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

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

\textsuperscript{1} Fred Hutchinson Cancer Research Center, Seattle, USA,
\textsuperscript{2} INSERM EA SERAIC, Université de Rennes I, France
\textsuperscript{3} British Columbia Cancer Research Center, Vancouver, Canada
\textsuperscript{4} BD Biosciences, USA
\textsuperscript{5} AdBrite Inc, San Francisco, USA
Flow Cytometry

- 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

- 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
```r
> 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))

<table>
<thead>
<tr>
<th>name</th>
<th>desc</th>
<th>range</th>
<th>minRange</th>
<th>maxRange</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P1 FSC-H</td>
<td>FSC-H</td>
<td>1024</td>
<td>0</td>
<td>1023</td>
</tr>
<tr>
<td>$P2 SSC-H</td>
<td>SSC-H</td>
<td>1024</td>
<td>0</td>
<td>1023</td>
</tr>
<tr>
<td>$P3 FL1-H</td>
<td></td>
<td>1024</td>
<td>1</td>
<td>10000</td>
</tr>
<tr>
<td>$P4 FL2-H</td>
<td></td>
<td>1024</td>
<td>1</td>
<td>10000</td>
</tr>
<tr>
<td>$P5 FL3-H</td>
<td></td>
<td>1024</td>
<td>1</td>
<td>10000</td>
</tr>
<tr>
<td>$P6 FL1-A</td>
<td>&lt;NA&gt;</td>
<td>1024</td>
<td>0</td>
<td>1023</td>
</tr>
<tr>
<td>$P7 FL4-H</td>
<td></td>
<td>1024</td>
<td>1</td>
<td>10000</td>
</tr>
<tr>
<td>$P8 Time</td>
<td>Time (51.20 sec.)</td>
<td>1024</td>
<td>0</td>
<td>1023</td>
</tr>
</tbody>
</table>

> keyword(frame, "$DATE")

```

```r
$\$DATE$
```

[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

```r
> 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:

> pData(set)

<table>
<thead>
<tr>
<th>name</th>
<th>Sample</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>060909.001</td>
<td>060909.001</td>
<td>empty</td>
</tr>
<tr>
<td>060909.002</td>
<td>060909.002</td>
<td>fitc compensation</td>
</tr>
<tr>
<td>060909.003</td>
<td>060909.003</td>
<td>pe compensation</td>
</tr>
<tr>
<td>060909.004</td>
<td>060909.004</td>
<td>apc compensation</td>
</tr>
<tr>
<td>060909.005</td>
<td>060909.005</td>
<td>7AAD compensation</td>
</tr>
</tbody>
</table>

> set[[2]]
flowFrame object with 8805 cells and 7 observables:
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

```r
> library(flowCore)
> library(flowViz)
> data(GvHD)
> densityplot(~ `FSC-H`, GvHD[8:21], filter=curv1Filter("FSC-H"))
```
The image shows a flow cytometry analysis software interface with various datasets and parameters. The interface includes tables and graphs for different flow set details, cell number, margin events, and time flow. Each dataset is represented by a unique identifier and includes parameters such as FSC-H, SSC-H, FL1-H, etc. The tables display data points for each parameter, and the graphs illustrate trends over time or other variables.
Conclusions & perspectives

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

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