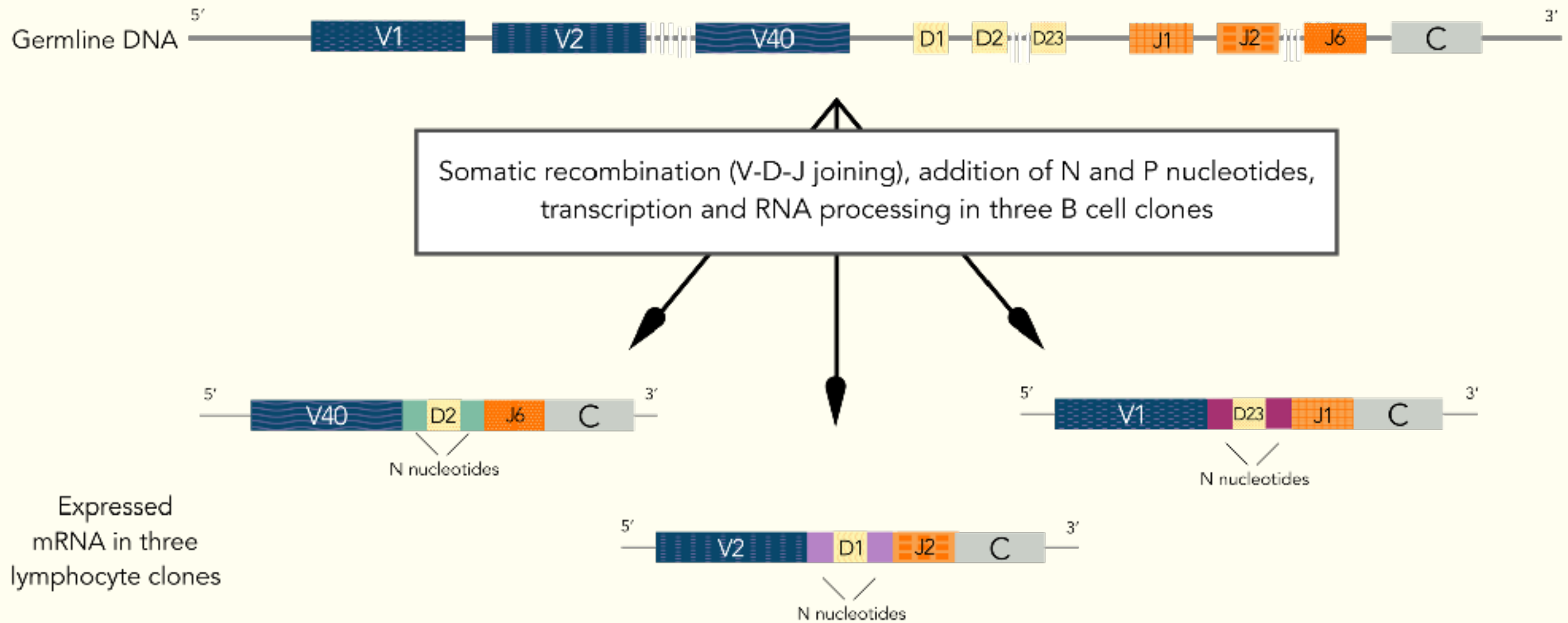
BCR



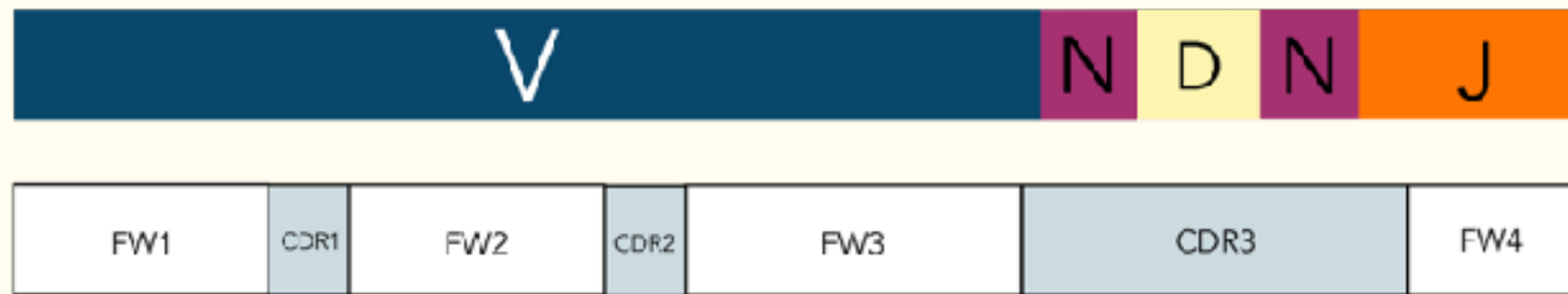
Immunoglobulin



Diversity of antigen receptor genes

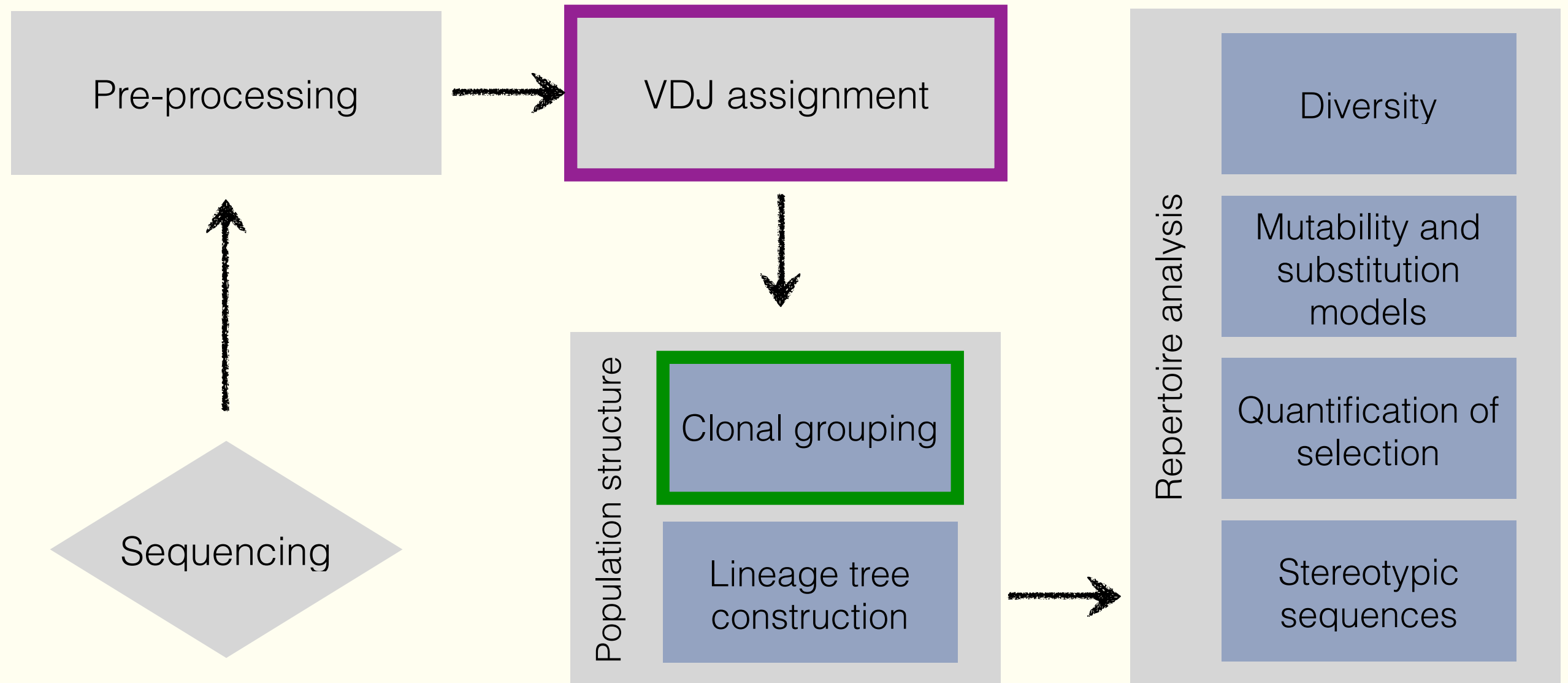


B-cell receptor genes



Comparison and evaluation of Rep-seq analysis tools

The pipeline for B-cell repertoire analysis



Selected tools

Algorithmic approach	IMGT/high V-quest	MixCR	Partis	Vidjil
	Alignment based	Alignment + K-mer	Modeling-based (HMMs)	K-mer (Alignment free)

Results for VDJ assignment

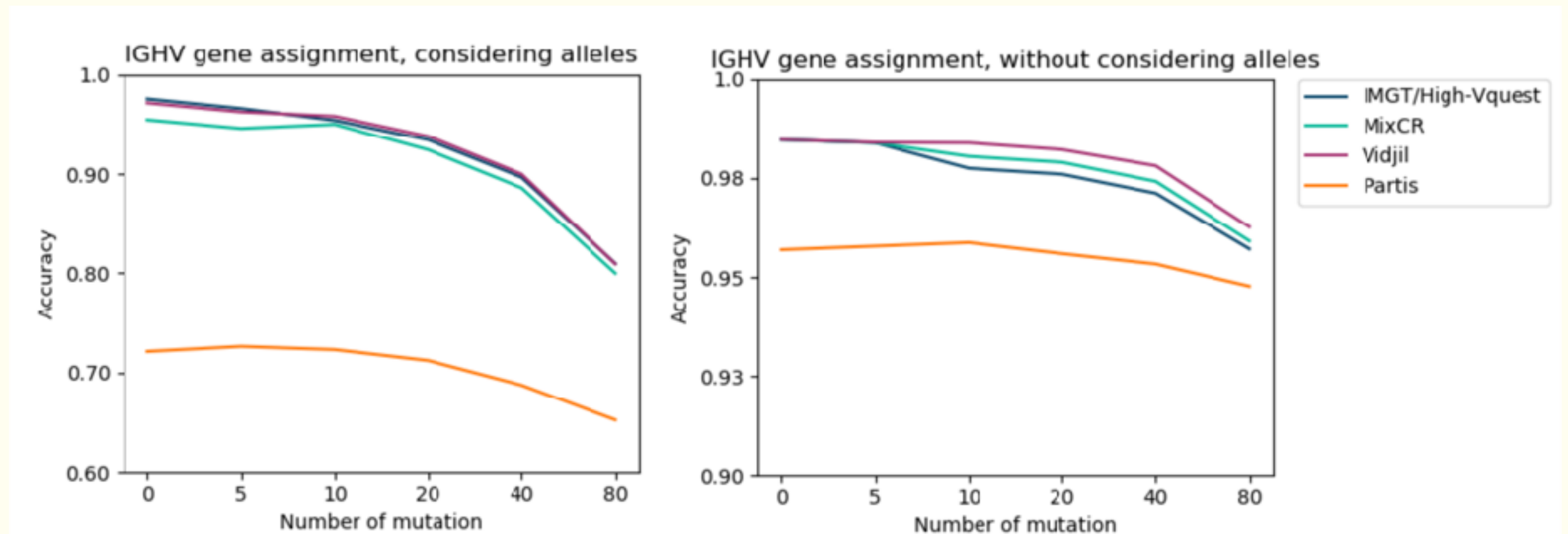
Simulated data

```
graph LR; A[VDJ assignment] --> B["Random V(N)D(N)J  
{0, 5, 10, 20, 40, 80}  
mutations.  
Total of 7151 sequences"]
```

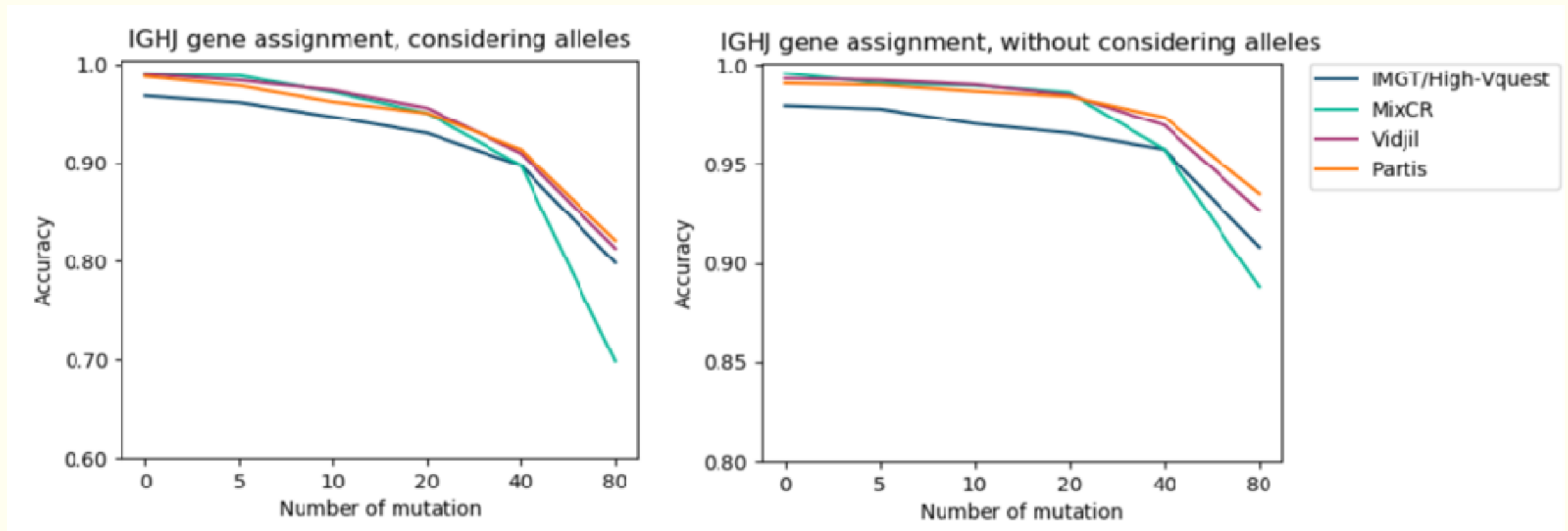
VDJ assignment

Random V(N)D(N)J
{0, 5, 10, 20, 40, 80}
mutations.
Total of 7151 sequences

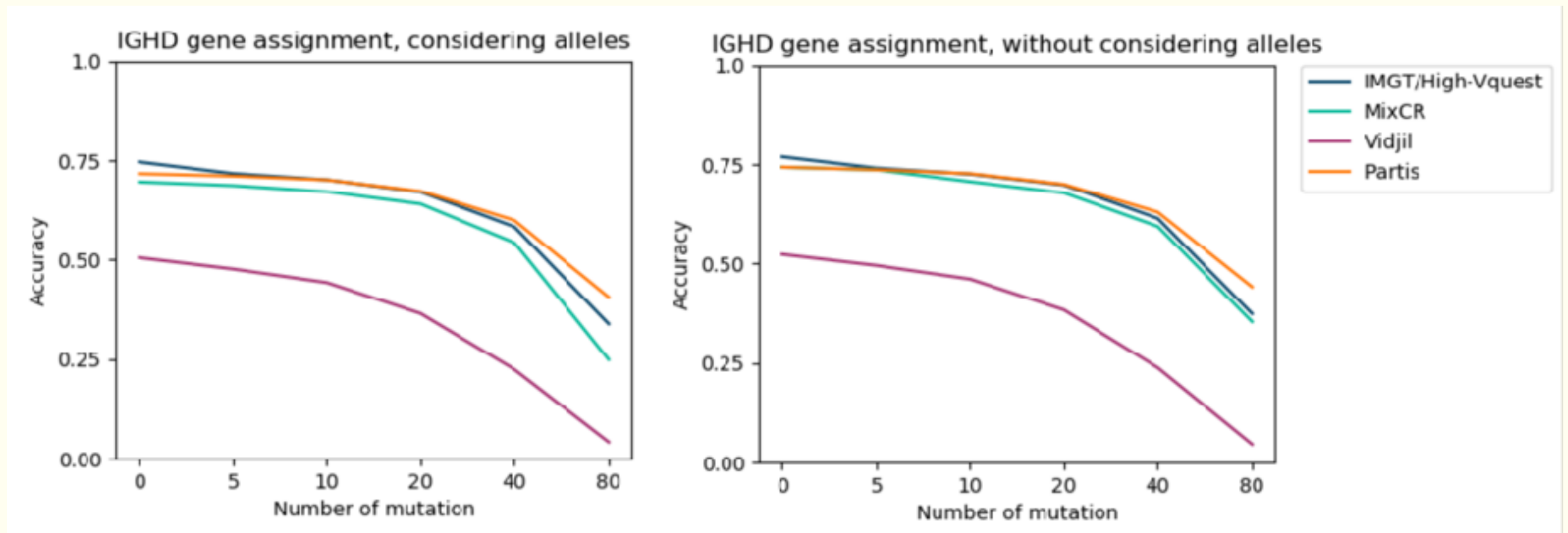
IGHV genes assignment with(out) considering alleles



IGHJ genes assignment with(out) considering alleles



IGHD genes assignment with(out) considering alleles



Computational time of VDJ assignment

Average computation time (m)	IMGT/high V-quest	MixCR	Partis	Vidjil
	48.6	0.07	19.8	0.02

Results for VDJ assignment

Real data



```
graph LR; A[VDJ assignment] --> B[IGHV assignment agreement];
```

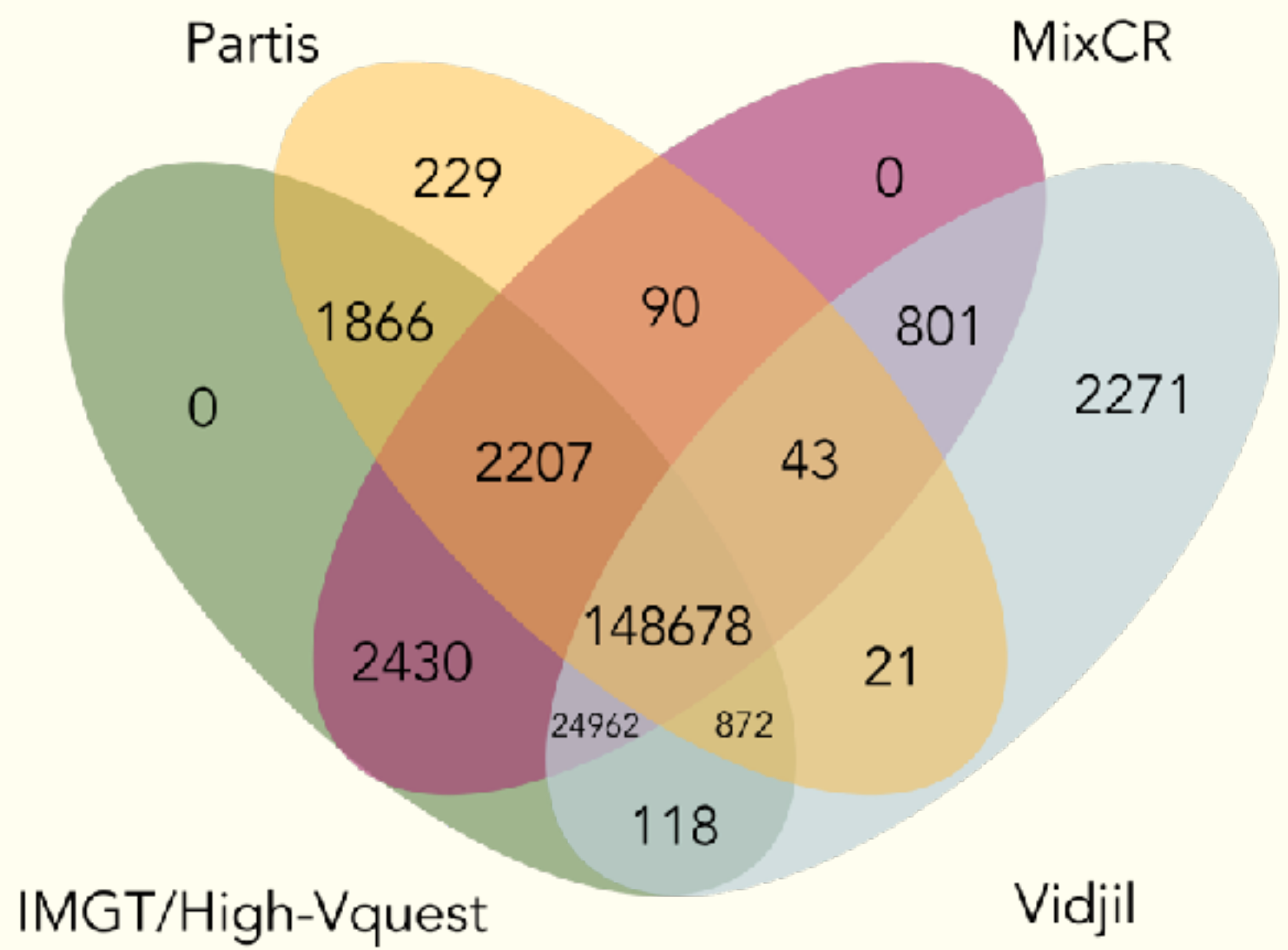
VDJ assignment

IGHV assignment
agreement

Real datasets

	# reads	# Sequences	# Unique Sequences
Patient 1	583147	528377	184588
Patient 2	398123	382194	229192
Patient 3	458094	399723	207636

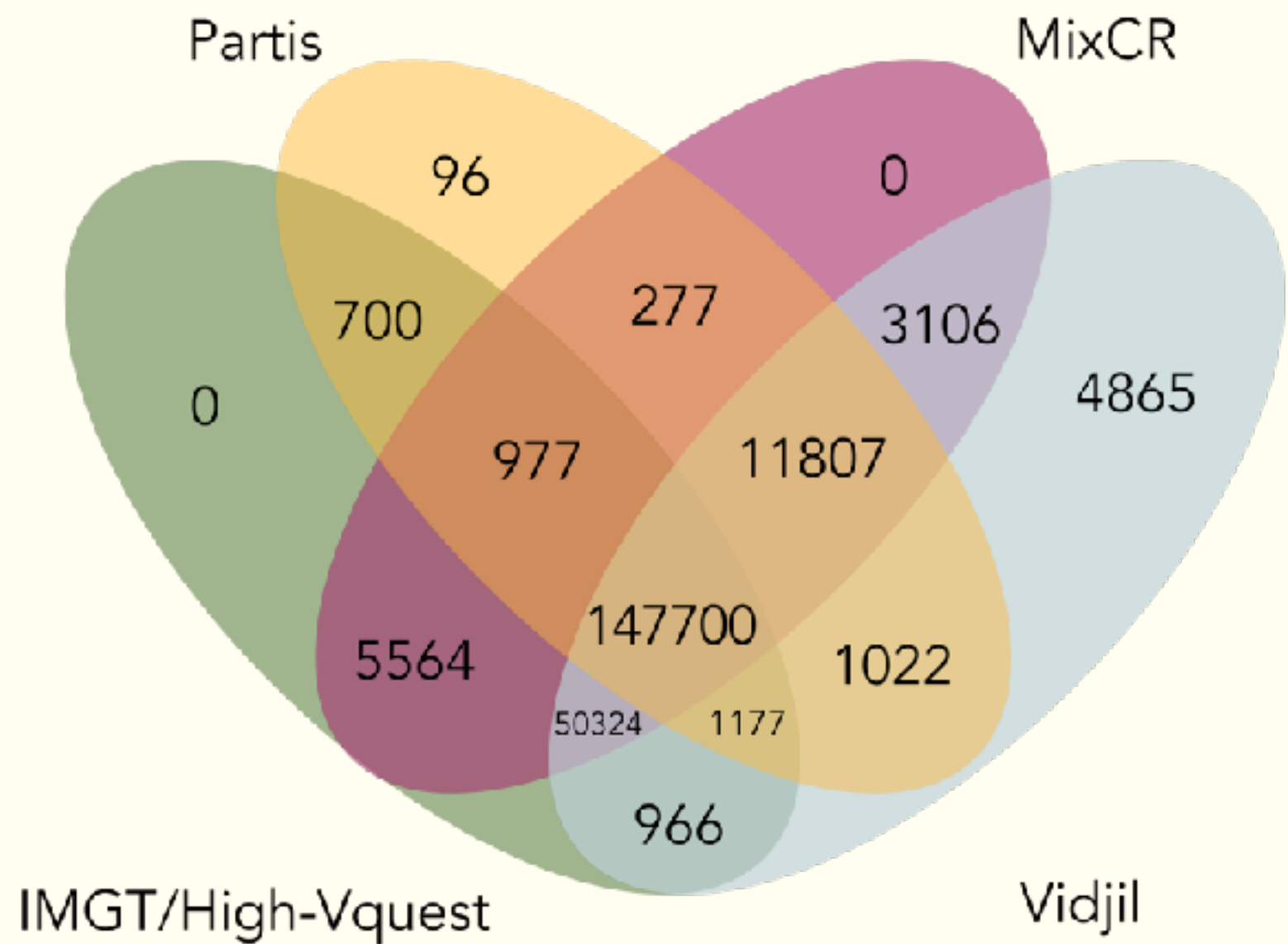
Patient 1



184588 sequences

4 different assignments : 185

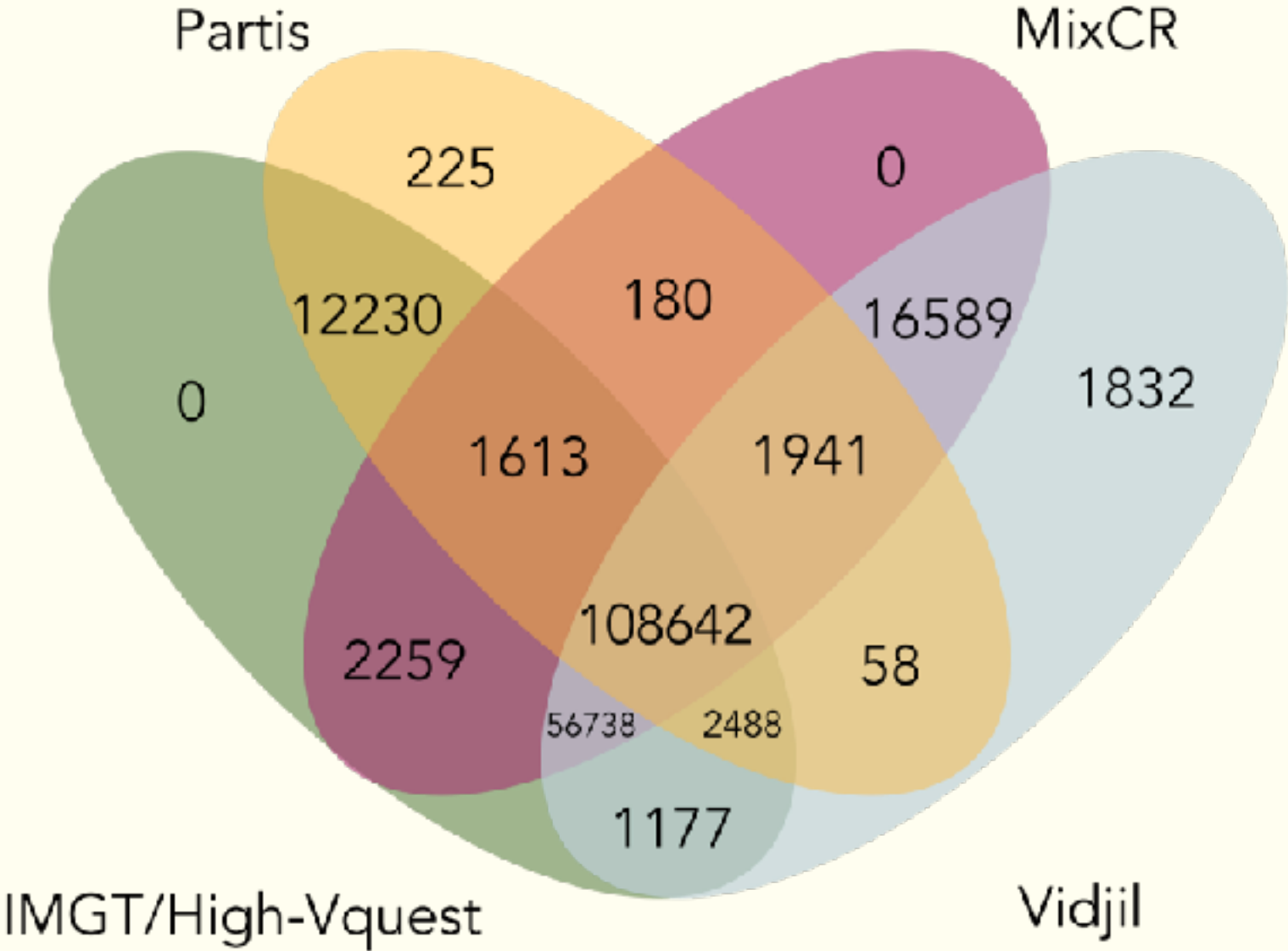
Patient 2



229192 sequences

4 different assignments : 611

Patient 3



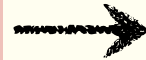
207636 sequences

4 different assignments : 1691

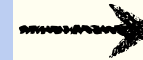
Results for clonal grouping

Simulated data

Clonal grouping



Simulated repertoire
based on patient data



2 major clones 39%, 46%
+ 4 minor clones
(between 0.1 and 0.5%)
+ 526 polyclonal background
27731 sequences

Performance of clonal grouping on simulated data

	High-Vquest	MixCR	Partis	Vidjil
Precision $P=TP/(TP+FP)$	1	1	0,9952	1
Recall $R=TP/(TP+FN)$	0,522	0,212	0,9354	0,278
F-score $F=2(P \times R)/(P+R)$	0,686	0,35	0,9643	0,435
# Cluster (expected # cluster = 532)	560	540	530	586
Computational time (m)	12960	4	240	2

Results for clonal grouping

Real data



```
graph LR; A[Clonal grouping] --> B[3 patient's sequences]
```

Clonal grouping

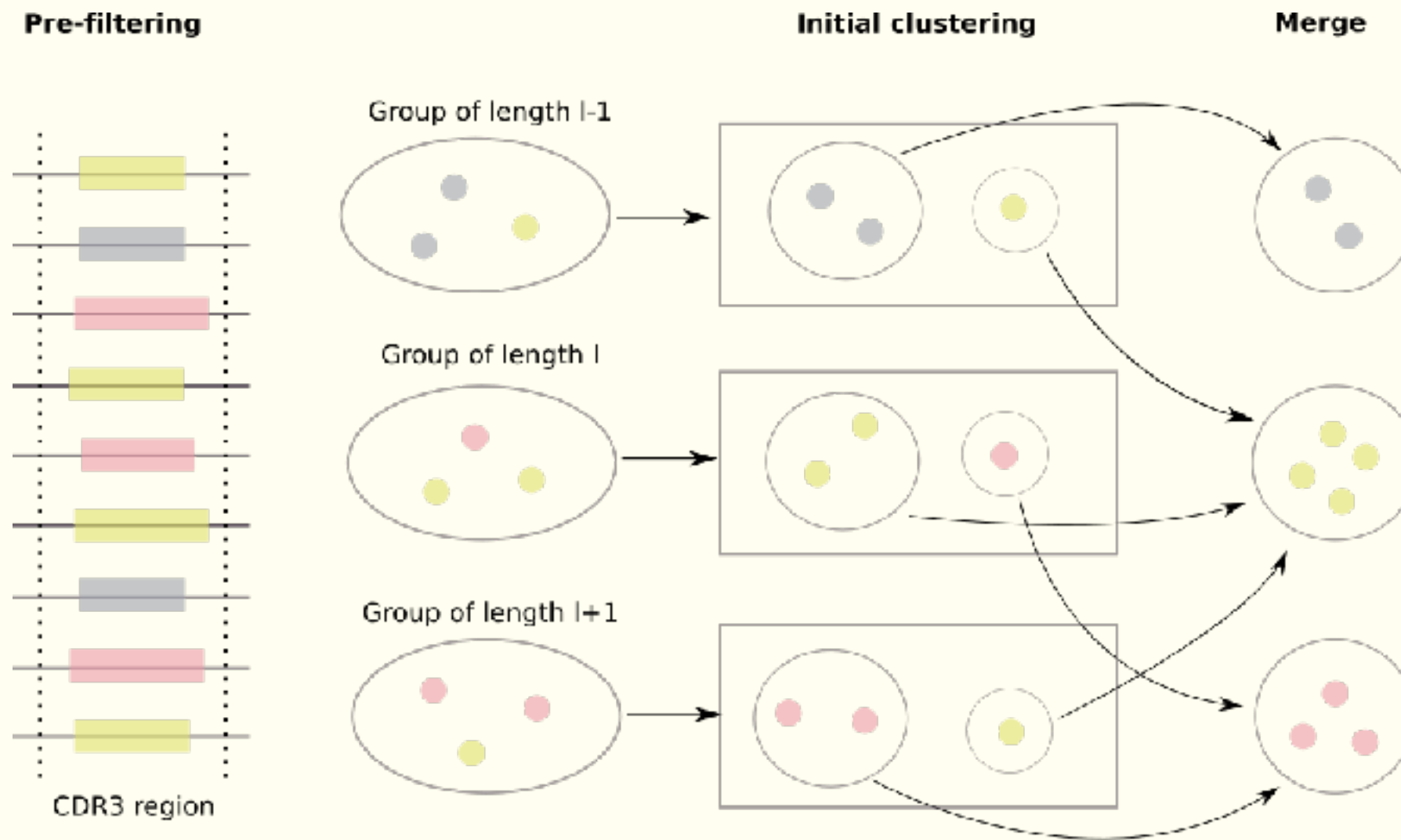
3 patient's sequences

Performance of clonal grouping on real data

	Patient 1		Patient 2		Patient 3	
	# Clone	Silhouette	# Clone	Silhouette	# Clone	Silhouette
High/V- quest	18724	0.29	36696	0.17	27398	0.10
MixCR	26918	0.15	30078	0.19	24835	0.14
Partis	2909	0.37	11875	0.29	6395	0.22
Vidjil	30719	0.03	66063	0,009	51735	0,004

MixClus: an ultra fast and sensitive algorithm for clustering BCR-IG sequences

Mixclus : a progressive cluster growth approach



Mixclus' performance on simulated data

27731 sequences

Number of expected cluster : 532

	Mixclus	Change-O	Partis	Vidjil
Precision	0.99	0.99	0,9952	1
Recall	0.99	0.99	0,9354	0,278
F-score	0.99	0.99	0,9643	0,435
# Cluster	560	540	530	586
Computational time (min)	0.02	0.2	240	2

Mixclus' performance on simulated data

464928 sequences

Number of expected cluster : 619

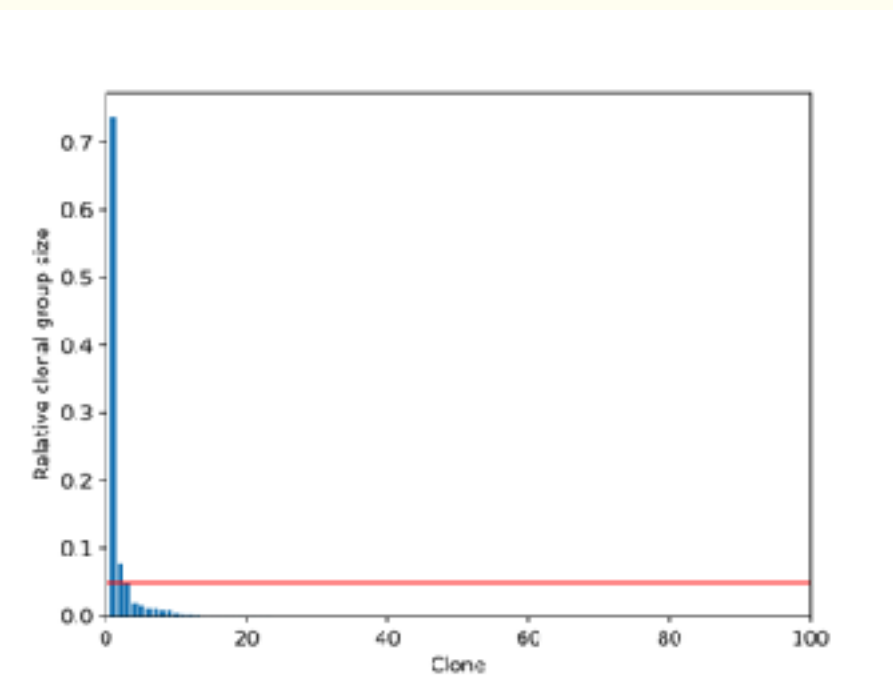
Tool	#clusters	Erreur VJ	precision	recall	F-score	Time(min)
Mixclus	622	0,001	0.99	0.99	0.99	0.51
Change-O	704	0,001	0.99	0.99	0.99	265.7
Partis	1169	0.08	0.96	0.89	0.92	743.2
Vidjil	69526	0,002	0.99	0,041	0,057	4

Mixclus' performance on real data

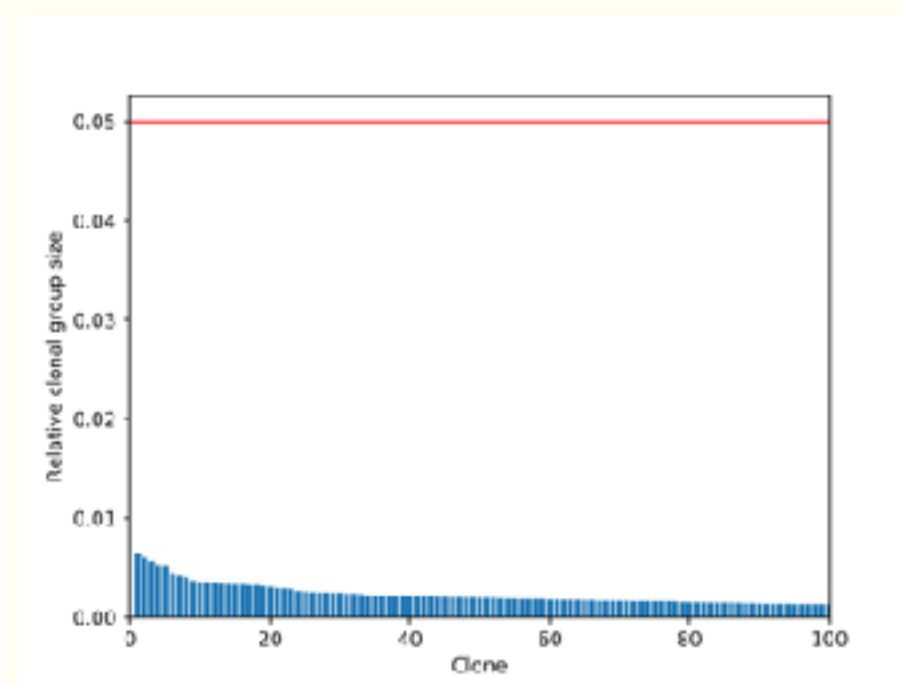
Tool	#clusters	Erreur VJ	silhouette	Time(min)
Patient 1				
Mixclus	3311	0.05	0.72	3.78
Change-O	3098	0.07	0.69	40.8
Partis	2909	0.2	0.37	544.86
Vidjil	30719	0.03	0.03	4
Patient 2				
Mixclus	16041	0.08	0.63	6
Change-O	11483	0.06	0.67	27.4
Partis	11875	0.36	0.29	639.68
Vidjil	66063	0.04	0,009	3
Patient 3				
Mixclus	7453	0.07	0.42	25
Change-O	7803	0.01	0.60	96
Partis	6395	0.53	0.22	494.73
Vidjil	51735	0.04	0,004	1

The biggest 100 clones' size distribution

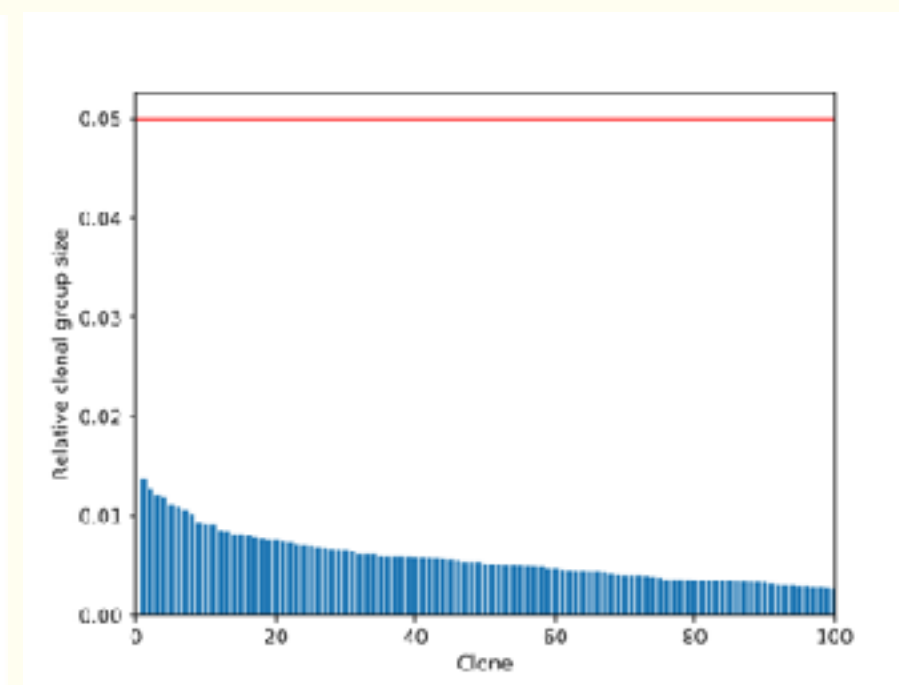
Clonality threshold : 5%.



Patient 1



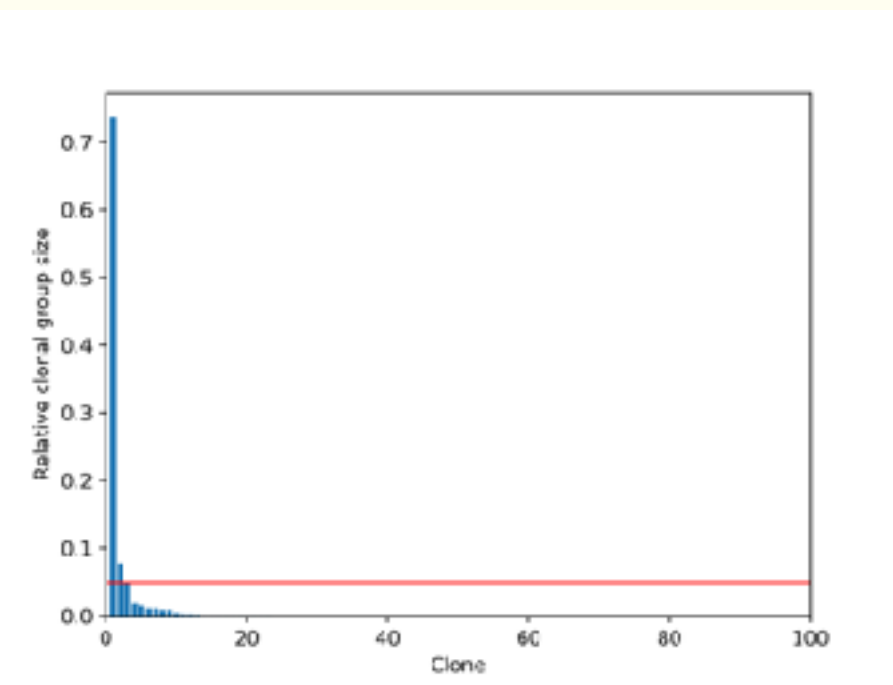
Patient 2



Patient 3

The biggest 100 clones' size distribution

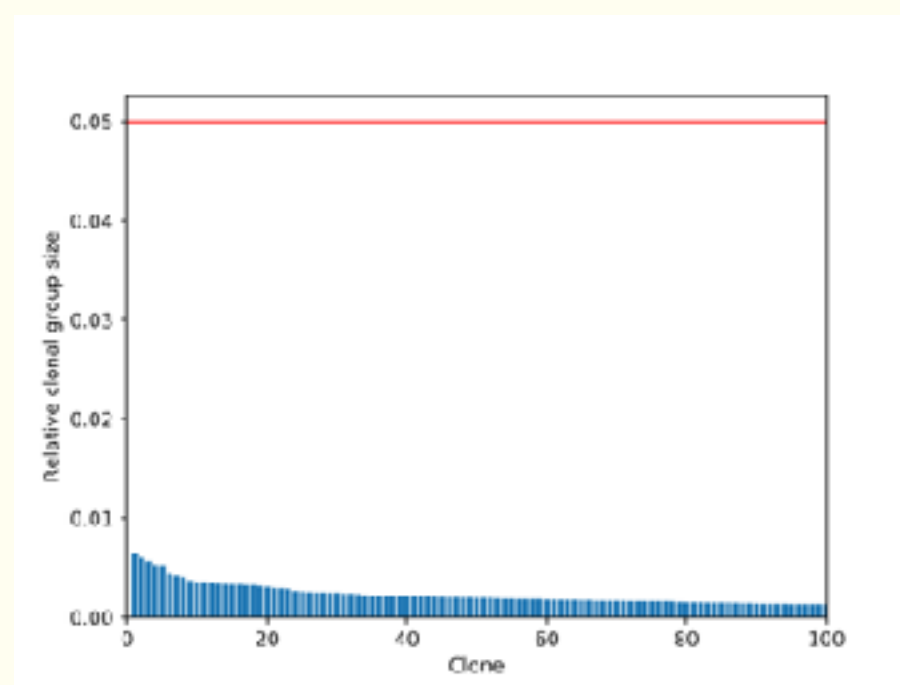
Clonality threshold : 5%.



Patient 1



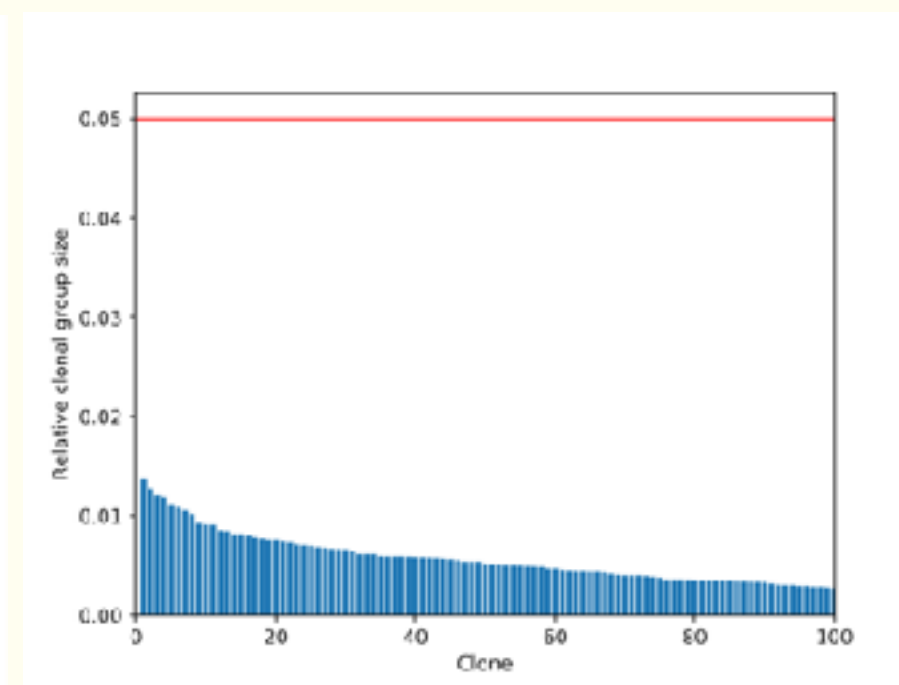
Monoclonal



Patient 2



Polyclonal



Patient 3



Oligoclonal / Polyclonal

Overall conclusion and perspective

Developing a pipeline of BCR-IG analysis to be used for clinical applications

Lineage tree construction