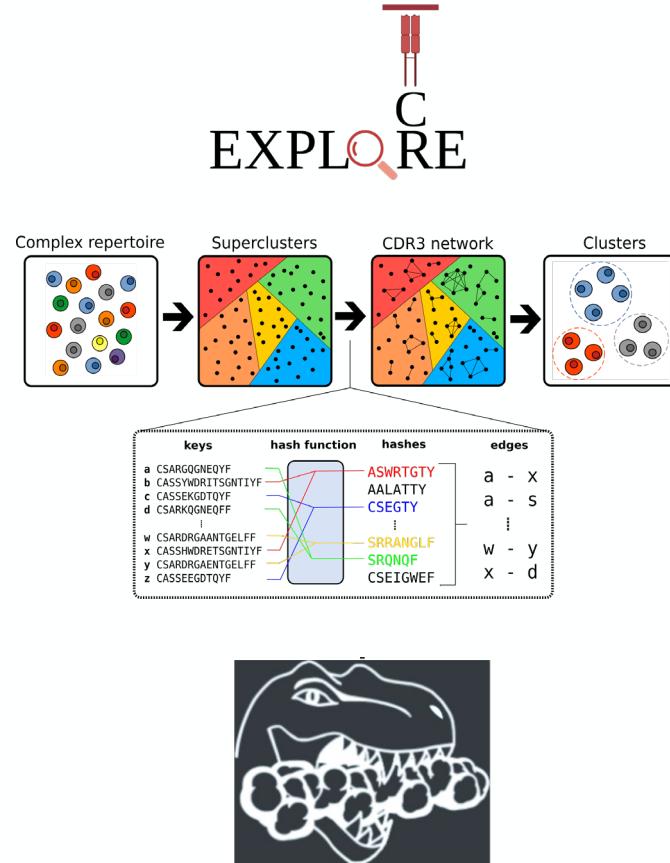


Overview of TCR workshop

- Introduction
 - 2:00 – 2:15 pm
- TCR-repertoire analysis with TCR_Explore
 - 2:15 – 3:00 pm
- Break (10min)
- Understanding diversity analysis and clustering
 - 3:10 – 4:00 pm
- Break (10min)
- Epitope prediction with TCRex
 - 4:10 – 5:00 pm
- Conclusion
 - 5:00 – 5:15 pm



Introduction

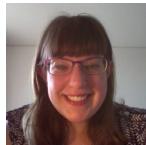
2:00 - 2:15pm



Kris Laukens lab
part of Adrem Data Lab
dept. Computer Sciences



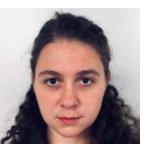
post-doctoral scientists



managers



pre-doctoral scientists



former PhD graduates



Scope: bioinformatics & biomedical data sciences

dept. Computer Sciences



University of Antwerp
| Adrem | Adrem Data Lab

MISSION: study of data science technology (information theory, data mining, artificial intelligence) to make sense of biomedical data



University of Antwerp
Precision Medicine Technologies
IOF Consortium



AUDACIS - Antwerp Unit for Data Analysis and
Computation in Immunology and Sequencing
University of Antwerp



University of Antwerp
MST | Microbial Systems Technology
Centre of Excellence



biomedical informatics
network Antwerpen

collaborations:



We are here!



Adrem
Adrem Data Lab
University of Antwerp

dept. Computer Science

biomina

bio- / medical
informatics network

What we do: **bio-data science**

algorithms	mining	applications
unsupervised pattern discovery	spectral data	infectious diseases
machine learning & deep learning	omics data	intensive care
	health data	immunology
	sensor data	oncology
		toxicology
		plant biotech
		pharmacy

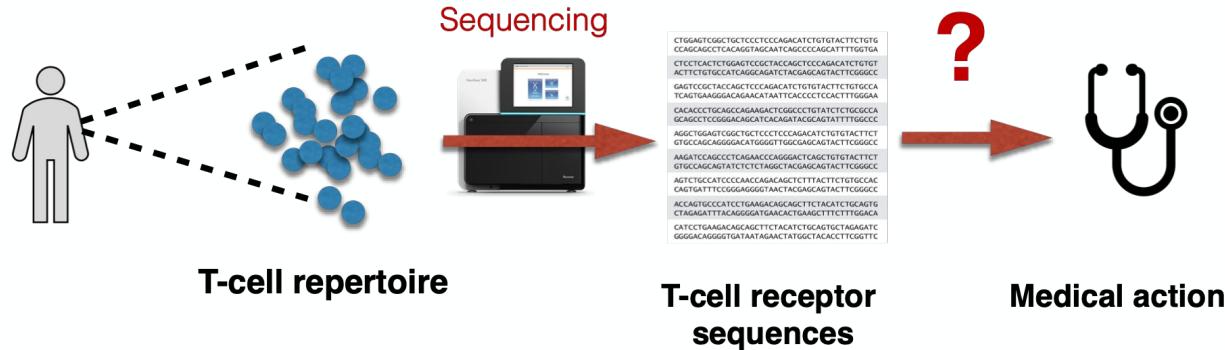


Collaborations:



industry

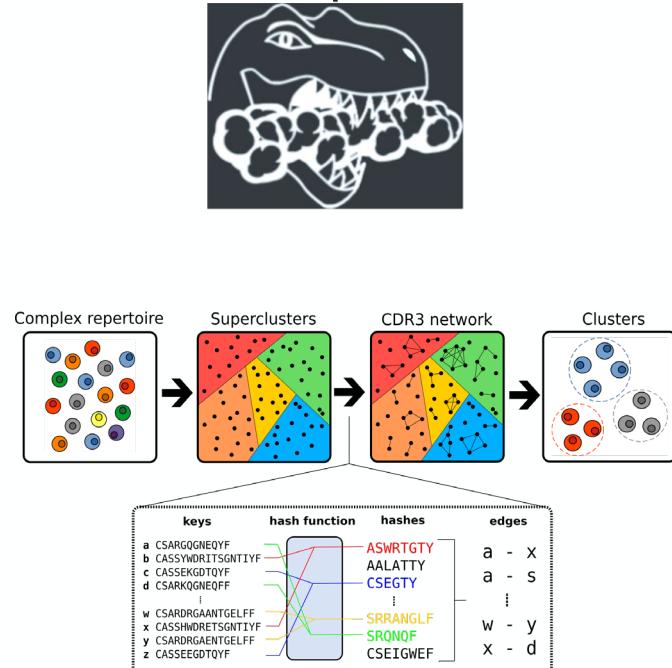
Machine learning to translate immune system sequences into clinical decisions



- The immune system of each individual is unique.
- It is composed of complex mix of white blood cells with unique genetic code
- We can now read this code, and "decode" it into an immune-map.

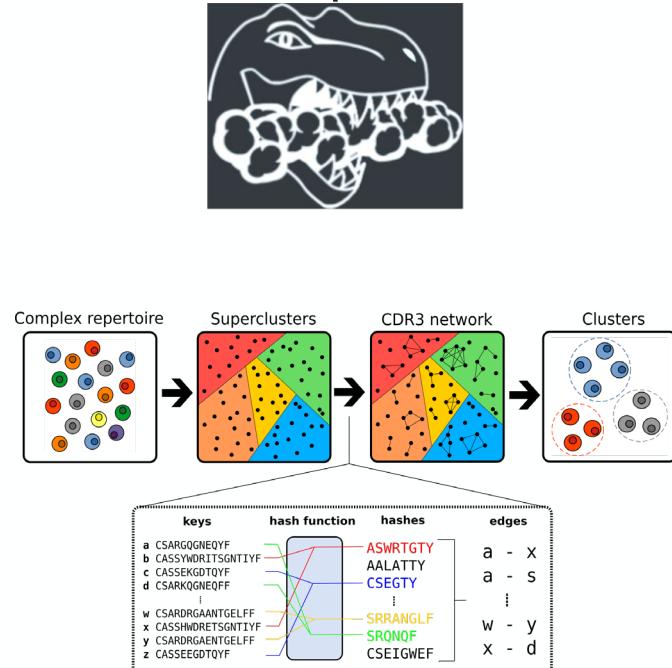
Machine learning to translate immune system sequences into clinical decisions

- Algorithms to decode the human immune receptor repertoire to predict, model and monitor
 - TCRex - Link TCR sequences to known epitopes
 - ClusTCR - Group TCR sequences by putative targets
 - ImReX - Link TCR sequences to unseen epitopes
- T cell / B cell immune map gives insight in
 - Past – past exposure to epitopes
 - Present – Monitor immune response to vaccination or active infection
 - Future – Predict response to epitope exposure



Machine learning to translate immune system sequences into clinical decisions

- Algorithms to decode the human immune receptor repertoire to predict, model and monitor
 - TCRex - Link TCR sequences to known epitopes
 - ClusTCR - Group TCR sequences by putative targets
 - ImReX - Link TCR sequences to unseen epitopes
- T cell / B cell immune map gives insight in
 - Past - past exposure to epitopes
 - Present - Monitor immune response to vaccination or active infection
 - Future - Predict response to epitope exposure

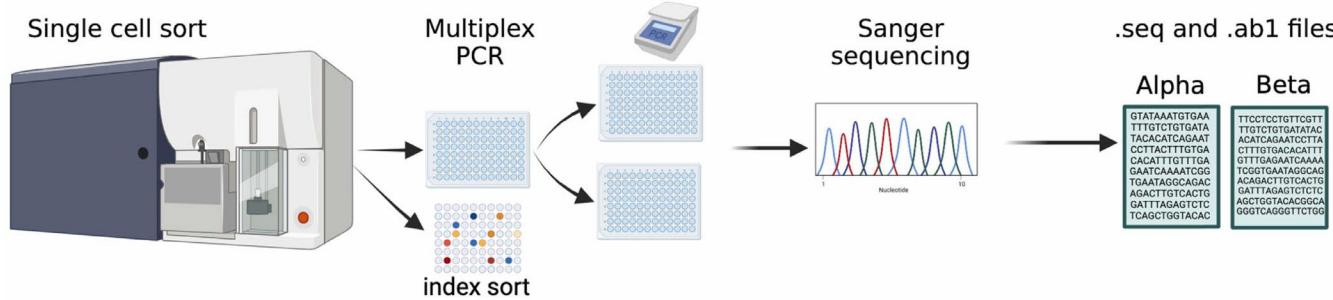


Types of experimental processes

- Paired Sanger Sequencing (Focus experiments)
- Bulk TCR-seq (Depth of repertoire)

Types of experimental processes

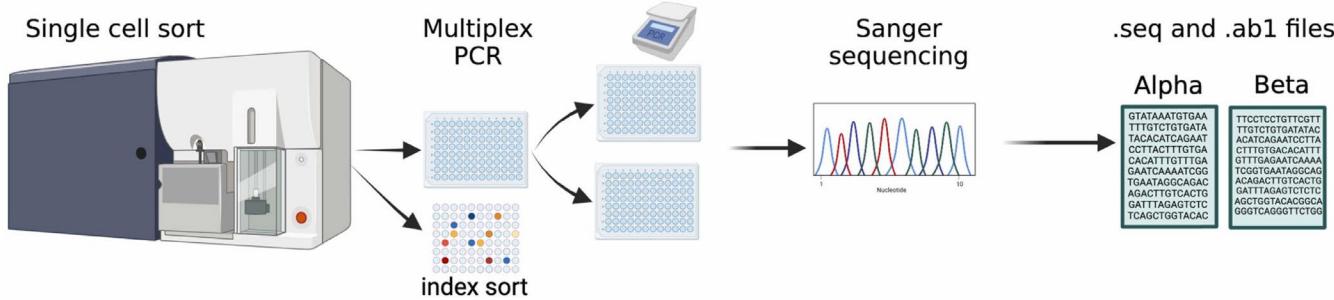
- Paired Sanger Sequencing (Focus experiments)



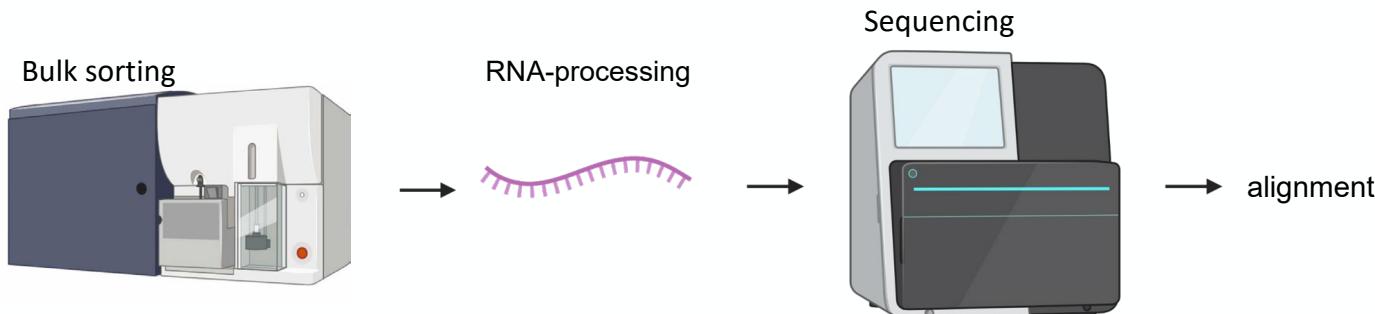
- Bulk TCR-seq (Depth of repertoire)

Types of experimental processes

- Paired Sanger Sequencing (Focus experiments)



- Bulk TCR-seq (Depth of repertoire)



Consideration for experimental questions

	Sanger sequencing	Bulk sequencing
Chain analysis	Paired chain	Single chain
# sequences	Up to 80 per plate Higher resolution	100,000 Depth of repertoire
Accuracy	Higher as no reconstruction of sequence	Lower to moderate Depends on protocol (alignment assumptions)
Time	Lab intensive	Less time intensive
Marker interrogation	Single cell and Protein expression Marker validation	Depends on sort (average within sample)

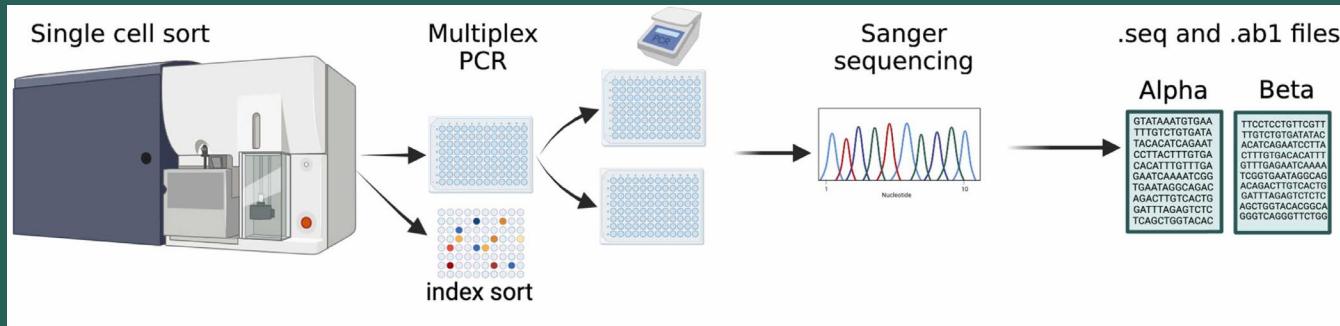
Consideration for experimental questions

	Sanger sequencing	Bulk sequencing
Chain analysis	Paired chain	Single chain
# sequences	Up to 80 per plate Higher resolution	100,000 Depth of repertoire
Accuracy	Higher as no reconstruction of sequence	Lower to moderate Depends on protocol (alignment assumptions)
Time	Lab intensive	Less time intensive
Marker interrogation	Single cell and Protein expression Marker validation	Depends on sort (average within sample)

Outcomes of ATCR pre-meeting workshop

- Paired TCR repertoire interrogation with TCR_Explore
 - TCR-explore.erc.monash.edu
- Understand diversity and TCR clustering in Python
 - Jupyter notebook (github)
- Predicting epitopes with TCRex (Beta chain)
 - Jupyter notebook (github)

TCR repertoire analysis



Sanger Sequencing TCR repertoire analysis with FACS index sort data
Presenter Kerry

Advantages of using TCR_Explore

Table 1. Comparison of QC and automation pipelines.

Step	Process	Current pipeline	TCR_Explore
1	Chromatogram visualisation	Manual	Manual
2	Conversion of .seq to .fasta files	Manual*	Automated
3	Merging of up to 50 files for IMGT sequence alignment	Manual	Automated
	Command-line program	Requires coding expertise	-
4	Classification of productive and non-productive sequences	Manual*	Partially automated
5	Pairing of $\alpha\beta$ or $\gamma\delta$ TCR sequences	Manual*	Automated based on Sequence ID naming
6	Removal of non-productive sequences	Manual*	Automated
If immunophenotyping data available; Index data using .fcs files			
7	Pairing of TCR sequence with immunophenotype	Manual*	Automated
8	Conversion of negative values to allow for log transformation	Manual*	Automated

*Potential for human errors

Pipeline for automating the QC Sanger sequencing

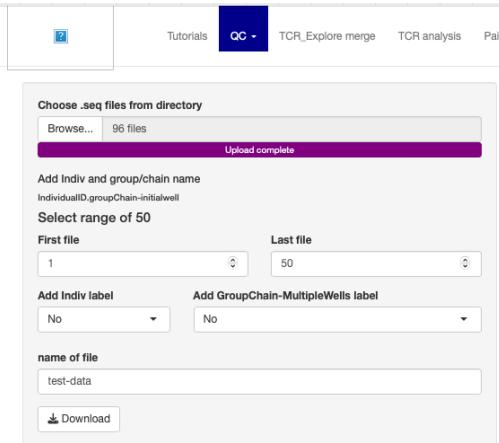
- .seq to .fasta files (TCR_Explore)
 - IndividualID.groupChain-initialwell
- Alignment in IMGT
 - https://www.imgt.org/IMGT_vquest/input
- Scoring the .ab1 files
- Pairing the TCR sequences
- Merging the paired files

QUALITY CONTROL

Part 1: Alignment of the files

Focused Sanger Sequencing experiments

- .seq to .fasta files
(TCR_Explore)
 - IndividualID.groupChain-initialwell



The screenshot shows the TCR_Explore merge interface. At the top, there are tabs: Tutorials, QC (selected), TCR_Explore merge (highlighted in blue), TCR analysis, and Paired TCR with Index data. Below the tabs, a message says "Choose .seq files from directory" with a "Browse..." button and a "96 files" indicator. A purple bar below says "Upload complete". To the right, a sequence viewer displays a list of sequences starting with "V1" and "IFN_B-H10_H12.seq#1". The sequences are long strings of nucleotides.

```

V1
> IFN_B-H10_H12.seq#1
CGCAAGAAGTAGACGTTTCCGGTCTCTAGCACGGTGAGCCGTGTCCTGGCCGAAGAACCTGCTATTGATAGGATAGT
CCTGGTGAAGCTGGCACAGAAATAAGCTGCCGAATCTCTCGCAGTACCTGCTGGATCCTCAGGATGCTGGGGCCCTCTT
GGGAAATTCAAGCAGAAAATCGGACTTCTCTCTTGCCAGTGACCCAGGCCAACTGAGTAATGAATCAGCCTCAGCCCCA
TGCTCTGGGTCTTGTGATACCAAGGACATGATTATCATGGTTATATCCTGGCACACAA
> IFN_B-H5_G12.seq#2
AGCGCAGGAGGACGTTTCCAGGTCTCTAGCACGGTGAGCCGTGTCCTGGCCGAAGAACCTGCTATTGATGCTACACC
GCTAGTCCTCTGGCTGCTGCACAGAAATCACAGCAGAGTCACCAAGCTCCAGGGAAATTGATGTAAGATTA
> IFN_B-H4_F12.seq#3
CGCAGAAGTGACCTTGTGGCTCTACAACGTGAGTCTGGTGCCTGTGCTAAAGAACCTGCTAGTGTGCCACCC
TAAGCTGCTGGCACAGAACTACAGGGCTGAGTCCCTGGCCAGCTCCAAGGAGCTATGTCAGTTCAAGATGAAAGTCACTGA
ACTGTTGAGCTGAGAACTGATCAGGGATGCTTCTTATCGCTA
> IFN_B-G9_E12.seq#4
ACAAGGGGGACGTTTCTGGCTCTAGCACGGTGAGCCGTGCCCAGGGAGACTGCTCGGAGGCTCGGAGC
CTGAGAGATATGCTGCTCTGGCTCATGGCTCATGTTGCTCACAGTCAG
> IFN_B-G4_D12.seq#5
CGCATGGTGGACGTTTCTGGCTCTAGCACGGTGAGCCGTGCCCAGGGAGACTGCTCGGAGGCTCGGAGC
ACTGGTGGACAGAAAGTAAAGAGCTGCTGGTGGGATGGCAGACTCTAGGGACAGGGAGAATTGCTGCTGCTG
GAGAGACACTGATCATCTA
  
```

- Alignment in IMGT
 - https://www.imgt.org/IMGT_vquest/input

QUALITY CONTROL

Part 2: Scoring the .ab1 files

Focused Sanger Sequencing experiments

- Scoring the .ab1 files

The screenshot shows a user interface for scoring .ab1 files. At the top, there are navigation tabs: Tutorials, QC (which is selected), TCR_Explore merge, TCR analysis, and Paired TCR with Index data. Below the tabs, a form allows users to choose .ab1 files from a directory, upload or update a file, and add individual or group chain labels. The 'name of file' field contains 'T00024'. A button to download a single chain file is also present. To the right, a table lists sequencing results with columns for name_temp, pa_score, len_pa, and pa_score_base.

name_temp	pa_score	len_pa	pa_score_base
T00024.IFN_B-H10_H12	462.437652587891	293	1.57828550371294
T00024.IFN_B-H5_G12	279.519134521484	149	1.87596734578177
T00024.IFN_B-H4_F12	188.629379272461	201	0.938454623246074
T00024.IFN_B-G9_E12	232.003082275391	129	1.79847350601078
T00024.IFN_B-G4_D12	43.0167655944824	180	0.238982031080458
T00024.IFN_B-G2_C12	226.287673950195	146	1.54991557500134
T00024.IFN_B-F10_B12	413.307525634766	304	1.35956422906173
T00024.IFN_B-F8_A12	507.873626708984	352	1.44282280315052
T00024.IFN_B-F4_H11	350.954071044922	193	1.81841487588042
T00024.IFN_B-E7_G11	499.907104492188	296	1.68887535301415
T00024.IFN_B-E4_F11	269.656341552734	148	1.82200230778875
T00024.IFN_B-D6_E11	263.757080078125	149	1.77018174549077
T00024.IFN_B-D5_D11	244.031494140625	147	1.66007819143282
T00024.IFN_B-D3_C11	259.793548583984	147	1.76730305159173
T00024.IFN_B-D2_B11	287.767852783203	181	1.5898776396862

QUALITY CONTROL

Part 3: Filtering, pairing and merging

Adding in the .ab1 scores

Tutorials QC ▾ TCR_Explore merge TCR analysis Paired TCR with Index data

Choose a dataset: own_data

Select file for IMGT datafile vquest-2.xls
Upload complete

Select file QC ab1 file T00024.ab1 QC 2023-04-30.csv
Upload complete

Download auto QC table

File name

IMGT create QC file Paired chain file Single chain file TCRdist output file

Fill in the 'clone_quality' column with lowercase: pass or fail

Information included Summary+JUNCTION

Check uploaded files QC file (manual interrogation) QC file with .ab1 scoring

Show 10 entries Search:

Sequence number	Sequence ID	V-DOMAIN Functionality	V-GENE and allele	V-REGION identity %	J-GENE and allele	J-REGION identity %	D-GENE and allele	JUNCTION frame	Sequence
1	45	T00024.CD8A-A3_A01.seq#96	productive	AV19*01	100	AJ40*01	98.31	in-frame	tgtgaccctggact
2	44	T00024.CD8A-A4_B01.seq#95	productive	AV29/DV5*04	99.3464050292969	AJ48*01	98.41	in-frame	tctgctgaaggtcc
3	43	T00024.CD8A-A7_C01.seq#94	unproductive	AV19*01	98.2142868041992	AJ23*01	96.23	out-of-frame	tttgtgtcccttgact
4	42	T00024.CD8A-A9_D01.seq#93	unproductive	AV16*01	70.5882339477539	AJ9*01	95.08	in-frame	gattgtgaaaccttg
5	41	T00024.CD8A-B4_E01.seq#92	productive	AV9-2*01,AV9-2*02	98.5074615478516	AJ42*01	98.48	in-frame	tgaaaccactcttt
6	40	T00024.CD8A-C2_F01.seq#91	productive	AV14/DV4*03	98.6899566650391	AJ42*01	89.39	in-frame	tcaggctgtgaatc
7	39	T00024.CD8A-C6_H01.seq#89	productive	AV41*01	99.212600780078	AJ34*01	96.55	in-frame	tttgtttatgtctgagc

Adding in the .ab1 scores

Choose a dataset: own_data

Select file for IMGT datafile vquest-2.xls
Upload complete

Select file QC ab1 file T00024.ab1 QC 2023-04-30.csv
Upload complete

Download auto QC table

File name

IMGT create QC file Paired chain file Single chain file TCRdist output file

Fill in the 'clone_quality' column with lowercase: pass or fail

Information included Summary+JUNCTION

Check uploaded files QC file (manual interrogation) **QC file with .ab1 scoring**

Show 10 entries Search:

JUNCTION (AA) (with frameshift)	pa_score	len_pa	pa_score_base	V.sequence.quality.check	chromatogram_check	clone_quality	comment
'IF	734.1672974	450	1.631482883	No issue flagged by IMGT	Very high	pass	
KLTF	605.3075562	381	1.588733743	No issue flagged by IMGT	Very high	pass	
CV#GGKLIF	-211.9836731	431	-0.491841469	Unproductive issue	Poor	fail	Either poor sequence quality or no sequence called
	746.0119629	452	1.650468944	Unproductive issue	Very high	fail	Unproductive high quality sequence, possibly resolvable
DGNLIF	572.8182373	297	1.928680934	No issue flagged by IMGT	Very high	pass	
NLIF	862.2457886	451	1.91185319	No issue flagged by IMGT	Very high	pass	
	658.0794678	344	1.913021709	No issue flagged by IMGT	Very high	pass	

Pairing the sequences

Tutorials QC TCR_Explore merge TCR analysis Paired TCR with Index data

Choose a dataset: own_data1

Completed QC file (.csv) TCR_Explore_merged2023.04.30.csv Upload complete

option for paired and TCRdist outputs
Alpha-beta or gamma-delta ab

Information included Summary+JUNCTION

Download paired chain file

File name

Paired chain file

IMGT create QC file Paired chain file Single chain file TCRdist output file

Copy CSV Excel Search:

clone_quality	V.sequence.quality.check	chromatogram_check	cloneCount
1 fail	No alignment	Poor	4
2 fail	No arrangement	Poor	1
3 fail	No issue flagged by IMGT	Low	1
4 fail	No issue flagged by IMGT	Poor	1
5 fail	Unproductive issue	Low	2
6 fail	Unproductive issue	Moderate	1
7 fail	Unproductive issue	Poor	6
8 fail	Unproductive issue	Very high	6
9 pass	No issue flagged by IMGT	High	2
10 pass	No issue flagged by IMGT	Very high	68

Showing 1 to 10 of 11 entries Previous 1 2 Next

Show 10 entries Search:

V.GENE_B	AV	AJ	AVJ	BV	BJ	BD	BVJ	BVDJ	AVJ.BVJ	AVJ.BVDJ
V29-1*01	AV19	AJ40	AV19.J40	BV29-1	BJ2-7	BD1	BV29-1.BJ2-7	BV29-1.BD1.BJ2-7	AV19.J40_BV29-1.BJ2-7	AV19.J40_BV29-1.BD1.BJ2-7
V9*01	AV29/DV5	AJ48	AV29/DV5.J48	Bv9	BJ1-5	-	BV9.BJ1-5	BV9.BJ1-5	AV29/DV5.J48_BV9.BJ1-5	AV29/DV5.J48_BV9.BJ1-5
V13*01	AV14/DV4	AJ42	AV14/DV4.J42	BV13	BJ1-2	BD1	BV13.BJ1-2	BV13.BD1.BJ1-2	AV14/DV4.J42_BV13.BJ1-2	AV14/DV4.J42_BV13.BD1.BJ1-2
V9*01	AV41	AJ34	AV41.J34	Bv9	BJ2-7	BD1	BV9.BJ2-7	BV9.BD1.BJ2-7	AV41.J34_BV9.BJ2-7	AV41.J34_BV9.BD1.BJ2-7

Merging multiple plates

TC EXPLORER Tutorials QC **TCR_Explore merge** TCR analysis Paired TCR with Index data

Select files to merge

Browse... 3 files Upload complete

Download table

Merge Multiple Files

Show 5 entries Search:

cloneCount Indiv.group Indiv group well V.GENE.and.allele_A J.GENE.and.allele_A JUNCTION_A

Favourites Recents Applications Desktop Documents Downloads

iCloud iCloud Drive Shared

Tags Red Orange Yellow Green Blue Purple

Choose Files to Upload

QC_pairing

Name	Date Modified	Size	Kind	Date Created	Date Added
T00024	27/04/2023	--	Folder	03/03/2023	27/04/2023
> Micromon	27/04/2023	--	Folder	03/03/2023	27/04/2023
E10630_paired.csv	26/04/2023	12 KB	co...es	26/04/2023	27/04/2023
SJS.TEN.paired.csv	03/03/2023	39 KB	co...es	03/03/2023	27/04/2023
T00016_paired.csv	26/04/2023	16 KB	co...es	26/04/2023	27/04/2023
T00024_paired.csv	26/04/2023	12 KB	co...es	26/04/2023	27/04/2023
T00024_scored.csv	27/04/2023	51 KB	co...es	03/03/2023	27/04/2023

All

TGTGCTTTGGACAAACAATGACATGCGTTT

TGCATCGTCAGAGTCGCGAATACTGGAGGCTC

TGTGCTGGCTGCGTGGAAGCCAAGGAAATCT

TGTGTGGTGGGGGGCGGACTGCCAGTAACT

TGTGGCACAGAGATCAACGACTTCAAGCTAGC

Showing 1 to 5 of 98 entries Previous 1 2 3 4 5 ... 20 Next

Merging multiple plates

TC EXPLORER

Tutorials QC ▾ TCR_Explore merge TCR analysis Paired TCR with Index data

Select files to merge

Browse... 3 files Upload complete

Download table

Merge Multiple Files

Show 5 entries Search:

cloneCount	Indiv.group	Indiv	group	well	V.GENE.and.allele_A	J.GENE.and.allele_A	JUNCTION_A	
All	All	A	All	.	All	All	All	
1	1	E10630.CD8	E10630	CD8	A5	AV38-2/DV8*01	AJ43*01	TGTGCTTGGACAAACATGACATGCGCTT
2	1	E10630.CD8	E10630	CD8	A9	AV26-1*01,AV26-1*02	AJ9*01	TGCATCGTCAGAGTCGCGAATCTGGAGGCTC
3	1	E10630.CD8	E10630	CD8	B10	AV35*01	AJ42*01	TGTGCTGGGCTCGTGGAAAGCCAAGGAAATCT
4	1	E10630.CD8	E10630	CD8	B4	AV12-1*01	AJ44*01	TGTGTGGTGGGGGGCGGCACTGCCAGTAACT
5	1	E10630.CD8	E10630	CD8	C8	AV30*05	AJ20*01	TGTGGCACAGAGATCAACGACTTCAAGCTCAG

Showing 1 to 5 of 98 entries Previous 1 2 3 4 5 ... 20 Next

TCR Analysis section

Part 4: Creating Figures with TCR_Explore

Challenge!

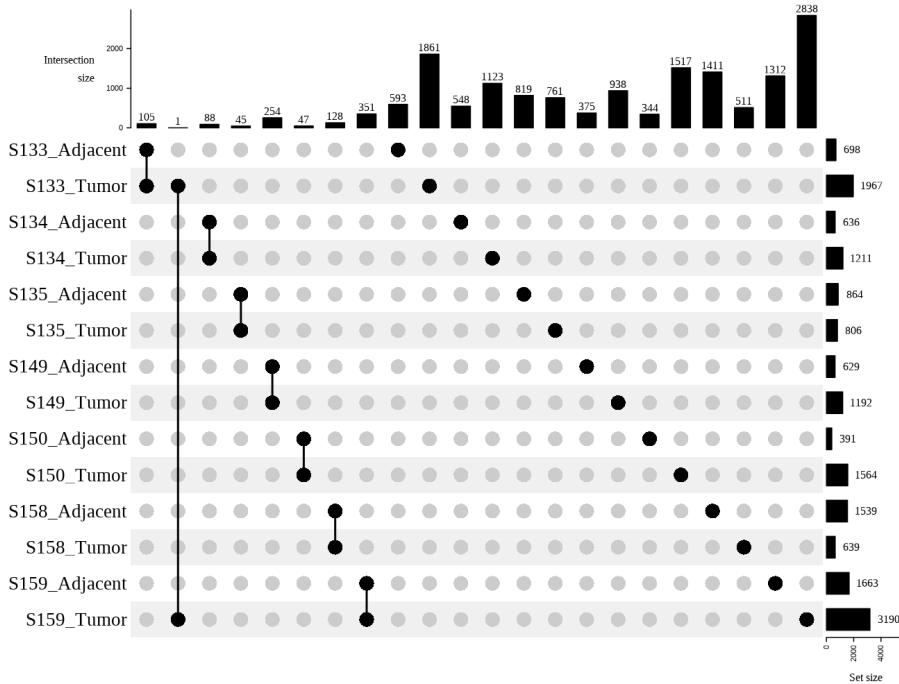
Part 4: Creating Figures with TCR_Explore

Interrogating a single cell RNA-seq dataset

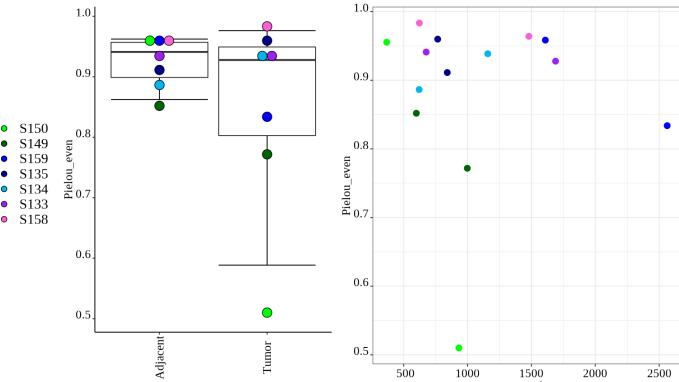
1. Upload to TCR Explore (analysis section)

- Overview (Treemap or Circos[®]/chord diagram)
 - Group
 - Individual and group analysis
- Motif
 - Length plot
- Diversity
- Overlap
 - Upset plot

Example figures



Patient specific response



QUALITY CONTROL

Part 5: Filtering, pairing and merging FACSsort Index data

Focused Sanger Sequencing experiments

- Merging with Flow cytometry Index sorting data.
- Cleaning the data
 - Adding in the dimensional reduction (UMAP)
 - Unsupervised clustering (k-means)
 - Remove negative counts to small integers
- Figure generation

Break

TCR repertoire analysis in python

Data parsing and diversity metrics

TCR clustering with ClusTCR

TCR specificity annotation with TCRex

Romi Vandoren

TCR repertoire analysis in python

Data parsing and diversity metrics

TCR clustering with ClusTCR

TCR specificity annotation with TCRex

Romi Vandoren

Preprocessing data

Why?

To ensure data quality



Input?

v_call	j_call	junction_aa	duplicate_count	frequency
TRBV6-2	TRBJ2-1	CASSNSDRTYGDNEQFF	33422.0	0.012503577447769263
TRBV24-1	TRBJ2-5	CATSSVLTQQETQYF	24502.0	0.009166496757382637
TRBV12-3	TRBJ2-3	CASSSRGLANTQYF	22361.0	0.008365522569252841
TRBV29-1	TRBJ2-7	CSVVGADTYEQYF	20930.0	0.007830168032487901
TRBV7-8	TRBJ1-1	CASSLGTALNTEAFF	20193.0	0.007554447352127482
TRBV7-2	TRBJ2-2	CASSRRHLGNTGELFF	18597.0	0.006957364304834091

↓ ↓ ↓ ↓

V- and J-gene CDR3 beta Clone count Clone frequency

Preprocessing data



Why?

To ensure data quality

Input?

Dataframe containing :

- CDR3 sequences
- V- and J-genes
- Clone counts
- Clone frequencies

How?

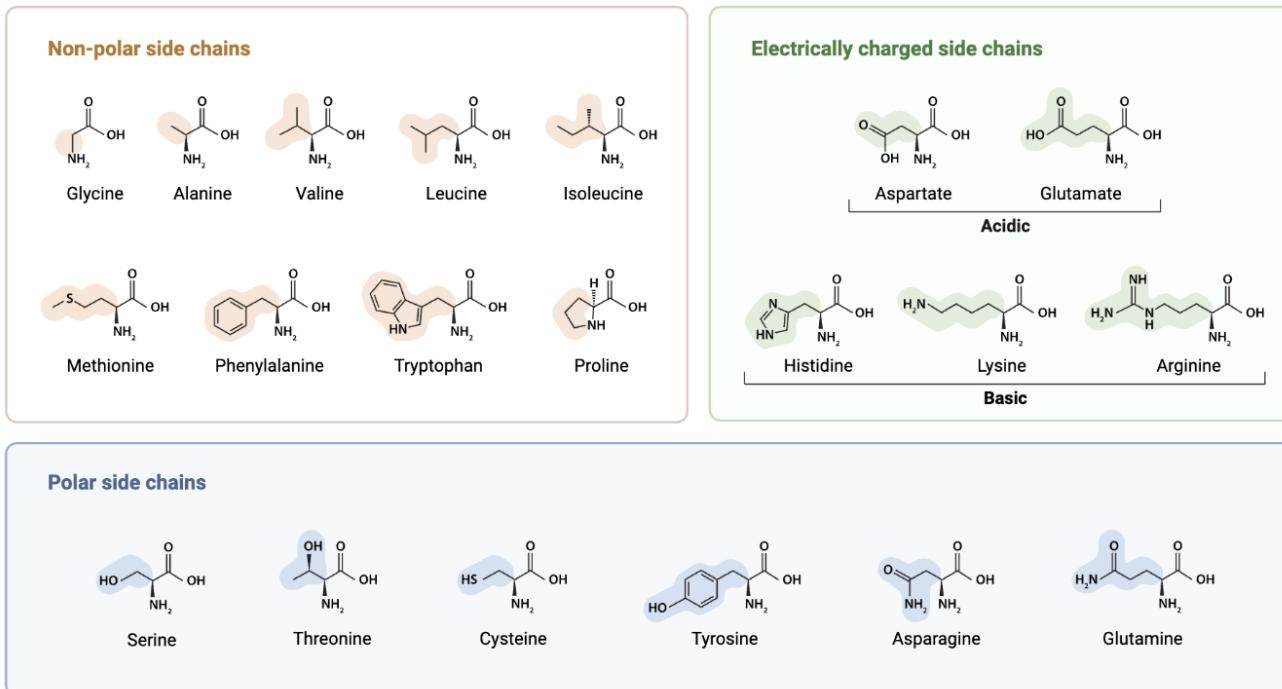
Several parsing steps:

- (1) Remove CDR3 β containing non-amino acid characters
- (2) Remove CDR3 β not starting/ending with C and F
- (3) Merge duplicate counts and frequencies
- (4) Remove duplicates

Preprocessing data

Parsing steps:

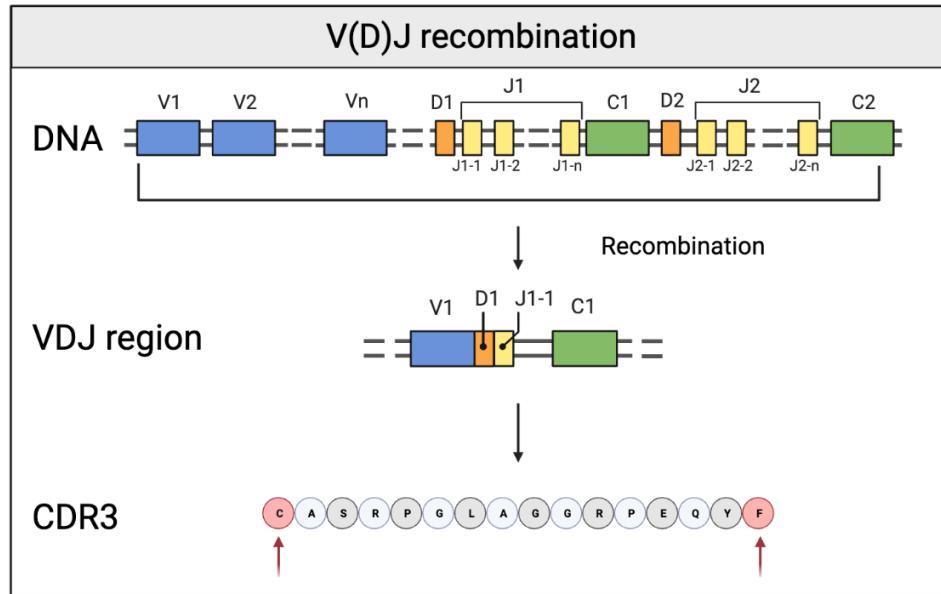
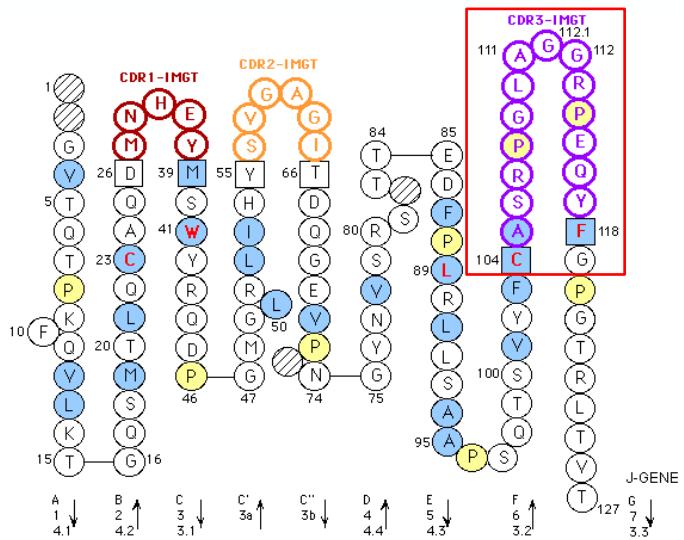
(1) Remove CDR3 β sequences containing non-amino acid characters



Preprocessing data

Parsing steps:

- (1) Remove CDR3 β sequences containing non-amino acid characters
- (2) Remove CDR3 β sequences not starting/ending with C and F



Preprocessing data

Parsing steps:

- (1) Remove CDR3 β sequences containing non-amino acid characters
- (2) Remove CDR3 β sequences not starting/ending with C and F
- (3) Merge duplicate counts and frequencies

v_call	j_call	junction_aa	duplicate_count	frequency
TRBV6-2	TRBJ2-1	CASSNSDRTYGDNEQFF	33422.0	0.012503577447769263
TRBV24-1	TRBJ2-5	CATSSVLTQQETQYF	24502.0	0.009166496757382637
TRBV12-3	TRBJ2-3	CASSSRGLANTQYF	22361.0	0.008365522569252841
TRBV29-1	TRBJ2-7	CSVVGADTYEQYF	20930.0	0.007830168032487901
TRBV24-1	TRBJ2-5	CATSSVLTQQETQYF	20193.0	0.007554447352127482
TRBV7-2	TRBJ2-2	CASSRRHLGNTGELFF	18597.0	0.006957364304834091

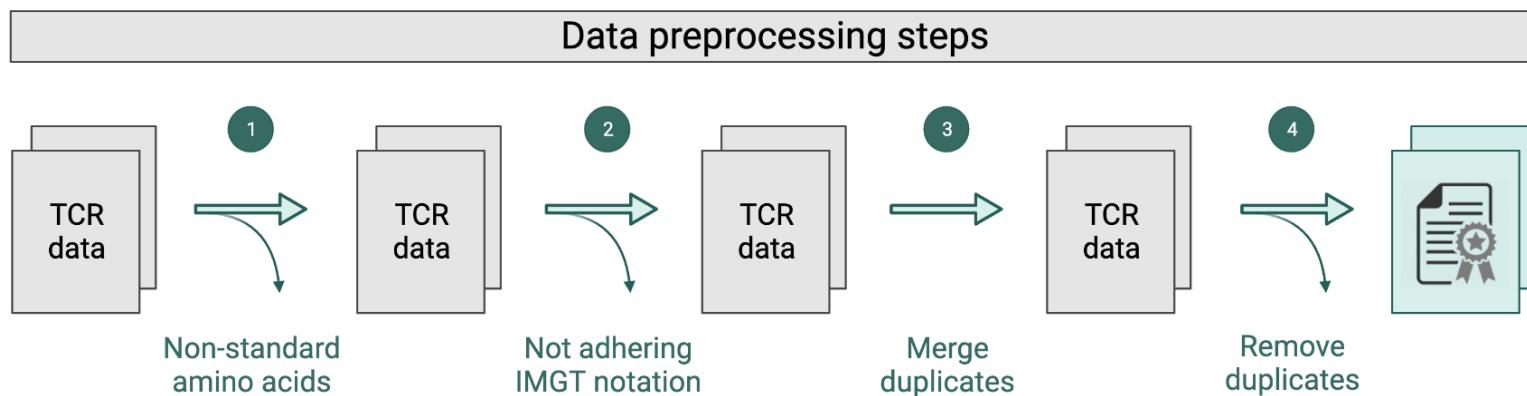
↓

TRBV24-1	TRBJ2-5	CATSSVLTQQETQYF	44695.0	0.016720944109510118
----------	---------	-----------------	---------	----------------------

Preprocessing data

Parsing steps:

- (1) Remove CDR3 β sequences containing non-amino acid characters
- (2) Remove CDR3 β sequences not starting/ending with C and F
- (3) Merge duplicate counts and frequencies
- (4) Remove duplicates



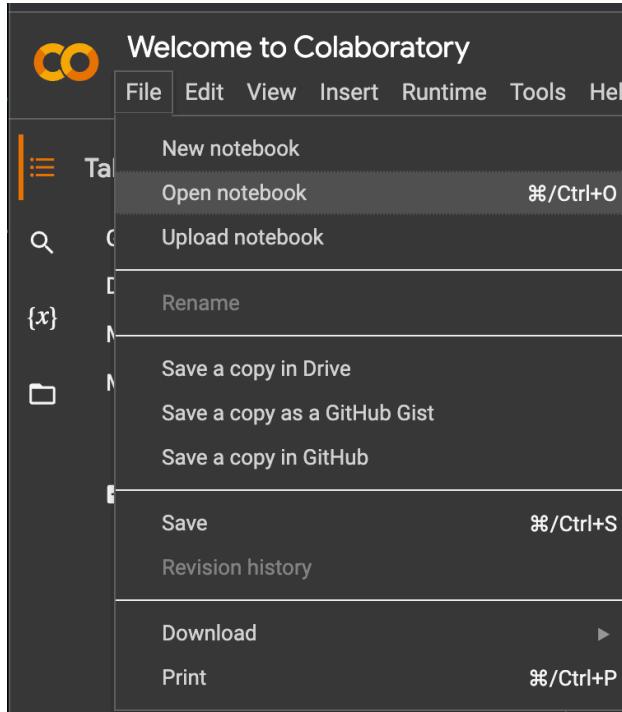
Google colab step-by-step

Part 1A: Parsing_and_diversity.ipynb

Data parsing

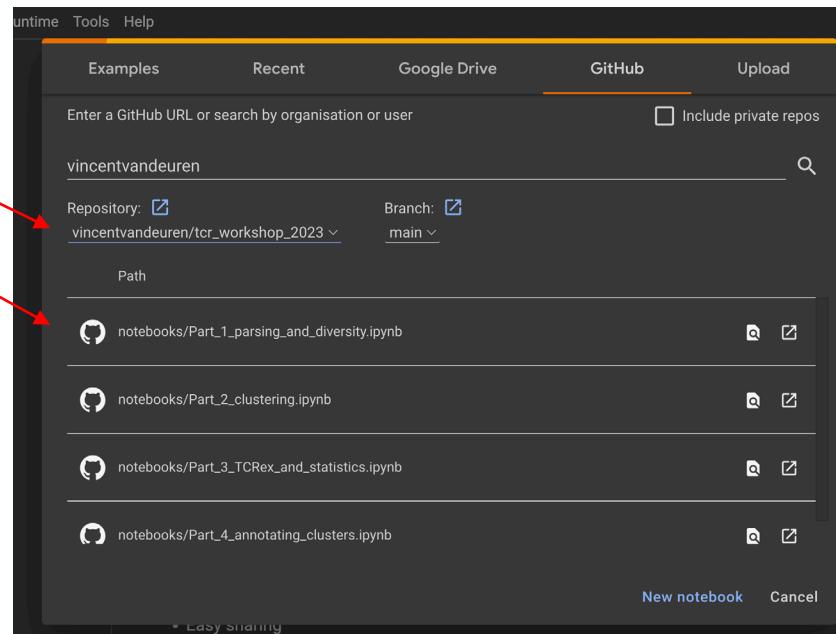
Working in google colab

- Open Google colab: <https://colab.research.google.com/>
- Open new notebook



Working in google colab

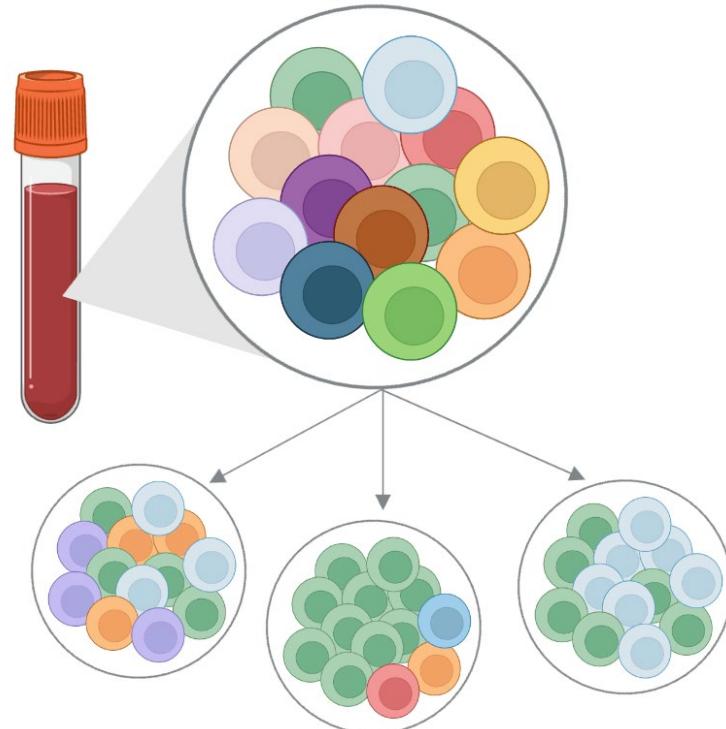
- Open Google colab: <https://colab.research.google.com/>
- Open new notebook
- Pick Github and enter: [vincentvandeuren](#), pick 'tcr_workshop_2023' from the dropdown



Quantifying repertoire diversity

Why?

- Describe immune fitness of an individual
- Indicator for specific disease
- Biomarker for therapy response
- Compare immunological state between repertoires



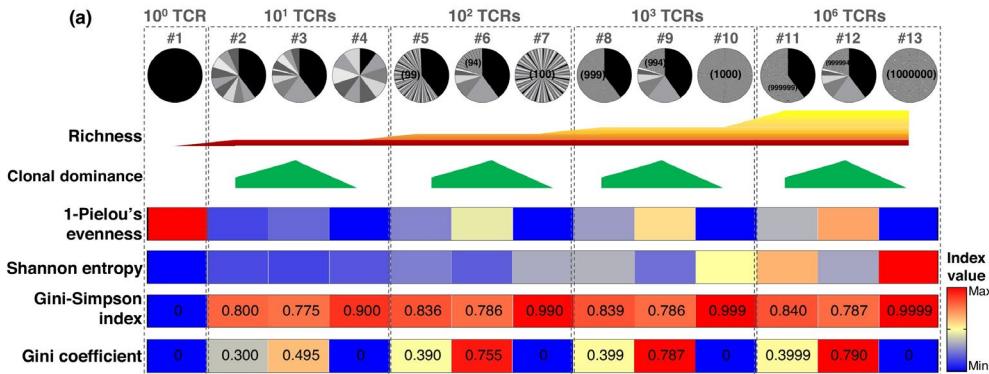
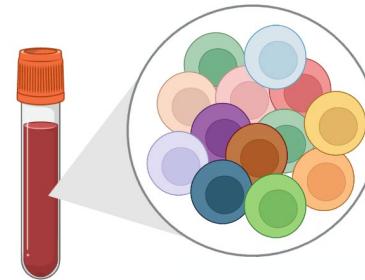
Quantifying repertoire diversity

Why?

- Immunological fitness

How?

- Many different metrics developed
 - Repertoire richness
 - Shannon Diversity
 - Pielou's Evenness
 - Gini-Simpson diversity
 - Gini coefficient
 - DE50



Chiffelle et al., *Curr Opin Biotechnol*, 2020.

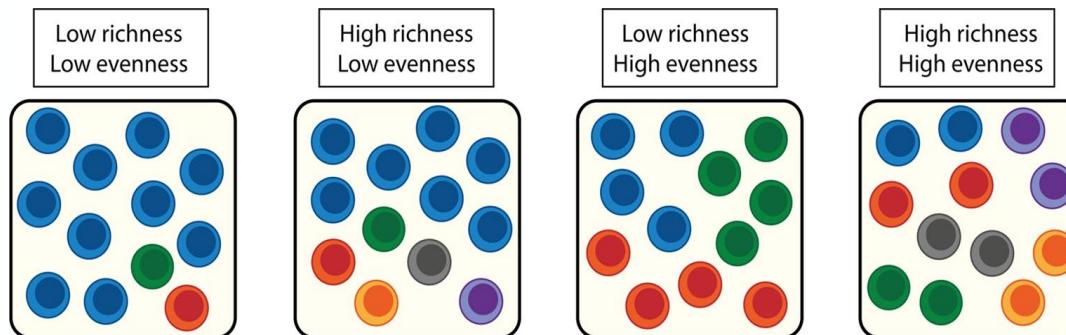
Quantifying repertoire diversity

Multiple diversity options?

- **Richness:** number of unique TCRs
 - Sensitive to the number of rare clonotypes
- **Evenness:** distribution spectrum of sequences
 - Strongly affected by clonal expansions
- **Fractions:** DE50



Use **panel** of diversity metrics for comprehensive overview



Quantifying repertoire overlap

Multiple overlap metrics

- **Jaccard**: Overlapping TCRs compared to total TCRs
 - Does not take frequency into account

$$Jaccardindex = J(i, j) = \frac{c_{ij}}{N_i + N_j - c_{ij}}$$

- **Morisita**: Overlapping TCRs and frequency distribution compared to total TCRs
 - Includes frequency in comparison

$$Morisita - Hornindex = MH(i, j) = \frac{2 \sum_{i=1}^N n_{1i} n_{2i}}{(\sum_{i=1}^S f_i^2 + \sum_{i=1}^S g_i^2) n_1 n_2}$$

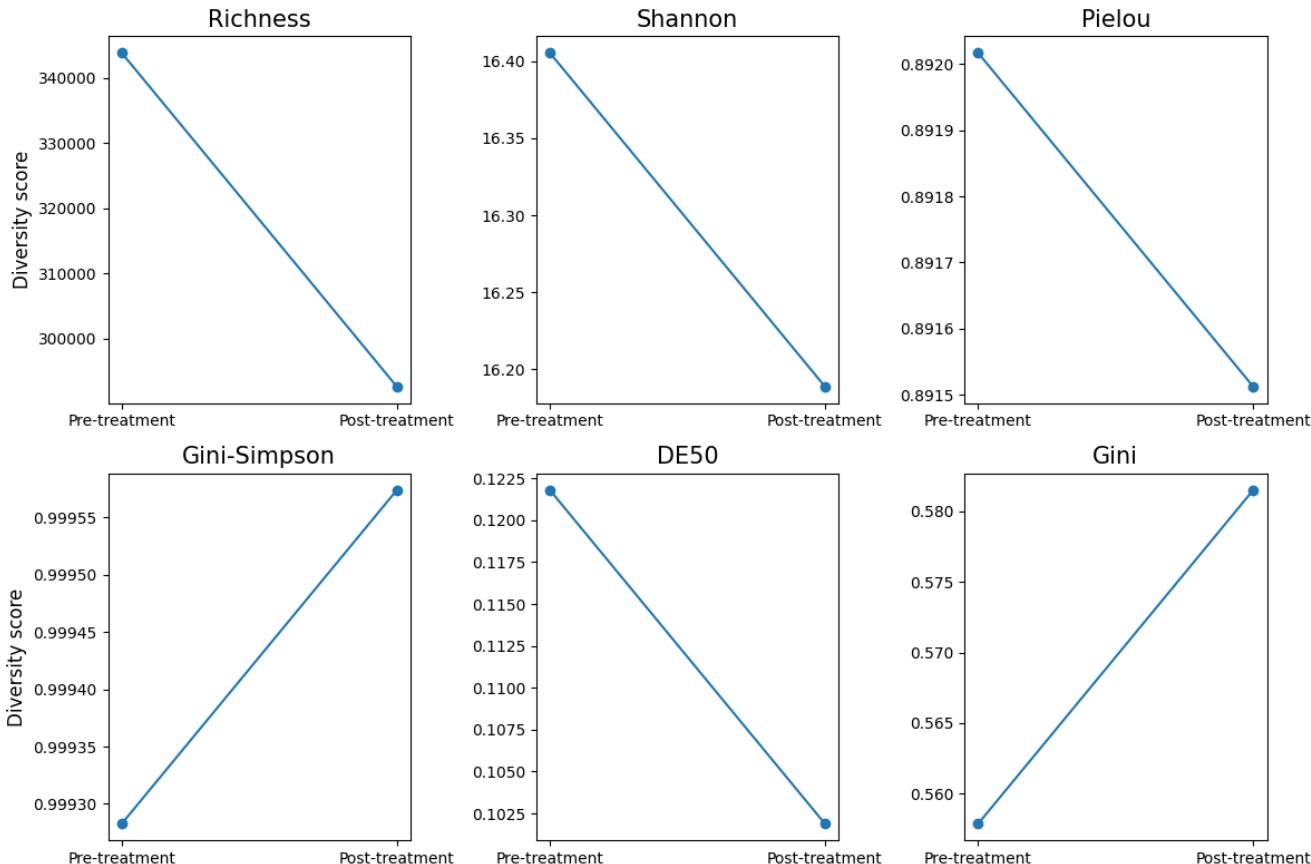
- **Public clones**: Which TCRs are present in more than 1 person?

Google colab step-by-step

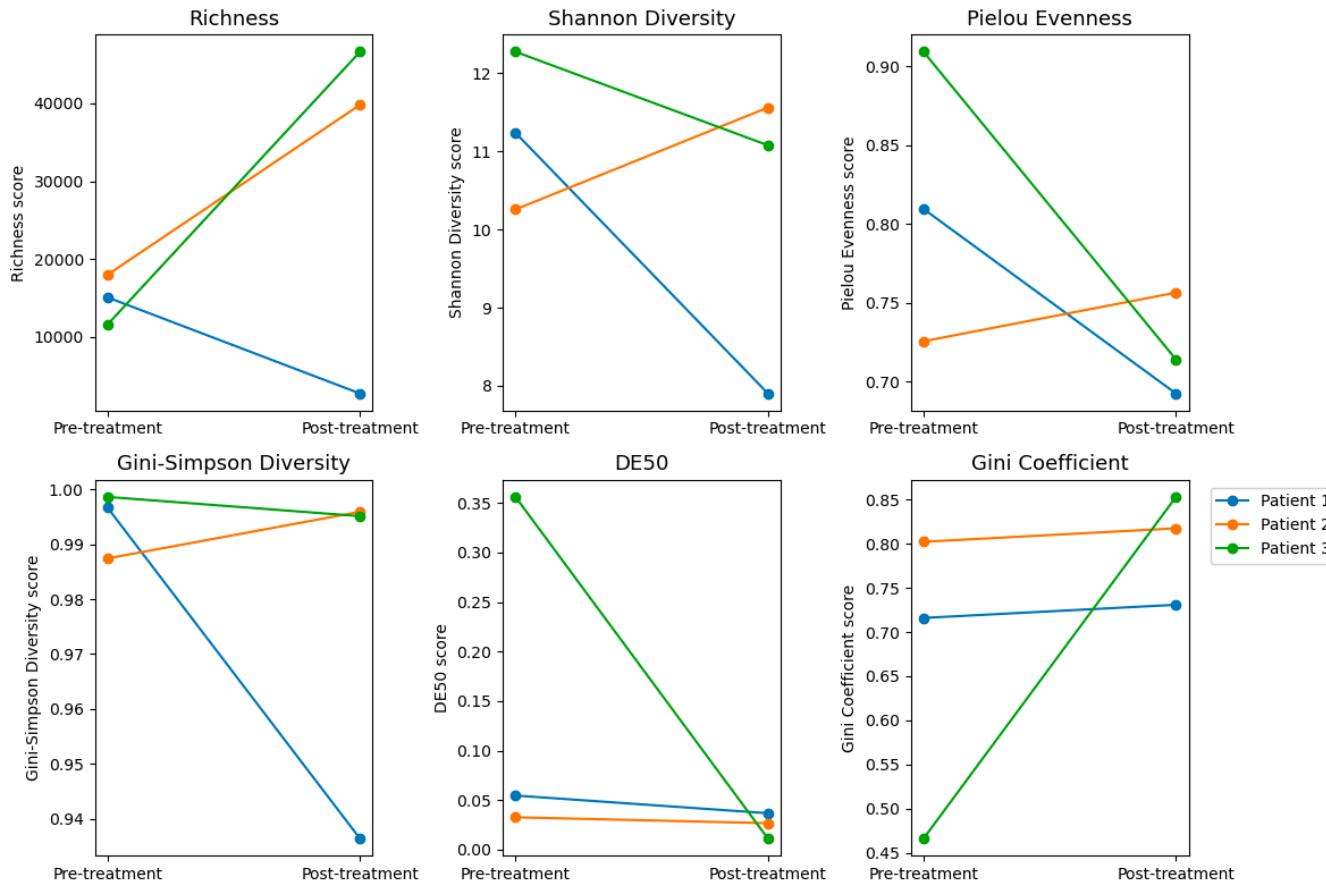
Part 1B: Parsing_and_diversity.ipynb

Diversity Metrics

Quantifying repertoire diversity: Results



Quantifying repertoire diversity: Patient comparison



TCR repertoire analysis in python

Data parsing and diversity metrics

TCR clustering with ClusTCR

TCR specificity annotation with TCRex

Romi Vandoren

TCR repertoire clustering

Why?

- Reduce enormous complexity of the repertoire

How?

- Cluster together TCR sequences with high sequence similarity

CASSQETRTDTQYF
CASSGTGGEETQYF
CASSRGGGRETQYF
CASSLGNEASYNEQFF
CASSQETGTDQYF
CASSRTGGRETQYF
CASSAGTSTDQYF
CASSLGYEASYNEQFF

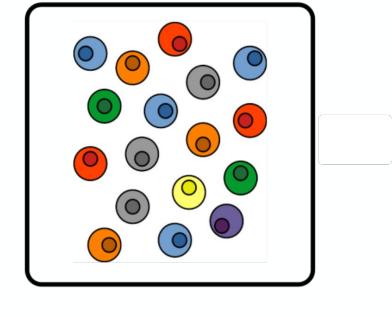


CASSQET**R**TDTQYF
CASSQET**G**TDTQYF
CASS**A**GT**S**STDQYF

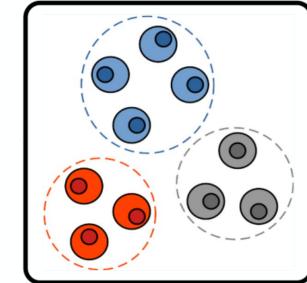
CASS**G**TG**E**ETQYF
CASS**R**GG**G**RETQYF
CASSRTGGRETQYF

CASSLGYEASYNEQFF
CASSLGNEASYNEQFF

Complex repertoire



Clusters

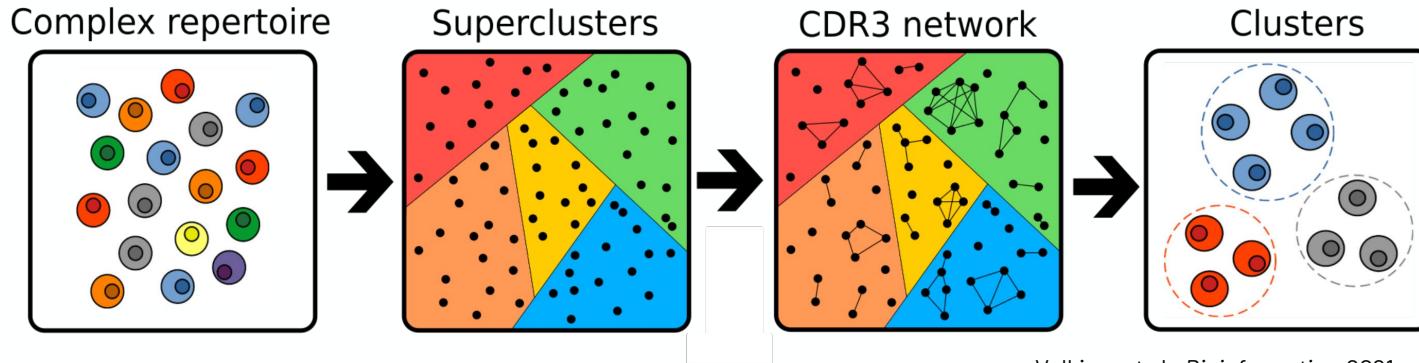


Repertoire clustering: ClusTCR

How?

Using a 2-step clustering process:

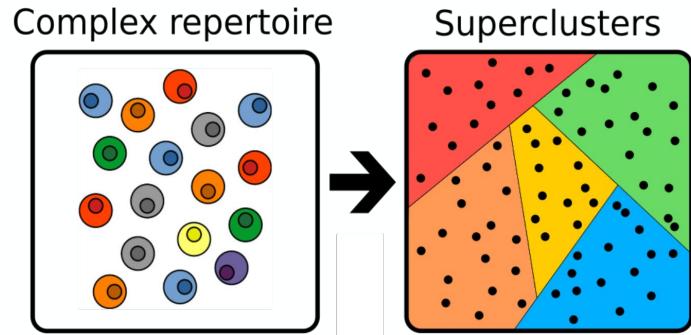
- Faiss Clustering Library (speed)
- Markov Clustering Algorithm (accuracy)



Repertoire clustering: ClusTCR

Faiss Clustering Library:

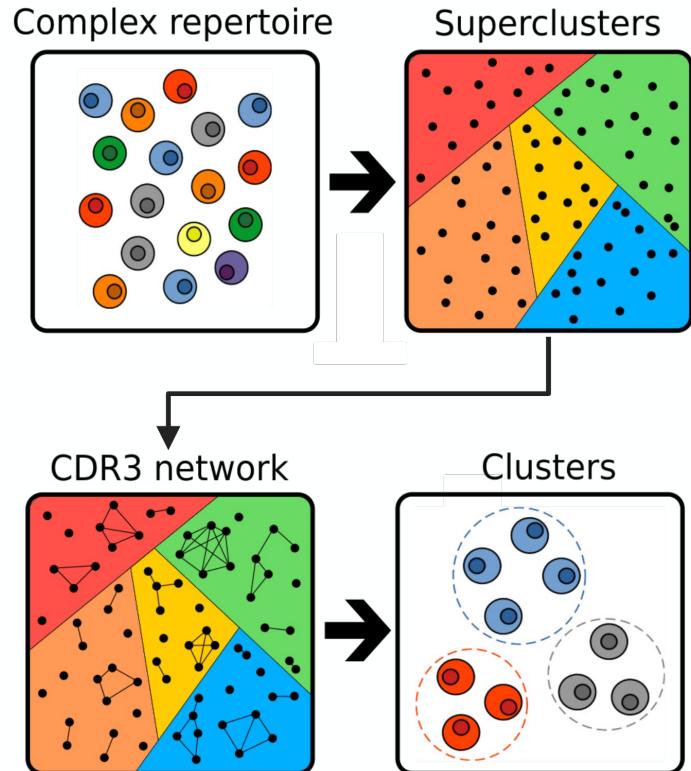
- Convert sequences into vectors
- Using CDR3 chemical properties
- K-means clustering
- Using all TCR sequences
- Create superclusters of similar characteristics



Repertoire clustering: ClusTCR

Markov Clustering Algorithm:

- Identify network structures in a graph
- CDR3 sequences are the nodes
- Edges represent maximum amino acid edit of 1
 - (Hamming distance)
- Hashing function to limit number of comparisons
- Divide superclusters into individual clusters



Repertoire clustering: ClusTCR

Results:

- **Summary**: dataframe detailing the cluster motif and size
- **Features**: dataframe describing several physicochemical properties of each of the TCR sequences that have been clustered
- **Cluster_df**: dataframe representing all TCR sequences and the cluster number they belong in and their associated features.

Google colab step-by-step

Part 2: Clustering.ipynb

Break

TCR repertoire analysis in python

Data parsing and diversity metrics

TCR clustering with ClusTCR

TCR specificity annotation with TCRex

Vincent Van Deuren

TCR specificity annotation with TCRex

- **Couple TCR sequences to their antigenic targets**
 - Annotation based on curated databases (e.g. VDJdb, IEDB, TCRMatch)
 - Machine learning algorithms for predicting TCR specificity (e.g. TCRex)

TCR specificity annotation with TCRex

- Couple TCR sequences to their antigenic targets
 - Annotation based on curated databases (e.g. VDJdb, IEDB, TCRMatch)
 - Machine learning algorithms for predicting TCR specificity (e.g. TCRex)

Front. Immunol., 29 November 2019

Sec. T Cell Biology

Volume 10 - 2019 | <https://doi.org/10.3389/fimmu.2019.02820>

Detection of Enriched T Cell Epitope Specificity in Full T Cell Receptor Sequence Repertoires



Sofie Gielis^{1,2,3},



Pieter Moris^{1,3†},



Wout Bittremieux^{1,3,4†},



Nicolas De Neuter^{1,2,3},



Benson Ogunjimi^{2,5,6,7},



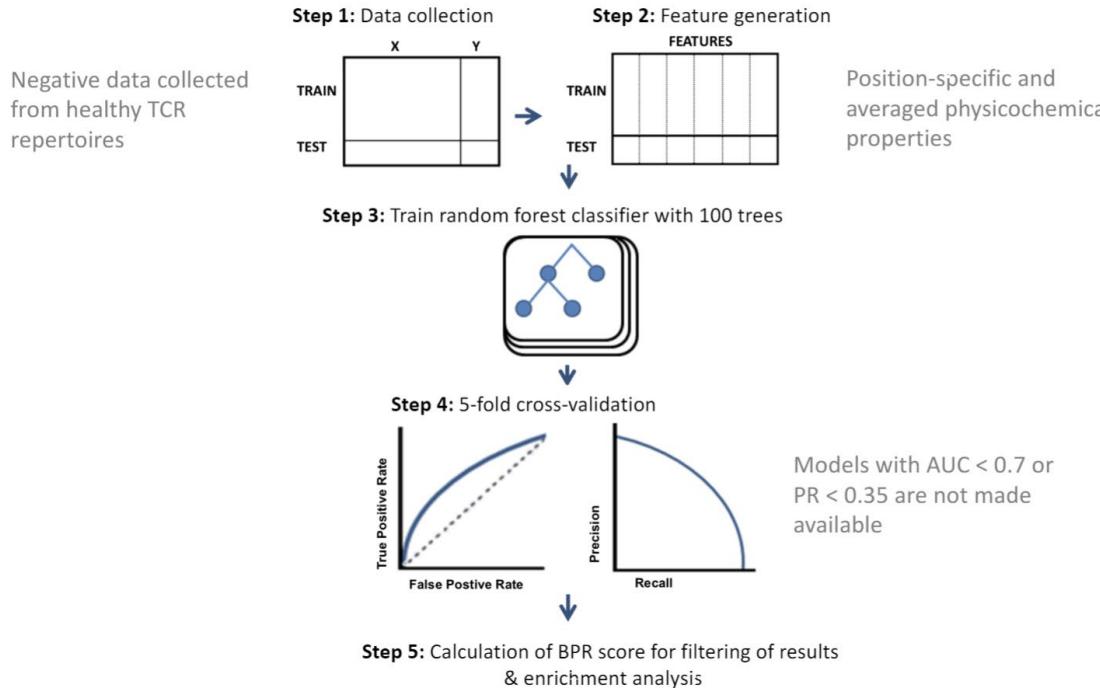
Kris Laukens^{1,2,3‡} and



Pieter Meysman^{1,2,3**}

TCR specificity annotation with TCRex

- What is TCRex?



TCR specificity annotation with TCRex

- What can TCRex do?
 - Predict interaction with known epitopes
 - 93 viral epitopes (47 against SARS-CoV-2), 5 cancer epitopes
 - Train new models:
 - Positive training data provided by user (max 500 sequences)
 - Negative training data provided by TCRex
- Limitations
 - Human TCR-beta sequences only
 - File limit: 50000 sequences per job

TCR specificity annotation with TCRex

Using TCRex:

1. Start a session: <https://tcrex.biodatamining.be/>
2. Upload TCR sequence data file, in correct format

Select your TCR sequence data file: No file selected.

3. Select your epitope models of interest
4. Submit!

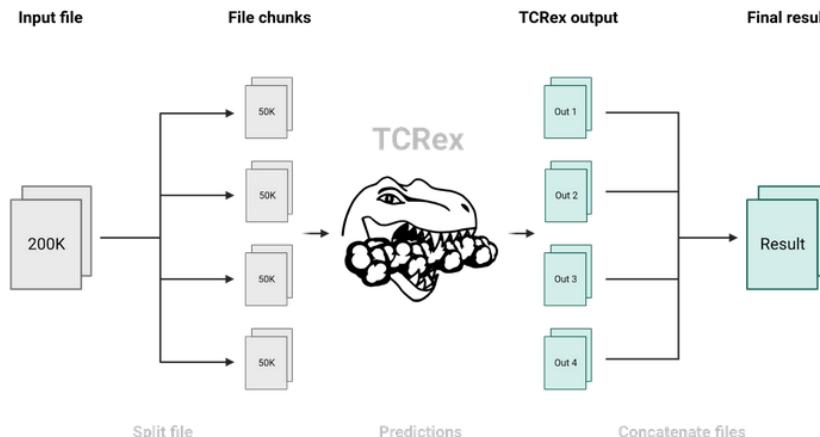
TCR specificity annotation with TCRex

Advanced TCRex options:

- **Model version:** use older versions for e.g. comparison with past results
- **IMGT parsing:** automatic parsing of V/J genes to standard format
- **Enrichment threshold:** threshold used for automatic enrichment analysis
- **Enrichment background:** upload a custom background repertoire file for the automatic enrichment analysis

TCR specificity annotation with TCRex

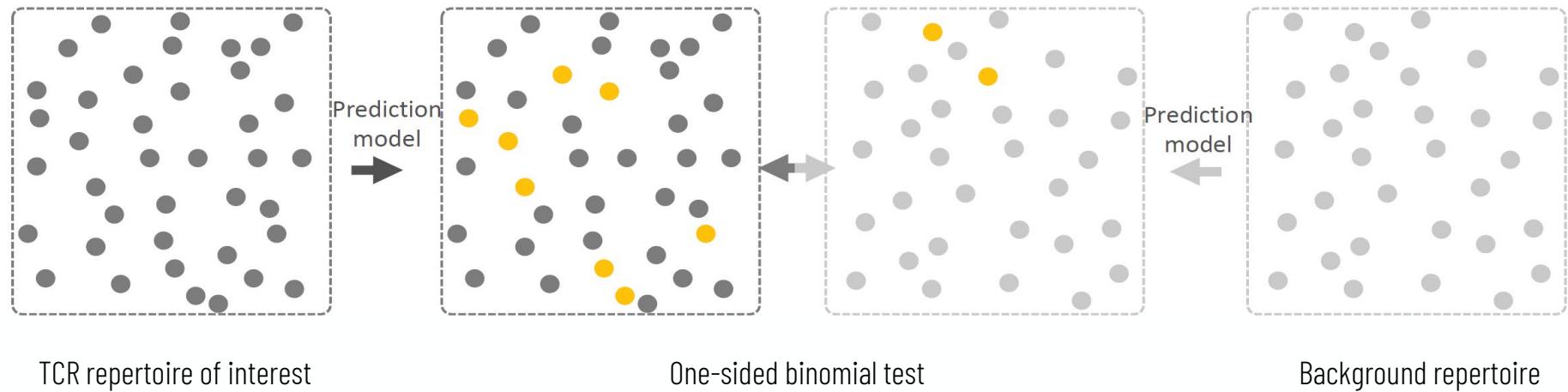
Caveat: maximum 50000 sequences per session, but multiple sessions can be started simultaneously



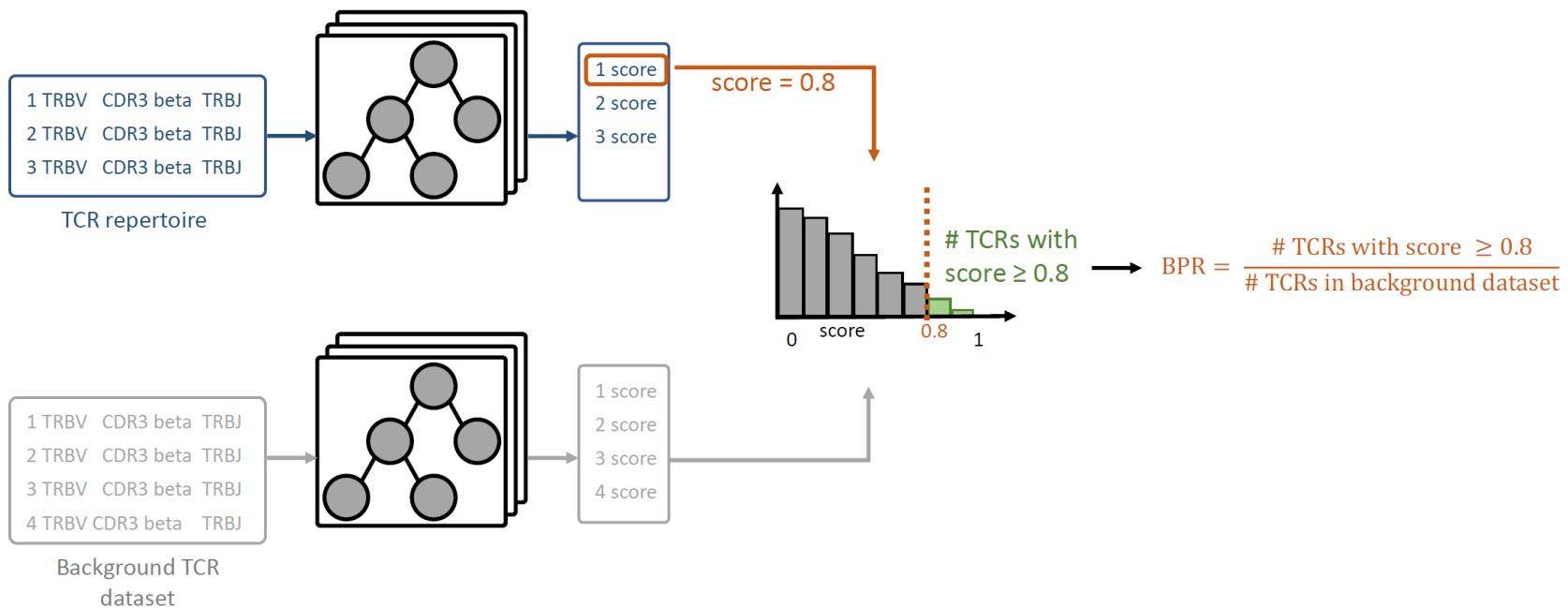
These functions for splitting and concatenating your data will be provided in the Jupyter notebook

Enrichment analysis based on TCRex annotations

Question: do we find more TCRs reactive against a particular epitope than expected



Addendum: Binding Positive Rate



Google colab step-by-step

Part_3_TCRex_and_statistics.ipynb

Google colab step-by-step

Part_4_annotation_clusters.ipynb

Take home messages

- **TCR_Explore**
 - Easy to use web interface
 - Streamlined QC process to aid in accuracy of Sanger Sequencing data
- **ClusTCR**
 - Easy to use
 - Fast compared to other state-of-the-art clustering algorithms
([GLIPH2](#), [iSMART](#) and [tcrdist](#))
- **TCRex**
 - Predicts TCR-epitope binding for a set of known, well-studied epitopes
 - Can be used for epitope-specificity enrichment analyses

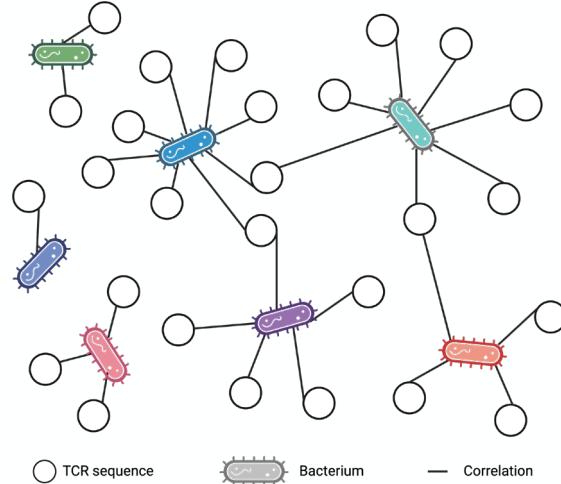
ATCR meeting: Current tool development

STEGO.R



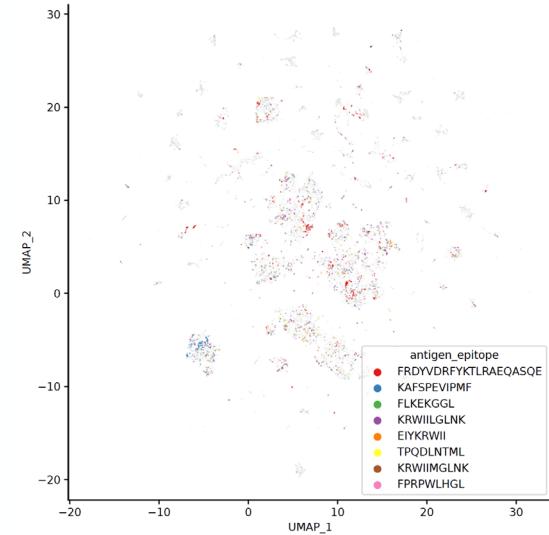
Talk: Thursday @12.00h
STEGO.R for (easy)
interrogation of combined
scTCR repertoire and scRNA-
seq data

Mapping the TCR-microbiome network



Talk: Wednesday @10.40h
Unravelling the complex TCR-
microbiome interactions network
in the colon

RapTCR



Visualizing the TCR sequence space
Poster

Acknowledgements



AUDACIS - Antwerp Unit for Data Analysis and
Computation in Immunology and Sequencing
University of Antwerp

Spokespersons

- Kris Laukens (strategic)
- Pieter Meysman (computational)
- Benson Ogunjimi (biomedical)

Postdoctoral researchers

- Esther Bartholomeus
- Hajar Besbassi
- My Ha
- Mariia Kuznetsova
- Kerry Mullan

PhD students

- Charlotte Adams
- Fabio Affaticati
- Ceder Dens
- Nicky De Vrij
- Sofie Gielis
- Anna Postovskaya
- Sebastiaan Valkiers
- Vincent Van Deuren
- Romi Vandoren
- Alexandra Vujkovic

Study nurse

- Karin Peeters

Supporting & collaborating scientists

- Wim Adriaensen (ITM)
- Philippe Beutels (CHERMID, UAntwerp)
- Wout Bittemieux (ADReM, UAntwerp)
- Peter Delputte (LMPH, UAntwerp)
- Niel Hens (CHERMID, UAntwerp)
- Eva Lion (LEH, UAntwerp)
- Peter Ponsaerts (LEH, UAntwerp)
- Evelien Smits (CORE, UAntwerp)
- Arvid Suls (MEDGEN/CMG, UAntwerp)
- Paul G Thomas (St Jude, USA)
- Pierre Van Damme (CEV, UAntwerp)
- Koen Vercauteren (ITM)

Funding

- Chan Zuckerberg Initiative



Contact details

- Romi Vandoren (romi.vandoren@uantwerpen.be)
- Kerry Mullan (Kerry.mullan@uantwerpen.be)
- Vincent Van Deuren (vincent.vandeuren@uantwerpen.be)