ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

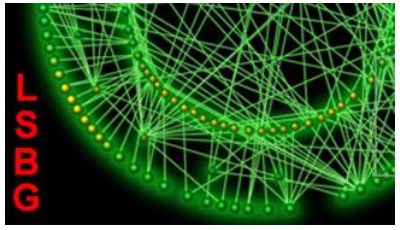
Biotechnology and Bioengineering

Laboratory of Systems Biology and Genetics

ASAP
Automated Single-cell Analysis Pipeline

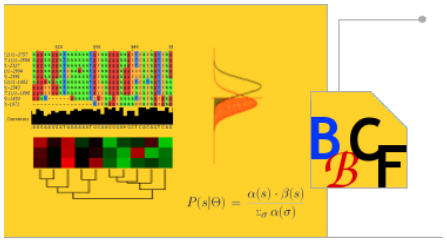
WEB-BASED PIPELINE FOR THE ANALYSIS AND INTERACTIVE VISUALIZATION OF SINGLE-CELL RNA-SEQ DATA

VINCENT GARDEUX



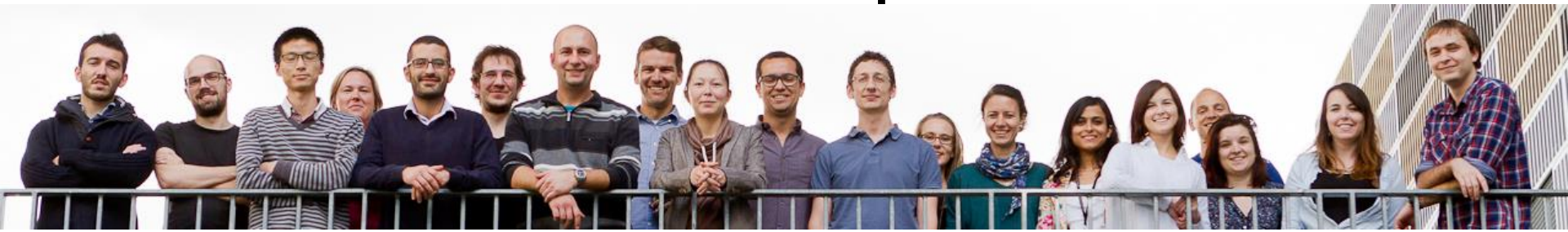
BART DEPLANCKE
PETRA C. SCHWALIE

FABRICE DAVID



Published in Bioinformatics: https://doi.org/10.1093/bioinformatics/btx337

Few words about Deplancke's lab



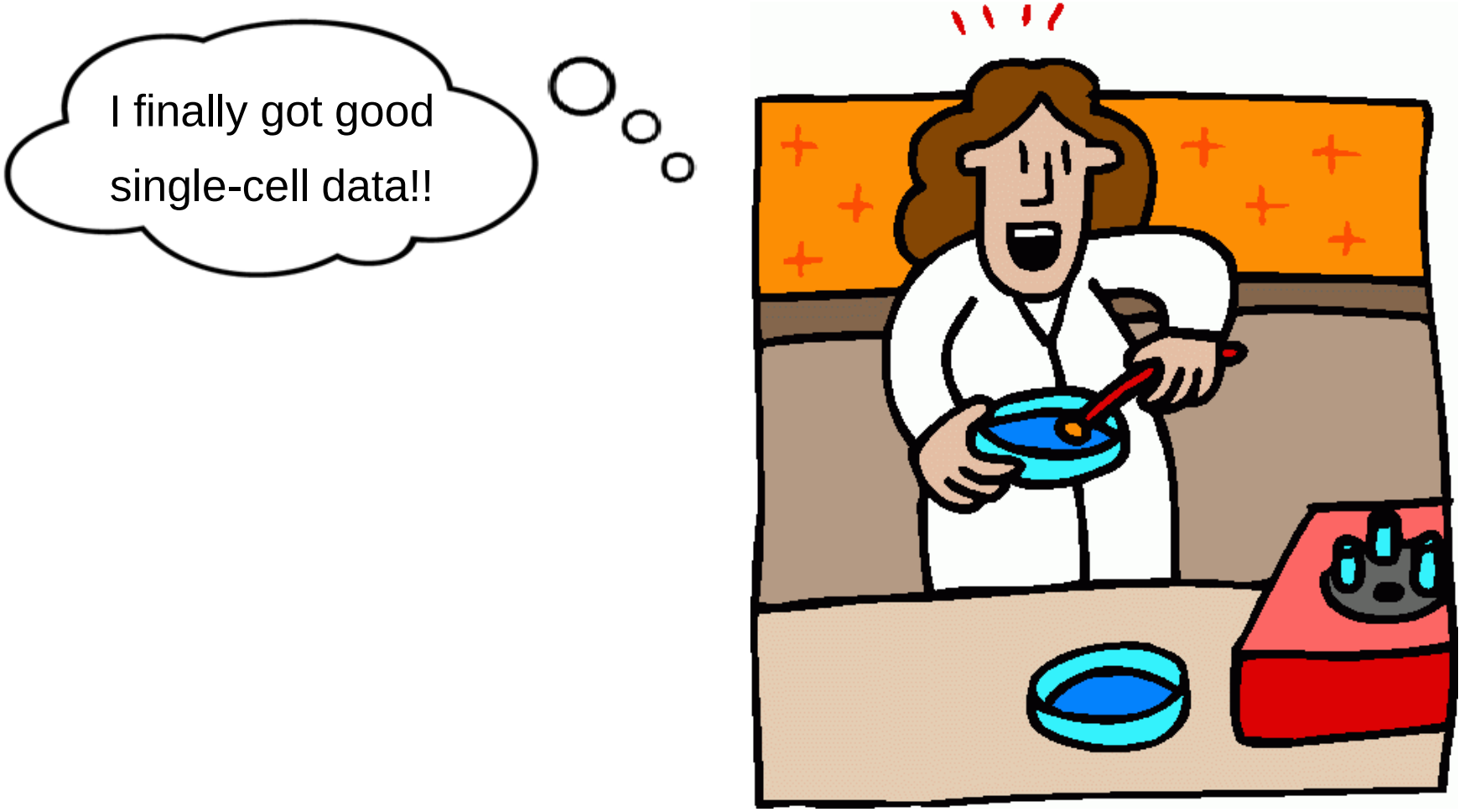
- Located at EPFL (Ecole Polytechnique Fédérale de Lausanne), Lausanne, Switzerland
- 1 Bart, 5 Postdocs, 11 (+ 2 shared) PhD students, 2 Master students, 2 lab tech
- Working in Fly genomics, Aging, Obesity, and Microfluidics tech

⇒RNAseq, ATACseq, ChIPseq, WES, Proteomics, …

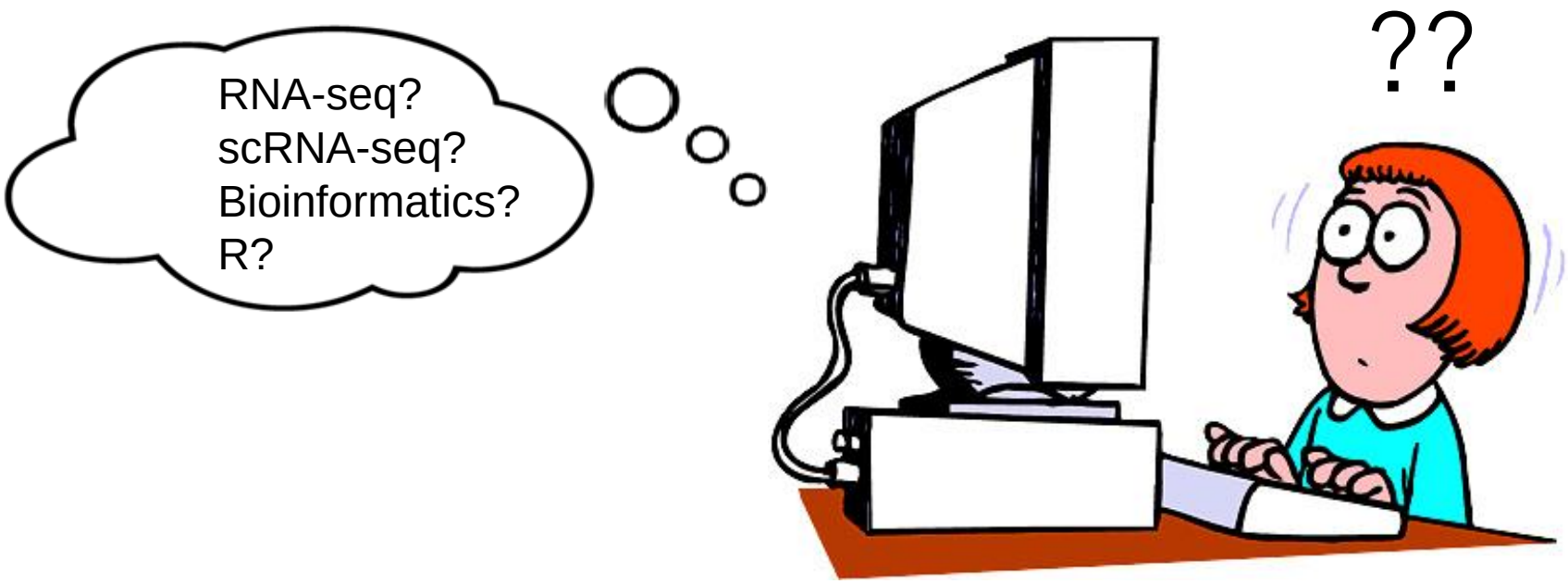
- Also, more and more datasets generated from single-cell technologies: Dropseq & Smartseq2

⇒And only 2 bioinformaticians!

Community needs: « I want to do scRNA-seq ! »



Community needs: « But how do I analyze scRNA-seq data? »



⇒ This is a typical bottleneck, and was routinely experienced in our own lab

scRNA-seq Computational Workflow

Quality control & filtering (I)

↓

Mapping to reference genome

↓

Read/transcript matrix
(raw counts/UMI or normalized data)

ASAP

↓

Quality control & filtering (II)

↓

Normalization, noise removal

↓

Study-specific downstream analyses
to generate new biological insight

Cell type identification
Cell type characterization
Gene network analysis
Kinetics of transcription

scRNA-seq Computational Workflow

Quality control & filtering (I)

↓

Mapping to reference genome

↓

Read/transcript matrix
(raw counts/UMI or normalized data)

ASAP

"Black box" fixed pipeline?

scRNA-seq pipeline may not be applicable to all datasets

nature International weekly journal of science

Archive › Volume 547 › Issue 7661 › Toolbox › Article

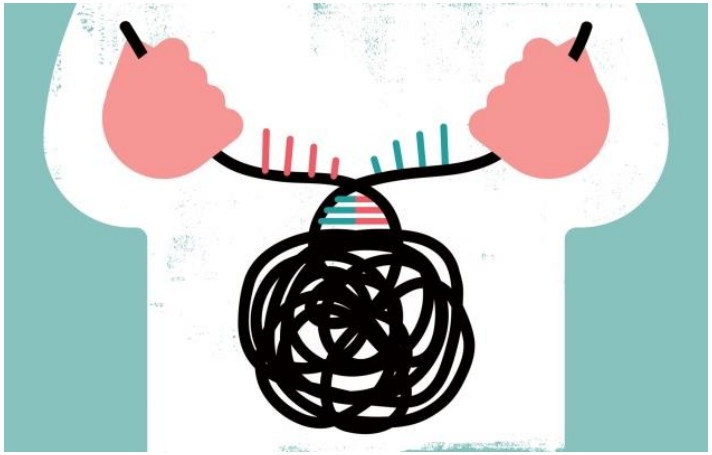
NATURE | TOOLBOX

Single-cell sequencing made simple

Data from thousands of single cells can be tricky to analyse, but software advances are making it easier.

Jeffrey M. Perkel

03 July 2017 | Corrected: 05 July 2017, 06 July 2017



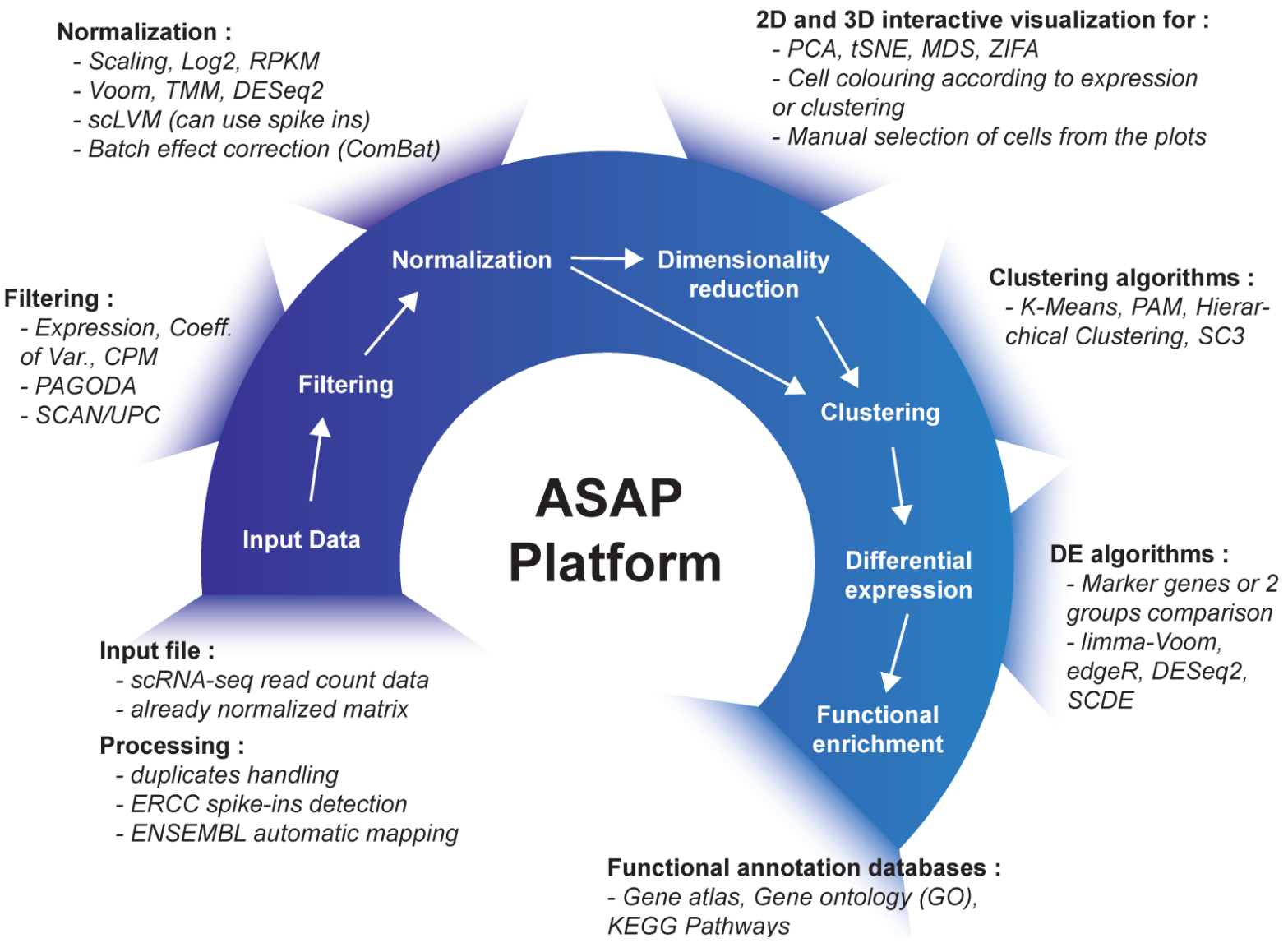
"The tools aren't perfect for every situation"

⇒ "A pipeline that excels at identifying cell types, for instance, might stumble with pseudo-time analysis"

"Appropriate methods are 'very data-set dependent'", says Sandrine Dudoit, (biostatistician at the University of California, Berkeley).

⇒ "The methods and tuning parameters may need to be adjusted to account for variables such as sequencing length"

ASAP: Automated Single-cell Analysis Pipeline



Technically, what is ASAP?

- **Centralized computational resources**: Ruby-on-rails server currently hosted at the BBCF

- Implementation of the "*delayed-jobs*" gem that allows job queuing management

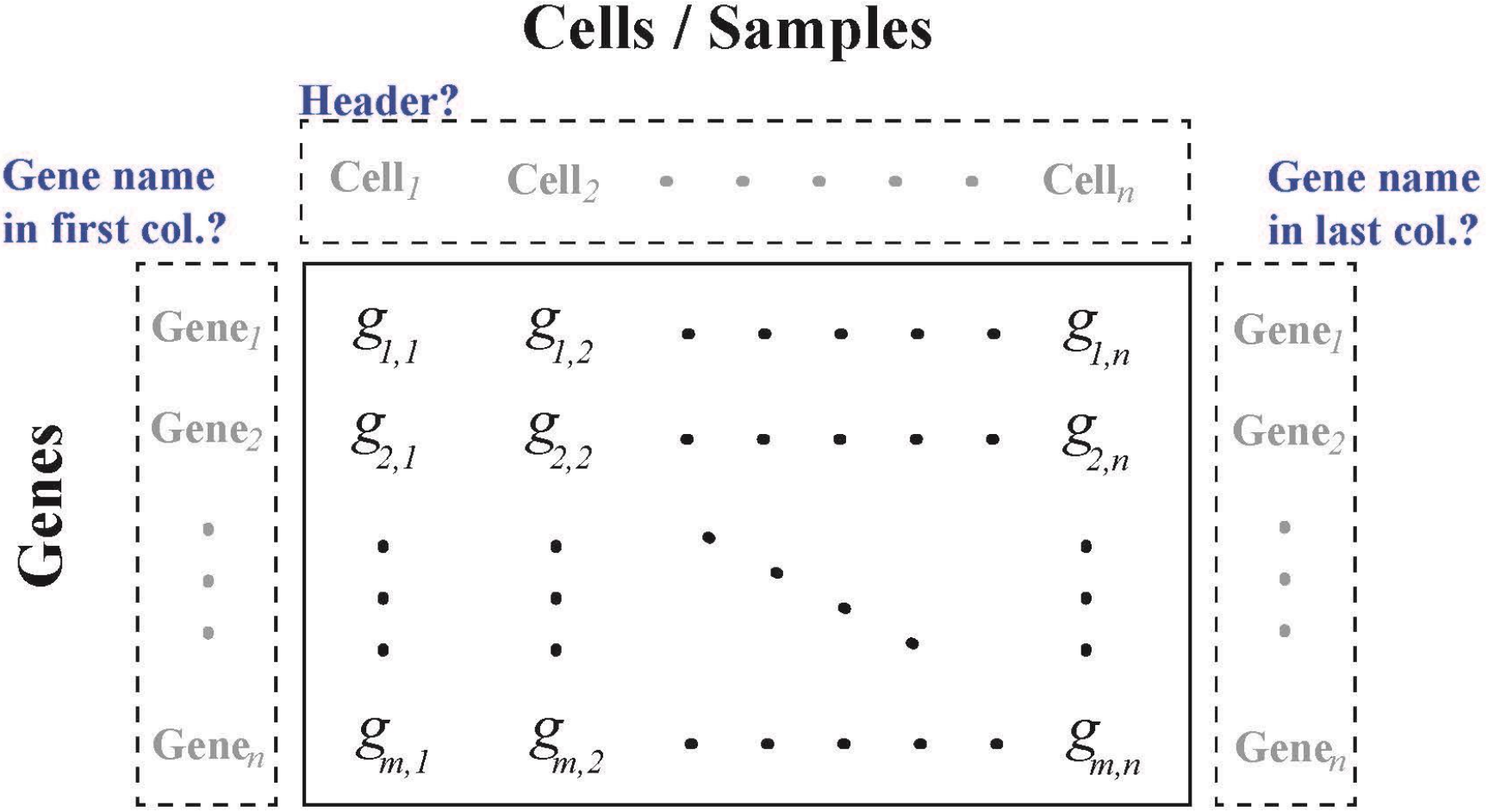
- Single-cell analysis scripts are written in R (mostly), Python (dimension reduction) and Java (parsing, functional enrichment)

⇒Generates JSON files that are interpreted by the browser

- **Interactive** and user-friendly web interface with 2D/3D visualization (using plotly JS) [currently moving to plotly webGL]

Dataset formatting

ASAP input file: count matrix or already normalized gene expression matrix



Demonstration

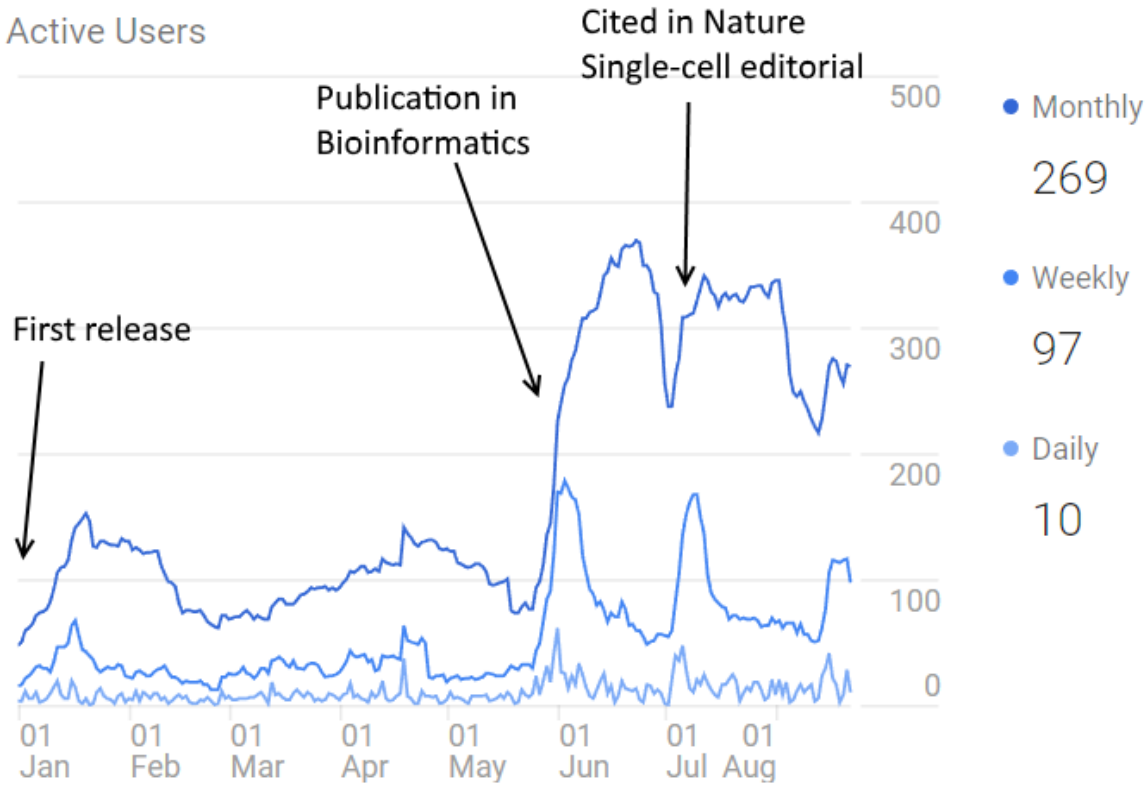
https://asap.epfl.ch

Conclusions

- To open state-of-the-art analysis algorithms to biology laboratories and non-bioinformaticians experts.
- Fast platform for checking and having an overview of every-day data output
- Still in development, many things can be done/added

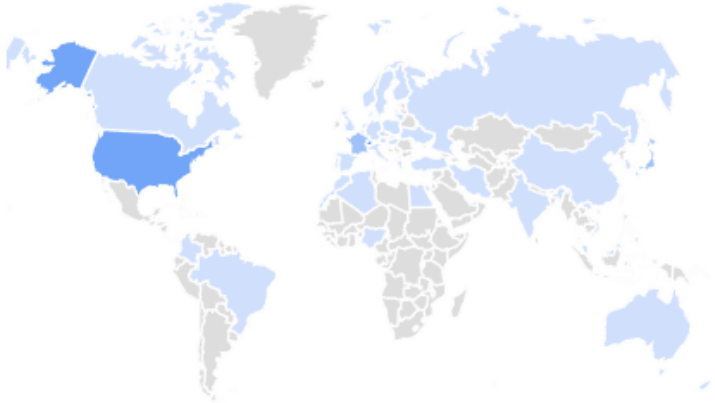
ASAP usage since first release

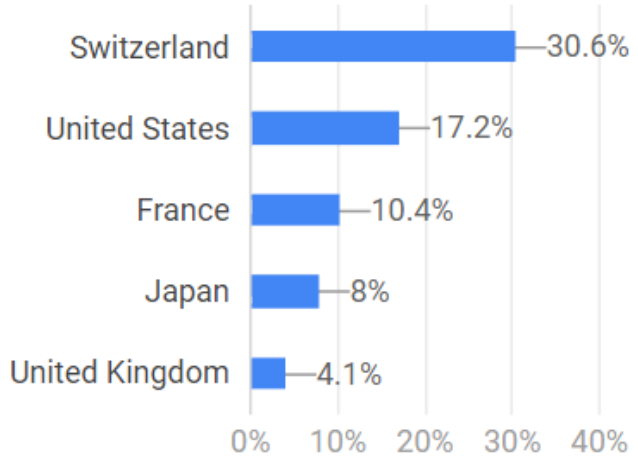
Users	Sessions	Bounce Rate	Session Duration
1.2K	3.7K	37.31%	9m 0s
↑2,941.5%	↑3,870.7%	↓1.9%	↓9.9%



- Monthly: 269
- Weekly: 97
- Daily: 10

Sessions by country



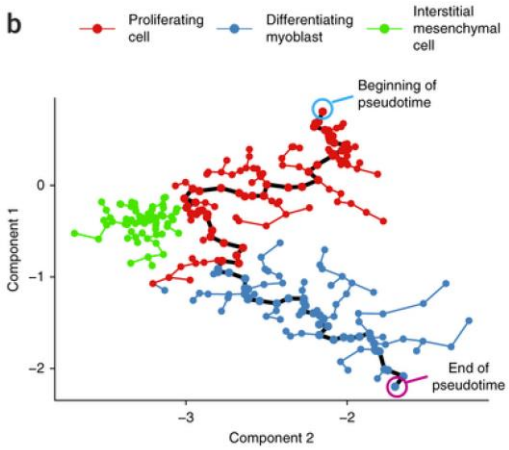


Future developments

- Improve visualization/responsiveness (front-end developer)

⇒ Many plots are in creation on the dev server

- Improve reproducibility (Docker, back-end dev)
- Allow projects to be collaboratively accessed/shared
- Add new tools / algorithms as they are now published (M3DROP, BPSC, …)
- Mapping dynamic processes / transcriptional trajectory (Monocle, …)



- Add other databases for functional enrichment (pharmgkb, oncogenes, …)
- Handle bigger datasets (>50k cells) for e.g. implementing HDF5 matrices