Benchmarking of T-cell receptor repertoire profiling methods reveals large systematic biases

Encarnita Mariotti-Ferrandiz, Associate Professor

i[3] Laboratory - UMRS 959 – Sorbonne Université

Paris, France

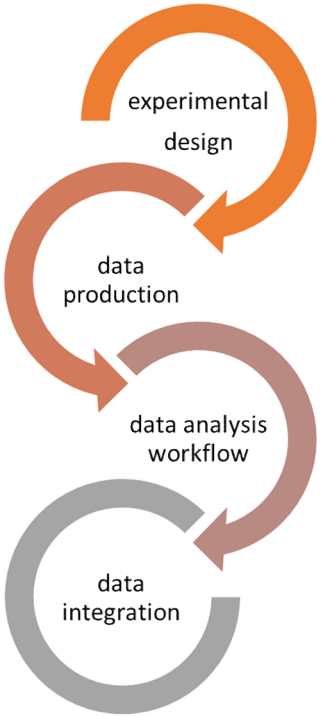
07/12/2020 – AIRR-C meeting

Biological resources WG pre-meeting

@ : encarnita.mariotti@sorbonne-universite.fr

Next generation sequencing for AIRR study

Quantitative description of rearrangement sequences



- T-cell sorting **or** total T cell
- Tissue, organ **or** blood
- DNA **or** RNA

- Appropriate method to address the question
  - TCR RepSeq method *Rosati et al. 2017; Barennes et al. 2020*

- Alignment tools *Bolotin et al 2015 ; Gerritsen et al. 2016*
- Statistical analysis workflow *Shugay et al. 2015; Nazarov et al. 2015; Greiff et al, 2015;*
- Statistical modelling *Mariotti-Ferrandiz et al. 2016; Chaara et al. FI, 2017; Miho et al, 2019 ; Greiff et al, 2020*

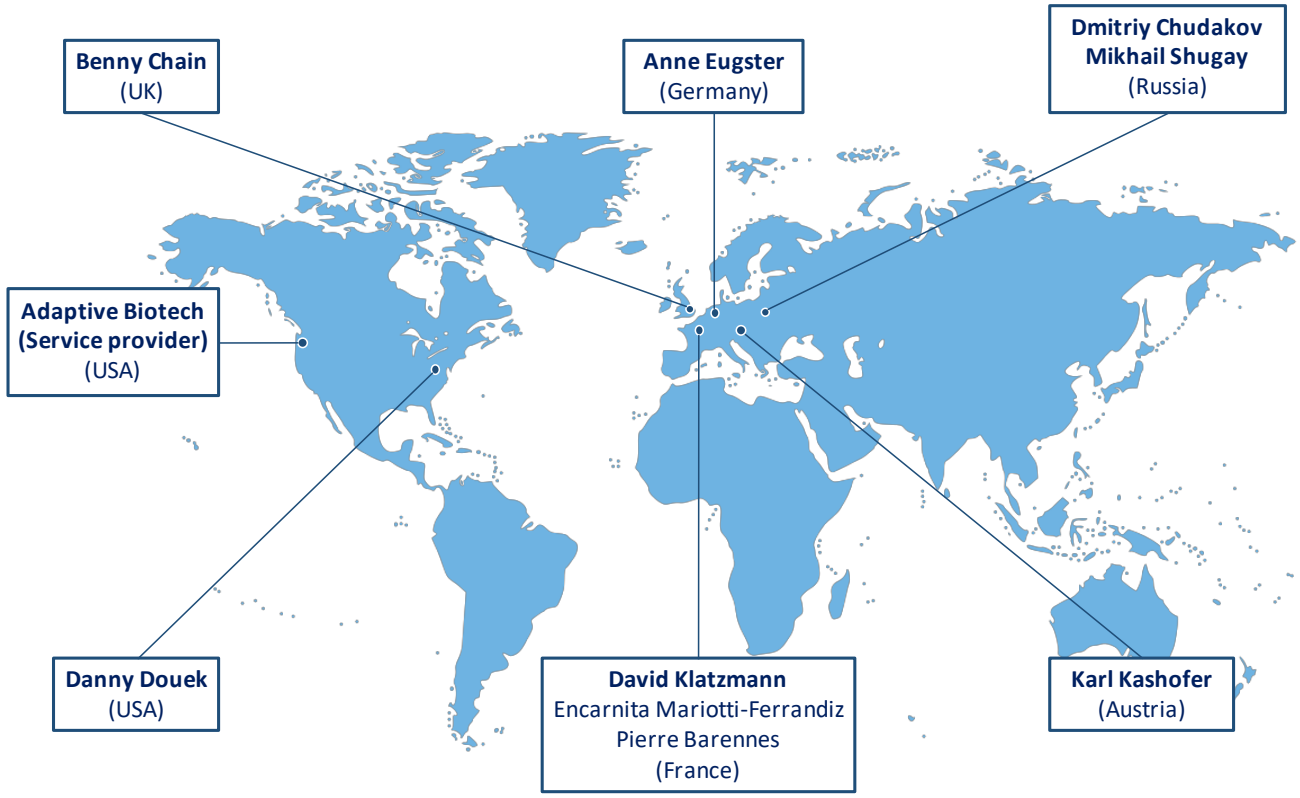
- TCR data sharing
- Standards from AIRR community *Rubelt et al. 2017; Breden et al. 2017*

Boudinot, P et al (2008) Molecular Immunology
Six, A et al. (2013) Frontiers in Immunology

Available methods

	template nature	*Amplification method*	*Applications*
Academic labs	RNA	RACE-PCR	bulk & single cell
	RNA/gDNA	Multiplex PCR VC/VJ	bulk & single cell
	RNA/gDNA	Multiplex-PCR VC/VJ	bulk
	gDNA	Multiplex-PCR VJ	bulk
	RNA	RACE-PCR	bulk
	RNA	RACE-PCR	Single cell
	RNA	RNASeq	Single cell
	RNA	RACE-PCR	Single cell

International collaboration



nature biotechnology
ANALYSIS
https://doi.org/10.1038/s41587-020-0656-3
Check for updates

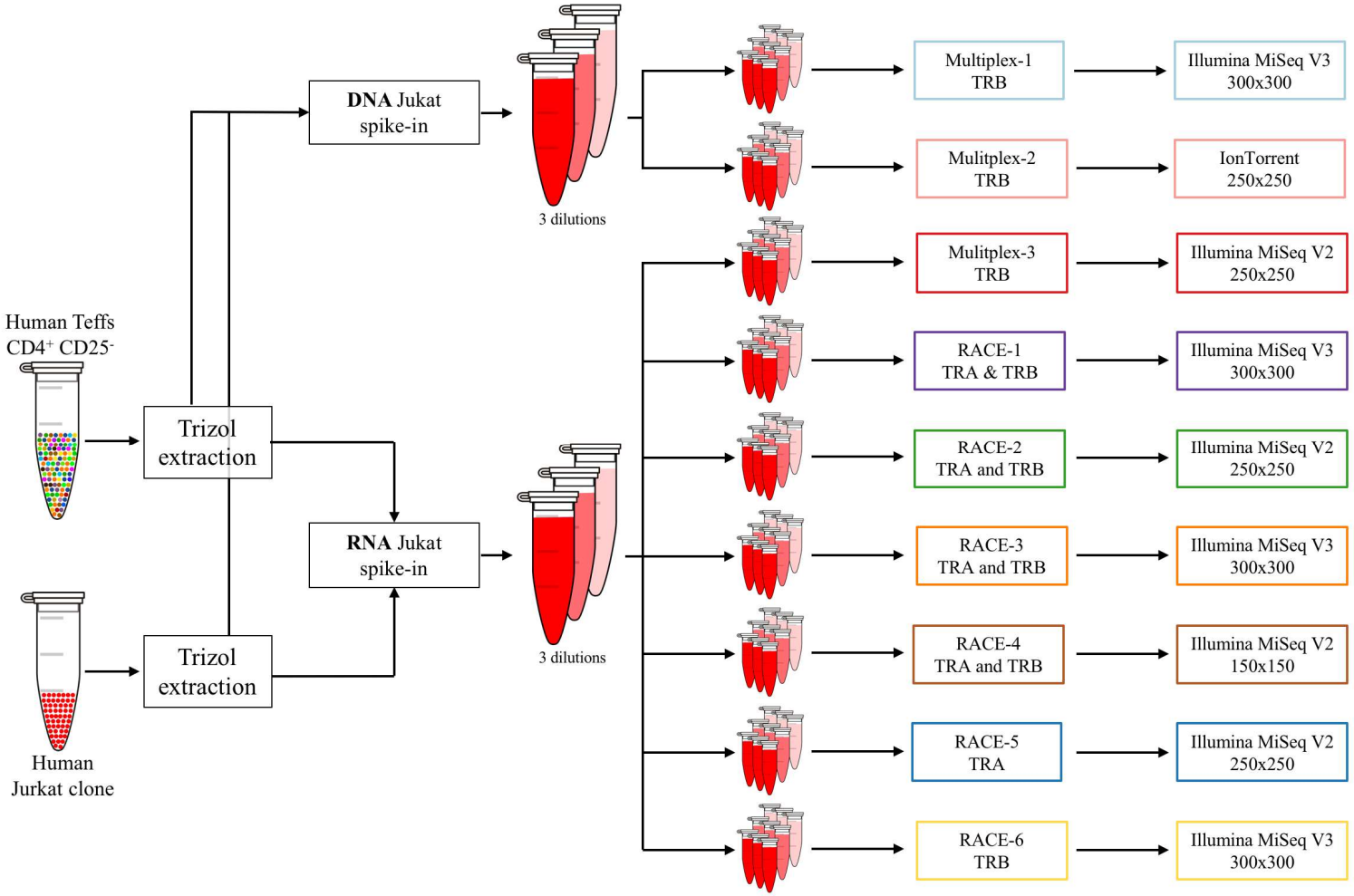
Benchmarking of T cell receptor repertoire profiling methods reveals large systematic biases

Pierre Barennes[1,2], Valentin Quiniou[1,2], Mikhail Shugay[3,4,5], Evgeniy S. Egorov[4], Alexey N. Davydov[6], Dmitriy M. Chudakov[3,4,5,6], Imran Uddin[7], Mazlina Ismail[7], Theres Oakes[7], Benny Chain[7], Anne Eugster[8], Karl Kashofer[9], Peter P. Rainer[10,11], Samuel Darko[12], Amy Ransier[12], Daniel C. Douek[12], David Klatzmann[1,2] and Encarnita Mariotti-Ferrandiz[1,2]✉

Pierre Barennes
PhD student

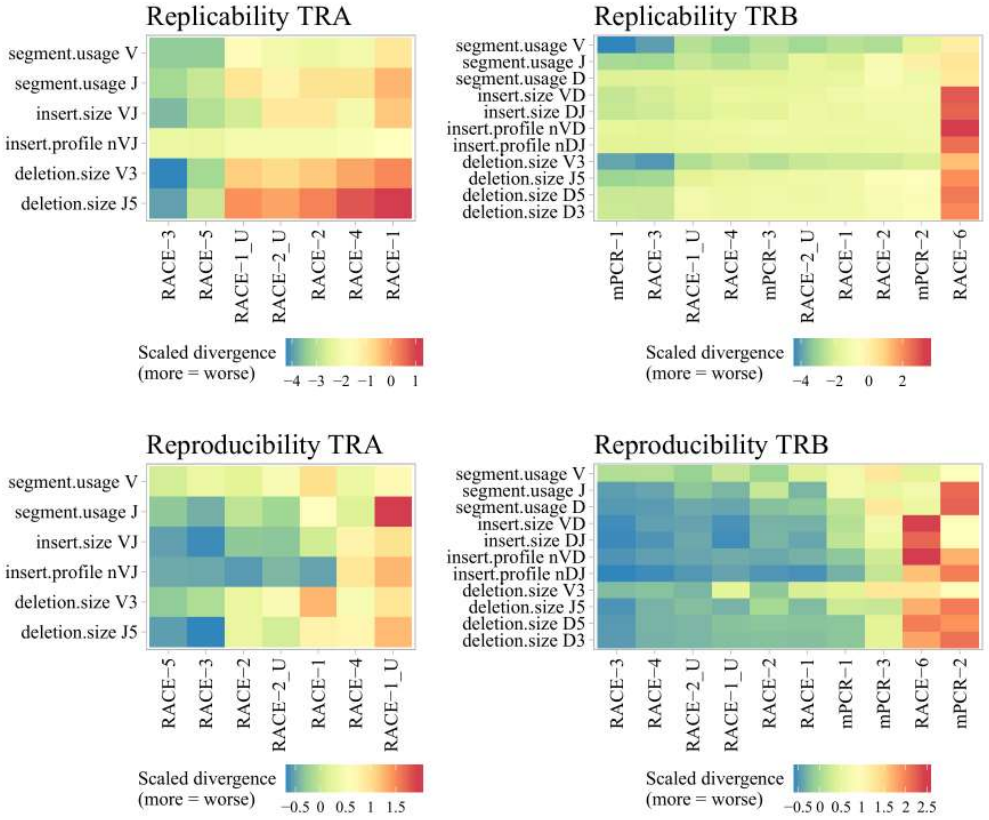
- Most used methods (publications)
- Voluntary based (challenge their method)
  - Academic labs
  - Commercially available kits
    - Service providers
    - Reagent providers

Experimental design



VDJ rearrangement

Replicability : similarity among data within the same method
Reproducibility : similarity among data between different methods

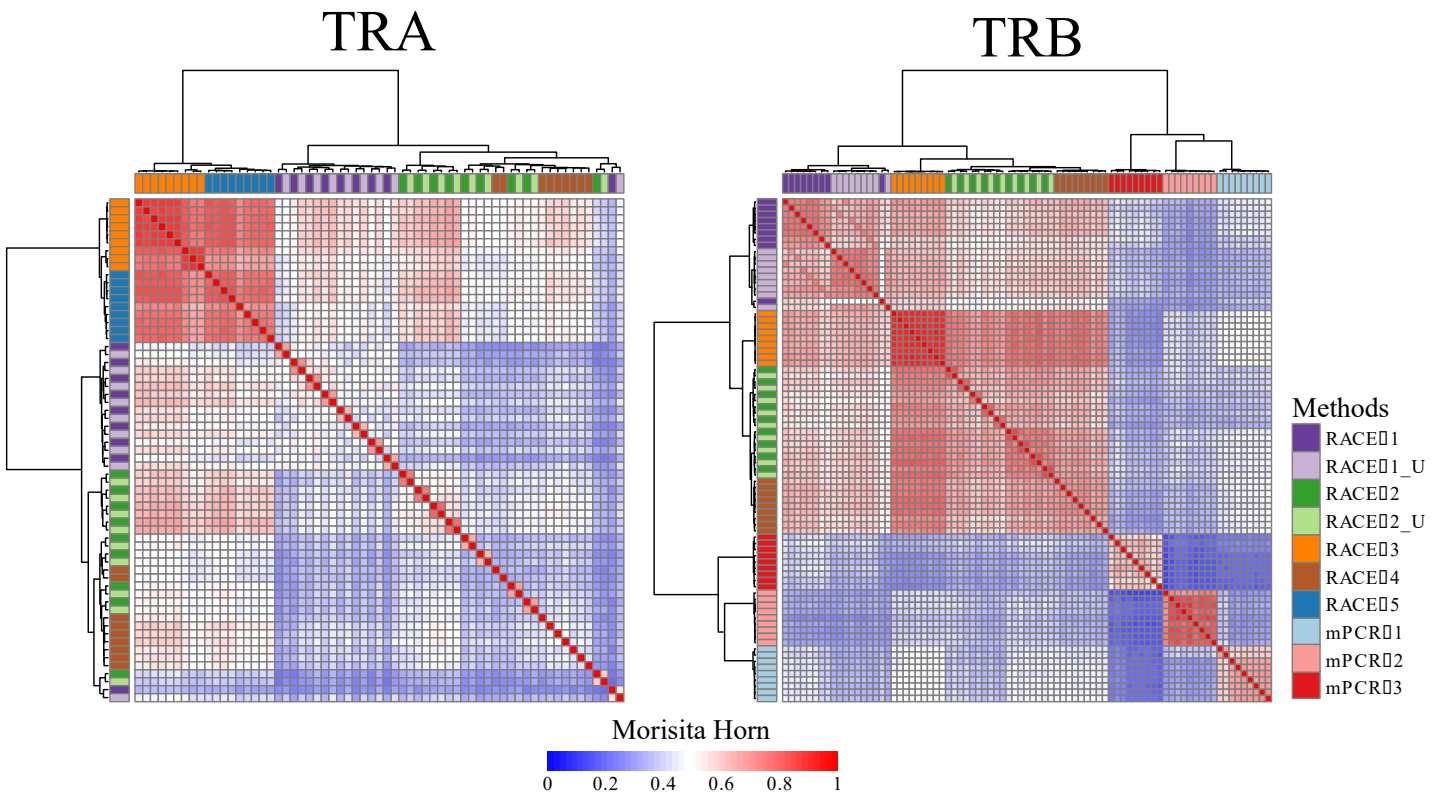


Methods differ between each other

- **basic parameters of the TCR repertoire**
- **inter-sample replicability and reproducibility**

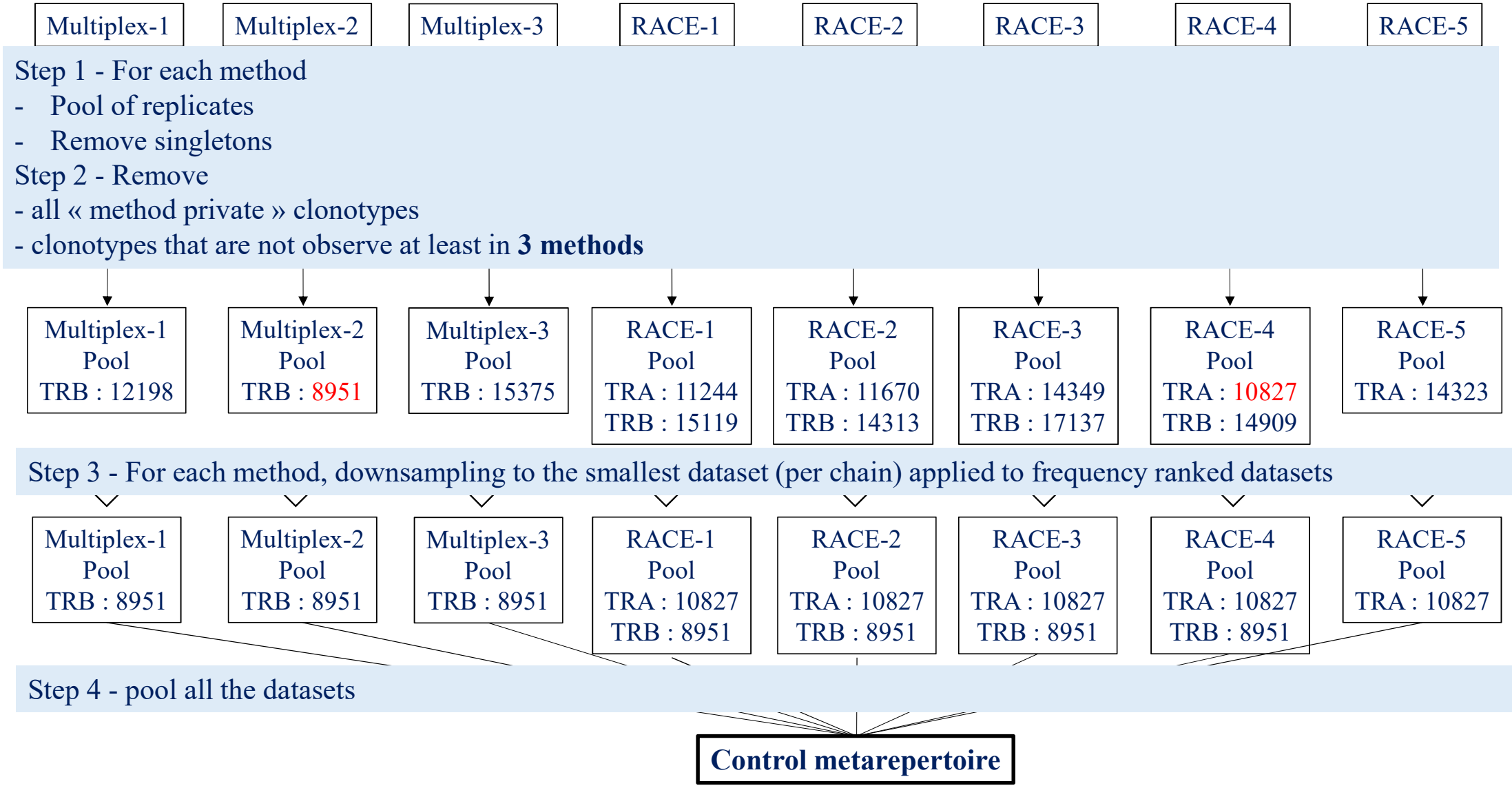
Columns are sorted according to the mean scaled distance (averaged over all rows) from the lowest (best replicability/reproducibility) to the highest (worst replicability/reproducibility)

Rearrangement similarity



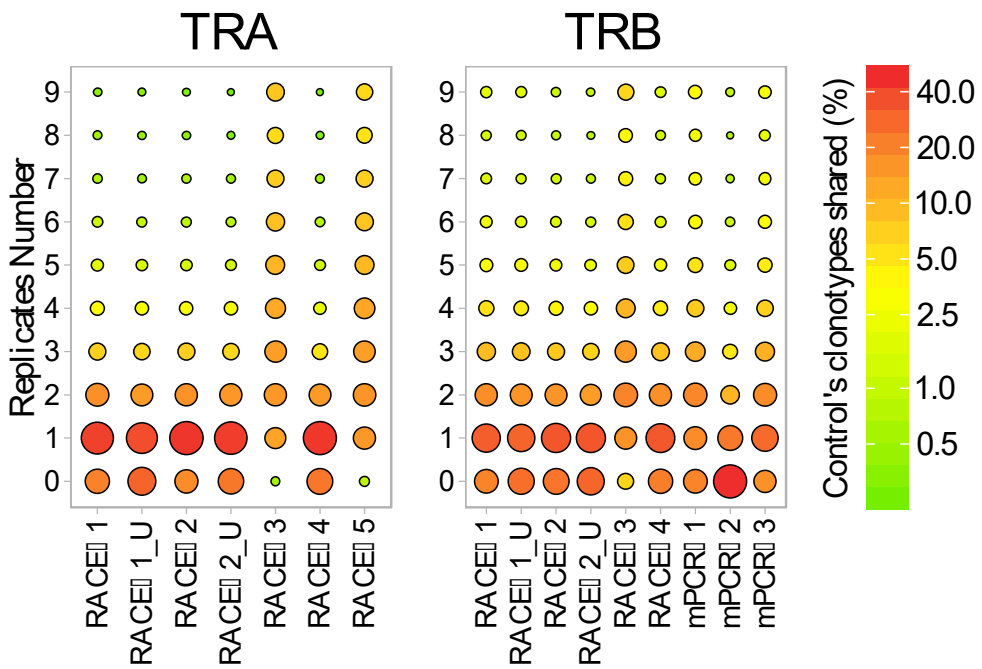
- In general, higher intra-method replicability for TRB than TRA chain, except RACE-3
- Major differences between 5'RACE and mPCR at the level of clonotype composition (same results using Jaccard distance)

Metarepertoire control to assess sensitivity & reproducibility



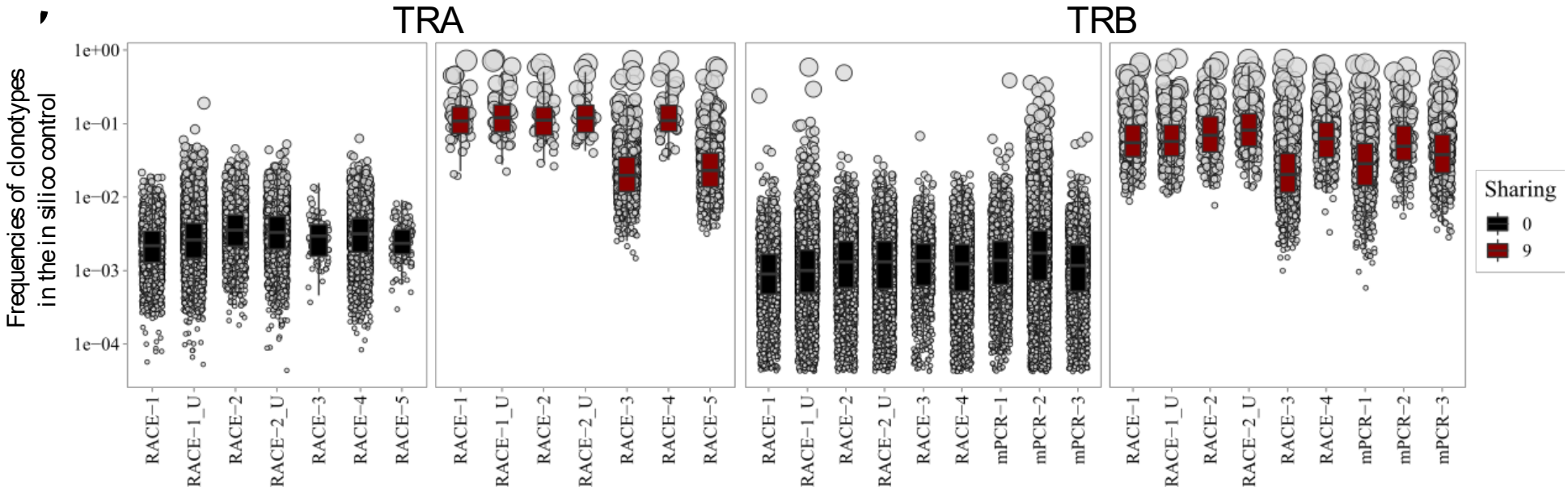
Replicability

% of clonotype from the metarepertoire control (MRC) found per method and per number of replicates (sharing)



- for TRA, RACE-3 and RACE-5 out-performed the other tested methods
  - capturing up to 20% of the MRCs in all 9 replicates (replicate number = 9)
  - missing (replicate number = 0) less than 1% of the MRCs
- for TRB, much less differences between the methods

Sensitivity



- RACE-3 and RACE-5 sensitivity are higher compared to all other methods

→ detect a larger proportion of clonotypes at lower abundances.

→ more obvious for TRA than for TRB

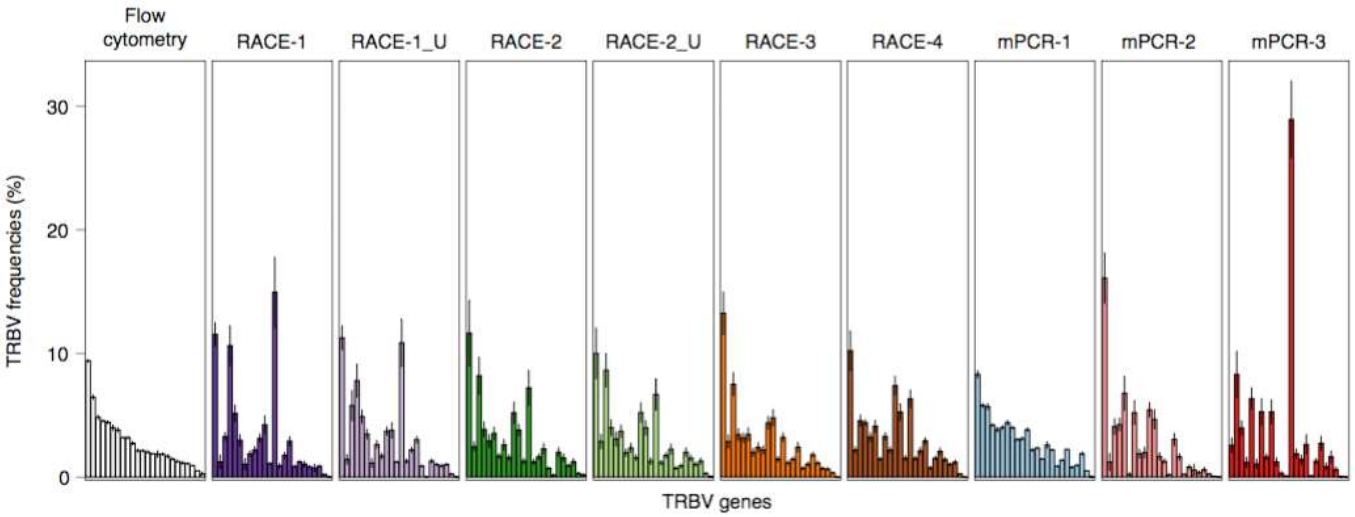
→ the other methods compared similarly between each other

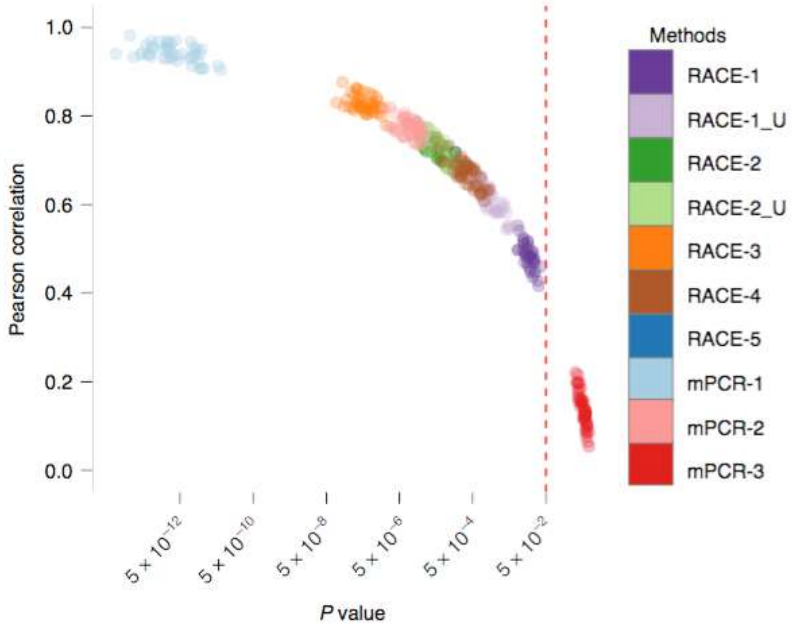
Summary

Table 1 | Comparative performance of the nine TCRseq molecular methods

TR chain	Method	Replicability	Reliability	Sensitivity	Cost per sample ($)	Controls and standards	Format type	fastq data availability
TRA	RACE-1	7	4	4	~230	-	Lab protocol	Yes
	RACE-1_U	4	5	4	~230	UMI	Lab protocol	Yes
	RACE-2	5	4	5	230–280	-	Service or kit	Yes
	RACE-2_U	4	5	5	230–280	UMI	Service or kit	Yes
	RACE-3	3	2	3	~150	-	Kit	Yes
	RACE-4	5	6	4	~150	-	Lab protocol	Yes
	RACE-5	2	3	3	~300	-	Lab protocol	Yes
TRB	mPCR-1	3	3	3	~350–550[a]	Synthetic TCRs	Service or kit	No
	mPCR-2	6	7	7	~25	-	Lab protocol	Yes
	mPCR-3	5	5	3	~350–550[a]	-	Service or kit	Yes
	RACE-1	6	5	4	~230	-	Lab protocol	Yes
	RACE-1_U	4	6	5	~230	UMI	Lab protocol	Yes
	RACE-2	6	6	6	230–280	-	Service or kit	Yes
	RACE-2_U	6	6	7	230–280	UMI	Service or kit	Yes
	RACE-3	2	2	3	~150	-	Kit	Yes
	RACE-4	3	5	4	~150	-	Lab protocol	Yes

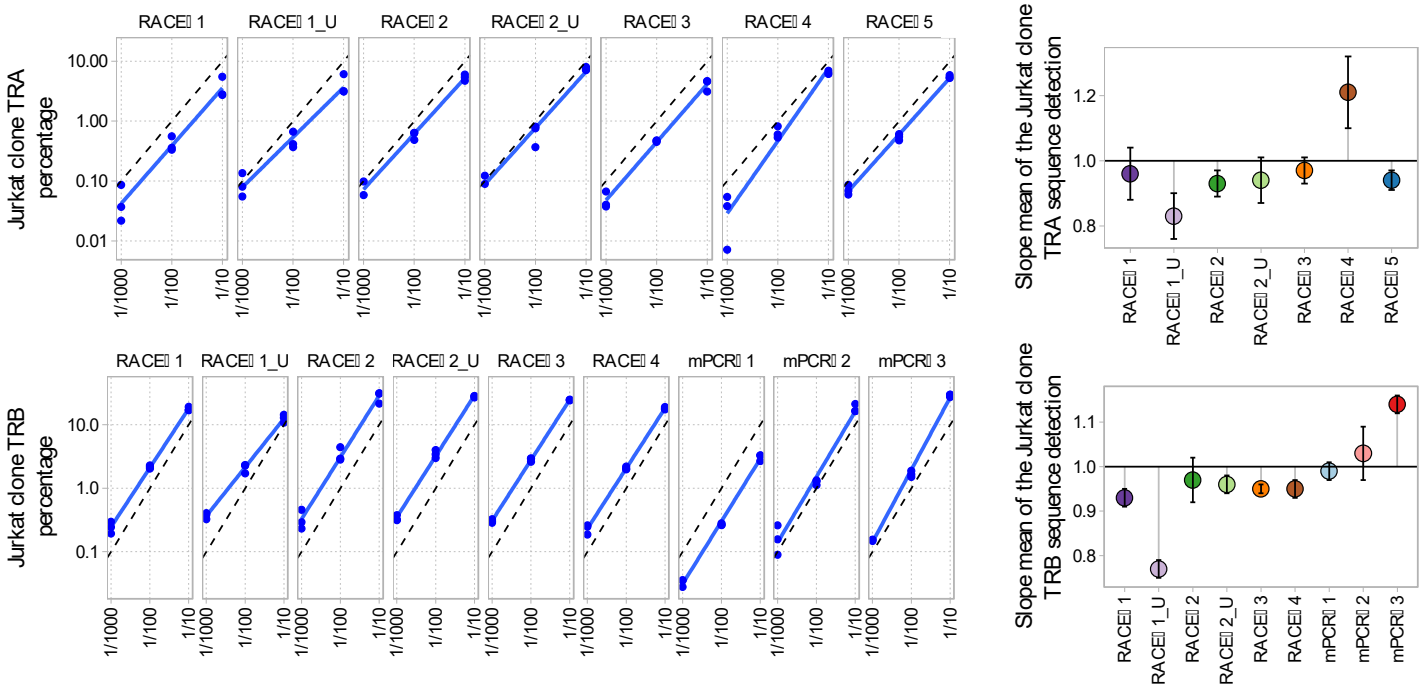
Controls - TRBV usage





- gDNA/mPCR methods have a good correlation with flow cytometry data
- RNA/RACE-based methods are variable, *r2* ranging from 0.4 to 0.8 ($P < 0.05$)
- mPCR-3 (RNA/mPCR-based) poorly correlates

Controls - clonal spike-in & detection sensitivity



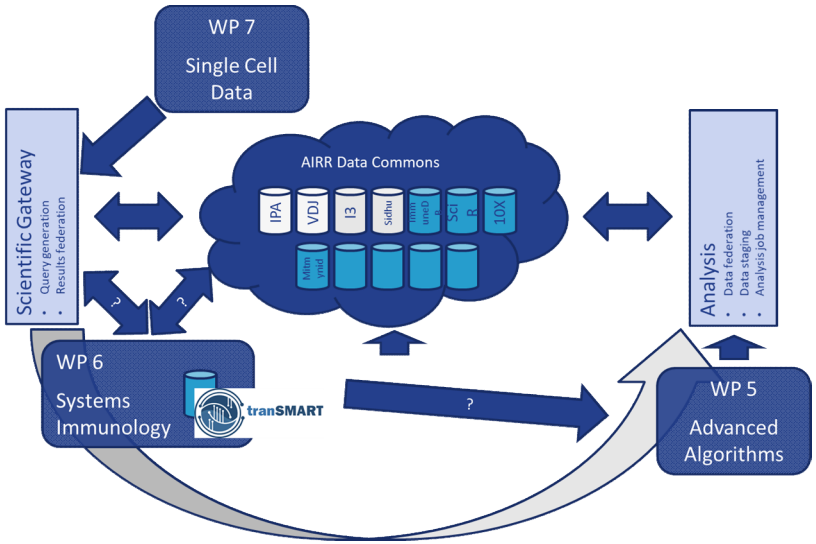
- All methods
  - TRA & TRB detected at all dilutions
  - TRA frequencies three times lower than expected
  - TRB frequencies three times higher than expected (opposite for mPCR-1)
  - Linearity between ratio maintained (except RACE-1_U, RACE-4 and mPCR-3)

Ongoing challenges

- We need federated tools to allow comparison and get valuable biological information out of those massive datasets

AIRR Community

iReceptor+ Sharing AIRR-seq data



Defining standards

- Data production
- Data storage & sharing
- Data exploration

Acknowledgments

i³ lab – D. Klatzmann

***i*mmunopathology (*i*P)**

David Klatzmann

Caroline Aheng

Michelle Rosenzwajg

Pierre Barennes

Gwladys Fourcade

Valentin Quiniou

Nicolas Coatnoan

Paul Stys

Hélène Vantomme

***i*ntegrative *i*mmunology (*i*2)**

Adrien Six

Nicolas Tchitchek

Vanessa Mhanna

Kenz Le Gouge

Aurélien Béliard

CIC-Bti

Michelle Barbié

DHU i2B

Francis Berenbaum

Jérémie Sellam

Collaborators

Anne Eugster (Dresden University, Germany)

Daniel C. Douek team (NIAID, USA)

Samuel Darko

Amy Ransier

Benny Chain team (UCL, UK)

Imran Uddin

Mazlina Ismail

Theres Oakes

Dmitriy M. Chudakov team (Skoltech, Russia)

Mikhail Shugay

Evgeniy S. Egorov

Alexey N. Davydov

Karl Kashofer (Graz Hospital, Austria)

Peter Rainer

External services

Abiba Doukani

Badreddine Mohand Oumoussa

Yannick Marie

Agnes Rastetter

Philippe Frogel

Véronique Dhennin

Julien Derop

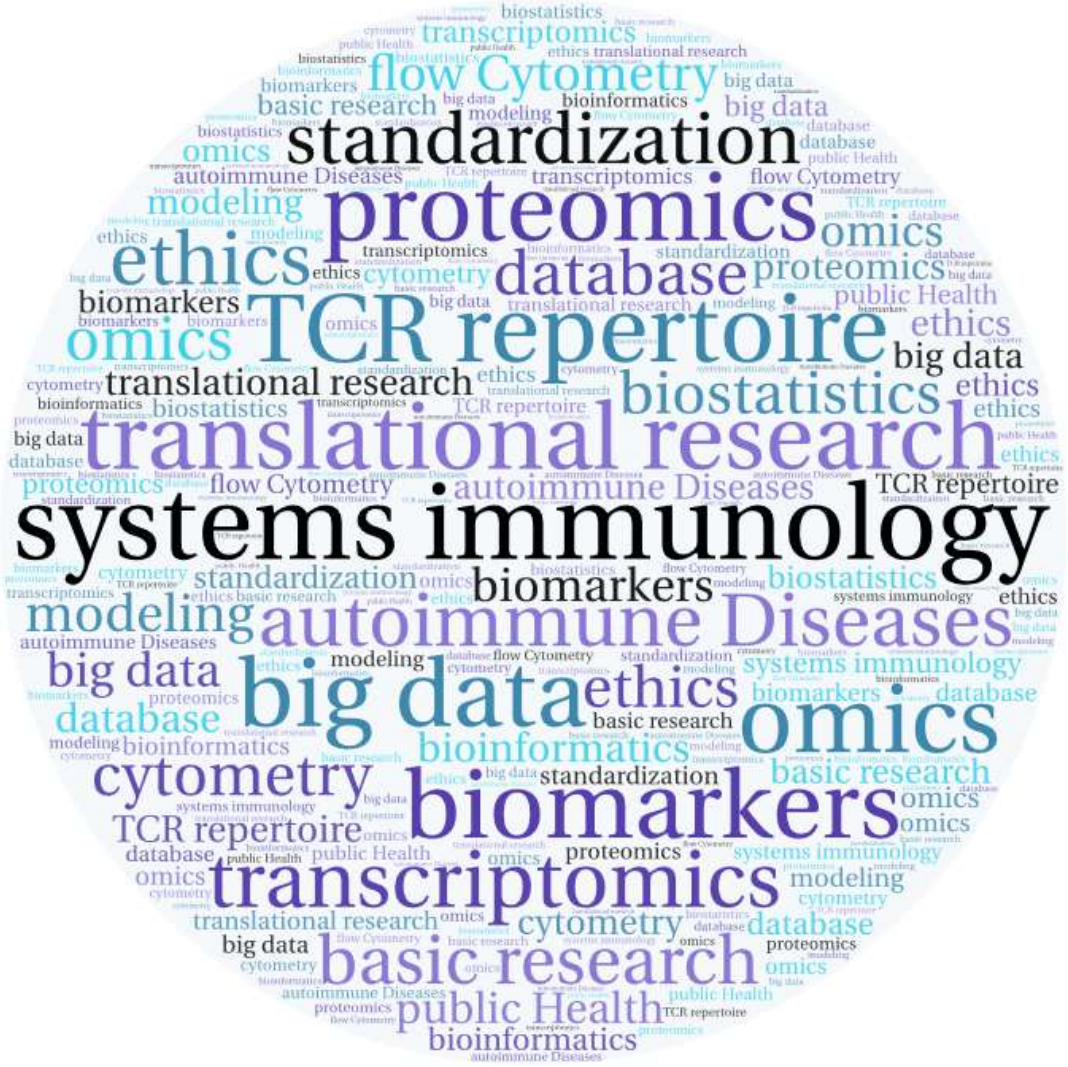
Amélie Bonnefond

iReceptor+

Gur Yaari (Bar Ilan University, Israel)

Felix Breden (SFU, Canada)

Brian Corrie


biostatistics
transcriptomics
flow Cytometry
ethics translational research
biomarkers
big data
basic research
bioinformatics
big data
modeling
standardization
database
omics
autoimmune Diseases
public Health
transcriptomics
flow Cytometry
modeling
proteomics
omics
ethics
transcriptomics
standardization
proteomics
ethics
cytometry
database
public Health
biomarkers
translational research
ethics
omics
TCR repertoire
big data
translational research
ethics
biostatistics
ethics
bioinformatics
biostatistics
TCR repertoire
ethics
big data
translational research
database
proteomics
flow Cytometry
autoimmune Diseases
TCR repertoire
systems immunology
cytometry
standardization
biomarkers
biostatistics
ethics
modeling
autoimmune Diseases
big data
modeling
flow Cytometry
standardization
systems immunology
big data
ethics
biomarkers
database
proteomics
omics
database
basic research
modeling
bioinformatics
bioinformatics
cytometry
big data
standardization
basic research
biomarkers
omics
TCR repertoire
omics
omics
database
public Health
proteomics
systems immunology
omics
transcriptomics
modeling
cytometry
translational research
omics
cytometry
database
big data
basic research
proteomics
omics
autoimmune Diseases
public Health
proteomics
public Health
TCR repertoire
bioinformatics

AIRR Community

https://www.antibodysociety.org/the-airr-community/

AIRR Community Meeting V: "Zooming in to the AIRR Community"

AIRR Community Meeting V – Zooming in to the AIRR Community! will include three 4-hour virtual sessions on **December 8th-10th, 2020** and will include AIRR-C Working Group & Sub-committee presentations, scientific sessions, interactive poster sessions and software tool demonstrations and networking events.

In addition to the main meeting, there are two free pre-meetings. Be sure to check out the pre-meeting links below for agenda and registration/connection details:

1) **"AIRR-seq in the Pandemic"** co-hosted by the AIRR Community and Tsinghua University in China will take place on December 5th/6th and

2) **"AIRR-seq Biological Standards and Workflows"** hosted the Biological Resources Working Group on December 7th.