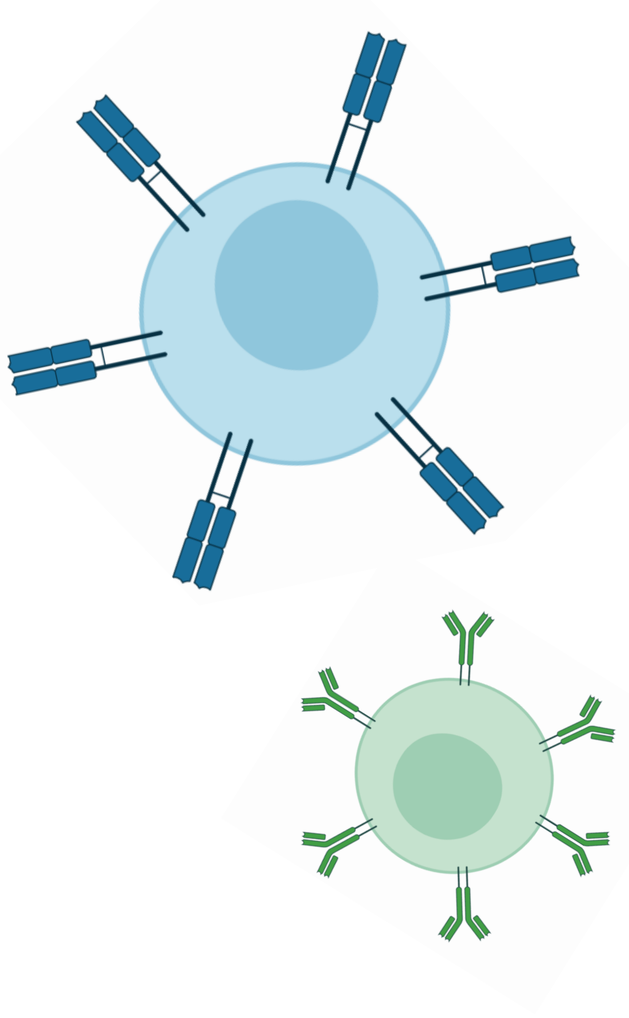
Intro to immune repertoire sequencing and analysis

Maggie Russell

TFCB 2023

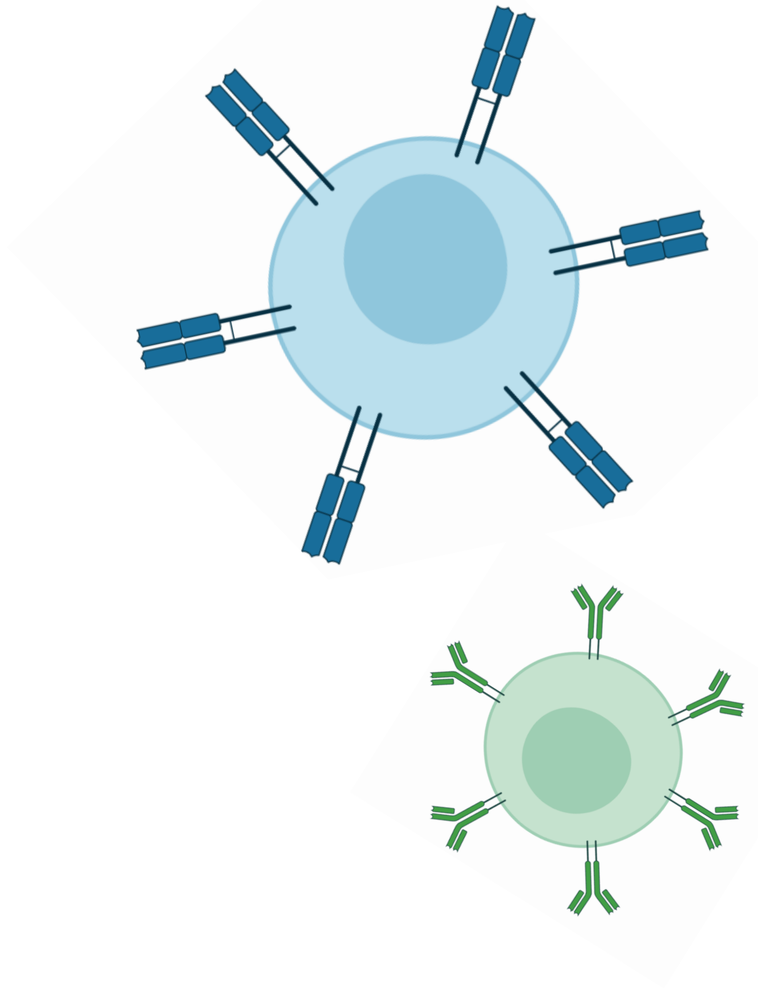
Lecture goals:

1. learn about immune repertoire sequencing
2. familiarize with immune repertoire data
3. work through an example analysis



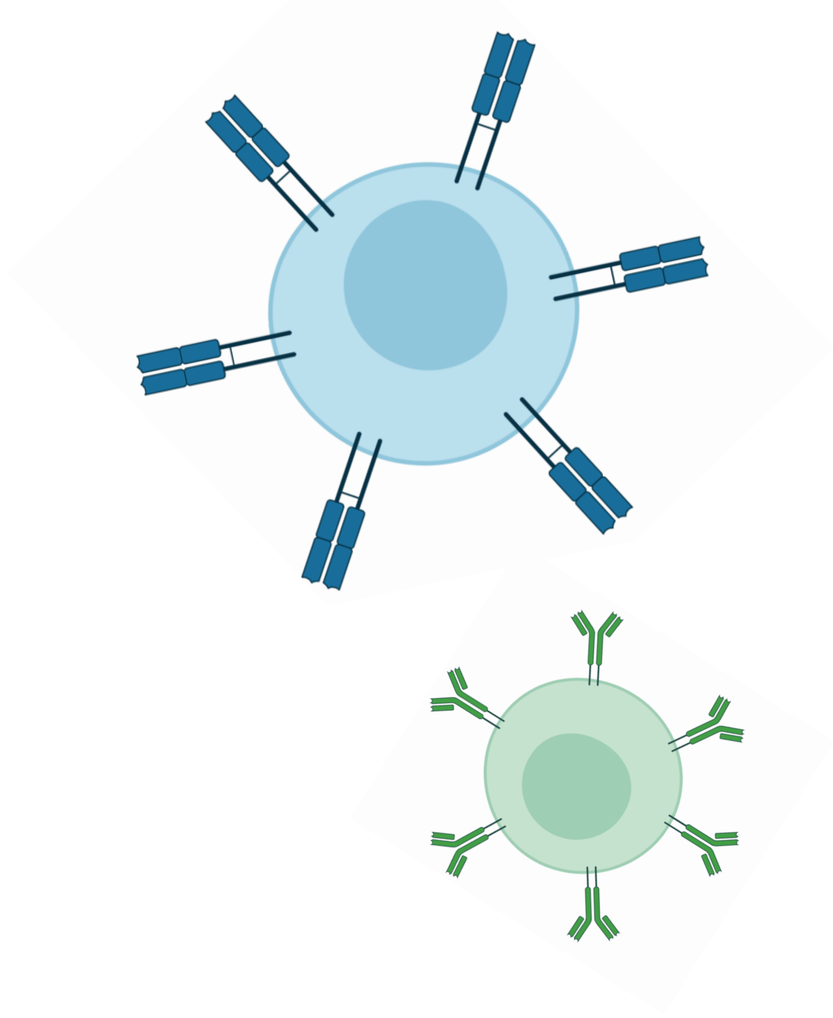
Lecture goals:

1. learn about immune repertoire sequencing

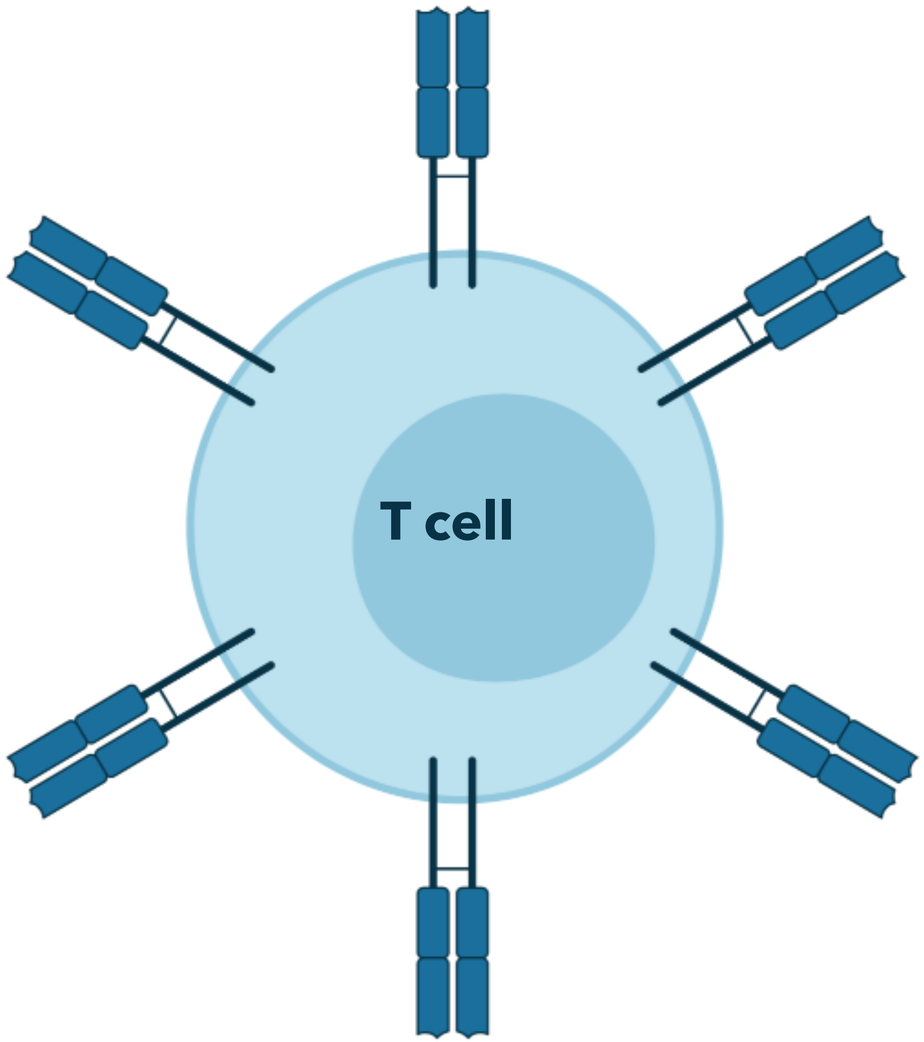


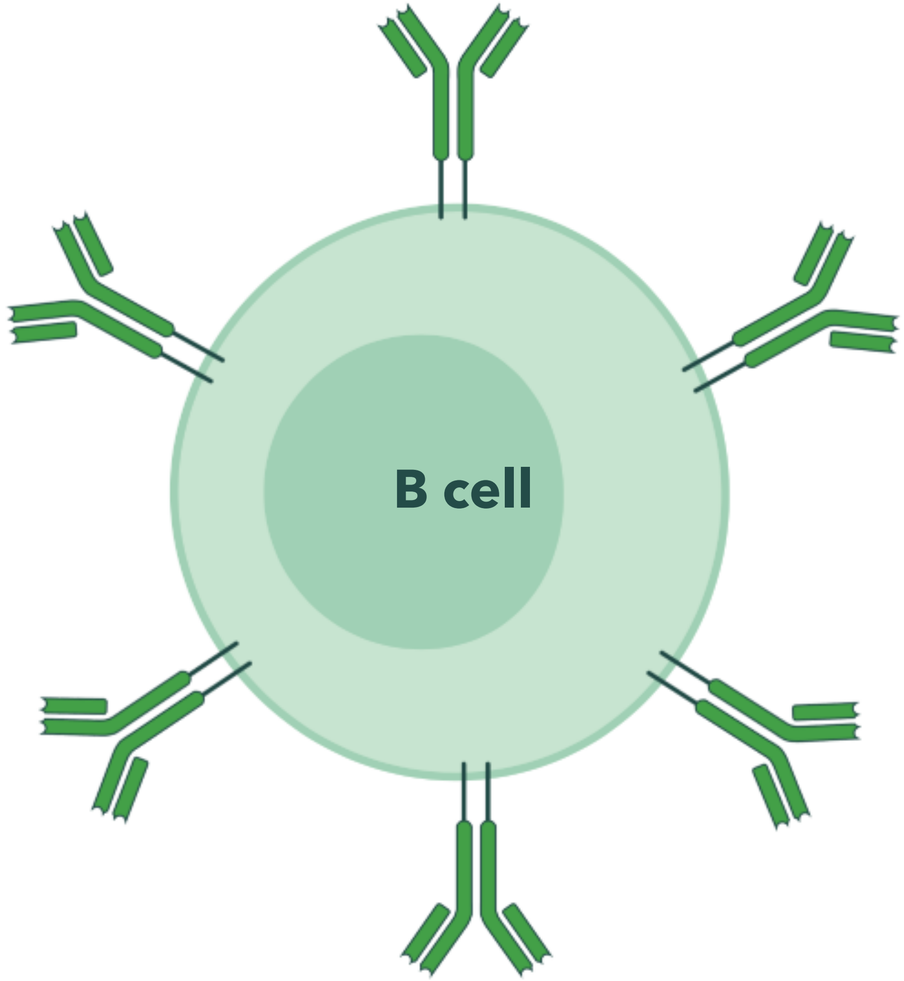
Lecture goals:

1. learn about immune repertoire sequencing
    - **what are immune repertoires?**

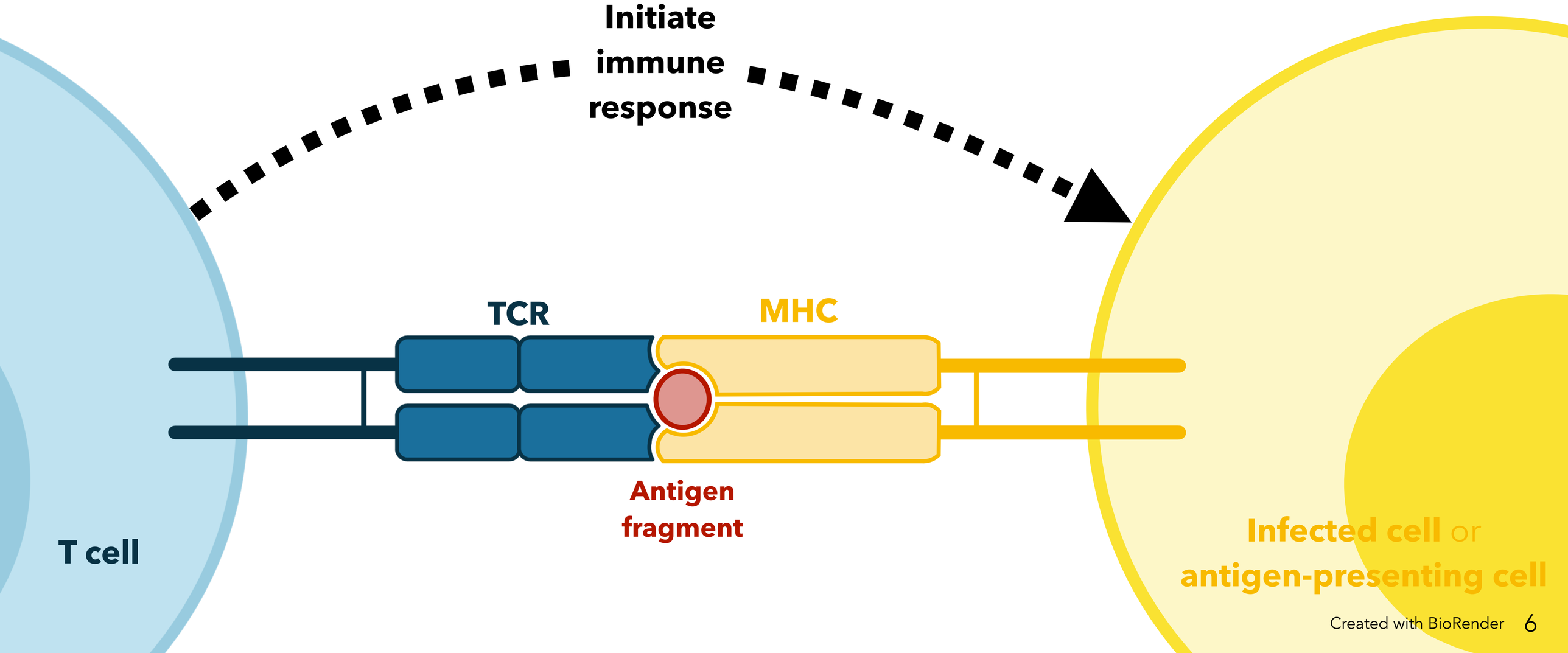


Adaptive immunity is essential for our survival

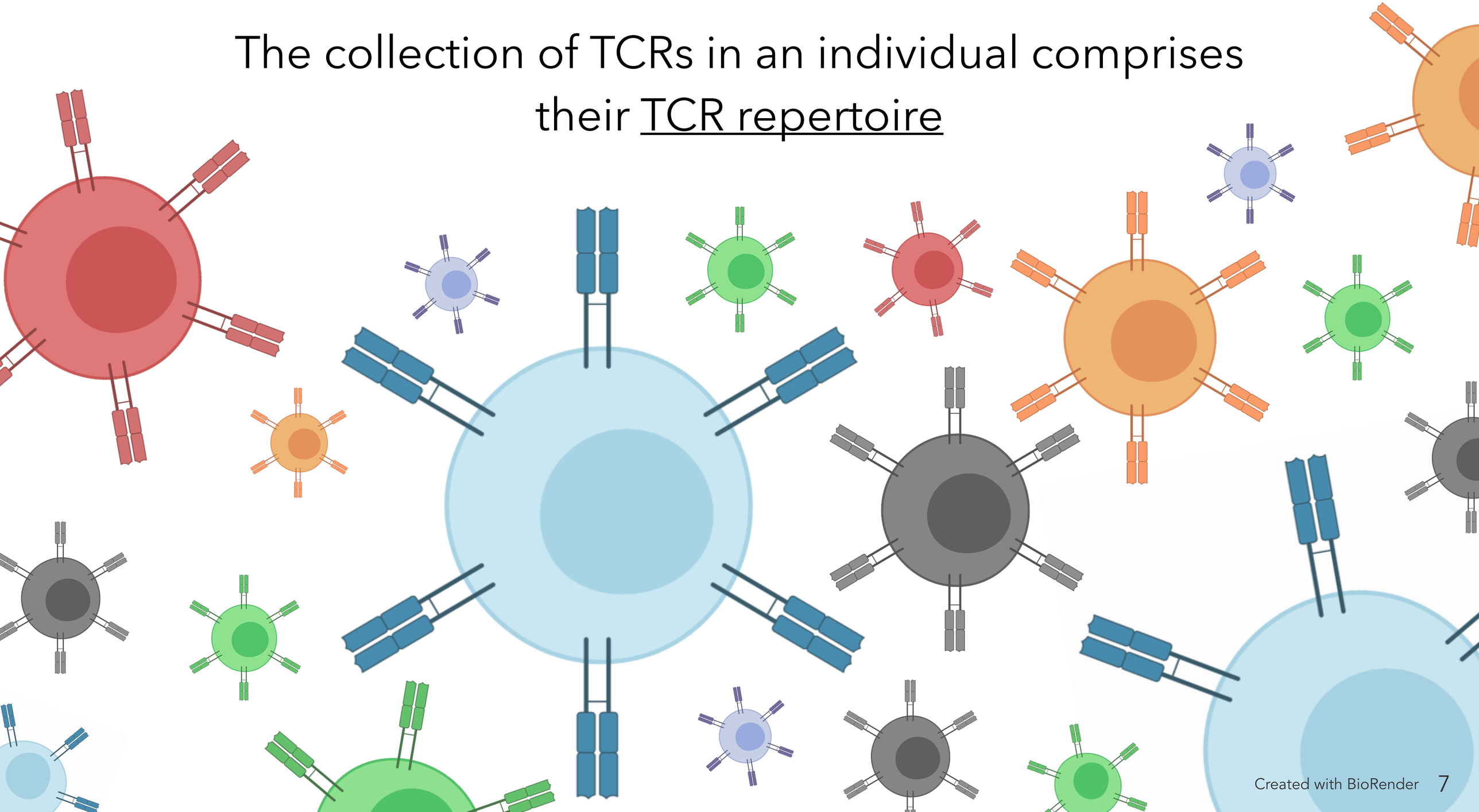




T cell receptors recognize antigen fragments bound to MHC

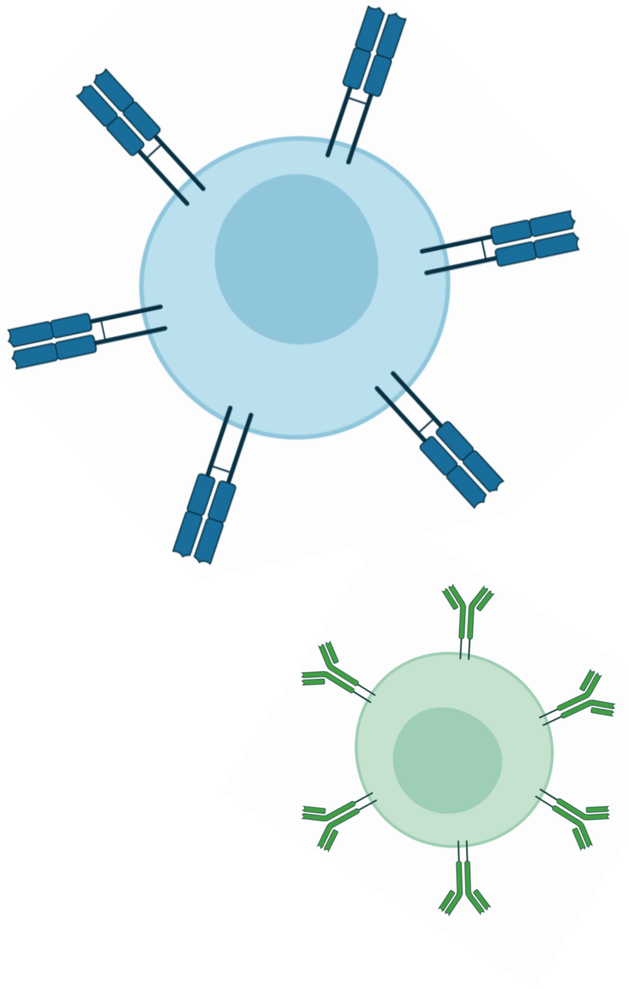


The collection of TCRs in an individual comprises their TCR repertoire



Lecture goals:

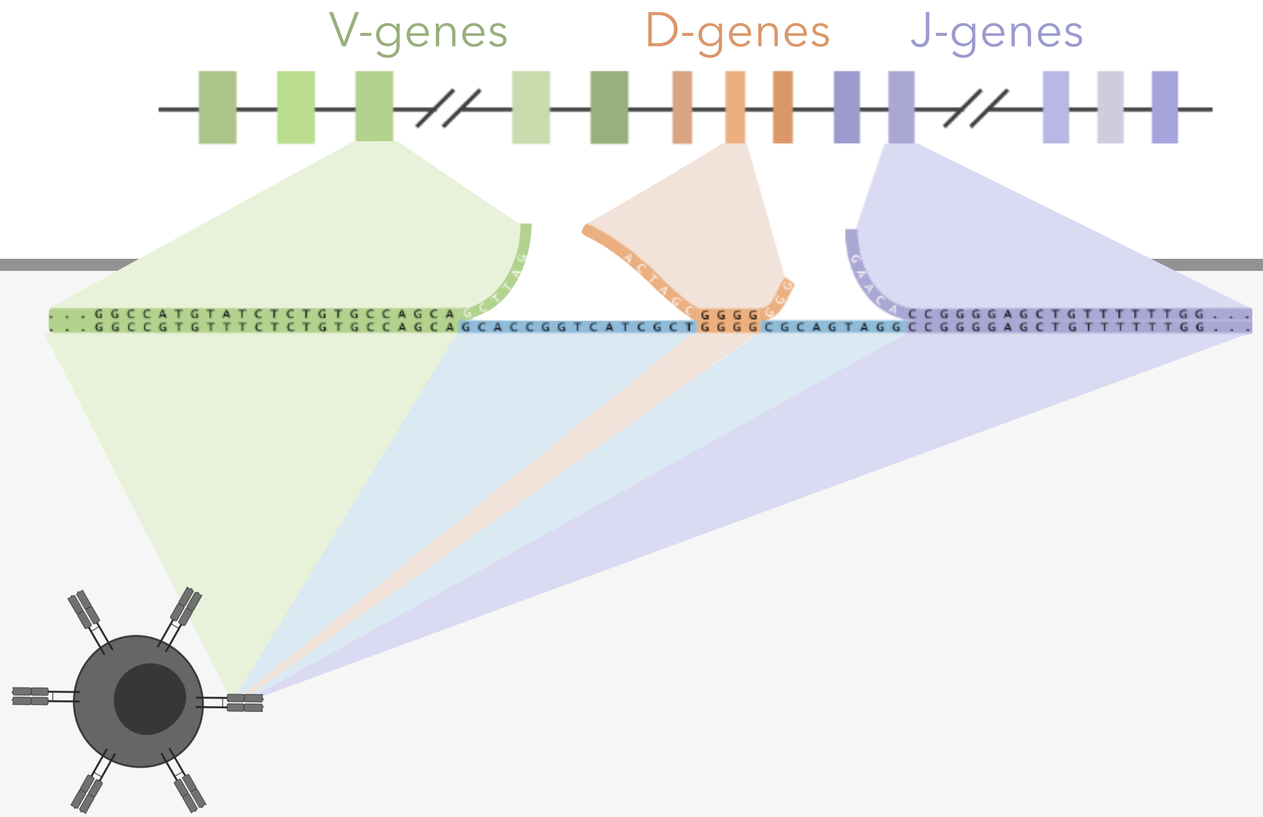
1. learn about immune repertoire sequencing
    - what are immune repertoires?
    - **how are they formed?**



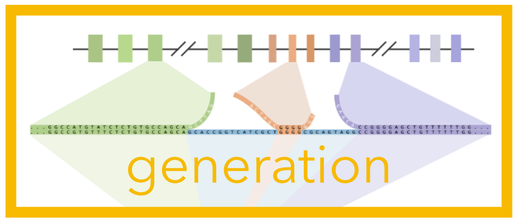
Repertoire composition is influenced by generation, selection, and exposures

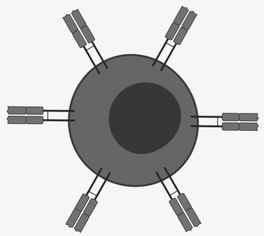
Let's use a water pipe as an analogy for TCR repertoire formation…

Repertoire composition is influenced by generation, selection, and exposures



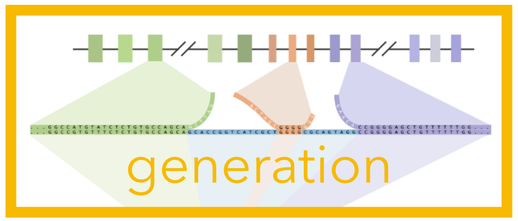
Repertoire composition is influenced by **generation**, selection, and exposures





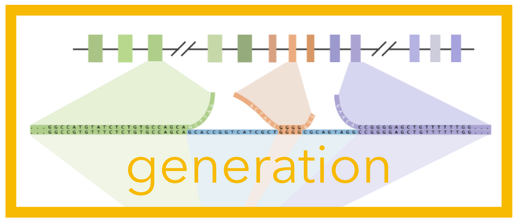
modified from Murugan et. al *PNAS* 2012
Created with BioRender

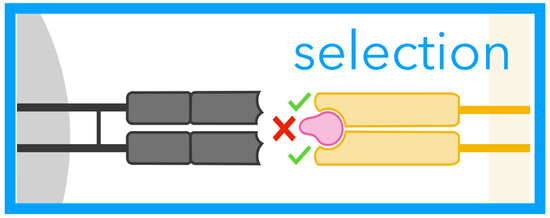
Repertoire composition is influenced by **generation**, selection, and exposures

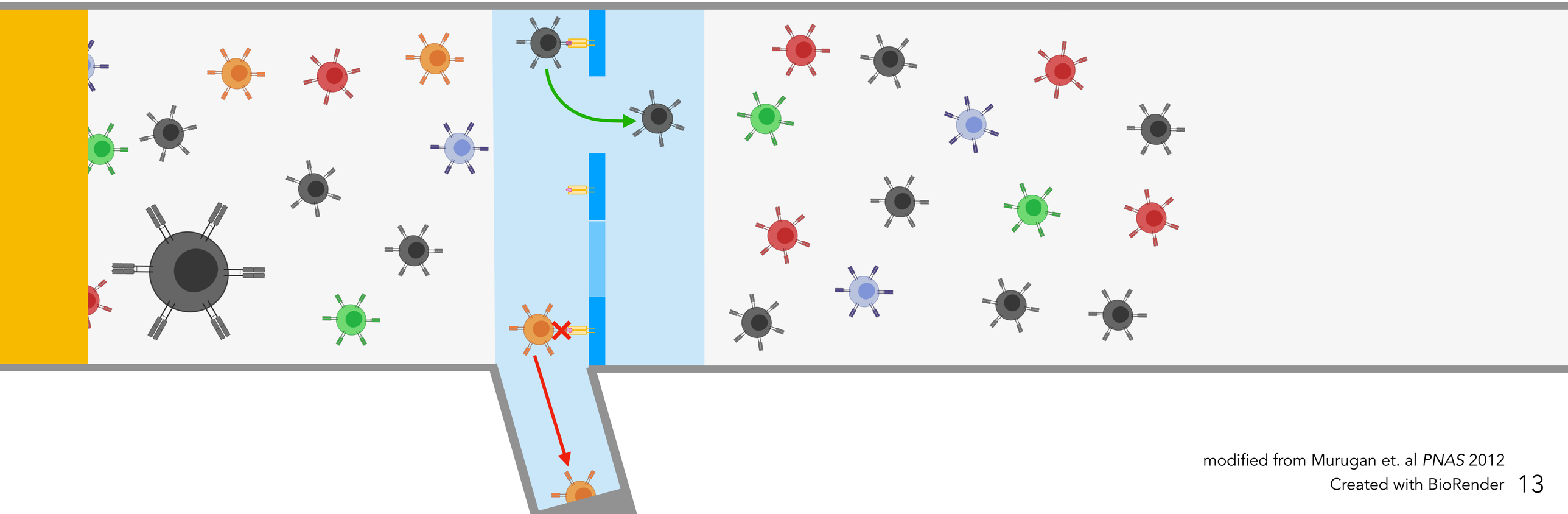


modified from Murugan et. al *PNAS* 2012
Created with BioRender

Repertoire composition is influenced by generation, **selection**, and exposures

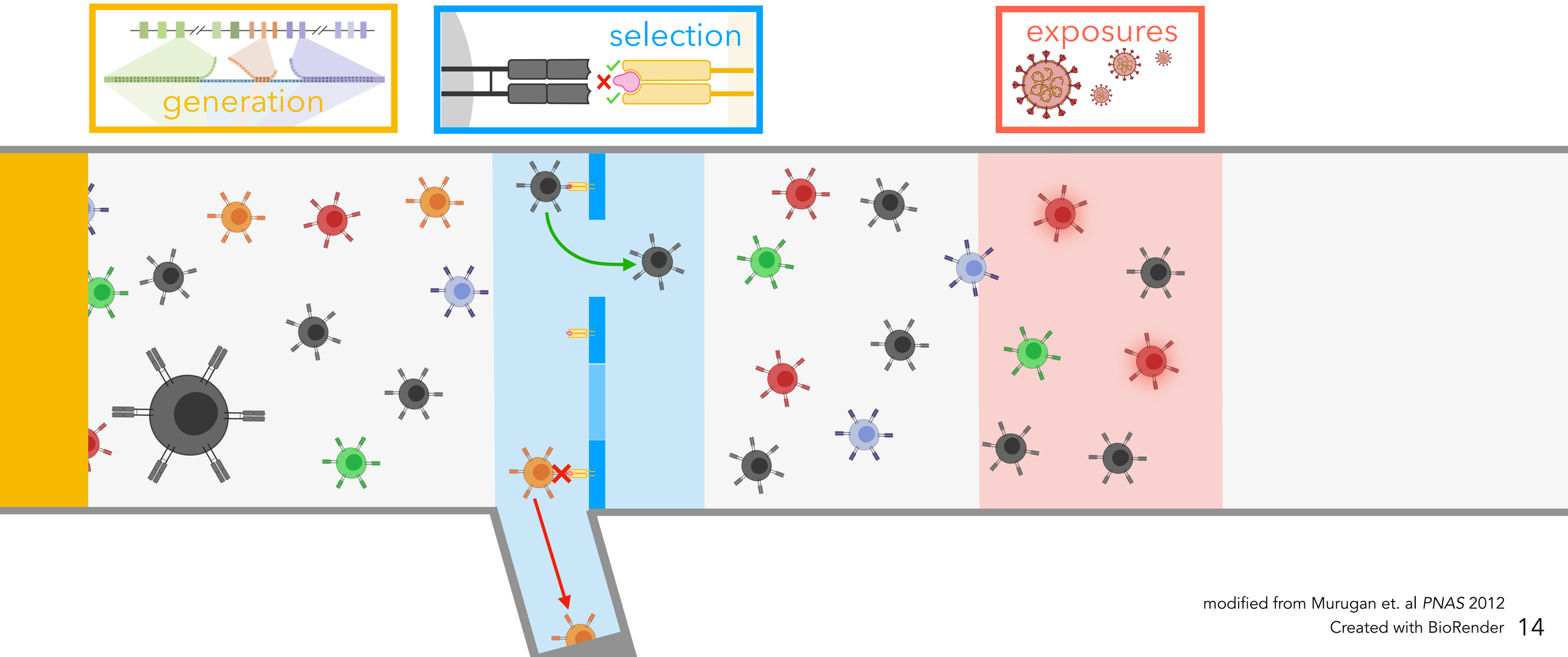






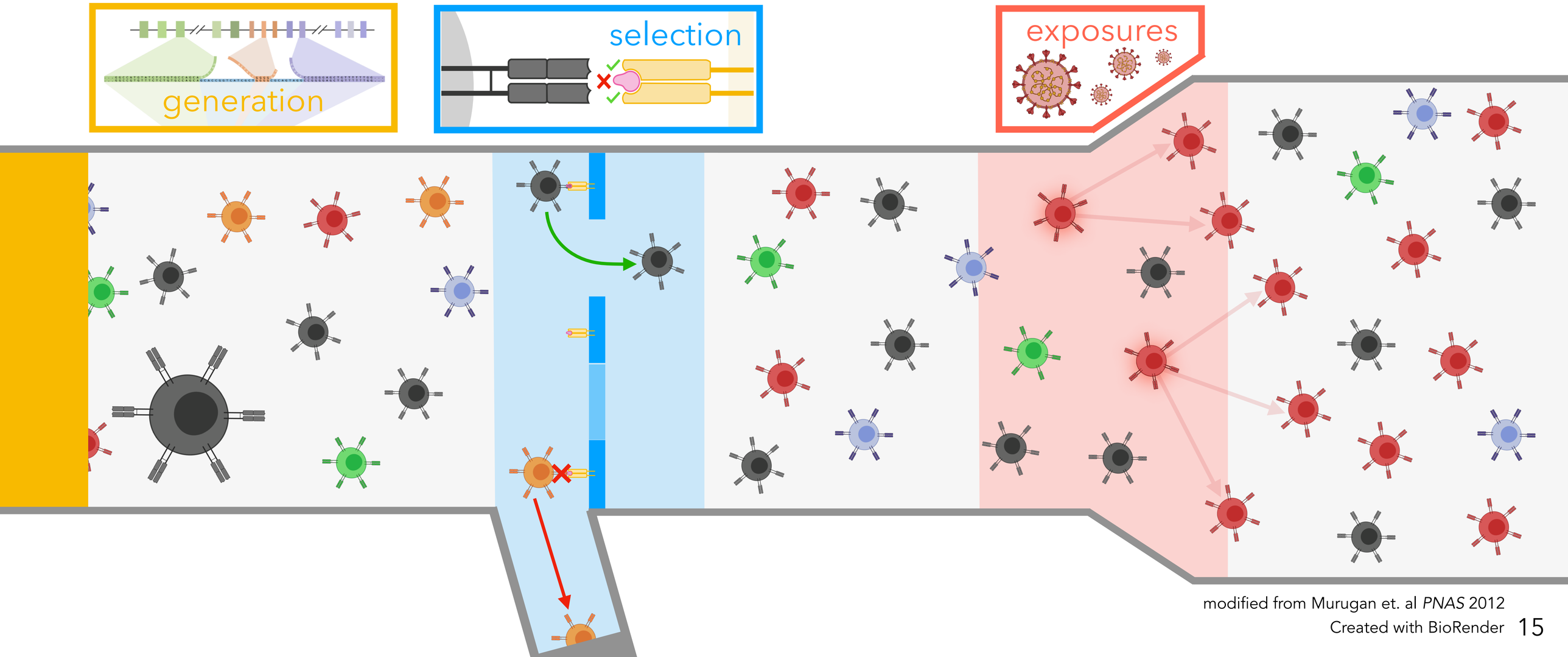
modified from Murugan et. al *PNAS* 2012
Created with BioRender

Repertoire composition is influenced by generation, selection, and **exposures**



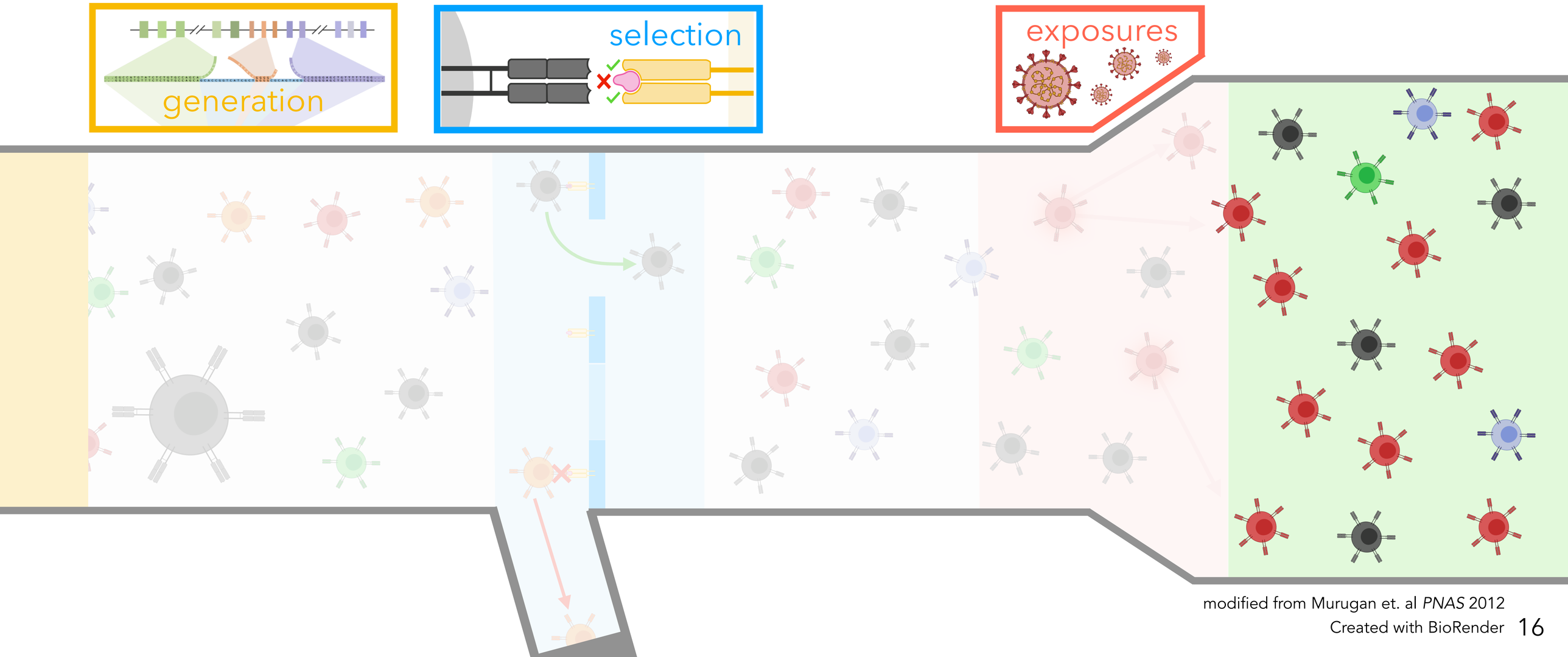
modified from Murugan et. al *PNAS* 2012
Created with BioRender

Repertoire composition is influenced by generation, selection, and **exposures**



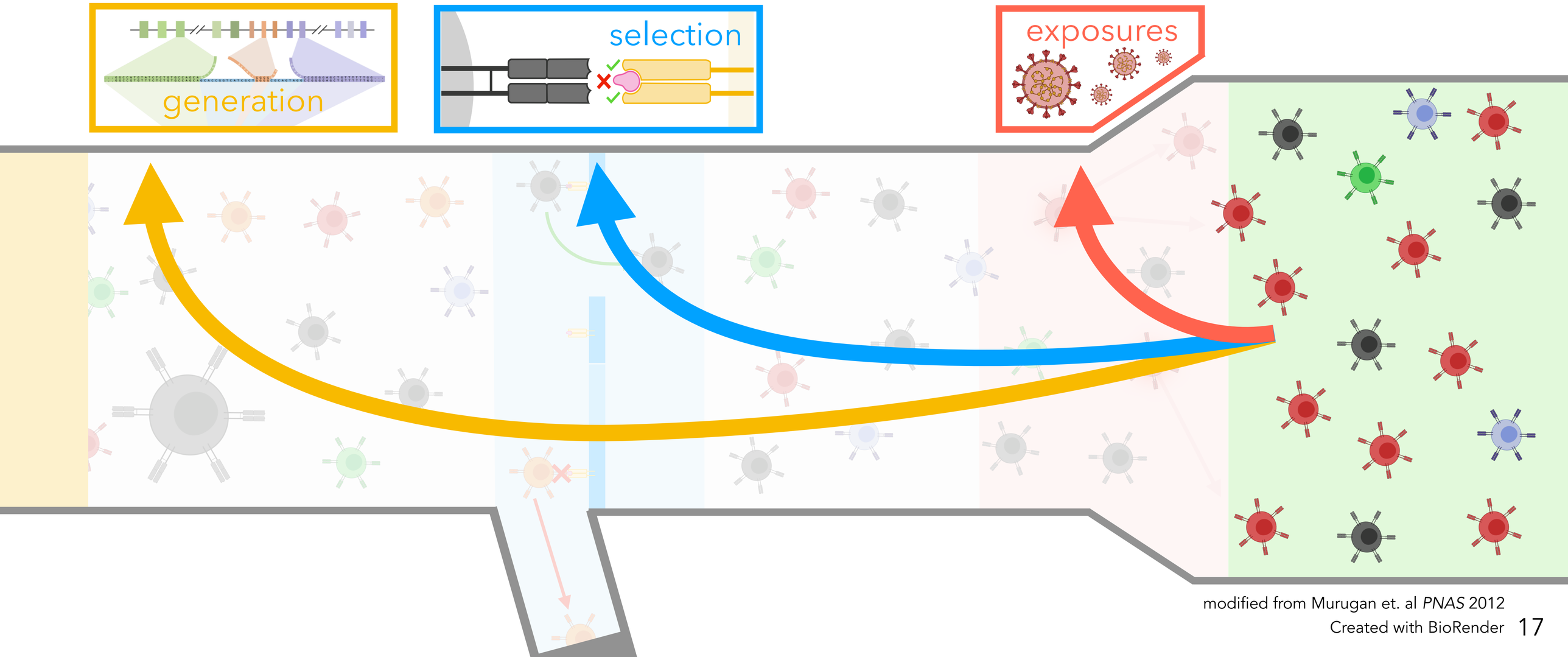
modified from Murugan et. al *PNAS* 2012
Created with BioRender

We can sample a **repertoire** using sequencing



modified from Murugan et. al *PNAS* 2012
Created with BioRender

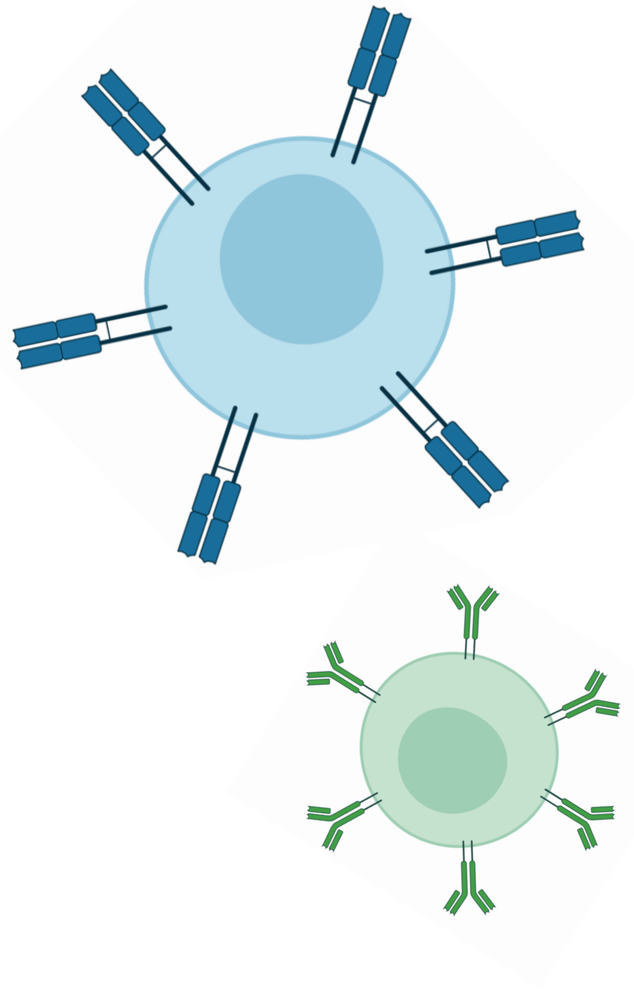
We can sample a **repertoire** using sequencing



modified from Murugan et. al *PNAS* 2012
Created with BioRender

Lecture goals:

1. learn about immune repertoire sequencing
    - what are immune repertoires?
    - how are they formed?
    - **how are they sequenced?**



Both **bulk sequencing** and **single-cell sequencing** are possible

	Bulk	**Single-cell**
Repertoire coverage *(e.g. total # of unique sequences that can be identified)*	**High**	**Low**

Both **bulk sequencing** and **single-cell sequencing** are possible

	Bulk	**Single-cell**
Repertoire coverage *(e.g. total # of unique sequences that can be identified)*	**High**	**Low**
Chain pairing *(e.g. each receptor consists of two protein chains)*	**No**	**Yes**

Both **bulk sequencing** and **single-cell sequencing** are possible

	Bulk	**Single-cell**
Repertoire coverage *(e.g. total # of unique sequences that can be identified)*	**High**	**Low**
Chain pairing *(e.g. each receptor consists of two protein chains)*	**No**	**Yes**
Sample size	**High**	**Low**

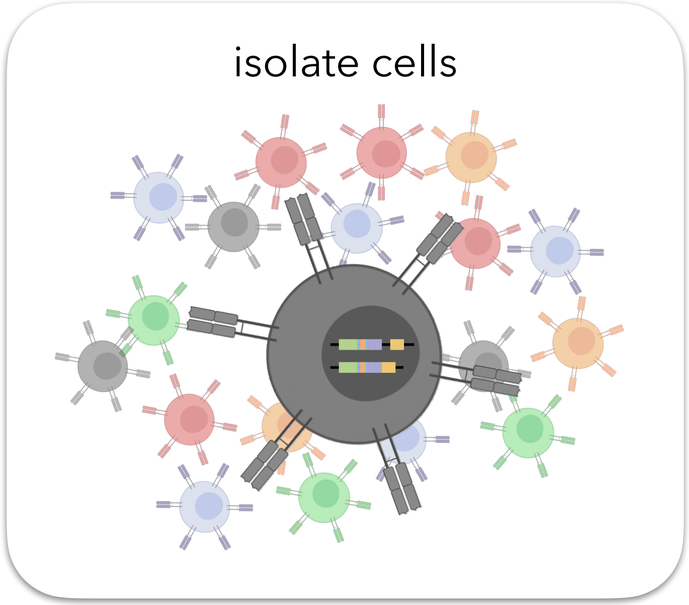
Both **bulk sequencing** and **single-cell sequencing** are possible

	Bulk	**Single-cell**
Repertoire coverage *(e.g. total # of unique sequences that can be identified)*	**High**	**Low**
Chain pairing *(e.g. each receptor consists of two protein chains)*	**No**	**Yes**
Sample size	**High**	**Low**
Cost	**Low**	**High**

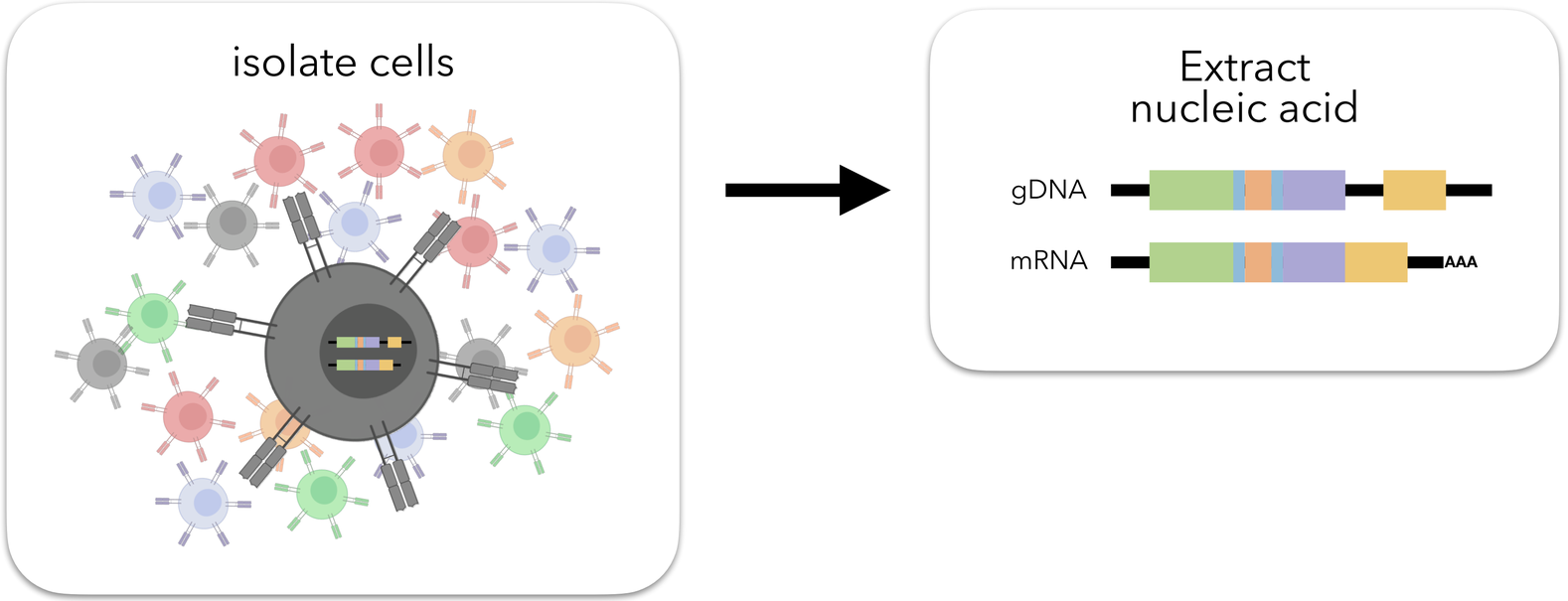
Both **bulk sequencing** and **single-cell sequencing** are possible

	Bulk	**Single-cell**
Repertoire coverage *(e.g. total # of unique sequences that can be identified)*	**High**	**Low**
Chain pairing *(e.g. each receptor consists of two protein chains)*	**No**	**Yes**
Sample size	**High**	**Low**
Cost	**Low**	**High**

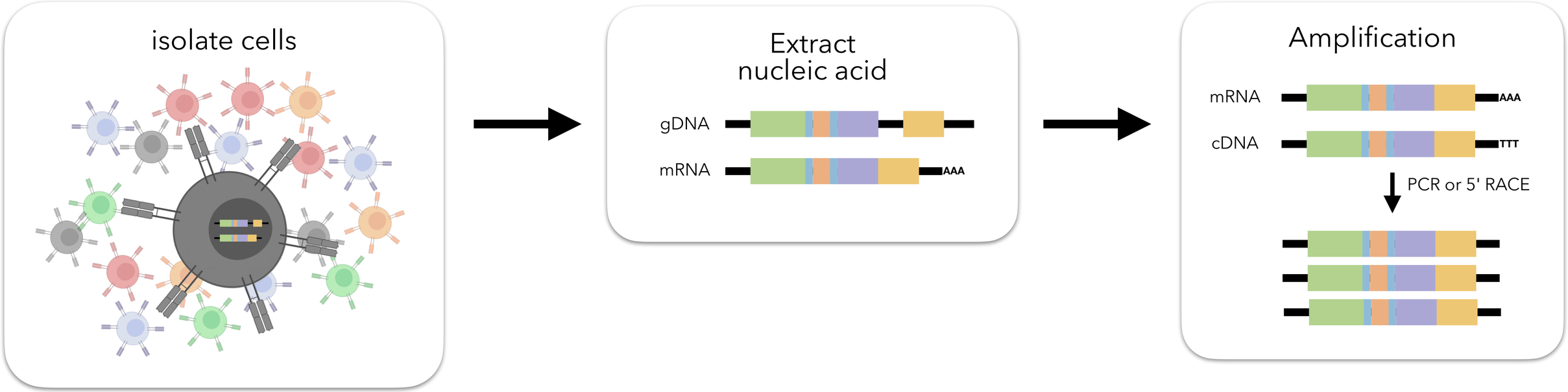
Bulk repertoire sequencing overview



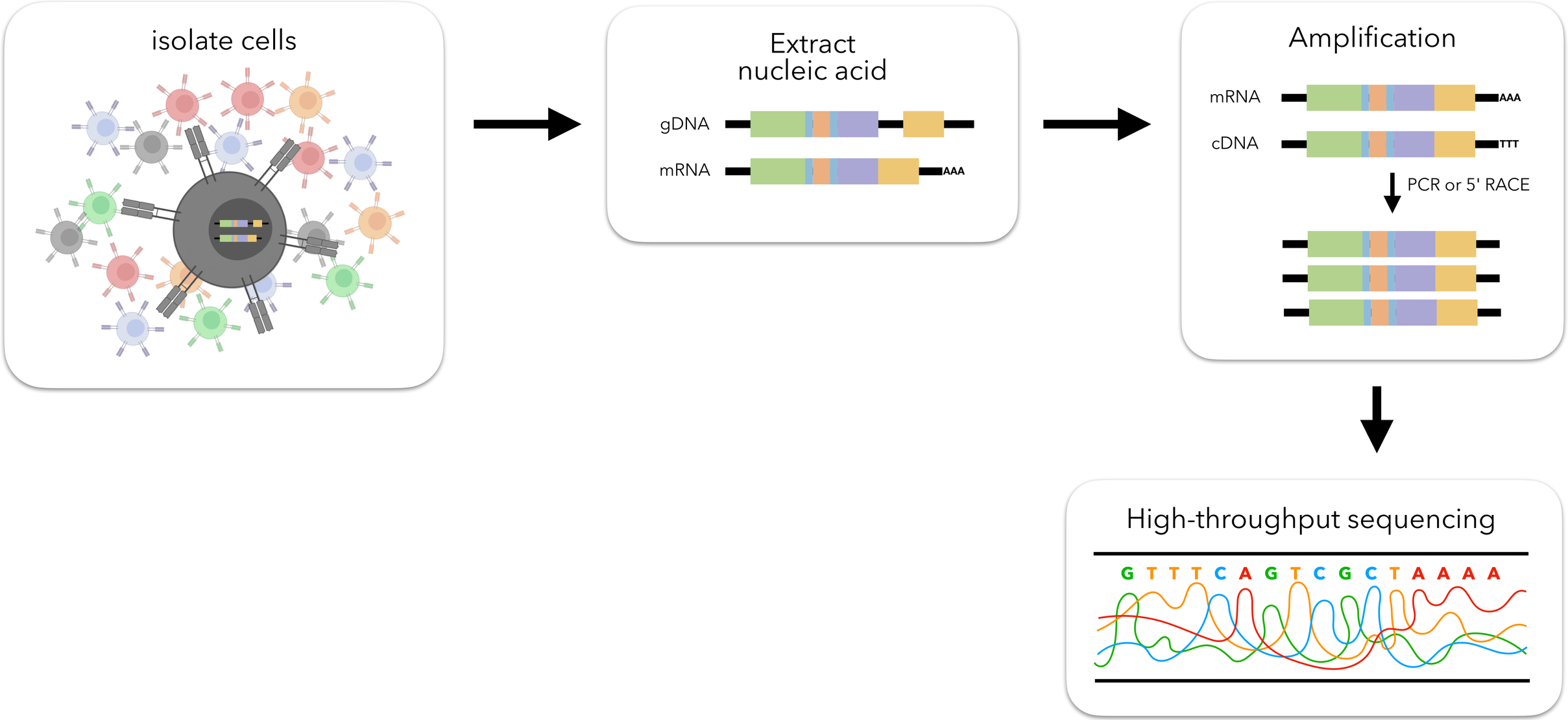
Bulk repertoire sequencing overview



Bulk repertoire sequencing overview

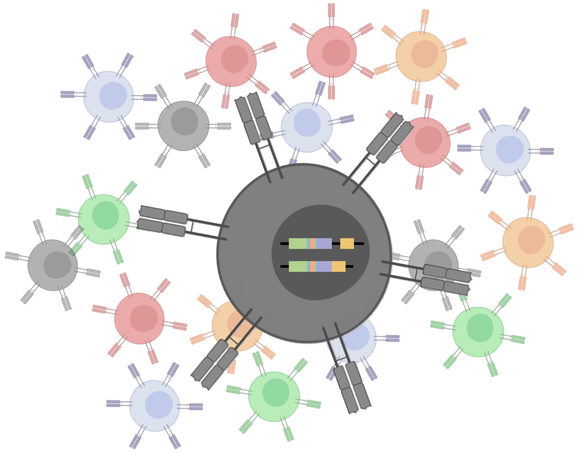


Bulk repertoire sequencing overview

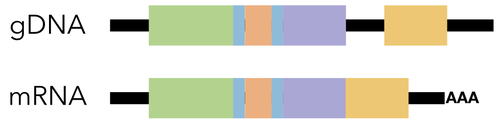


Bulk repertoire sequencing overview

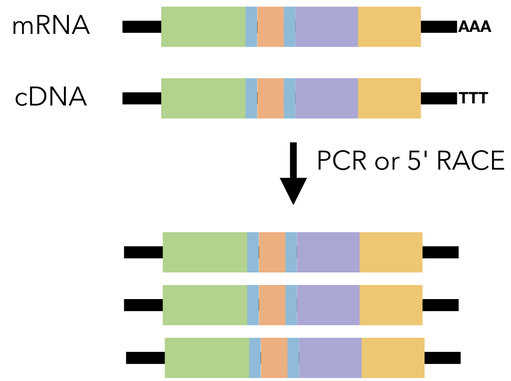
isolate cells



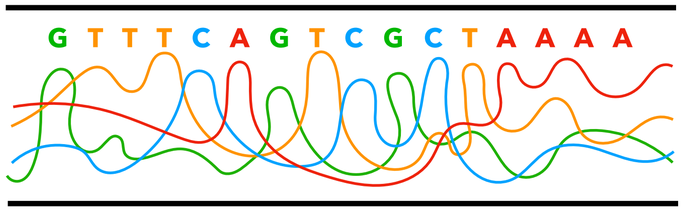
Extract
nucleic acid



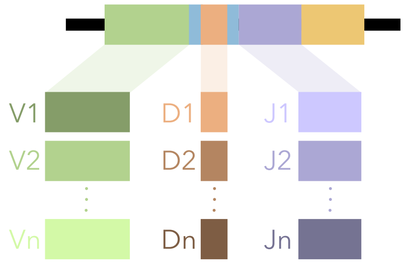
Amplification



High-throughput sequencing



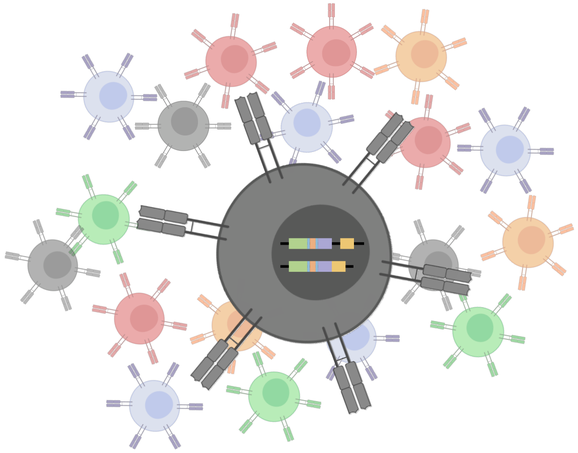
Data preprocessing



(filtering, alignment, annotation,
clustering, etc.)

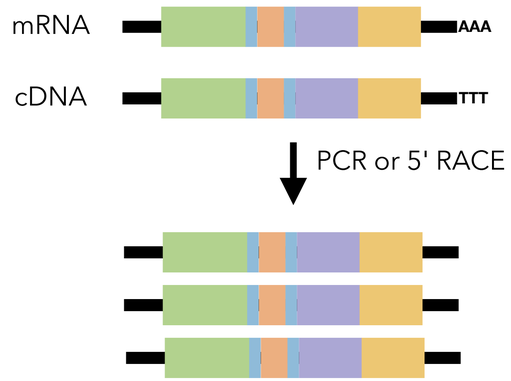
Bulk repertoire sequencing overview

isolate cells

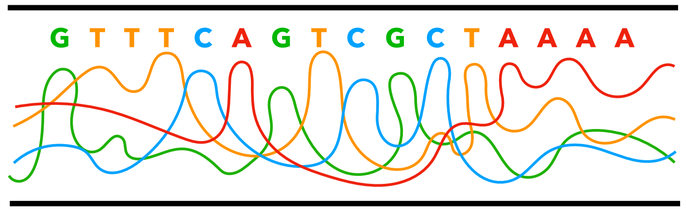


Extract
nucleic acid

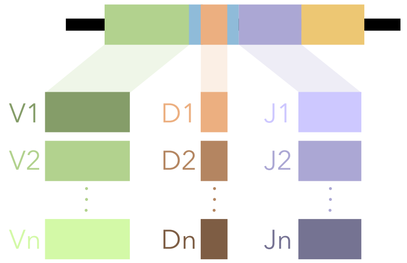
Amplification



High-throughput sequencing



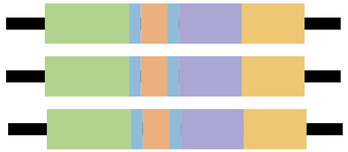
Data preprocessing



(filtering, alignment, annotation, clustering, etc.)

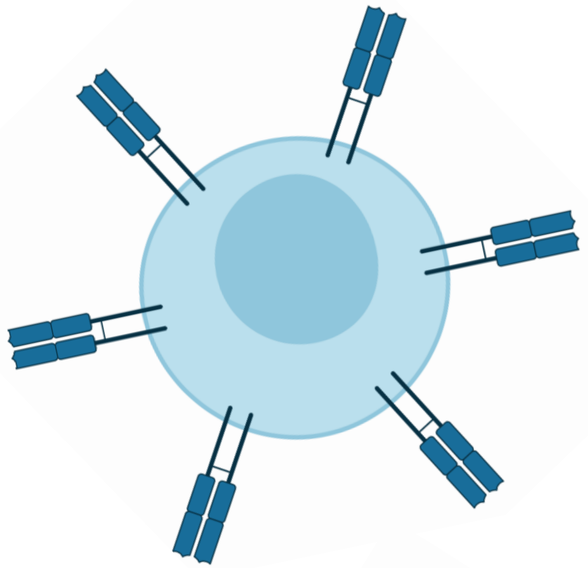
Data analysis

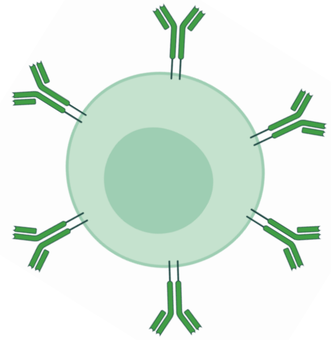
Processed **bulk repertoire sequencing** example output



cdr3_nucseq	cdr3	v_gene	d_gene	j_gene	v_trim	d0_trim	d1_trim	j_trim	vd_insert	dj_insert	vd_insert_nucs	dj_insert_nucs
<chr>	<chr>	<chr>	<chr>	<chr>	<int>	<int>	<int>	<int>	<int>	<int>	<chr>	<chr>
TGTGCCAGCAGCTTGAATCACGAGCAGTACTTC	CASSLNHEQYF	TRBV5-6*01	TRBD2*02	TRBJ2-7*01	1	3	13	5	4	0	AATC	
TGCGCCAGCAGCTTGGCAGAGACCCAGTACTTC	CASSLAETQYF	TRBV5-1*01	TRBD1*01	TRBJ2-5*01	2	9	0	4	0	0		
TGCGCCAGTCGAGCGGCGAGCTCCTACAATGAGCAGTTCTTC	CASRAASSYNEQFF	TRBV5-1*01	TRBD2*01	TRBJ2-1*01	9	6	5	0	4	2	GTCG	GC
TGTGCCAGCAGCTTAAATCTGGTGAGGTACGAGCAGTACTTC	CASSLNLVRYEQYF	TRBV7-2*01	TRBD2*02	TRBJ2-7*01	2	11	1	4	8	0	AATCTGGT	
TGTGCCTGGTCAGGGGGCCCAAACACTGAAGCTTTCTTT	CAWSGGPNTEAFF	TRBV30*01	TRBD1*01	TRBJ1-1*01	5	4	0	2	1	3	T	ACC
TGTGCCACCGAACGAGGGCCCCAAGAGACCCAGTACTTC	CATERGPQETQYF	TRBV2*03	TRBD1*01	TRBJ2-5*01	10	5	3	1	7	2	CCGAACG	CC
TGTGCCAGCATAGCGGGAGGTGAGCAGTTCTTC	CASIAGGEQFF	TRBV28*01	TRBD2*02	TRBJ2-1*01	7	6	3	9	1	2	T	GG
TGTGCCTGGAGCTCCCTCCCTGGCGGGGAGAACAATGAGCAGTTCTTC	CAWSSLPGGENNEQFF	TRBV30*01	TRBD2*01	TRBJ2-1*01	3	7	3	5	11	3	CTCCCTCCCTG	AGA
TGTGCCAGCAGTTATCAGGTCACTGAAGCTTTCTTT	CASSYQVTEAFF	TRBV6-6*02	TRBD1*01	TRBJ1-1*01	4	4	5	4	2	2	AT	TG
TGTGCCAGCGGCCCAGGGCTCGGATACAATCAGCCCCAGCATTTT	CASGPGLGYNQPQHF	TRBV5-5*01	TRBD2*02	TRBJ1-5*01	7	12	0	3	5	8	GGCCC	ATAGGCTC
TGTGCCAGTGCGGGATTCTATGGCTACACCTTC	CASAGFYGYTF	TRBV6-1*01	TRBD1*01	TRBJ1-2*01	9	7	2	4	3	3	TGC	TTA
TGTGCCAGTGCAGGGTACACCGGGGAGCTGTTTTTT	CASAGYTGELFF	TRBV2*03	TRBD1*01	TRBJ2-2*01	9	4	4	3	2	2	TG	TG
TGTGCCATCAGTGAATACAATGAGCAGTTCTTC	CAISEYNEQFF	TRBV10-3*01	TRBD1*01	TRBJ2-1*01	3	3	6	8	2	0	AT	
TGTGCCATCAGTAACACCGGGGAGCTGTTTTTT	CAISNTGELFF	TRBV10-3*02	TRBD2*02	TRBJ2-2*01	6	5	10	2	0	0		
TGTGCCAGTAGCCCTACCCGGTCTGGAAACACCATATATTTT	CASSPTRSGNTIYF	TRBV19*01	TRBD2*02	TRBJ1-3*01	7	5	9	1	4	5	GCCC	GGCCC
TGTGCCACCAGCAGAGGATCGGGGCTAGCGGGTGTTGAGCAGTTCTTC	CATSRGSGLAGVEQFF	TRBV15*02	TRBD2*01	TRBJ2-1*01	1	4	4	9	8	3	GATCGGGG	TGT
TGTGCCAGCAGCTTAACGGTGGGGTCAGGAGAGACCCAGTACTTC	CASSLTVGSGETQYF	TRBV7-3*01	TRBD2*01	TRBJ2-5*01	1	9	3	4	4	5	CGGT	GGACT

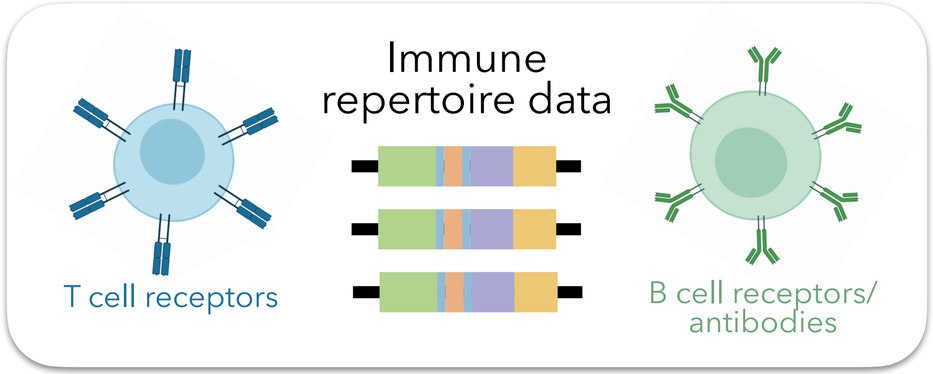
Lecture goals:





1. learn about immune repertoire sequencing
    - what are immune repertoires?
    - how are they formed?
    - how are they sequenced?
    - **what are some common areas of repertoire analysis?**

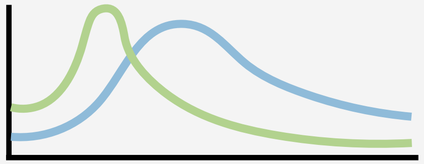
Immune repertoire analyses often focus on diversity, architecture, evolution, or convergence



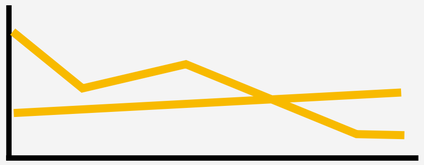
Immune repertoire analyses often focus on **diversity**, architecture, evolution, or convergence

Underlying mechanisms of diversity generation

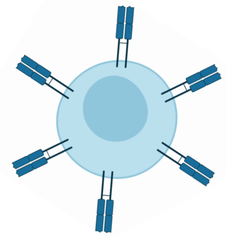
probabilistic sequence annotation



recombination statistics to learn about generation and selection

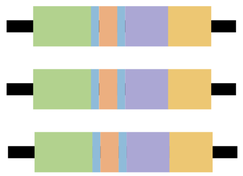


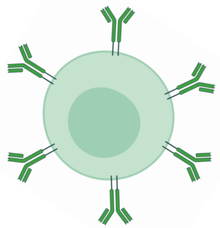
clonotype diversity dynamics



T cell receptors

Immune repertoire data

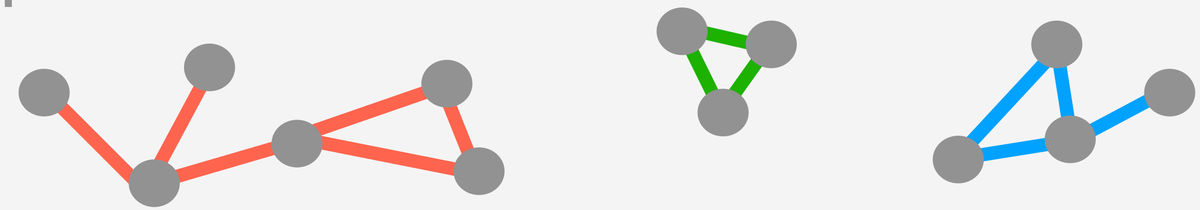




B cell receptors/ antibodies

Immune repertoire analyses often focus on diversity, **architecture**, evolution, or convergence

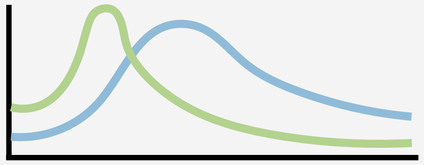
Repertoire architecture



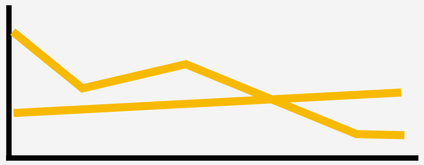
defining antigen recognition breadth using network analysis

Underlying mechanisms of diversity generation

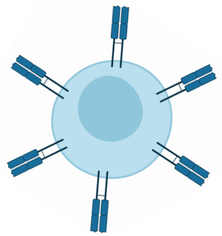
probabilistic sequence annotation



recombination statistics to learn about generation and selection

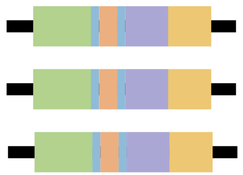


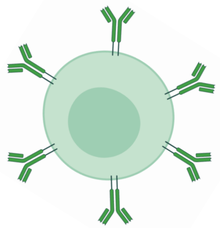
clonotype diversity dynamics



T cell receptors

Immune repertoire data

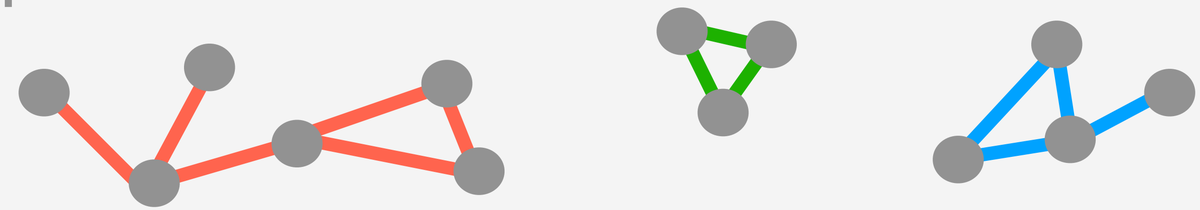




B cell receptors/ antibodies

Immune repertoire analyses often focus on diversity, architecture, **evolution**, or convergence

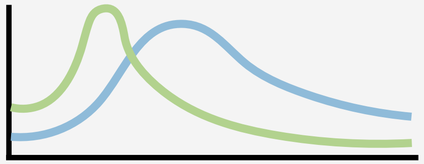
Repertoire architecture



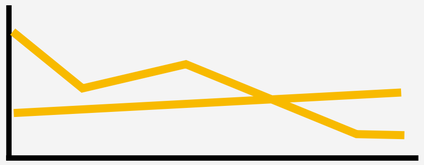
defining antigen recognition breadth using network analysis

Underlying mechanisms of diversity generation

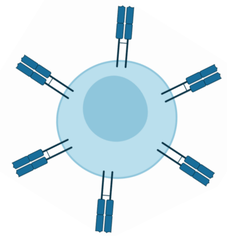
probabilistic sequence annotation



recombination statistics to learn about generation and selection

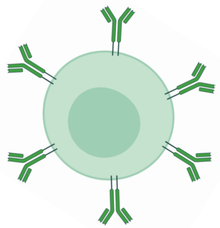


clonotype diversity dynamics



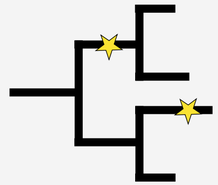
T cell receptors

Immune repertoire data

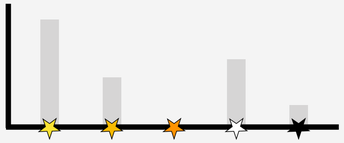


B cell receptors/ antibodies

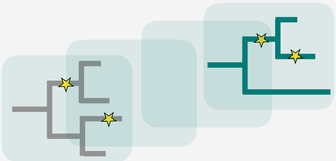
Antibody evolution



reconstruction of phylogenetic trees



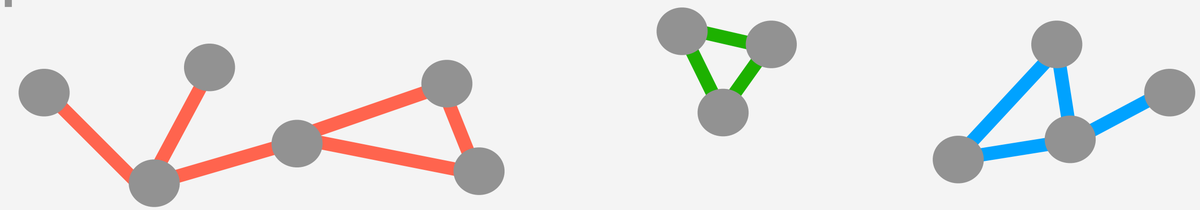
inferring mutation statistics



simulating of antibody repertoire evolution

Immune repertoire analyses often focus on diversity, architecture, evolution, or **convergence**

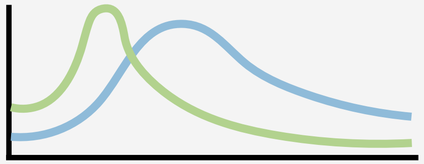
Repertoire architecture



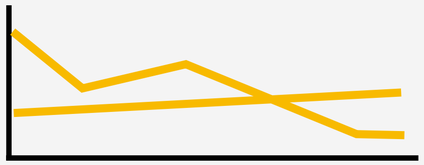
defining antigen recognition breadth using network analysis

Underlying mechanisms of diversity generation

probabilistic sequence annotation

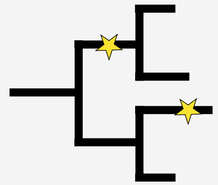


recombination statistics to learn about generation and selection

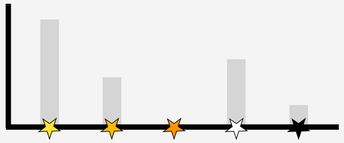


clonotype diversity dynamics

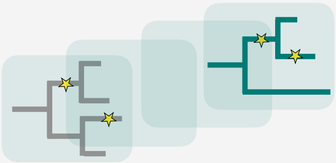
Antibody evolution



reconstruction of phylogenetic trees



inferring mutation statistics



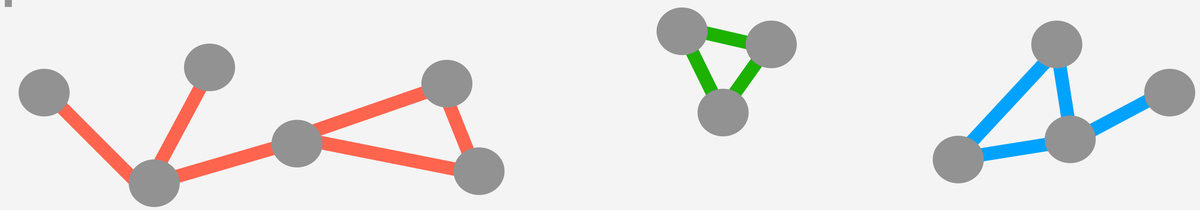
simulating of antibody repertoire evolution

Molecular convergence

exploring cross-individual sequence similarity and convergence

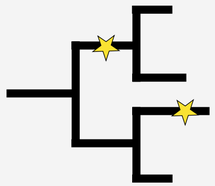
Immune repertoire analyses often focus on diversity, architecture, evolution, or convergence

Repertoire architecture

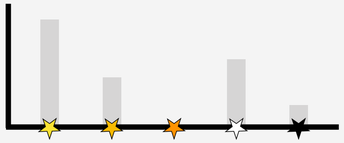


defining antigen recognition breadth using network analysis

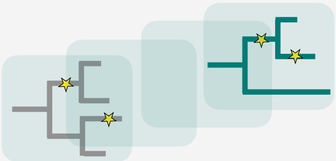
Antibody evolution



reconstruction of phylogenetic trees



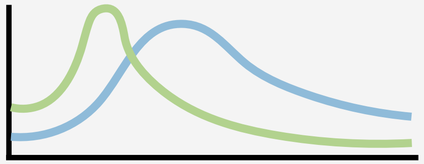
inferring mutation statistics



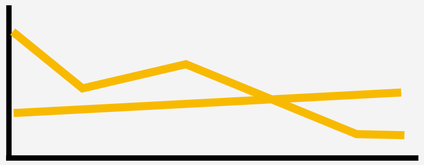
simulating of antibody repertoire evolution

Underlying mechanisms of diversity generation

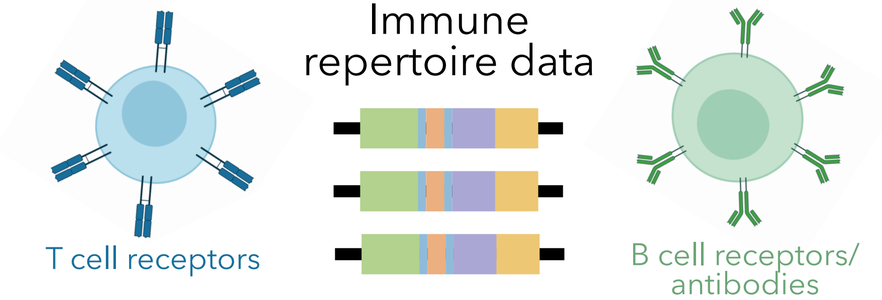
probabilistic sequence annotation



recombination statistics to learn about generation and selection



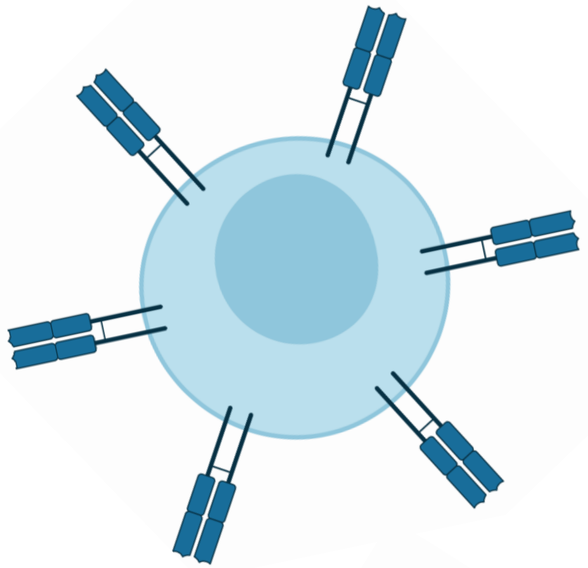
clonotype diversity dynamics

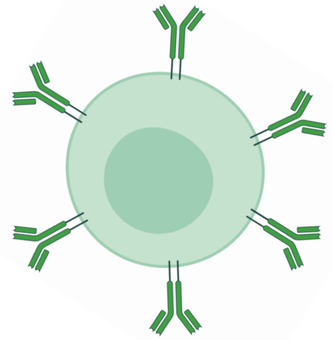


Molecular convergence

exploring cross-individual sequence similarity and convergence

Lecture goals:





1. learn about immune repertoire sequencing
    - what are immune repertoires?
    - how are they formed?
    - how are they sequenced?
    - what are some common areas of repertoire analysis?
2. familiarize with immune repertoire data
3. work through an example analysis