

# Investigating CD8 T cell antigen reactivities

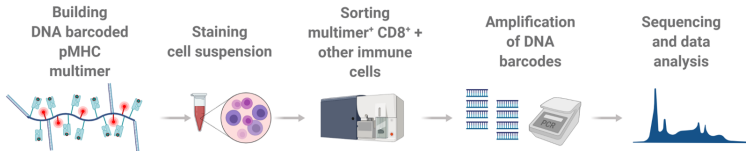
Rasmus, Mikkel, Lasse - Group 19

2022-05-07

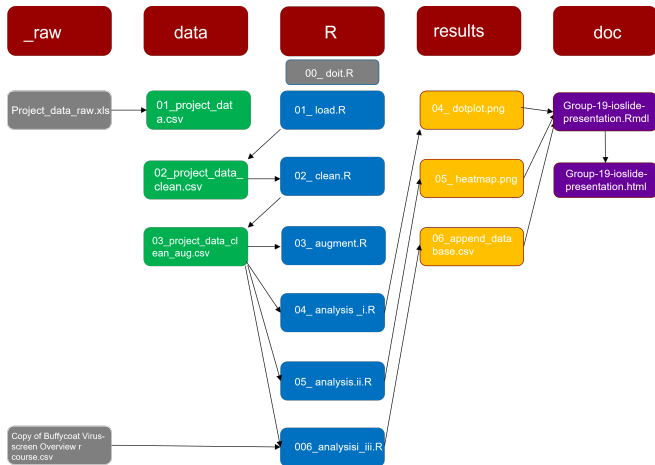
# Introduction

- ▶ Sequencing -> Barracoda pipeline -> our wrangling + visualization
- ▶ Aim: To build a pipeline of data wrangling and visualizations after barracoda pipeline to explore sequence hits

## T cell identification from blood



# Flow chart or flowchart



# Project data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	barcode	sample	count.1	input.1	input.2	input.3	log_fold_c_p	-log10(p)	masked_p	-log10(ma	count.norm	input.norm	HLA	Origin	Peptide	Sequence	
2	A188200	BC372	20	221	128	172	-1,19684	0,356135	0,448386	1	0	0,001852	0,004247	A0101	CMV	v9	YSEHPTFTSQY
3	A198200	BC372	62	325	203	292	-0,22758	0,929139	0,031919	1	0	0,00574	0,00672	A0101	CMV	v15	VTEHDTLLY
4	A208200	BC372	20	167	132	155	-1,02688	0,475387	0,322953	1	0	0,001852	0,003779	A0101	HHV-1	a42	VLEETSVMVL
5	A218200	BC372	29	260	170	215	-0,98056	0,475387	0,322953	1	0	0,002685	0,005301	A0101	FLU-A	a43	YVLDLHLIV
6	A228200	BC372	42	307	185	212	-0,56483	0,827113	0,082426	1	0	0,003888	0,005752	A0101	CMV	a44	SLYNTVATL
7	A238200	BC372	48	369	247	291	-0,74802	0,66238	0,178893	1	0	0,004444	0,007466	A0101	HadV-C	a45	YMLDLQPETT
8	A248200	BC372	55	327	196	277	-0,36081	0,929139	0,031919	1	0	0,005092	0,006538	A0201	FLU-A	v3	GILGFVFTL
9	A258200	BC372	77	528	348	456	-0,6195	0,776893	0,109639	1	0	0,007128	0,010954	A0201	EBV	v5	CLGGLTMTV
10	A188201	BC372	27	174	116	132	-0,47362	0,929139	0,031919	1	0	0,0025	0,003473	A0201	EBV	v6	GLCTLVAML
11	A198201	BC372	25	158	90	139	-0,4466	0,929139	0,031919	1	0	0,002314	0,003153	A0201	EBV	v10	FLYALALL
12	A208201	BC372	44	84	50	63	1,33699	0,169828	0,76999	0,169828	0,76999	0,004073	0,001609	A0201	CMV	v13	NLVPVMVATV
13	A218201	BC372	38	112	54	71	0,878481	0,455815	0,341211	0,455815	0,341211	0,003518	0,001907	A0201	EBV	v16	YVLDLHLIV
14	A228201	BC372	41	215	138	167	-0,16921	0,952783	0,021006	1	0	0,003796	0,004268	A0201	CMV	a46	FLYALALL
15	A238201	BC372	24	119	63	87	0,025416	0,971028	0,012768	0,971028	0,012768	0,002222	0,00218	A0201	B19	a47	RLNELLAYV
16	A248201	BC372	9	44	34	43	-0,26739	0,952783	0,021006	1	0	0,000833	0,001006	A0201	CMV	a48	RMLGDVMAV
17	A258201	BC372	38	160	107	113	0,16969	0,929139	0,031919	0,929139	0,031919	0,003518	0,003128	A0201	EBV	a49	FLGDDPSA
18	A108226	BC372	332	618	405	502	1,293653	0,106509	0,972615	0,106509	0,972615	0,030735	0,012535	A0201	HHV-1	a50	FLGGHVAVL
19	A118226	BC372	151	341	231	306	0,950207	0,325186	0,487868	0,325186	0,487868	0,013979	0,007233	A0201	VZV	a51	ILMWEAVTL
20	A128226	BC372	80	301	234	284	0,121582	0,933802	0,029745	0,933802	0,029745	0,007406	0,00681	A0201	EBV	a52	STLGEVCLF
21	A108224	BC372	236	810	634	695	0,295916	0,929139	0,031919	0,929139	0,031919	0,021848	0,017799	A0201	HadV-C	a53	TIHDIILCEV
22	A148226	BC372	36	208	131	136	-0,22391	0,93328	0,029988	1	0	0,003333	0,003892	A0201	VACV	a54	GIFEDRAPV
23	A158226	BC372	40	202	109	130	0,049369	0,965493	0,015251	0,965493	0,015251	0,003703	0,003575	A0201	VACV	a55	TLGIVCPI
24	A168226	BC372	103	632	411	515	-0,42464	0,929139	0,031919	1	0	0,009535	0,0128	A0201	EBV	a56	NMLSTVLGV
25	A178226	BC372	24	93	55	80	0,253196	0,929139	0,031919	0,929139	0,031919	0,002222	0,001862	A0201	B19	a57	NLLTTPKFT
26	A108227	BC372	167	484	366	457	0,511414	0,776893	0,109639	0,776893	0,109639	0,01546	0,010847	A0201	CMV	a58	YLQGNWWTL
27	A118227	BC372	266	621	417	534	0,927777	0,325186	0,487868	0,325186	0,487868	0,024625	0,012943	A0201	EBV	a59	AMLLRLRI
28	A128227	BC372	113	335	221	264	0,633288	0,635667	0,19677	0,635667	0,19677	0,010461	0,006743	A0201	HPV	a60	FIPQYLSAV
29	A108225	BC372	300	687	545	615	0,852618	0,382559	0,417302	0,382559	0,417302	0,027773	0,015382	A0201	HPV	a61	YLLEMLWRL
30	A148227	BC372	81	355	255	313	-0,02557	1	0	1	0	0,007499	0,007634	A0201	NWV	a62	AIMDKNIIL
31	A158227	BC372	56	326	236	315	-0,48463	0,919387	0,036501	1	0	0,005184	0,007257	A0201	HMPV	a63	LLDFRVFMGV
32	A168227	BC372	51	181	132	168	0,245948	0,929139	0,031919	0,929139	0,031919	0,004721	0,003982	A0201	HHV-1	a64	KVDDTFYVY
33	BC372	BC383	BC385	BC386	BC387	BC388	BC389	BC390	BC391	BC393	BC394	BC395	N.T.C.F.				

Figure 1: Raw excel-file contains several sheets

## Loading data

```
setwd("/cloud/project")
```

```
data <- read_excel("data/_raw/project_data_raw.xlsx")
```

```
data
```

```
## # A tibble: 118 x 17
```

```
##   barcode sample count.1 input.1 input.2 input.3 log_fc
```

```
##   <chr>      <chr>      <dbl>    <dbl>    <dbl>    <dbl>
```

```
## 1 A18B200 BC372         20      221      128      172
```

```
## 2 A19B200 BC372         62      325      203      292
```

```
## 3 A20B200 BC372         20      167      132      155
```

```
## 4 A21B200 BC372         29      260      170      215
```

```
## 5 A22B200 BC372         42      307      185      212
```

```
## 6 A23B200 BC372         48      369      247      291
```

```
## 7 A24B200 BC372         55      327      196      277
```

```
## 8 A25B200 BC372         77      528      348      456
```

```
## 9 A18B201 BC372         27      174      116      132
```

```
## 10 A19B201 BC372         25      158        90      139
```

## Loading data - merging sheets

```
setwd("/cloud/project")

# Accessing all excel sheets
sheet <- excel_sheets("data/_raw/project_data_raw.xlsx")

# Creating a list of individual data frames for each sheet
data_frame <- lapply(setNames(sheet, sheet),
                      function(x) read_excel("data/_raw/pr
                                             sheet = x))

# Attaching individual data frames together
data_frame <- bind_rows(data_frame,
                        .id = "Sheet")

data_frame
```

```
## # A tibble: 1,770 x 18
```

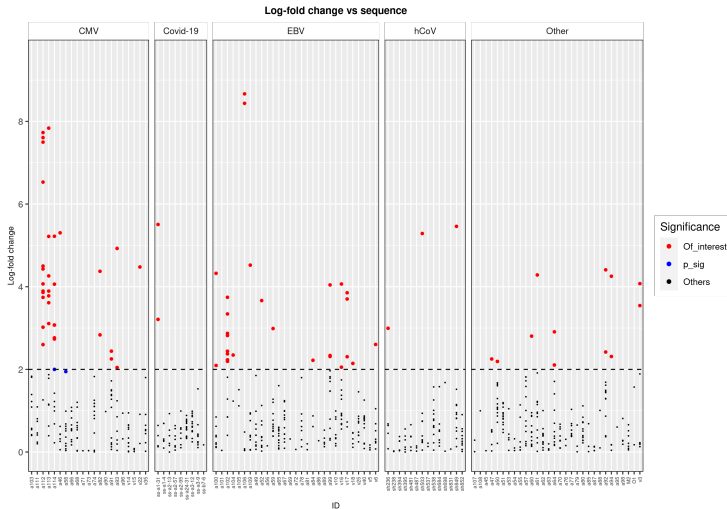
```
##   Sheet barcode sample count.1 input.1 input.2 input.3
```

```
##   <chr> <chr>    <chr>    <dbl>  <dbl>  <dbl>  <dbl>
```

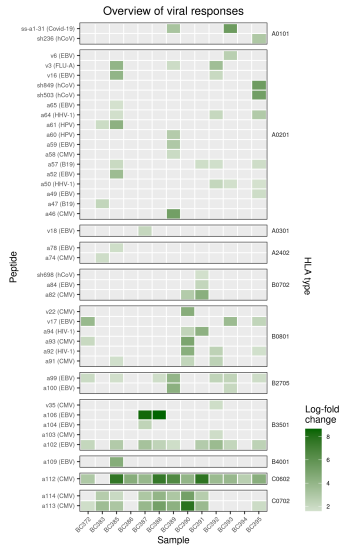
```
## 1 PG372 A18P200 PG372      20     221     128     172
```

# Viral responses in multiple samples?

- Goal: Visualize whether specific CD8 T cells that recognize the same viral epitope are found in multiple samples.



## Simple overview of viral responses





## Donor response database

- The database has a special format

[illegible]

## Appending new responses on the response database

- ▶ Remove irrelevant data ( $\log\_fold\_change > 2$ )
- ▶ Match peptides in the database sheet with the the ones from the dataset
- ▶ Assign a database row position to the matched peptides
- ▶ Create a dataframe with the same amount of rows and add matching peptides on corresponding rows
- ▶ Append it on the database

# Adding database row positions to matching data

```
> sequence_matches
```

```
[1] 36 NA NA NA NA NA NA 22 35 NA NA NA NA NA NA N
A NA 37 NA NA NA NA NA NA NA NA NA NA NA 29 NA NA N
A NA NA NA NA NA NA 60 41 NA NA NA NA NA NA 60 NA N
A NA NA NA NA 25 36 NA NA NA NA NA NA NA 36 NA NA N
A
[68] NA NA 22 35 NA NA NA NA NA NA NA
> |
```



barcode	sample	log <sub>2</sub> fold change	p	HLA	Origin	Peptide	Sequence	id	sequence_matches
A18281	BC372	3.853277	2.155473e-18	B*08:01	BBV	v17	RAKTKQLL	v17 (BBV)	36
A128232	BC372	2.044920	9.801409e-04	B*08:01	CMV	j83	SLNTIATL	j83 (CMV)	NA
A182815	BC372	2.190579	9.654926e-04	B*50:01	BBV	j192	LLGSLUNV	j192 (BBV)	NA
A178234	BC372	3.020955	2.809886e-08	C*06:02	CMV	j112	YLANGGFL	j112 (CMV)	NA
A218201	BC385	2.251417	2.530436e-04	A*02:01	B19	j47	RNLLEKVF	j47 (B19)	NA
A108235	BC385	5.612496	5.964419e-12	C*07:02	CMV	j118	FLGAGALV	j118 (CMV)	NA
A118235	BC385	2.761118	5.106079e-07	C*07:02	CMV	j114	MMLRDRWL	j114 (CMV)	NA
A248200	BC385	4.075380	1.521824e-15	A*02:01	FLU A	v8	GIQSPVPL	v8 (FLU A)	22
A218201	BC385	4.065788	4.538075e-15	A*02:01	BBV	v16	YVDMGVV	v16 (BBV)	35
A128226	BC385	3.664117	6.496718e-13	A*02:01	BBV	j42	STEVKCYL	j42 (BBV)	NA
A178226	BC385	2.014044	1.019286e-03	A*02:01	B19	j57	NLITTPVT	j57 (B19)	NA
A108225	BC385	4.284511	3.245033e-17	A*02:01	MPV	j61	YLEMKWRL	j61 (MPV)	NA
A182815	BC385	2.871852	1.499926e-07	B*50:01	BBV	j192	LLGSLUNV	j192 (BBV)	NA
A148234	BC385	4.522632	1.432462e-18	B*06:01	BBV	j109	TLDMRTTA	j109 (BBV)	NA
A178234	BC385	7.608918	2.275886e-44	C*06:02	CMV	j112	YLANGGFL	j112 (CMV)	NA
A108235	BC385	5.108481	1.485786e-09	C*07:02	CMV	j118	FLGAGALV	j118 (CMV)	NA
A178234	BC386	4.500294	3.502106e-18	C*06:02	CMV	j112	YLANGGFL	j112 (CMV)	NA
A128228	BC387	2.144890	2.257469e-04	A*03:01	BBV	v18	RUMAAQVY	v18 (BBV)	37

# Adding values in the right rows

Sequence	sequence_matches	BC372	BC385	BC387	BC389	BC390	BC391	BC393	BC395	BC392
1 RAKFKQLL	36	3.855277	N/A	N/A	N/A	N/A	N/A	3.799042	2.305808	N/A
2 GILGPFVTL	22	N/A	4.075380	N/A	N/A	N/A	N/A	N/A	N/A	3.543469
3 YVLDHLNV	35	N/A	4.065788	N/A	N/A	N/A	N/A	N/A	N/A	2.052760
4 RLRAEAQYK	37	N/A	N/A	2.14489	N/A	N/A	N/A	N/A	N/A	N/A
5 FYVALALLL	29	N/A	N/A	N/A	5.304188	N/A	N/A	N/A	N/A	N/A
6 ILKEPVHGV	60	N/A	N/A	N/A	N/A	2.834547	4.373389	N/A	N/A	N/A
7 ELRRKMMYIM	41	N/A	N/A	N/A	N/A	4.477543	N/A	N/A	N/A	N/A
8 GLCTIVAML	25	N/A	N/A	N/A	N/A	N/A	N/A	2.602879	N/A	N/A

+

sequence_matches	match.ncut_1_xxxx_xxxx donor sequence database
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26

## Final product

[illegible]

Questions?