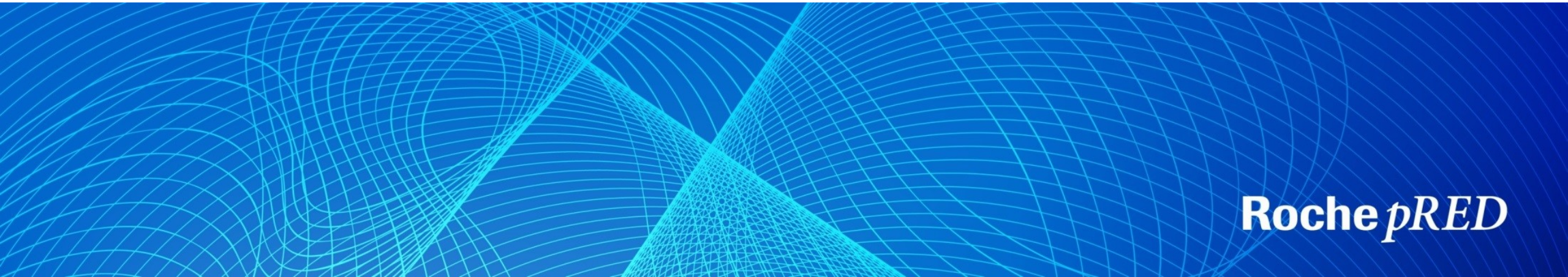
scRNA-Seq analysis in Cancer Immunotherapy Pharma Research

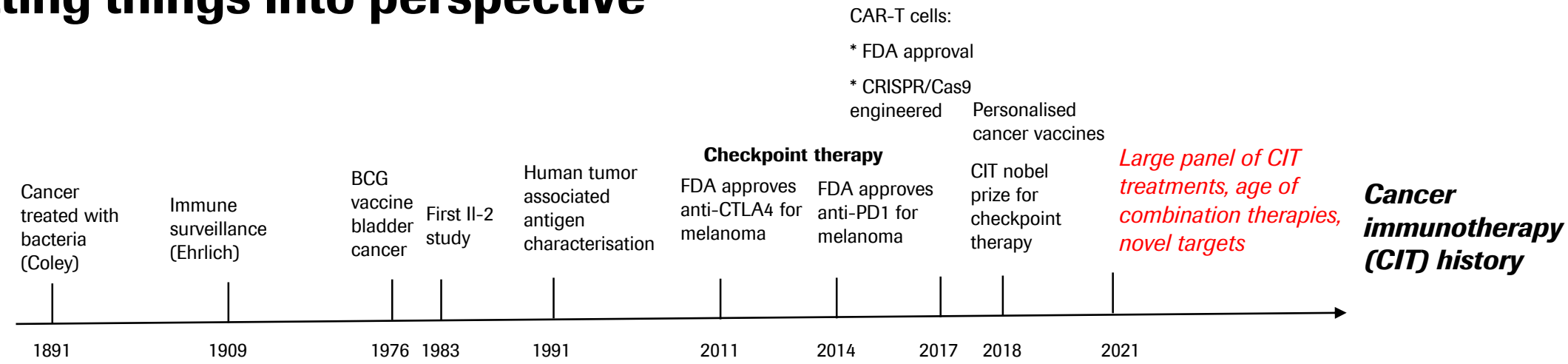
Petra Schwalie

A computational biologist working in single-cell RNA-seq-based cancer immunotherapy research at pRED Basel (Pharmaceutical Sciences, BiOmics and Pathology, Bioinformatics & Exploratory Data Analyis)

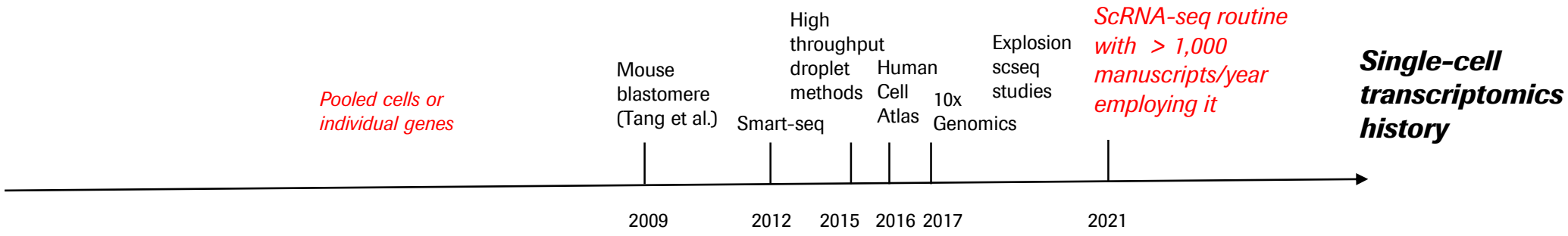
07. 05. 2021, Guest Lecture Uni Basel *"Mathematical and Computational Biology In Drug Discovery"*



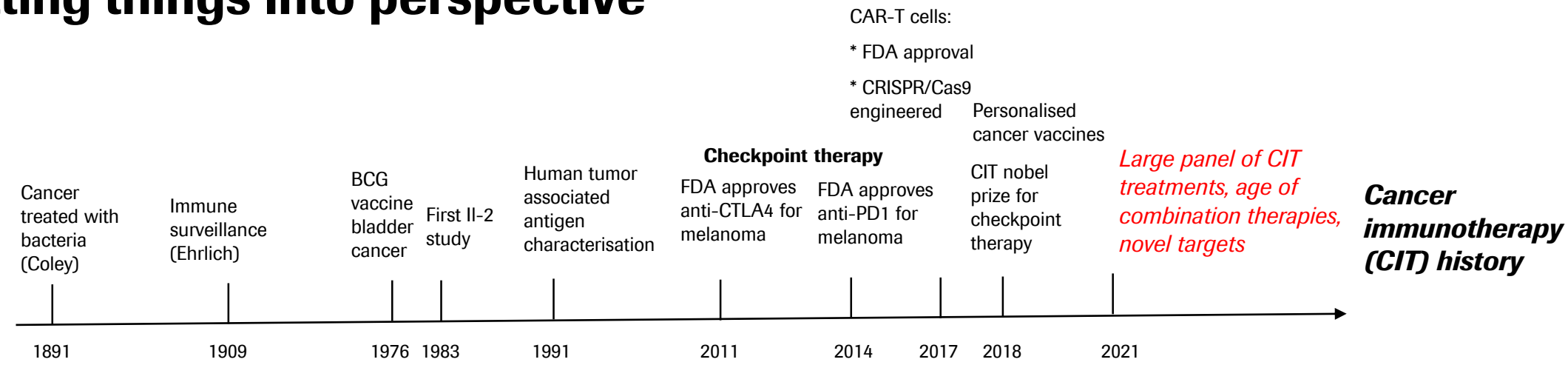
Putting things into perspective

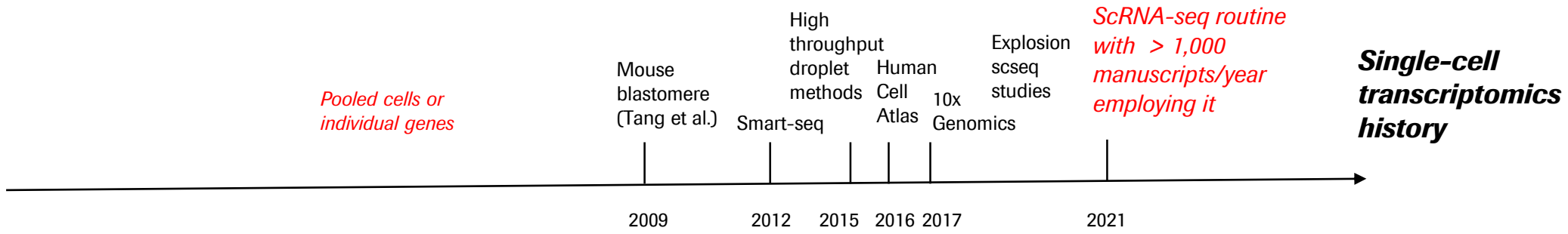


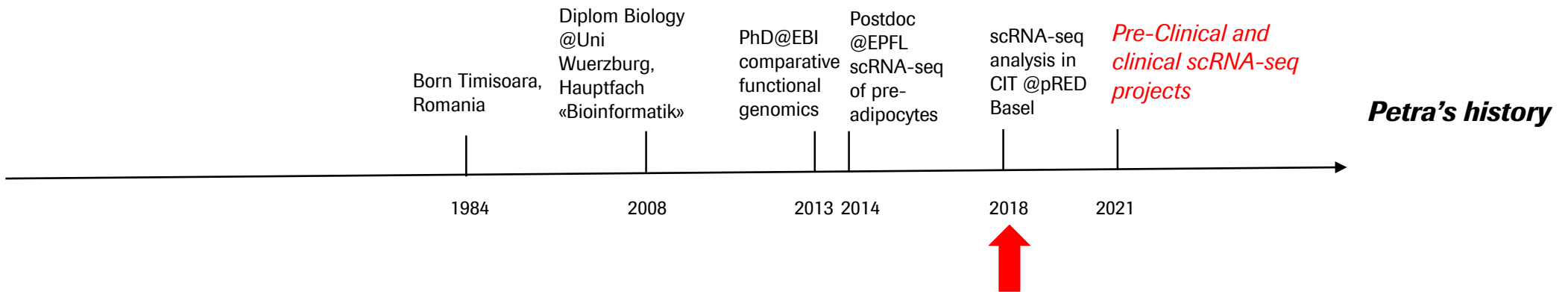
Putting things into perspective



Putting things into perspective







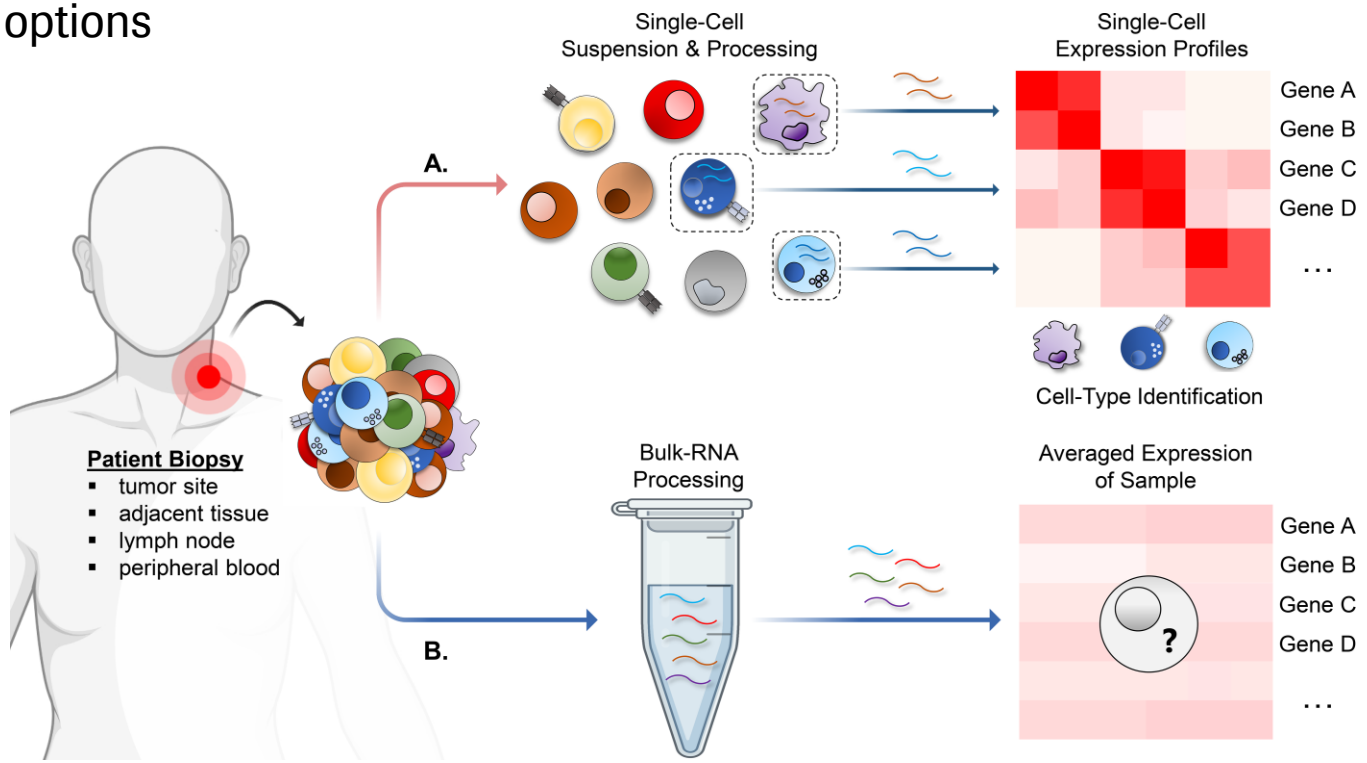
Cancer immunotherapy has been around for a while, what is the issue?

- Most of the patients treated with CIT do not respond or eventually relapse
- Adverse events of CIT treatment (up to 30% severe)
- Explosion of immunotherapies & combination options
- Cost and technical hurdles
- Targeting the tumor is challenging,
  especially “immune desert” ones
- Biological challenges:
  - Tumor microenvironment is complex
  - Immune system is complex

Cancer immunotherapy has been around for a while, what is the issue?

- Most of the patients treated with CIT do not respond or eventually relapse
- Adverse events of CIT treatment (up to 30% severe)
- Explosion of immunotherapies & combination options
- Cost and technical hurdles
- Targeting the tumor is challenging,
  especially “immune desert” ones
- Biological challenges:
  - Tumor microenvironment is complex
  - Immune system is complex

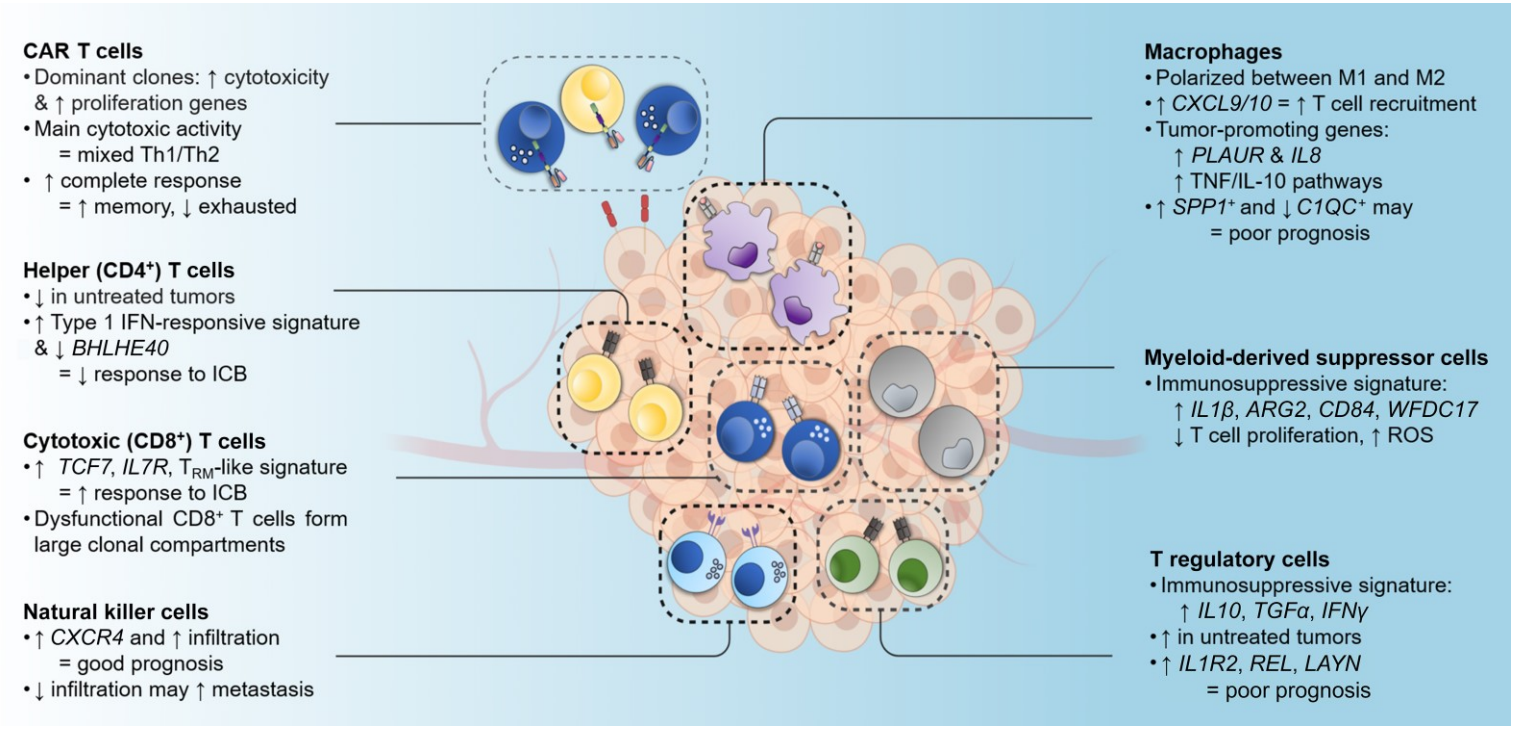
How can single-cell RNA-seq help?



Guruprasad et al. JEM 2020

What did single-cell RNA-seq already deliver in this context?

- better understanding of immune cell types (among others) and their characteristics
- unique transcriptional programs in the tumor microenvironment
- large heterogeneity inside single tumors and from one tumor to another
- revealed possible CIT response predictors and resistance mechanisms

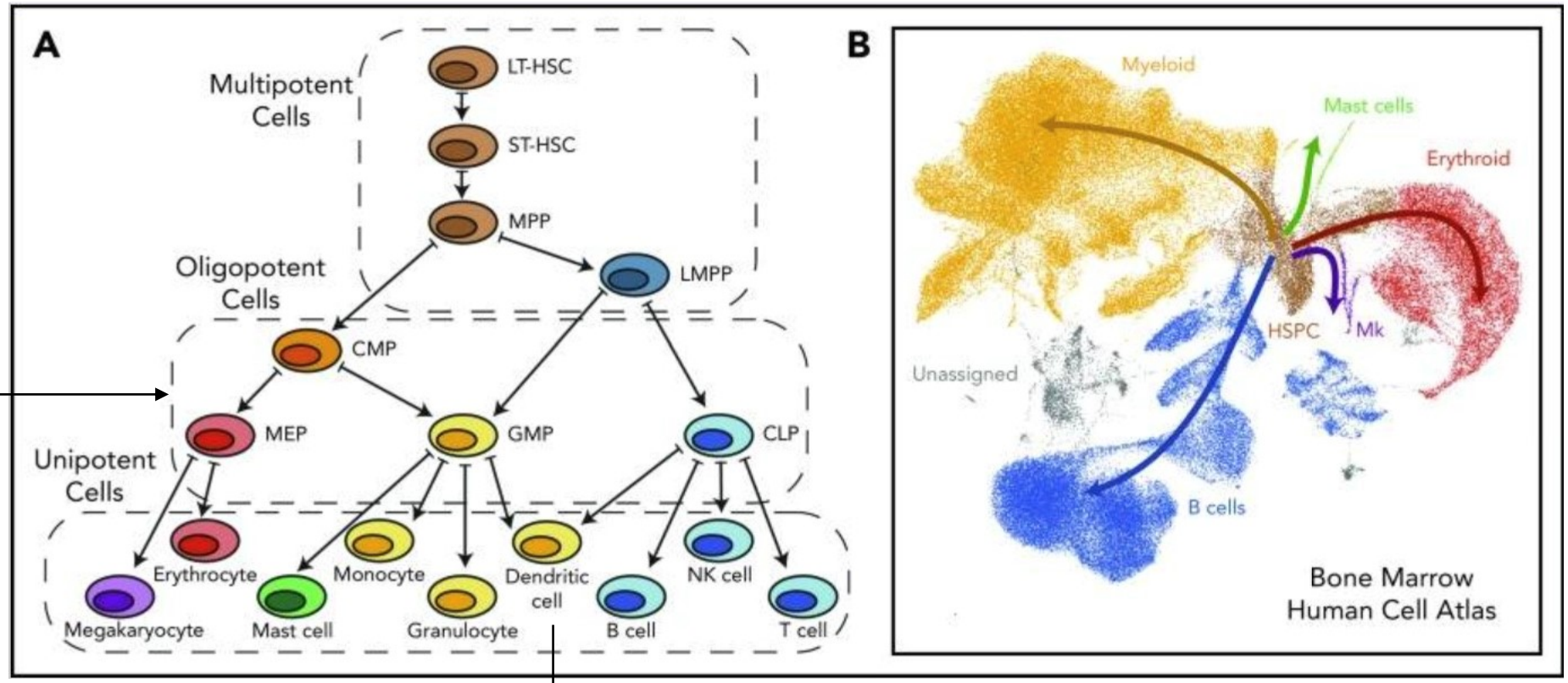


Guruprasad et al. JEM 2020

scRNA-seq reshaped parts of the hematopoietic tree

Continuous differentiation landscapes, with little or no discrete differentiation stages and smooth transitions across the cell states

Unipotent (cells dominate the middle tier of hematopoietic progenitor,



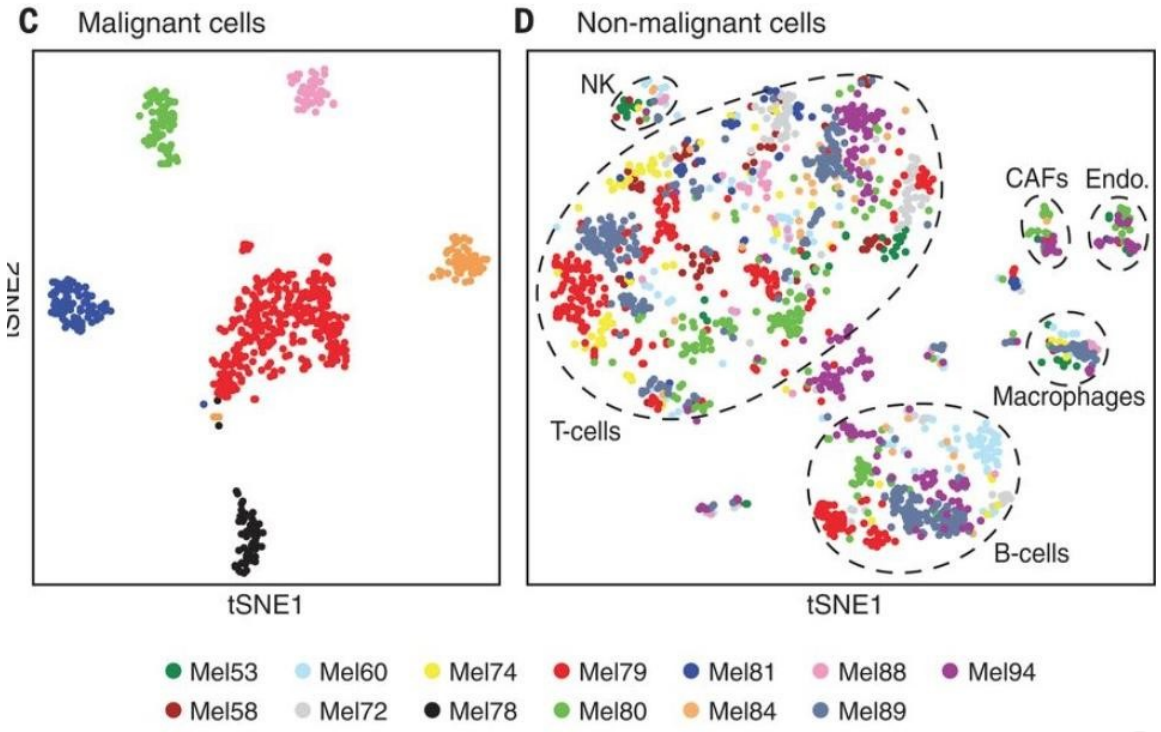
High resolution DC subtypes

Watcham et al. Blood 2019

scRNA-seq in melanoma

[Tirosh et al. Science 2016]

- Reference study for scRNA-seq in ONC, despite large limitations (<5,000 cells across 19 patients), provided many long-standing insights:
- **Inter-patient variability highly different malignant cells vs. non malignant cells**:
  - Malignant cells - individual patient-specific clusters
  - Immune, endothelial cells and fibroblasts cluster per cell types
- Malignant cells in same patient are heterogenous, with main axes:
  - cell cycle
  - spatial context (not malignant-specific)
  - drug-resistance programs
- **A gradient of T cell phenotypes** rather than discrete cell subtypes/states: naïve/memory to cytotoxic/activated/exhausted, high **expression of coinhibitory receptors by most cytotoxic cells in the tumor**

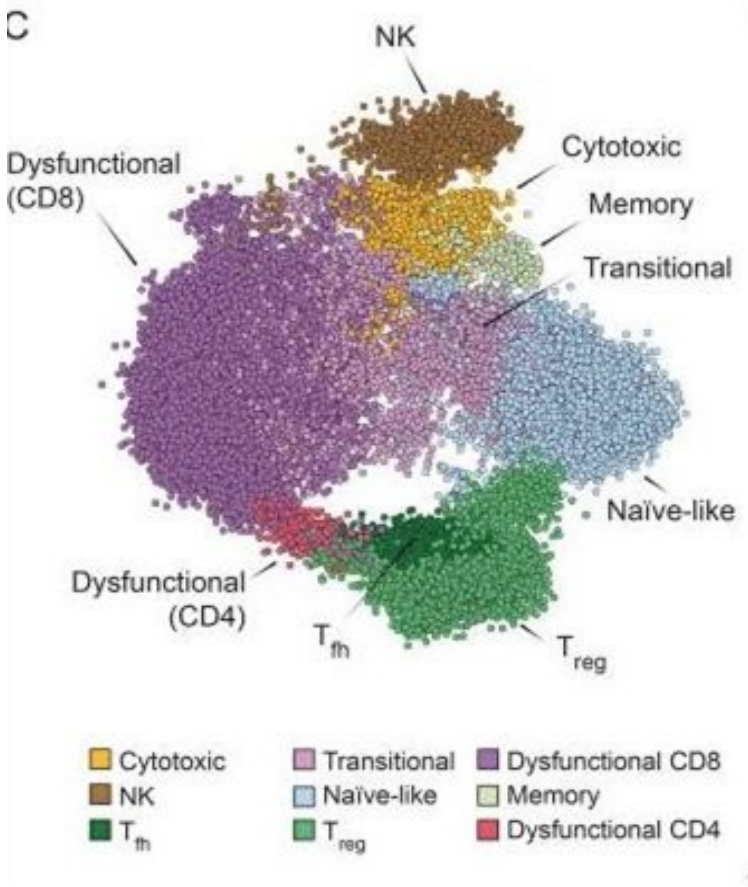


scRNA-seq in melanoma

[Li et al. Cell 2019]

Higher resolution insight into the gradient of T cell phenotypes

- Focused on T cell characterization, large nr. of cells (46,612 immune cells; 29,825 T cells, 25 patients)
- Immune cell subtypes largely shared but relative abundance highly variable across patients
- Conserved trajectories of CD8 T cells
  - **Dysfunctional T cells**
    - part of a wide differentiation spectrum
    - highest levels of T cell expansion
    - ongoing proliferation
    - inhibitory expression receptor
    - likely drivers of tumor reactivity (*ex vivo* experiments, Ifng and TNFa secretion)
  - **Cytotoxic T cells** - not linked to dysfunction, not proliferative, more abundant PBMCs, less tumor reactive *ex vivo*



So what do we actually do? Where do we employ scRNA-seq?

External data - e.g. healthy and disease
Some internal data (mainly disease)

Disease-specific cell types and states?

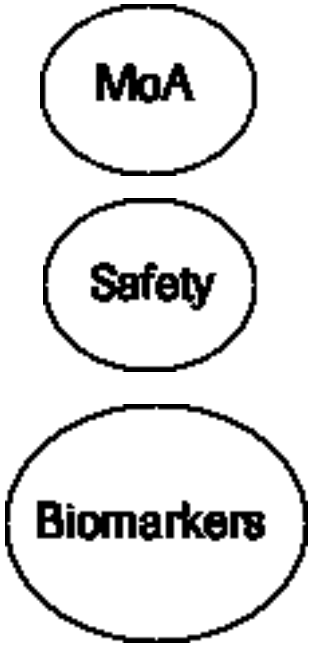
Genes/pathways specifically (co)-expressed in disease context?

Heterogeneity across individuals (target expression, pathway expression, cell types)?

Subclasses of individuals?

So what do we actually do? Where do we employ scRNA-seq?

Mainly internal data: in vitro experiments (2D, 3D culture)
or in vivo experiments (typically mouse)



What are the changes induced by X in specific cell populations?

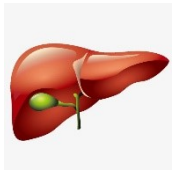
Which cell types react most and how?

Which are the most consistent changes across different model systems?

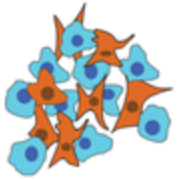
Which changes can we expect to see in the clinic?

So how does scRNA-seq work? Experimental workflow

Tissue dissociation

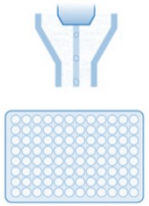


Solid tissue



Dissociated cells/nuclei

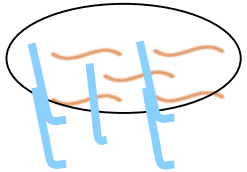
Single cell capture – transcriptome sequencing



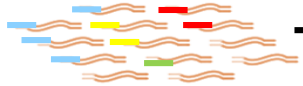
Flow cytometry
96-384 cells

Throughput

Single cells/nuclei



cDNA



Pooled cDNA libraries

NGS

Cell/nuclei suspension

Droplet microfluidics
1000s of cells

Typically highly standardised/automatised platforms applied internally (e.g. 10x Genomics Chromium)

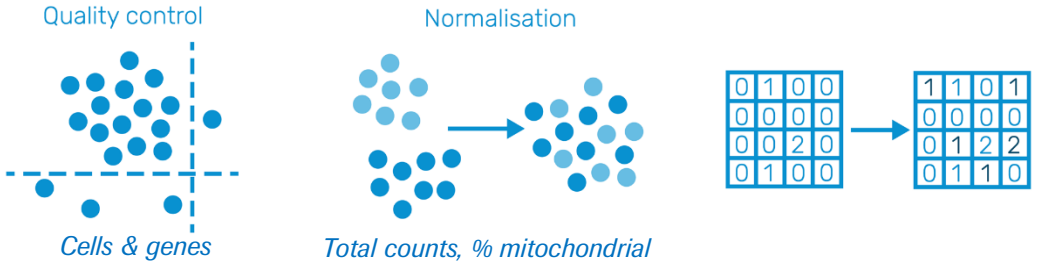
Adapted from Megana Prasad, pRED

So how does scRNA-seq work? Computational workflow

Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol*. 2019 Luecken & Theis.

From short reads to gene-cell matrix

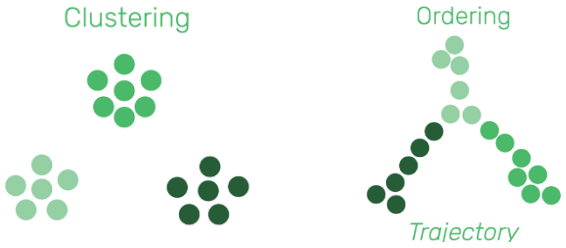
QC, filtering & normalisation



Strive for automatisation & standardisation, F.A.I.R. data (findable, accessible, interoperable, reusable)

Dimensionality reduction & visualization

Clustering & cell assignment



Downstream analysis

Adapted from Zapia et al. PLoS Computational Biology 2018

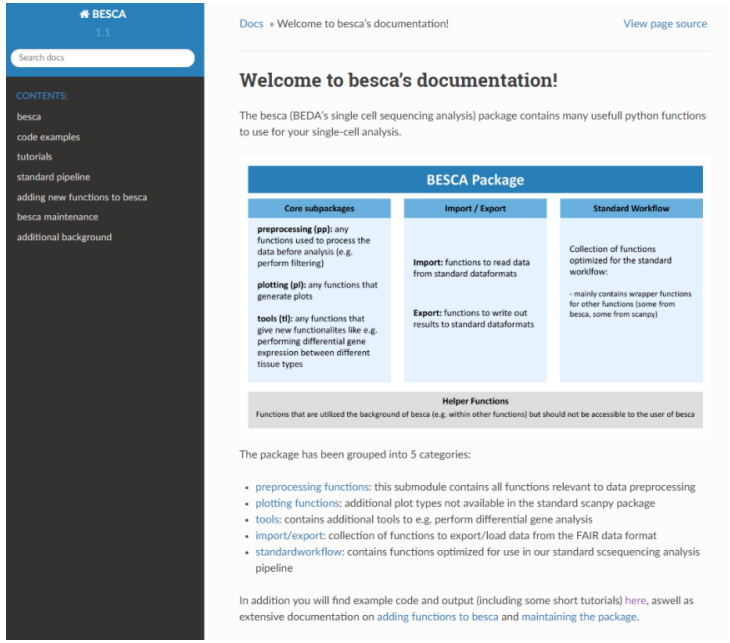
How do we work?

- Reproducible & standardized scRNA-seq analysis *!but also ability to customise per needs*
- Flexible workflow and usage of state-of-the art methods
- Keeping up to date with new developments both on the experimental and methodological side
  - Internship projects (Master & PhD level)
    - cell annotation
    - cell deconvolution
    - perturbation prediction
    - RNA to protein prediction
    - ligand-receptor interactions

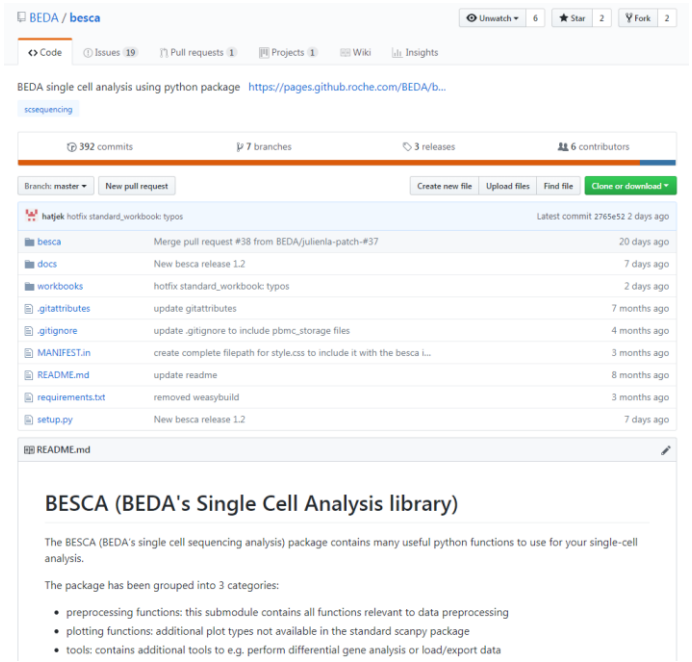
Code sharing and automatization

Besca (BEDA's Single Cell Analysis Toolkit) to streamline single cell transcriptomics analyses

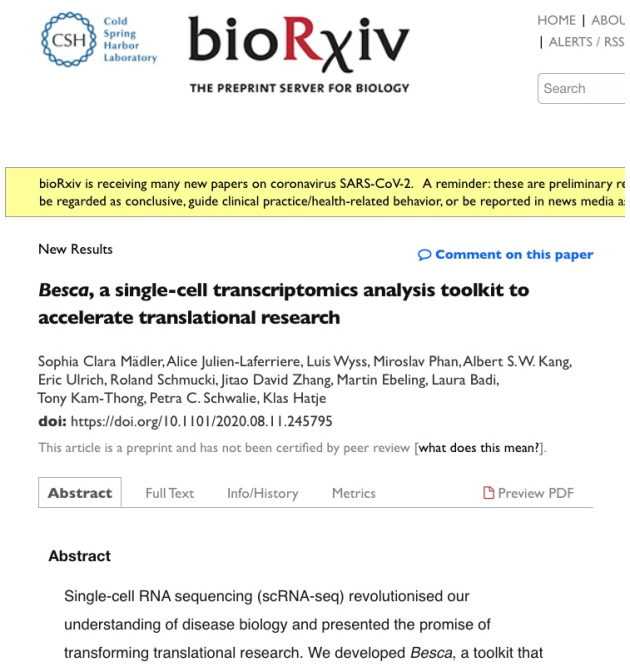
Besca documentation page



Besca on github



Besca preprint

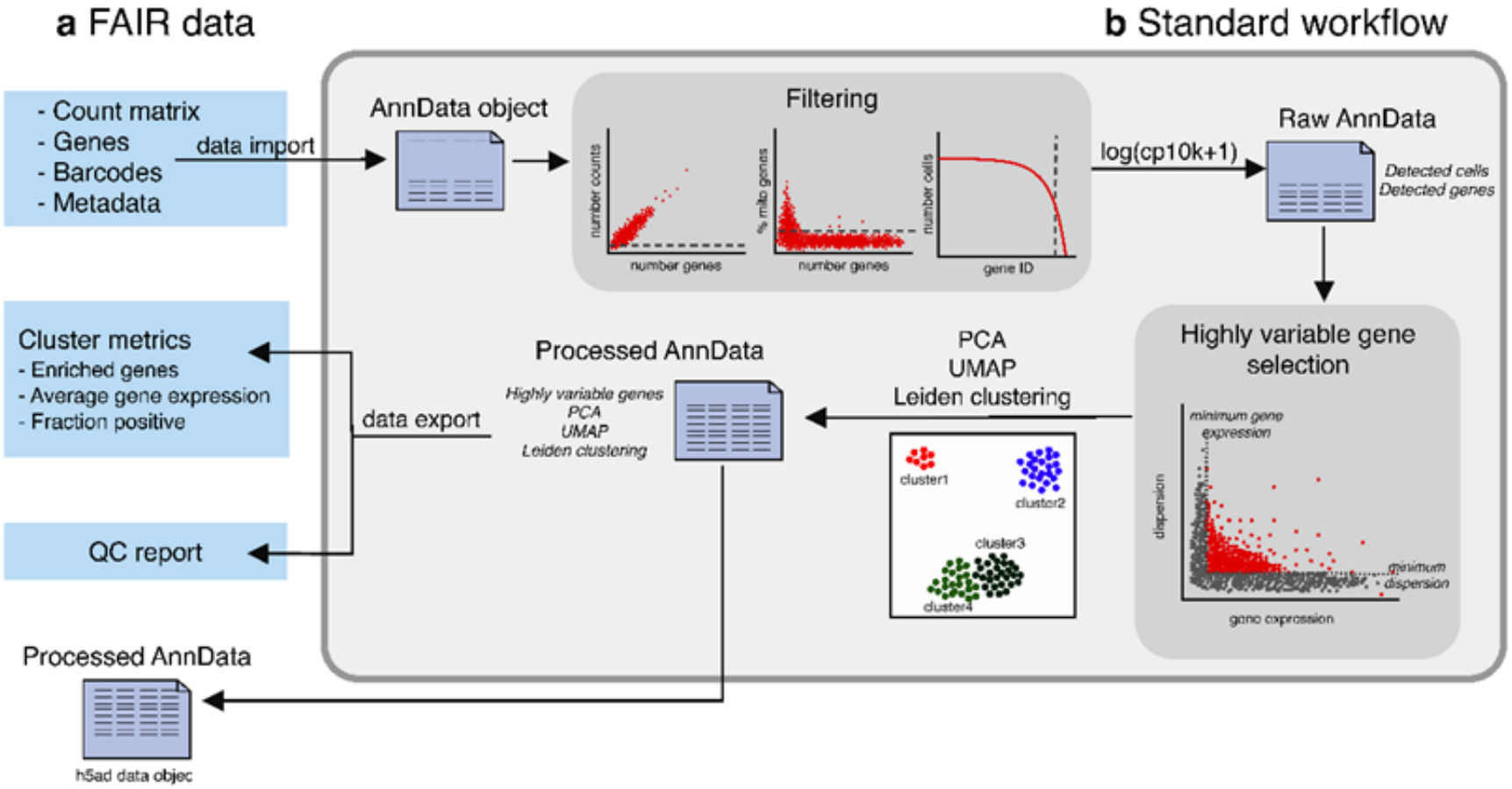


Sophia Clara Mädler, Alice Julien-Laferriere, Luis Wyss, Miroslav Phan, Albert S. W. Kang, Eric Ulrich, Roland Schmucki, Jitao David Zhang, Martin Ebeling, Laura Badi, Tony Kam-Thong, Petra C. Schwalie, **Klas Hatje**

Please find additional information here: https://github.com/bedapub/besca | Maedler et al. *bioRxiv*. 2020

A standard workflow as part of Besca (BEDA's Single Cell Analysis Toolkit)

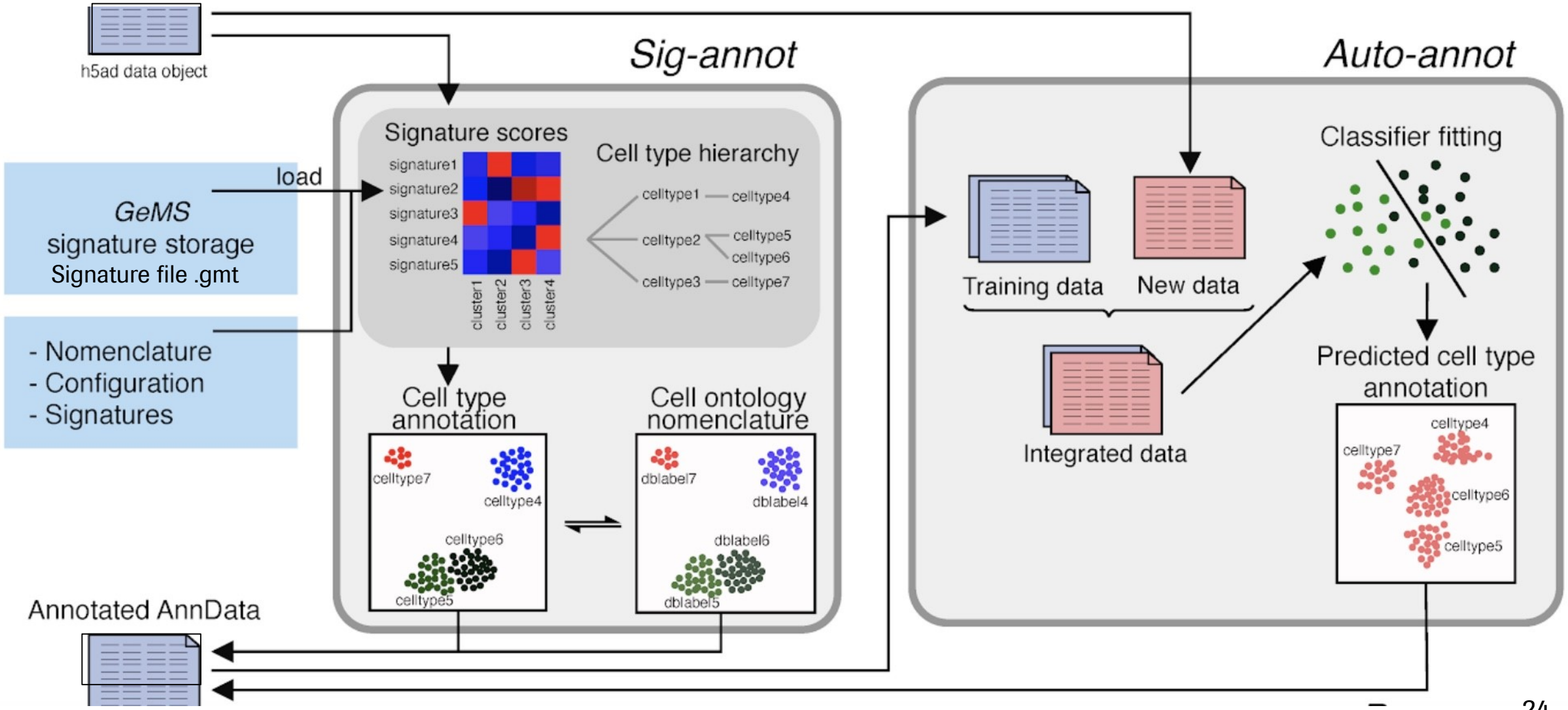
A toolkit to streamline single cell transcriptomics analyses according to current best practices



Source: Maedler et al. *bioRxiv*. 2020 | https://github.com/bedapub/besca | https://github.com/bedapub/bescape

Automatised cell annotation as part of Besca (BEDA's Single Cell Analysis Toolkit)

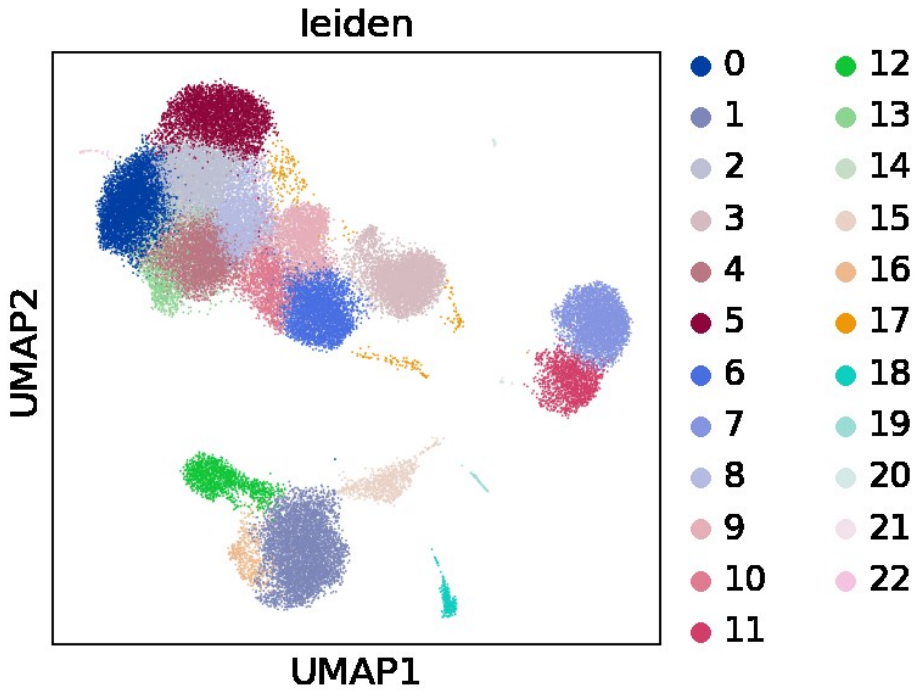
Addressing one of the bottlenecks in scRNA-seq analysis



Source: Maedler et al. *bioRxiv*. 2020

Why cell type annotation?

- How does treatment X affect B cells?
- Is the abundance of regulatory T cells different upon treatment?



Kotliarov et al. 2020 PBMC dataset, 20 individuals *(reanalysis besca by Sophia Maedler)*

Why cell type annotation?

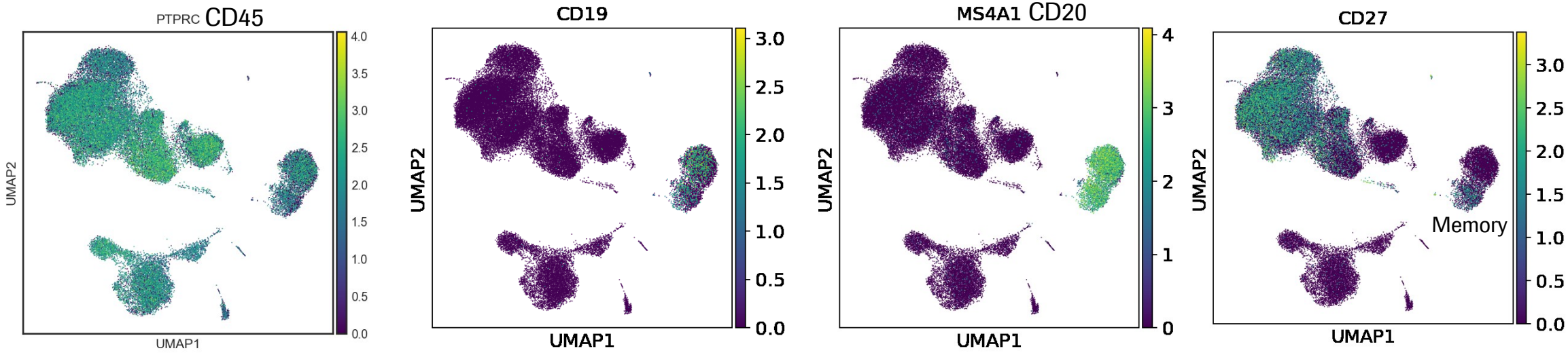
- scRNA-seq is unbiased => gene expression profile of all cells in a sample

  (unlike FACS, in scRNA-seq cell types are typically not predefined)

- Cell types need to be recognized based on their transcriptome
  - "easy" → "complex" task (e.g. B cells vs. T cells → memory vs. effector CD8+ T cells)
  - requires
    - an understanding of the **universe of cell types** (and/or states)
      - a common vocabulary of cell types (CELL ONTOLOGY)
      - an agreement of characteristic cell types features (at transcriptome level) (SIGNATURES/MARKERS/REFERENCE GENE EXPRESSION)
    - a way to **match the transcriptome** to that understanding
      - on a cell per cell basis
      - on a cluster basis

Challenge III: Mapping single cells to a reference atlas ("Eleven grand challenges in single-cell data science" Genome Biol 2020)

Cell type annotation “v1” – manual marker-based, per cluster



Immune cell

B cell

Naive B-cell: CD27 low

Memory B-cell: CD27 high

Annotation from high level cell types down to specific cell types

Limitations

- Laborious => does not scale to processing a large number of samples
- Requires expert knowledge & is prone to bias & error
- Poorly reproducible across studies
- Typically do not cover continuous cell states

“v2” – automatic marker-based annotation

The manual process can be automatized: marker enrichment => call clusters/cells algorithmically

- Implemented in besca as Sig-Annot
- Publicly available tools such as SCINA, DigitalCellSorter, CellAssign

“v3” – similarity of gene expression to reference

Use previous **reference data** (e.g. bulk gene expression of sorted cells) & assign cell types based on similarity (e.g. Spearman correlation, cosine distance)

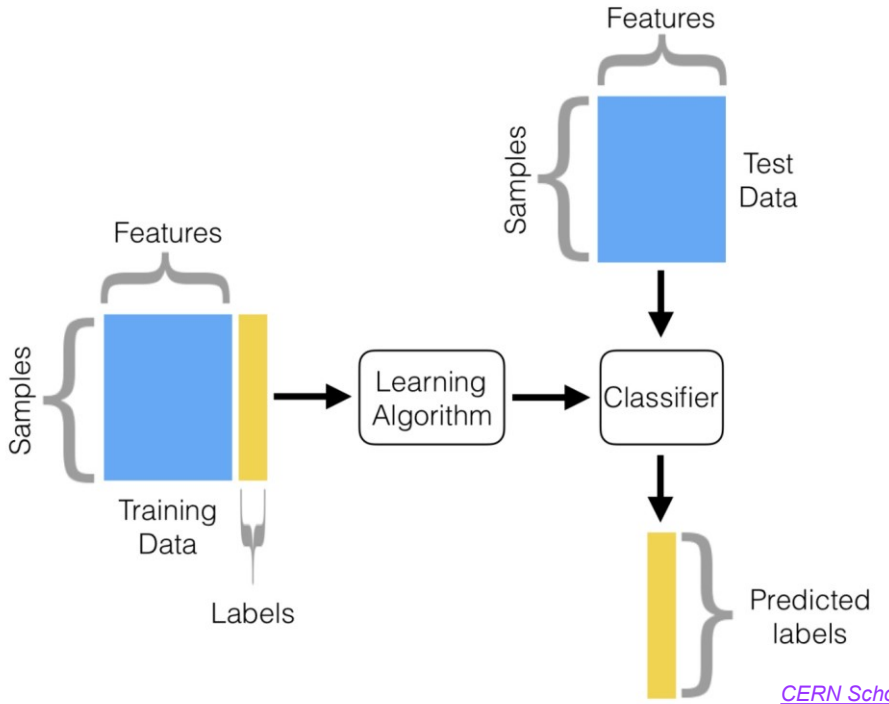
- Publicly available tools e.g. singleR (hierarchical process, large collection of reference datasets), scMatch (can use ontology), CHETAH (hierarchical), scMCA (mouse cell atlas-specific), scmapCell

Cell type annotation “v4” – machine-learning based

Luis Wyss

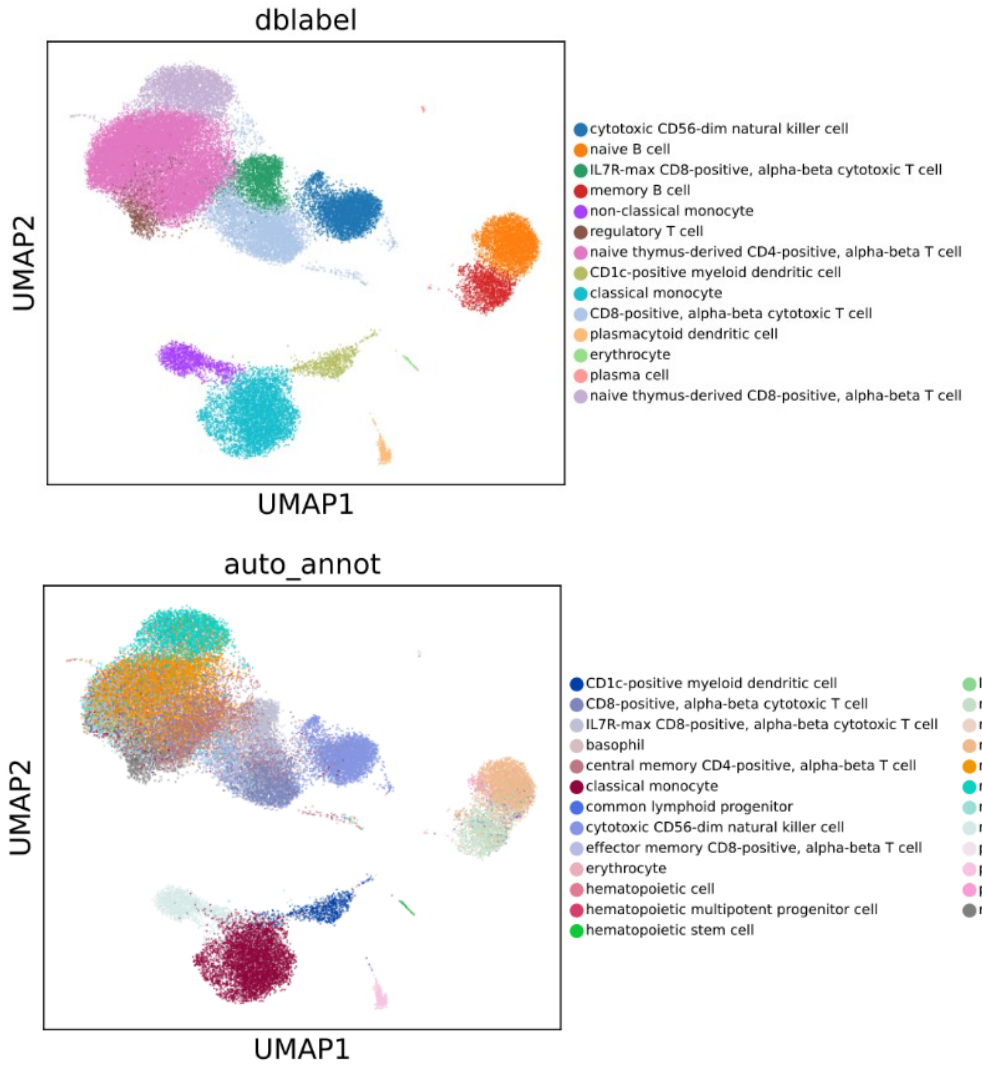
Use previous data to **train a model to automatically “learn”** characteristics of a cell & predict the type of a new one

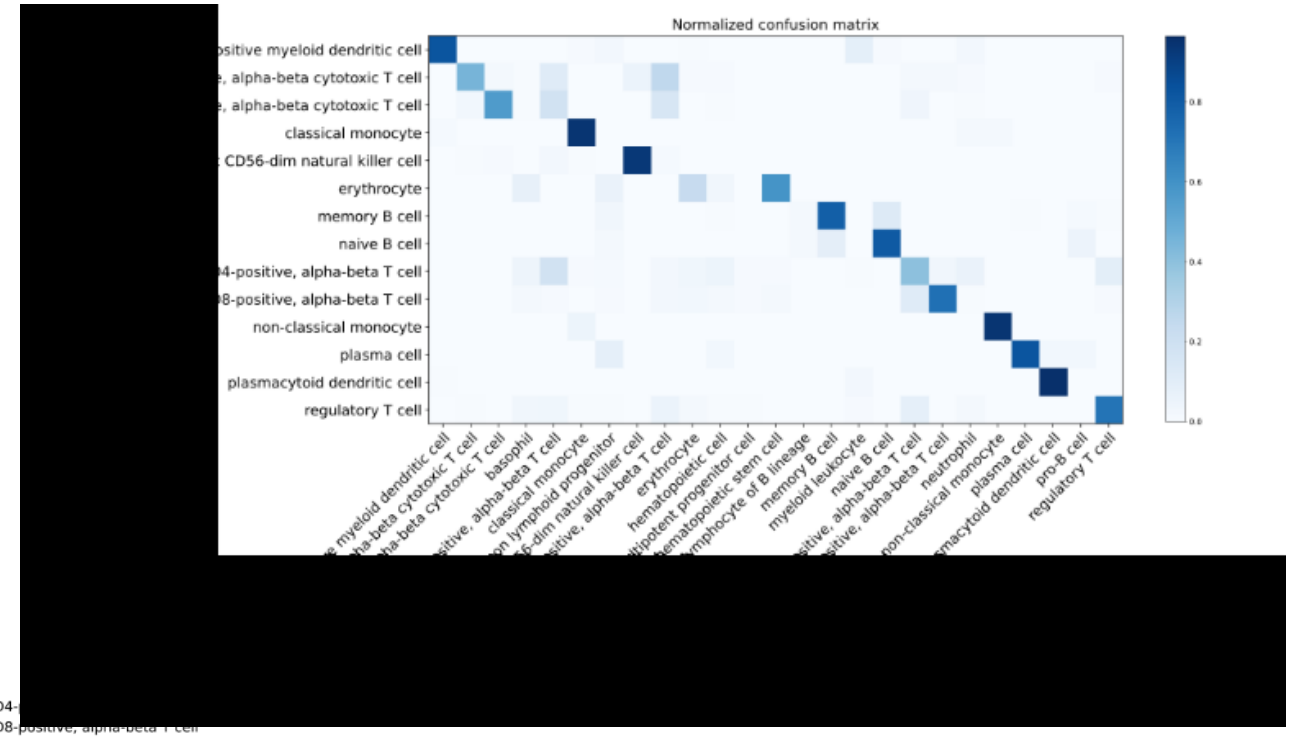
- linear discriminant analysis (*LDA – e.g.* scID), support vector machine (*SVM – e.g.* scPred), random forest (*RF – e.g.* Seurat , SingleCellNet), Xgboost (e.g. CASTLE)
- **Implemented in besca as Auto-Annot (SVM, logistic regression)**
- Deep learning tools – e.g. scANVI (variational inference), MARS (meta-learning; claimed to generalize well to unannotated experiments and to identify cell types that were never seen during training)



Cell type prediction in PBMCs

Luis Wyss





Source: Maedler et al. *bioRxiv*. 2020

Dataset from Granja et al 2019 was used to predict cell types in Kotliarov et al 2020.

How do we work?
Engage with the single cell community outside Roche

- **Open source** libraries
  - Besca (https://github.com/bedapub/besca)
  - Bescape (https://github.com/bedapub/bescape)

- **Human cell atlas** contribution
  - Besca preprint (DOI: 10.1101/2020.08.11.245795)

- Student **interns**
  - Stephan, Gregor, Sophia, Albert, Luis, Miro, Eric, Andreea, Anthony, Mariia, Demeter

- Academic **guest visits**

  Sarah Teichmann (Sanger Institute, Cambridge)

  Fabian Theis (Helmholtz Center, Munich)

  Niko Beerenwinkel (ETH Zurich, Basel)

- **EMBL-EBI Industry Programme Workshop** on Understanding Single Cell Atlases

Acknowledgements

Single-cell squad, BEDA colleagues and all collaborators

- **Klas Hatje**
- **Sophia Maedler**
- **Tony Kam-Thong**
- **Martin Ebeling**
- **Nikolaos Berntenis**
- **Anthony Sonrel**
- **Miroslav Phan**
- **Luis Wyss**
- **Eric Ulrich**
- **David Jitao Zhang**
- **Alex Seitz**
- **Llucia Alberti Servera**
- **Roland Schmucki**
- **Marco Berrera**
- **Juliane Siebourg-Polster**
- **Laura Badi**
- **Fabian Birzele**
- Isabelle Wells
- Javier Gayan
- Daniel Marbach
- Chia-Huey Ooi
- Andreea Ciuprina
- Bhavesh Soni
- Mariia Bilous
- Christian Holland
- Alberto Valdeolivas Urbelz
- Said Aktas
- Tai-Hsien Ou Yang
- Ying He
- Andreas Roller
- Marina Bacac
- Sylvia Herter
- Ramona Schlenker
- Alison Ribeiro
- Jitka Somandin
- Tamara Huesser
- Vinko Tosevski
- Emilio Yangueez
- Sarah Diggelmann
- Christian Klein
- Laura Codarri Deak
- Iryna Dekhtiarenko
- Stephan Gasser
- Mi He
- Steffen Dettling
- Sabine Hoves
- Sebastian Fuchs
- Andreas Frei
- Malte Steinberg
- Floriane Point
- Marisa Mariani
- Inga Clausen
- Nadine Kumpesa
- Megana Prasad
- Jelena Kuehn Georgijevic
- Olivia Spleiss
- Claudia Bossen
- Alexia Phedonos

+ many more colleagues providing data & advice!

Doing now what patients need next

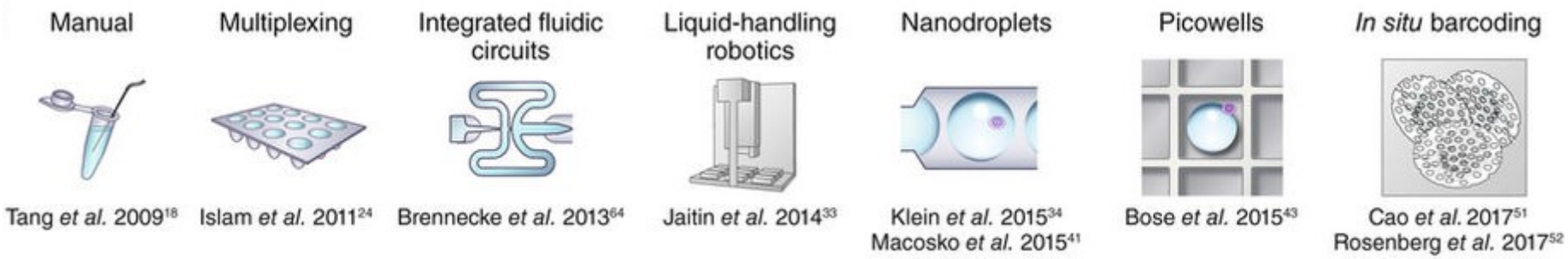
References

- Gohil et al. Nat. Reviews Clinical Onc (https://www.nature.com/articles/s41571-020-00449-x)
- Waldman et al. Nat. Reviews Immunol (https://www.nature.com/articles/s41577-020-0306-5)
- Guruprasad et al. JEM 2020 (https://rupress.org/jem/article/218/1/e20201574/211615/The-current-landscape-of-single-cell)
- Watchman et al. Blood 2019 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6440294/)
- Laehneman et al. Genome Biol. 2020 (https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-1926-6)
- Kiselev et al. Nat Rev Genetics 2019 (https://www.nature.com/articles/s41576-018-0088-9)
- Andrews et al. Mol. Aspects Medicine 2018 (https://www.sciencedirect.com/science/article/pii/S0098299717300493) Identifying Cell Populations with RNA-seq
- Sversson et al. Nat. Protocols 2018 (https://www.nature.com/articles/nprot.2017.149)
- Macaulay & Voet, PLoS Genetics 2014 (https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1004126)
- Scanpy: Wolf et al. Genome Biol. 2017 (https://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1382-0)
- Human Cell Atlas https://www.humancellatlas.org/
- Hicks et al. Biostatistics 2018 (https://academic.oup.com/biostatistics/article/19/4/562/4599254)
- On a cell's identity – Wagner et al. Nat Biotech 2016 (https://www.nature.com/articles/nbt.3711)

BACKUP

Single-cell RNA-seq challenges

* **Cell Capture:** throughput, automation, cell stress

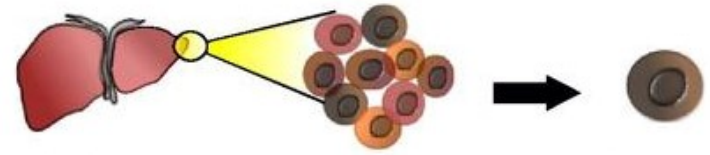


* **Small quantities:** obtain enough material for an accurate readout without introducing biases

* **Data analysis & interpretation:** sparseness, noise, high dimensionality, batch effects, doublets, …

⇒ Gene 'dropouts' (readouts of 0) in which a gene is observed at a moderate expression level in one cell but is not detected in another cell

Experimental workflow



(micromanipulation, laser-capture microdissection, FACS, microdroplets, microfluidics)

Single Cell Isolation

Cell lysis

RNA

cDNA

Inclusion of:

- *unique molecular identifiers (UMIs) for reducing technical noise*
- *cell-specific barcodes for multiplexing*

Reverse transcription & second-strand synthesis

Amplified cDNA

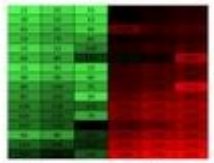
Amplification *(PCR or in vitro transcription)*

Library preparation

Sequencing Library

CAAGTTCCTACAGCTA
AGTCCATGCCCATCCG
AATCGGACTTCAGCCT
GACCTAAGCCATCAGA
AATCCTAGCATCCAGC
ACCGTTACATCAACAG
ATTCGATAACGACCAT
CATGCCATTGACGATT

Sequencing

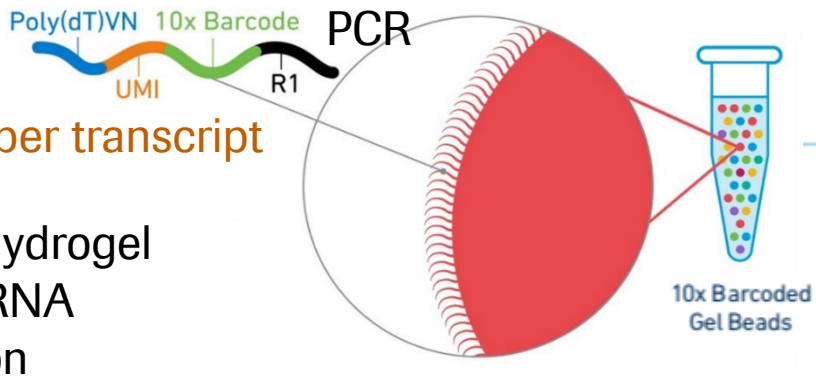


Single-cell Expression Profiles

adapted from WIkipedia

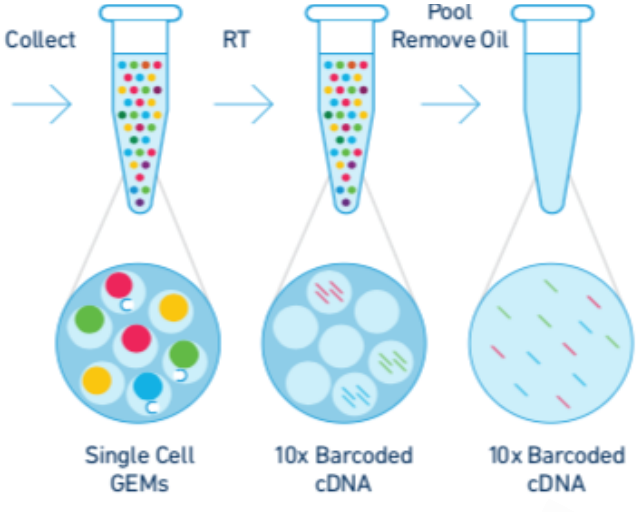
10X Genomics experimental workflow

Capture 1 per cell



Barcoded hydrogel beads – mRNA hybridisation

cells need to be dissociated from tissue before loading into 10x



poly(dT) sequence: catches the poly-A tail at the 3' end of transcripts

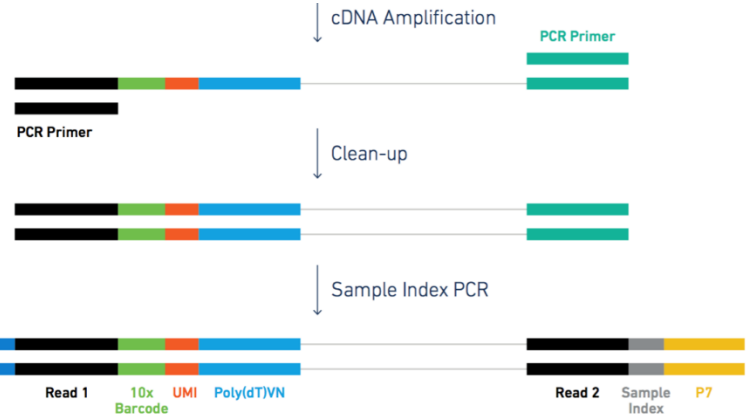
UMI (unique molecular identifier): unique barcode that labels each transcript

10X barcode: barcode that is consistent throughout a bead and marks that cell

sample index: barcode to identify sample

R1 sequence: adaptor sequence for PCR reactions/sequencing

R2 sequence: adaptor sequence for PCR reactions/sequencing



combine different samples and sequence

Doing now what patients need next