THE BIOCONDUCTOR PACKAGE FLOWCORE, A SHARED DEVELOPMENT PLATFORM FOR FLOW CYTOMETRY DATA ANALYSIS IN R

N. Le Meur[1,2], F. Hahne[1], R. Brinkman[3], B. Ellis[5], P. Haaland[4], D. Sarkar [1], J. Spidlen[3], E. Strain[4], R. Gentleman[1]

[1] Fred Hutchinson Cancer Research Center, Seattle, USA,

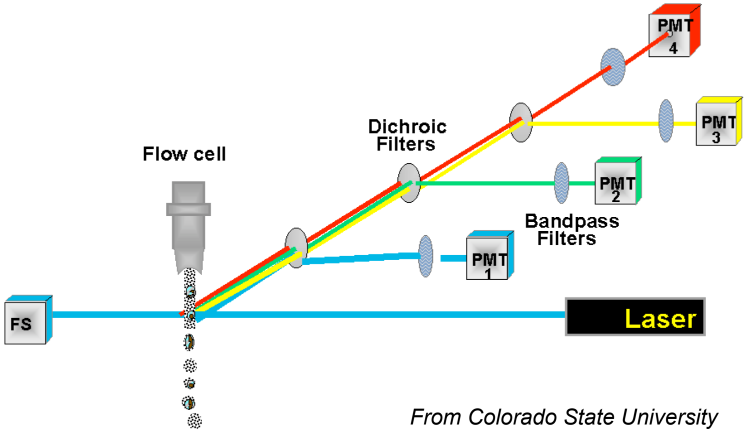
[2] INSERM EA SERAIC, Université de Rennes I, France

[3] British Columbia Cancer Research Center, Vancouver, Canada

[4] BD Biosciences, USA

[5] AdBrite Inc, San Francisco, USA

Flow Cytometry



From Colorado State University

- Immuno-typing
- Cell count
- DNA count
- Pathogen detection

- Healthcare
- Microbiology
- Agro-science
- Industry

Challenges

- High throughput multi-factorial data
  - Data management
  - Time management
- Reproducibility
  - Automation
  - Standardization

→ `flowCore` and Co

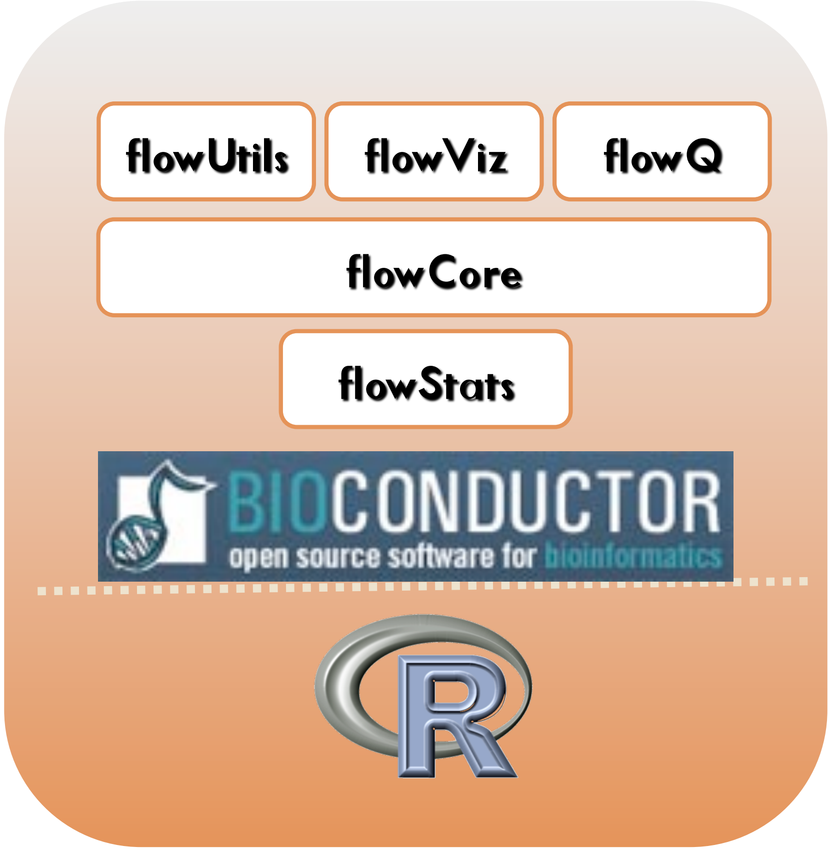
flowCore is…

- a **Bioconductor package** providing support for flow data to the R statistical programming language
- a **shared development platform** for statistical software to analyze (high-throughput) flow cytometry data
- a collection of **data structures**, associated **methods** and **functions** for the standard operations in flow data analysis
- one **implementation of** the Gating-ML, Transformation-ML and Compensation-ML **standards**
- platform independent
- extendable

flowCore is not …

- a **GUI** tool designed for interactive use or small scale data inspection
- a collection of **ready to use workflows** (although one can combine the tools offered by flowCore into workflows by means of scripts)
- a **data base** (although it can speak to almost all data bases via the standard interfaces)

flowCore and Co



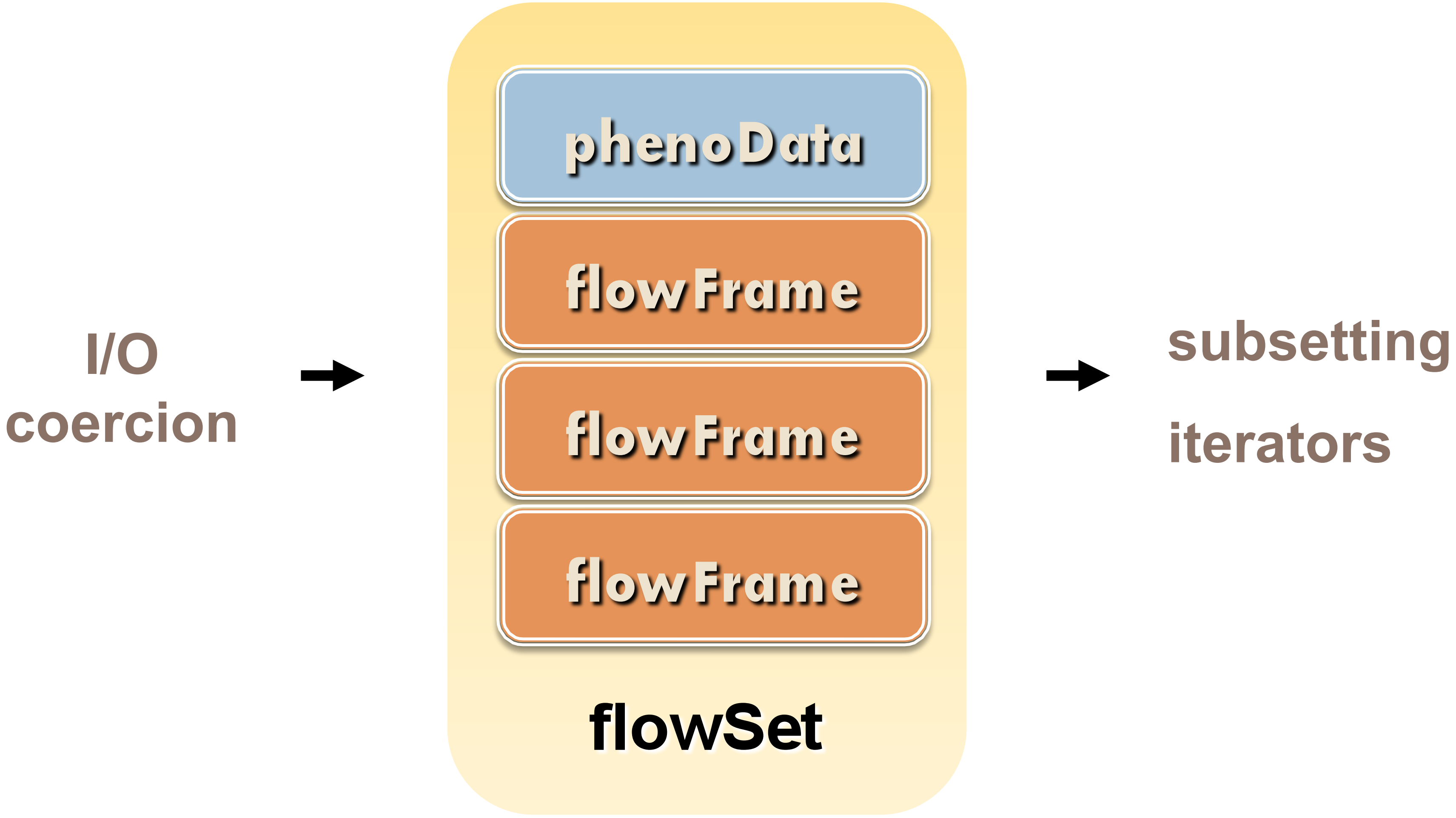
flowUtils flowViz flowQ

flowCore

flowStats

- basic data structures, standard flow operations
- I/O, data base access
- visualization of flow data
- Quality assessment, quality control
- statistical methods
- Annotation, bioinformatics tools
- general purpose tools

Data structures



I/O
coercion

phenoData

flowFrame

flowFrame

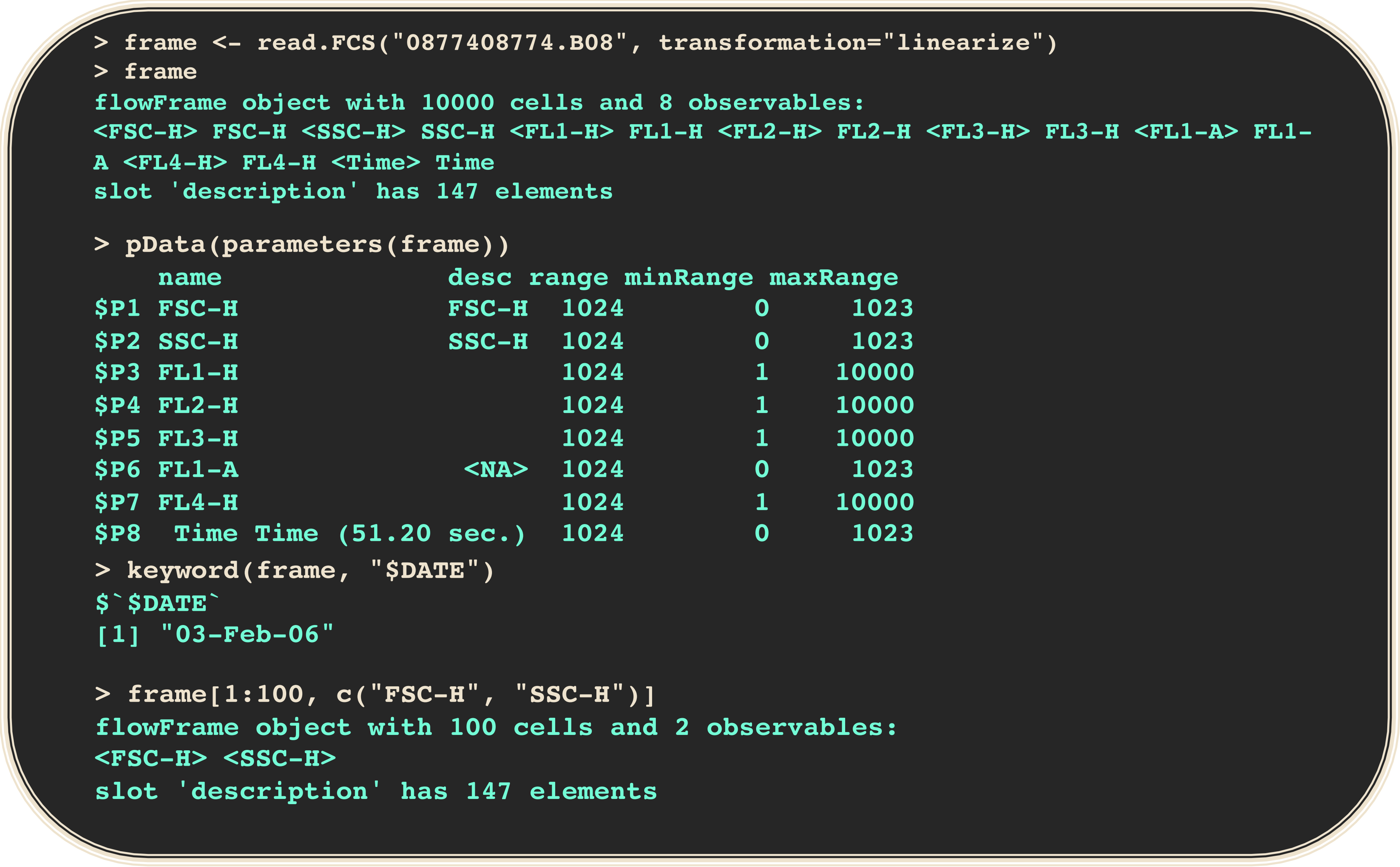
flowFrame

flowSet

subsetting

iterators

flowFrame



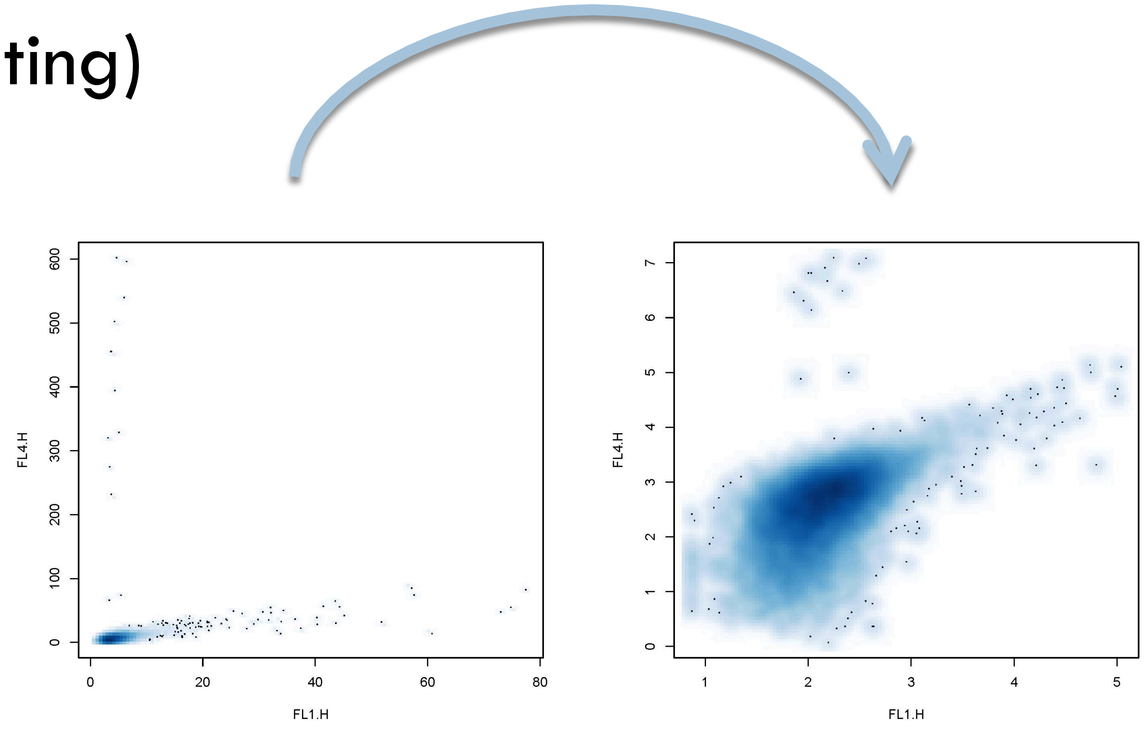
```
> frame <- read.FCS("0877408774.B08", transformation="linearize")
> frame
flowFrame object with 10000 cells and 8 observables:
<FSC-H> FSC-H <SSC-H> SSC-H <FL1-H> FL1-H <FL2-H> FL2-H <FL3-H> FL3-H <FL1-A> FL1-
A <FL4-H> FL4-H <Time> Time
slot 'description' has 147 elements

> pData(parameters(frame))
     name              desc range minRange maxRange
$P1 FSC-H             FSC-H  1024        0     1023
$P2 SSC-H             SSC-H  1024        0     1023
$P3 FL1-H                    1024        1    10000
$P4 FL2-H                    1024        1    10000
$P5 FL3-H                    1024        1    10000
$P6 FL1-A              <NA>  1024        0     1023
$P7 FL4-H                    1024        1    10000
$P8  Time Time (51.20 sec.)  1024        0     1023
> keyword(frame, "$DATE")
$`$DATE`
[1] "03-Feb-06"

> frame[1:100, c("FSC-H", "SSC-H")]
flowFrame object with 100 cells and 2 observables:
<FSC-H> <SSC-H>
slot 'description' has 147 elements
```

flowSet

```
> set <- read.flowSet(pattern="060909")
> set
A flowSet with 5 experiments.

An object of class "AnnotatedDataFrame"
  rowNames: 060909.001, 060909.002, ..., 060909.005  (5 total)
  varLabels and varMetadata description:
    name: Name

  column names:
  FSC.H SSC.H FL1.H FL2.H FL3.H FL1.A FL4.H

> pData(set)
                 name Sample         Type
060909.001 060909.001               empty
060909.002 060909.002   fitc compensation
060909.003 060909.003     pe compensation
060909.004 060909.004    apc compensation
060909.005 060909.005   7AAD compensation

> set[[2]]
flowFrame object with 8805 cells and 7 observables:
<FSC.H> FSC.H <SSC.H> SSC.H <FL1.H> FL1.H <FL2.H> FL2.H
<FL3.H> FL3.H <FL1.A> FL1.A <FL4.H> FL4.H
slot 'description' has 129 elements
```

Standard operations

- *`compensate()`*
- *`transform()`*
- Filtering (or gating)

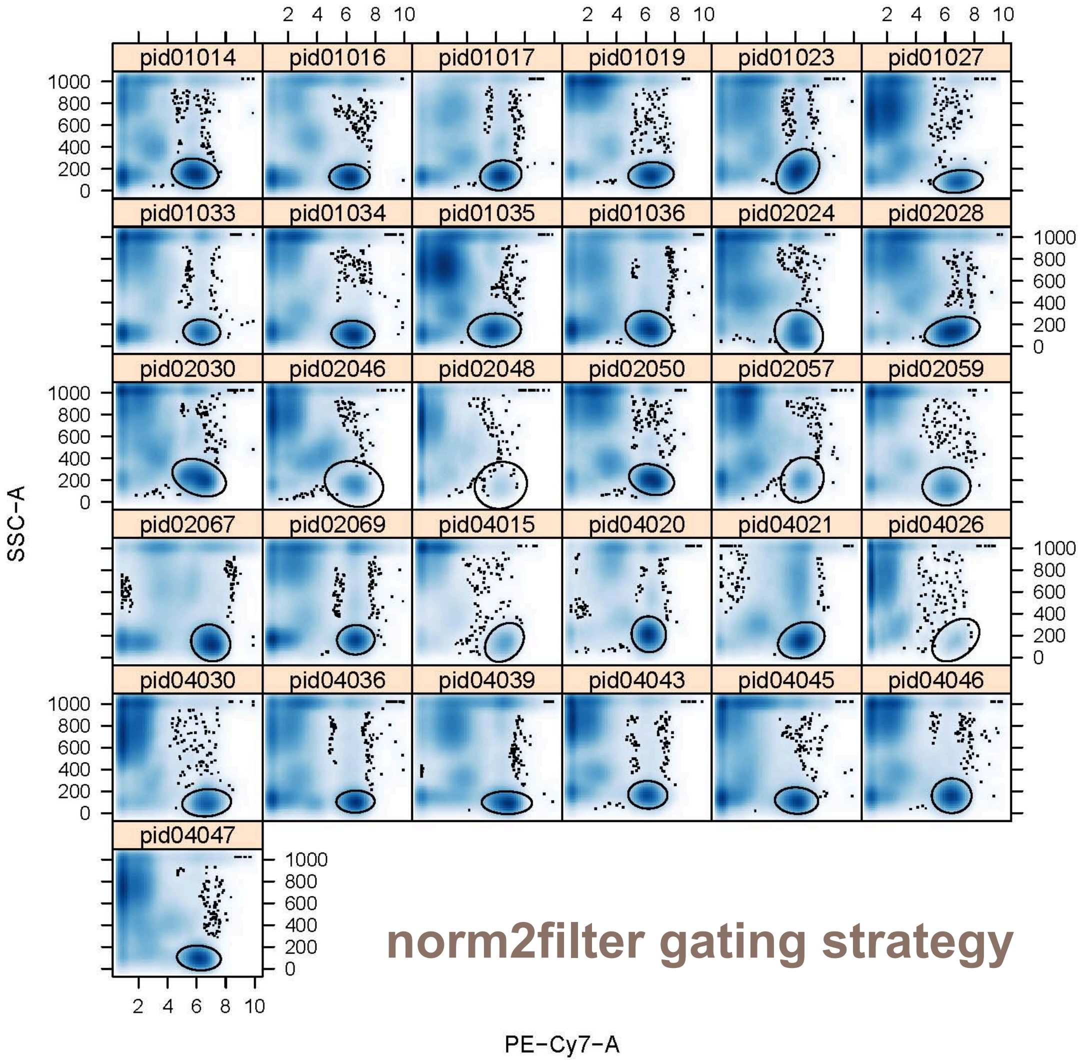


`arcsinTransform()`

Filtering and gating

- Defined by constant coordinates in the parameter space
  - rectangle
  - ellipsoide
  - quadratic
  - polytope
  - polygon
- Data driven gate (filter)
  - sampleFilter, random sampling of events to include
  - kmeansFilter
  - norm2Filter, fitting of bivariate normal distribution
  - curvFilter, high density regions

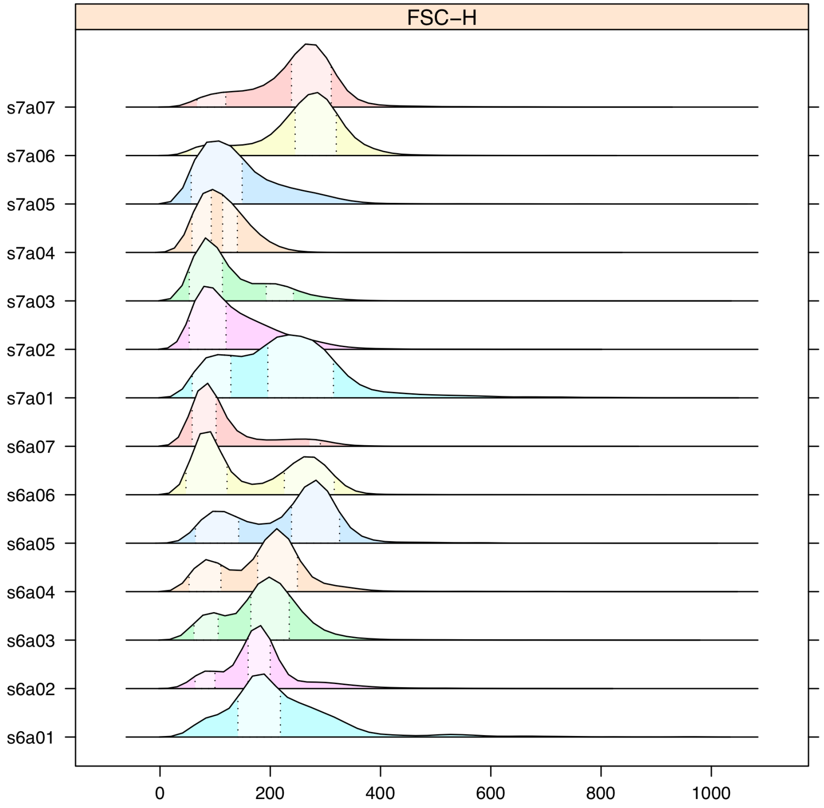
Data driven gate: an example



norm2filter gating strategy

Visualization with flowViz

- multivariate plots of flowFrames and flowSets using lattice-type graphics from *flowViz*
  - conditional variables (e.g. frames in a flowSet)
  - grouping variables (e.g. treatments)
  - gates, filters
  - customization



```
> library(flowCore)
> library(flowViz)
> data(GvHD)
> densityplot(~ `FSC-H`, GvHD[8:21],
     filter=curv1Filter("FSC-H"))
```

flowQ

flow set details

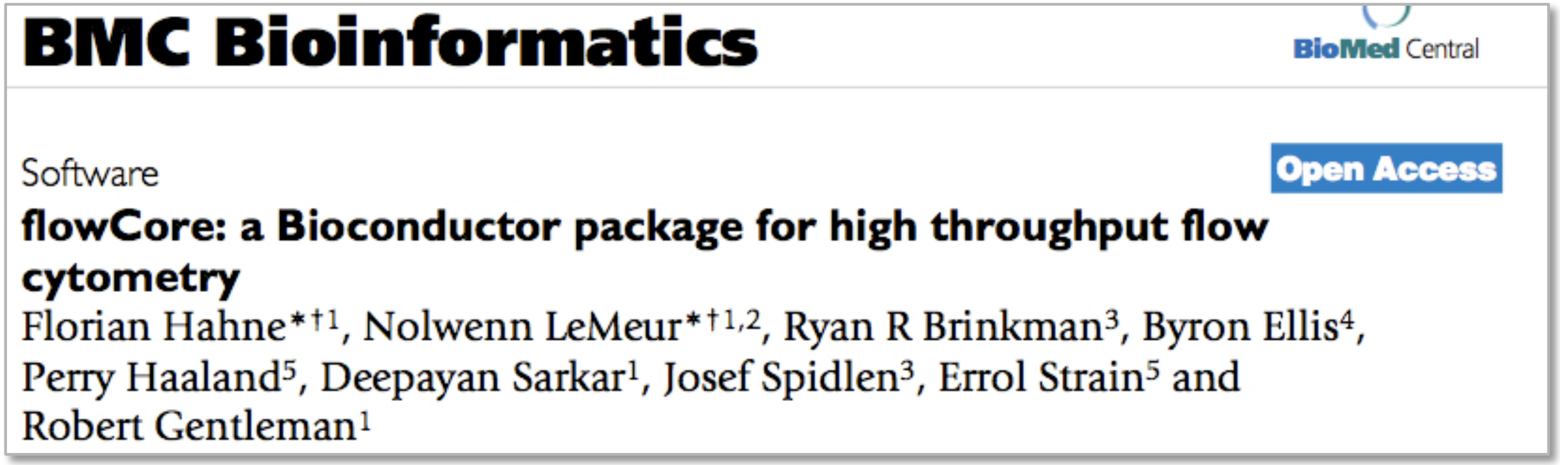
name	desc	minRange	maxRange
FSC-H	FSC-Height	0	1023
SSC-H	SSC-Height	0	1023
FL1-H	CD15 FITC	0.881	1023
FL2-H	CD45 PE	0.881	1023
FL3-H	CD14 PerCP	1	1023
FL2-A	NA	0	1023
FL4-H	CD33 APC	1	1023
Time	Time (51.20 sec.)	0	1023

flow set details	cell number	margin events			time line		time flow
1 s5a01	3400	●	FSC-H 0.4%	SSC-H 2%	●	FL1-H 0	●
2 s5a02	3400	●	FSC-H 0.2%	SSC-H 0.8%	●	FL1-H 0	●
3 s5a03	3400	●	FSC-H 0.3%	SSC-H 0.8%	●	FL1-H 0	●
4 s5a04	8600	●	FSC-H 0.2%	SSC-H 0.4%	●	FL1-H 0	●
5 s5a05	10000	●	FSC-H 0.9%	SSC-H 0.6%	●	FL1-H 0	●
6 s5a06	3800	●	FSC-H 2%	SSC-H 3%	●	FL1-H 0	●
7 s5a07	14000	●	FSC-H 4%	SSC-H 2%	●	FL1-H 7.1	●

Patient	Visit	Days	Grade	name
1	7	33	3	s5a07

Conclusions & perspectives

- A **shared development platform** for statistical software to analyze (high-throughput) flow cytometry data



BMC Bioinformatics

Software

flowCore: a Bioconductor package for high throughput flow cytometry

Florian Hahne*†[1], Nolwenn LeMeur*†[1,2], Ryan R Brinkman[3], Byron Ellis[4], Perry Haaland[5], Deepayan Sarkar[1], Josef Spidlen[3], Errol Strain[5] and Robert Gentleman[1]

Acknowledgements

- Robert Gentleman
- Florian Hahne
- Deepayan Sarkar
- Ryan Brinkman
- Joseph Spidlen
- Perry Haaland
- Byron Ellis
- Nathalie Théret
- Michel Le Borgne

FRED HUTCHINSON
CANCER RESEARCH CENTER

BC Cancer Agency
CARE & RESEARCH
An agency of the Provincial Health Services Authority

UMR IRISA

cnrs UNIVERSITÉ DE RENNES 1 INSA RENNES ens CACHAN BRETAGNE

Inserm
Institut national
de la santé et de la recherche médicale

CONTRE LE CANCER
LA LIGUE
101 comités