ImmunoMind.

Easy, fast, and practical AIRR analysis. Exploration of single-cell and bulk immune repertoire data in R using Immunarch with application to immunotherapy

Sponsored by the Adaptive Immune Receptor Repertoire Community of the Antibody Society

Biography

Lead Bioinformatician (T-cells, AI/ML)

Dr. Chudakov’s Laboratory
of Adaptive Immunity

- tcR – Immunarch’s predecessor
- Worked on AS, MRD
- T-cell biology
- Research group on Machine Learning in AIRR

Biography

Lead Bioinformatician (T-cells, AI/ML)

Dr. Chudakov’s Laboratory
of Adaptive Immunity

- tcR – Immunarch’s predecessor
- Worked on AS, MRD
- T-cell biology
- Research group on Machine Learning in AIRR

Director of AI

AI Infrastructure Startup

- Product management
- Team management

Biography

Lead Bioinformatician (T-cells, AI/ML)

Dr. Chudakov's Laboratory
of Adaptive Immunity

- tcR – Immunarch's predecessor
- Worked on AS, MRD
- T-cell biology
- Research group on Machine Learning in AIRR

Director of AI

AI Infrastructure Startup

- Product management
- Team management

Co-Founder & CEO

ImmunoMind

- Immunarch
    - >200 citations
    - >45,000 downloads
- UC Berkeley SkyDeck
- Multi-omics analytics platform for
  CAR-T cell/TCR therapies

Highlights – citations and happy users

Nature Medicine

Characteristics of anti-CD19 CAR T cell infusion products associated with efficacy and toxicity in patients with large B cell lymphomas. Deng et al.

Nature

B cells and tertiary lymphoid structures promote immunotherapy response. Helmink et al.

Journal of Clinical Investigation

T cell repertoire remodeling following post-transplant T cell therapy coincides with clinical response. Smith et al.

Science

Transient rest restores functionality in exhausted CAR-T cells through epigenetic remodeling. Weber et al.

Cancer Discovery

A Burned-Out CD8+ T-cell Subset Expands in the Tumor Microenvironment and Curbs Cancer Immunotherapy. Sanmamed et al.

Blood

Spatiotemporal Assessment of Immunogenomic Heterogeneity in Multiple Myeloma. Merz et al.

Testimonials

“The platform has been incredibly helpful with analyzing and interpreting our data. When I reached out with questions, Vadim was very helpful and responsive. I'm grateful to ImmunoMind for making this critical analysis program.”

— Senior Researcher, TOP-5 Pharma company

“Using the platform has been extremely helpful in producing beautiful publication ready plots for visualizing how clonotypes change across different timepoints. The platform takes hours off data analysis by providing a tool to easily analyze multiple data files and run biomarker discovery in a very smooth and quick way.”

— Molly B. El Alam, MPH. Research Assistant, Department of Radiation Oncology, The University of Texas M.D. Anderson

Philosophy

✓ **Principle I. Seamless work with data and methods**

Many data formats – many analysis methods – one library. Support all popular and clinically-important data formats and analysis methods

Philosophy

- ✓ **Principle I. Seamless work with data and methods**

  Many data formats – many analysis methods – one library. Support all popular and clinically-important data formats and analysis methods

- ✓ **Principle II. Minimize cognitive load to help focus on science**

  To help people focus on science, minimize the cognitive load required for coding, installation, file parsing and visualizations

Philosophy

✓	**Principle I. Seamless work with data and methods**	Many data formats – many analysis methods – one library. Support all popular and clinically-important data formats and analysis methods
✓	**Principle II. Minimize cognitive load to help focus on science**	To help people focus on science, minimize the cognitive load required for coding, installation, file parsing and visualizations
✓	**Principle III. Data type-agnostic**	Support the seamless work with data of any type – data frames, data tables, or databases.

Philosophy

✓	**Principle I. Seamless work with data and methods**	Many data formats – many analysis methods – one library. Support all popular and clinically-important data formats and analysis methods
✓	**Principle II. Minimize cognitive load to help focus on science**	To help people focus on science, minimize the cognitive load required for coding, installation, file parsing and visualizations
✓	**Principle III. Data type-agnostic**	Support the seamless work with data of any type – data frames, data tables, or databases.
✓	**Principle IV. Data nature-agnostic**	Support the data of any nature – bulk or single-cell.

How to install Immunarch

CRAN

Common installation method directly from CRAN. Installs the latest release. Consider using RStudio. In case of troubles check https://immunarch.com for a comprehensive list of solutions

```
install.packages("immunarch")
```

GitHub

Advanced installation method from source code. Install the latest release or pre-release, not published on CRAN yet

```
library(devtools)
install_github("immunomind/immunarch")
```

Docker

Use the virtual machine to work with Immunarch. Useful if you work on servers and if you don’t want to install >100 R packages on your machine

Link: https://hub.docker.com/r/immunomind/immunarch-docker

Input data formats

Immunarch **supports all popular** and critical TCR and BCR preprocessing and analysis **formats**

Immunarch **automatically detects input formats** without additional inputs from users, i.e., user just need to pass the path to the file or entire folder

Supported formats

- Any AIRR-formatted files
- TRUST4
- 10X Genomics
- MiXCR, all versions
- ImmunoSEQ
- IMGT
- MiGEC
- MiTCR
- VDJtools
- ArcherDX
- CATT and more to come

Input data formats

Automatic detection of file formats

```
> immdata = repLoad("../immunarch-test-files/data/test_files//")

== Step 1/3: loading repertoire files... ==

Processing "../immunarch-test-files/data/test_files//" ...
  -- [1/9] Parsing "../immunarch-test-files/data/test_files///10x_filtered_contig_annotations.csv" -- 10x (filt.contigs)
  [!] Removed 1087 clonotypes with no nucleotide and amino acid CDR3 sequence.

  -- [2/9] Parsing "../immunarch-test-files/data/test_files///airr.tsv" -- airr
  [!] Removed 2013 clonotypes with no nucleotide and amino acid CDR3 sequence.

  -- [3/9] Parsing "../immunarch-test-files/data/test_files///archer.tsv" -- archer
 0s  -- [4/9] Parsing "../immunarch-test-files/data/test_files///immunoseq_1.txt" -- immunoseq
  -- [5/9] Parsing "../immunarch-test-files/data/test_files///migec_orig.txt" -- migec

  -- [6/9] Parsing "../immunarch-test-files/data/test_files///migmap.txt" -- migmap

  [!] Removed 36 clonotypes with no nucleotide and amino acid CDR3 sequence.

  -- [7/9] Parsing "../immunarch-test-files/data/test_files///mixcr.1.txt" -- mixcr
  -- [8/9] Parsing "../immunarch-test-files/data/test_files///rtcr.txt" -- unsupported format, skipping

  -- [9/9] Parsing "../immunarch-test-files/data/test_files///vidjil.txt" -- unsupported format, skipping
```

```
repLoad("path/to/your/folder")
```

Immunarch data format – structure

```
immdata = repLoad("path/to/your/folder")
```

immdata structure:

- immdata$data – list of immune repertoire tables
  One row – one clonotype
- immdata$meta – table with metadata

Immunarch data format – tables

Clones	Proportion	CDR3.nt	CDR3.aa	V.name	D.name
36	0.0026722090	TGTGCGAGAGACTCCTTTTATGGGGGAGTAAGTCAGTT...	CARDSFYGGVSQFDPW	IGHV7-4-1*02	IGHD3-16*01,IGHD3-16*02
33	0.0024495249	TGTGCGAAAGAGGGACTGTGGTACGGGGGGAACTGGT...	CAKEGLWYGGNWFDPW	IGHV3-23*04	IGHD3-10*01
31	0.0023010689	TGTGCGAAAGAGGGACTGTGGTACGGGGGGAACTGGT...	CAKEGLWYGGNWFDPW	IGHV3-23*04	IGHD3-10*01
31	0.0023010689	TGTGCGAATCAATGGGTGGCTCGGGAAATTGGCCCACG...	CANQWVAREIGPRRGGYW	IGHV3-23*01,IGHV3-23D*01	IGHD3-10*01
29	0.0021526128	TGTGCGAGAAGCACTTCAAGGCGGAATACTATAATTCG...	CARSTSRRNTIIRGRVWLDPW	IGHV7-4-1*02	IGHD3-10*01
27	0.0020041568	TGTGCGAGAGACTCCTTTTATGGGGGAGTAAGTCAGTT...	CARDSFYGGVSQFDPW	IGHV7-4-1*02	IGHD3-16*01,IGHD3-16*02
27	0.0020041568	TGTGCAAAAGAAAAACGGGTGAACAGTTATGGTTATTT...	CAKEKRVNSYGYFYFDYW	IGHV3-9*01	IGHD5-18*01,IGHD5-5*01
25	0.0018557007	TGTGCAAAAGAAAAACGGGTGAACAGTTATGGTTATTT...	CAKEKRVNSYGYFYFDYW	IGHV3-9*01	IGHD5-18*01,IGHD5-5*01

```
immdata = repLoad("path/to/your/folder")
```

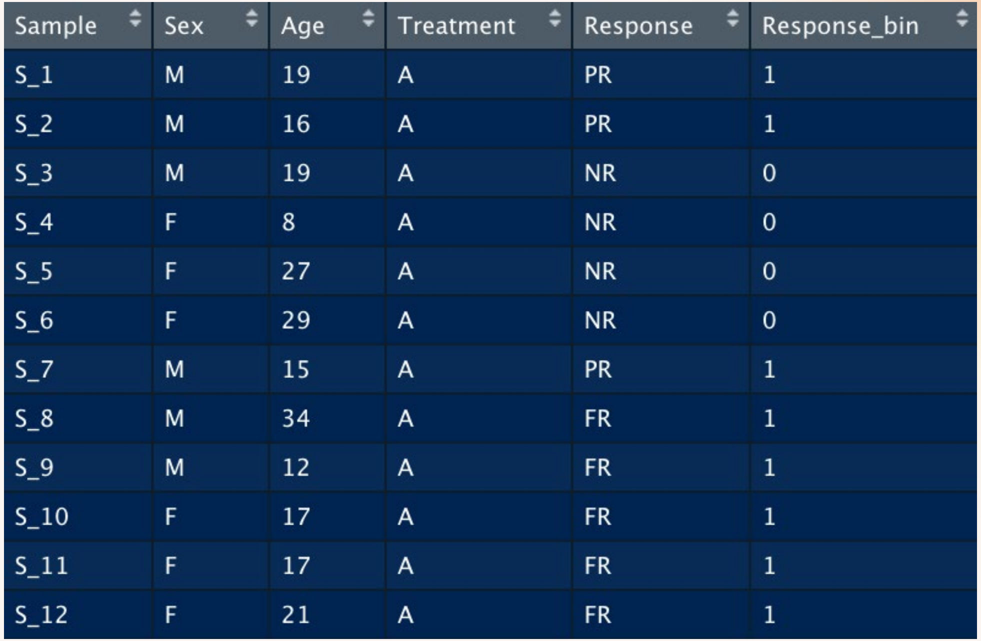
immdata structure:

- immdata$data – list of immune repertoire tables
  One row – one clonotype
- immdata$meta – table with metadata

File format columns

- Clones, Proportion
- CDR3.nt, CDR3.aa
- V.name, D.name, J.name
- Additional columns: C.name, full sequence, CDR1-2 and FRs

Immunarch data format – metadata



Sample	Sex	Age	Treatment	Response	Response_bin
S_1	M	19	A	PR	1
S_2	M	16	A	PR	1
S_3	M	19	A	NR	0
S_4	F	8	A	NR	0
S_5	F	27	A	NR	0
S_6	F	29	A	NR	0
S_7	M	15	A	PR	1
S_8	M	34	A	FR	1
S_9	M	12	A	FR	1
S_10	F	17	A	FR	1
S_11	F	17	A	FR	1
S_12	F	21	A	FR	1

```
immdata$meta
```

Exploratory data analysis (EDA)

Goal: give a quick overview of the data for quality control and sanity check purposes

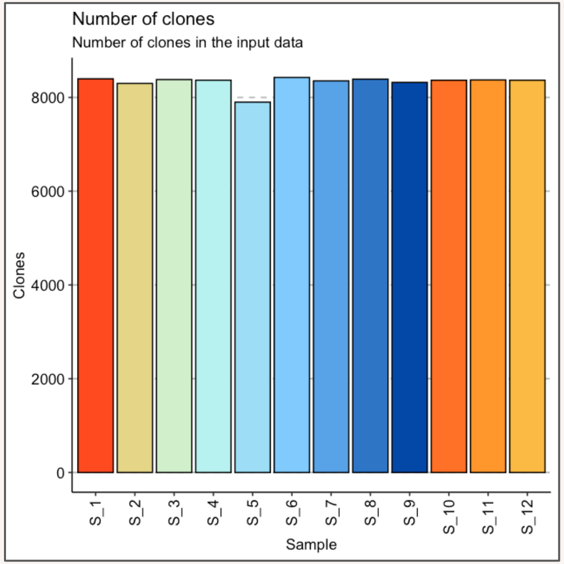
Methods: number of **clones** or clonotypes, distribution of CDR3 lengths, distribution of abundances

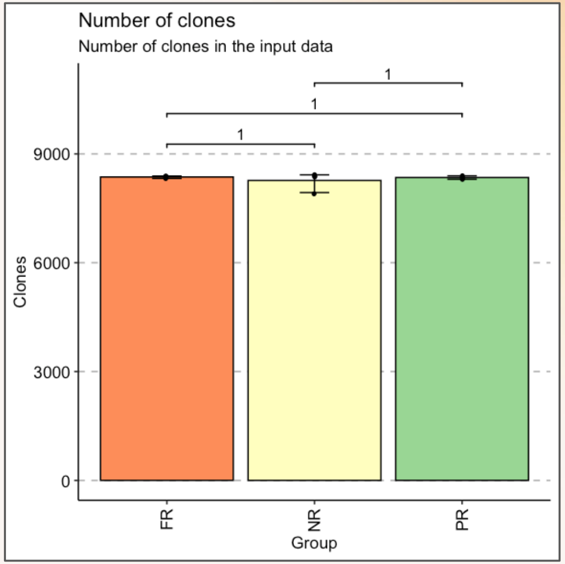
Number of clones

Estimate the number of clones in each sample to find low-quality samples and create a set-up for downsampling

Exploratory data analysis (EDA)

Number of clones





```
res = repExplore(immdata$data, "clones")
vis(res) # OR
repExplore(immdata$data, "clones") %>% vis()
```

```
repExplore(immdata$data, "clones") %>%
  vis(.by = "Response", .meta =
immdata$meta)
```

Exploratory data analysis (EDA)

Goal: give a quick overview of the data
for quality control and sanity check purposes

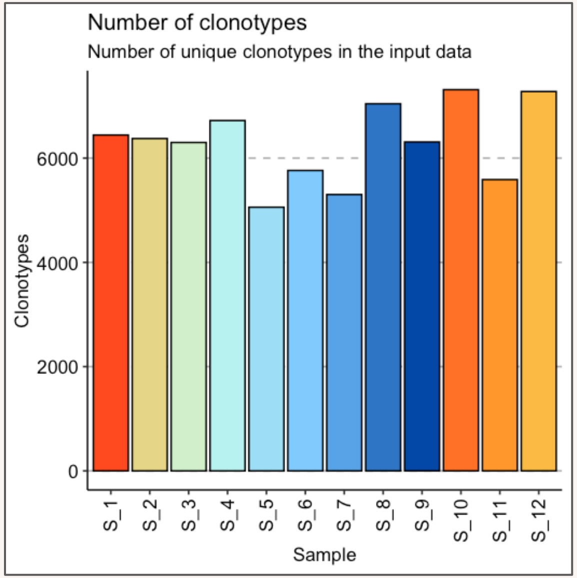
Methods: number of clones or **clonotypes**,
distribution of CDR3 lengths, distribution of abundances

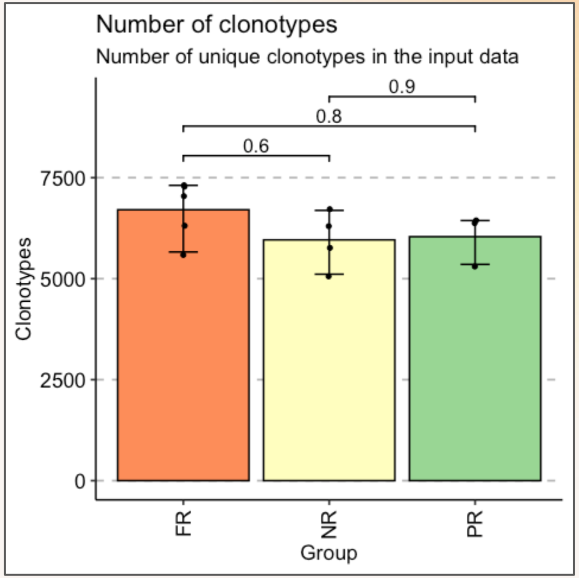
Number of clonotypes

Estimate the number of unique clonotypes in each sample to find low-quality samples and create a set-up for downsampling

Exploratory data analysis (EDA)

Number of clonotypes





```
repExplore(immdata$data, "volume") %>%
  vis()
```

```
repExplore(immdata$data, "volume") %>%
  vis(.by = "Response", .meta =
immdata$meta)
```

Exploratory data analysis (EDA)

Goal: give a quick overview of the data
for quality control and sanity check purposes

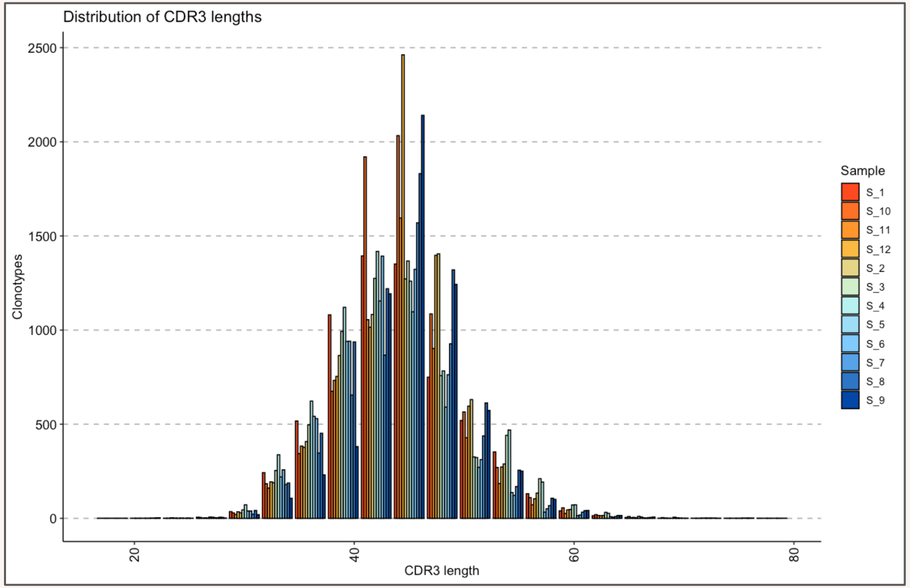
Methods: number of clones or clonotypes,
distribution of CDR3 lengths, distribution of abundances

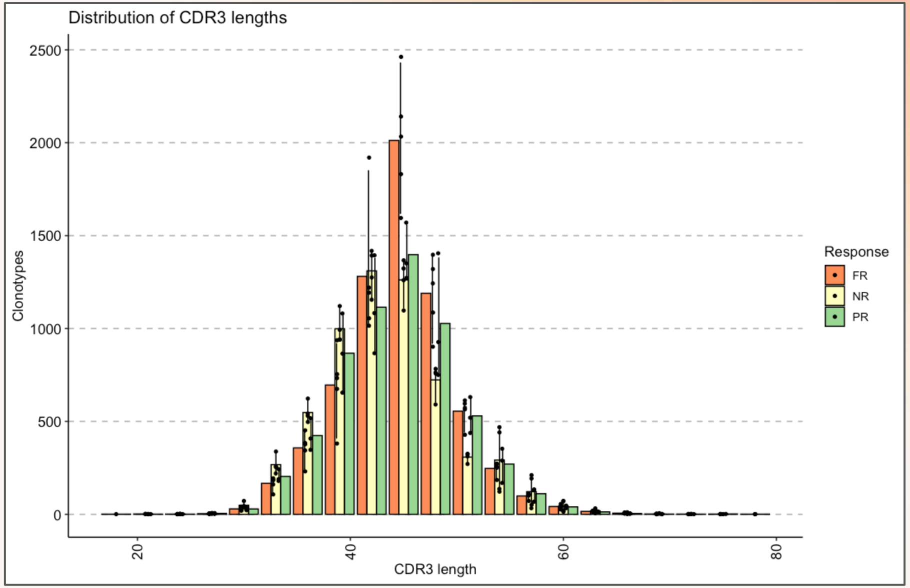
Distribution of CDR3 lengths

Estimate the distribution of lengths of nucleotide or amino acid sequences of CDR3 to find anomalies and differences between groups

Exploratory data analysis (EDA)

Distribution of CDR3 lengths





```
repExplore(immdata$data, "len") %>% vis()
```

```
repExplore(immdata$data, "len") %>%
   vis(.by = "Response", .meta =
immdata$meta)
```

Exploratory data analysis (EDA)

Goal: give a quick overview of the data for quality control and sanity check purposes

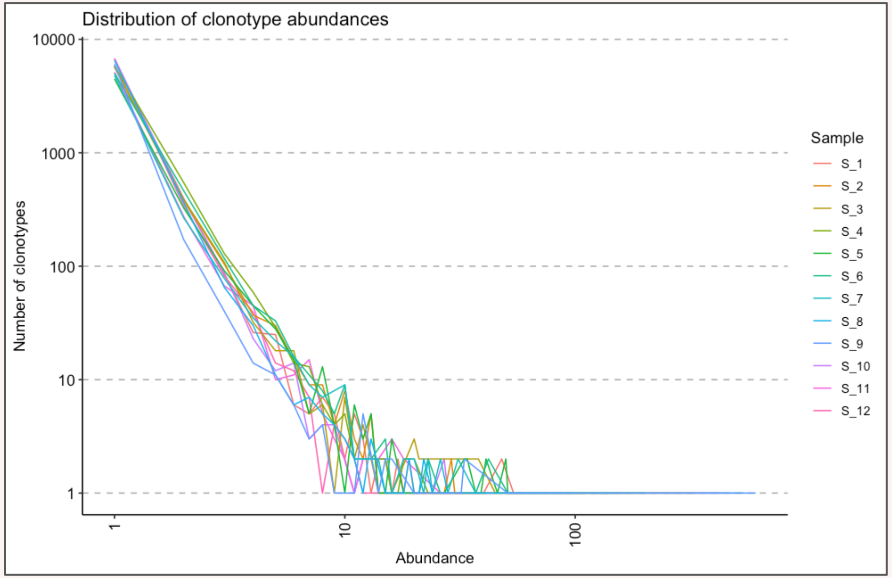
Methods: number of clones or clonotypes, distribution of CDR3 lengths, **distribution of abundances**

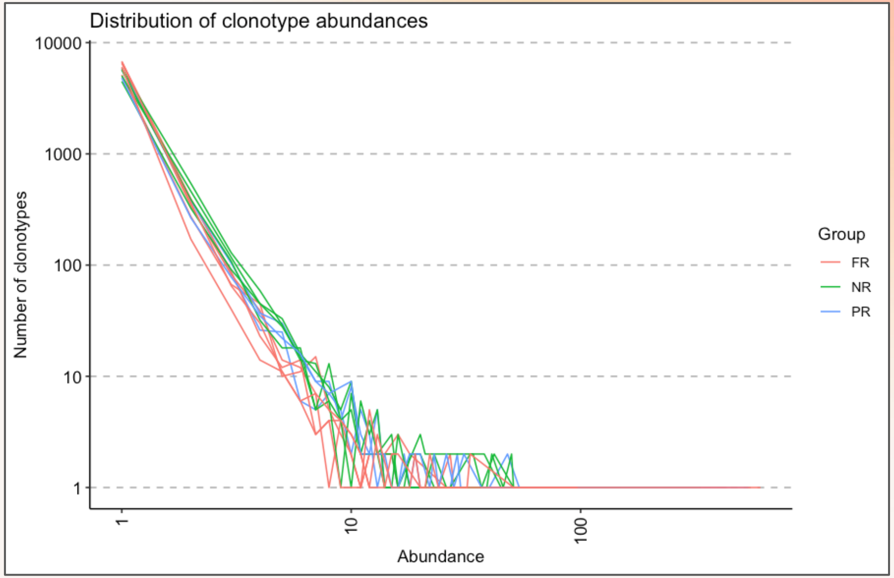
Distribution of clonotype abundances

Estimate the abundances of clonotypes, i.e., how frequent clonotypes with specific number of clones

Exploratory data analysis (EDA)

Distribution of clonotype abundances





```
repExplore(immdata$data, "count") %>% vis()
```

```
repExplore(immdata$data, "count") %>%
   vis(.by = "Response", .meta =
immdata$meta)
```

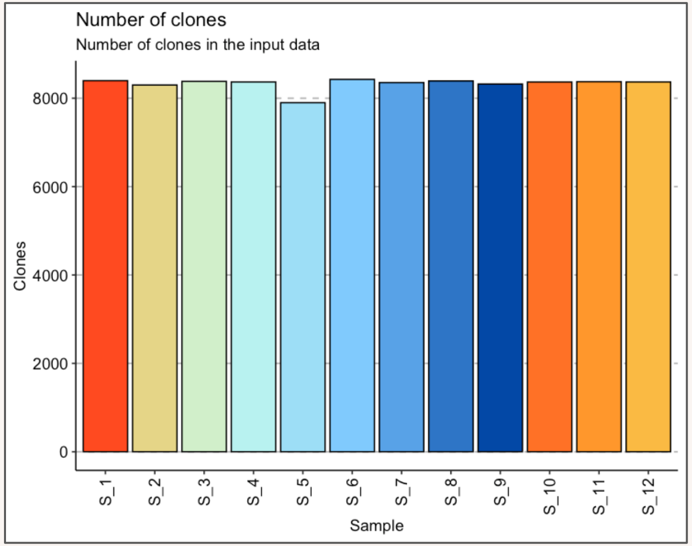
Downsampling

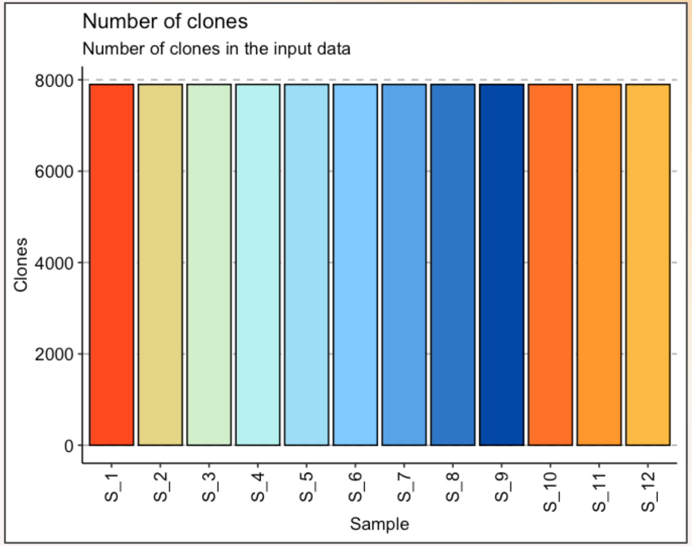
Goal: make data samples comparable

Details: preprocess the data to make samples equal or at least similar in size, and accelerate the subsequent data analysis

Downsampling

No downsampling vs. with downsampling





```
repExplore(immdata$data, "clones") %>%
  vis()
```

```
repSample(immdata$data, "downsample") %>%
  repExplore("clones") %>% vis()
```

Public clonotype analysis ("overlap")

Goal: estimate the similarity of samples using the number of shared or "public" clonotypes

Applications: tumor-specific clonotype discovery, response prediction

Highlighted publications

1. Preprocessing purposes: find groups of clonotypes shared between samples or tumors and samples, and annotate or analyse them further.
2. TCR Repertoire Analysis Reveals Mobilization of Novel CD8+ T Cell Clones Into the Cancer-Immunity Cycle Following Anti-CD4 Antibody Administration Aoki et al.
3. Greater extent of blood-tumor TCR repertoire overlap is associated with favorable clinical responses to PD-1 blockade Aoki et al.
4. Evaluating T-cell cross-reactivity between tumors and immune-related adverse events with TCR sequencing: pitfalls in interpretations of functional relevance Cottrell et al.

Public clonotype analysis ("overlap")

Goal: estimate the similarity of samples using the number of shared or "public" clonotypes

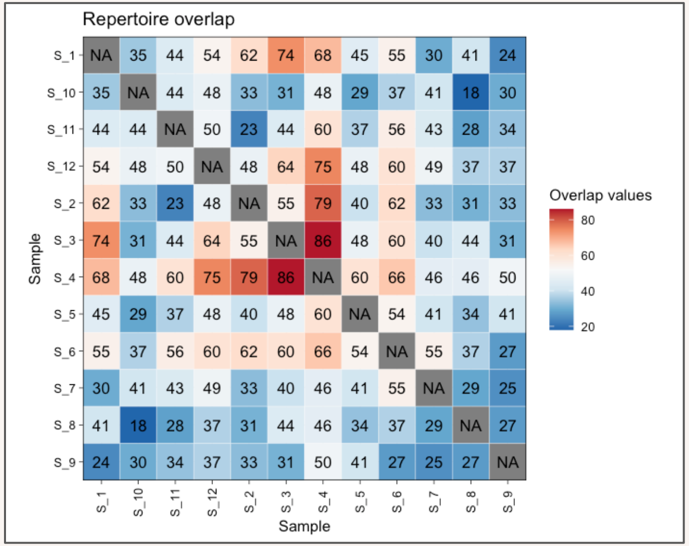
Methods: shared clonotypes, Jaccard index, Morisita-Horn index

Shared clonotypes

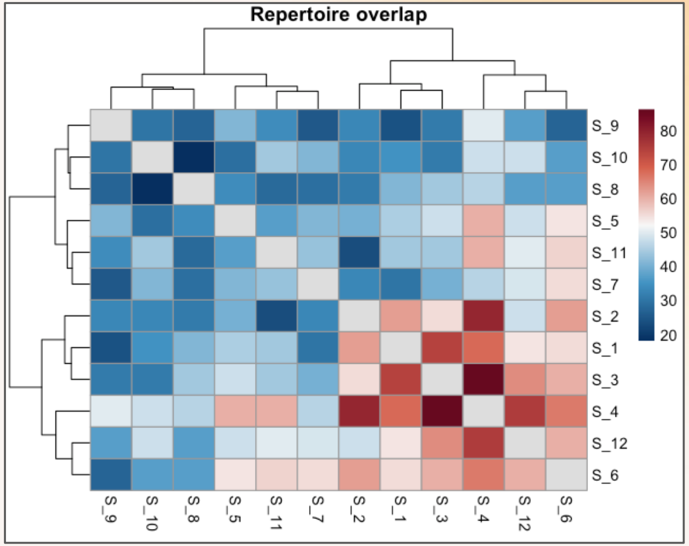
Estimate the number of shared clonotypes between samples. Works great if you downsampled data or would like to quickly understand the landscape. Use CDR3, CDR3+V, CDR3+V+J. Find clonotypes of interest

Exploratory data analysis (EDA)

Shared clonotypes



Sample	S_1	S_10	S_11	S_12	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9
S_1	NA	35	44	54	62	74	68	45	55	30	41	24
S_10	35	NA	44	48	33	31	48	29	37	41	18	30
S_11	44	44	NA	50	23	44	60	37	56	43	28	34
S_12	54	48	50	NA	48	64	75	48	60	49	37	37
S_2	62	33	23	48	NA	55	79	40	62	33	31	33
S_3	74	31	44	64	55	NA	86	48	60	40	44	31
S_4	68	48	60	75	79	86	NA	60	66	46	46	50
S_5	45	29	37	48	40	48	60	NA	54	41	34	41
S_6	55	37	56	60	62	60	66	54	NA	55	37	27
S_7	30	41	43	49	33	40	46	41	55	NA	29	25
S_8	41	18	28	37	31	44	46	34	37	29	NA	27
S_9	24	30	34	37	33	31	50	41	27	25	27	NA



```
repOverlap(immdata$data) %>% vis()
```

```
repOverlap(immdata$data) %>%
   vis(.plot = "heatmap2")
```

Public clonotype analysis ("overlap")

Goal: estimate the similarity of samples using the number of shared or "public" clonotypes

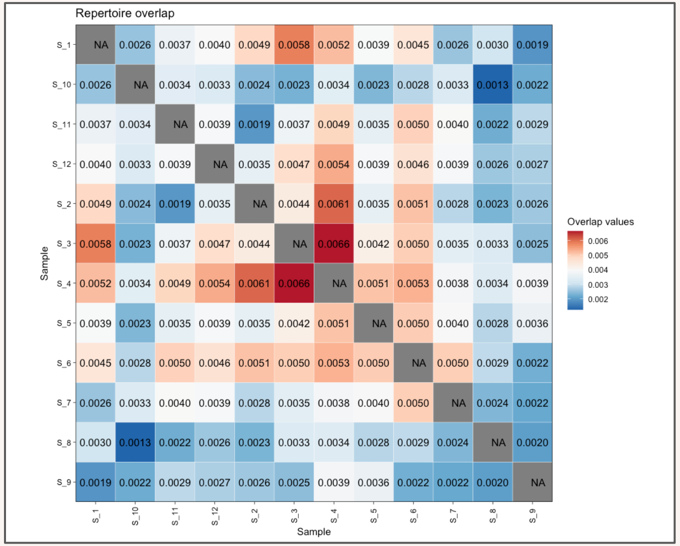
Methods: shared clonotypes, **Jaccard index**, Morisita-Horn index

Jaccard index

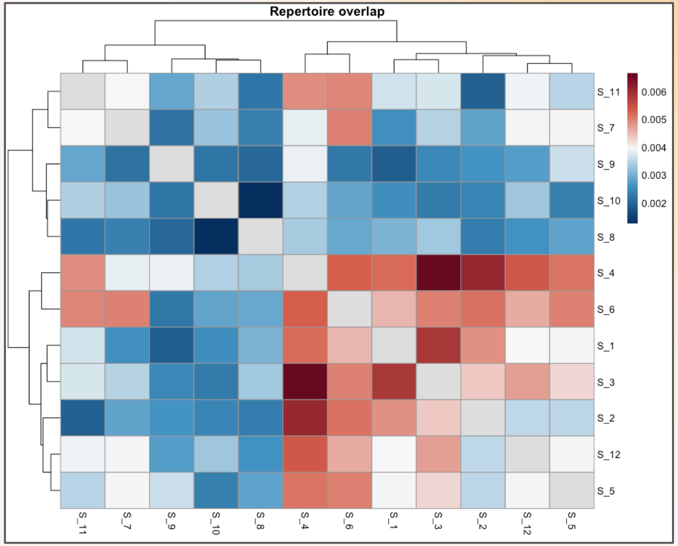
Estimate the similarity of samples using the number of shared clonotypes. Doesn’t take into account the abundance

Public clonotype analysis ("overlap")

Jaccard index



Sample	S_1	S_10	S_11	S_12	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9
S_1	NA	0.0026	0.0037	0.0040	0.0049	0.0058	0.0052	0.0039	0.0045	0.0026	0.0030	0.0019
S_10	0.0026	NA	0.0034	0.0033	0.0024	0.0023	0.0034	0.0023	0.0028	0.0033	0.0013	0.0022
S_11	0.0037	0.0034	NA	0.0039	0.0019	0.0037	0.0049	0.0035	0.0050	0.0040	0.0022	0.0029
S_12	0.0040	0.0033	0.0039	NA	0.0035	0.0047	0.0054	0.0039	0.0046	0.0039	0.0026	0.0027
S_2	0.0049	0.0024	0.0019	0.0035	NA	0.0044	0.0061	0.0035	0.0051	0.0028	0.0023	0.0026
S_3	0.0058	0.0023	0.0037	0.0047	0.0044	NA	0.0066	0.0042	0.0050	0.0035	0.0033	0.0025
S_4	0.0052	0.0034	0.0049	0.0054	0.0061	0.0066	NA	0.0051	0.0053	0.0038	0.0034	0.0039
S_5	0.0039	0.0023	0.0035	0.0039	0.0035	0.0042	0.0051	NA	0.0050	0.0040	0.0028	0.0036
S_6	0.0045	0.0028	0.0050	0.0046	0.0051	0.0050	0.0053	0.0050	NA	0.0050	0.0029	0.0022
S_7	0.0026	0.0033	0.0040	0.0039	0.0028	0.0035	0.0038	0.0040	0.0050	NA	0.0024	0.0022
S_8	0.0030	0.0013	0.0022	0.0026	0.0023	0.0033	0.0034	0.0028	0.0029	0.0024	NA	0.0020
S_9	0.0019	0.0022	0.0029	0.0027	0.0026	0.0025	0.0039	0.0036	0.0022	0.0022	0.0020	NA



```
repOverlap(immdata$data, "jaccard") %>%
  vis()
```

```
repOverlap(immdata$data, "jaccard") %>%
  vis(.plot = "heatmap2")
```

Public clonotype analysis ("overlap")

Goal: estimate the similarity of samples using the number of shared or "public" clonotypes

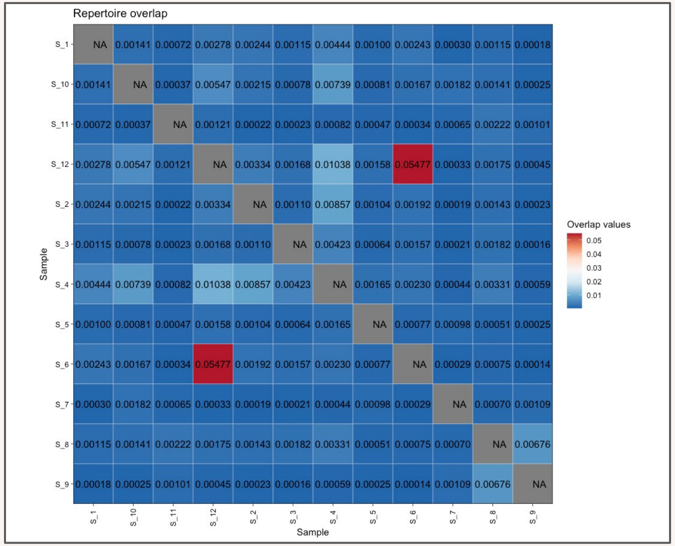
Methods: shared clonotypes, Jaccard index, **Morisita-Horn index**

Morisita-Horn index

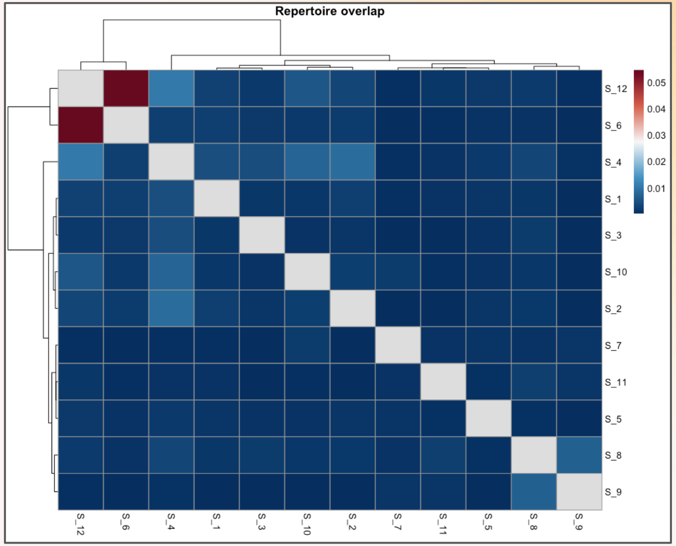
Estimate the similarity of samples using the number of shared clonotypes and their abundances. I.e., more abundant clonotypes affect the similarity more

Public clonotype analysis ("overlap")

Morisita-Horn index



	S_1	S_10	S_11	S_12	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9
S_1	NA	0.00141	0.00072	0.00278	0.00244	0.00115	0.00444	0.00100	0.00243	0.00030	0.00115	0.00018
S_10	0.00141	NA	0.00037	0.00547	0.00215	0.00078	0.00739	0.00081	0.00167	0.00182	0.00141	0.00025
S_11	0.00072	0.00037	NA	0.00121	0.00022	0.00023	0.00082	0.00047	0.00034	0.00065	0.00222	0.00101
S_12	0.00278	0.00547	0.00121	NA	0.00334	0.00168	0.01038	0.00158	0.05477	0.00033	0.00175	0.00045
S_2	0.00244	0.00215	0.00022	0.00334	NA	0.00110	0.00857	0.00104	0.00192	0.00019	0.00143	0.00023
S_3	0.00115	0.00078	0.00023	0.00168	0.00110	NA	0.00423	0.00064	0.00157	0.00021	0.00182	0.00016
S_4	0.00444	0.00739	0.00082	0.01038	0.00857	0.00423	NA	0.00165	0.00230	0.00044	0.00331	0.00059
S_5	0.00100	0.00081	0.00047	0.00158	0.00104	0.00064	0.00165	NA	0.00077	0.00098	0.00051	0.00025
S_6	0.00243	0.00167	0.00034	0.05477	0.00192	0.00157	0.00230	0.00077	NA	0.00029	0.00075	0.00014
S_7	0.00030	0.00182	0.00065	0.00033	0.00019	0.00021	0.00044	0.00098	0.00029	NA	0.00070	0.00109
S_8	0.00115	0.00141	0.00222	0.00175	0.00143	0.00182	0.00331	0.00051	0.00075	0.00070	NA	0.00676
S_9	0.00018	0.00025	0.00101	0.00045	0.00023	0.00016	0.00059	0.00025	0.00014	0.00109	0.00676	NA



```
repOverlap(immdata$data, "morisita") %>%
  vis()
```

```
repOverlap(immdata$data, "morisita") %>%
  vis(.plot = "heatmap2")
```

Clonality analysis

Goal: estimate and compare the differences in abundances of clonotypes between the samples. Very similar to diversity analysis

Applications: therapy response prediction

Highlighted publications

1. TCR repertoire characteristics predict clinical response to adoptive CTL therapy against nasopharyngeal carcinoma Wang et al.
2. Combined TCR Repertoire Profiles and Blood Cell Phenotypes Predict Melanoma Patient Response to Personalized Neoantigen Therapy plus Anti-PD-1 Poran et al.
3. T cell receptor repertoire features associated with survival in immunotherapy-treated pancreatic ductal adenocarcinoma Hopkins et al.
4. Polyfunctional tumor-reactive T cells are effectively expanded from non-small cell lung cancers, and correlate with an immune-engaged T cell profile De Groot et al.

Clonality analysis

Goal: estimate and compare the differences in abundances of clonotypes between the samples

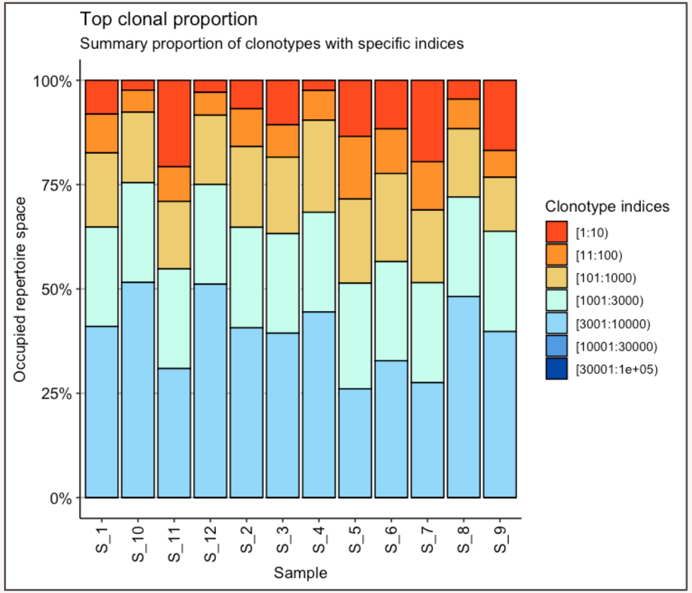
Methods: proportions of the **most** and least **abundant clonotypes**, relative abundance

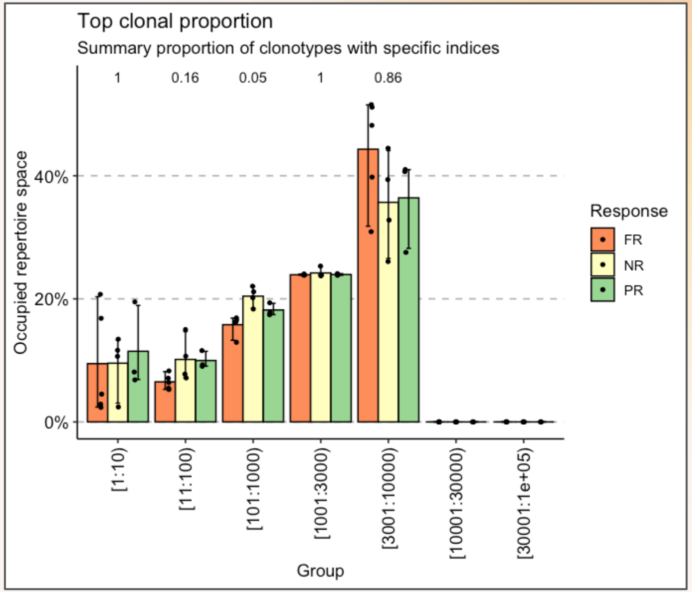
Proportion of the most abundant clonotypes

Compute the proportion of the sample occupied by the pool of the most abundant / most prevalent clonotypes

Clonality analysis

Proportion of the most abundant clonotypes





```
repClonality(immdata$data, "top") %>% vis()
```

```
repClonality(immdata$data, "top") %>%
   vis(.by = "Response", .meta =
immdata$meta)
```

Clonality analysis

Goal: estimate and compare the differences in abundances of clonotypes between the samples

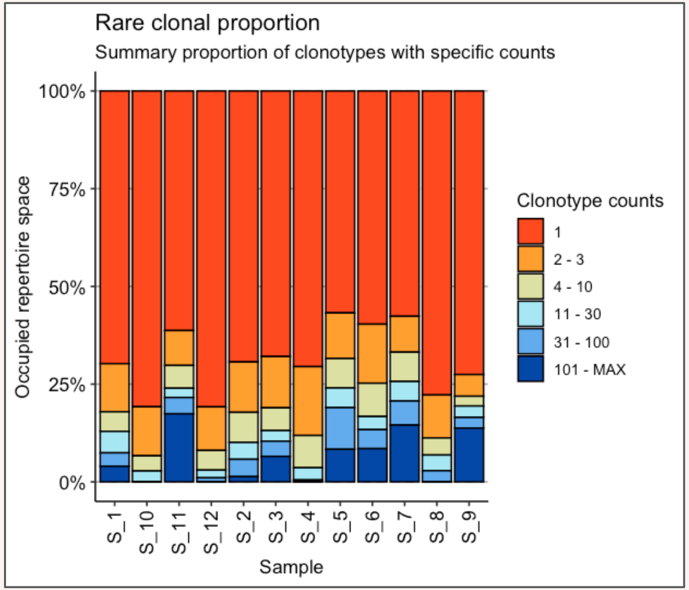
Methods: proportions of the most and **least abundant clonotypes**, relative abundance

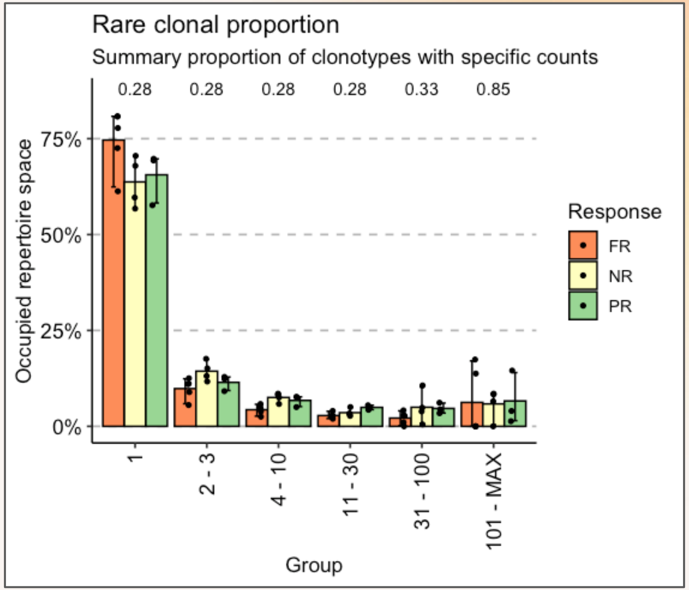
Proportion of the least abundant clonotypes

Compute the proportion of the sample occupied by the pool of the least abundant / rare clonotypes

Clonality analysis

Proportion of the least abundant clonotypes





```
repClonality(immdata$data, "rare") %>% vis()
```

```
repClonality(immdata$data, "rare") %>%
   vis(.by = "Response", .meta =
immdata$meta)
```

Clonality analysis

Goal: estimate and compare the differences in abundances of clonotypes between the samples

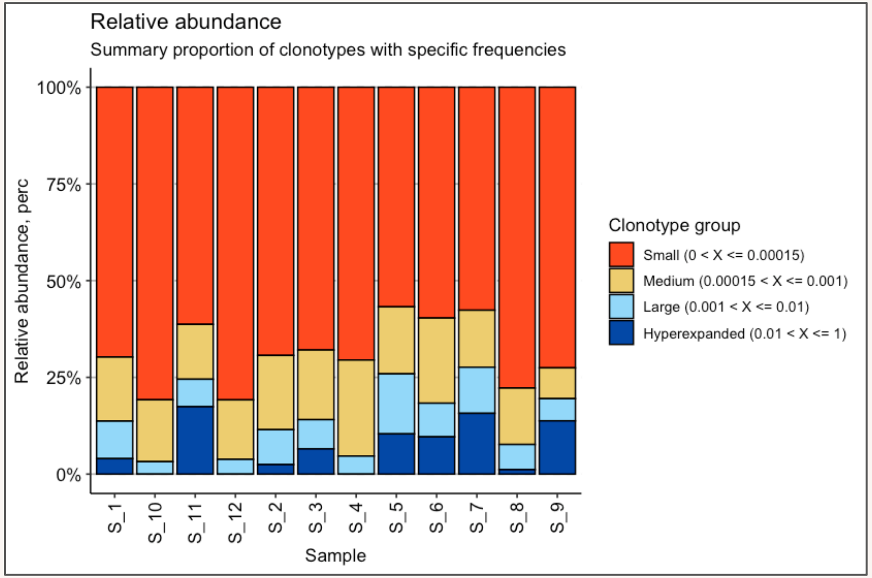
Methods: proportions of the most and least abundant clonotypes, **relative abundance**

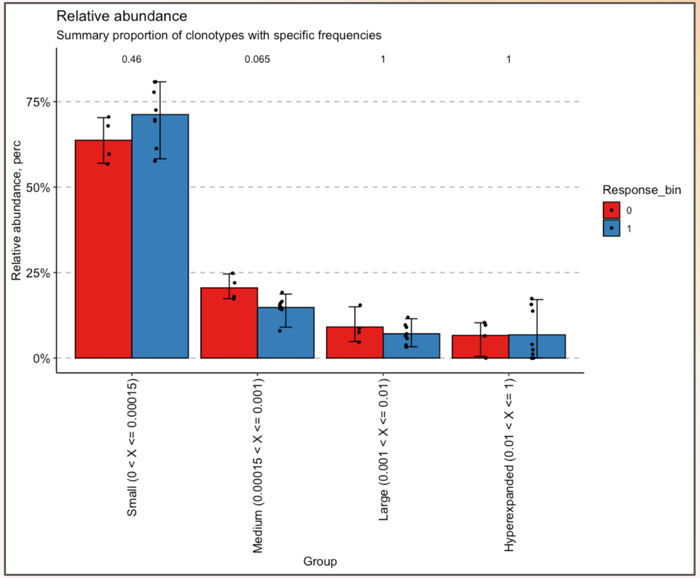
Relative abundance

Compute the proportion of the sample occupied by the clonotypes of given sizes. I.e., estimate the "architecture" of the sample

Clonality analysis

Relative abundance





```
repClonality(immdata$data, "homeo") %>%
  vis()
```

```
repClonality(immdata$data, "homeo") %>%
  vis(.by = "Response", .meta =
immdata$meta)
```

Diversity analysis

Goal: estimate and compare the diversity of clonotypes in samples

Applications: patient selection, response prediction, prognosis biomarkers

Highlighted publications

1. Characteristics of anti-CD19 CAR T cell infusion products associated with efficacy and toxicity in patients with large B cell lymphomas Deng et al.
2. Transient rest restores functionality in exhausted CAR-T cells through epigenetic remodeling Weber et al.
3. T-cell receptor repertoire analysis for the diagnosis and treatment of solid tumor: A methodology and clinical applications Li et al.
4. The T cell receptor repertoire of tumor infiltrating T cells is predictive and prognostic for cancer survival Valpione et al.

Diversity analysis

Goal: estimate and compare the diversity of clonotypes in samples

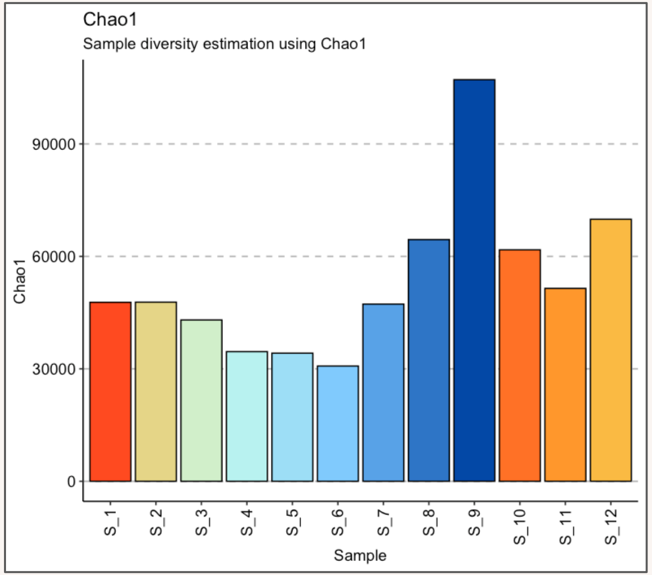
Methods: **Chao1**, rarefaction analysis, Hill numbers

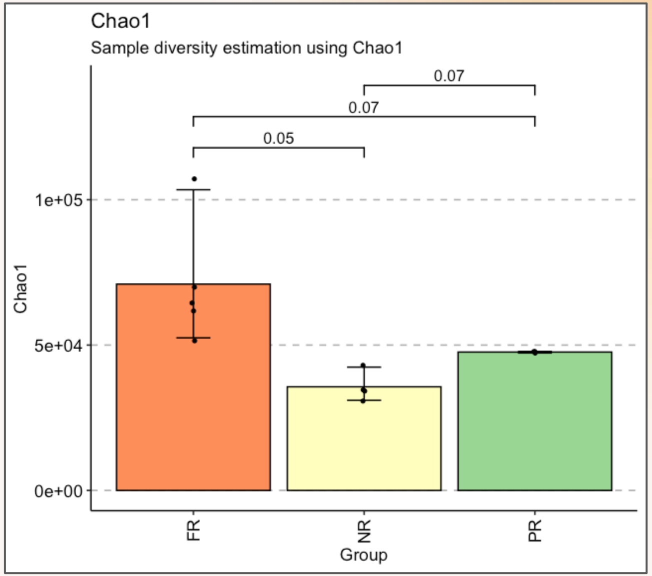
Chao1

Estimate the number of "species" (e.g., clonotypes / cells with different sequences) in the sample using nonparameteric asymptotic estimator

Diversity analysis

Chao1





```
repDiversity(immdata$data, "chao1") %>%
  vis()
```

```
repDiversity(immdata$data, "chao1") %>%
  vis(.by = "Response", .meta =
immdata$meta)
```

Diversity analysis

Goal: estimate and compare the diversity of clonotypes in samples

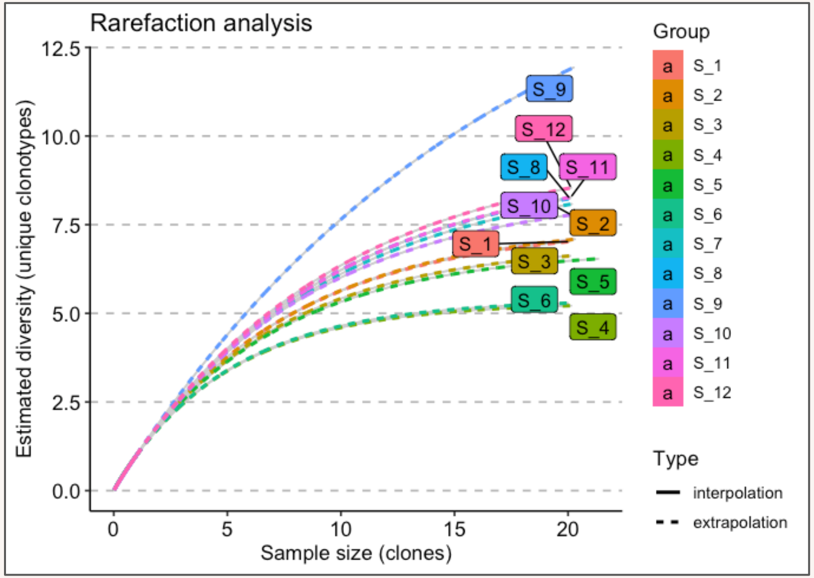
Methods: Chao1, **rarefaction analysis**, Hill numbers

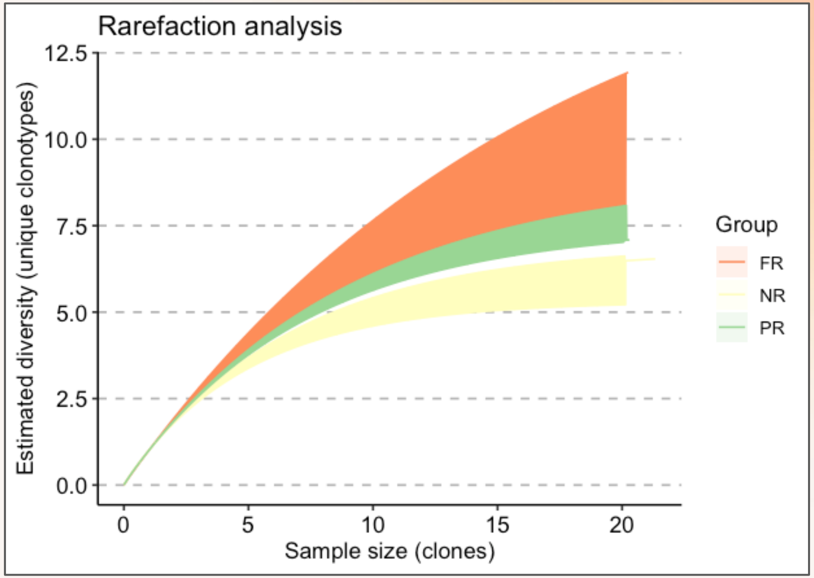
Rarefaction analysis

Estimate the number of clonotypes using extrapolation techniques. Provides more details in comparison than Chao1

Diversity analysis

Rarefaction analysis





```
repDiversity(immdata$data, "raref") %>%
  vis()
```

```
repDiversity(immdata$data, "raref") %>%
  vis(.by = "Response", .meta =
immdata$meta)
```

Diversity analysis

Goal: estimate and compare the diversity of clonotypes in samples

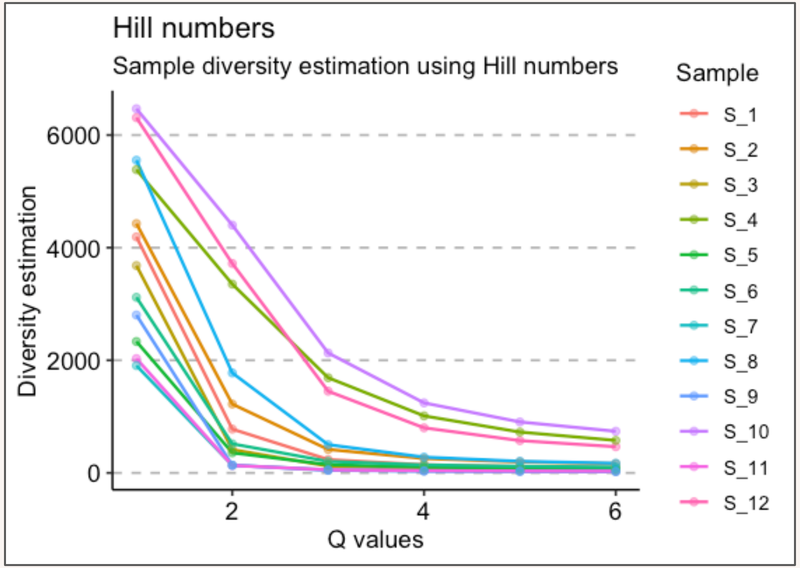
Methods: Chao1, rarefaction analysis, **Hill numbers**

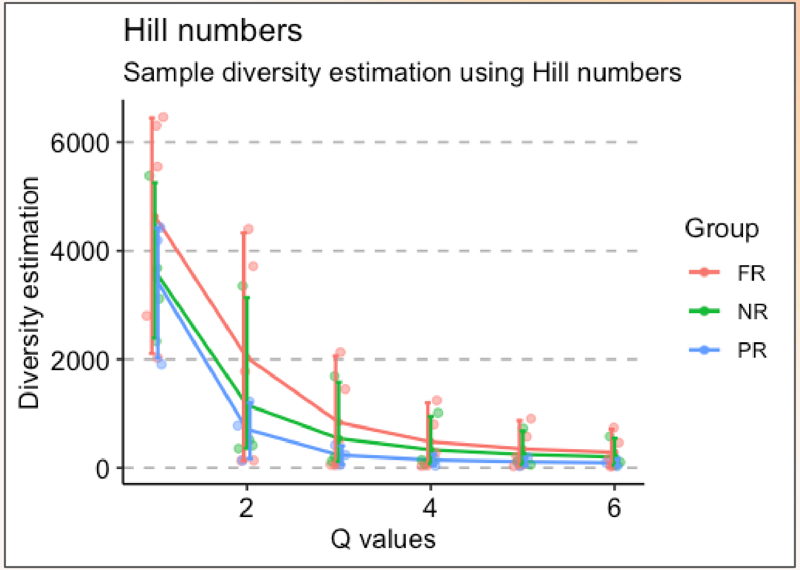
Hill numbers

Assess the structure of the sample's clonality – what clonotypes are driving the distribution of clonotype abundances. The idea is similar to the relative abundance

Diversity analysis

Hill numbers





```
repDiversity(immdata$data, "hill") %>%
  vis()
```

```
repDiversity(immdata$data, "hill") %>%
  vis(.by = "Response", .meta =
immdata$meta)
```

Gene usage analysis

Goal: estimate the frequency of Variable, Diversity and Joining gene segments to characterize the samples by the usage of specific gene segments and families

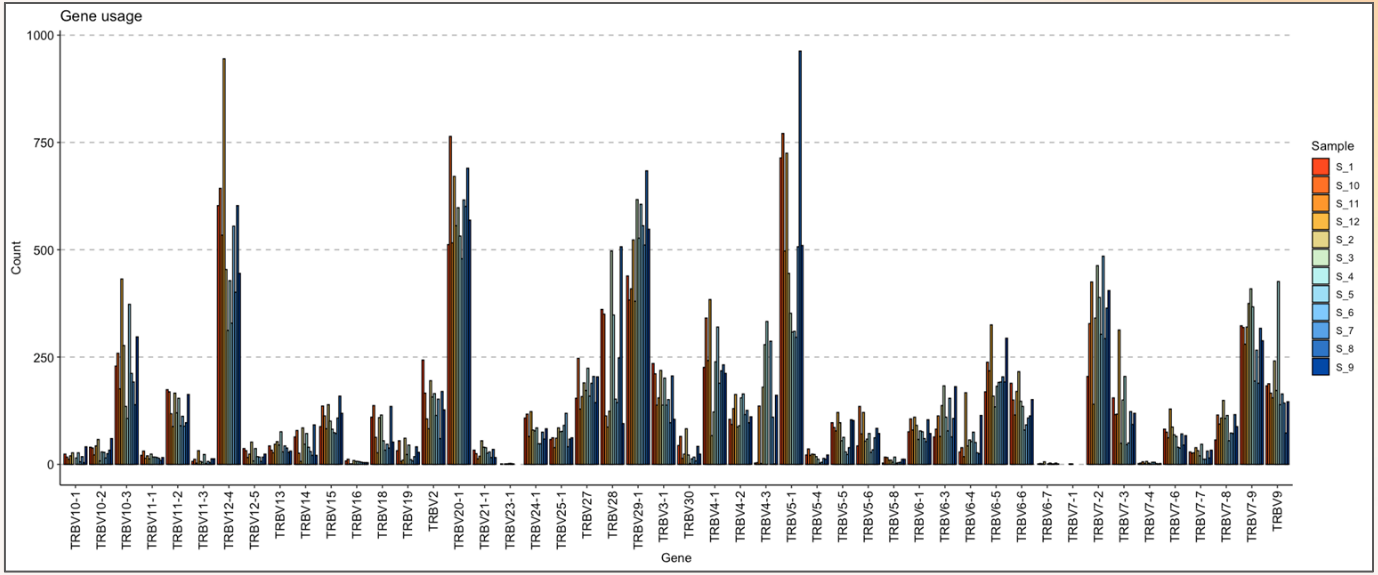
Applications: disease associations, patient selection, response prediction, prognosis biomarkers

Highlighted publications

1. A High-Avidity T-cell Receptor Redirects Natural Killer T-cell Specificity and Outcompetes the Endogenous Invariant T-cell Receptor Landoni et al.
2. Altered Repertoire Diversity and Disease-Associated Clonal Expansions Revealed by T Cell Receptor Immunosequencing in Ankylosing Spondylitis Patients Hanson et al.

Gene usage analysis

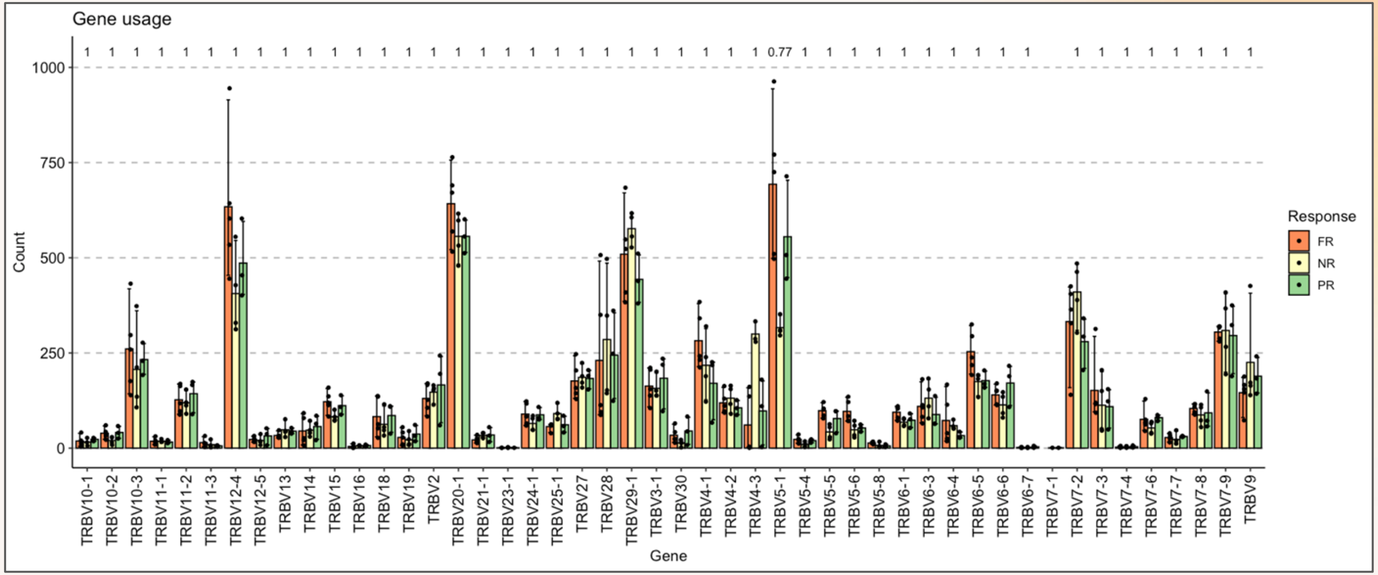
V-/D-/J-gene segments statistics



```
geneUsage(immdata$data) %>% vis()
```

Gene usage analysis

V-/D-/J-gene segments statistics



```
geneUsage(immdata$data) %>%
   vis(.by = "Response", .meta =
immdata$meta)
```

Annotation and tracking

Goal: characterize clonotypes in the samples using external ("annotation") and internal ("tracking") databases with clonotype information, such as links to specific diseases

Applications: CAR-T assessment and optimization, response prediction

Highlighted publications

1. Annotation – characterize the clonotypes of interest, identified on the previous analysis steps or in some other way. Support for VDJdb, McPAS, PIRD databases
2. Tracking – assess the persistence and expansion of CAR-T cells
3. Weighting tumor-specific TCR repertoires as a classifier to stratify the immunotherapy delivery in non–small cell lung cancers Han et al.
4. Clonal kinetics and single-cell transcriptional profiling of CAR-T cells in patients undergoing CD19 CAR-T immunotherapy Sheih et al.

Annotation and tracking

Goal: characterize clonotypes in the samples using external ("annotation") and internal ("tracking") databases with clonotype information, such as links to specific diseases

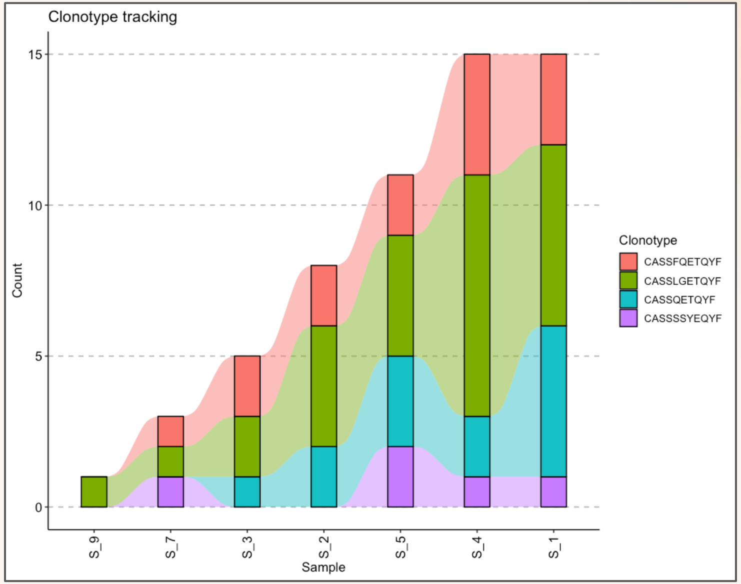
Methods: **annotation**, tracking

Annotation

Link clonotypes in the samples to specific conditions using external databases and gain insights into the behavior and content of immune repertoires

Annotation and tracking

Annotation (timepoints)



```
db = dbLoad(file_path, "mcpas", "Human",
"TRB", "Colorectal cancer")
dbAnnotate(immdata$data, db, "CDR3.aa",
"CDR3.beta.aa") %>% vis()
```

Annotation and tracking

Goal: characterize clonotypes in the samples using external ("annotation") and internal ("tracking") databases with clonotype information, such as links to specific diseases

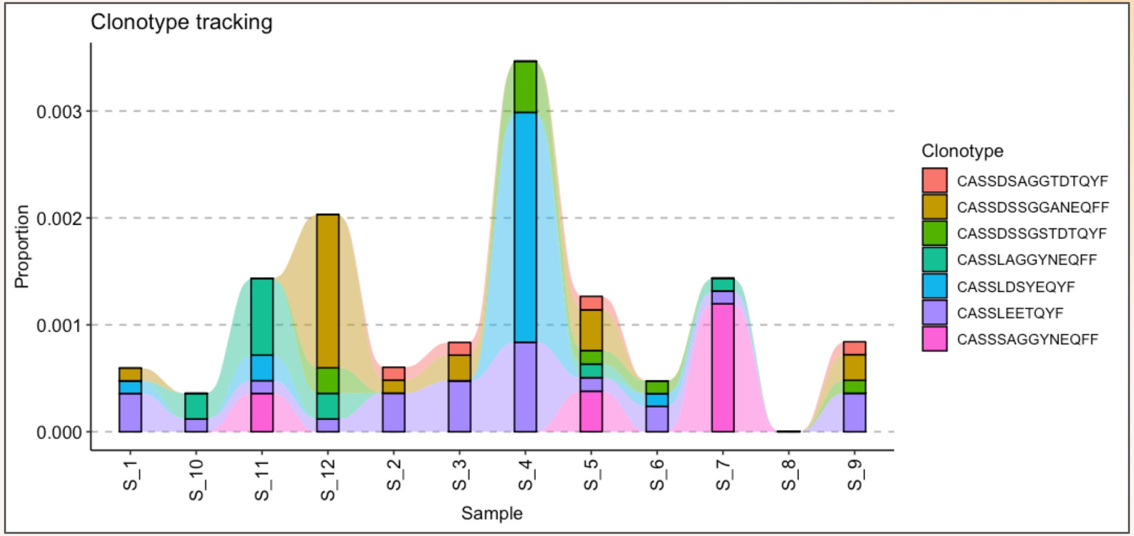
Methods: annotation, **tracking**

Tracking

Track specific clonotypes across samples or time points to gain insights into the immune repertoire dynamics of a tumor, cell therapy, etc.

Annotation and tracking

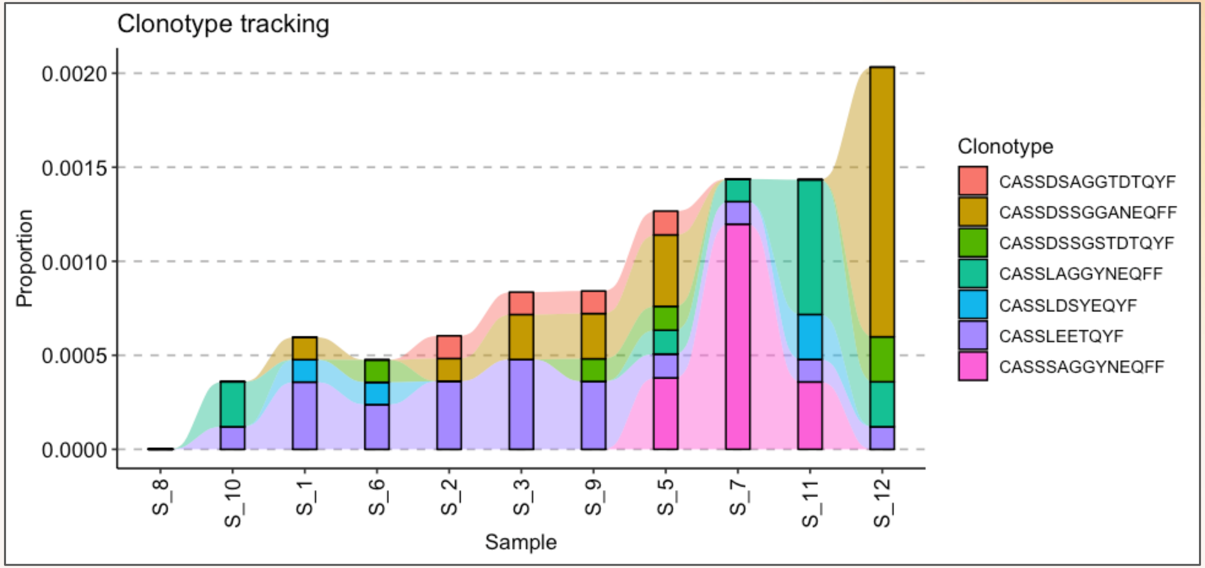
Tracking of target clonotypes



```
trackClonotypes(immdata$data, target,
.col = "aa") %>%
   vis(.plot = "smooth")
```

Annotation and tracking

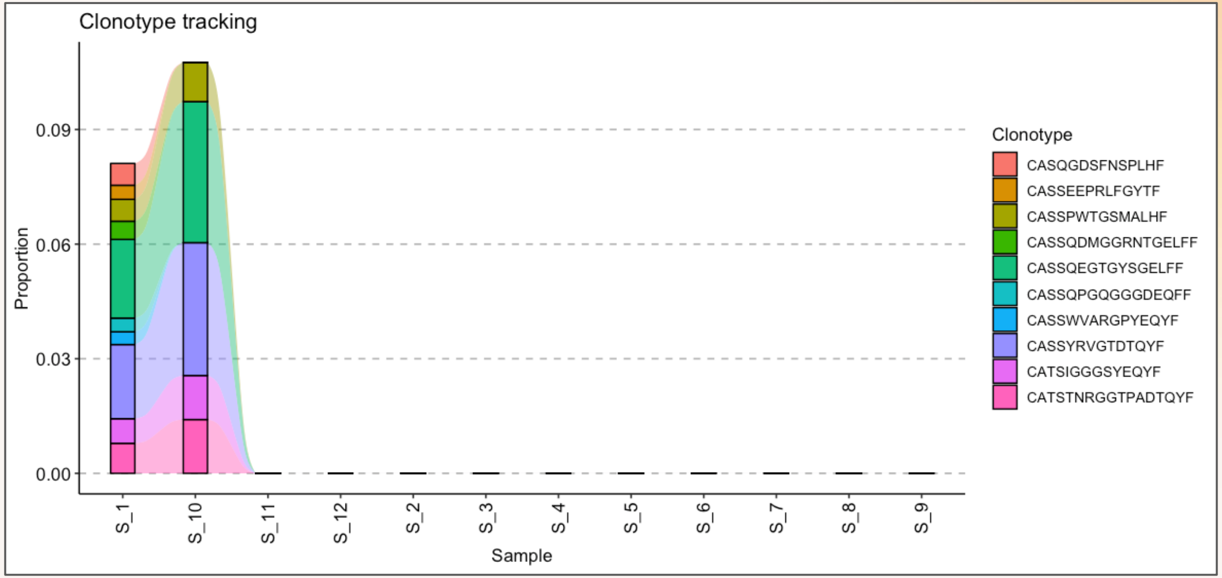
Tracking between time points



```
trackClonotypes(immdata$data, target,
.col = "aa") %>%
   vis(.plot = "smooth", .order = sample_order)
```

Annotation and tracking

Tracking the most abundant clonotypes from a specific repertoire



```
trackClonotypes(immdata$data, list(1, 10),
.col = "aa") %>%
   vis(.plot = "smooth", .order = sample_order)
```

Future of Immunarch and advanced topics I

Note: https://immunarch.com contains tutorials for all previous and some of the advanced topics

Data-specific analytics

1. **B-cell Receptor Lineage Trees.** Will be released in November, 2021. Contact us for the development version
2. **Single-cell immunogenomics, i.e., paired-chain data.** The basic support is already released (overlaps, diversity, tracking). We will be working on improvements, so suggestions are welcome
3. **Single-cell transcriptomics, CITE-seq, etc.** AIRR and SC (bulk as well) data integration. Currently on the ImmunoMind's analytics platform, but we plan to open-source our internal software tools

Future of Immunarch and advanced topics II

Note: https://immunarch.com contains tutorials for all previous and some of the advanced topics

Advanced statistical analysis

1. **Advanced gene usage.** Therapy response modelling and biomarker discovery of gene associations using gene usage
2. **Advanced diversity.** Advanced diversity indices and immune repertoire modelling
3. **Post-analysis.** Clustering of immune repertoires by overlaps or gene usages
4. **Public repertoire analysis.** Analysis of abundance statistics of shared or annotated clonotypes. E.g., the overall abundance of autoimmune clonotypes in samples
5. **Kmer and sequence motif analysis.** Sequence clustering and motif discovery

Future of Immunarch and advanced topics III

Note: https://immunarch.com contains tutorials for all previous and some of the advanced topics

Improvements

1. **User experience.** More verbose error messages with helpful information on how to solve issues. Helpful filtering and preprocessing functions
2. **More convenient visualizations.** Data class to stop passing ".meta" to charts
3. **More tutorials and learning materials.** We value education a lot, so check our websites and socials for more tutorials and reports on CAR-T

Communication & Support

GitHub Issues

Preferred method of communication. Bring us questions, bugs, and feature suggestions by opening a ticket on GitHub

Link: https://github.com/immunomind/immunarch/issues

Email

For sharing and discussing sensitive information

Email: support@immunomind.io

Contact me about your projects!

CAR-T cell / TCR therapies, other immunotherapies

Vadim I. Nazarov, CEO & Co-Founder
vadim@immunomind.io | +1 415 800 28 92

Website: https://immunomind.io
LinkedIn: https://www.linkedin.com/in/vdnaz/

Immunarch tutorials: https://immunarch.com

GitHub (stars are appreciated!)**:**
https://github.com/immunomind/immunarch/