

QA in high throughput flow cytometry

Robert Gentleman

Nishant Gopalakrishnan

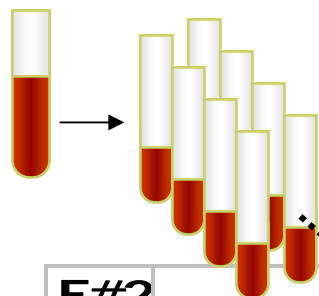
and a host of others

Existing Software+Stragies

- N. Le Meur *et al*, Data Quality Assessment of Ungated Flow Cytometry Data in High Throughput Experiments, *Cytometry, Part A*, 71A, 393-403, 2007
- flowCore/flowViz/flowQ all available from the Bioconductor project
- now in development **iFlow** a graphical user interface for the analysis of flow cytometry data

Applications

- 384 well plate format, for either stored samples, cells vs different drugs, etc.
- bar coded samples (multiplexing 25 samples in a single run)
- longitudinal studies, response to HIV vaccine, response to organ transplant, drugs, etc

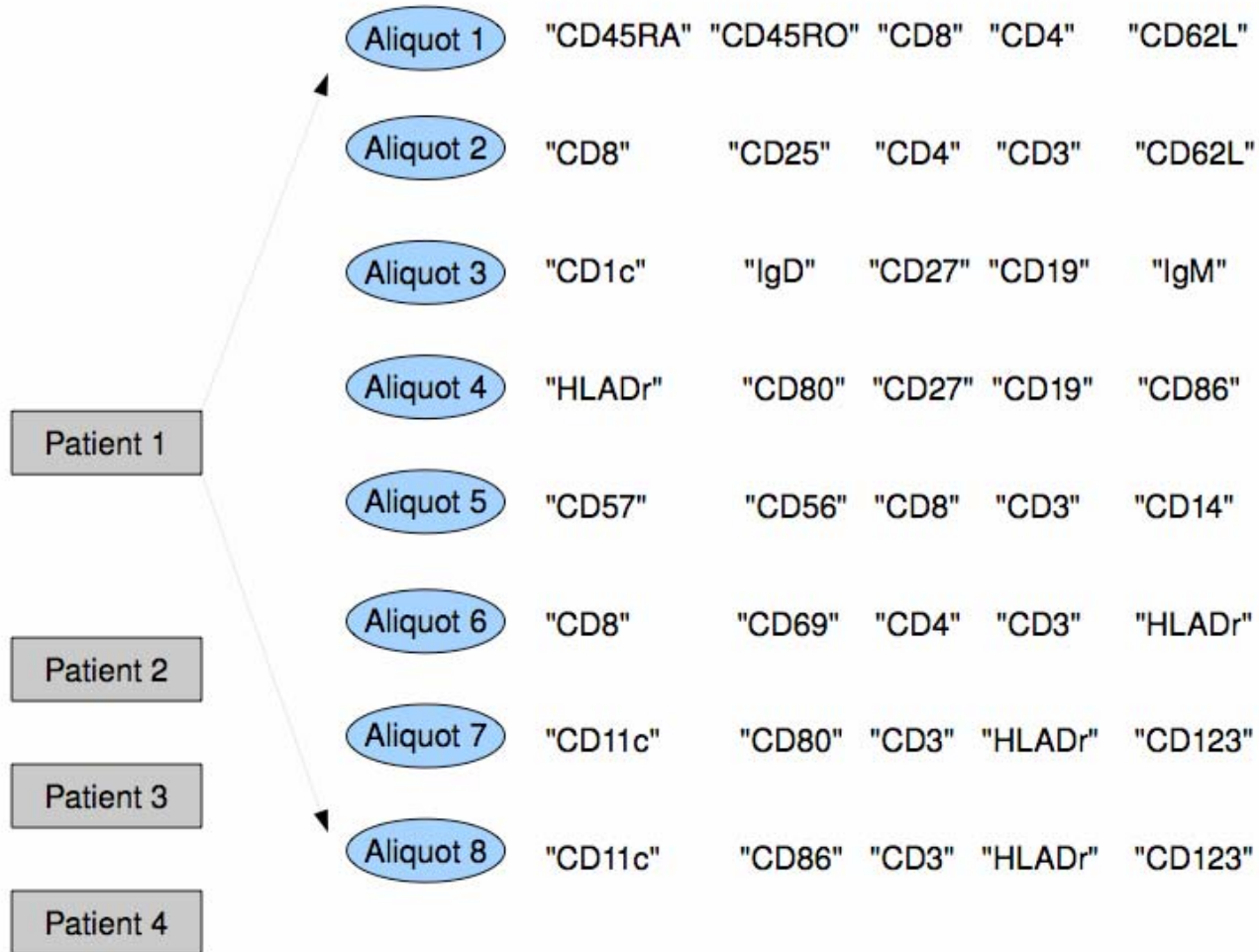


Time points

Stains

E#2		1	2	3	4	5	6	7	8	9	10	11	12
A	Leukocytes	-6	0	14	21	28	35	42	49	56	63	70	77
B	T cells	-6	0	14	21	28	35	42	49	56	63	70	77
C	NK cells	-6	0	14	21	28	35	42	49	56	63	70	77
D	B cells	-6	0	14	21	28	35	42	49	56	63	70	77
E	TCR	-6	0	14	21	28	35	42	49	56	63	70	77
F	1Act Marker	-6	0	14	21	28	35	42	49	56	63	70	77
G	2Act Marker	-6	0	14	21	28	35	42	49	56	63	70	77
H	3Act Marker	-6	0	14	21	28	35	42	49	56	63	70	77

Stain Information

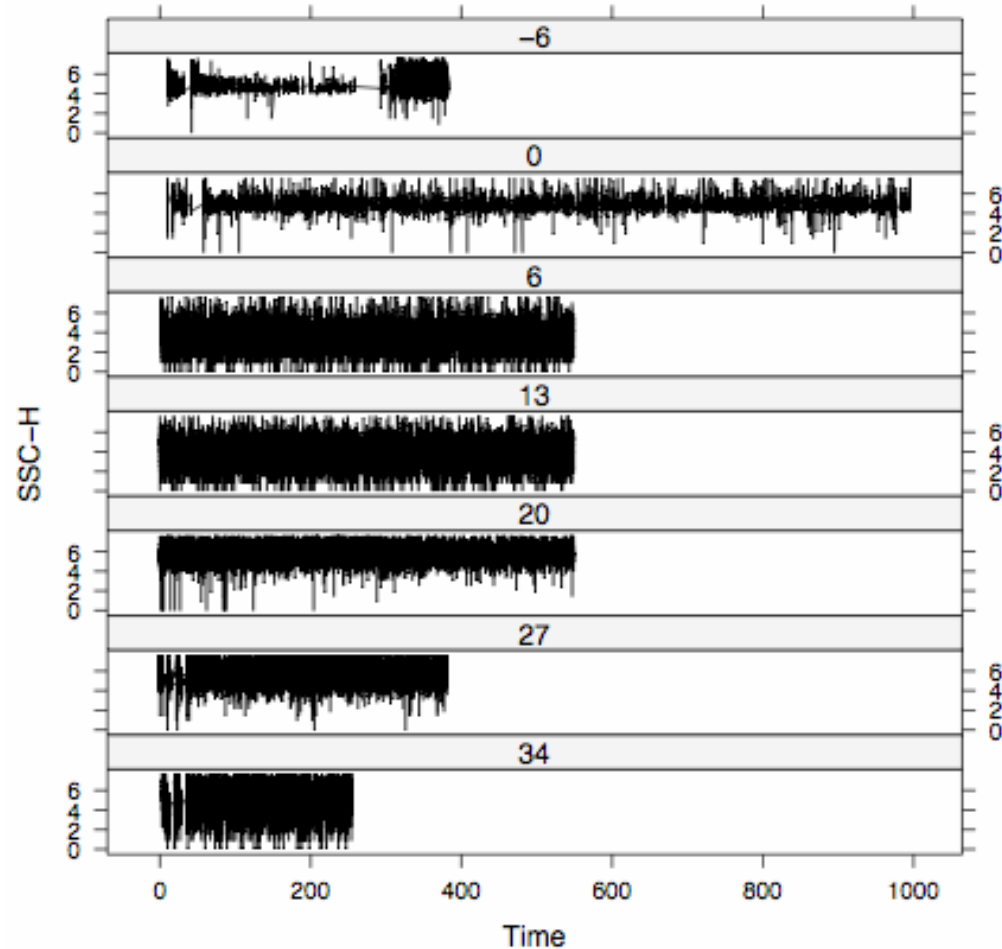


Some Tools

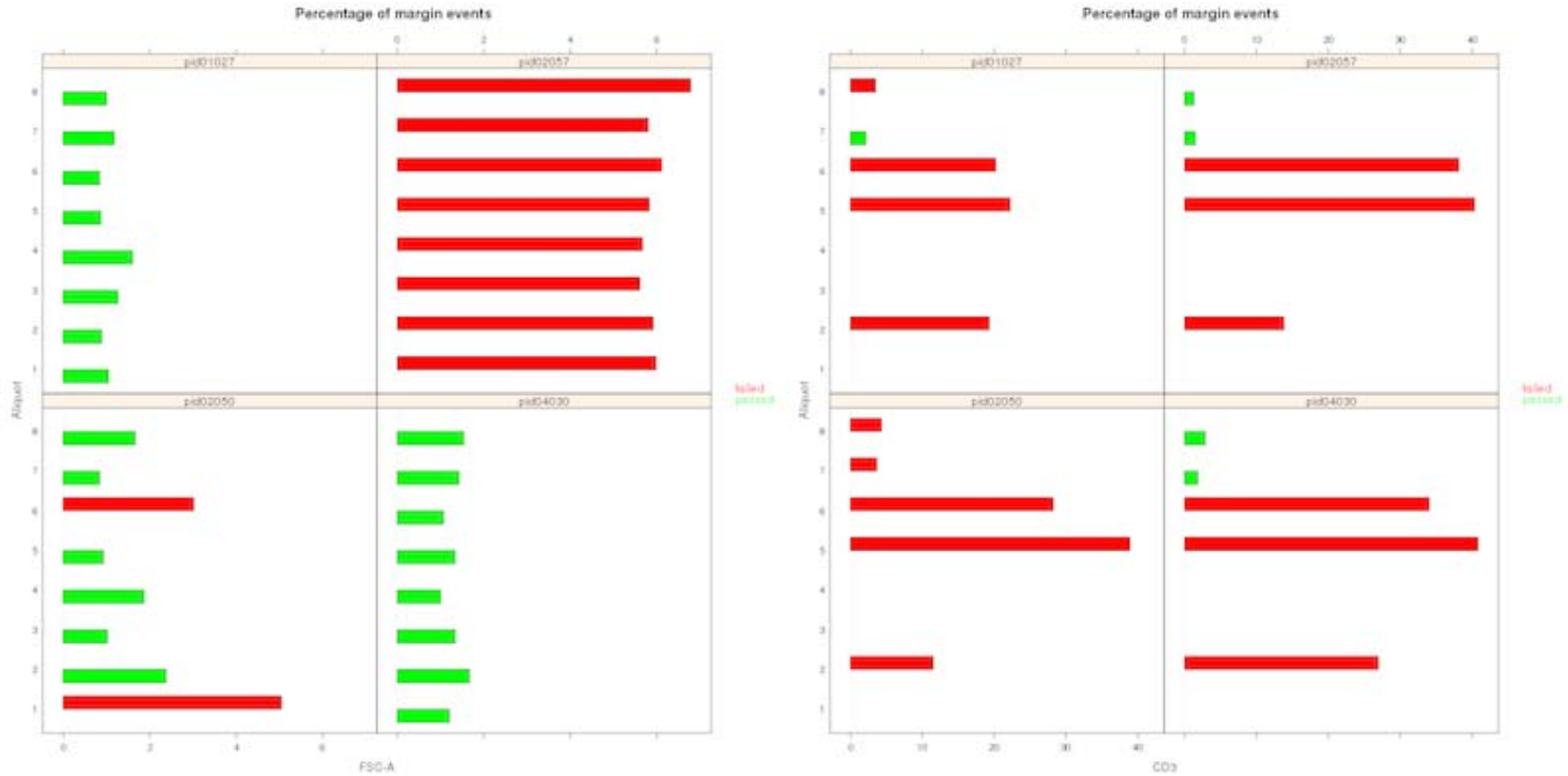
- Visualization methods are essential for flow cytometry
- basic problems
 - acquisition: time counter
 - boundary events: signal too low or too high
- since we get a very large number of measurements on different cells concentrating on the distribution/density seems prudent
 - ECDF plots
 - density plots
 - distances between estimates of distributions

Time of events

- almost all flow cytometers keep track of the time a cell was processed
- odd patterns in the acquisition times may indicate problems



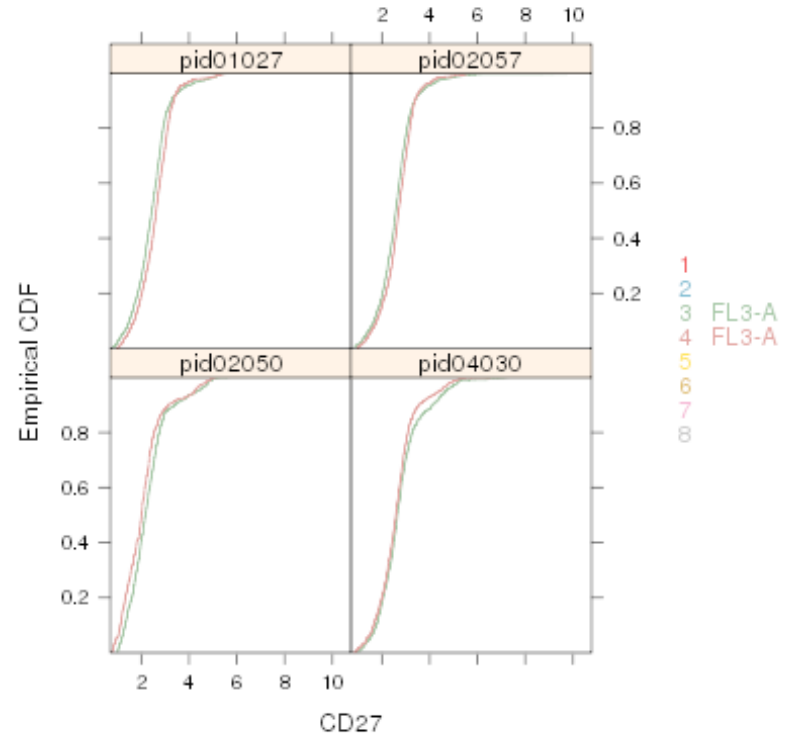
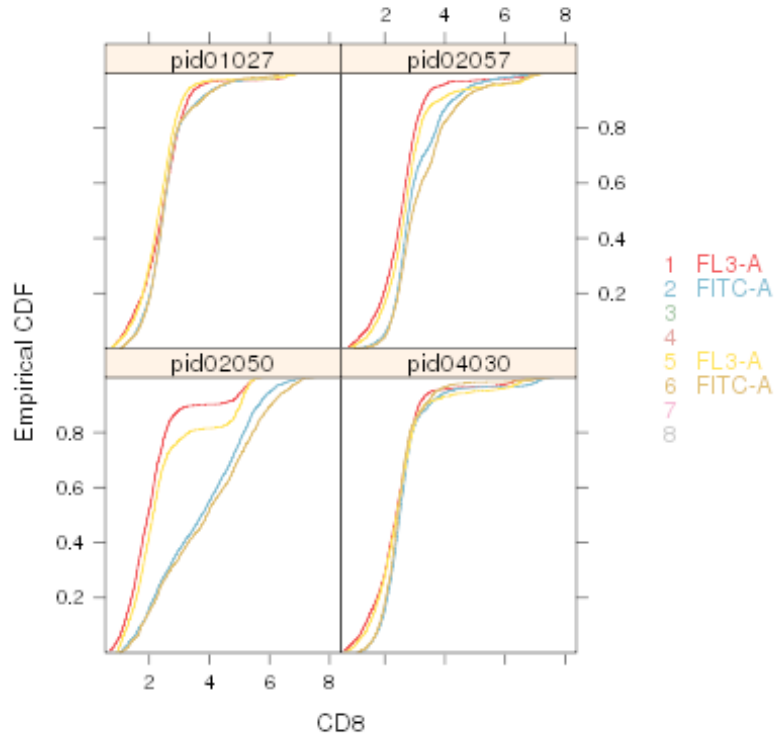
Boundary Events



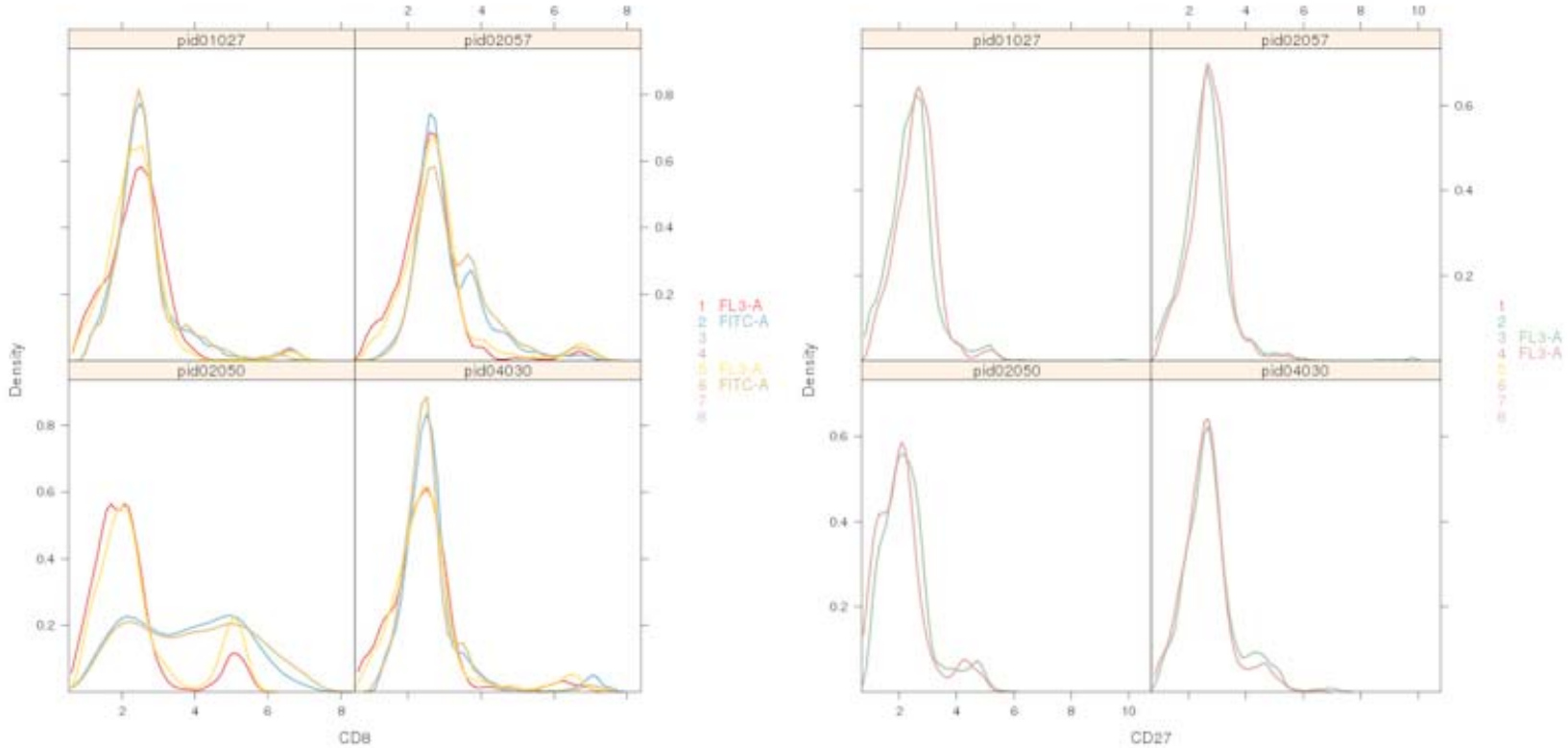
Philosophy Change

- the previous two methods are independent of what is in the run
- we suspect problems if either we see lots of boundary events, or oddities in timing, regardless of the content
- the next set of tools is based more on the notion that somethings should be the same, and attempting to detect when they are not

ECDF plots



Density plots



Kullback-Leibler Distance

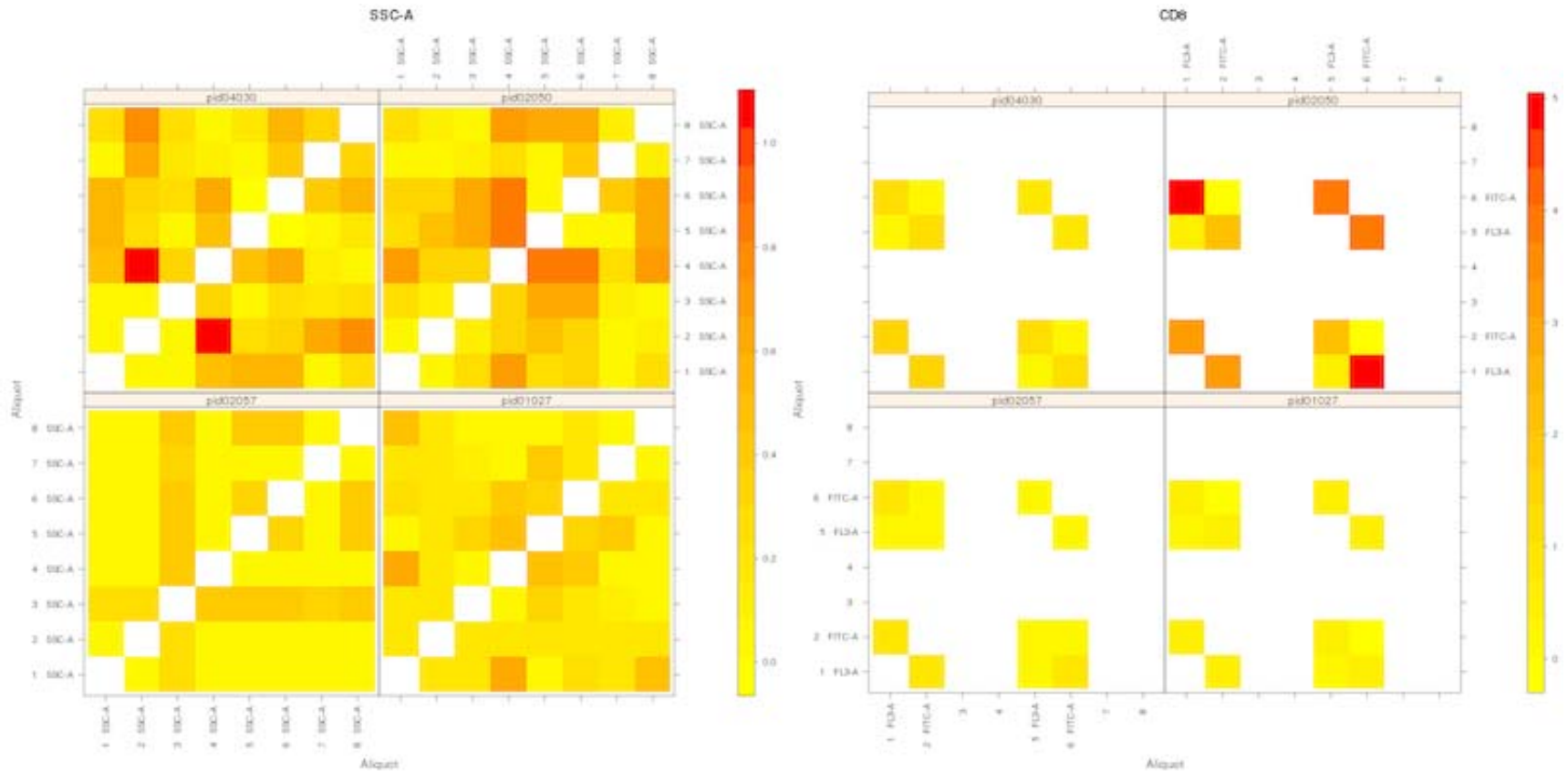
- an important question is how different are two density

The Kullback-Leibler Information between densities f_1 and f_2 is defined as

$$KLI = \int \frac{f_1(x)}{f_2(x)} f_2(x) dx$$

- one approach is to use the KL-distance (often symmeterized) ■

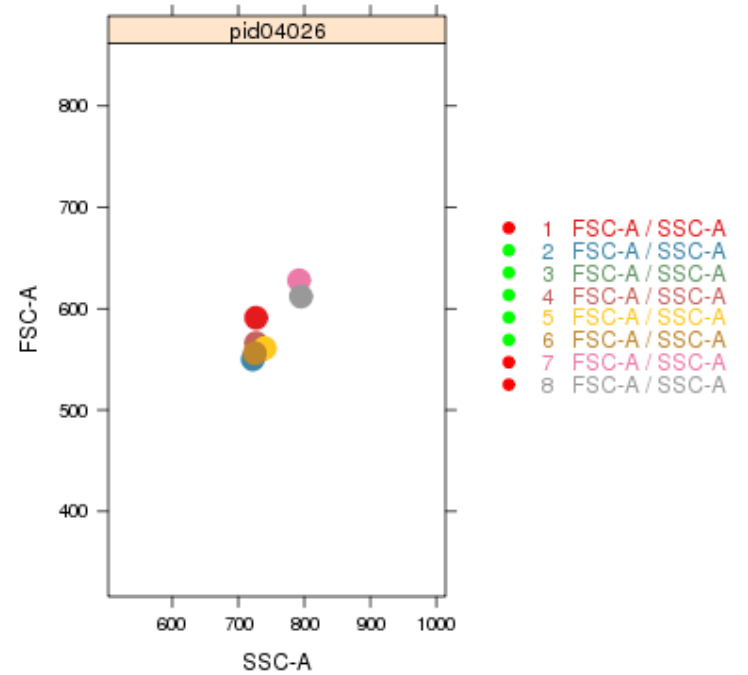
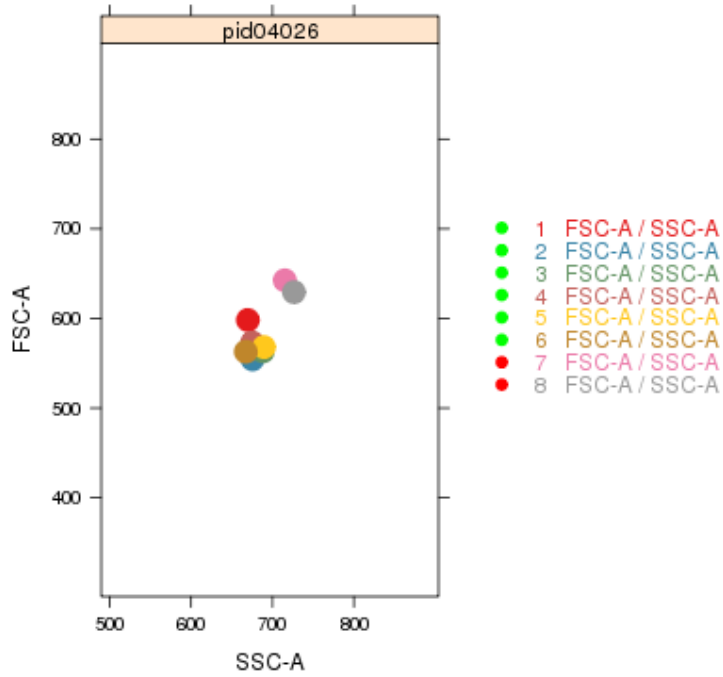
Heatmaps



Summary Statistics

- we are also trying to assess the value of different summary statistics
 - mean/median
- in order to be able to compare them across aliquots and look for outliers
- we use some standard outlier tests from the **outlier** and **mvoutlier** packages

Mean/median



Thanks

- Florian Hahne
- Deepayan Sarkar
- Nolwenn LeMeur
- Kyongryun Lee
- Adam Asare
- Vicki Seyfert
- Ryan Brinkman
- Matt Wand