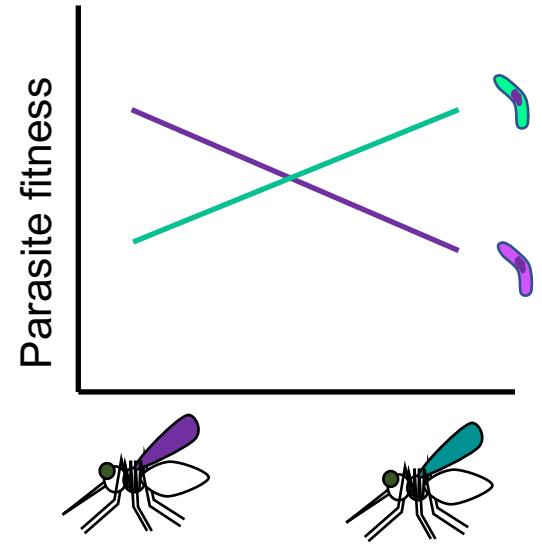
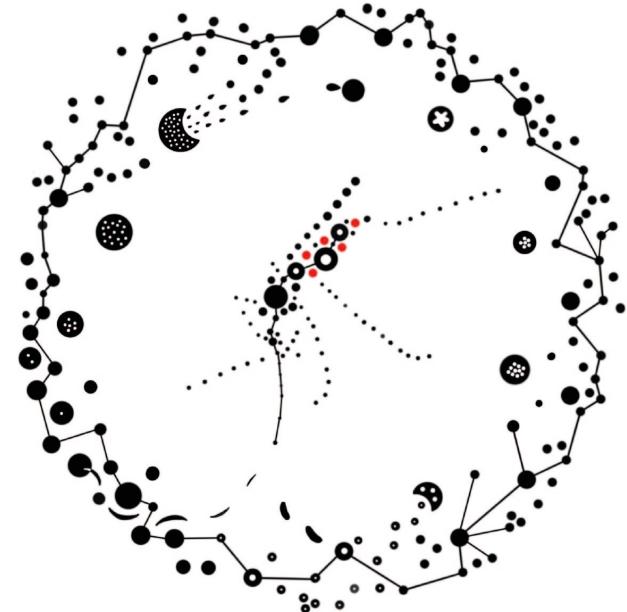


Data integration and visualisation in R for the malaria cell atlas

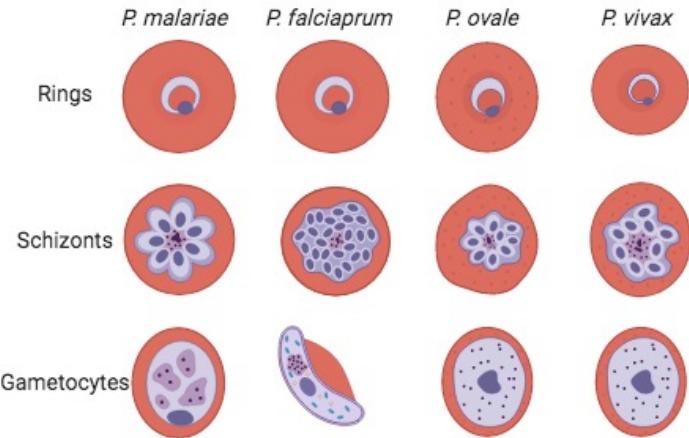
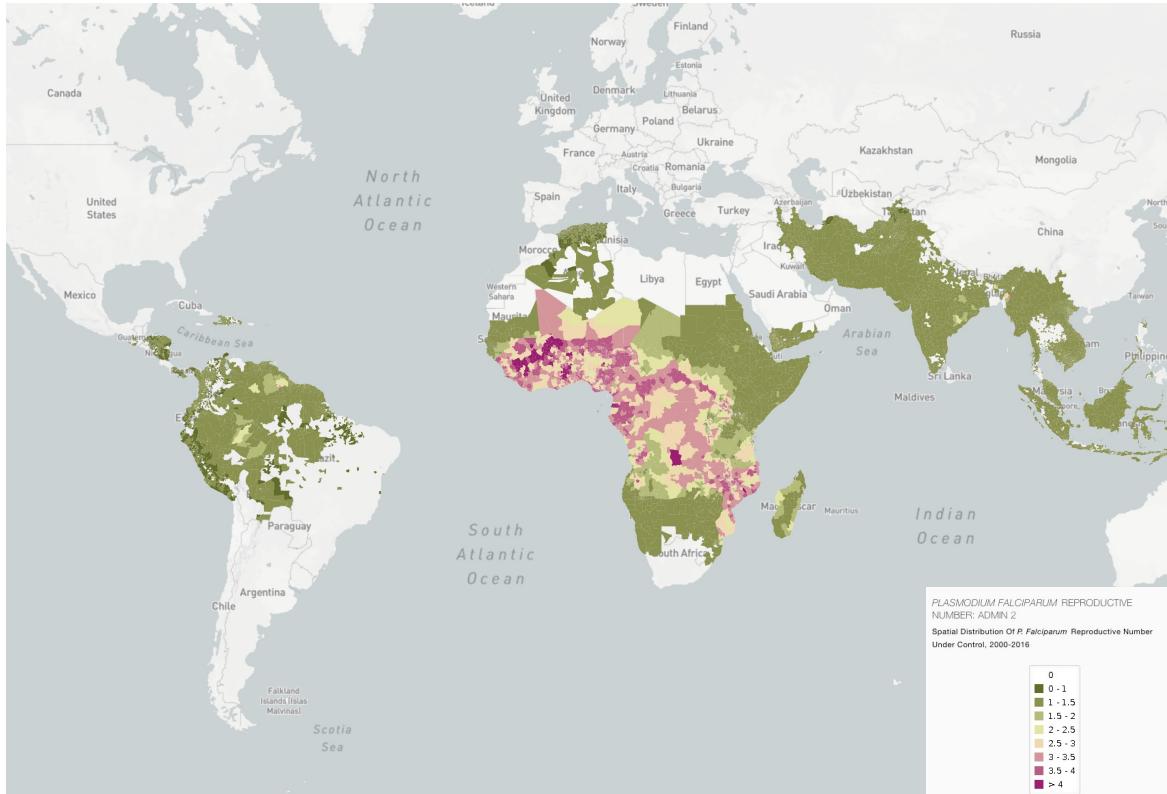
Virginia Howick

University of Glasgow
Biodiversity, Animal Health, Comparative Med
Wellcome Centre for Integrative Parasitology

African Bioinformatics Network
28 February 2022



Malaria is caused by a unicellular protozoan parasite



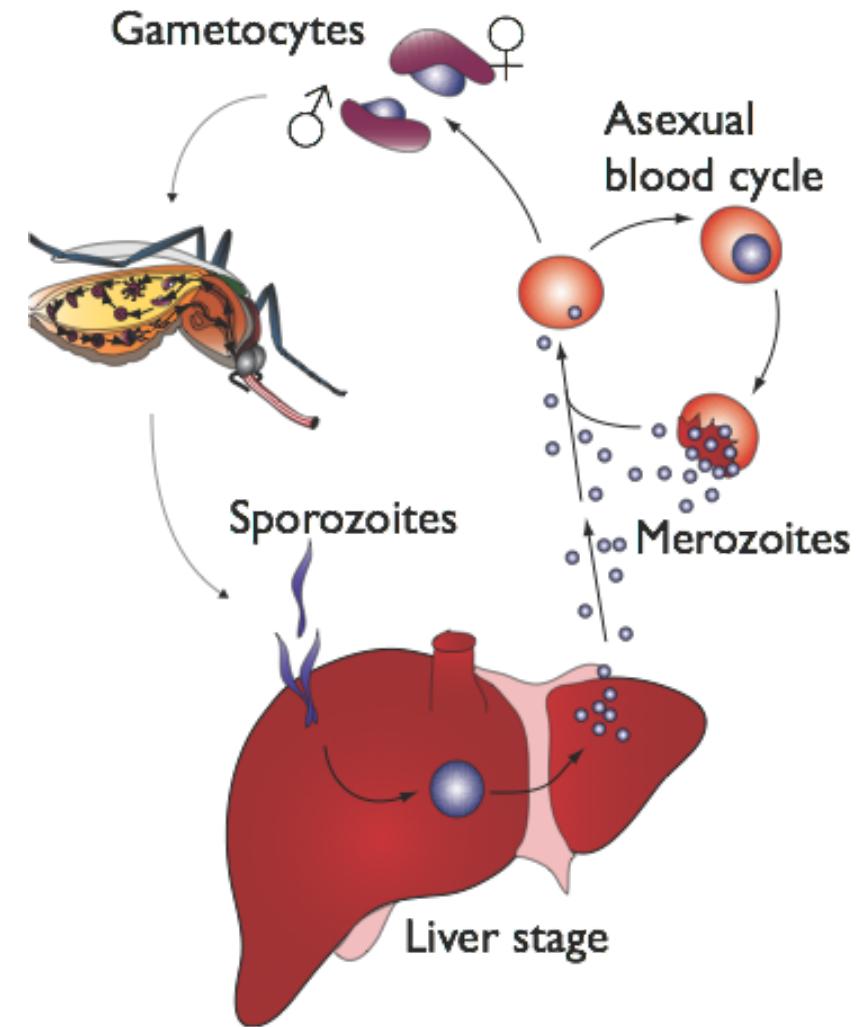
Malaria is caused by a unicellular protozoan parasite and is transmitted by mosquitoes in the genus *Anopheles*

Nearly half of the world's population is at risk of malaria with 241 million reported cases and over 600k deaths in 2020

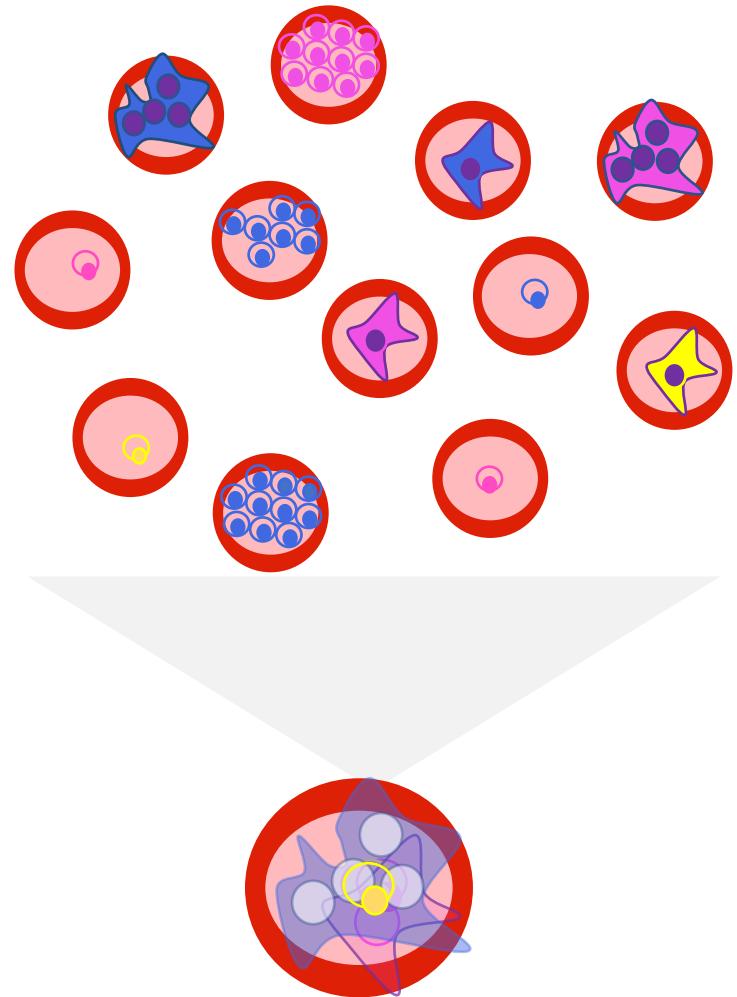


The *Plasmodium* life cycle

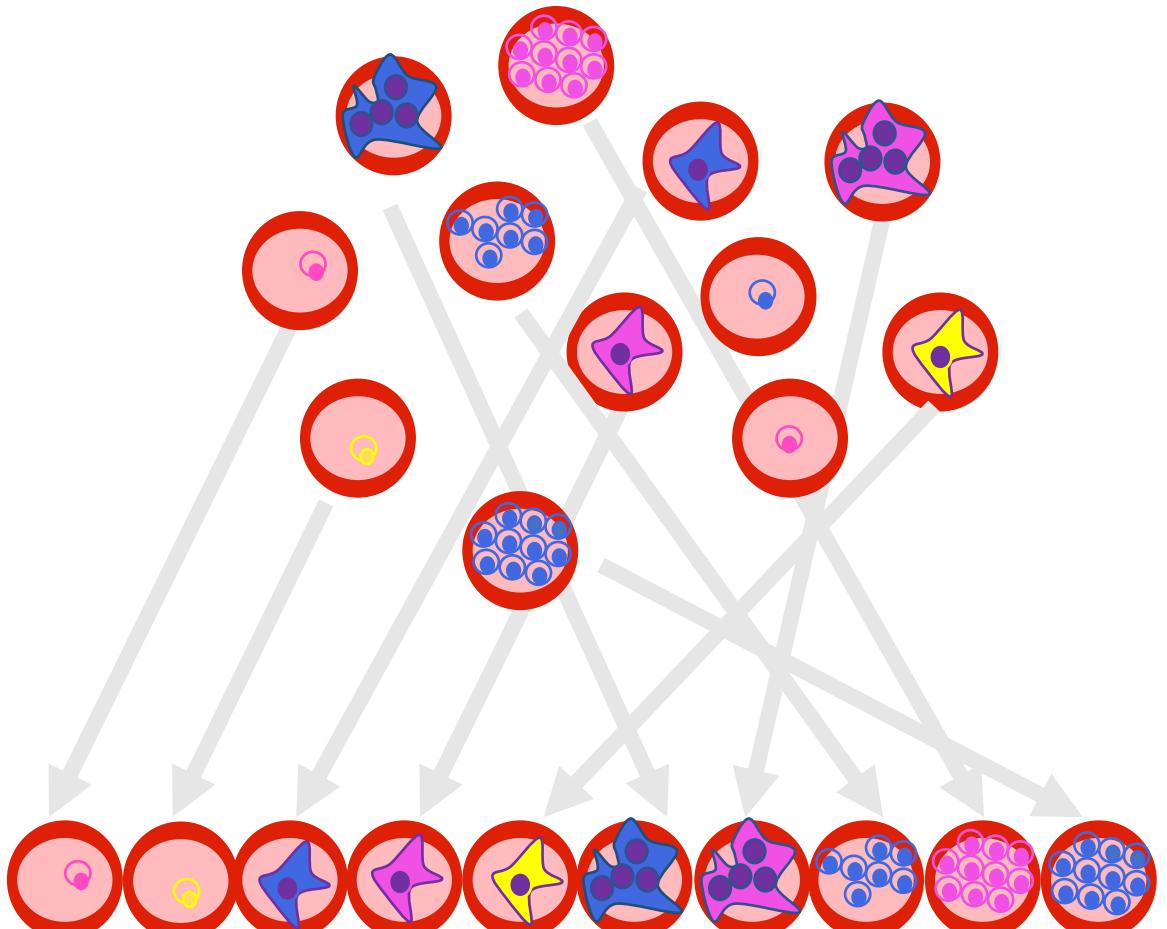
- The parasite develops in the mammalian liver and blood, as well as the mosquito midgut and salivary glands
- It differentiates into a series of morphologically distinct forms that alternate between invasive, replicative and sexual stages



Bulk



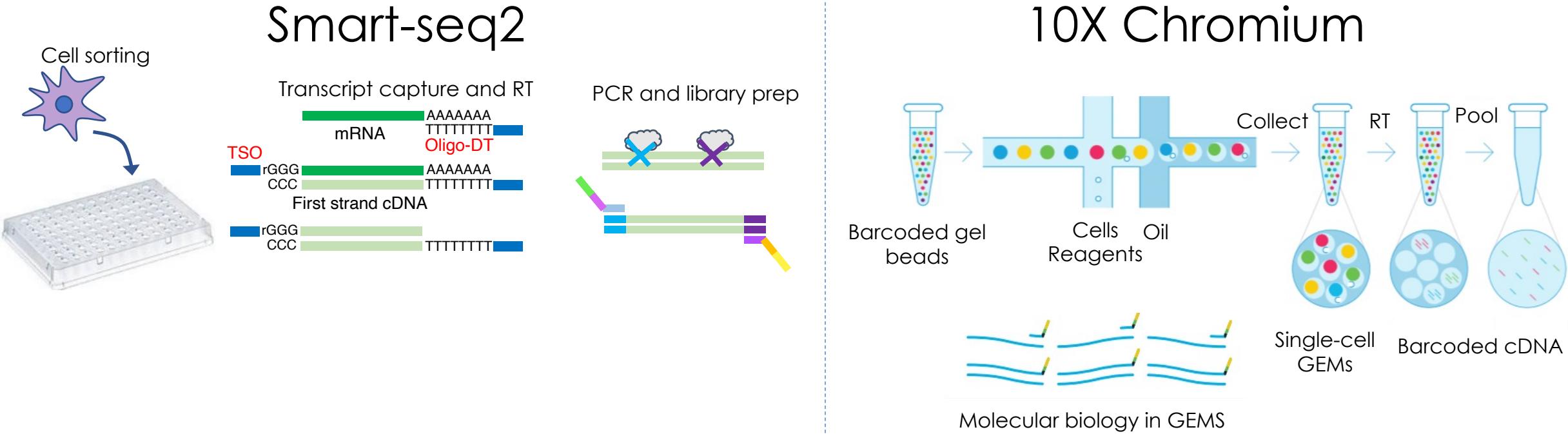
Single-cell



Single-cell genomics:

- Improves the temporal resolution of transcriptomes
- Allows for deconvolution of genotype from development and cell-state

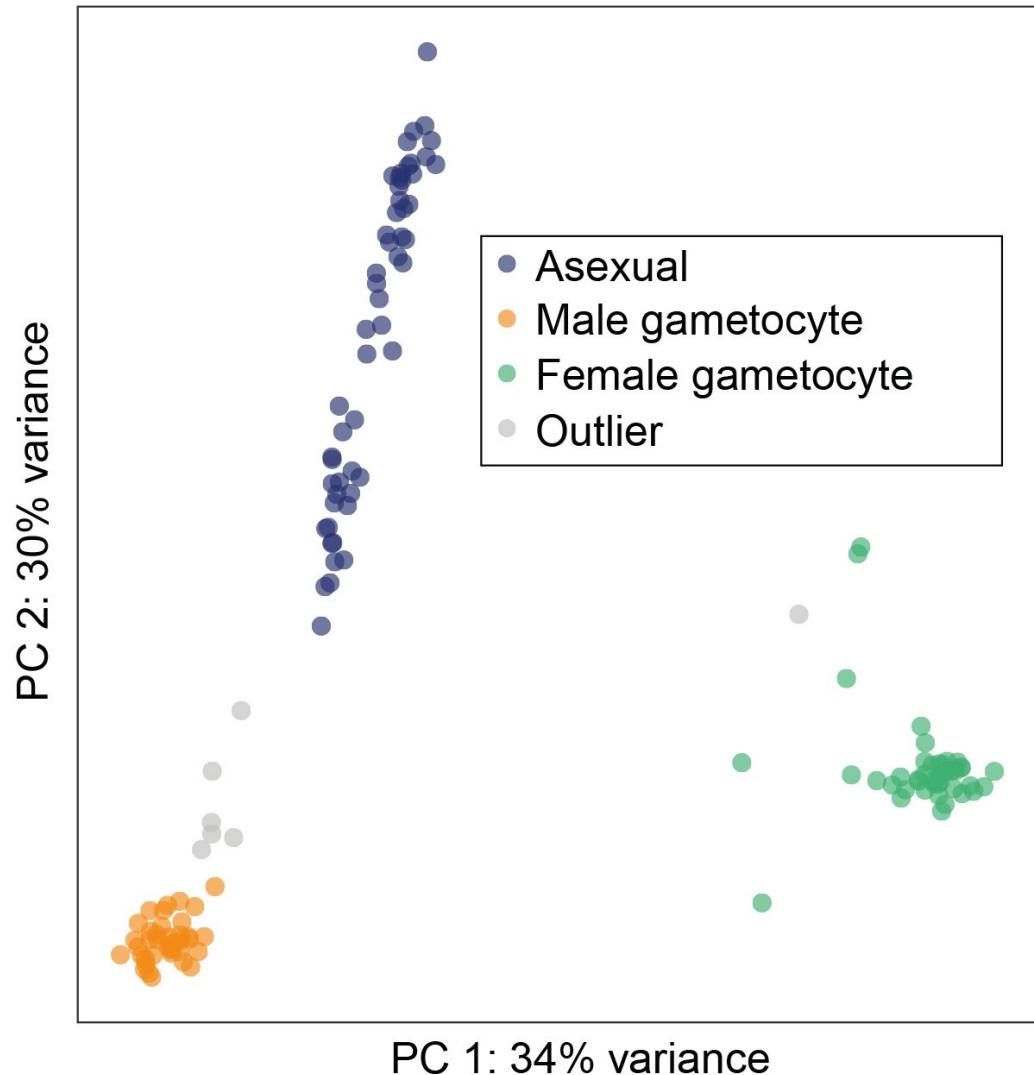
Single-cell RNA-seq methods



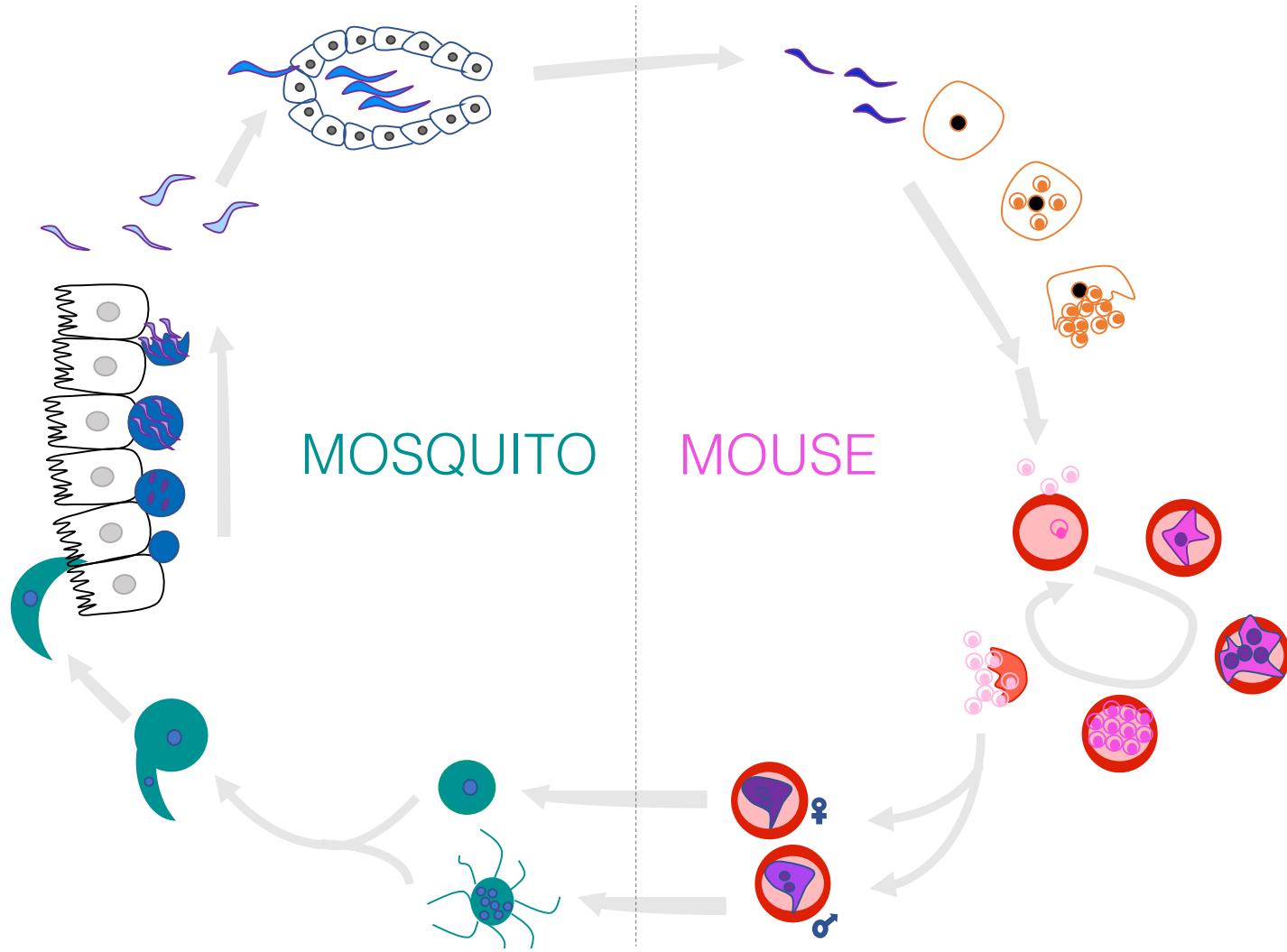
- ✓ Full length
- ✓ Better detection (genes/cell)
- ✓ Low through-put (96 or 384 well plates)

- ✓ 3' protocol
- ✓ Lower detection (genes/cell)
- ✓ Up to 10k cells per inlet

scRNA-seq in *Plasmodium*



- Modified Smart-seq2 protocol
- Cells sorted into plates
- Late-asexuals and mature gametocytes profiled in *P. falciparum* and *P. berghei*



Andy Russell

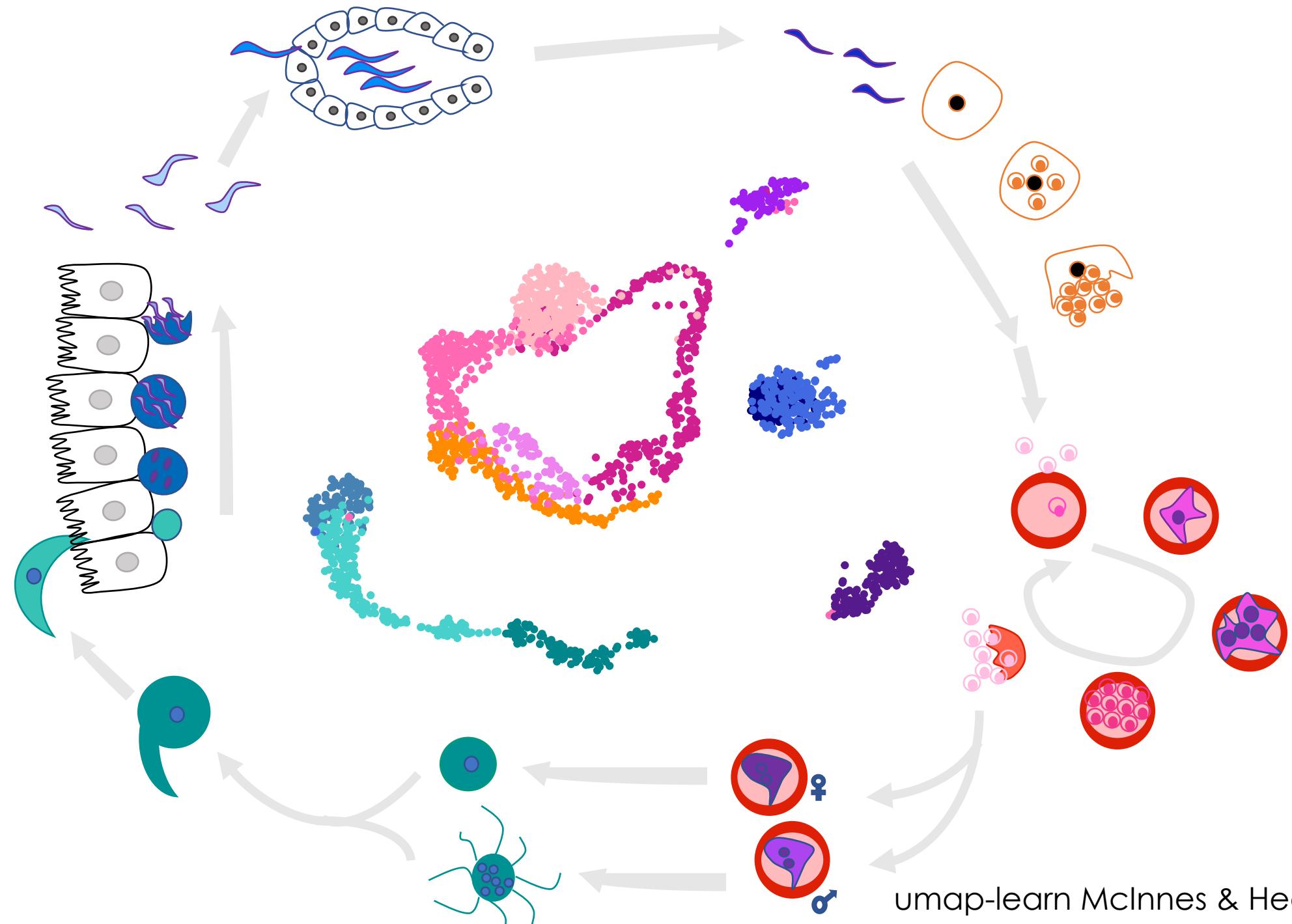


Arthur Talman



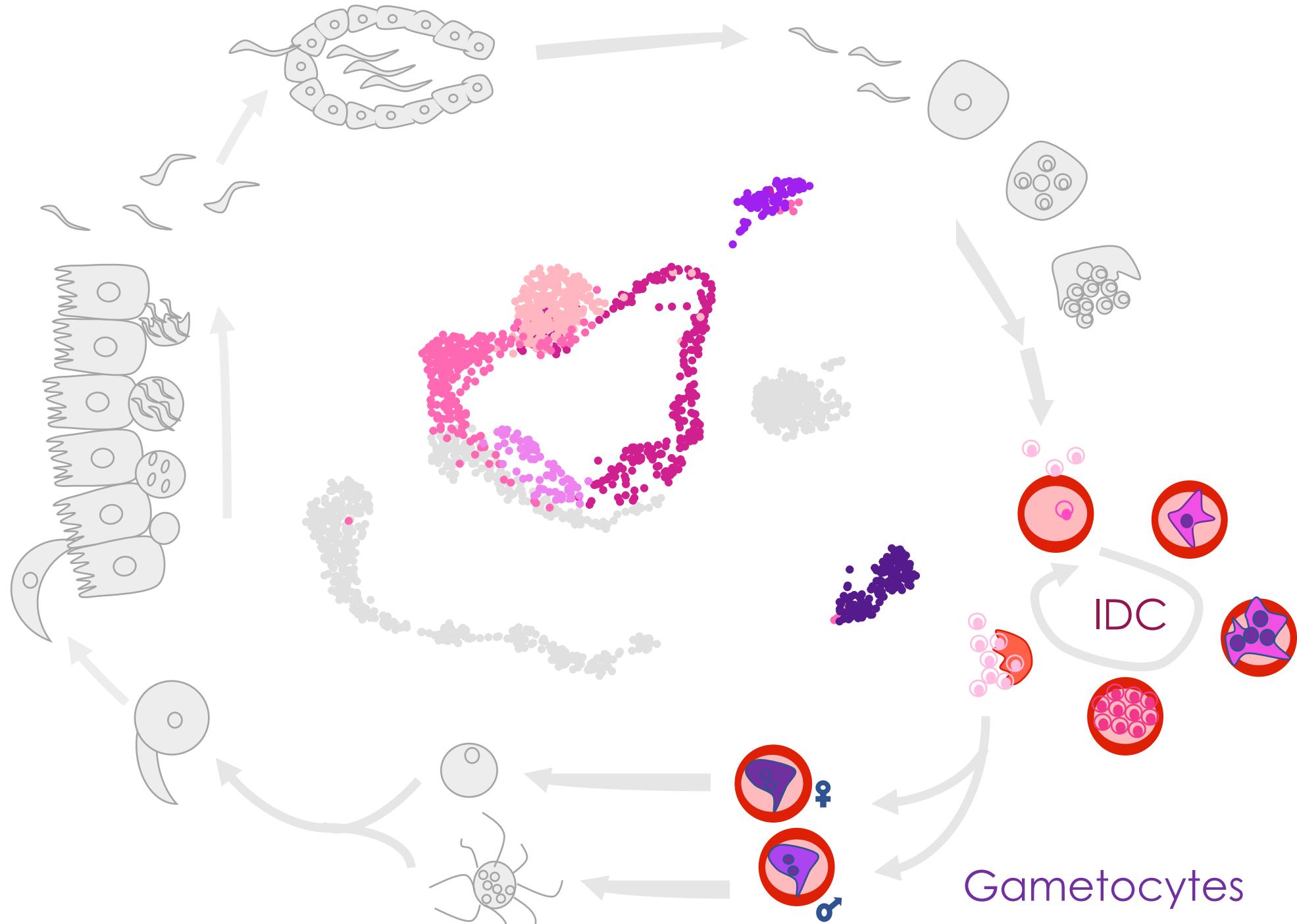
Mara Lawniczak

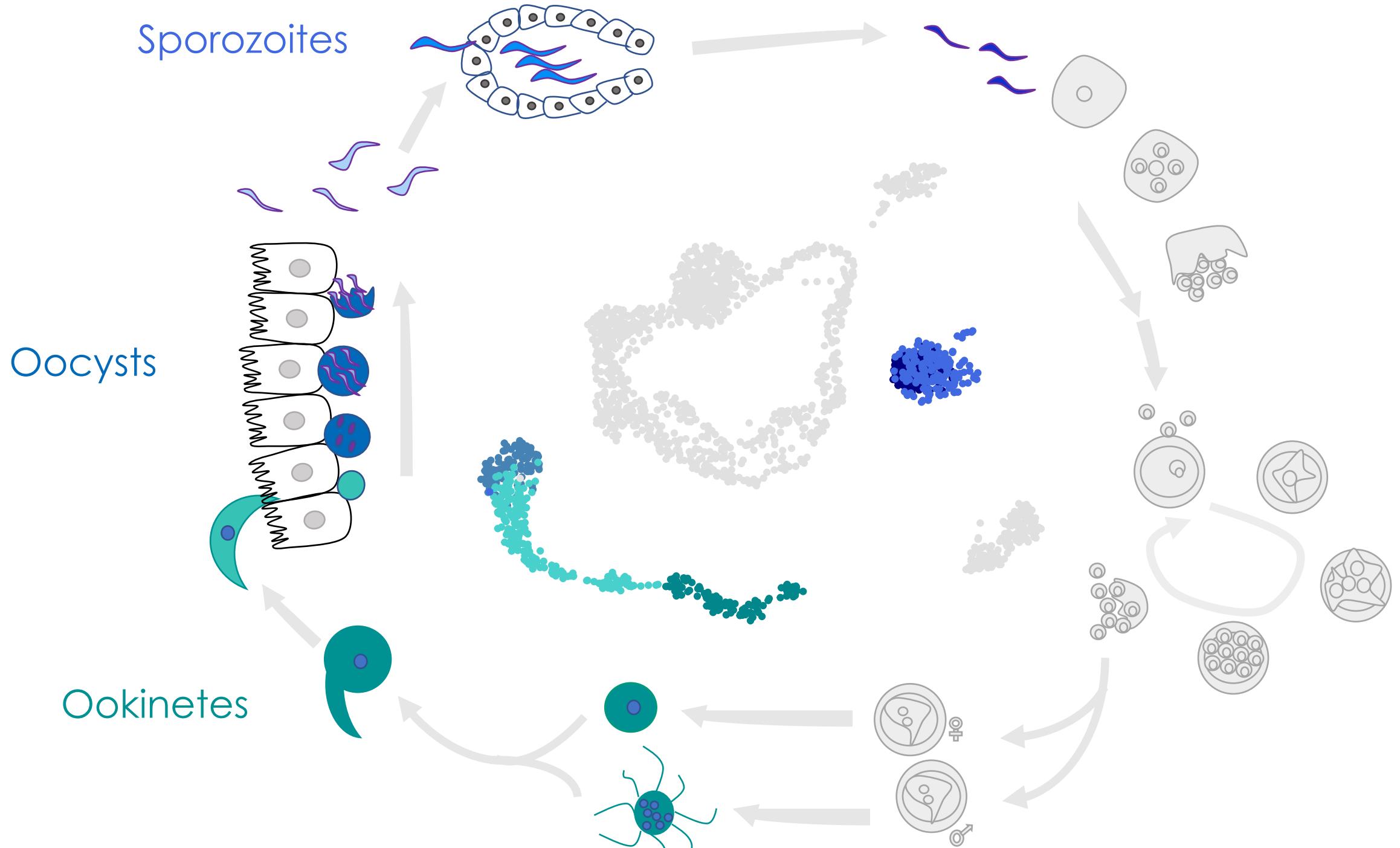
The goal of the Malaria Cell Atlas is to provide a resource of single cell transcriptomic data across the full life cycle of the parasite

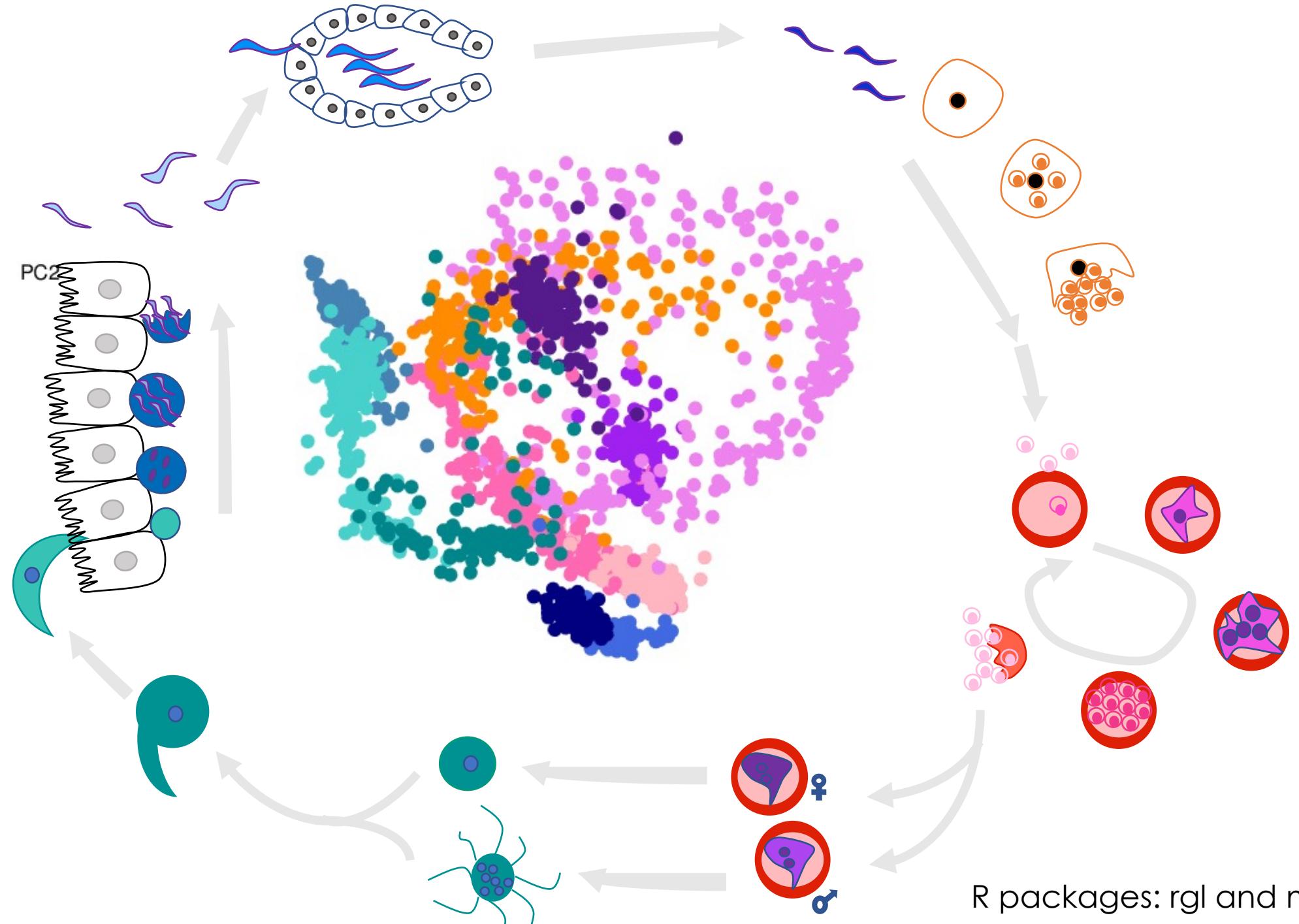


umap-learn McInnes & Healy 2018

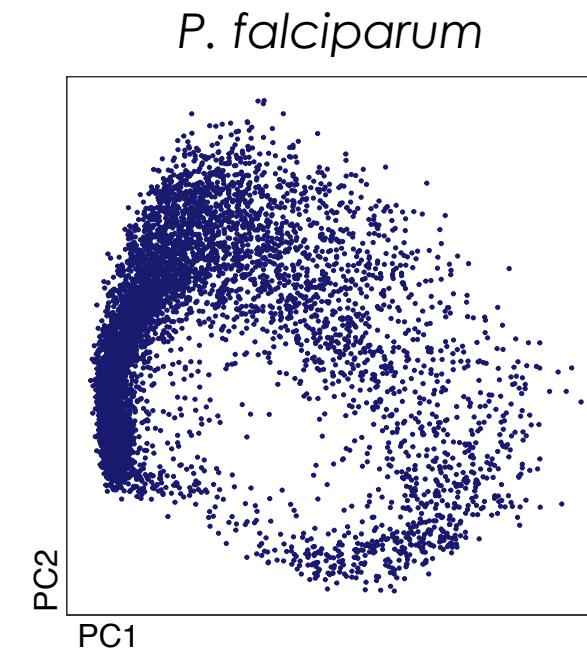
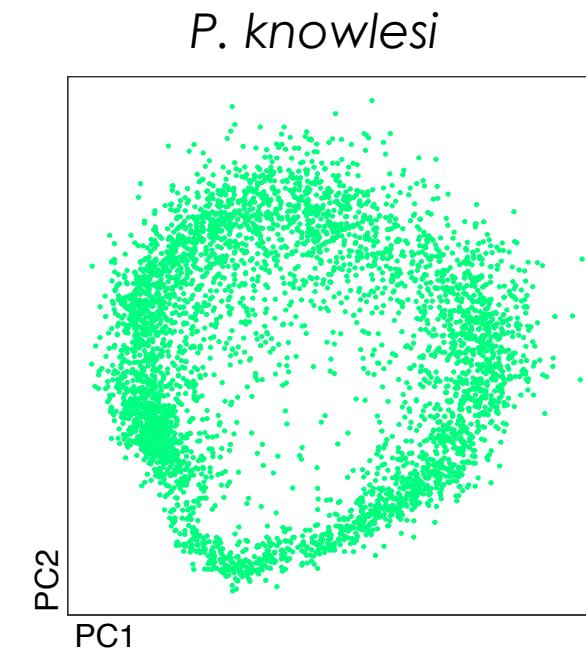
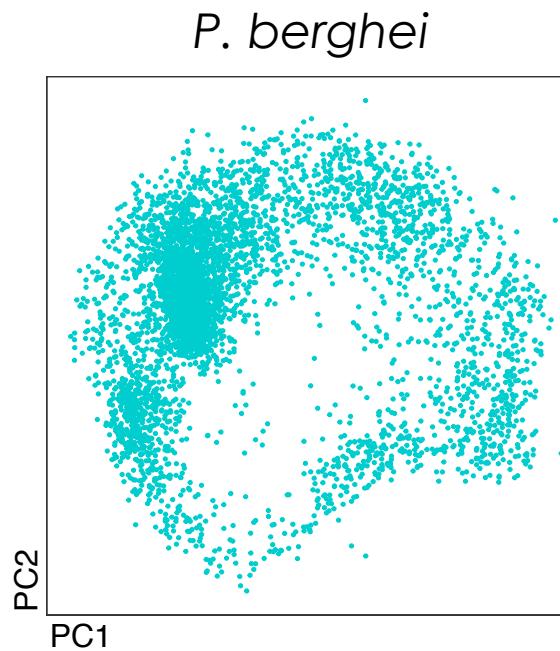




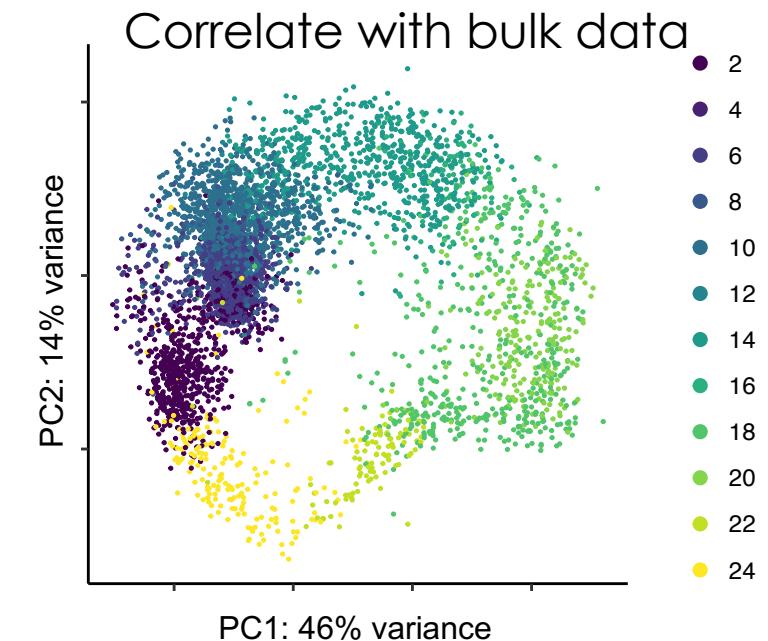
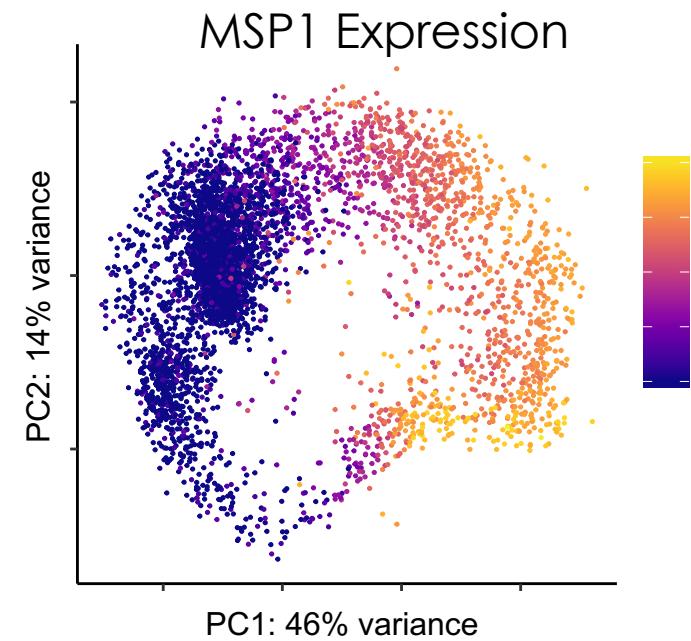
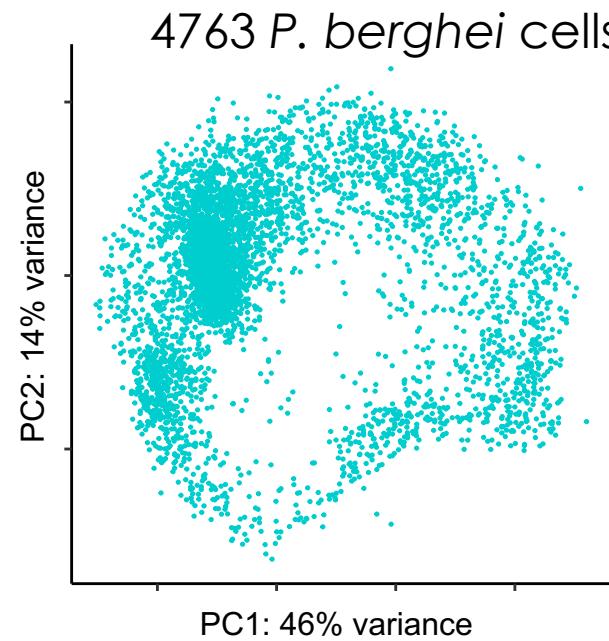




R packages: rgl and magick



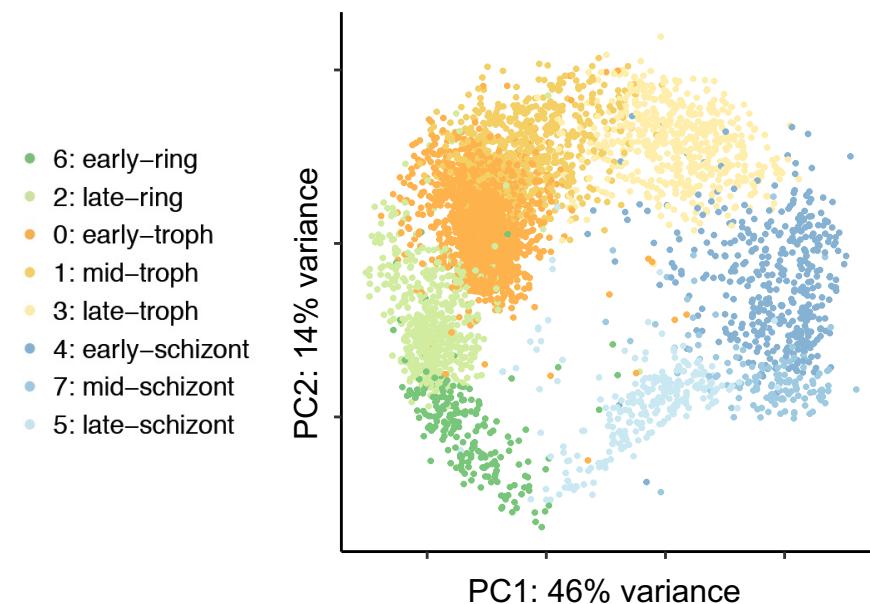
P. berghei development corresponds with marker gene expression and bulk studies



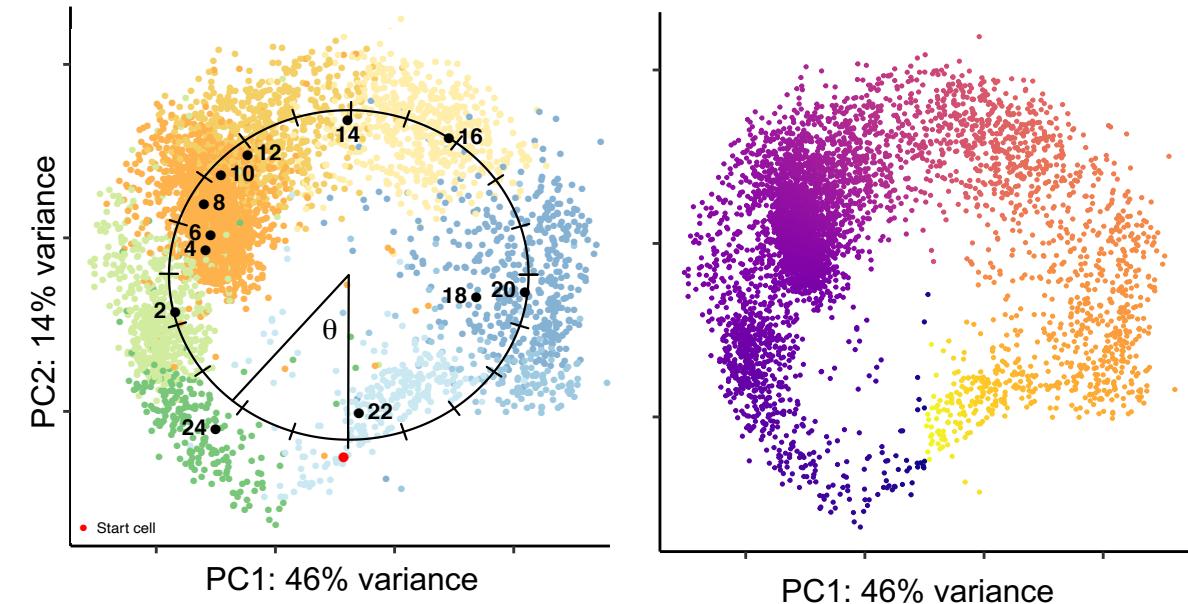
Hoo et al (2016)

Grouping and ordering cells in development

Cluster cells into groups of stages

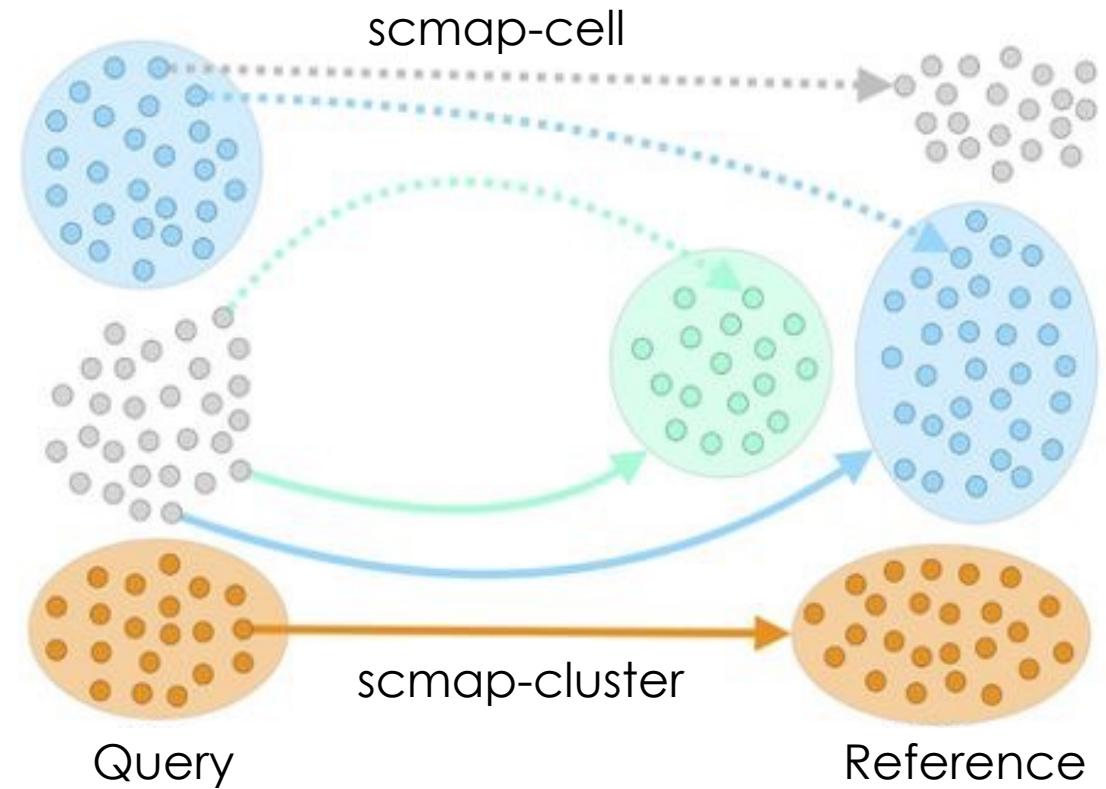


Order cells in development with 'clock' pseudotime

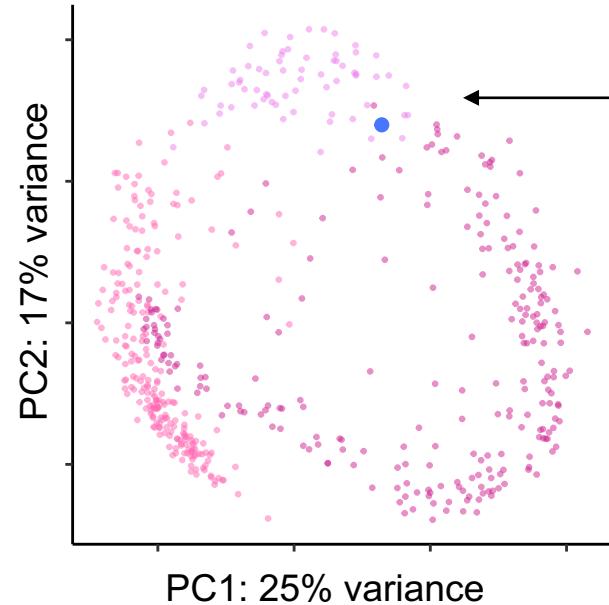


Using scmap to compare single-cell datasets

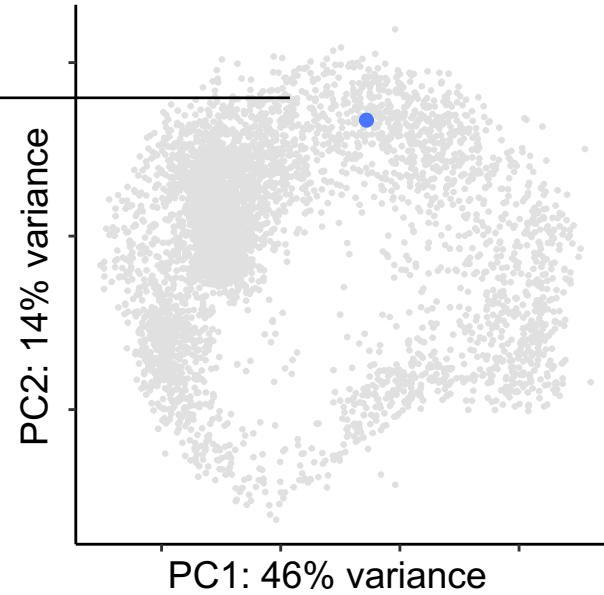
Scmap-cell allows for matching of individual cells across data sets using an approximate nearest neighbor search of a reference data set.



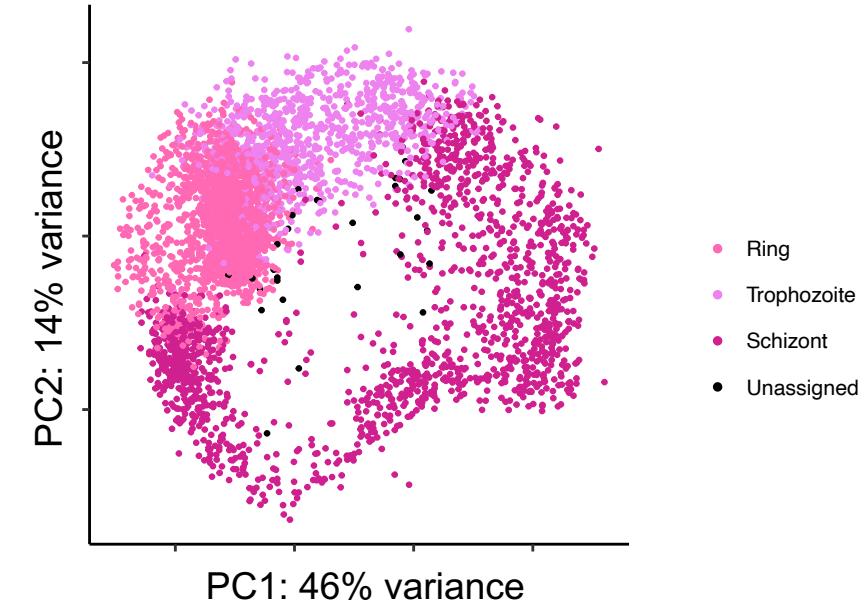
Mapping cells across technologies



Index data set:
538 *P. berghei* IDC cells from SS2

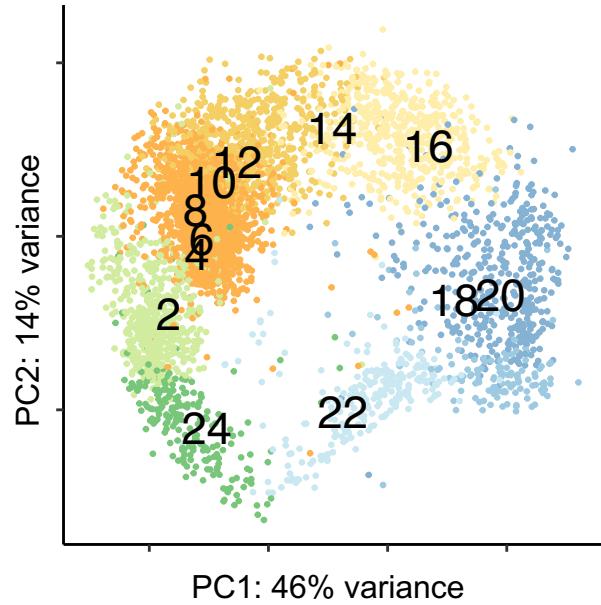


Query data set:
4763 *P. berghei* IDC cells from 10x

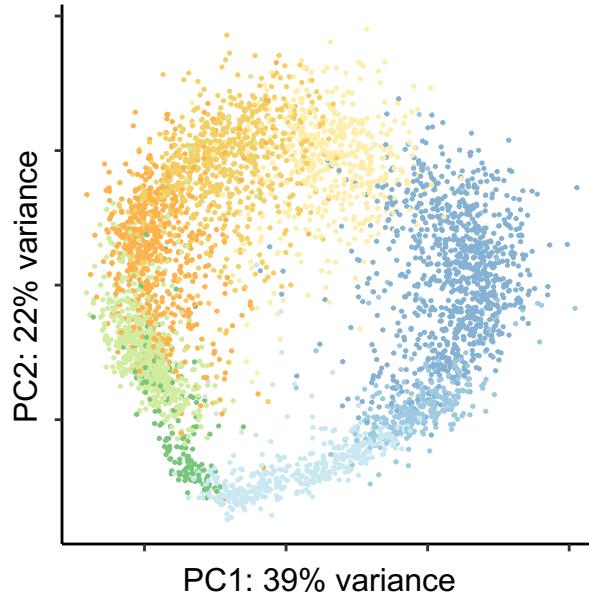


Query data set assigned:
Each 10x cell given stage assignment
based on the SS2 cell it mapped to

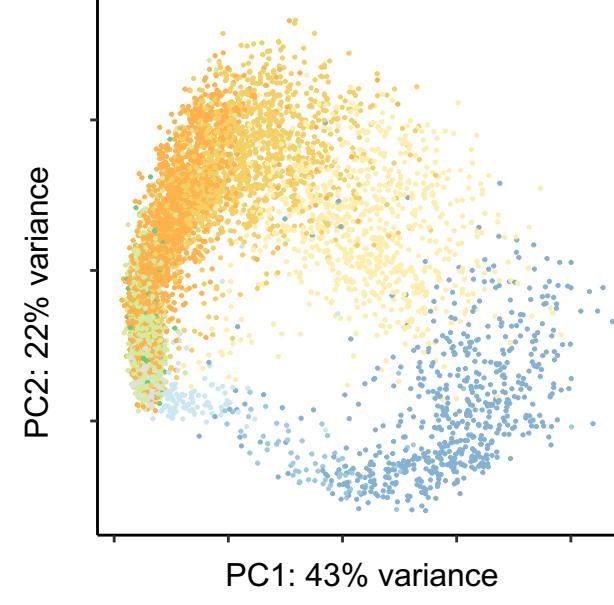
Mapping cells across species



Index data set:
4763 *P. berghei* IDC cells from 10x



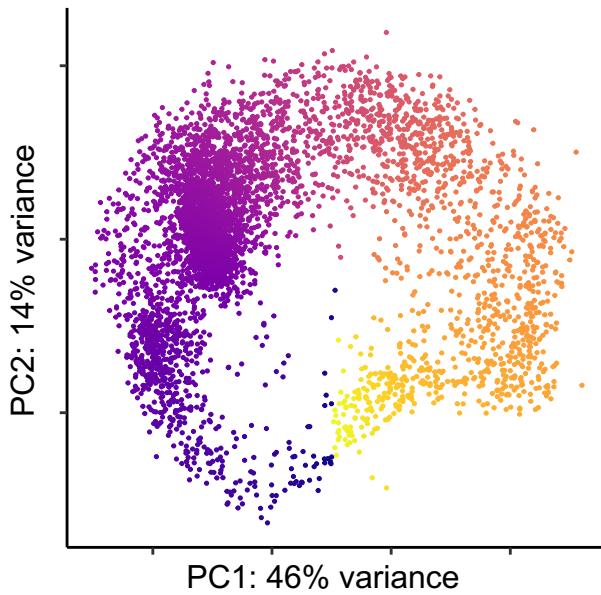
Query data set:
4237 *P. knowlesi* cells from 10x



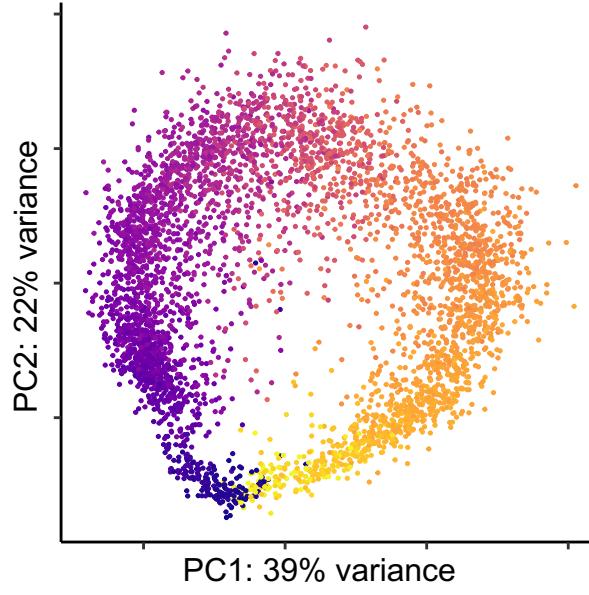
Query data set:
6737 *P. falciparum* cells from 10x

- 6: early-ring
- 2: late-ring
- 0: early-troph
- 1: mid-troph
- 3: late-troph
- 4: early-schizont
- 7: mid-schizont
- 5: late-schizont

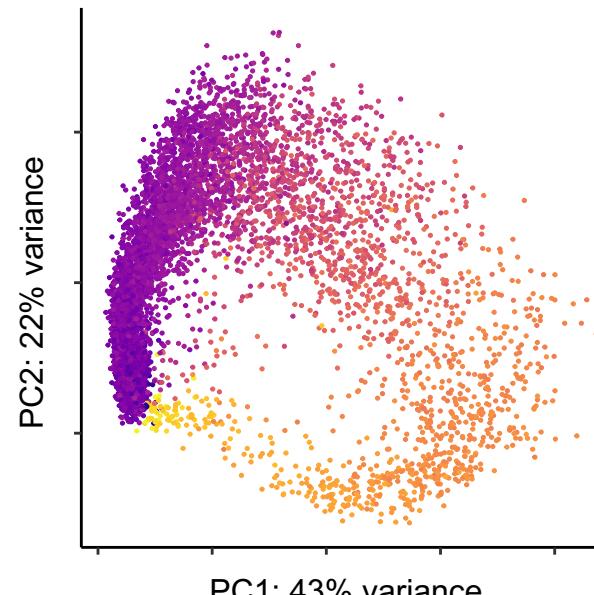
Mapping cells across species



Index data set:
4763 *P. berghei* IDC cells from 10x

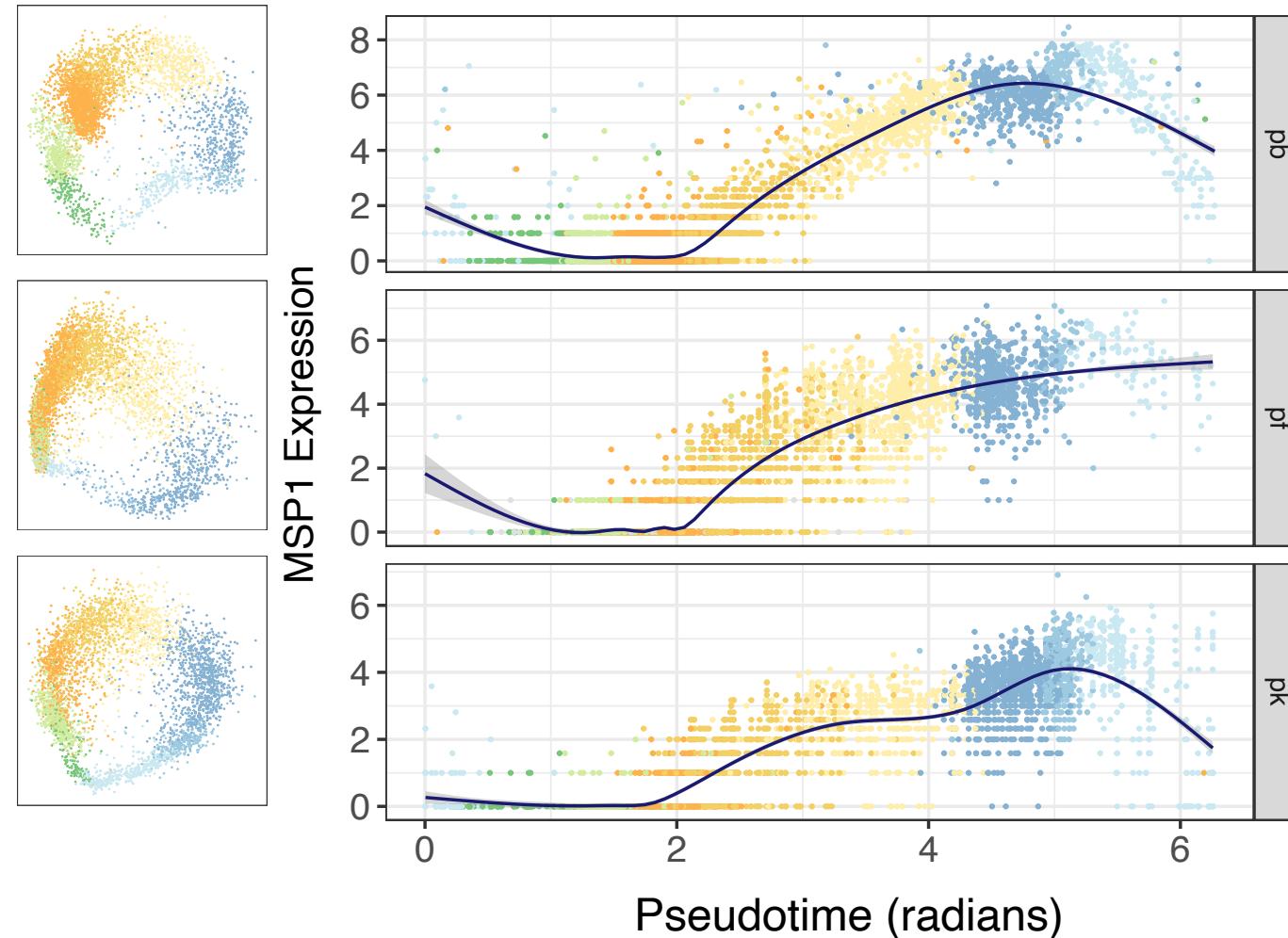


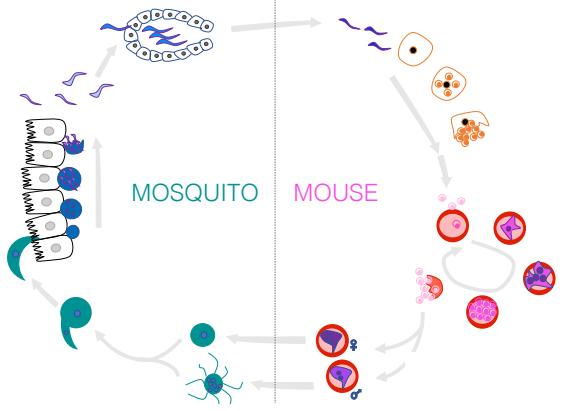
Query data set:
4237 *P. knowlesi* cells from 10x



Query data set:
6737 *P. falciparum* cells from 10x

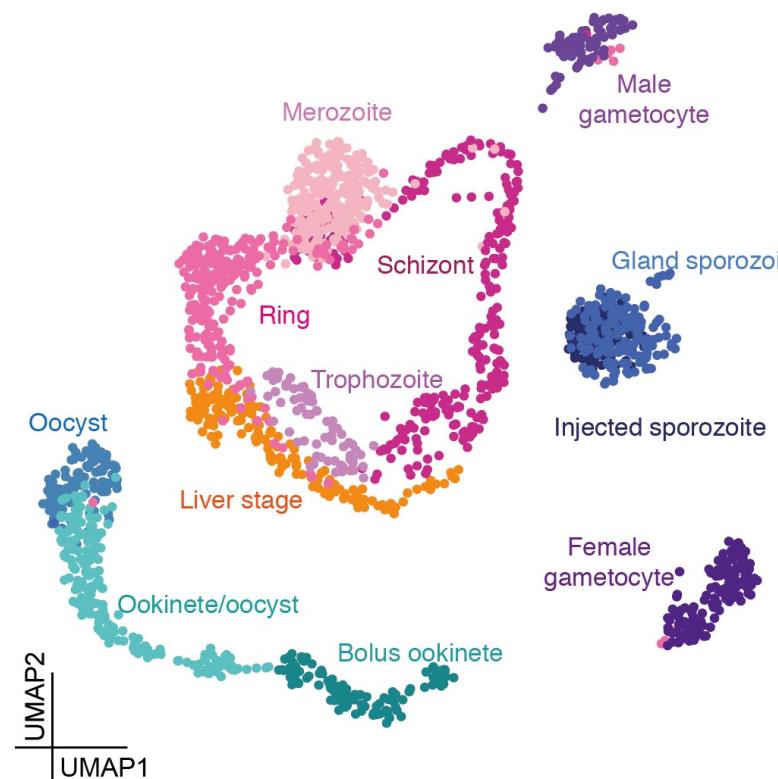
Mapping cells across species



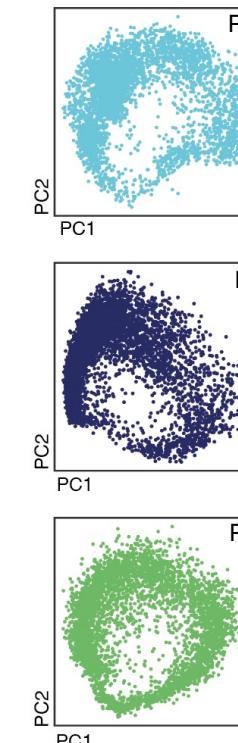


The goal of the Malaria Cell Atlas is to provide a resource of single cell transcriptomic data across the full parasite life cycle

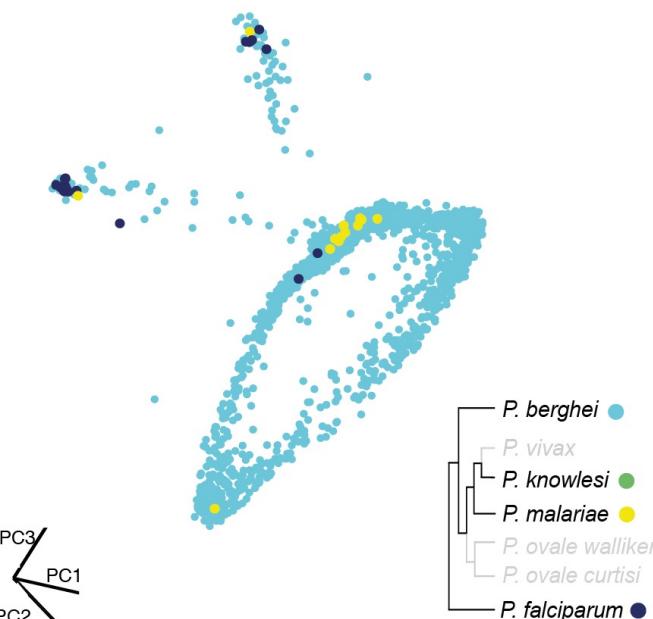
In vivo model system



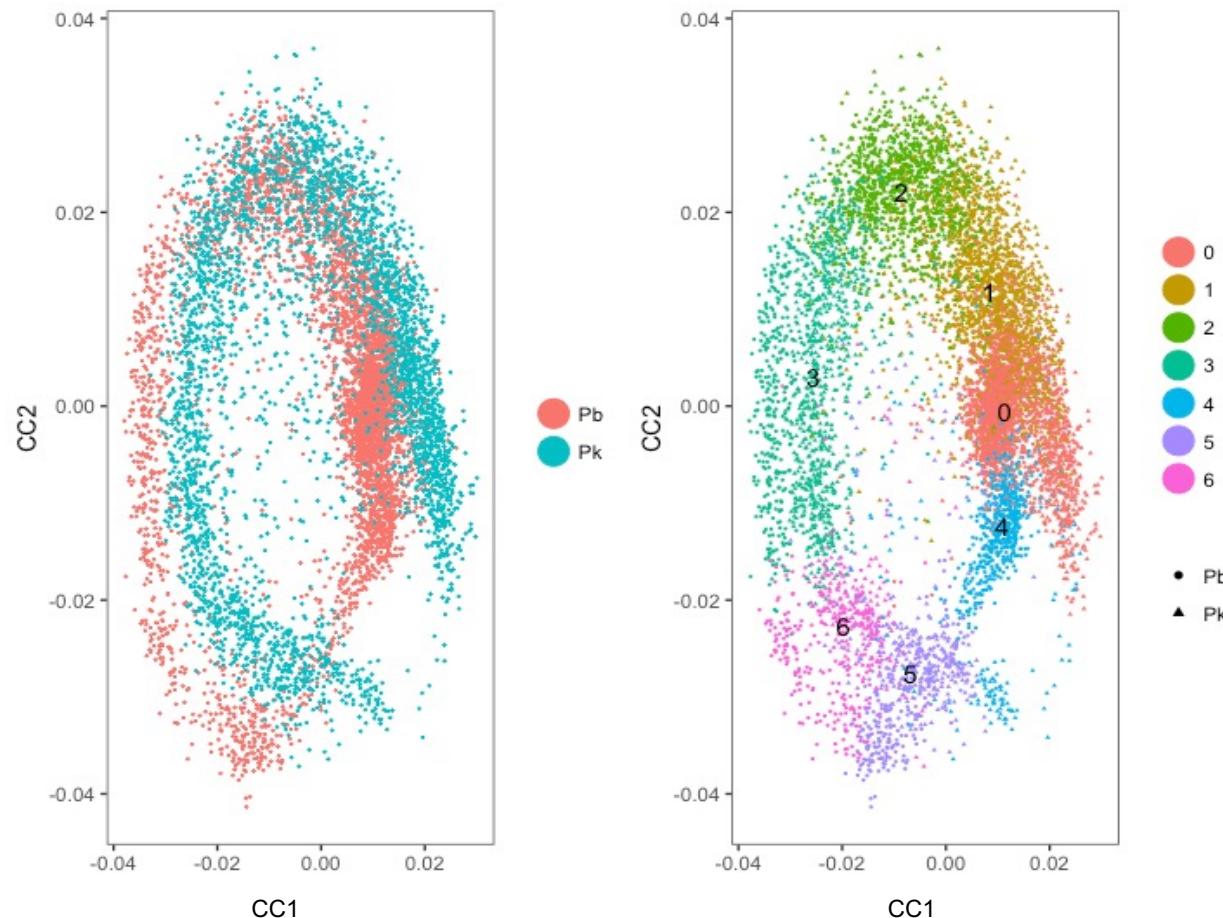
In vitro culture



Clinical isolates

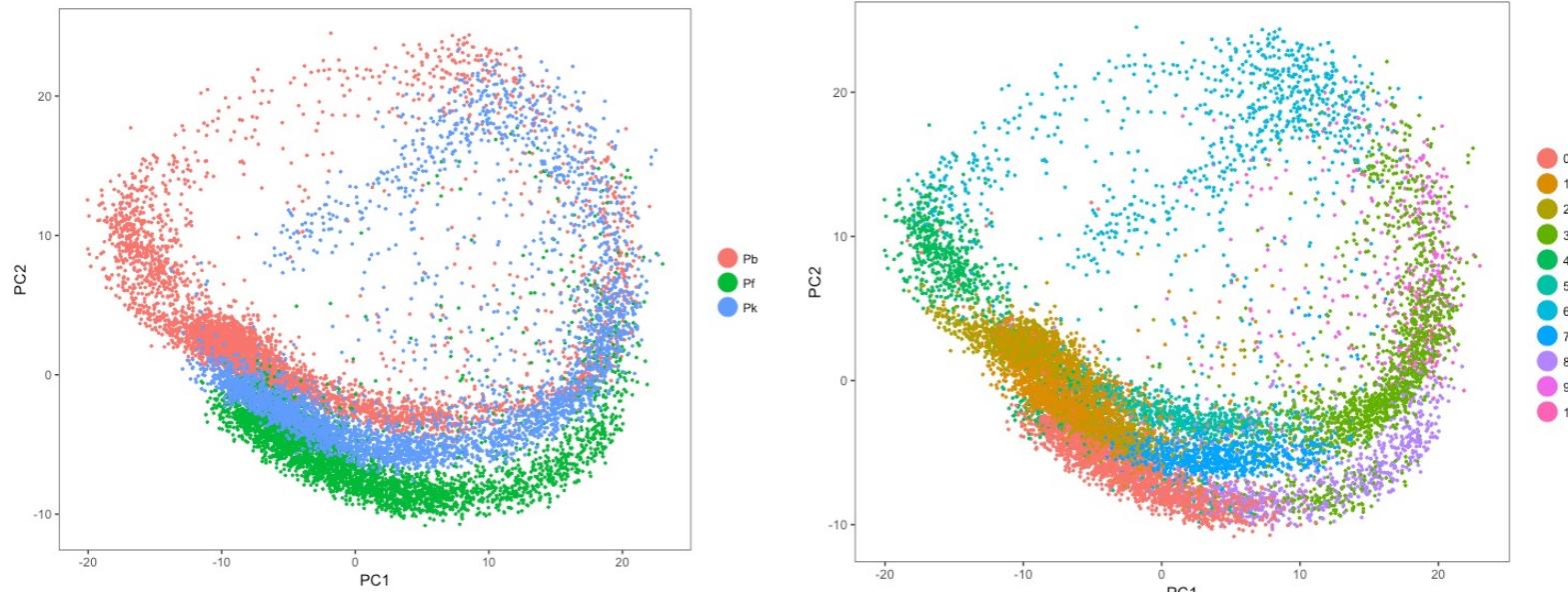


Behind the scenes: how do you classify the same cell-types across species?



Seurat's Canonical correlation analysis working well for 2 species, but not all 3

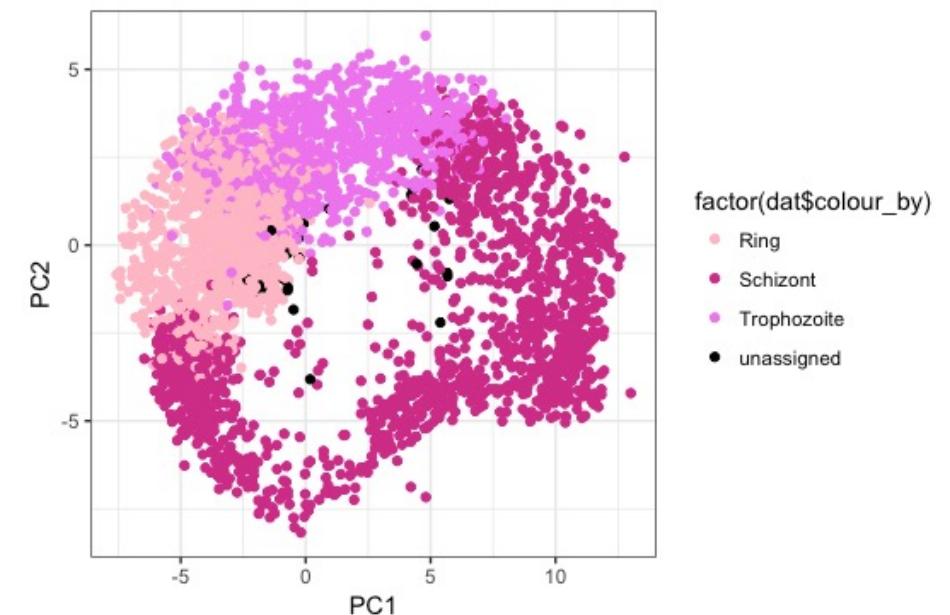
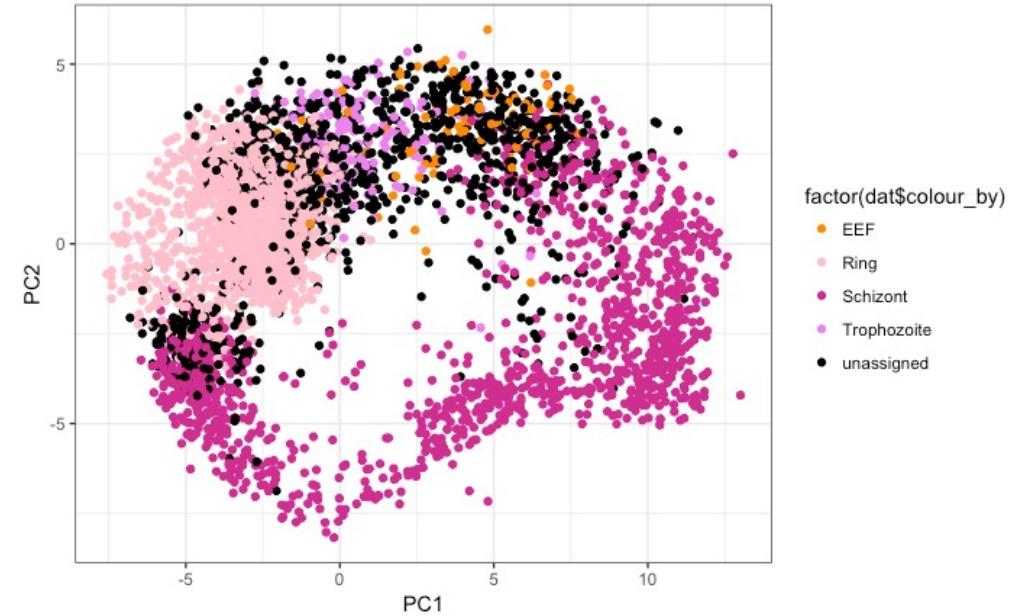
Behind the scenes: how do you classify the same cell-types across species?



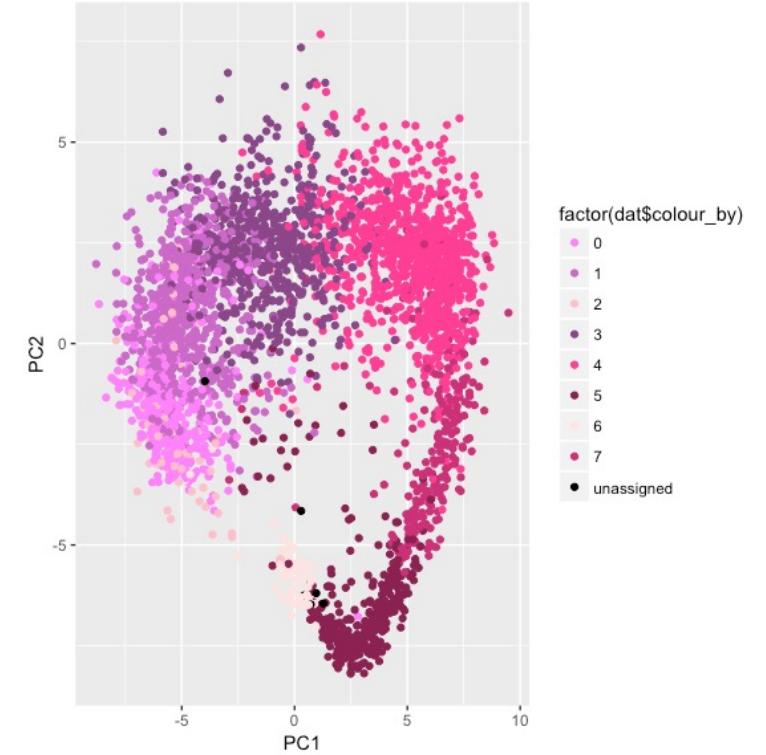
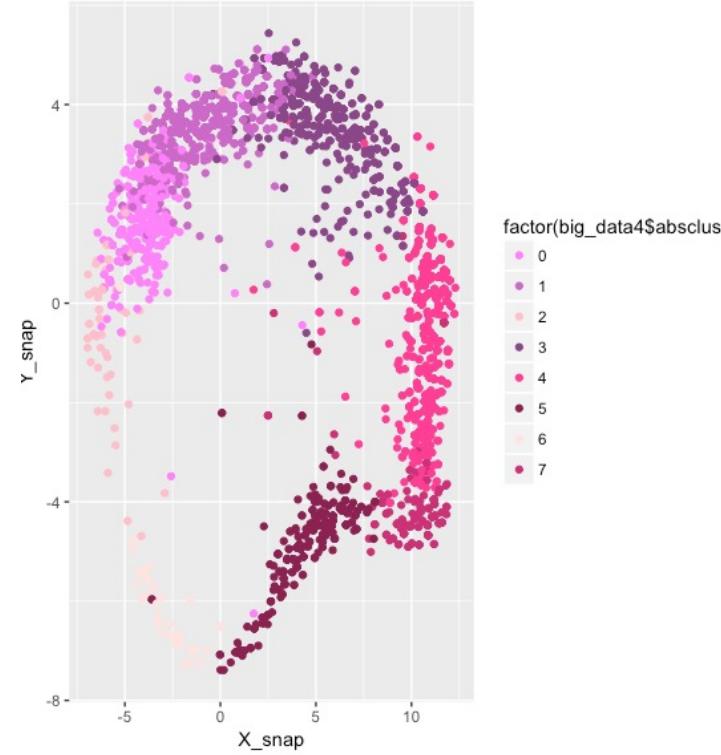
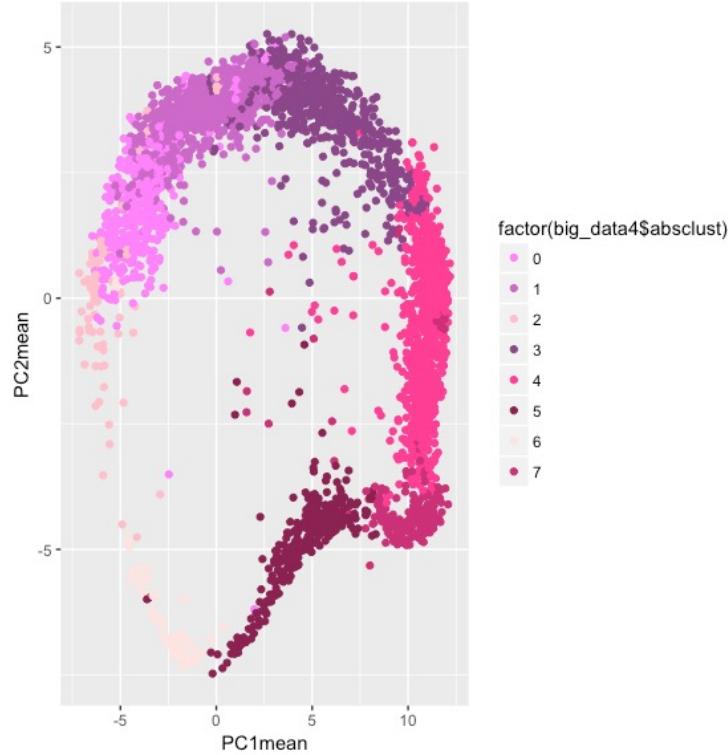
Seurat's Canonical correlation analysis working well for 2 species, but not all 3

Behind the scenes: how do you classify the same cell-types across species?

- SCmap default not working because clusters are not discrete
- Modify to use the top cell assignment only



Behind the scenes: how do you classify the same cell-types across species?



Get a mean of the PC coordinates from the top three nearest neighbours. Then snap nearest Pb cell in the Pb PCA and use this cell for cluster assignment.

Key resources

- Scmap tutorial: scmap.sanger.ac.uk/scmap/
- Seurat data integration:
satijalab.org/seurat/articles/integration_introduction.html
- Github for today's talk:
https://github.com/vhowick/abn_talk/upload
- Malaria cell atlas website: www.malariacellatas.org

Analysis View

- Principal component analysis
- UMAP

Parasite stage

- All stages
- Liver stages
- Blood stages
- Gametocytes
- Ookinetes/Oocysts
- Sporozoites

Colour By

- Stage
- Gene

Gene

Color for Minimum Expression



Color for Maximum Expression

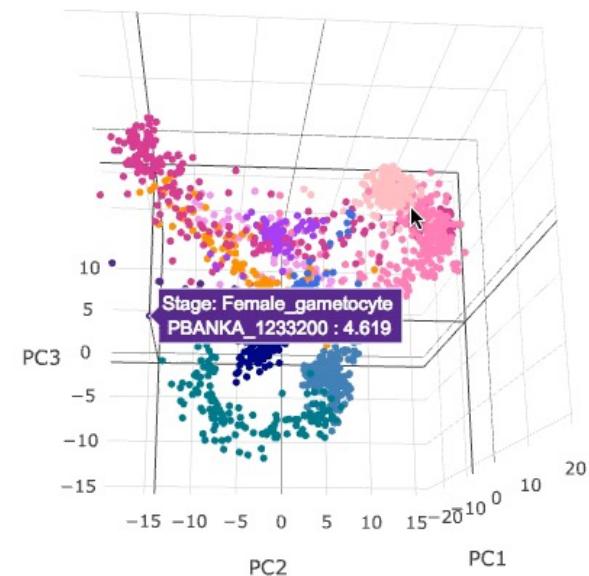


Cell Size:

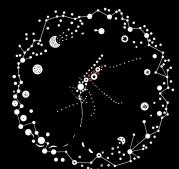


Mouse over points to see stage and expression level.

Plasmodium berghei Plasmodium falciparum Anopheles Mouse



- Liver_stage
- Merozoite
- Ring
- Trophozoite
- Schizont
- Male_gametocyte
- Female_gametocyte
- Ookinete
- Oocyst
- Gland_sporozoite
- Injected_sporozoite



Acknowledgements



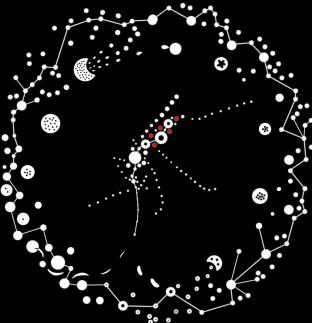
Mara Lawniczak



Andy Russell



Arthur Talman



www.malariacellatlas.org

Howick, Russell et al *Science* (2019) PMID: 31439762

Real, Howick et al *Nature Comms* (2021) PMID: 34045457



Haynes Heaton

Adam Reid

Martin Hemberg

Matt Berriman

Hellen Butungi

Lisa Verzier



Logo by Alex Cagan



Lawniczak group



Core cytometry team