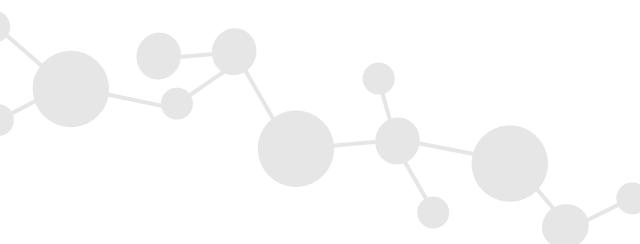


Identification of melanoma-specific T cell clonotypes in a murine model

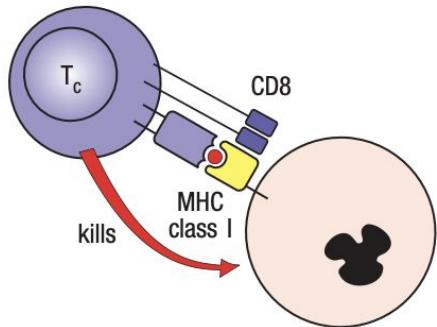


Student: *Ksenia Lupyr*

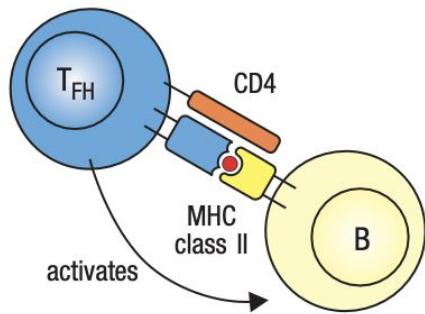
Research Advisor: *Dmitry Chudakov*

Research Co-advisor: *Olga Britanova*

Introduction

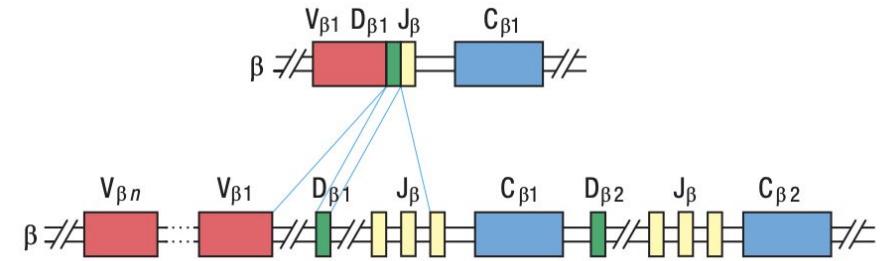


CD8+ cells kill infected and **cancer** cells

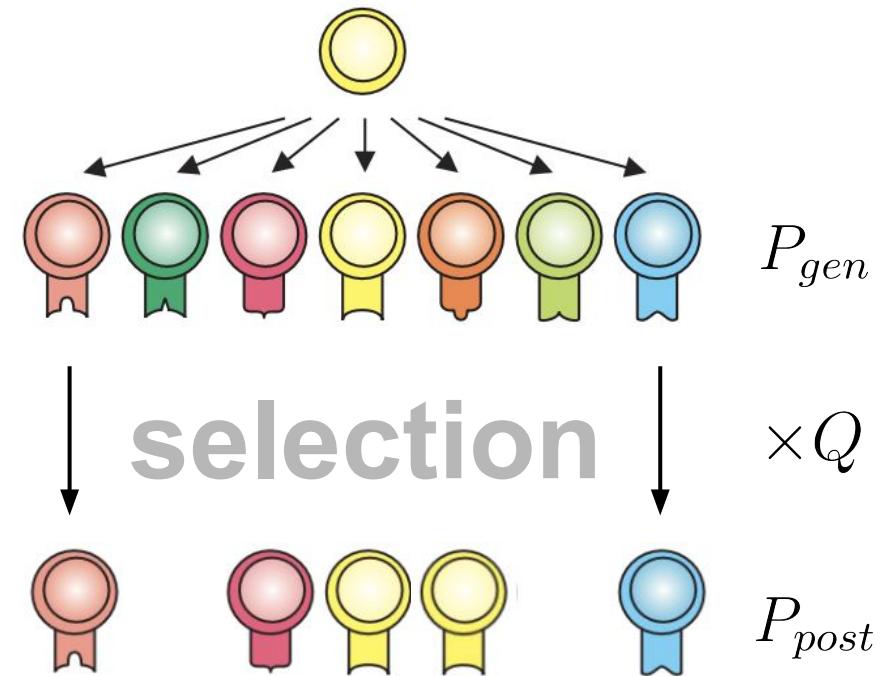


CD4+ cells regulate immune response by activating (**Th**) and inhibiting (**Treg**)

Some TCRs are shared between individuals (**public** TCRs) but most of them are specific for donor



$$P_{gen}(\sigma) = P_V \times P_{DJ} \times P_{del} \times P_{ins}$$

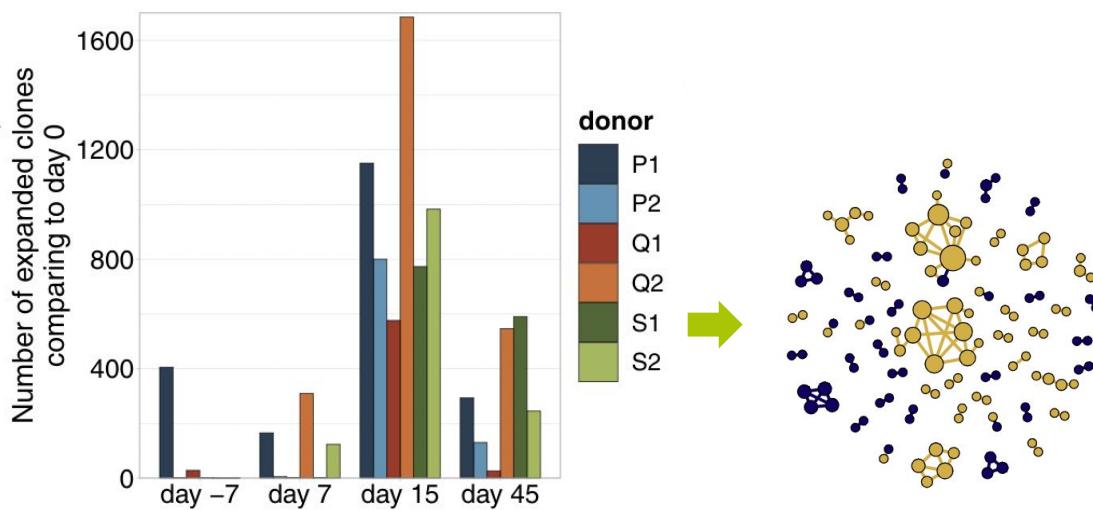




Introduction

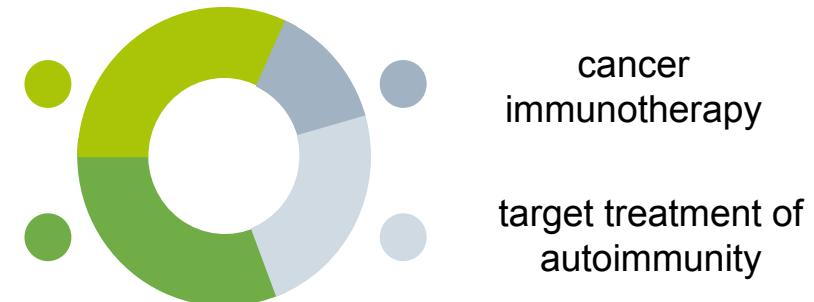
Immune system responds to the challenge by numerous T cells
many of which have homologous TCRs

Example: Yellow fever vaccination



Identification of specific TCRs

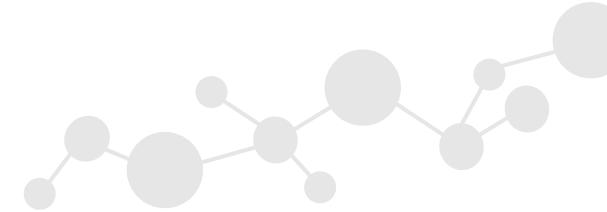
fundamental
research
vaccine
design



Challenges in the area

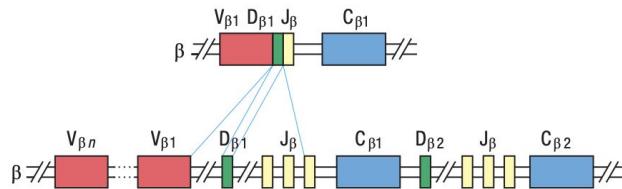
the number of TCRs with established specificity is limited

the quality of available data is not always properly controlled

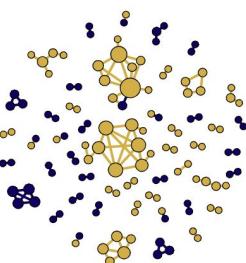


Introduction

How to identify condition specific TCRs?



$$P_{gen}(\sigma) = P_V \times P_{DJ} \times P_{del} \times P_{ins}$$



Find shared TCRs
with low generation
probability

OLGA, SONIA

Find clusters of
similar TCRs

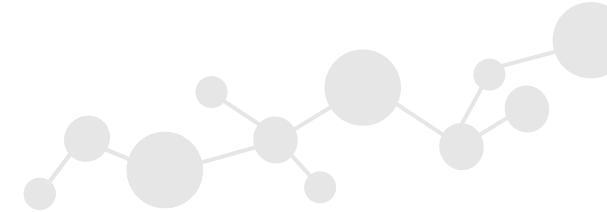
TCRdist, GLIPH2, etc

Find clonotypes with significantly high
number of neighbors, taking into account
their generation probabilities

ALICE

Limitations:

- Only human model
- Time consuming and not user-friendly

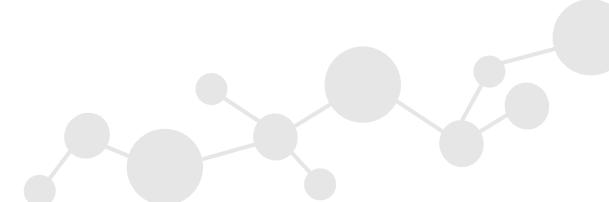


Aim and objectives

The aim of this study is to reveal the key melanoma-associated murine TCRs

Objectives

- 1 Develop an R library for the identification of statistically significant clusters of similar TCRs from a mouse or human TCR repertoire
- 2 Train generation probability model with IGOR on non-functional data and compare it with standard OLGA model
- 3 Compare developed tool with existing methods
- 4 Develop a procedure for working with a set of samples
- 5 Find melanoma specific TCRs with developed method

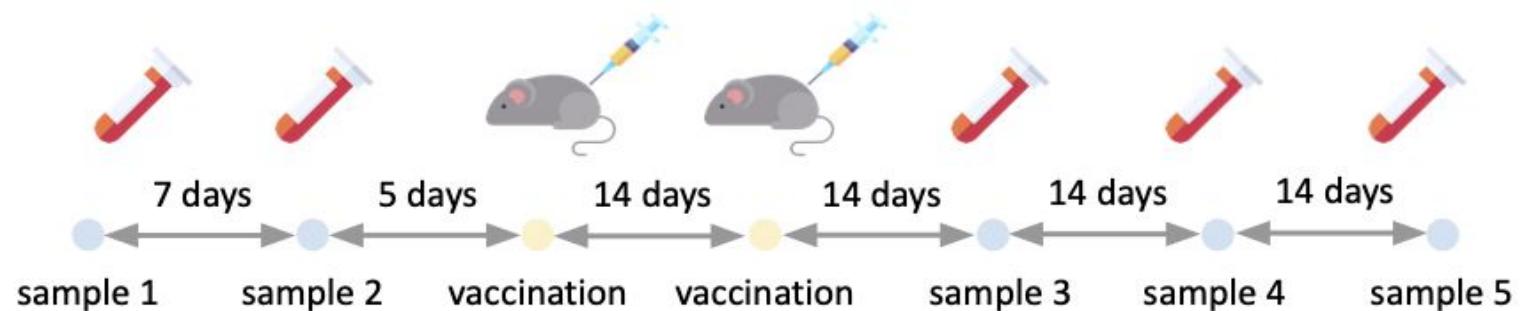


Results and discussion: TCRgrapher

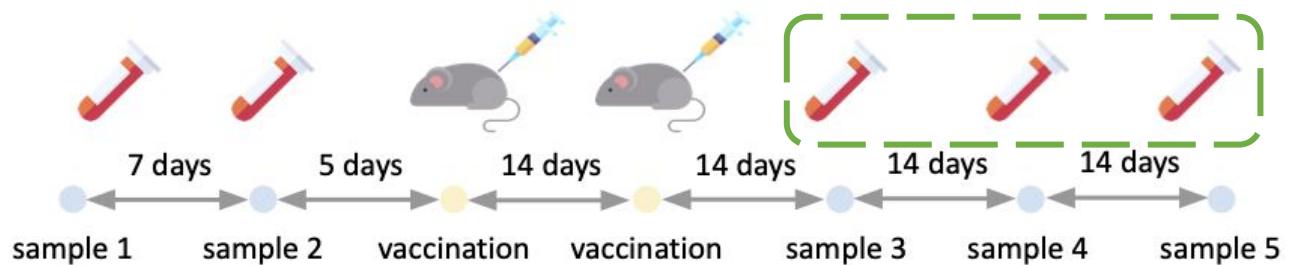
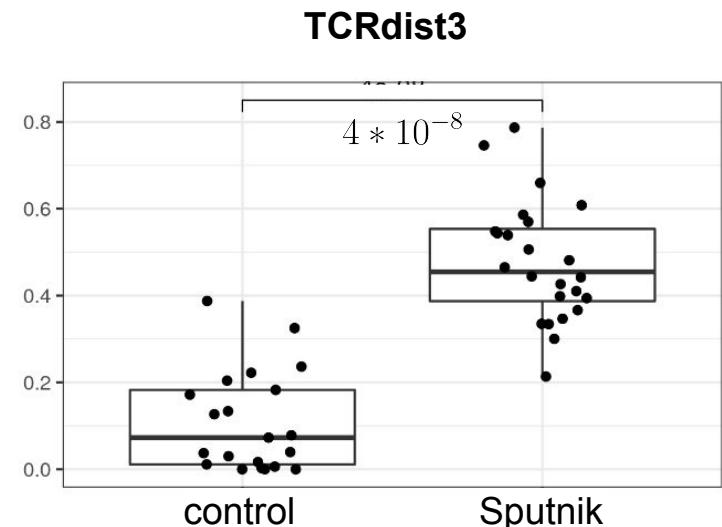
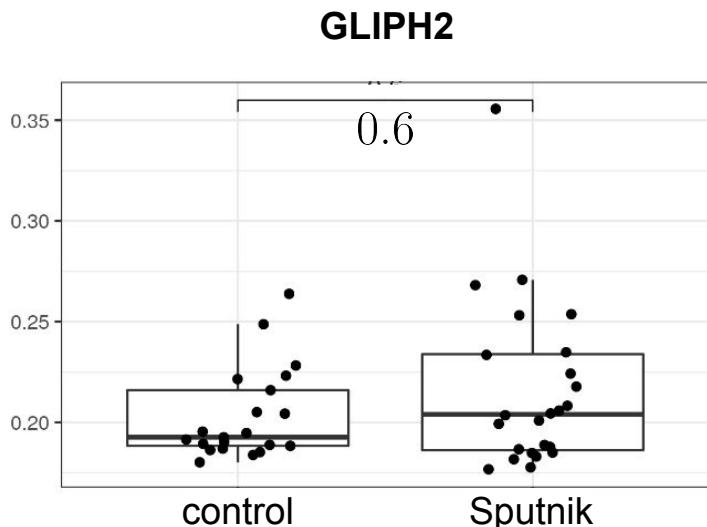
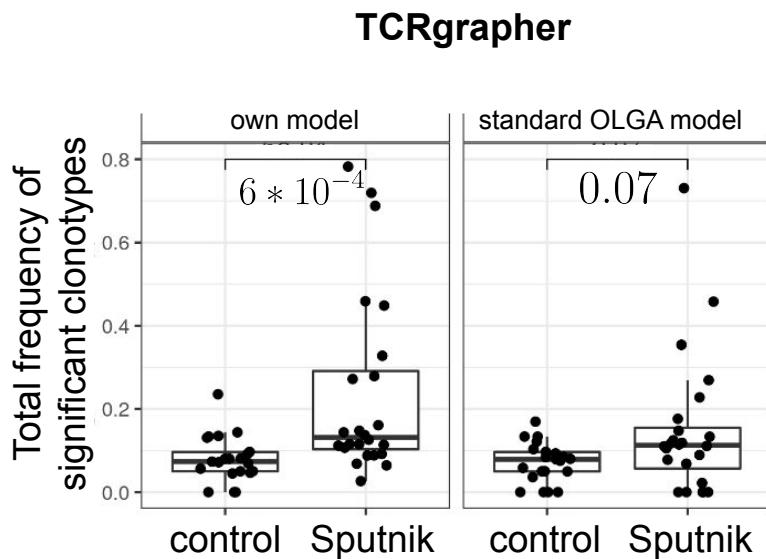
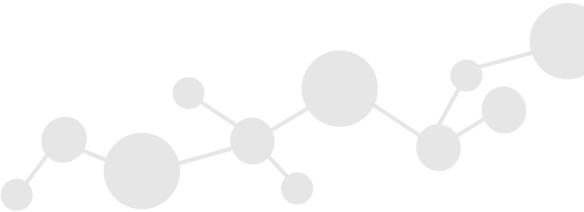
R package for identifying condition associated T cell clones

	TCRgrapher	ALICE	TCRdist3	GLIPH2
both for human and murine data	+	✗	+	+
doesn't require control	+	+	✗	+
fast analysis of large datasets	+	✗	✗	✗
working principle	probability + neighborhood	probability + neighborhood	neighborhood + comparison with control	probability + neighborhood

TCRgrapher was tested on TCR
repertoires of **mice vaccinated**
with Sputnik V

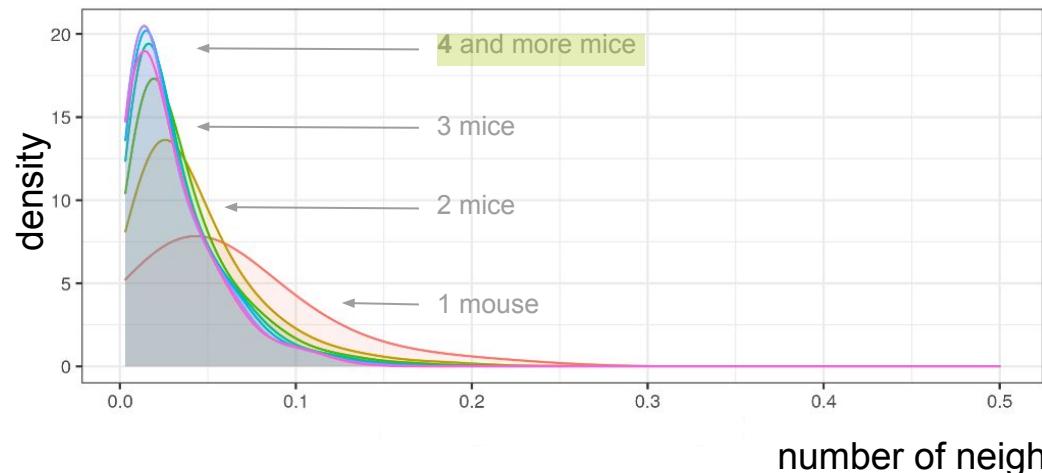


Analysis of individual mice after vaccination

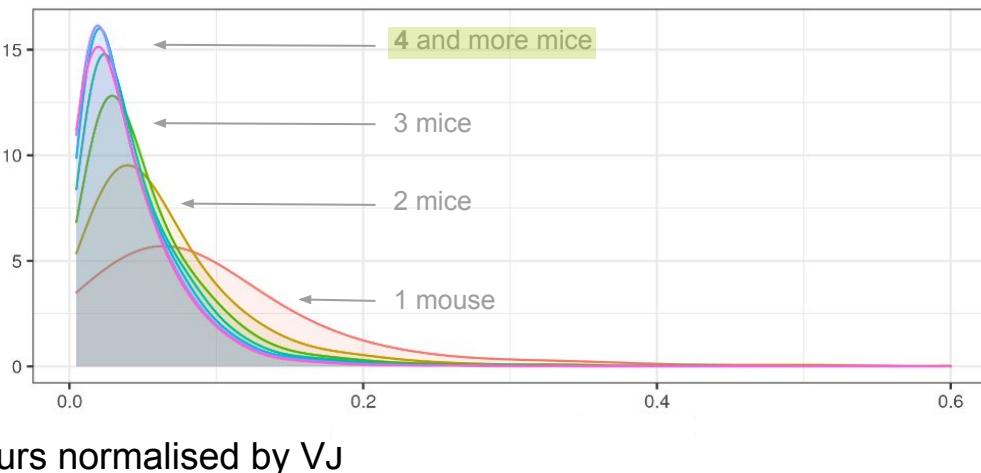


TCRgrapher: How number of samples affects the result

before vaccination

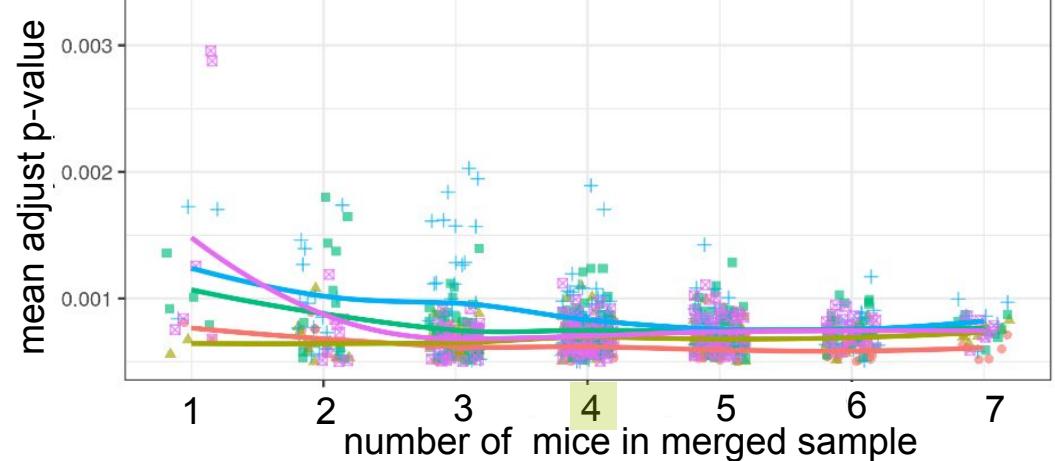
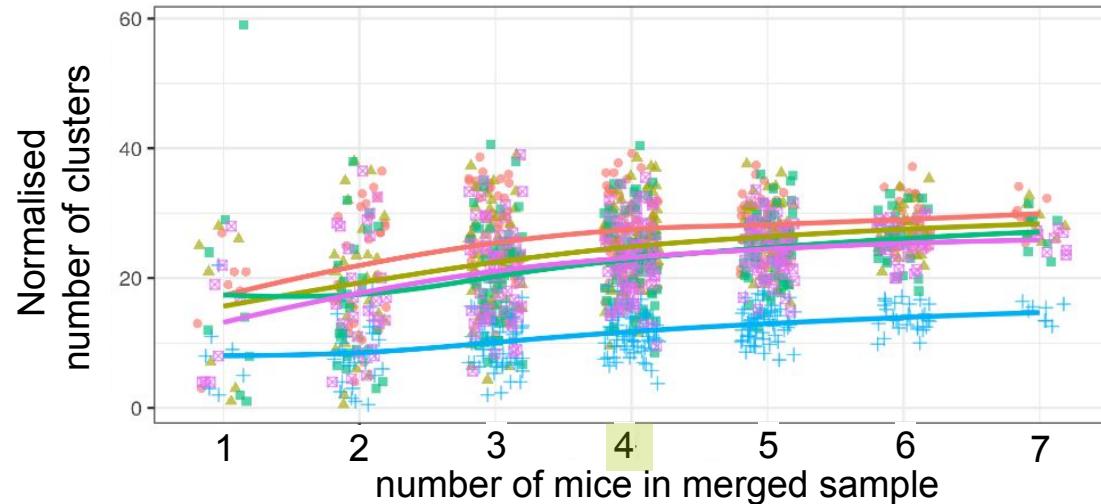


after vaccination



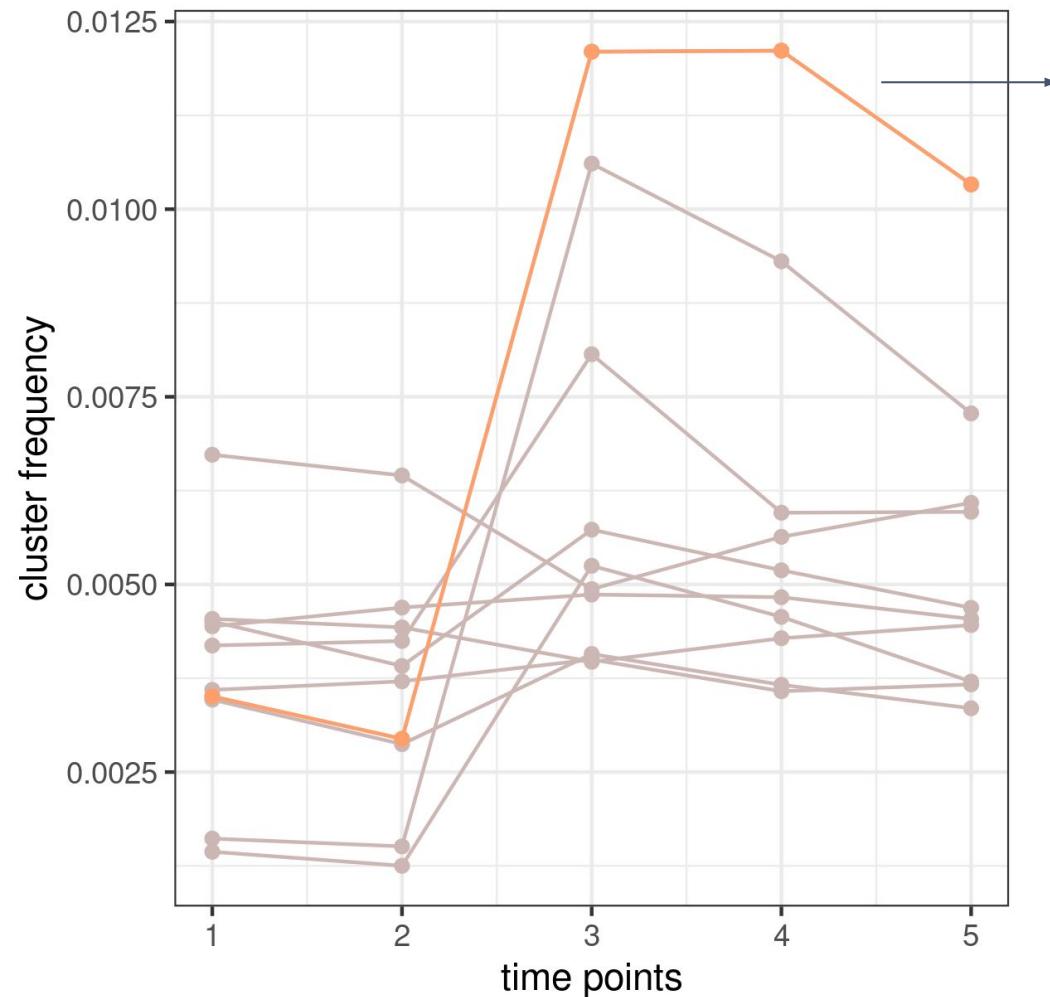
number of mice in merged sample

1
2
3
4
5
6
7

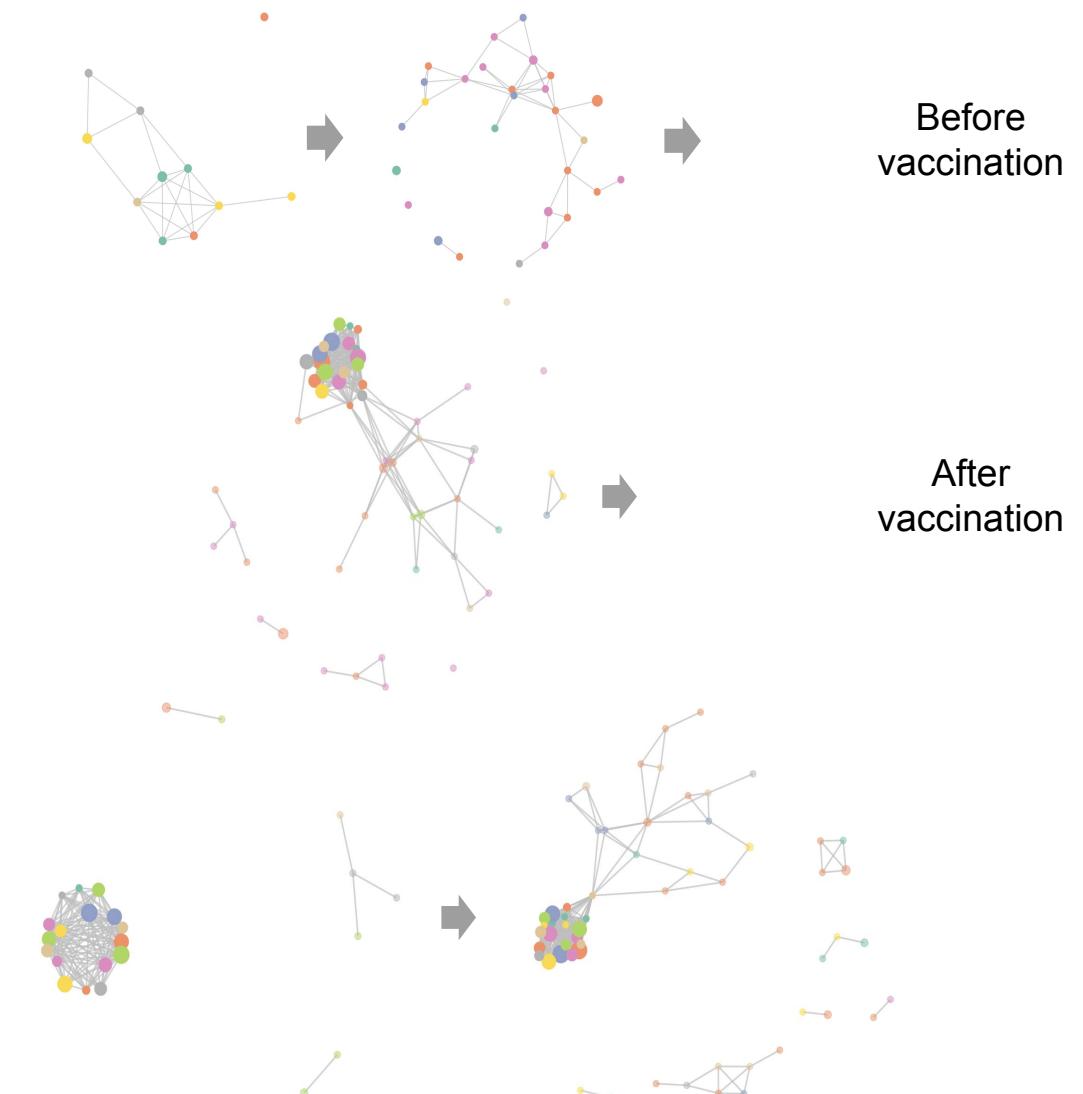


Cluster formation after vaccination

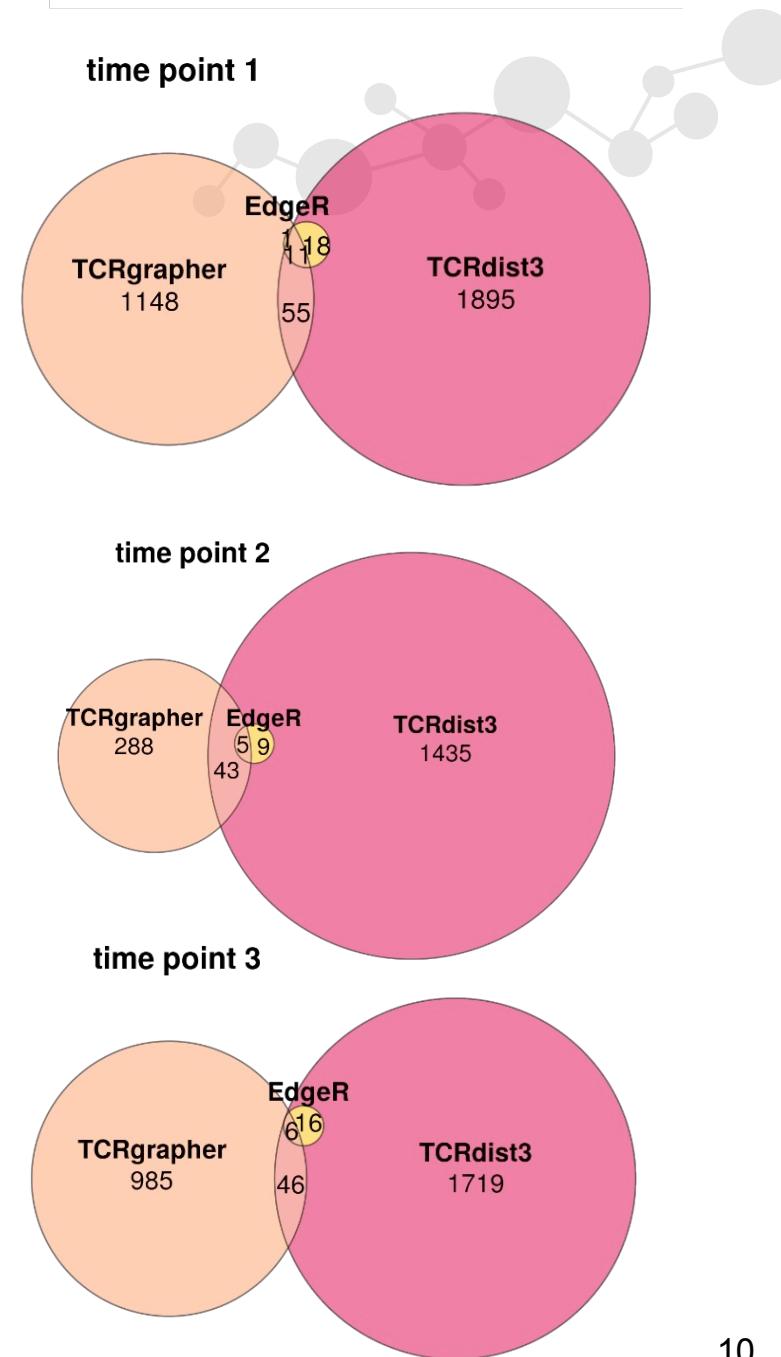
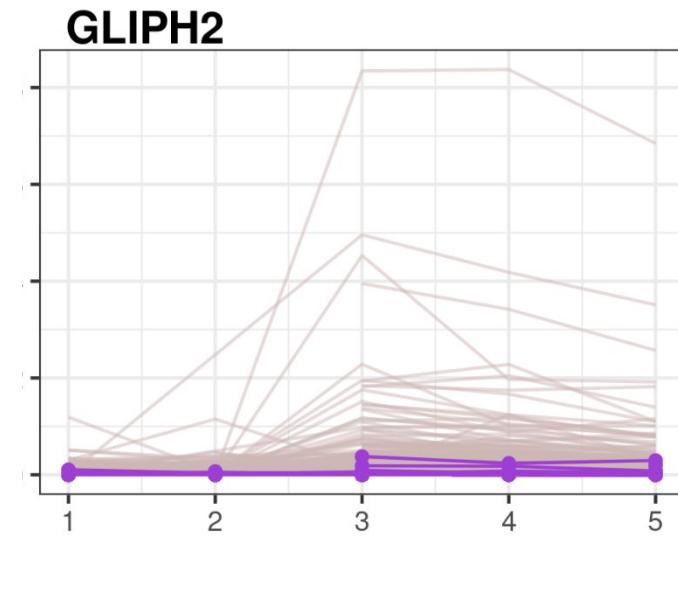
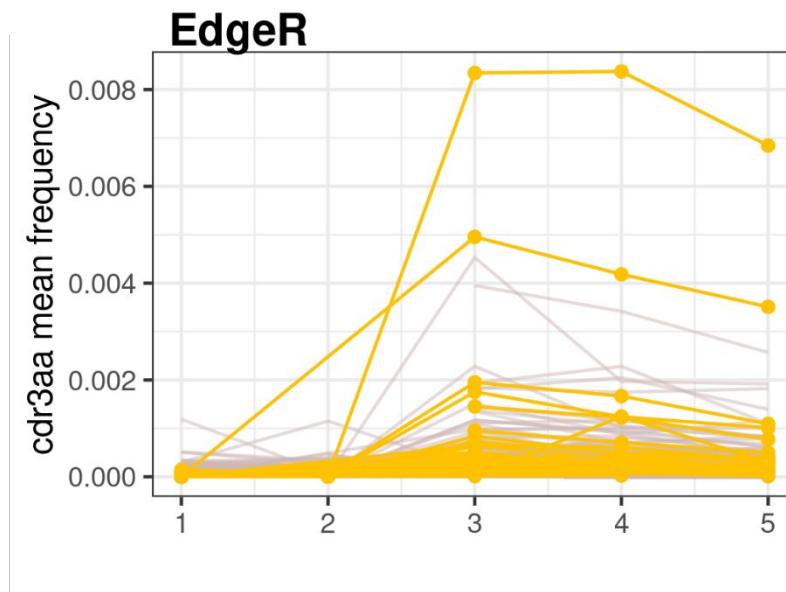
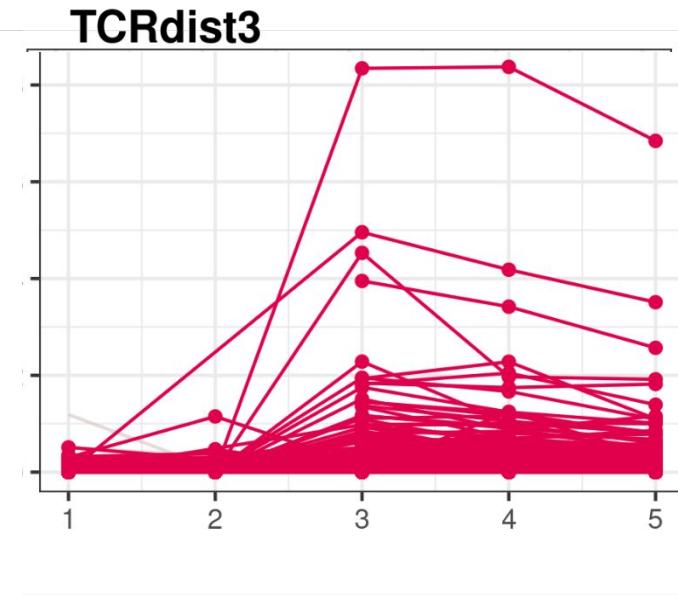
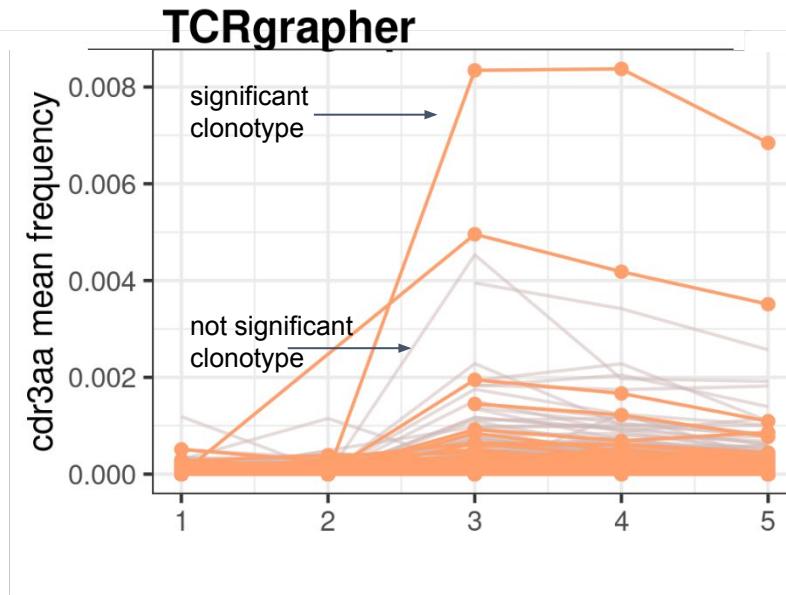
Top 10 clusters after vaccination

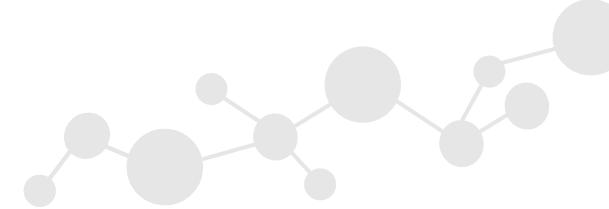


The most abundant cluster after vaccination



Analysis result of merged samples





Introduction



Ugur Sahin

Exploiting the Mutanome for Tumor Vaccination

John C. Castle, Sebastian Kreiter, Jan Diekmann, et al.
Cancer Res 2012

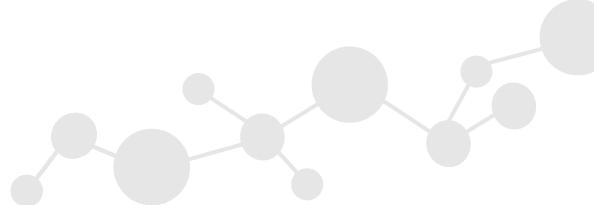
Mutant MHC class II epitopes drive therapeutic immune responses to cancer

Kreiter et al.
Nature. 2015 April 30;

- The first mouse tumor exome capture study, identifying **B16F10** somatic point mutations
- Authors explored T cell activity against neoepitopes with mutations. Response was mainly **CD4+**
- Two peptides were tested for antitumoral activity. In both cases growth of tumour was inhibited

peptide id	peptide name	sequence
p5	B16-M05-Eef2	FVVKAYLPVNESFAFTADLRSNTGGQA
p12	B16-M12-Gnas	TPPPSEEAMPFEFNGPAQGDHSQPPLQV
p17	B16-M17-Tnpo3	VVDRNPQFLDPVLAYLMKGKCEKPLAS
p20	B16-M20-Tubb3	FRRKAFLHWYTGEAMDEMEFTEAESNM
p22	B16-M22-Asf1b	PKPDFSQLQRNILPSNPRVTRFHINWD
p25	B16-M25-Plod1	STANYNTSHLNNNDVWQIFENPVDWKEK
p27	B16-M27-Obsl1	REGVELCPGNKYEMRRHGTTHSLVIHD
p28	B16-M28-Ppp1r7	NIEGIDKLTQLKKPFLVNNKINKIENI
p30	B16-M30-Kif18b	PSKPSFQEFDWENVSPELNSTDQPFL
p33	B16-M33-Pbk	DSGSPFPAAVILRDALHMARGLKYLHQ
p44	B16-M44-Cpsf3l	EFKHIKAFDRTFANNPGPMVVATPGM
p47	B16-M47-Rpl13a	GRGHILLGRLAAIVGKQVLLGRKVVVVR
p48	B16-M48-Def8	SHCHWNDLAVIPAGVVHNDFEPRKVS
p50	B16-M50-Sema3b	GFSQPLRRLVLHVVSAAQAEERLARAEE

Results and discussion

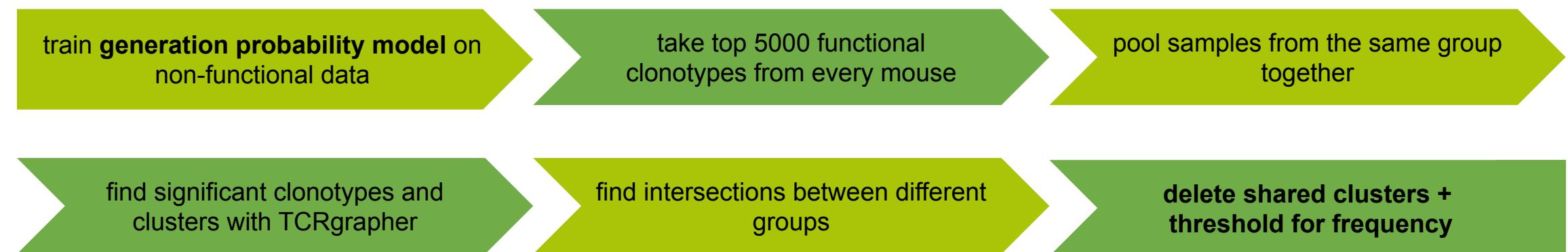


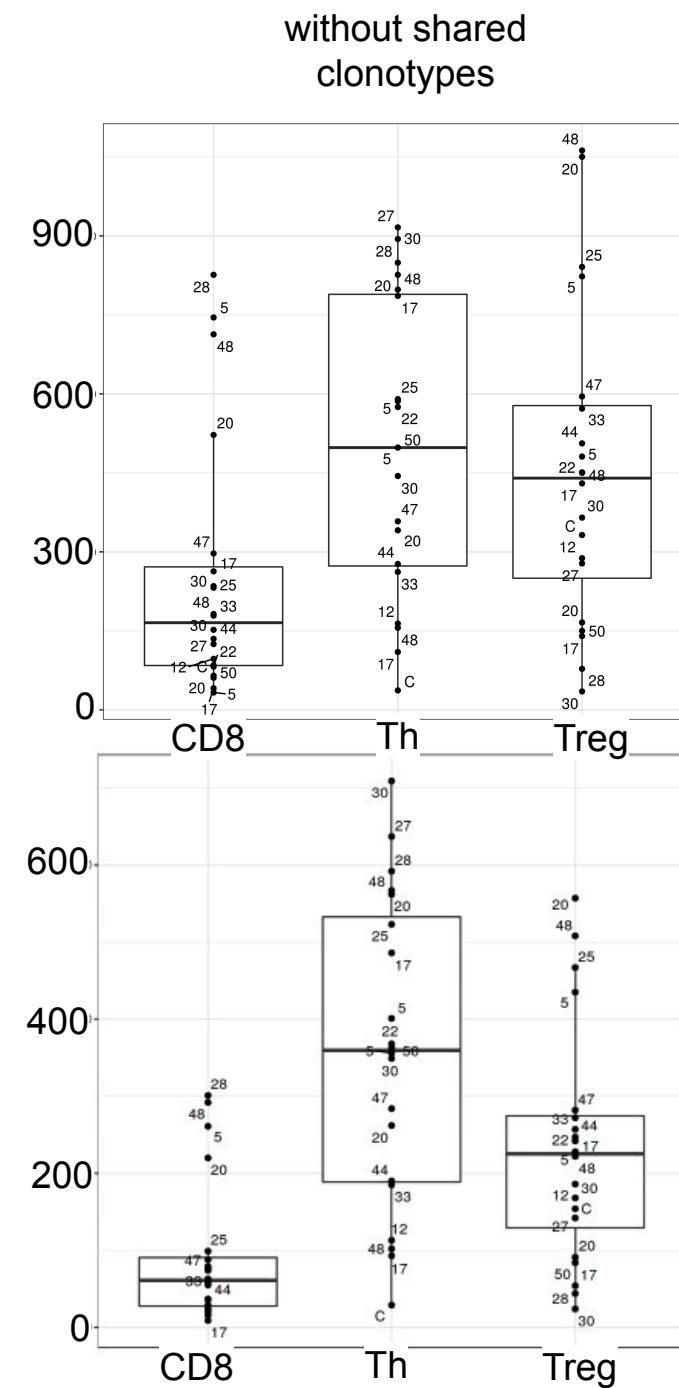
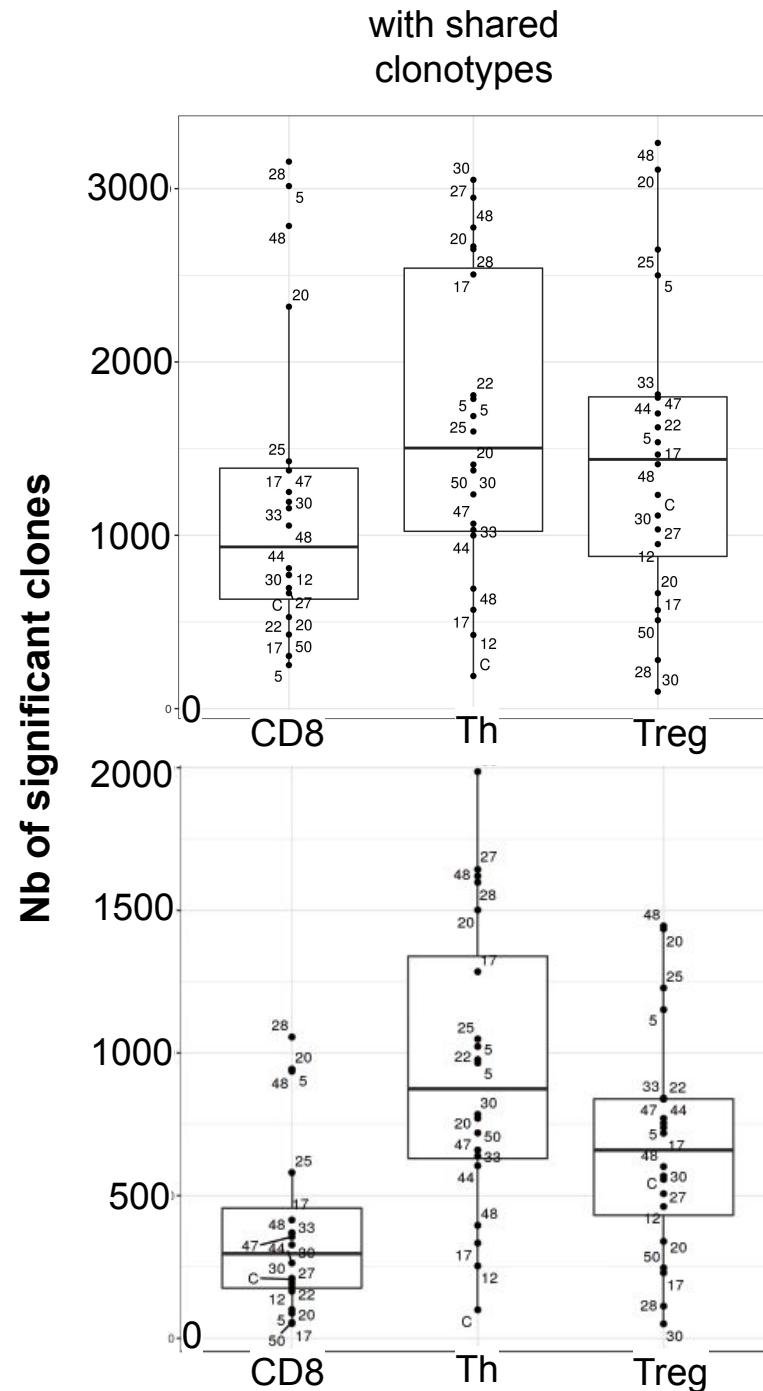
Experiment design

14 peptides from B16 melanoma cells
5 or 3 mice per group, one peptide per group
5 peptides had two experimental mice groups



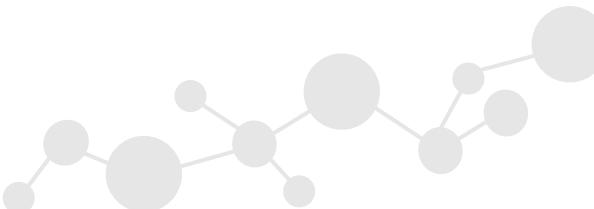
Analysis pipeline



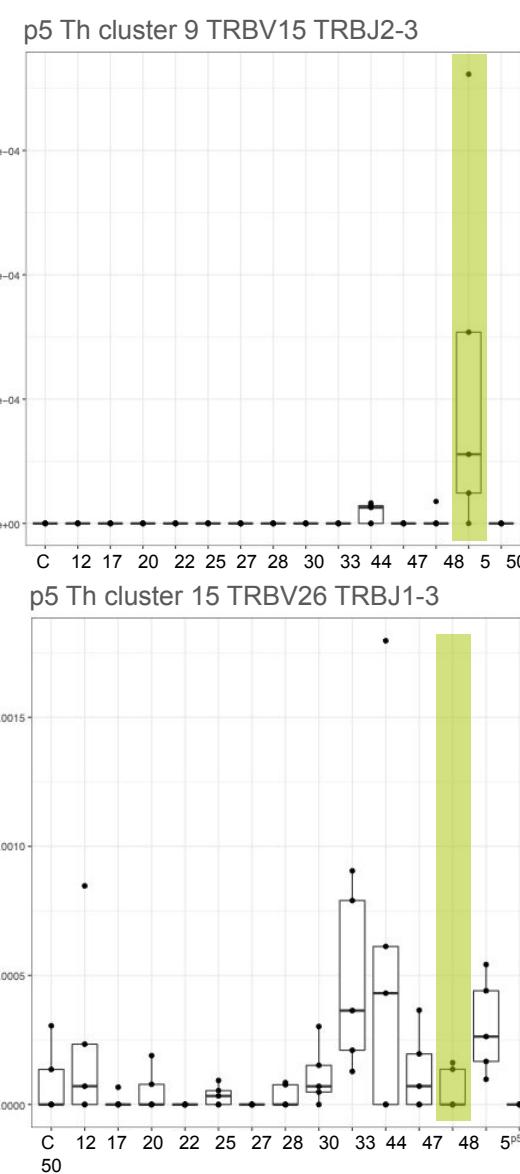
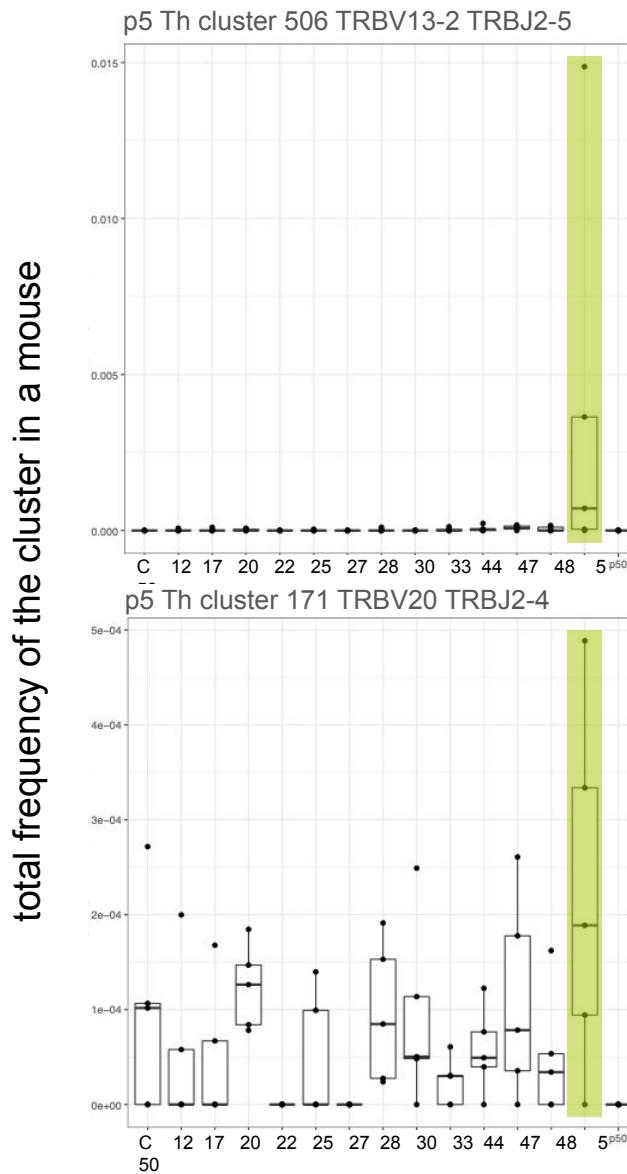


OLGA
standard
model

Own
models
(for every
cell type)



Clusters with total frequency more than 0.25% showed high specificity



p5
examples

> 0.25 %

< 0.25 %

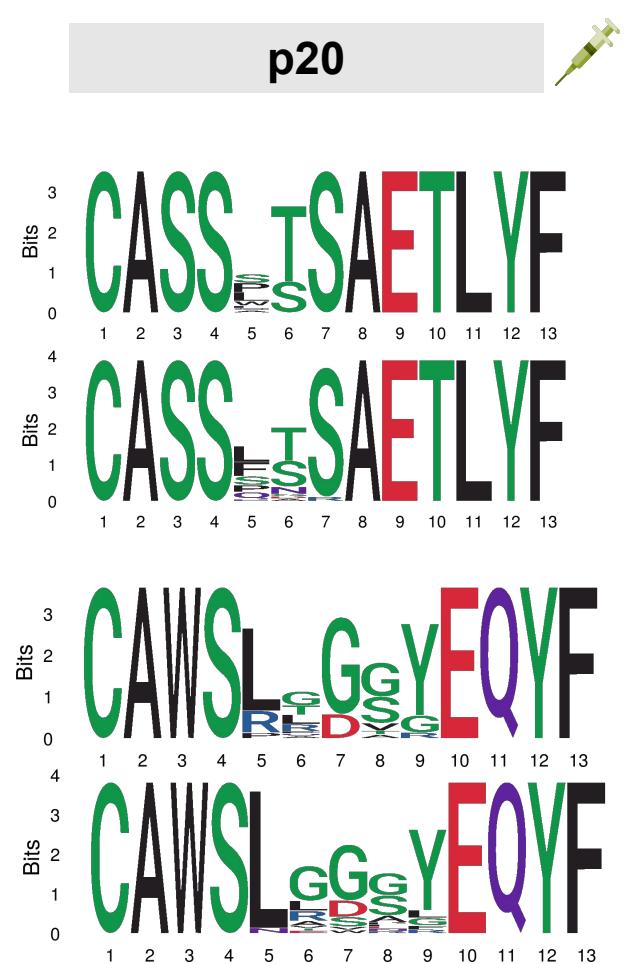
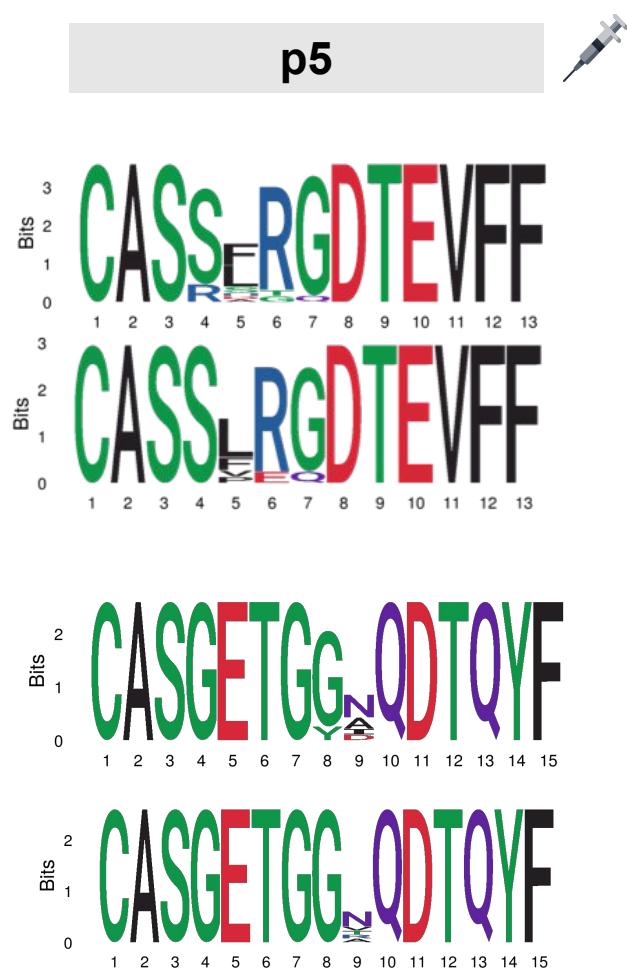
number of clusters in Th cells
with total frequency more than
0.0025

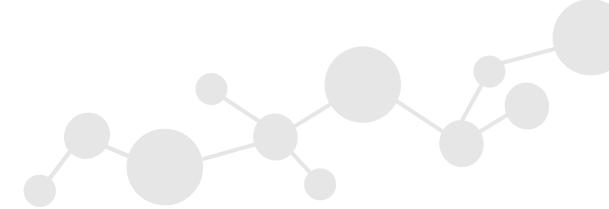
Peptide	Type	Replica	Nb of identified clusters
12	Th	0	9
17	Th	1	1
17	Th	2	11
20	Th	1	4
20	Th	2	16
22	Th	0	6
25	Th	0	27
27	Th	0	12
28	Th	0	16
30	Th	1	14
30	Th	2	17
33	Th	0	3
44	Th	0	7
47	Th	0	10
48	Th	1	4
48	Th	2	13
5	Th	1	6
5	Th	2	13
50	Th	0	13
C	Th	0	1

Results and discussion



Examples of reproducible clusters in Th

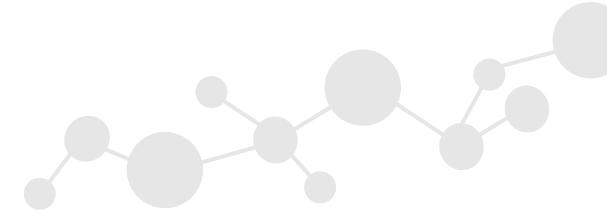




Conclusions

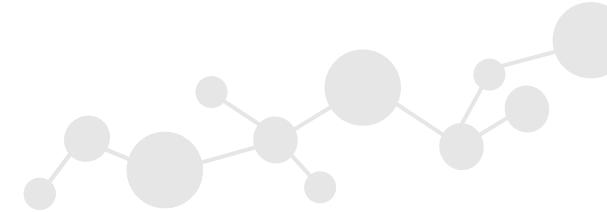
1. A method, TCRgrapher, for the identification of condition-associated TCRs in human and murine repertoires **was developed**. TCRgrapher **does not require control samples and is high-performance** compared with ALICE, GLIPH2 and TCRdist3 especially on large datasets.
2. The sufficient number of biological replicates was determined. TCR repertoire from **four mice** was sufficient to obtain the majority of antigen-specific TCR beta variants in the given experimental setting.
3. **Generation probabilistic models** were inferred for the repertoires obtained from PBMC and sorted CD8+, Th and Tregs cell subsets. The models trained on the available data **showed better results in comparison with OLGA standard model**: they recorded a statistically significant response to the vaccination with Sputnik and decreased false positive rate of identified melanoma-associated clonotypes.
4. Cluster searching approaches (TCRgrapher, TCRdist3) provided much more information than just searching for expanded clonotypes (EdgeR). TCRdist3 seemed to be the best choice if the control group presents.
5. The importance of cluster frequency in determining its specificity in the context of this work was shown. **Total frequency threshold** for the cluster was adjusted. CDR3 motifs of the TCR clusters specific to B16F10 melanoma antigens above this threshold were identified.
6. Significantly overrepresented TCR clusters in the repertoire of CD8+, Treg and Th cell subsets from the mice vaccinated with melanoma peptides were identified. The largest clusters were in Th, as expected. The mean number of clusters per vaccinated group was 11.

Melanoma-associated TCR database with identified clusters is available online
<https://github.com/KseniaMIPT/B1610-melanoma-associated-murine-TCRs>



Outlook

- Writing article about TCRgrapher and analysis of TCR repertoires of mice vaccinated with Sputnik V is in process
- Database with TCRs specific to melanoma peptides will be used in the future research
- My colleagues and I are going to write an article about anti-CTL4 therapy in a murine model and information about TCRs with established specificity will be used in it



Acknowledgements

I express sincere thanks to my colleagues for experiment design, library preparation and supervision

George V. Sharonov

Irina. A. Shagina

Diana V. Yuzhakova

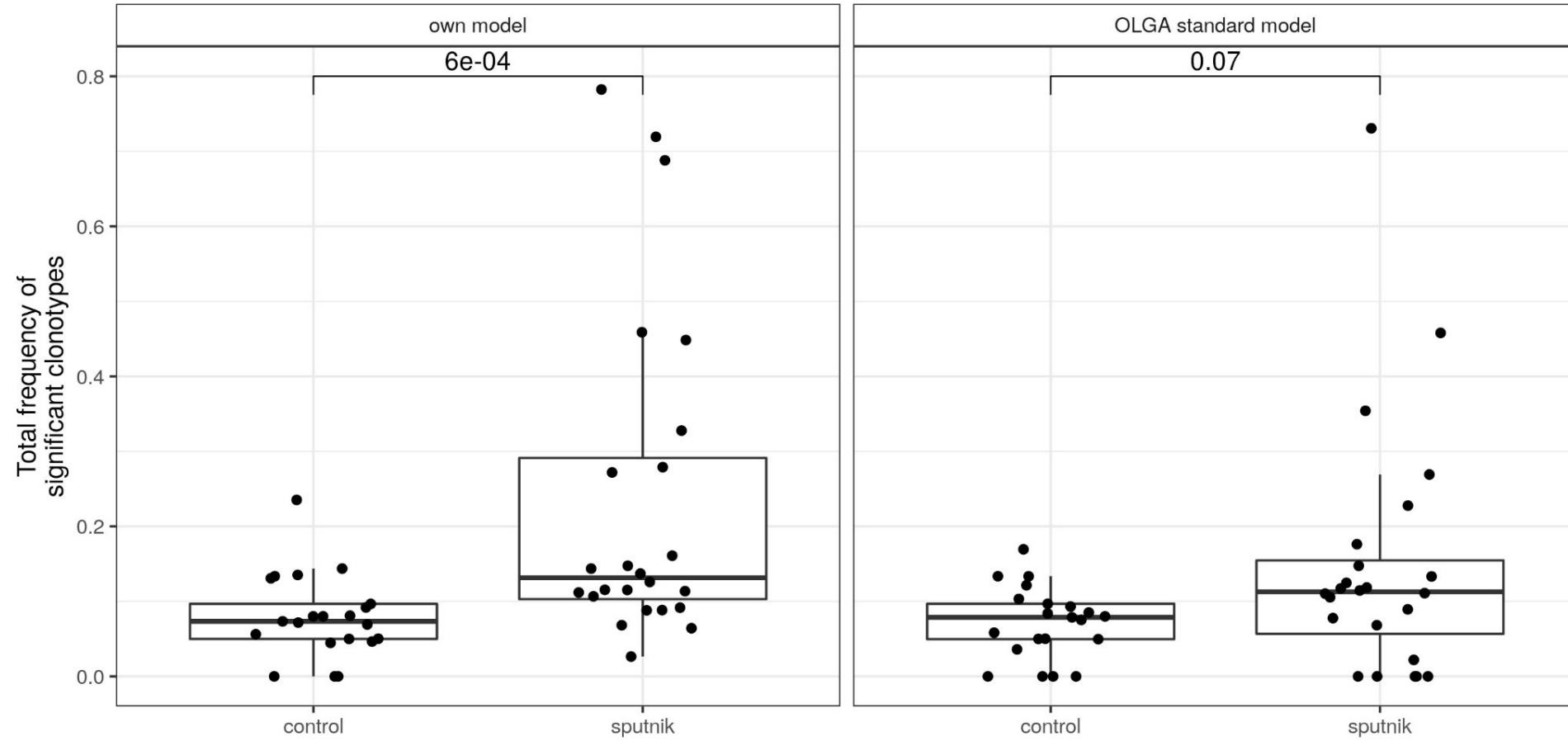
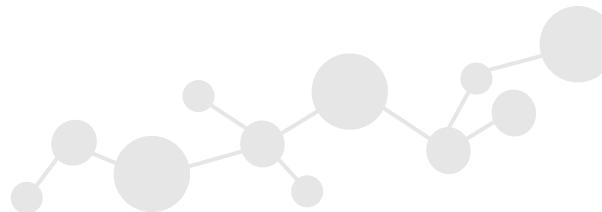
Anna V. Izosimova

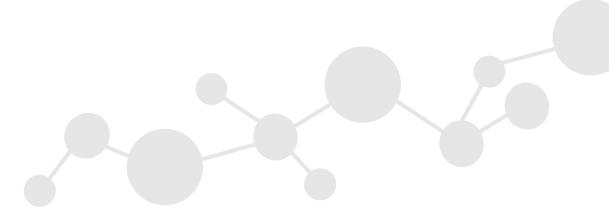
Pavel V. Shelyakin

Olga V. Britanova

Dmitry M. Chudakov

Outliers: possible explanation





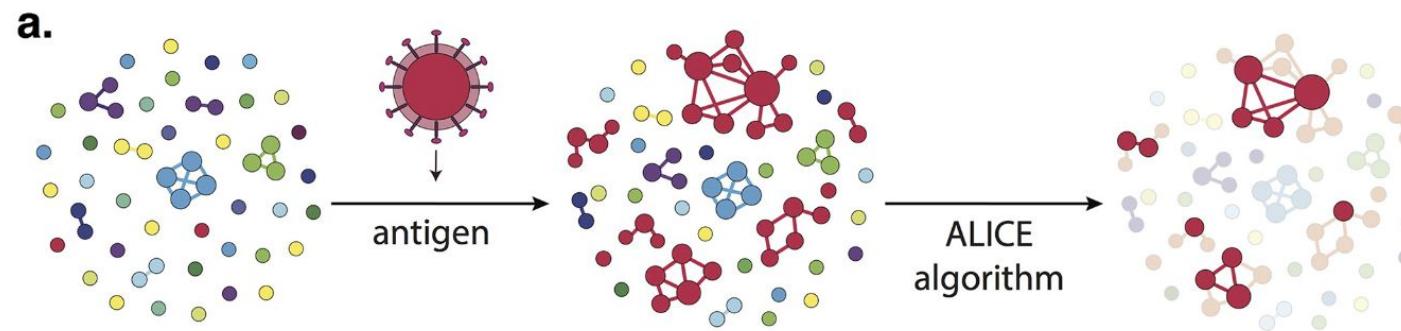
ALICE statistical model

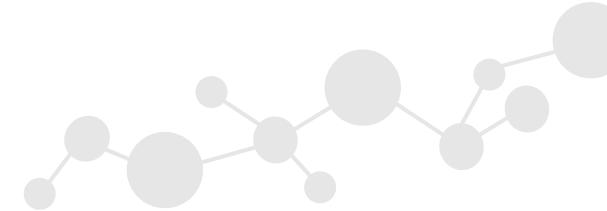
H_0 : The number of neighbors (d) of the sequence σ in the repertoire does not differ from random.

A neighbor is a sequence with the same VJ combination as σ , differing by one mismatch

$$P(d \mid \sigma) = e^{-\lambda} \frac{\lambda^d}{d!}$$

$$\lambda = n \sum_{\sigma'} Q P_{gen}(\sigma')$$





Abundance information

To include that information Pogorelyy suggested to replace number of neighbours (d) by the sum of transformed reads number over neighbours:

$$s = \sum_{i=1}^d f(c_i),$$

where c — number of reads account for the i th neighbour; $f(c)$ — transformation.

There are several variants of transformation: $f(c) = c$ corresponds to the sum of reads over all neighbours; $f(c) = \log(c)$ gives the sum of their logarithms. A number of reads per clonotype usually follows a power law. So, it can be useful to use a logarithmic scale.

The following definition was used

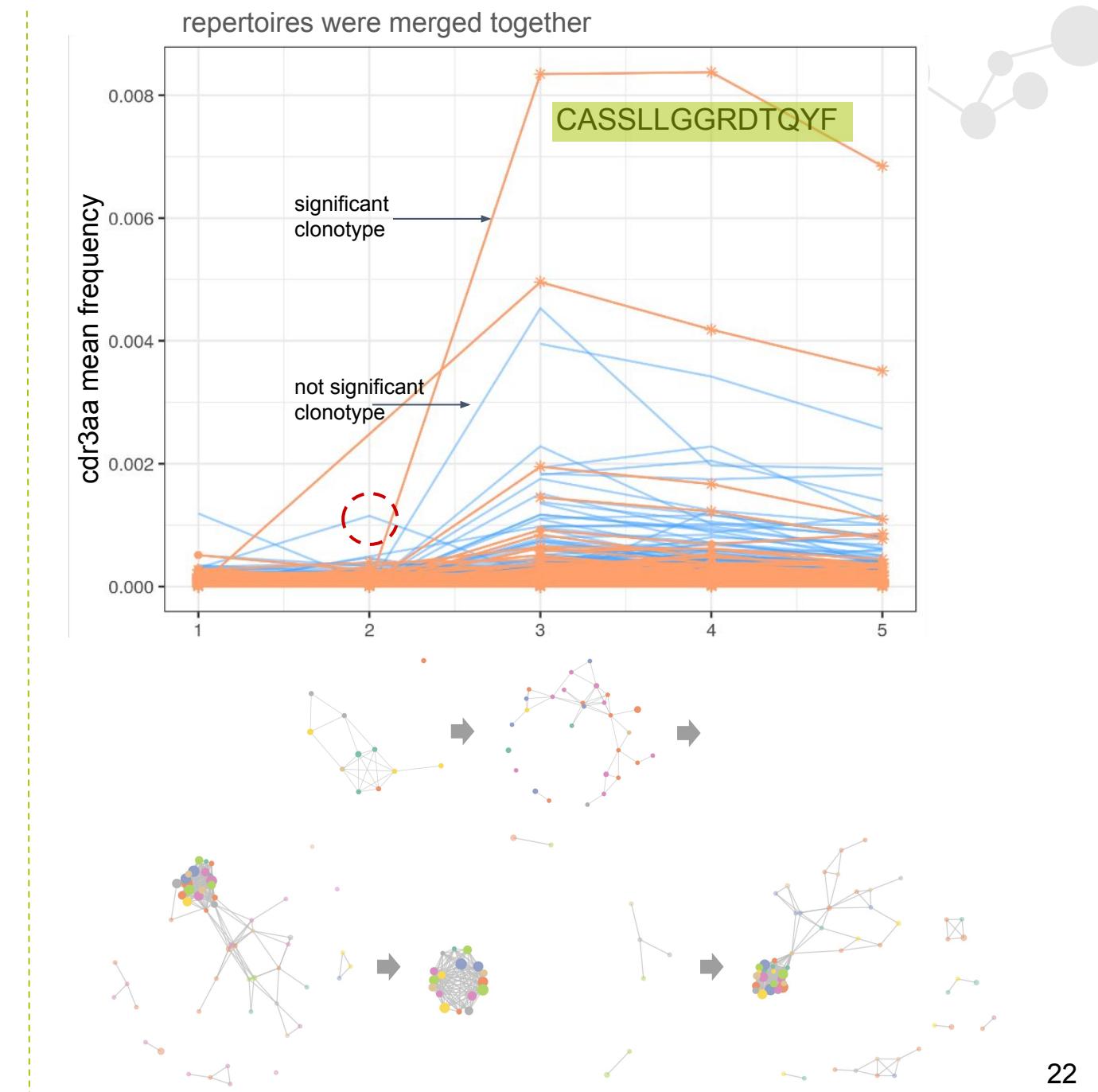
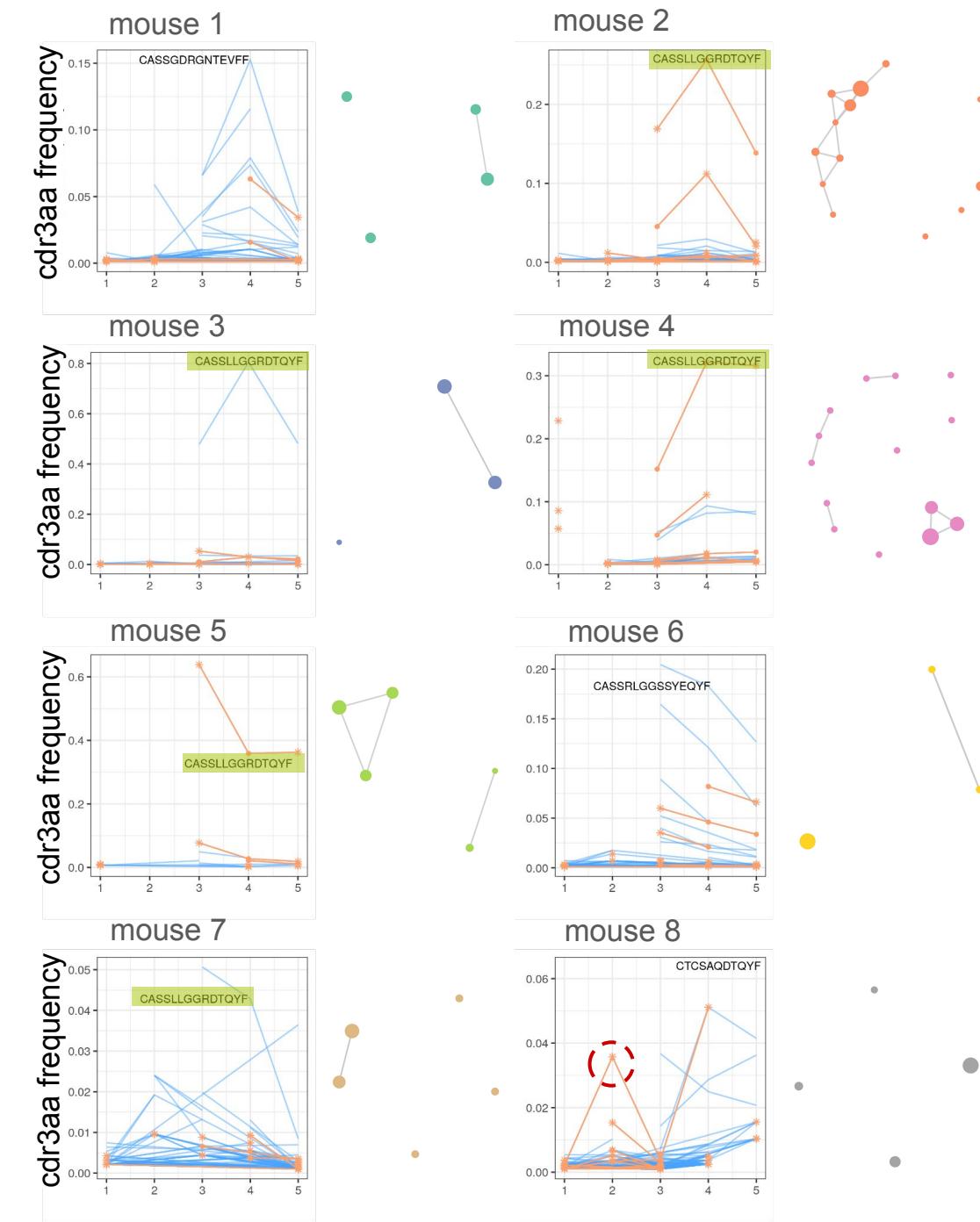
$$P(s|\sigma) = \sum_d P(s|d)P(d|\sigma),$$

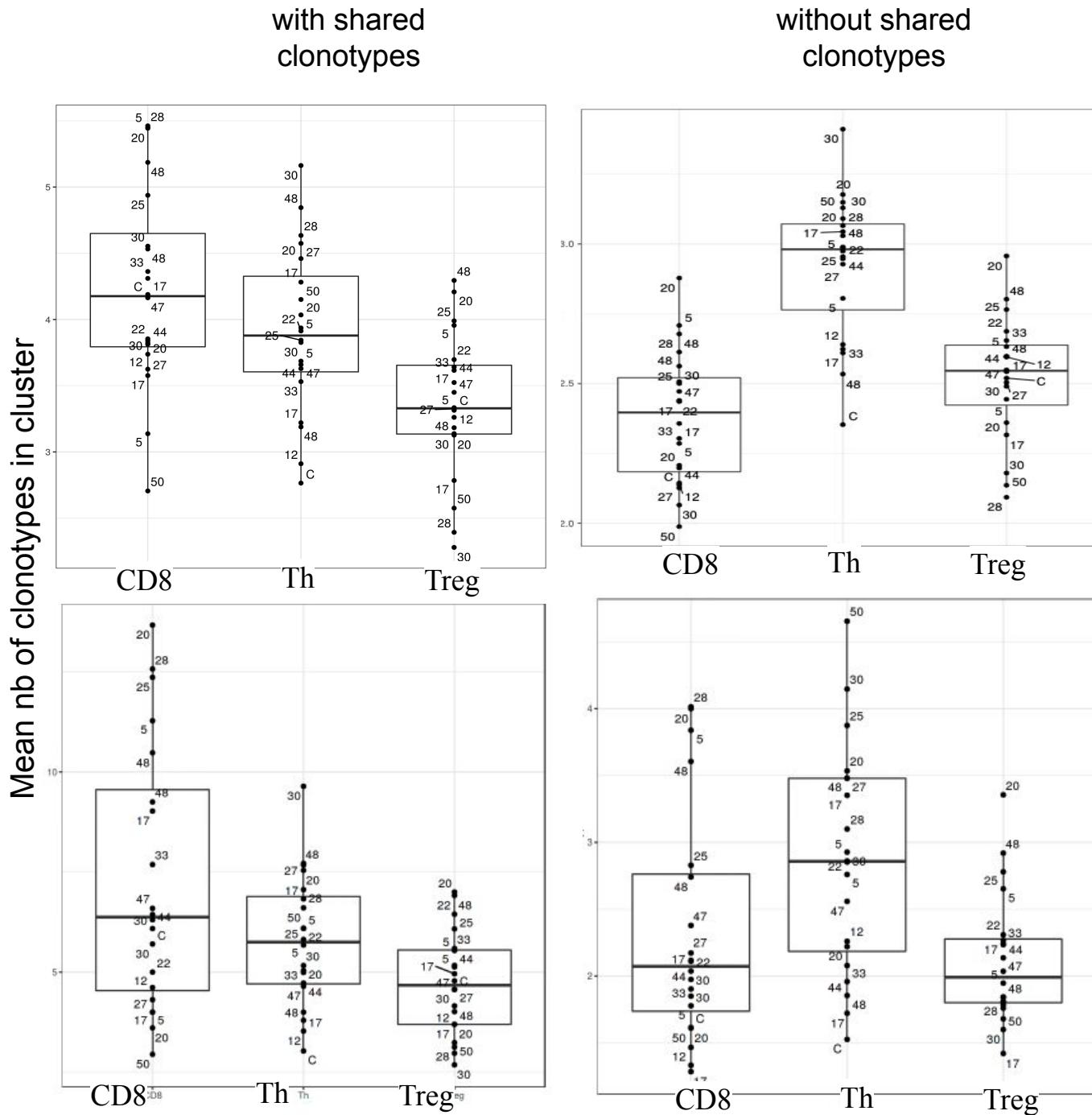
$P(s|d)$ is a probability density function (PDF) of random variables sum. As is known, the PDF of the sum of two independent random variables is the convolution of their two PDFs

$$f_{x+y}(z) = \int_{-\infty}^{+\infty} f_x(x) f_y(z - x) dx$$

$P(s|d)$ is a d -fold convolution of $P_f(f)$ PDF of reads in a given sample. For example

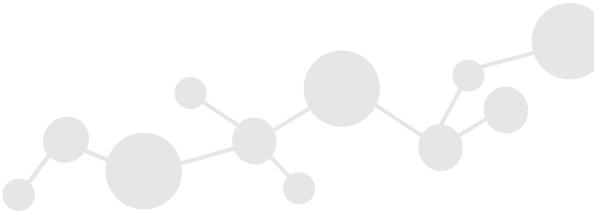
$$P(s|d = 1) = P_f(f) \text{ or } P(s|d = 2) = \sum_f P_f(f) P_f(2f - f) = \sum_f P_f(f) P_f(f).$$

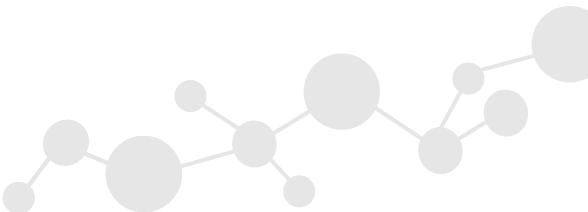
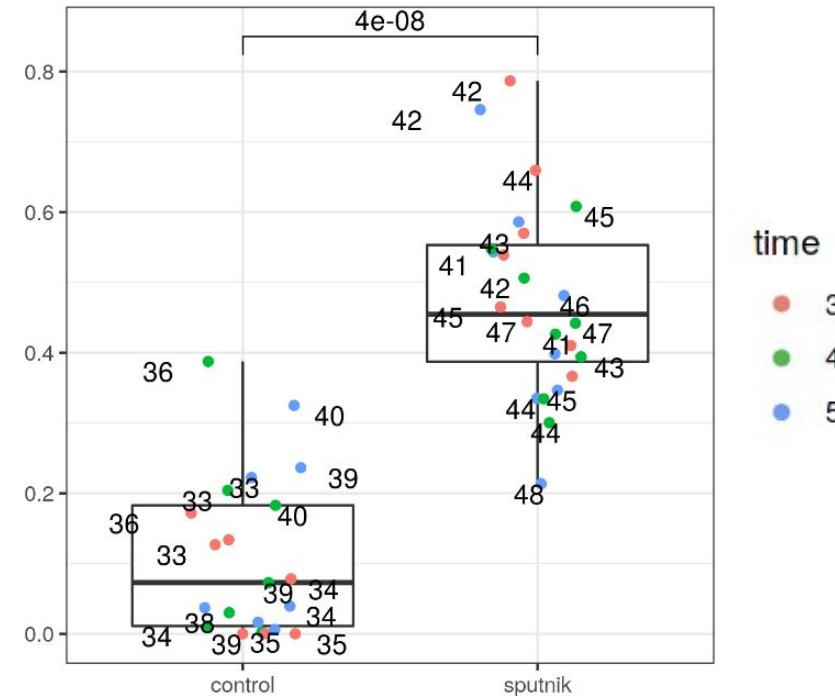
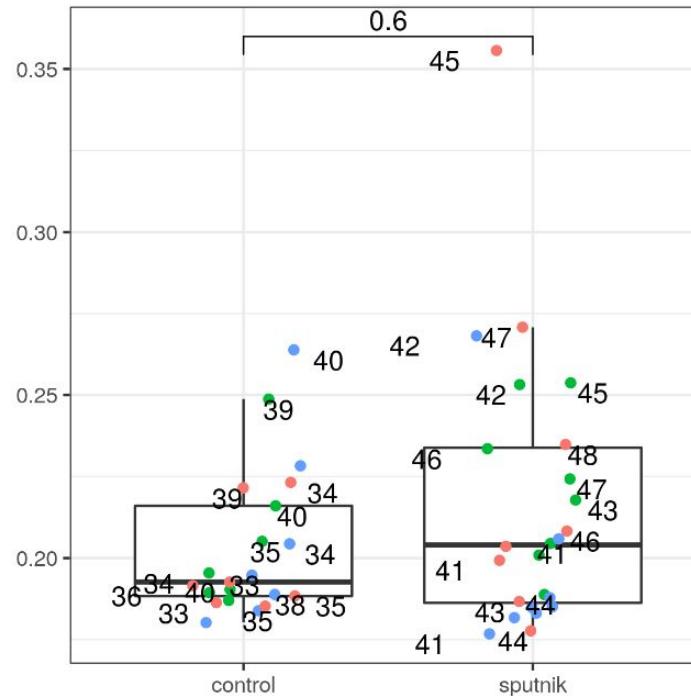
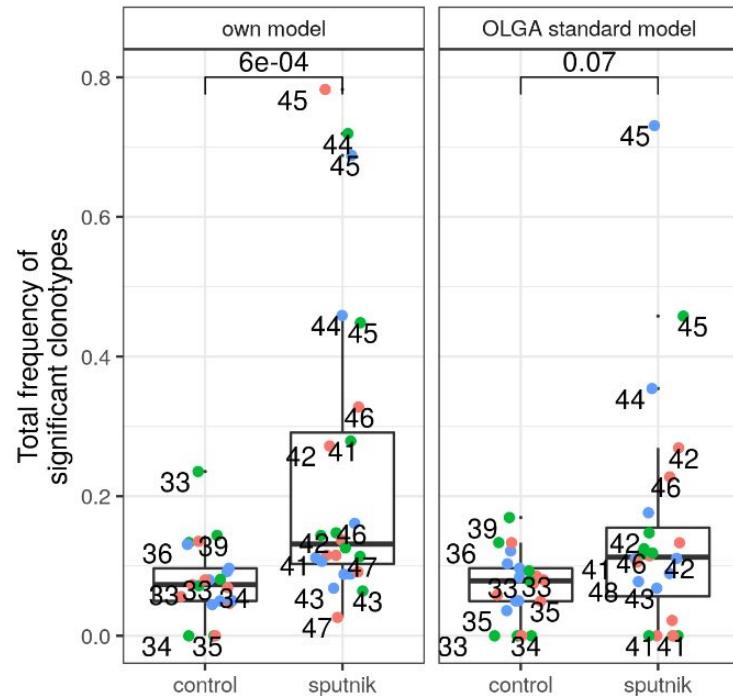


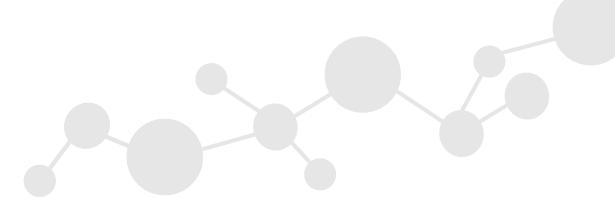
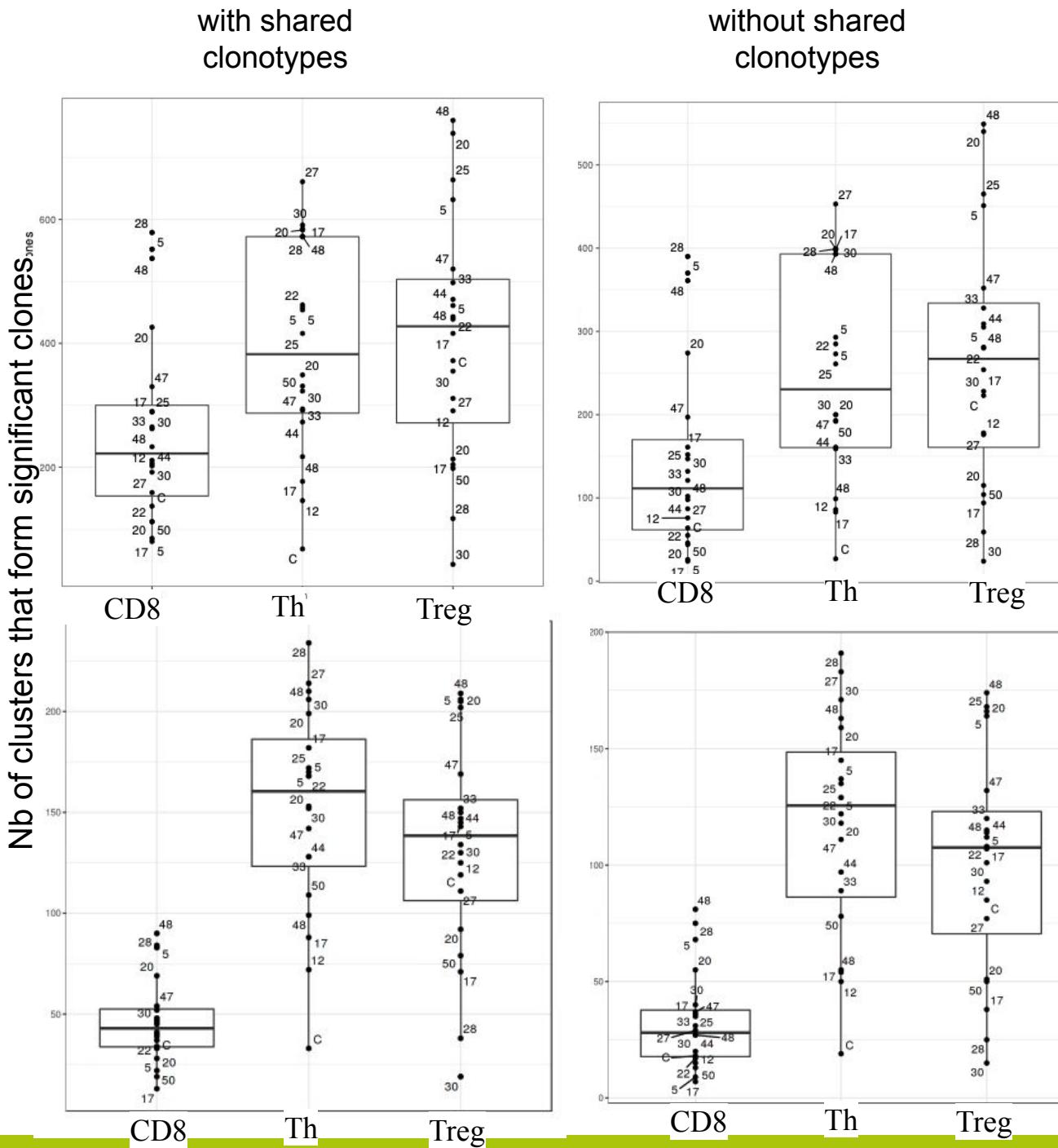


OLGA
standard
model

Own
model





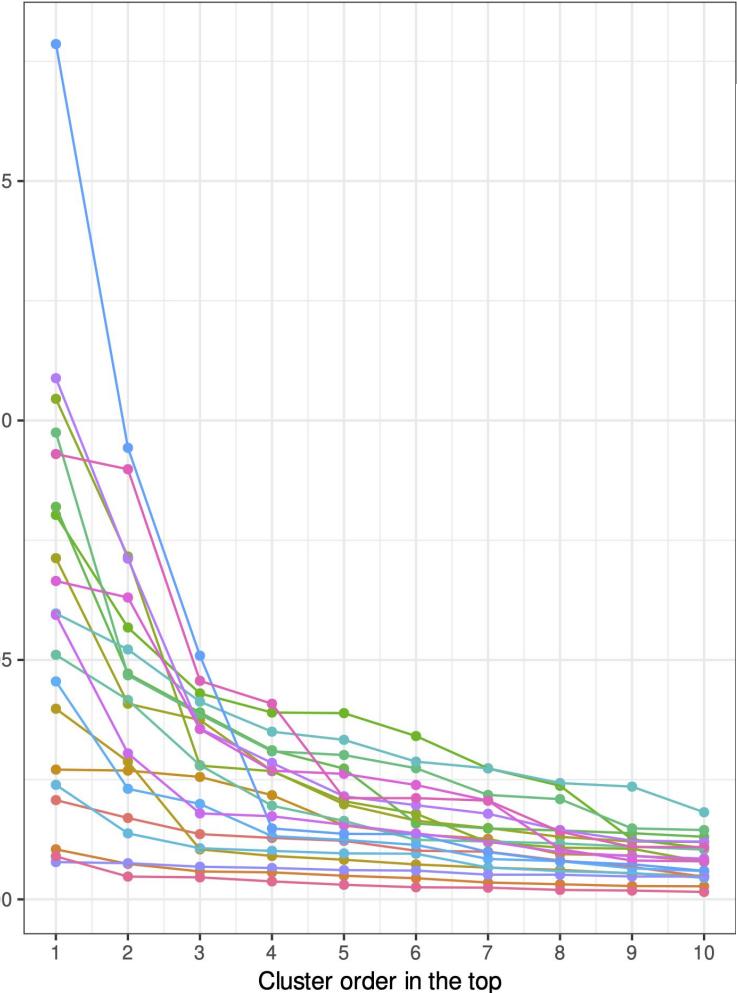


OLGA
standard
model

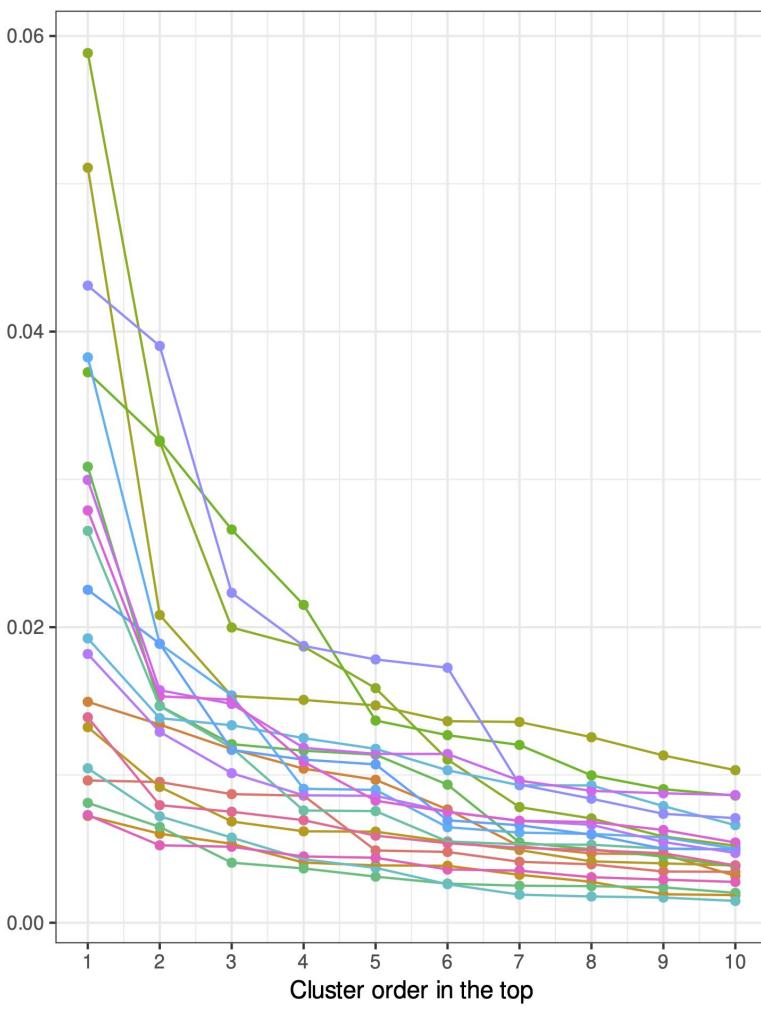
Own
model

Th

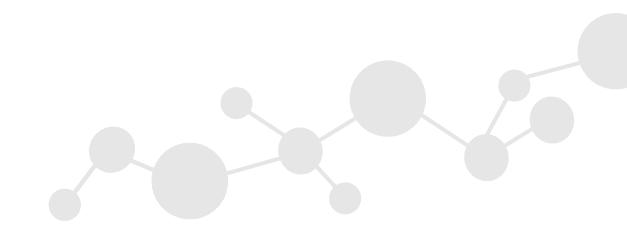
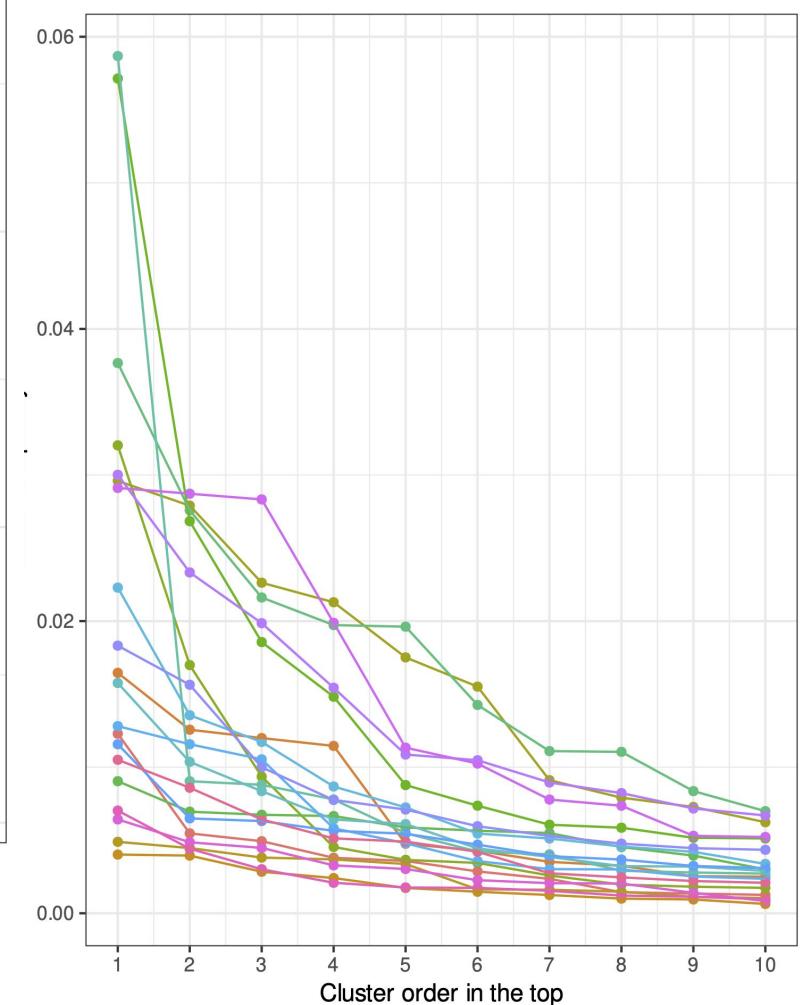
Cluster frequency



Treg



CD8



Peptide and group

- 12_0
- 17_1
- 17_2
- 20_1
- 20_2
- 22_0
- 25_0
- 27_0
- 28_0
- 30_1
- 30_2
- 33_0
- 44_0
- 47_0
- 48_1
- 48_2
- 5_1
- 5_2
- 50_0
- C_0



Methods

- Pre-processing was made with **migec**, **mixcr**, **vdjtools**
- For the task R library **TCRgrapher** was developed
<https://github.com/KseniaMIPT/tcrgrapher>
- For TCR generation probability calculation TCRgrapher uses **OLGA** and its statistical models
<https://github.com/statbiophys/OLGA>
- As an option, **SONIA** can be used for TCR generation probability and Q calculation
<https://github.com/statbiophys/SONIA>
- **R** was used for additional analysis and visualisation

