

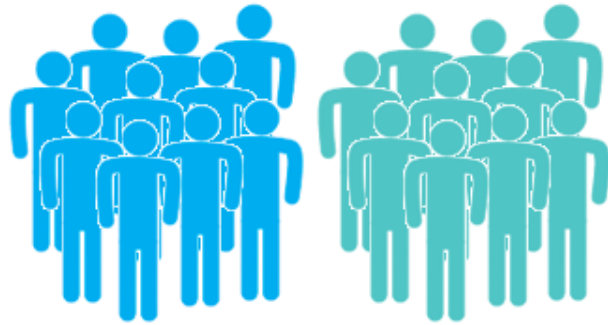
# Post-clustering analyses

---

S. GRANJEAUD



# What is the question?



An experiment starts with a question:

$A \neq B ?$

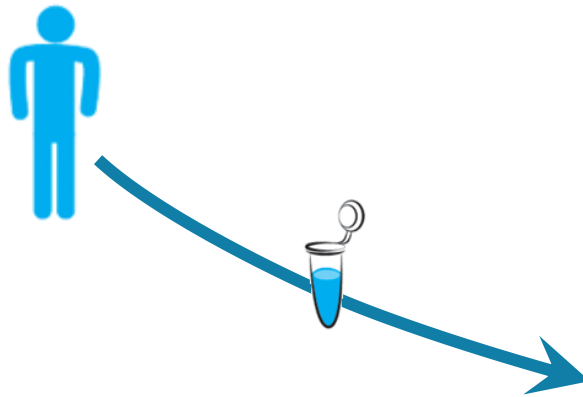


The experiment







Data:

FCS files, scRNAseq files...



**1 FCS**

	Marker-1	2	3	4	5	... X
Cell-1						
Cell-2						
Cell-3						
Cell-4						
Cell-5						
...						
...						

-  Create Bunch
-  Setup Batch
-  Tune Params
-  Process

**Define the amount of cells per FCS file**

**Extract a sample of cells**

**Select the channel to process**

**Batch adjust**

**Set the method to adjust batch effect**

**Set the percentile to adjust batch effect**

Exclude the zeroes from percentiles

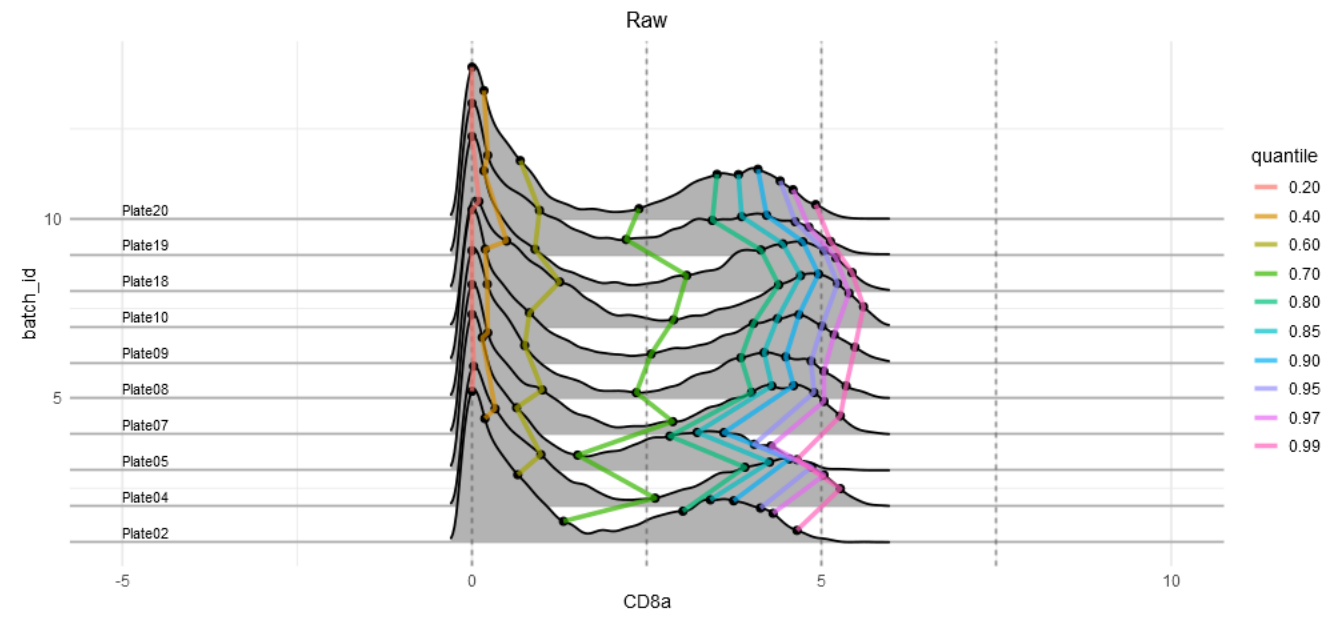
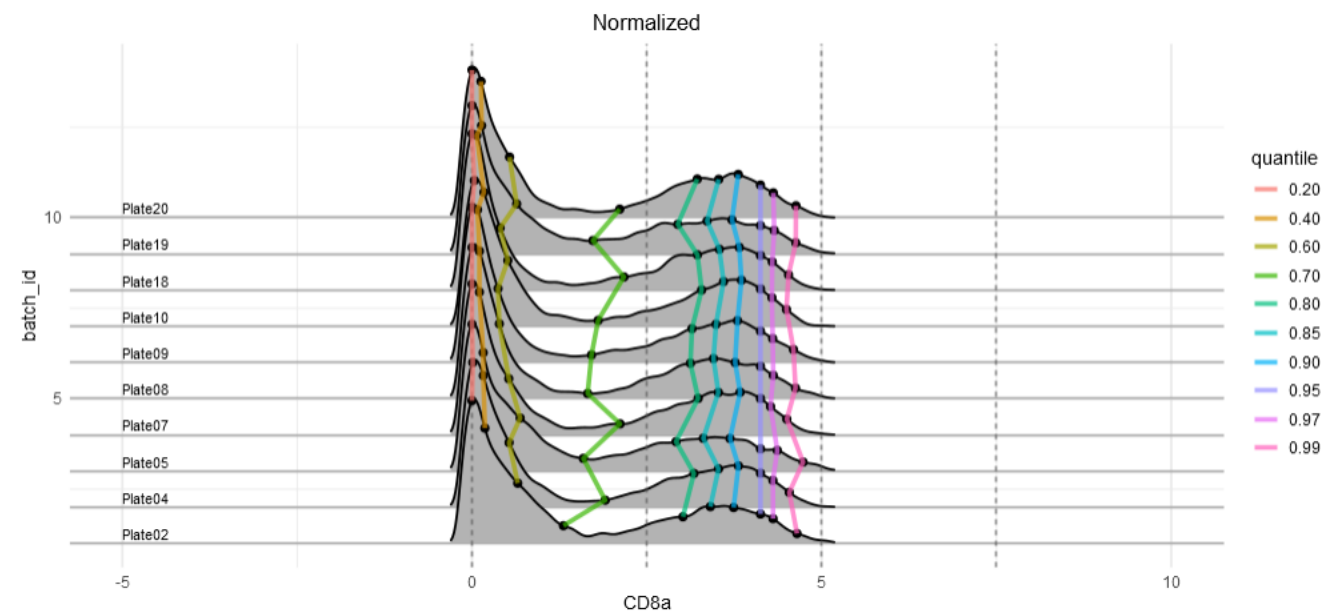
Apply to transformed intensities

**Transform**

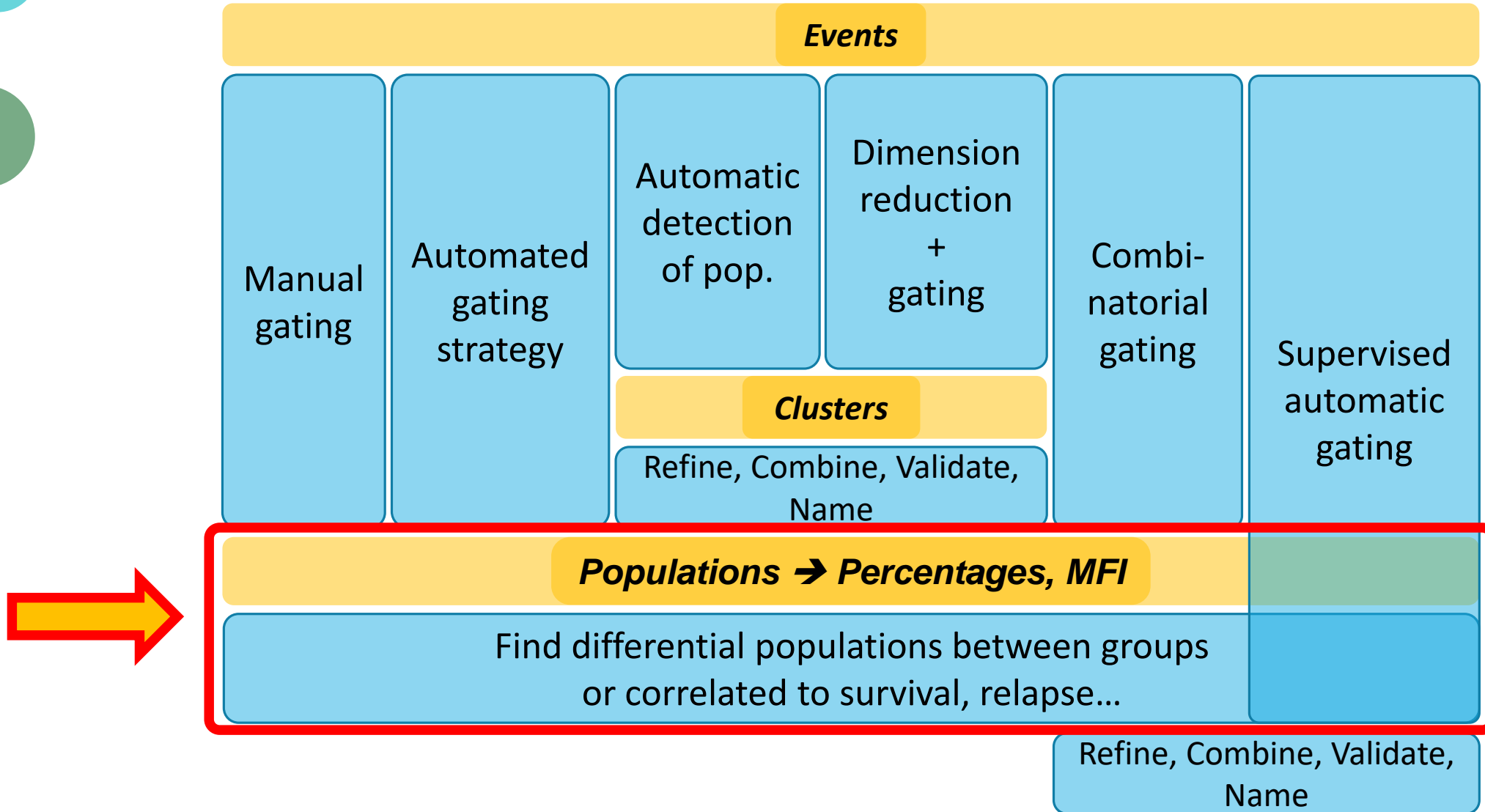
**Set the function**

**Set the cofactor**

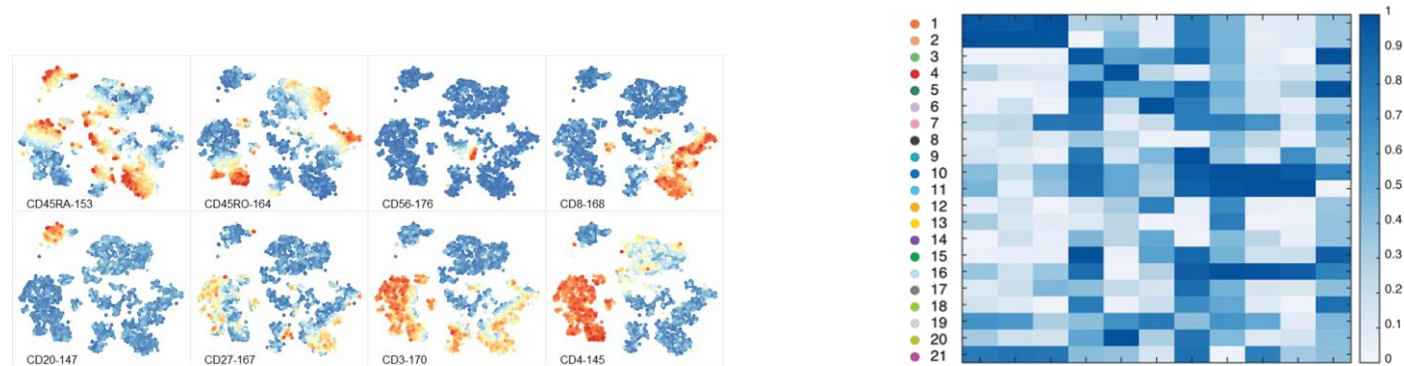
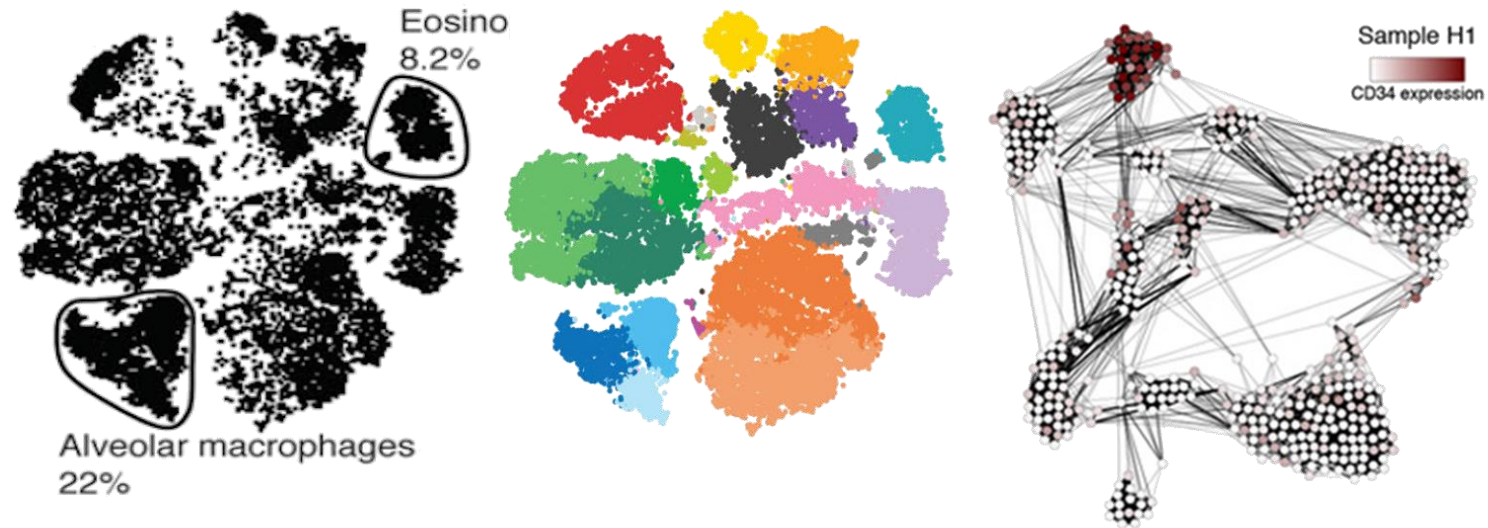
**Graphical options**



# Analyses Overview



# 1 Picture vs 1000 Words



Most illustrations from Mair et al 2016 and Kimball et al 2017

# Visualisation, the basics

Screen

X, Y coord.

Color

Shape

Size



Link the points => graph, tree

# What is on a dot plot?

---

What are those points on the scatter plot, graph?

- Single cells or centers of cell group

What is the meaning of the color

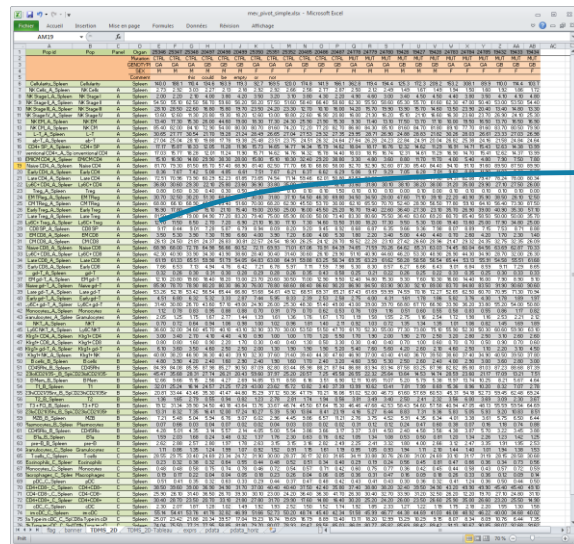
- Group membership or expression level of a marker
- Raw or transformed expression level

Is the distance on the screen meaningful?

How to evaluate the distance between dots?

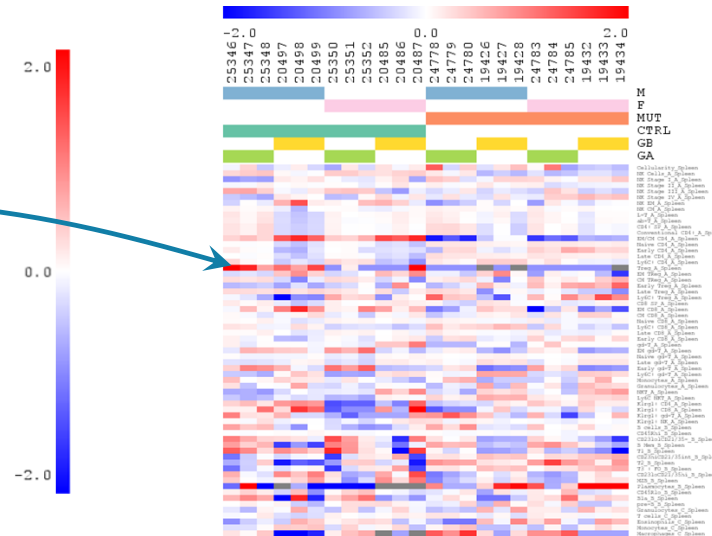
# From matrix to image

## Heatmap



The image shows a screenshot of a Microsoft Excel spreadsheet. The spreadsheet contains a large matrix of numerical data. The columns are labeled with various identifiers, and the rows are labeled with similar identifiers. The data is organized in a grid format, with each cell containing a numerical value. The spreadsheet is titled 'mem\_prot\_complex.xls' and is open in the Microsoft Excel application.

Scale  
Convert





# What is on a heatmap?

## What are those rectangles?

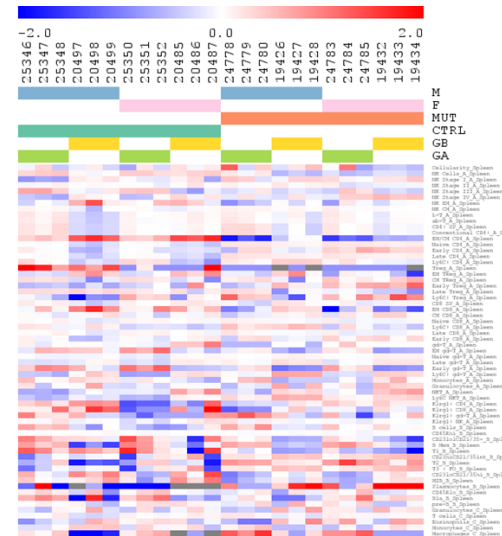
- Columns = ? Rows = ?
- Cell groups x markers or Cell groups x patients

## What is the meaning of the color?

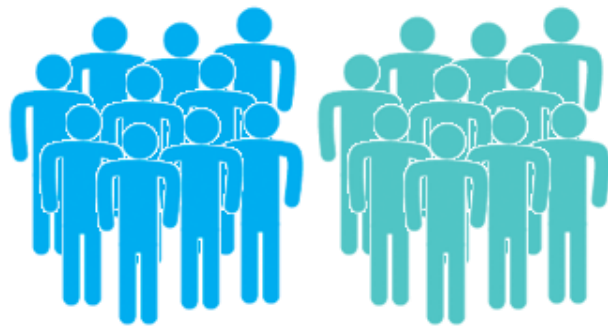
- Expression level of a marker or percentages
- Raw or transformed value
- What is the scaling?

## Criterion for arranging the rows/columns?

- User defined or hierarchical clustering
- How to evaluate the distance?



# What is the question?



FCS	Acq	Batch	Group	Patt	Age	...

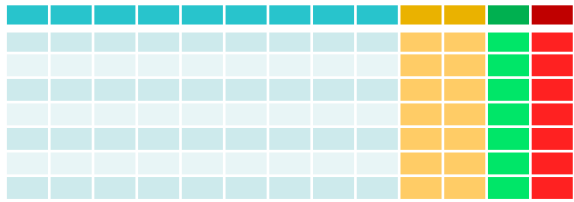


1 FCS

	Mrk1	Mrk2	Mrk...	Mrkm
Cell-1				
Cell-2				
Cell-3				
Cell-4				
Cell-5				
...				
...				

# Extract features

---



Markers

Reduced Dimensions

Clustering

Concatenation (FCS id)

Extract features

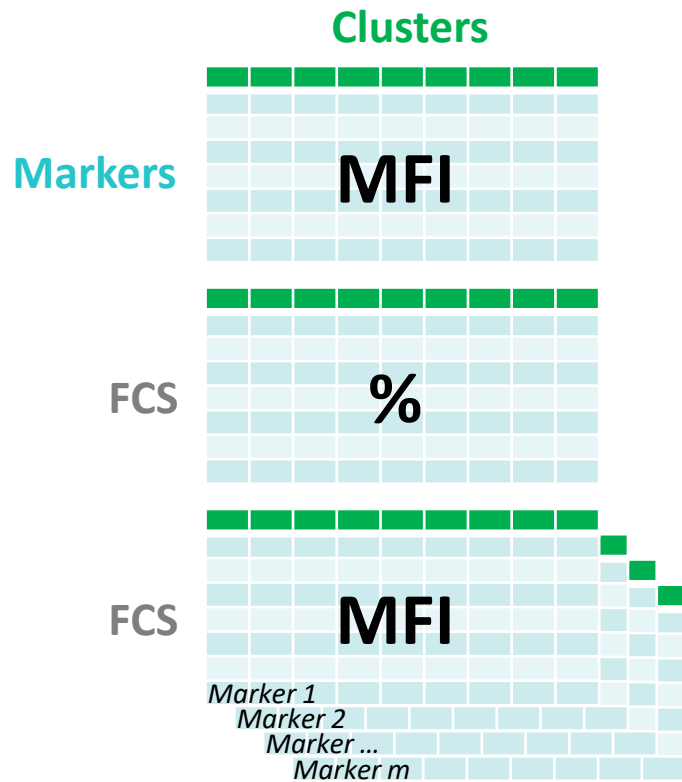
per cluster:

- MFI

per cluster and FCS:

- cell counts => cell abundance
- MFI

# Extract features



Label clusters

Explore %

Explore MFI

Differential  
Abundance

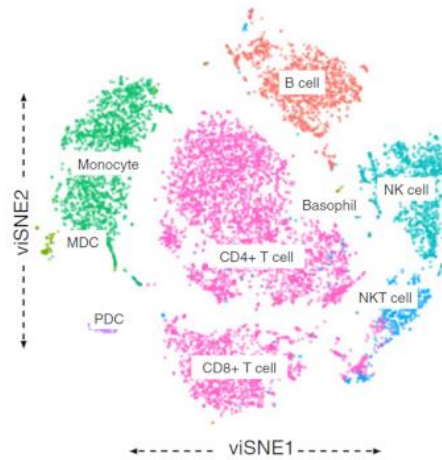
Differential  
State

# Dataset







## Article

- [Pubmed](#)

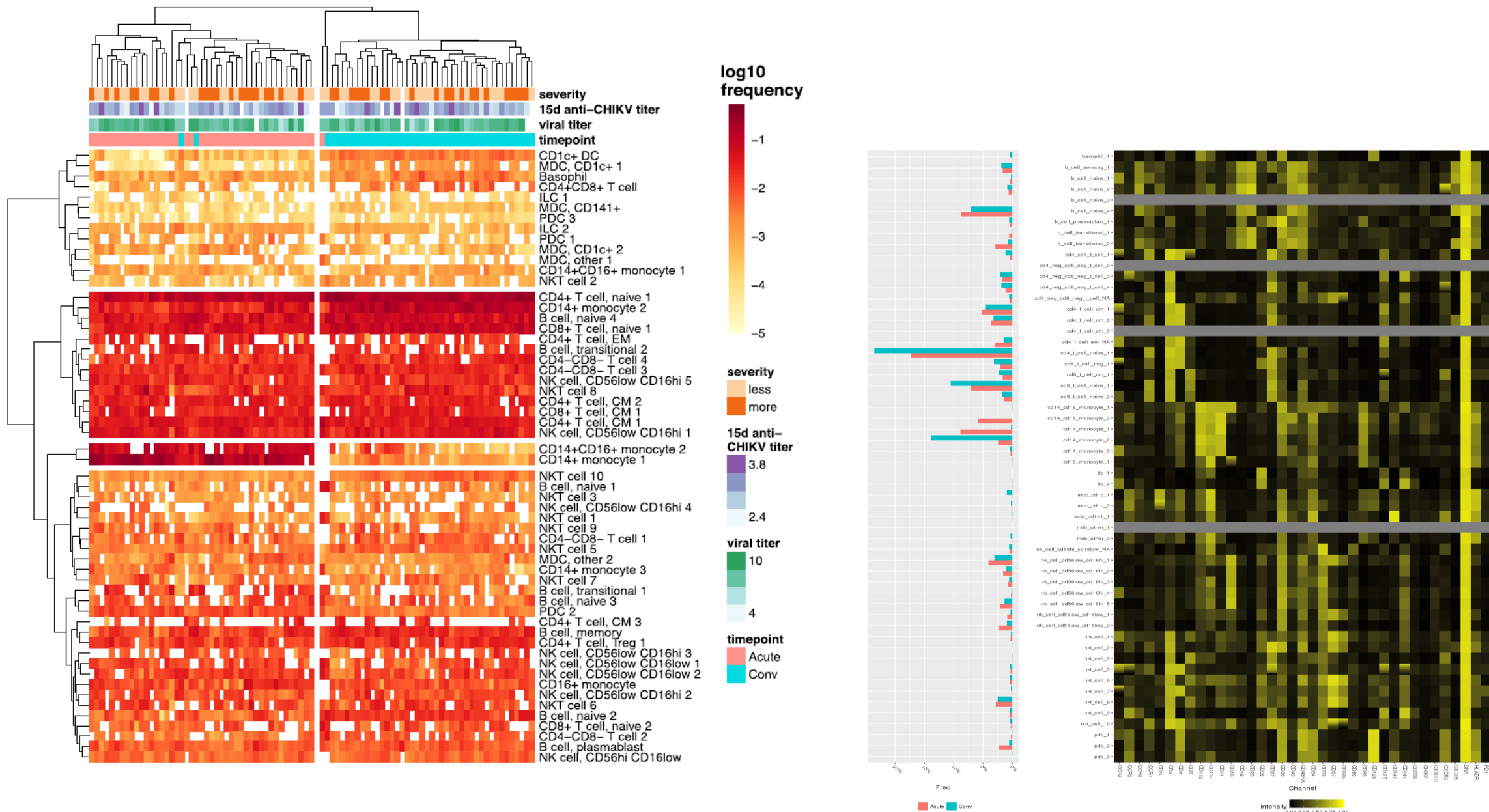


## Comprehensive innate immune profiling of chikungunya virus infection in pediatric cases

Daniela Michlmayr<sup>1,†</sup>, Theodore R Pak<sup>2,†</sup> , Adeeb H Rahman<sup>2,3</sup>, El-Ad David Amir<sup>2,3</sup>, Eun-Young Kim<sup>4</sup> , Seunghee Kim-Schulze<sup>2,3</sup>, Maria Suprun<sup>5</sup>, Michael G Stewart<sup>4</sup>, Guajira P Thomas<sup>4</sup>, Angel Balmaseda<sup>6</sup>, Li Wang<sup>2</sup>, Jun Zhu<sup>2</sup> , Mayte Suárez-Fariñas<sup>2,5</sup>, Steven M Wolinsky<sup>4</sup>, Andrew Kasarskis<sup>2</sup> & Eva Harris<sup>1,\*</sup> 

- We performed whole-blood RNA-seq, 37-plex **mass cytometry** of peripheral blood mononuclear cells (**PBMCs**), and serum cytokine measurements of **acute- and convalescent-phase** samples obtained from **42 children** naturally infected with CHIKV. Semi-supervised classification and clustering of single-cell events into **57 sub-communities** of canonical leukocyte phenotypes revealed a **monocyte-driven response** to acute infection, with the greatest expansions in “**intermediate**” **CD14++CD16+ monocytes** and an **activated subpopulation of CD14+ monocytes**. **Increases in acute-phase CHIKV** envelope protein E2 expression were highest for **monocytes and dendritic cells**.

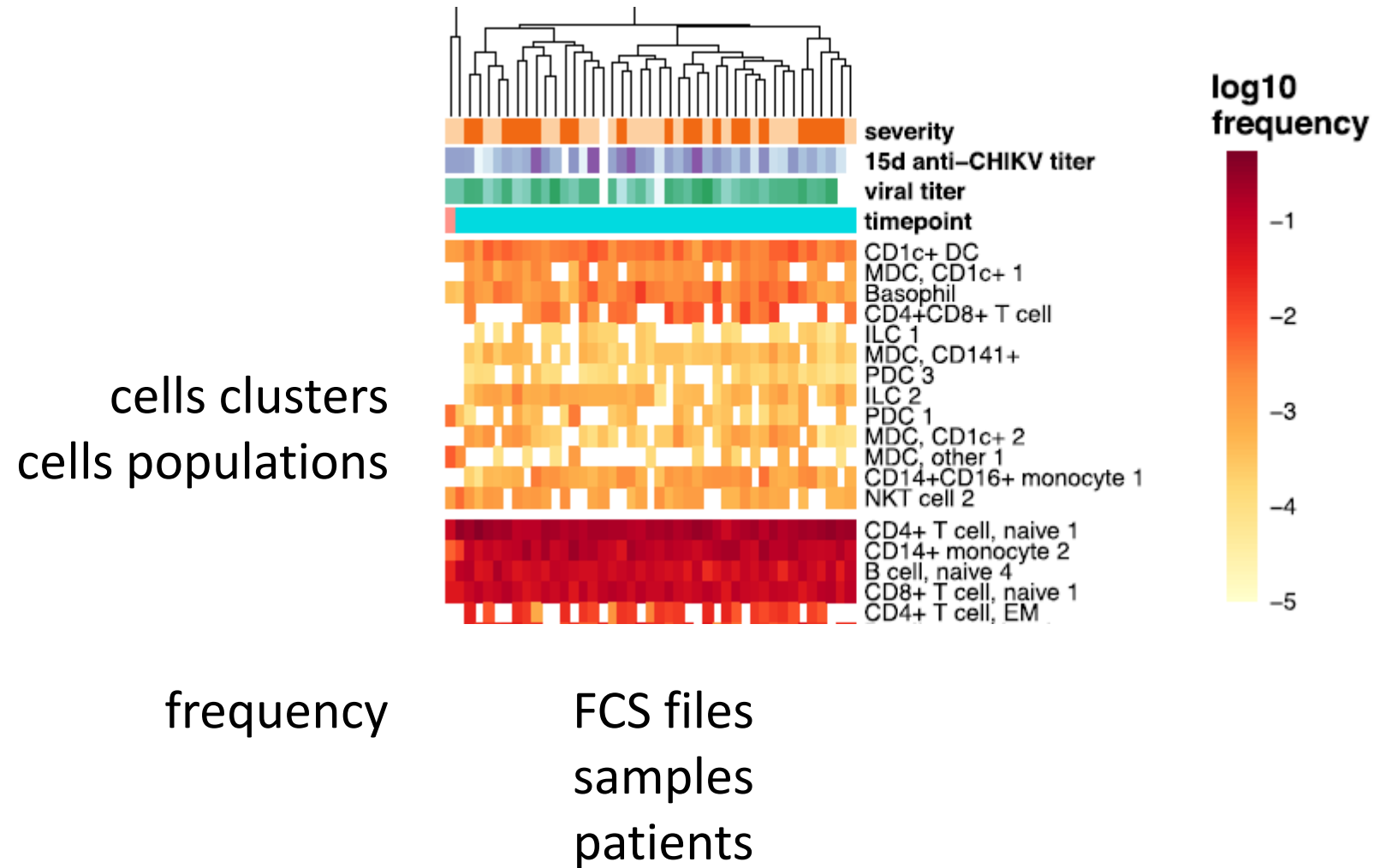
# Dataset - Example



# Matrix of data

- Different elements in rows and columns: clusters, markers, patients
- Many contents: count, percent, MFI
- Each content has a large dynamic range
- The dynamic range is larger than the information we look for

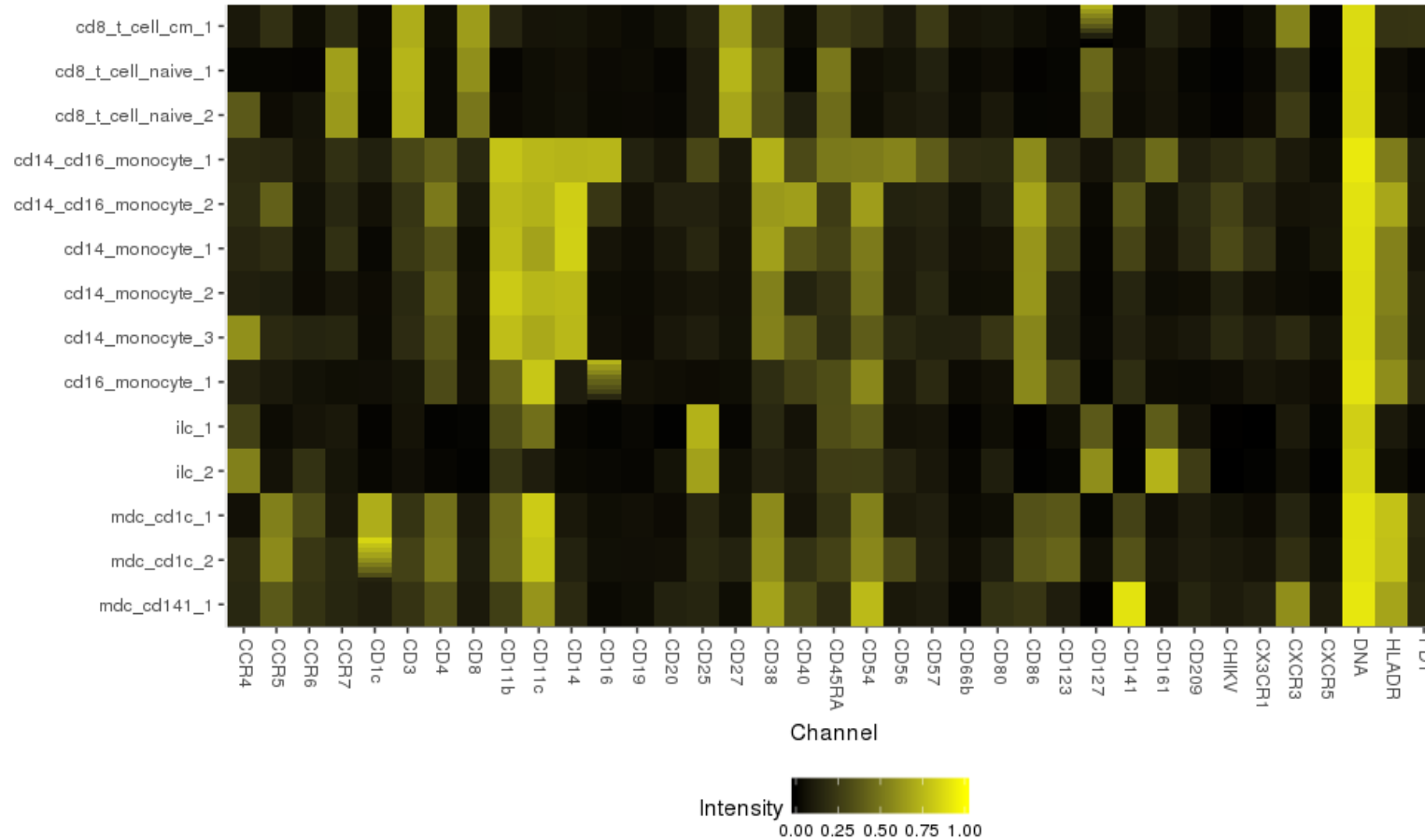
# Heatmap of frequency





# Heatmap of MFI

FCS markers  
proteins



cells clusters  
cells populations

intensity

# Matrix of data

- Different elements in rows and columns: clusters, markers, patients
- Many contents: count, percent, MFI
- Each content has a large dynamic range
  
- What is the information we are looking in the graphics?
- **What is the question?**
- How the content should be processed to become information?
- The content must be scaled in order to show the information
- The scaling depends on the content and the elements in rows and columns

# Methods for scaling

- min to max
  - $\text{min} = 0, \text{max} = 1$
- center
  - $\text{mean} = 0$  or  $\text{median} = 0$
- standardize
  - centering
  - divide by standard deviation
- apply to all data, but determine coefficients on all or a subset of data?

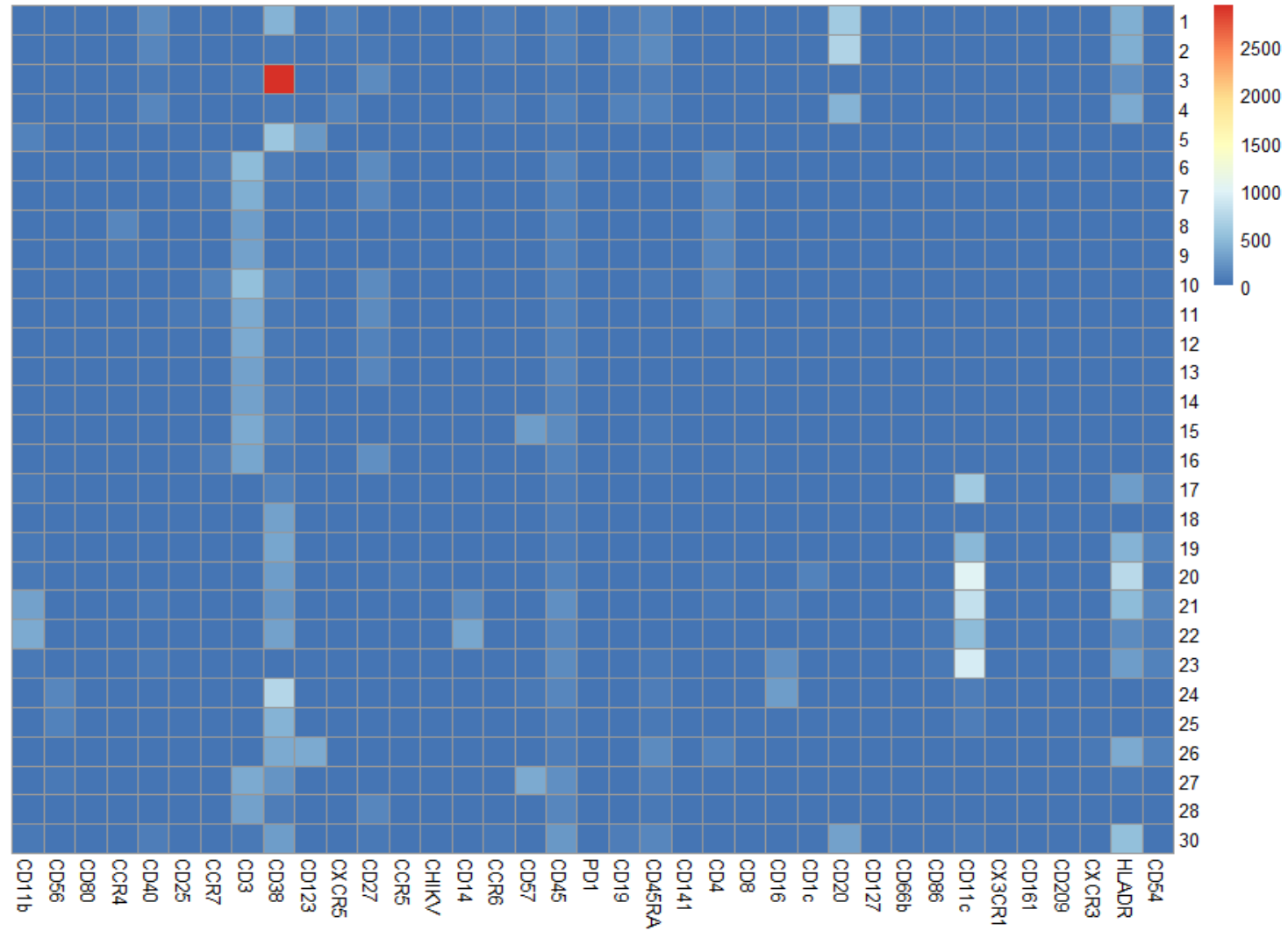
# Data to visualize = Matrix of MFI

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL
1	cluster_i	CD11b	CD56	CD80	CCR4	CD40	CD25	CCR7	CD3	CD38	CD123	CXCR5	CD27	CCR5	CHIKV	CD14	CCR6	CD57	CD45	PD1	CD19	CD45RA	CD141	CD4	CD8	CD16	CD1c	CD20	CD127	CD66b	CD86	CD11c	CX3CR1	CD161	CD209	CXCR3	HLADR	CD54
2	K1	20.64	5.3266	14861	10.153	156.48	7.2641	16.557	7.8572	426	6.0576	108.46	0.2031	2.2559	2.2927	3.7343	65.04	4.5073	103.93	0.3367	84.279	123.92	3.213	1.7888	0.6821	18.039	8.8621	600.55	3.2587	3.396	2.5884	4.6494	0.0535	2.5136	0.1234	1.1209	387.73	11.585
3	K2	23.981	6.2879	3.7006	11.475	138.22	7.9443	7.3281	6.5878	39.229	2.2805	51.6	56.046	3.9683	3.0324	4.951	66.599	5.6596	105.33	1.0379	109.13	164.79	4.8735	3.001	1.6452	21.683	24.581	681.7	4.1348	4.8948	4.014	22.714	0	2.7695	0.2089	4.8724	405.42	21.246
4	K3	7.7701	4.4041	1.4501	3.5945	29.57	23.338	17.675	37.04	2936.8	2.6209	0	151.96	0.6032	0	3.1067	1.0417	2.044	57.547	0.3839	31.809	79.242	0.8648	1.3245	0	1.7107	0.4553	4.4681	0	0	21.552	3.334	3.1216	6.8557	0.9792	5.3602	184.95	16.61
5	K4	19.778	4.7692	1.2052	8.6627	132.27	4.116	20.696	4.621	56.842	3.0394	88.724	0.5824	1.7315	2.4596	3.3021	72.283	3.948	92.339	0.0069	91.8	105.35	2.5665	1.4831	0.3668	13.814	8.4242	440.23	2.3983	3.3481	1.7123	5.4924	0	1.4312	0	1.4893	371.92	11.737
6	K5	96.655	0.8045	0	1.5568	1.325	19.328	3.6643	12.828	571.44	253.6	1.9037	0	3.4056	0	1.4858	0.0563	0.4706	33.321	0	0	4.127	0	0	0	0	0	0	0	0	0	30.79	0	0.3746	0	0	1.6339	8.5792
7	K6	0	5.7129	0	5.3518	1.1559	6.0124	62.26	482.35	87.811	0	0	171.56	0	0	0.7264	0.5924	5.181	135.3	0.6562	0.2501	16.723	1.0459	147.73	9.399	0.7145	0	0.4232	20.183	0	0	0.0514	0	2.1642	0	0.0484	1.7031	0.4854
8	K7	0	4.4829	0	13.981	0.7861	4.0033	40.405	389.65	50.159	0	0	135.17	0	0	0.1052	0.5782	3.1044	94.564	1.5586	0	4.8535	0.3589	129.32	2.0664	0.3464	0	0.1134	17.32	0	0	0	0	2.1263	0	0.6617	1.1962	0
9	K8	0.2649	5.6312	1.0098	131.86	3.0103	11.05	3.8045	269.66	5.0243	0	0	0.2149	3.6419	0	0.1826	2.1832	3.6906	94.843	5.1466	0	2.715	0.4085	138.27	2.1449	0.4925	0	0.505	26.05	0	0	0	0	2.1614	0	0.8755	2.9536	4.9932
10	K9	1.103	7.5616	0.085	36.982	2.4876	7.8136	4.8136	305.59	14.978	0	0	1.9926	2.3941	0	1.6449	1.4983	6.1927	94.924	4.708	0.5509	20.798	0.6483	121.17	2.1874	1.9266	0.3208	0.6179	18.012	0	0	0.3741	0	4.0294	0	1.182	5.131	4.6492
11	K10	0	4.8961	0	0.4984	0.0236	4.1797	89.277	520.01	110.19	0	0	169.38	0	0	0.8001	0	3.3682	95.466	0	0.2233	31.809	0.9149	128.36	2.1175	0.7018	0	0	18.327	0	0	0.0643	0	1.338	0	0	0.8664	0
12	K11	0	5.1009	0	11.665	1.2536	41.047	30.21	380.67	46.686	0	0	150.96	0.2421	0	0.6141	0.4184	3.4444	89.897	1.4097	0	11.222	0.341	114.71	1.9578	0.3657	0	0	1.8429	0	0	0	0	1.4986	0	0	2.2146	0.7448
13	K12	1.4749	3.1938	0	0.4414	0.2309	3.696	10.436	370.32	44.515	0	0	90.194	3.0925	0	1.32	0.3636	5.7177	111.33	0.3007	0	18.835	0	0	0	0.969	0	0	7.7017	0	0	1.5144	0	8.904	0	6.1514	4.2652	3.6683
14	K13	0.8914	1.6804	0	0.7193	0	2.4401	7.3958	301.96	39.075	0	0	123.61	6.6436	0	0.6447	0.1777	5.9175	122.9	4.1285	0	3.351	0	0	30.425	0.2637	0	0.7089	9.4407	0	0	0.7427	0	2.7404	0	12.512	5.6058	4.4999
15	K14	1.9772	5.0121	0	0.8447	0.0836	4.2725	3.1276	306.51	62.611	0	0	1.6778	11.722	0	2.1527	0.9919	14.758	112.7	4.382	0	3.2521	0	0	26.189	1.6843	0	0.0564	1.0847	0.785	0	0.9945	0	4.9928	0	3.3212	13.436	4.8061
16	K15	5.2168	19.618	0.1839	0.4968	0.537	13.44	3.0661	356.89	101.14	3.4204	0	2.6882	3.7888	0	10.105	0.7746	265.39	148.68	1.1549	2.2655	43.008	0	1.541	26.662	7.0244	1.8353	0.7793	1.1442	12.19	0	3.8676	1.058	9.2807	0	1.0146	21.016	5.2546
17	K16	0	1.6608	0	0	0	2.977	70.067	339.47	47.981	0	0	180.43	0	0	1.2631	0	4.4466	110.59	0	0	46.272	0	0	30.176	0.6328	0	0	16.004	0	0	1.067	0	1.51	0	6.0684	1.9466	1.225
18	K17	31.047	3.1074	0	3.989	21.532	1.59	1.7497	7.5903	97.552	18.591	0.003	0.6601	3.5318	0	8.3728	1.1328	3.2357	74.859	2.6652	0.9873	19.593	5.8228	13.25	0.2717	8.361	1.0735	1.5488	0	0	20.382	606.4	1.01	0.6418	0	1.3781	275.15	59.32
19	K18	11.592	5.9401	0	2.9047	4.29	5.051	2.8679	8.7276	297.99	0	0	0.1185	0.0838	0	1.2629	0.0615	3.1041	60.969	0	0.2436	22.347	0	0.3716	0	7.4128	0	0	0	0	0	14.834	0.712	9.2202	0	0.3041	3.7455	5.1342
20	K19	29.92	3.2282	1.123	5.5788	18.267	2.4054	3.0615	7.7214	336.59	15.077	1.5831	0.1712	15.262	0.9062	9.2538	2.2271	2.974	81.621	3.9732	0.3086	16.83	21.532	18.944	0.4999	2.4499	1.2525	3.9372	0	0	12.324	468.75	1.671	0.6347	1.2033	4.3182	438.27	111.41
21	K20	38.836	3.5987	0	0.9955	2.3307	4.0923	2.1229	12.368	284.05	15.896	0	0.9322	30.257	0.2226	3.6876	7.687	3.8166	101.42	4.9732	0.4565	11.665	5.7048	34.014	1.3199	1.0288	91.73	0.9203	0	0	11.011	1013.8	0.0179	0.5003	0.7207	3.113	740.88	54.898
22	K21	301.96	4.538	1.0764	8.1465	57.087	2.4652	2.4352	8.9226	217.68	24.409	1.0068	1.1274	7.7818	2.4372	173.88	1.0335	9.5237	191.87	4.1096	1.6966	19.116	8.7956	32.028	1.7915	88.065	1.5536	3.8975	0	0.8154	47.14	816.28	2.4608	1.2649	0.9717	1.255	483.77	145.81
23	K22	356.24	3.1708	0.0405	6.6226	17.704	2.4669	4.3285	9.3024	302.09	10.957	1.2104	0.5318	3.1566	4.0393	327.72	0.5532	5.261	134.36	2.2697	0.1459	12.149	4.5854	23.546	0.0219	1.855	0	2.0975	0	0	31.69	498.15	3.0686	0.7013	0.7217	0.164	168.48	70.028
24	K23	42.418	3.0072	0.2017	2.9354	31.381	0.9108	0.3469	4.2894	26.205	29.01	0.0305	0.1494	2.3182	0	10.112	0.3827	8.253	169.34	4.178	1.6177	30.702	3.1645	16.798	1.5769	179.29	1.1505	1.9426	0	0.1981	37.407	926.95	1.3993	0.2987	0	0.4789	265.91	106.3
25	K24	20.601	120.1	2.2096	1.2567	8.5138	11.547	3.2338	11.992	711.81	1.93	0	0.8022	0.1092	0	5.9754	0.1544	32.608	122.44	0	3.4121	62.341	1.4427	2.1757	1.9968	278.29	2.9114	0.395	0.1806	2.7893	0	86.707	1.8403	24.744	0.082	0	7.7847	8.5648
26	K25	16.043	100.91	2.7413	3.4643	2.3665	6.9362	3.1429	7.8232	419.99	0.1601	0	1.1718	1.6497	0	2.923	0.1951	6.3873	84.441	0	0	29.647	0	0	7.3004	0	0	0	0	0.0009	0	63.526	0.5668	17.078	0	1.4584	6.5403	7.8762
27	K26	12.667	4.9847	0	1.2674	4.4191	3.4543	7.5963	7.6099	377.82	372.22	3.2141	0	23.313	0	2.0377	0.8288	1.8237	63.507	0.233	2.0157	155.35	8.0728	93.893	3.3742	1.4403	0.6458	0	0	0.005	0.152	4.0803	0	1.4579	0.6947	43.2	353.63	97.587
28	K27	13.481	53.605	1.8708	3.6962	4.7178	20.243	6.6171	356.72	209.05	7.5432	0.0216	10.021	4.3629	0	22.534	3.6774	357.98	190.08	1.5423	6.7789	72.37	1.9257	7.3038	9.8684	16.896	4.9433	2.5622	5.822	17.252	0.8035	16.435	1.2797	17.29	0.4327	2.4267	45.128	10.319
29	K28	0.8737	5.855	0	3.5187	1.3289	6.6138	23.102	301.14	87.052	0.3478	0	130.29	1.3643	0	3.8316	0.9664	10.819	143.65	1.708	1.6596	35.758	0.0317	18.581	10.655	3.4538	0.7009	1.2709	7.7155	0.4752	0	3.0414	0.4256	5.2085	0	7.6483	8.5216	2.7553
30	K30	43.232	13.337	4.7273	14.922	78.358	12.284	25.115	57.368	292.87	8.2823	17.744	51.253	6.8674	3.2616	11.486	33.749	16.523	261.37	2.7664	77.509	125.41	6.9998	16.679	6.5089	27.732	11.371	303.16	8.6027	6.8994	7.8429	48.984	2.4408	7.4007	2.1489	12.291	515.86	26.645

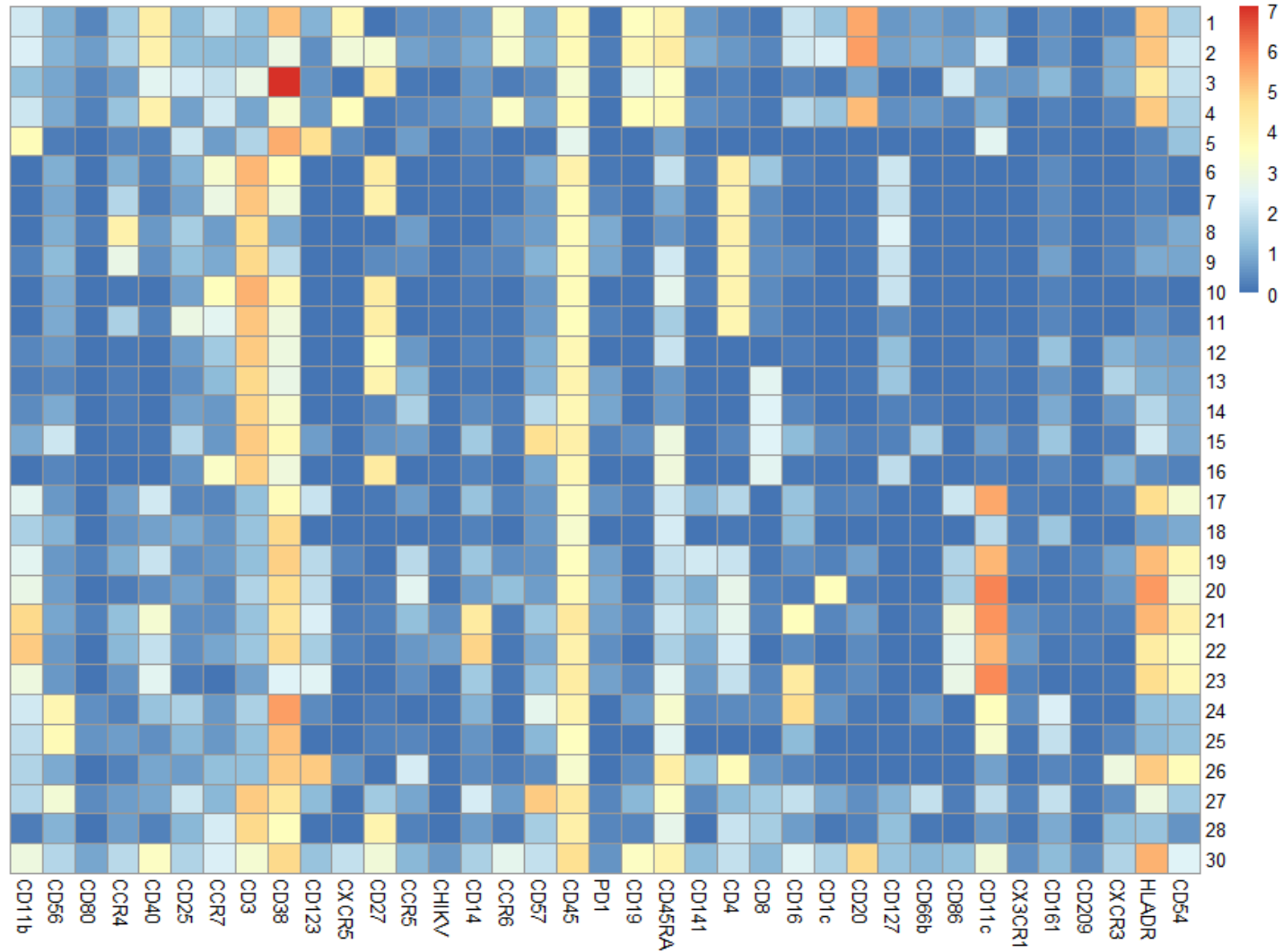
# Question = Annotate clusters

- MFI is based on transformed intensity (usually asinh)
  - MFI of untransformed intensities can be transformed at this stage
- Goal: determine the cell population in a cluster from its expression profile (aka MFI)
  - Determine which marker is high or low in a cluster
- Hypothesis: the expression of a marker ranges from low to high in the set of clusters
- Scaling: min max each marker
$$\min(\text{marker}_i\text{scaled}) = 0, \max(\text{marker}_i\text{scaled}) = 1$$
$$\text{marker}_i\text{scaled} = \frac{(\text{marker}_i - \min(\text{marker}_i))}{(\max(\text{marker}_i) - \min(\text{marker}_i))}$$

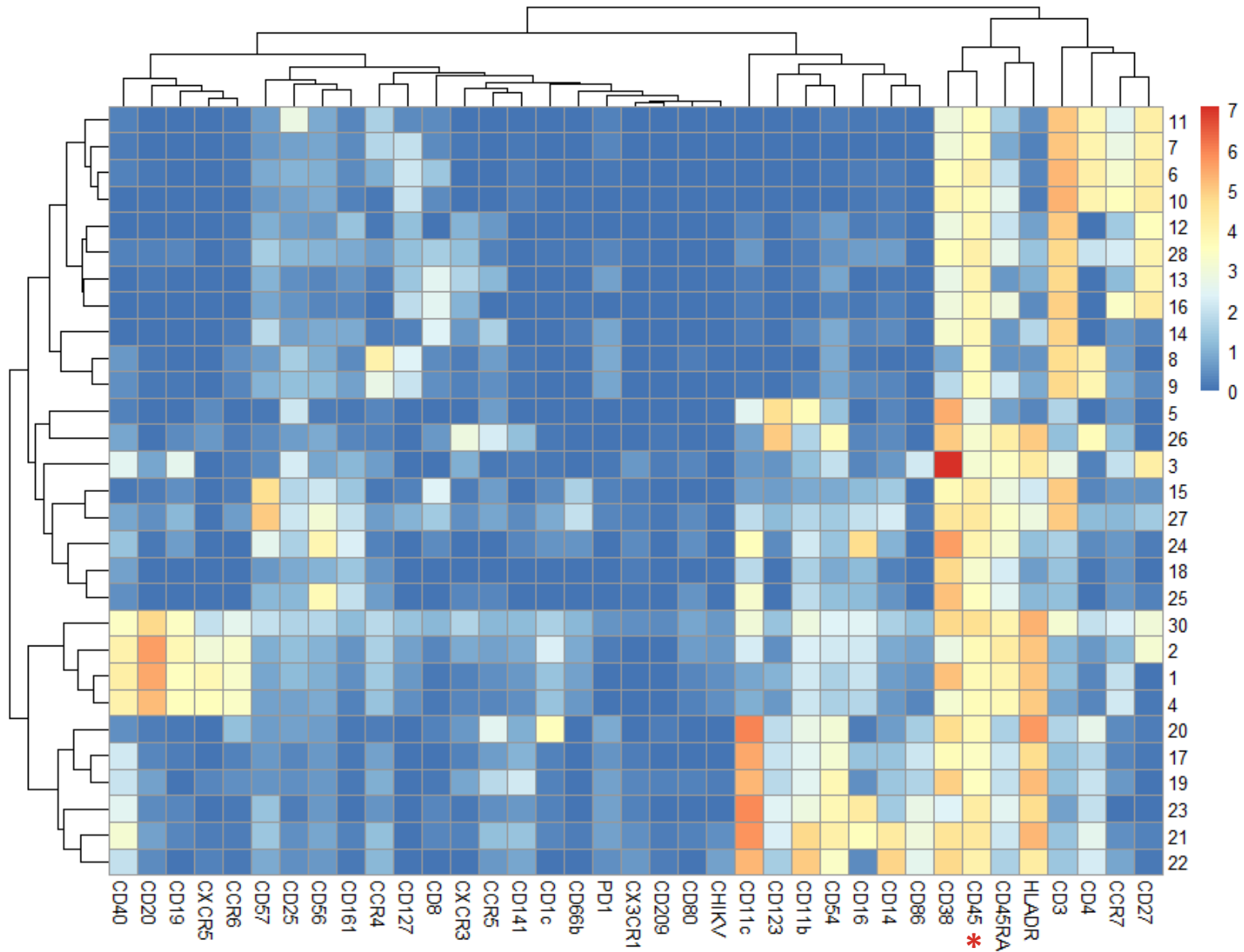
# Heatmap of MFI



# Transformed MFI

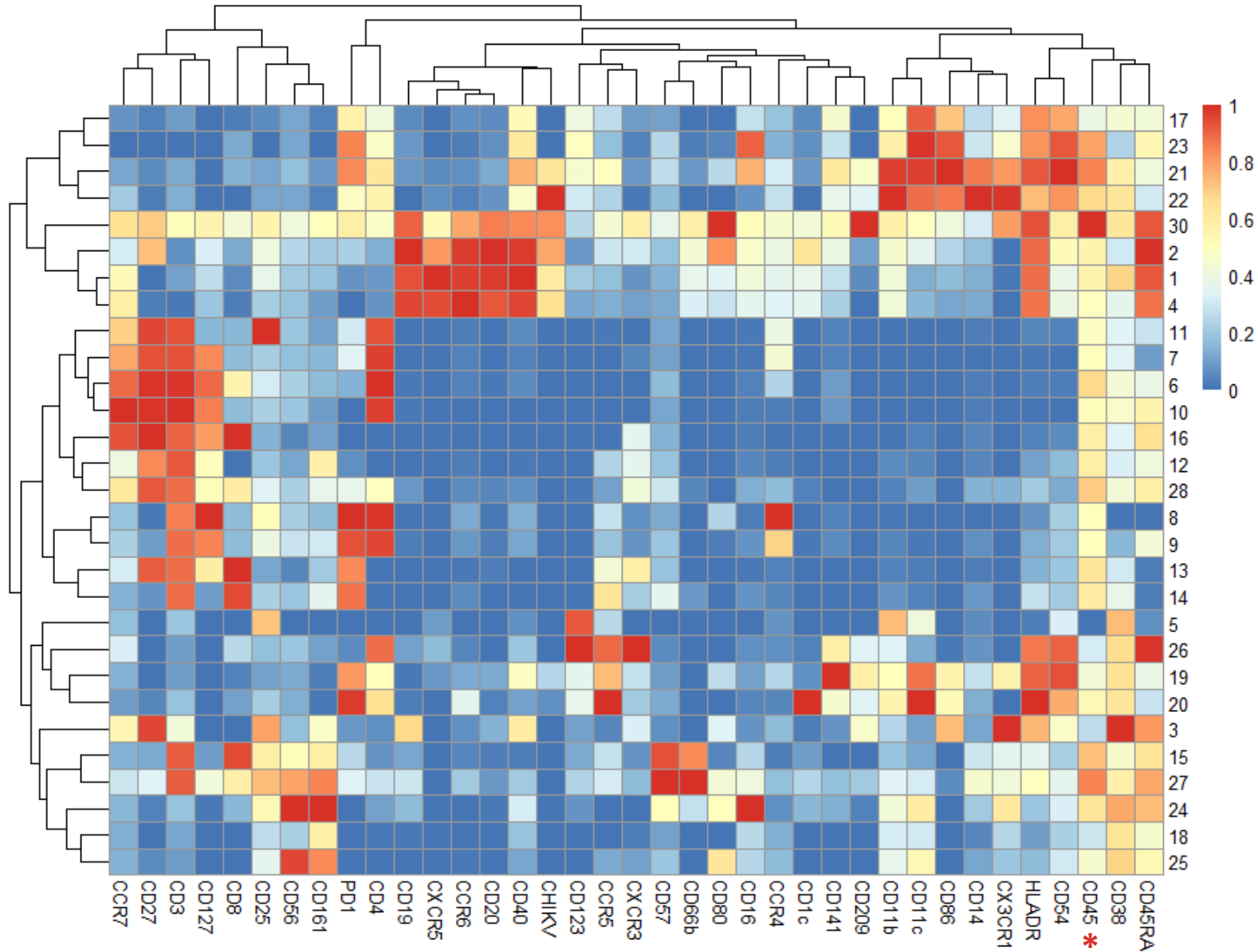


# Transformed MFI with hierar. clustering

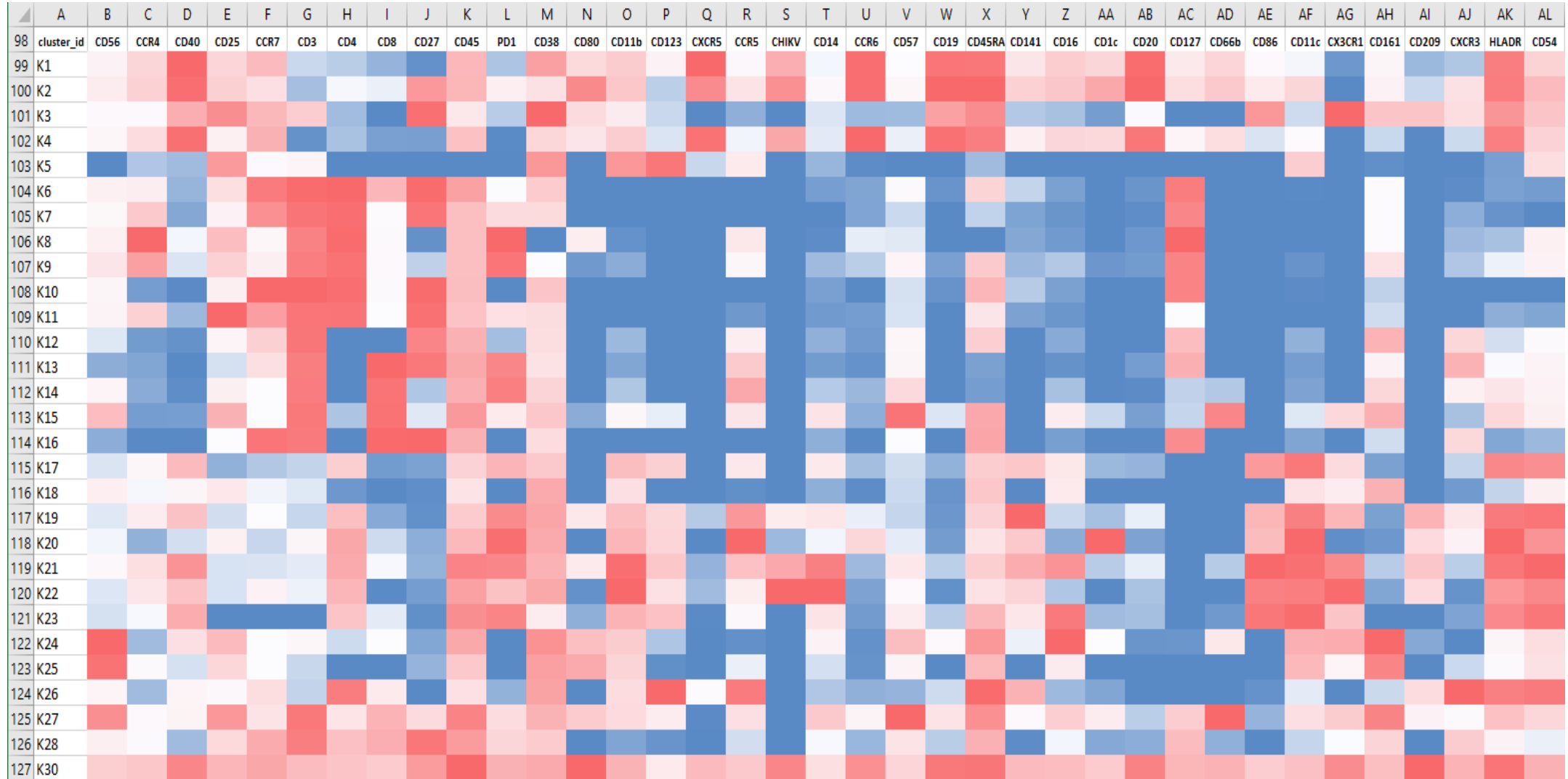




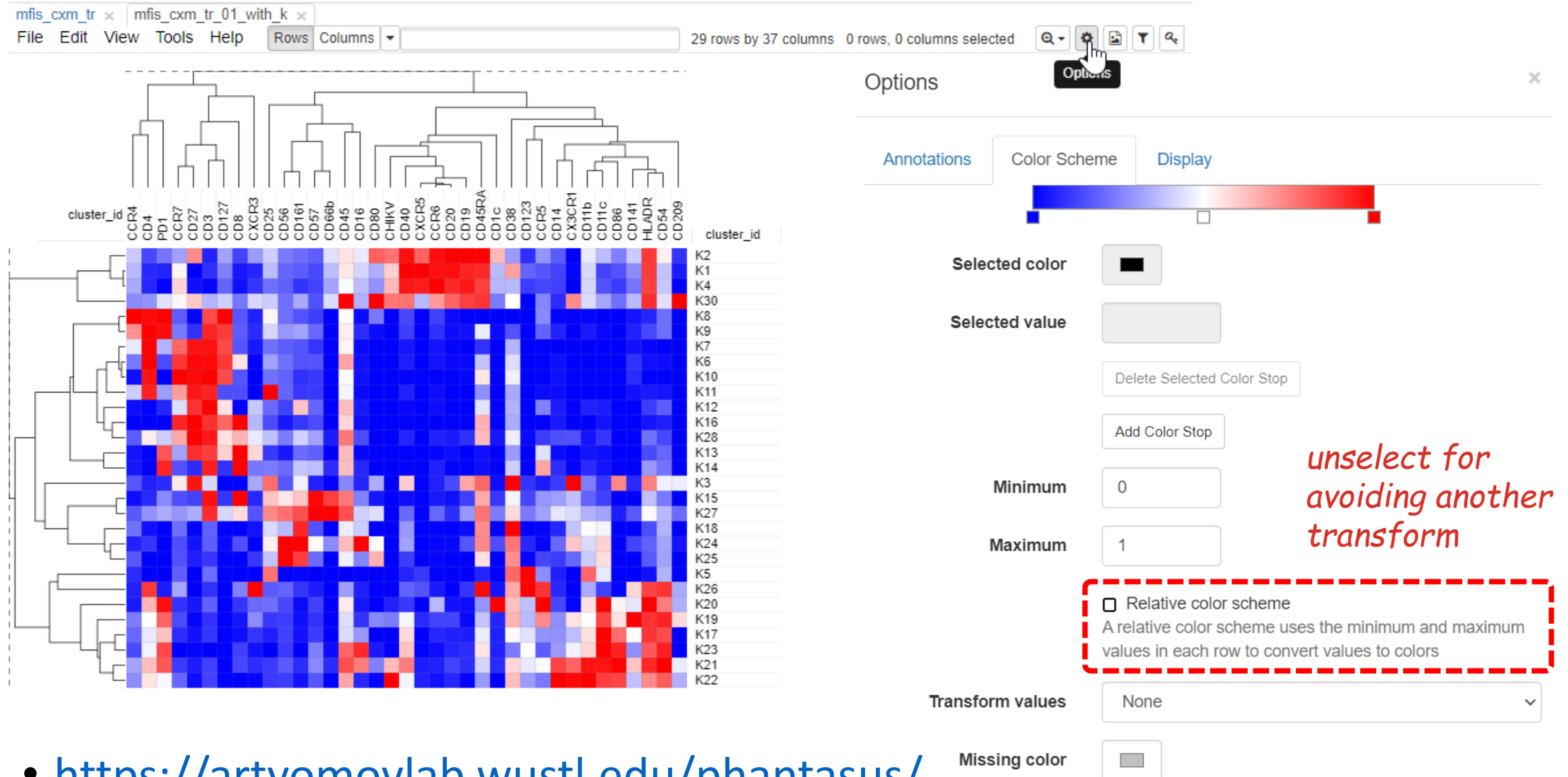
# Scaled transformed MFI with hcl.



# Let's do it with Excel!



# Let's view it with Phantasus



- <https://artyomovlab.wustl.edu/phantasus/>

# Data to visualize = Matrix of abundance

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL		
1	cluster_id	cluster_name	1758	161758	61760	61760	61773	61785	61790	61790	61793	61793	61794	61800	61800	61802	61822	61822	61823	61823	61824	61824	61857	61857	61828	61828	61829	61829	61838	61838	61839	61839	61839	61839	61839	61839	61839	61839	61839	61839
2	K1	B Cell (CD	1166	5478	9089	17594	1626	5907	1286	11142	2551	4729	934	12315	403	5129	2904	6603	4611	13501	1048	2977	199	5945	2592	15839	2530	4832	2406	5843	4670	31867	2492	7861	3993	5478	1686	119		
3	K2	B Cell (Me	187	3225	685	2867	296	4498	608	8120	813	3156	381	3181	84	2357	1394	4731	1376	4382	125	1360	237	2672	778	9155	243	3322	1818	13596	3238	9130	1156	6384	765	3883	2669	118		
4	K3	B Cell (Pla	287	1102	411	543	85	1560	890	853	494	436	153	978	118	434	333	661	174	866	129	273	289	276	628	876	49	269	2341	1656	1010	482	967	2084	242	278	600	17		
5	K4	B Cell_una	2184	19361	8841	20565	3530	7658	2531	17003	2930	6143	1031	5804	912	16404	9928	13396	7959	8540	699	2089	532	8795	2397	18981	924	11726	7123	17249	14074	14212	3316	10389	5249	9088	6014	114		
6	K5	Basophil	39	99	260	717	13	2334	185	322	0	0	0	548	0	425	0	0	708	2039	17	200	0	135	0	3458	44	450	0	0	11	731	0	182	160	361	0	4		
7	K6	CD4+ CD8-	0	22718	4306	22357	1324	15555	7155	26491	6379	33177	11112	37394	3091	32680	12473	36528	16463	31171	1946	10405	0	15019	0	22620	0	708	0	0	0	23847	0	0	0	0	0	0	7	
8	K7	CD4+ T Ce	5153	9110	6484	8315	2680	14915	2437	15099	9530	19052	4605	15702	1271	3471	6703	15501	12022	9815	3317	6033	1588	22731	5952	14550	2535	10727	6777	29794	9545	3309	5702	15164	10148	10324	12360	231		
9	K8	CD4+ T Ce	872	770	320	607	369	1871	1553	3249	3427	2257	582	1088	539	421	1358	1506	1219	829	676	694	332	1010	715	770	1510	1972	620	1086	534	138	1264	2794	2445	2915	2723	35		
10	K9	CD4+ T Ce	128	163	174	641	80	305	447	594	222	67	93	214	42	424	170	174	281	326	80	85	75	189	80	207	439	411	243	415	358	377	157	318	312	428	179	5		
11	K10	CD4+ T Ce	5977	17655	9211	15457	1809	20934	3217	12460	3248	9455	7107	23023	2177	30077	8399	25936	5482	16296	2559	6665	1412	28114	2306	15023	1752	18713	6072	38773	14486	27766	5094	20615	11304	18409	8209	330		
12	K11	CD4+ T Ce	1618	2729	1824	3592	541	4681	926	3694	2154	4703	1078	3441	434	3637	1609	5180	2844	4820	765	2102	521	7279	1525	5931	814	3530	1997	10305	3411	5018	1823	6868	3188	5605	2934	7		
13	K12	CD4- CD8-	3158	12606	4044	10848	2997	19347	3638	10474	5682	11616	3848	12581	1367	9610	4750	10347	7929	13861	1535	5813	3270	23885	6314	26402	3304	12649	5479	23470	6223	11035	4285	16144	8733	15527	3736	125		
14	K13	CD8+ T Ce	338	1243	429	2141	342	3327	634	2869	813	3326	1118	4419	644	1546	635	2611	1665	2967	709	1145	382	4297	1447	10635	257	2464	1992	7670	470	1199	462	1940	1018	1826	1168	33		
15	K14	CD8+ T Ce	178	602	67	189	112	580	231	549	190	411	174	342	222	320	90	197	316	468	211	130	228	669	517	1270	430	423	329	945	103	105	84	195	140	227	197	4		
16	K15	CD8+ T Ce	552	1585	105	312	733	2400	281	1549	2302	2069	454	1847	291	1739	321	607	2993	3143	104	418	344	4614	767	2669	2097	4416	139	808	436	528	477	769	443	574	1090	43		
17	K16	CD8+ T Ce	2079	11915	2799	6565	2468	24243	1945	14198	4777	31314	10556	45748	893	18087	7549	30201	16450	25147	1394	8754	621	27688	3008	27110	2297	18683	1266	13029	4310	17230	4024	15604	11280	17166	7395	250		
18	K17	CM- HLAD	1342	3285	1527	2032	109	1597	864	1416	1452	2541	1679	1380	157	868	261	535	1926	1386	425	234	843	445	3174	786	116	291	1194	1782	1643	384	671	4763	884	681	1041	35		
19	K18	CM- unas	488	6255	2687	11096	898	1945	6806	4565	1163	3392	793	2335	464	2473	866	2724	910	4142	419	2872	341	2165	996	2129	2129	3065	3956	4575	1184	2410	1008	2200	1996	1949	1114	34		
20	K19	Dendritic	177	1126	89	504	77	267	51	171	260	189	756	176	66	106	96	296	414	156	166	50	0	55	557	120	30	57	89	137	136	61	74	1579	214	83	293	4		
21	K20	Dendritic	0	530	113	372	0	1076	55	1399	0	2013	0	1276	0	540	119	2719	0	1977	0	390	0	770	0	1407	19	392	72	999	50	611	0	514	0	842	0	13		
22	K21	Monocyte	1372	1544	2760	181	916	247	3142	899	3666	554	2198	933	1796	245	1333	376	3490	625	521	174	1092	212	3468	1222	2741	269	3084	17606	1748	3089	2935	1810	2571	771	2761	11		
23	K22	Monocyte	22329	59594	22444	17604	14331	36709	28184	48024	64320	33766	24242	59408	13234	13999	15210	31837	28353	43119	6180	8649	8860	17941	22760	39625	20172	6455	14676	23533	7541	12539	9734	19961	11094	19693	18526	295		
24	K23	Monocyte	706	909	339	1925	147	1057	56	2088	1461	1866	690	1365	211	708	312	633	2023	666	661	477	1197	746	5753	2269	21	110	398	2800	1285	618	241	848	273	1662	197	7		
25	K24	NK Cell (C	1592	6786	828	13438	5891	12305	12519	12036	14456	10900	8152	13634	1474	6904	8227	9610	16047	15814	2122	2667	2026	4728	4648	6488	1499	8967	1557	22335	3048	9066	9826	22268	11137	25143	5249	125		
26	K25	NK Cell (C	1786	6745	2083	6418	3077	10133	4214	13489	2416	2879	934	2070	709	2270	3718	2482	3478	14232	1340	556	310	1505	1526	1465	2792	1675	940	6961	1516	1831	4474	5631	5589	2898	1725	145		
27	K26	Plasmacyt	798	1568	1058	1353	970	933	1092	795	2908	1049	1122	847	390	236	2539	1446	1132	447	210	125	264	735	717	685	440	265	399	662	901	736	1378	2100	996	1030	1458	14		
28	K27	Root_una	1082	7188	1583	4268	3443	14678	4397	16253	7660	48164	4626	27174	2692	18450	4678	18082	5171	7233	1254	6147	1686	24563	4044	21357	3794	10824	27	924	3138	8792	405	6461	3482	5266	4048	145		
29	K28	T Cell_una	21	17	0	0	0	14	0	0	0	0	0	14	0	0	0	10	12	0	0	0	0	11	29	49	51	0	0	116	0	24	0	0	0	0	0	0		
30	K30	Debris	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		

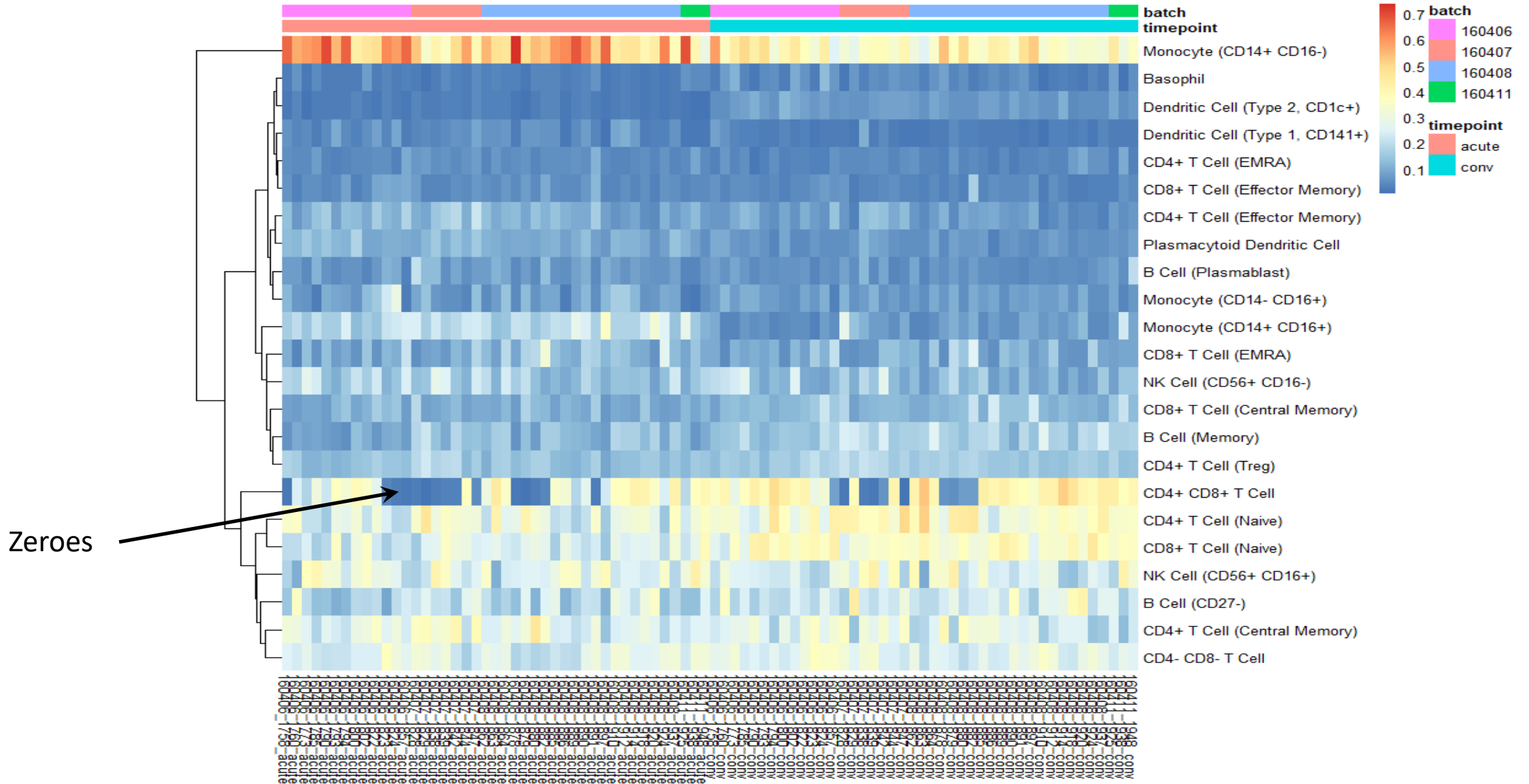
# Question = Find difference of frequency

- Frequency is based on transformed intensity (usually logarithm)
- Goal: determine the cell population/cluster whose frequency varies in relation to groups of patients
  - relates to the univariate statistical test to be performed
- Hypothesis: the frequency of a cluster ranges from low to high in the set of samples (FCS)
- Scaling: log2 transform and center to average of **each cluster i**  
$$\log2\_freq\_i = \log2(freq\_i)$$
$$\log2\_freq\_i\_scaled = \log2\_freq\_i - \text{mean}(\log2\_freq\_i)$$



# Transformed frequency

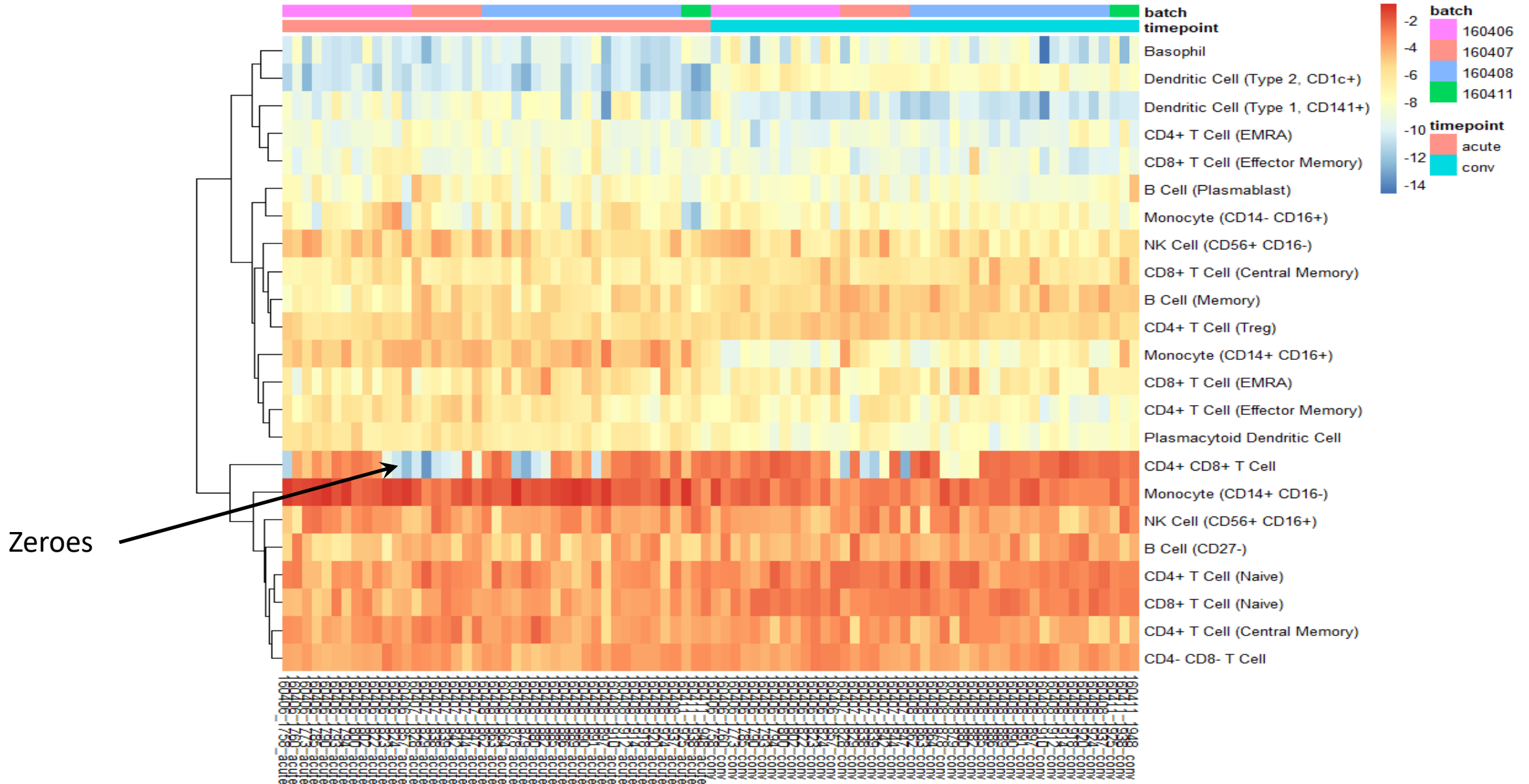
sqrt of frequency





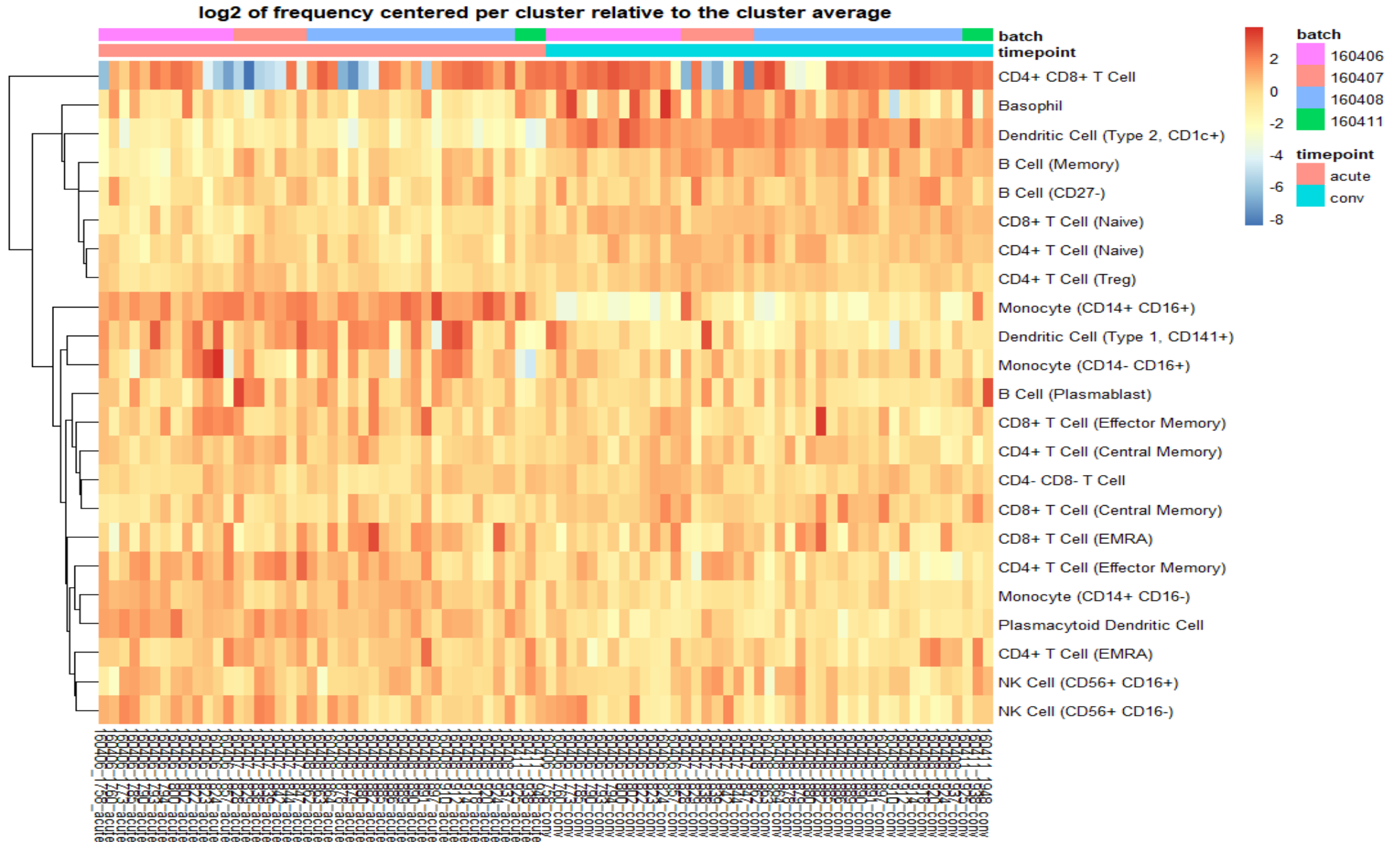
# Transformed frequency

log2 of frequency

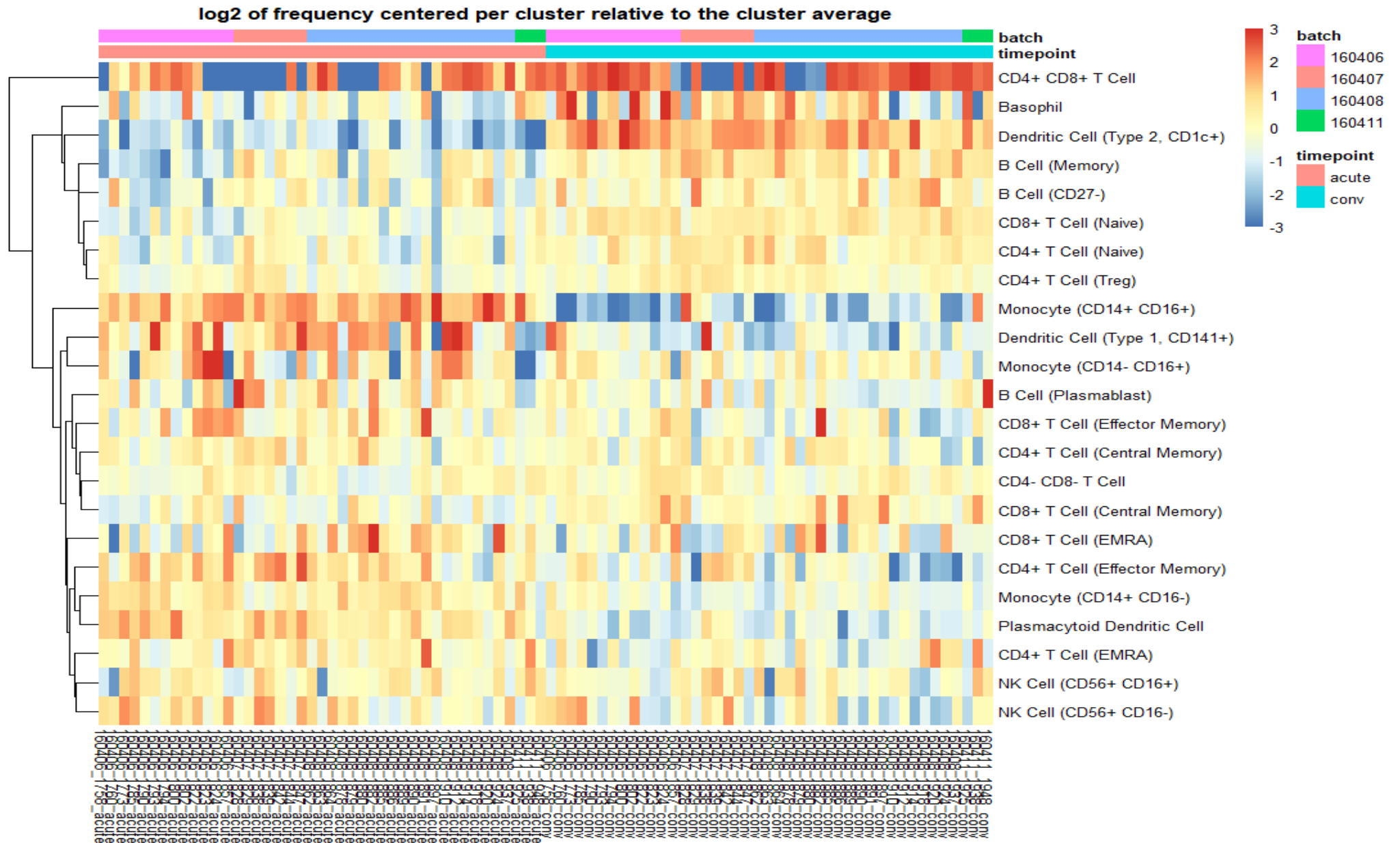




# Transformed then centered frequency



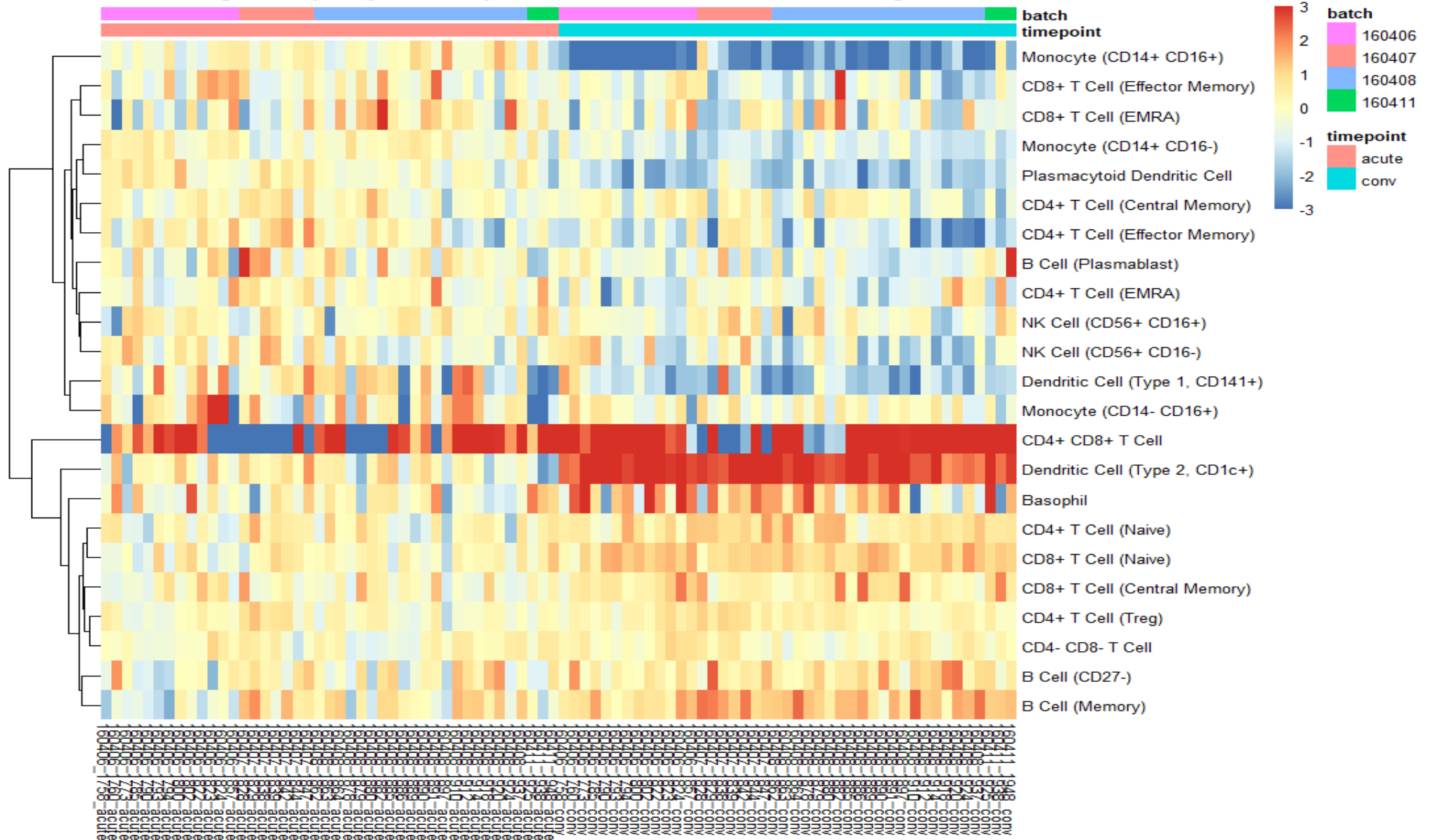
# Transformed then centered frequency



monocyte-driven response to acute infection, with the greatest expansions in “intermediate” CD14++CD16+ monocytes and an activated subpopulation of CD14+ monocytes

# Transformed then centered frequency

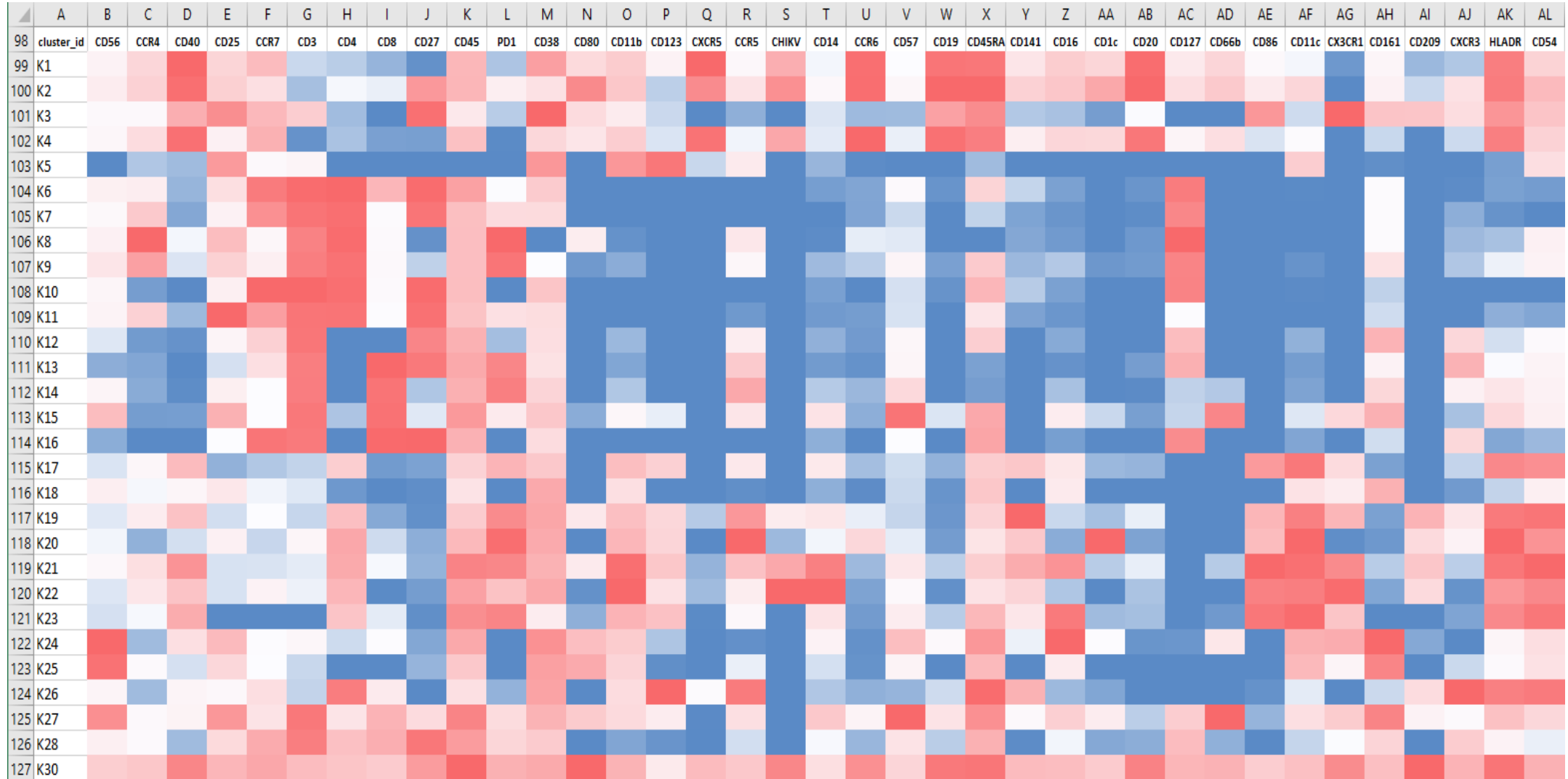
log2 of frequency centered per cluster relative to the "acute" average



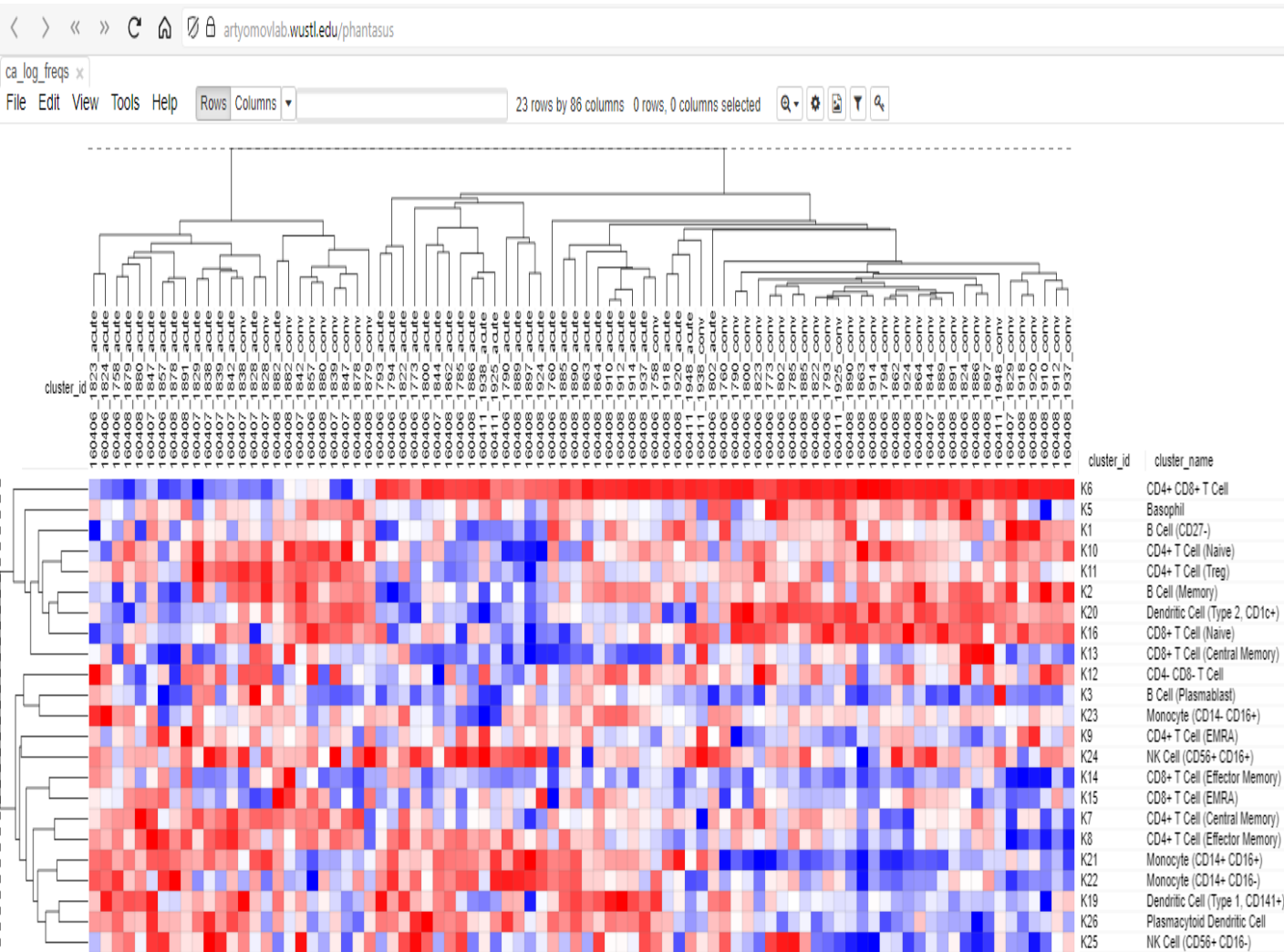
monocyte-driven response to acute infection, with the greatest expansions in "intermediate" CD14++CD16+ monocytes and an activated subpopulation of CD14+ monocytes



# Let's do it with Excel!



# Let's view it with Phantasia



0 rows, 0 columns selected

Options

Annotations Color Scheme Display

Selected color

Selected value

Delete Selected Color Stop

Add Color Stop

Minimum 0

Maximum 1

Transform values None

Missing color

Relative color scheme  
A relative color scheme uses the minimum and maximum values in each row to convert values to colors

*unselect for avoiding another transform*

- <https://artyomovlab.wustl.edu/phantasia/>

# Conclusion

- Heatmap is a classical way to visualize/explore the variation in frequency of clusters in relation to groups of patients
- MFI need to be transformed and scaled from 0 to 1
- Caveats of pre-gated markers
- Counts need to be transformed into frequencies
- Frequencies are usually log transformed and centered
  - centering is performed per cluster
  - centering on the overall mean or on the mean of a reference group
  - scaling can also be performed
- Caveats of zero counts

# Linear scale

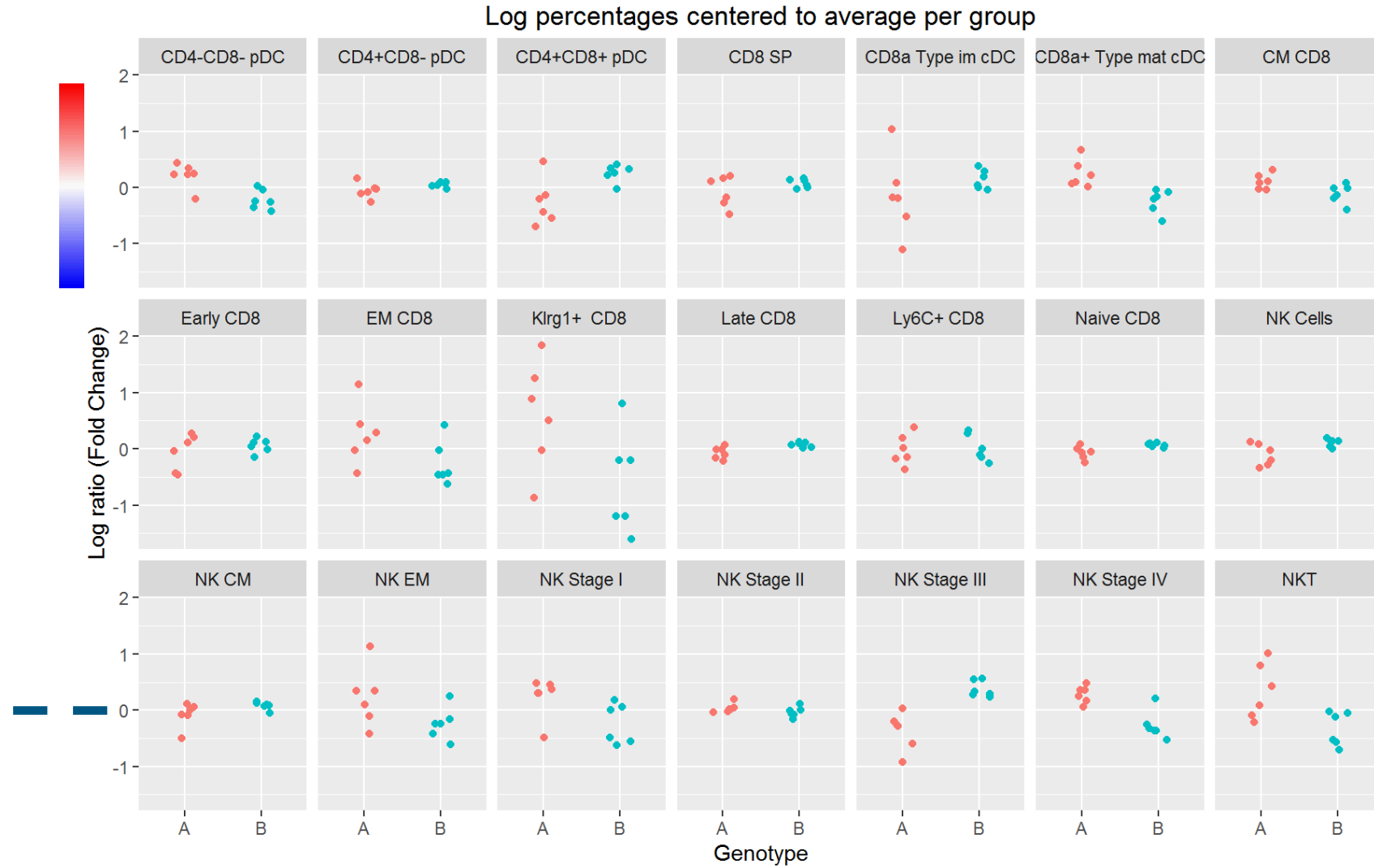




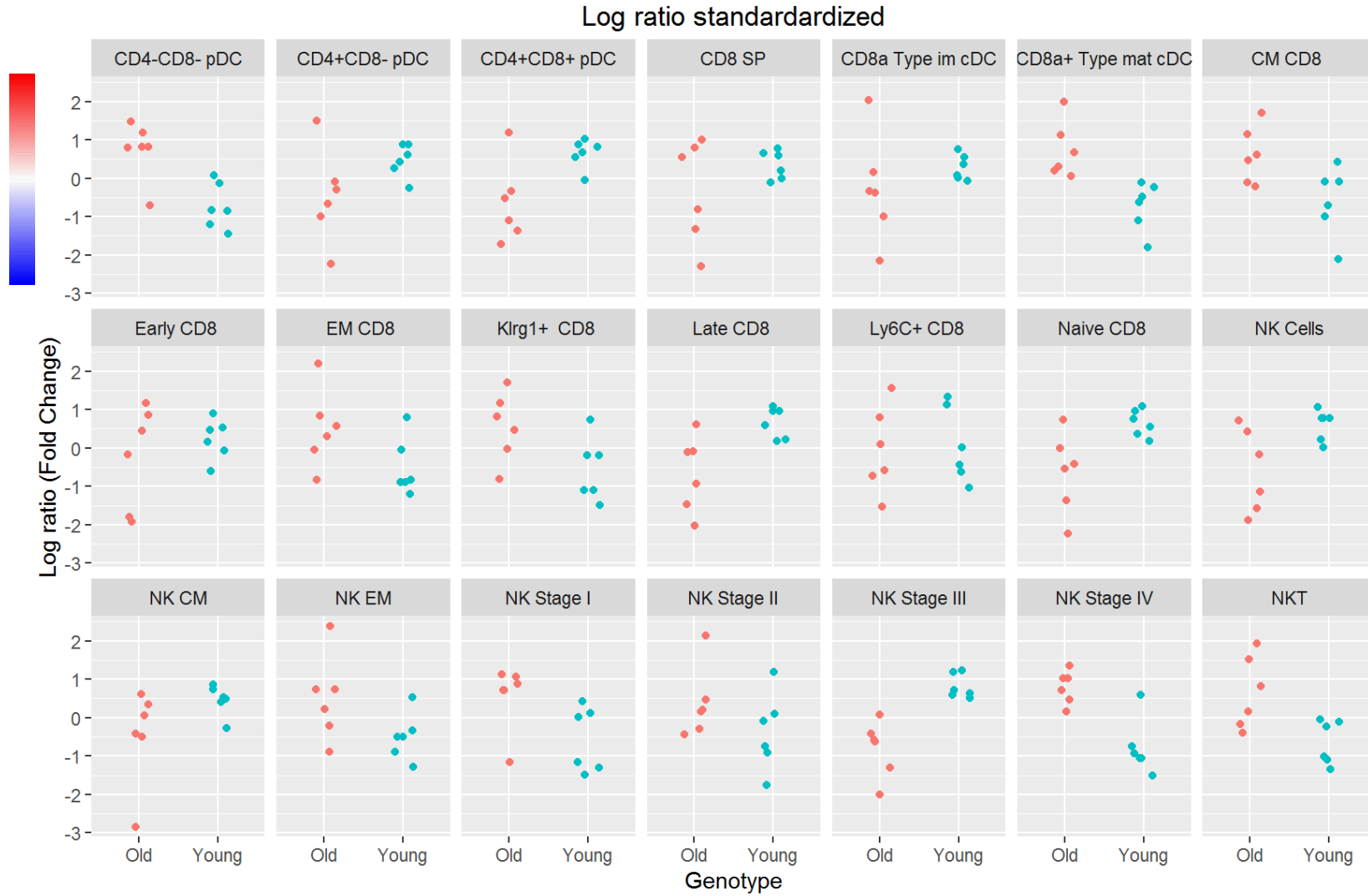
# Log scale



# Centered log values



# Standardize?



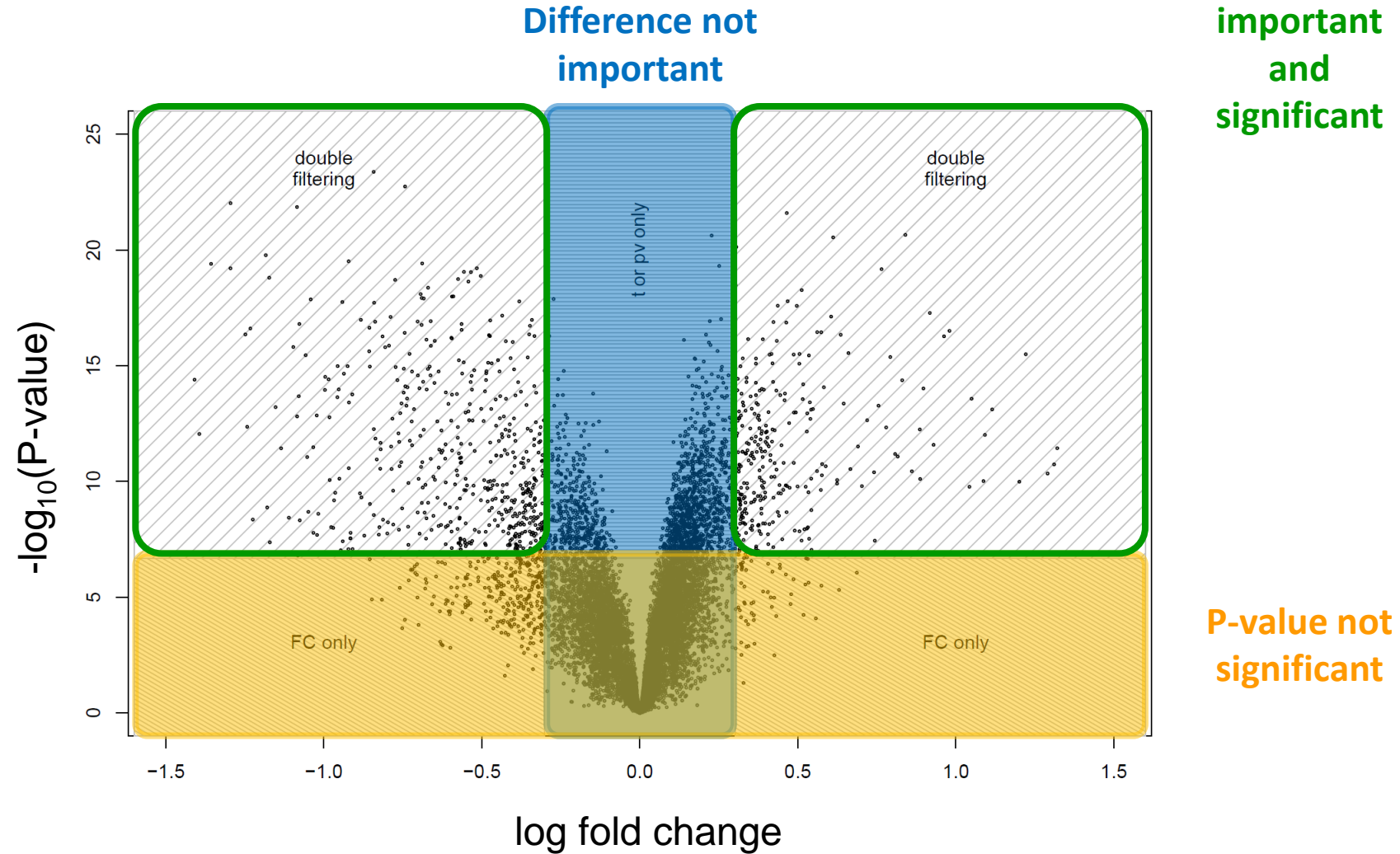
# Log2 properties

- log2 is proportional to log10
- 1 qRT-PCR cycle  $\sim$  x 2
- ratio  $\Rightarrow$  addition  
x 2  $\Rightarrow$  + 1
- log is symmetric: +100% = x 2  
- 50% = / 2
- log2 stabilizes the dispersion
- $\log_2(a / b) = \log_2(a) - \log_2(b)$

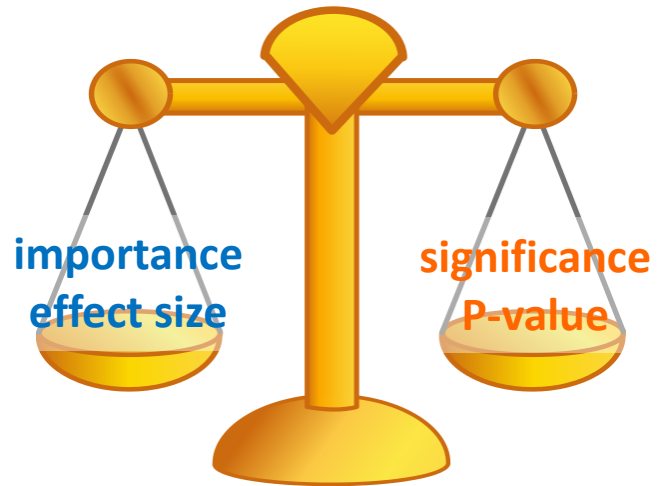
# What is "significance"?

- Statistical significance is not the same as practical importance.
- P-value does not tell whether the result is of a **practical importance**.
- Statistics does not tell us whether we are right. It tells us the chances of being wrong.
- Any particular threshold for declaring significance is arbitrary.

# FC vs P: Volcano plot

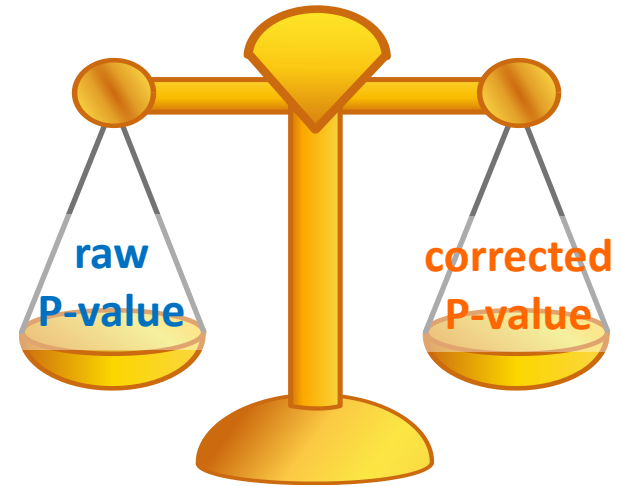


# Statistical trade-off



**Volcano Plot**

**report important  
and significant**



**False Discovery Rate**

**correct  
multiple tests**

# Grille de tests

Conditions d'application	variable gaussienne ou $N > 30$ ANOVA : variances homogènes	variable non gaussienne de préférence $N \geq 5$
Échantillons	test paramétrique	test non paramétrique
Un seul échantillon	Test t de Student à une norme	Test Wilcoxon à une norme
Deux échantillons indépendants	Test t de Student	Test de permutations Test de Mann - Whitney Test de Kruskal - Wallis
Plus de deux échantillons indépendants	ANOVA 1 facteur ANOVA 2 facteurs	Test de permutations Test de Kruskal - Wallis
Deux Séries appariées	Test t de Student apparié	Test du signe Test de Wilcoxon
Plus de deux séries appariées	ANOVA à mesures répétées	Test de Friedman
Mesure de l'association entre variables	Coefficient de corrélation de Pearson	Coefficient de corrélation par rangs de Spearman Coefficient de concordance de Kendall
Test de normalité	Test Shapiro - Wilk	Non Applicable



# La statistique, c'est PAS compliqué !

- Estimer

- Modéliser, expliquer
- Vérifier l'adéquation du modèle aux mesures

- Décider

- Risquer
- $\alpha$  : déclarer une différence à tort  
⇔ p-value
- $\beta$  : manquer une différence à tort  
⇔ puissance

		DECISION	
		<i>Fail to Reject <math>H_0</math></i>	<i>Reject <math>H_0</math></i>
ACTUAL	$H_0$ True	<b>Correct Decision</b> Confidence Interval = $1 - \alpha$	Type I Error <i>Producer Risk</i> $\alpha$ -Risk False Positive
	$H_a$ True	Type II Error <i>Consumer Risk</i> $\beta$ -Risk False Negative	<b>Correct Decision</b> <b>Power = <math>1 - \beta</math></b>

- La complexité vient de la multitude des applications de la statistique au cours de développement de cette science
- $H_0$ : Null Hypothesis    $H_a$ : Alternative Hypothesis

# La statistique, c'est PAS compliqué !

- Test d'hypothèse
  - assigner un modèle aux données
    - design (2, 3 ou 4 groupes, régression linéaire), nature du bruit, effets contrôlés
  - vérifier les conditions d'application a priori ou a posteriori l'adéquation du modèle aux mesures (résidus)
  - calculer la valeur d'un indice statistique adéquat
  - placer cette valeur dans la distribution de cet indice sous l'hypothèse nulle
  - mesurer la proportion d'indices plus grand : p-value

# La statistique, c'est PAS compliqué !

- P-value
  - garde-fou qui écarte l'aléa
  - autorise l'interprétation
- Interpréter, c'est conclure scientifiquement sur
  - l'importance de l'effet,
  - pas sur l'importance de la p-value

To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of.

[Ronald Fisher 1938](#)

# Take home messages

- P-value = False Positive rate
- Power = 1 - False Negative Rate
- P-value does not tell us whether we are right. It tells us the chances of being wrong.
- P-value does not tell whether the result is of a practical importance.
- Report difference and dispersion, or use confidence intervals.

# Analysis process

This is what you learn in school & textbooks

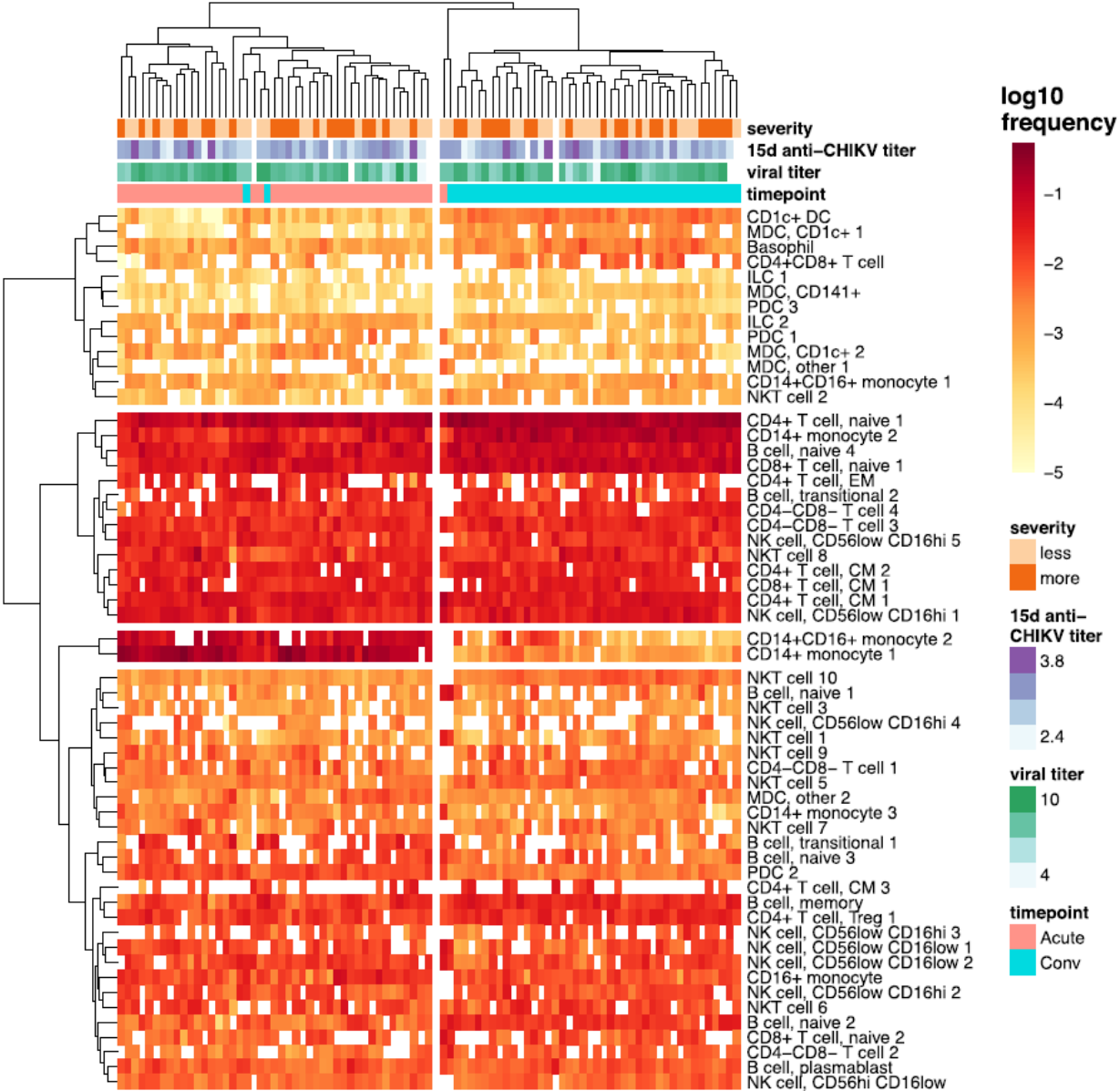
```
# (ideal) data analysis process
raw_data = GET(data)
proc_data = PROCESS(raw_data)
SUMMARY(proc_data)
PLOT(proc_data)
model = FIT_MODEL(proc_data)
prediction = PREDICT(model)
PRINT(prediction)
> "Woo-hoo! validated model =)"
```

This is what you learn in the real world

```
# (real) data analysis process
raw_data = GET(data)
clean_data = CLEAN(data)
proc_data = PROCESS(clean_data)
while (QUALITY(proc_data) != "good") {
    clean_data = CLEAN(proc_data)
    proc_data = PROCESS(clean_data)
    # while loop may run indefinitely
}
SUMMARY(proc_data)
PLOT(proc_data)
model = FIT_MODEL(proc_data)
prediction = PREDICT(model)
PRINT(prediction)
> "Ooops! model sucks =("
```



# Dataset - Example



Is this visualization the best to identify differences?

Is the frequency optimally transformed?

# Clustering: use tSNE coordinates or not?

