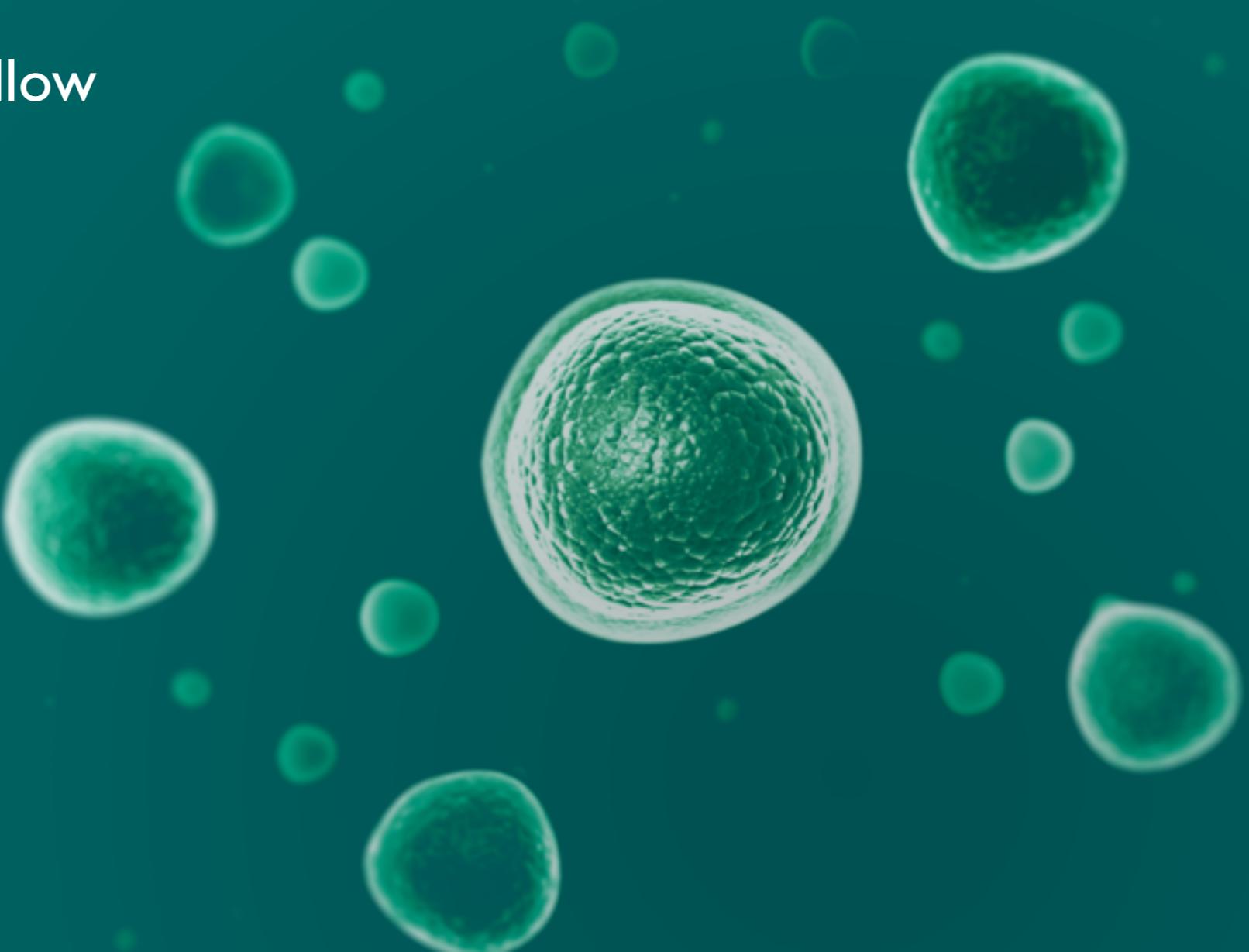


scater: an R/Bioconductor package for pre-processing, quality control, normalisation and visualisation of single-cell RNA-seq data

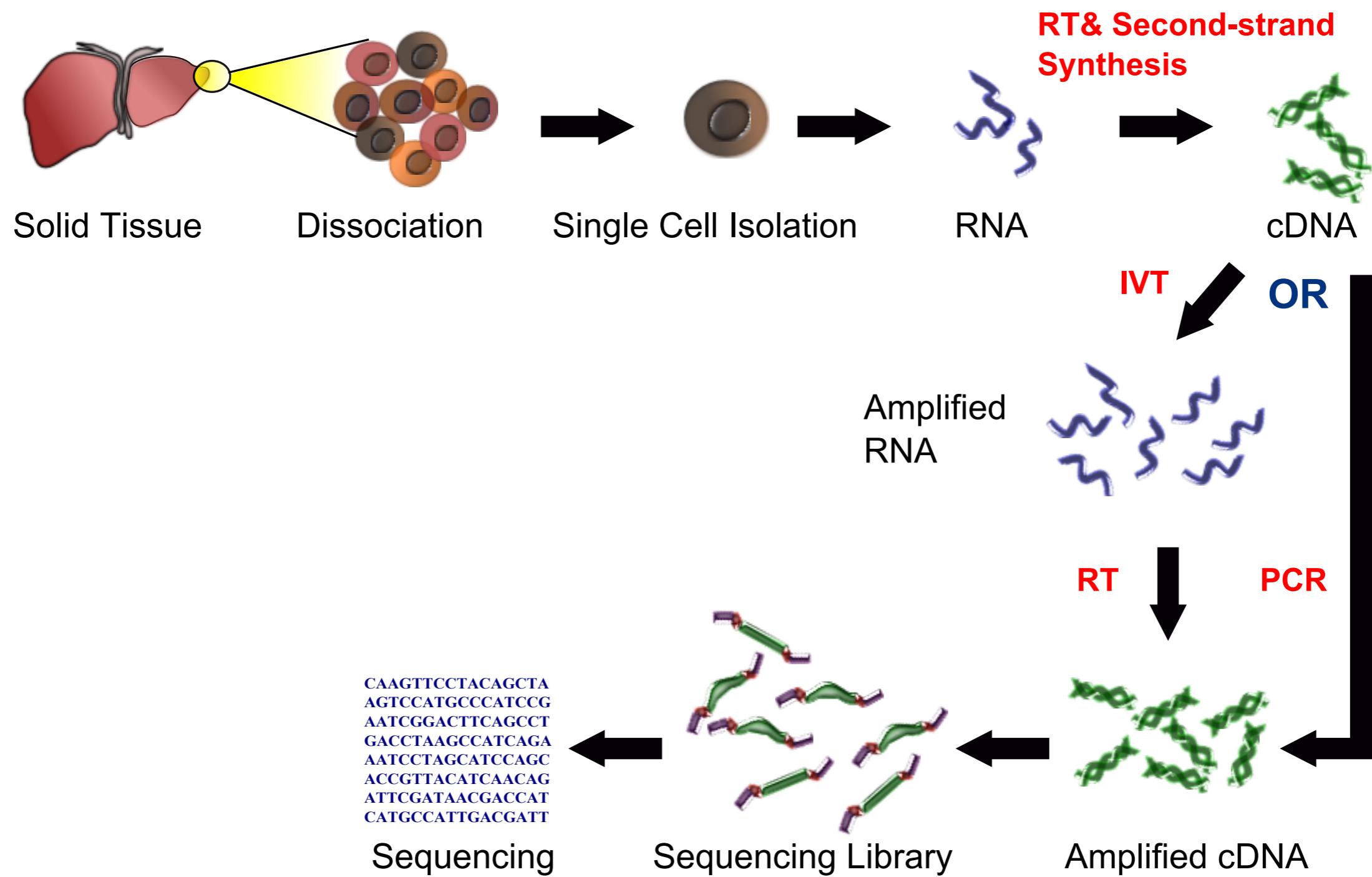
Davis McCarthy
NHMRC Early Career Fellow
Stegle Group, EMBL-EBI

@davisjmcc

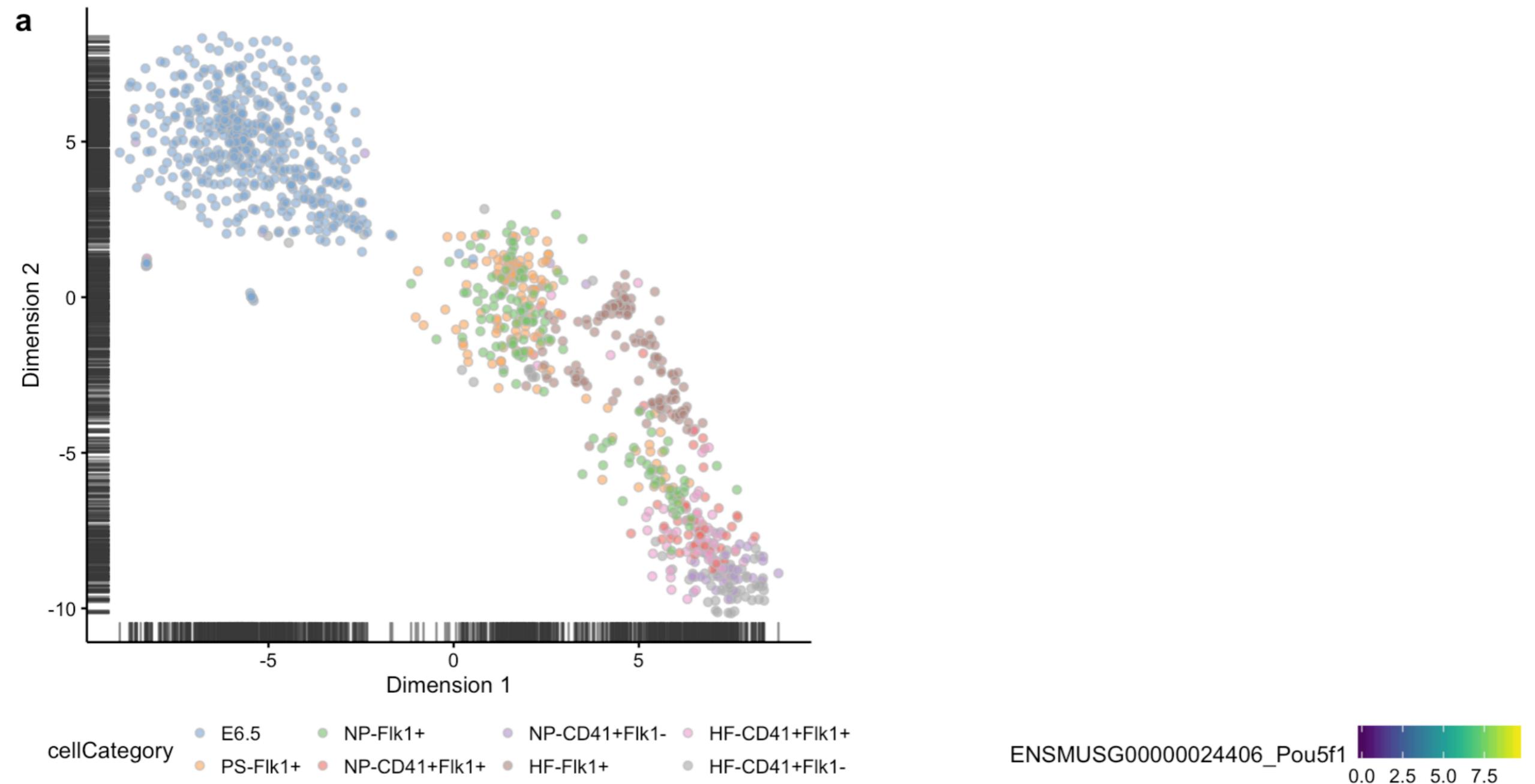
www.ebi.ac.uk



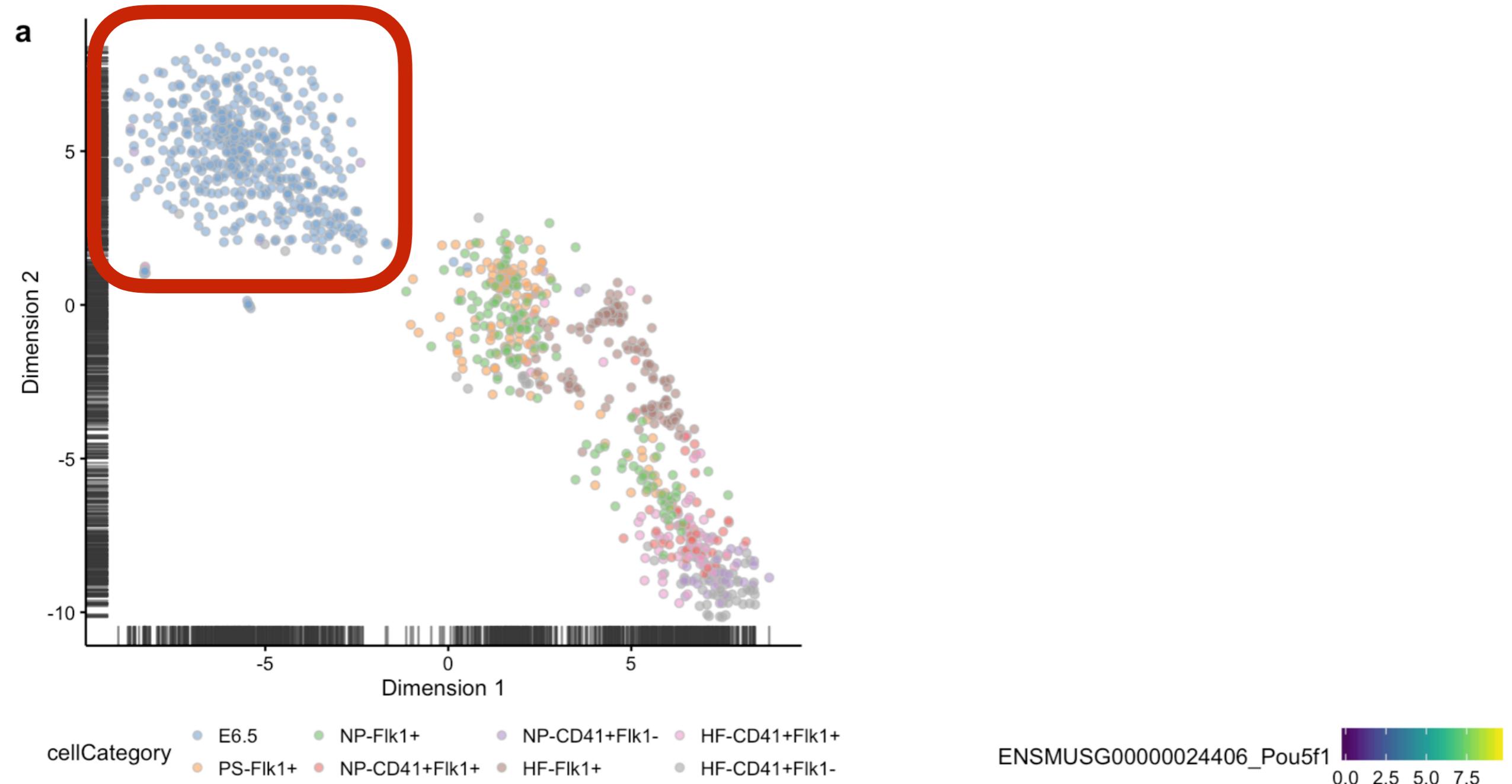
Single-cell RNA sequencing



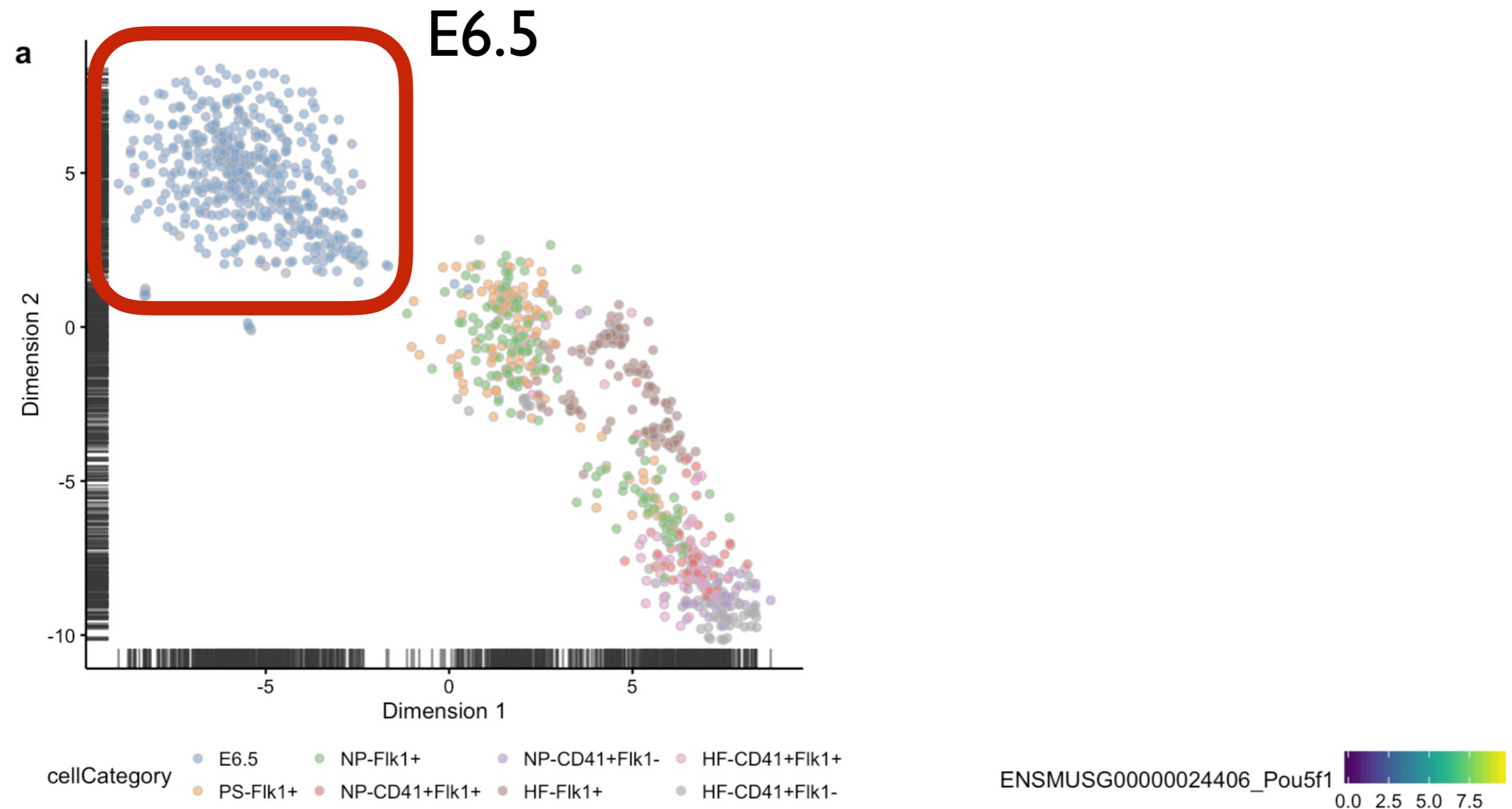
Differentiation to mouse mesoderm (Scialdone et al, *Nature*, 2016)



Differentiation to mouse mesoderm (Scialdone et al, *Nature*, 2016)



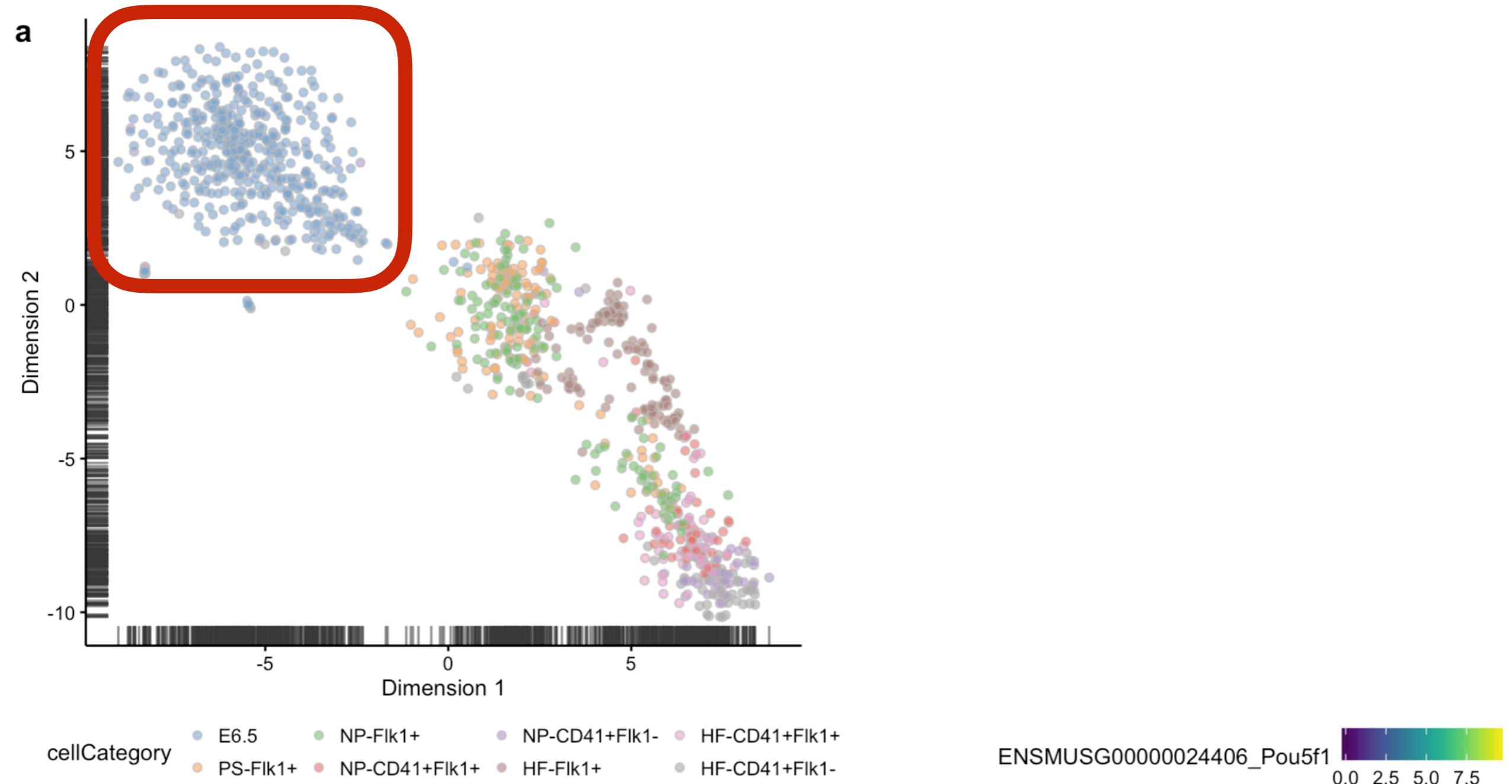
Differentiation to mouse mesoderm (Scialdone et al, *Nature*, 2016)



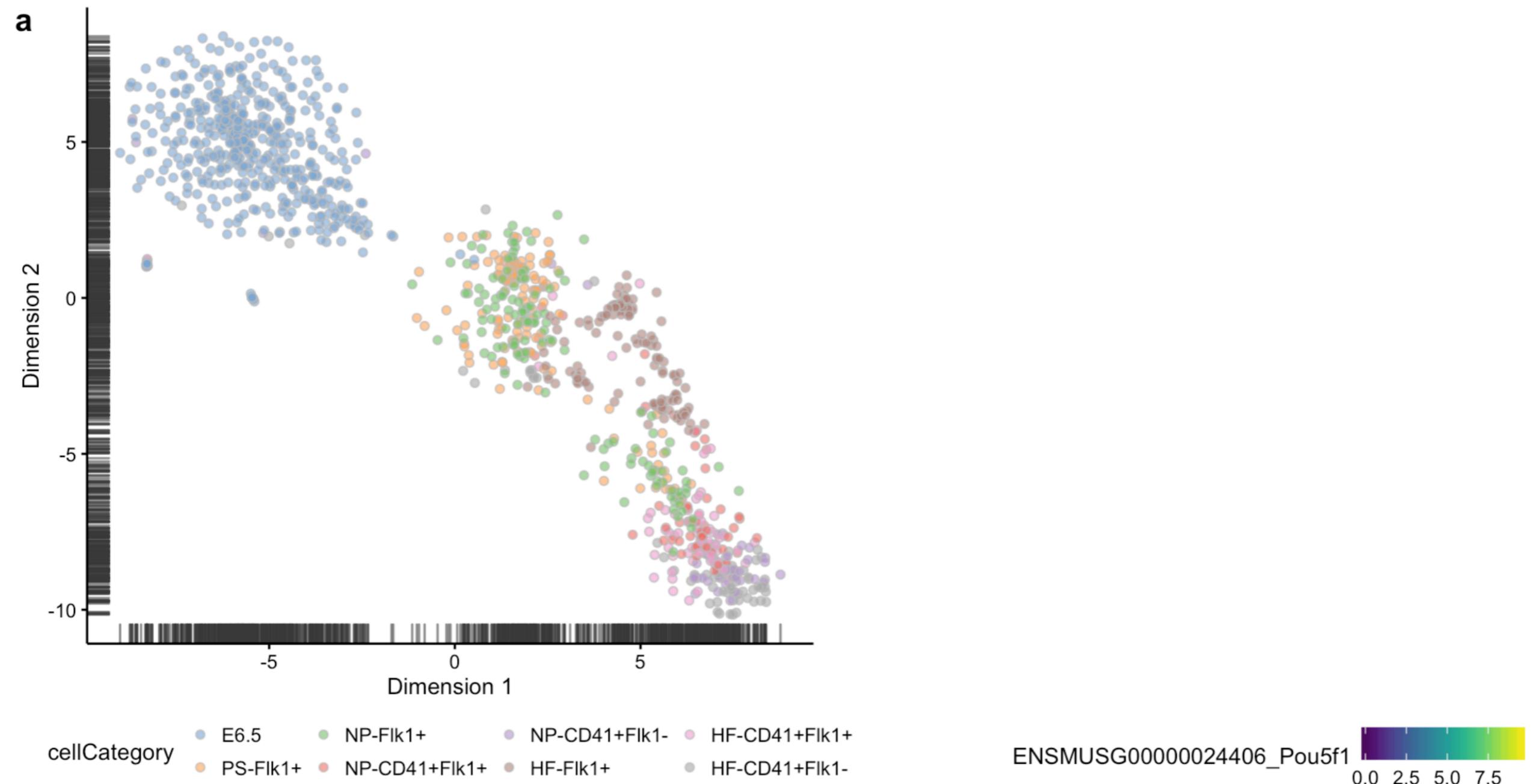
Visualisation with scater package

EMBL-EBI

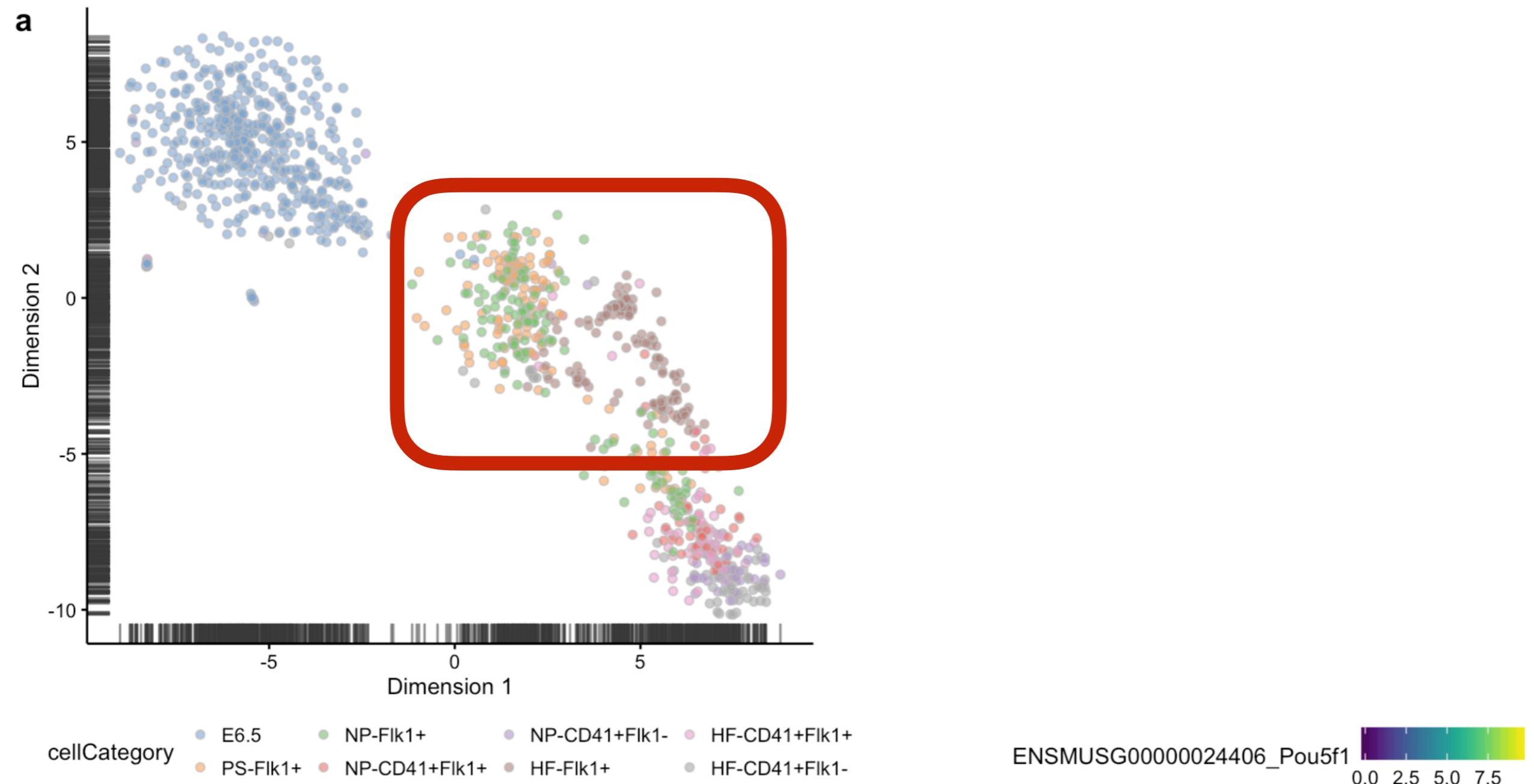
Differentiation to mouse mesoderm (Scialdone et al, *Nature*, 2016)



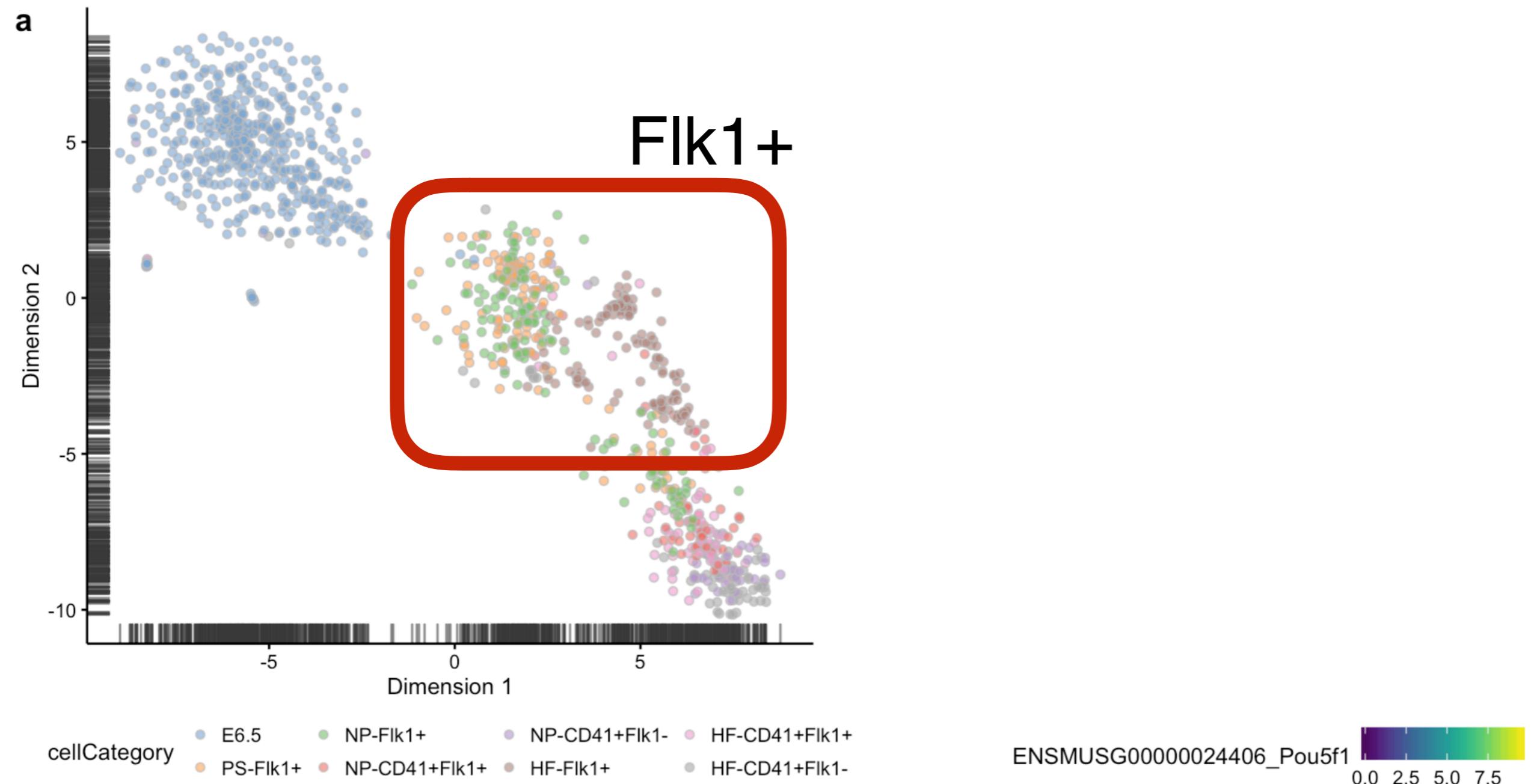
Differentiation to mouse mesoderm (Scialdone et al, *Nature*, 2016)



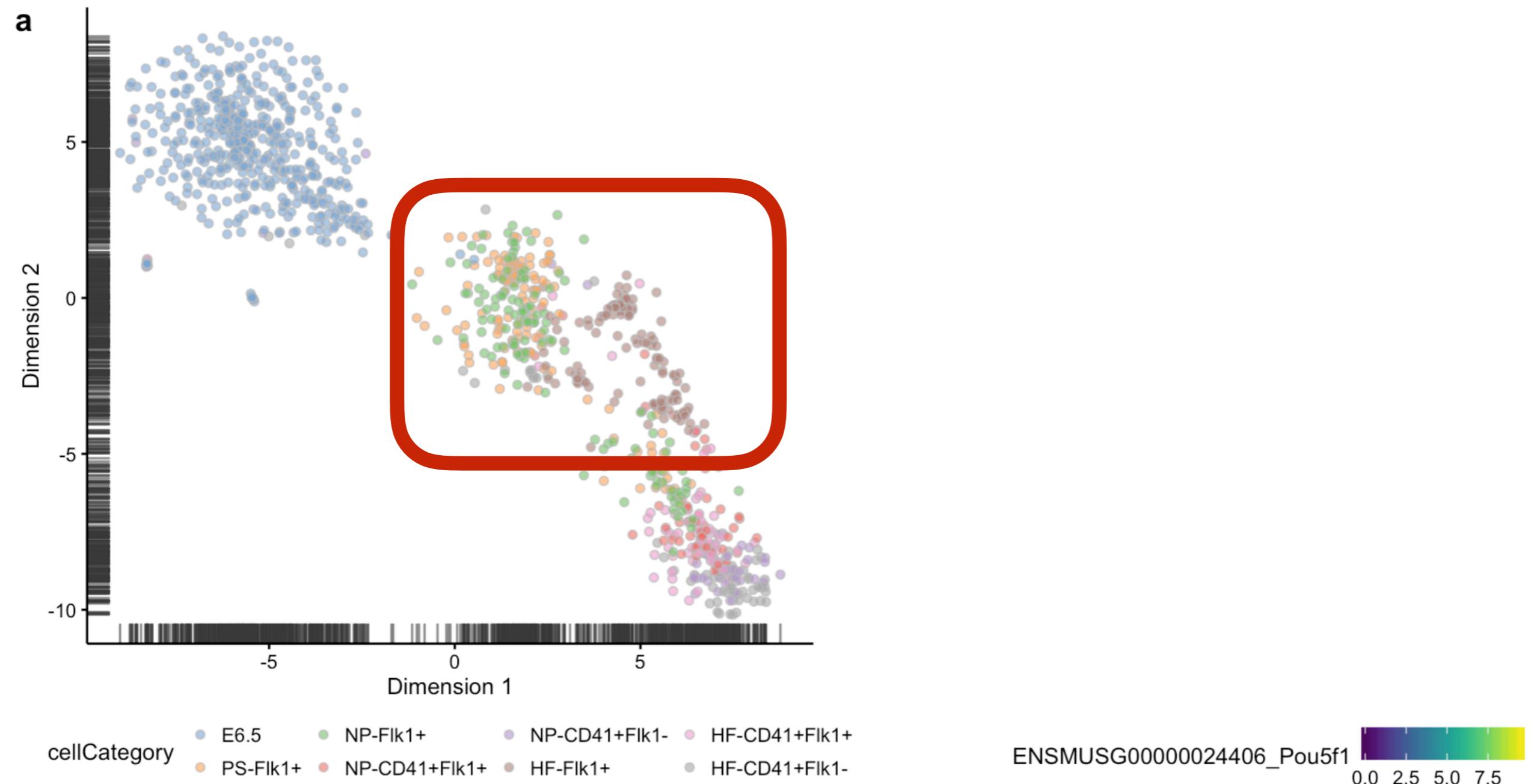
Differentiation to mouse mesoderm (Scialdone et al, *Nature*, 2016)



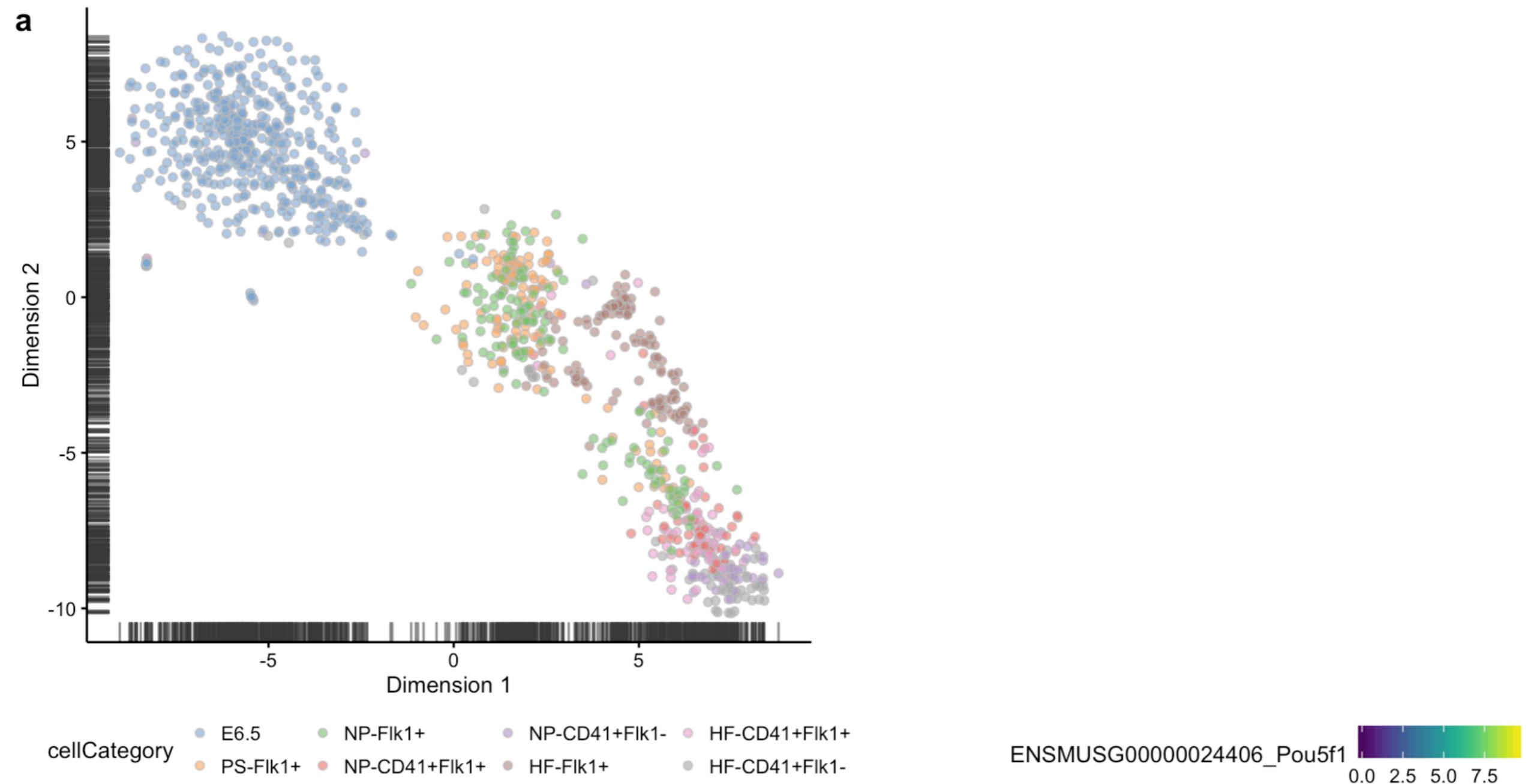
Differentiation to mouse mesoderm (Scialdone et al, *Nature*, 2016)



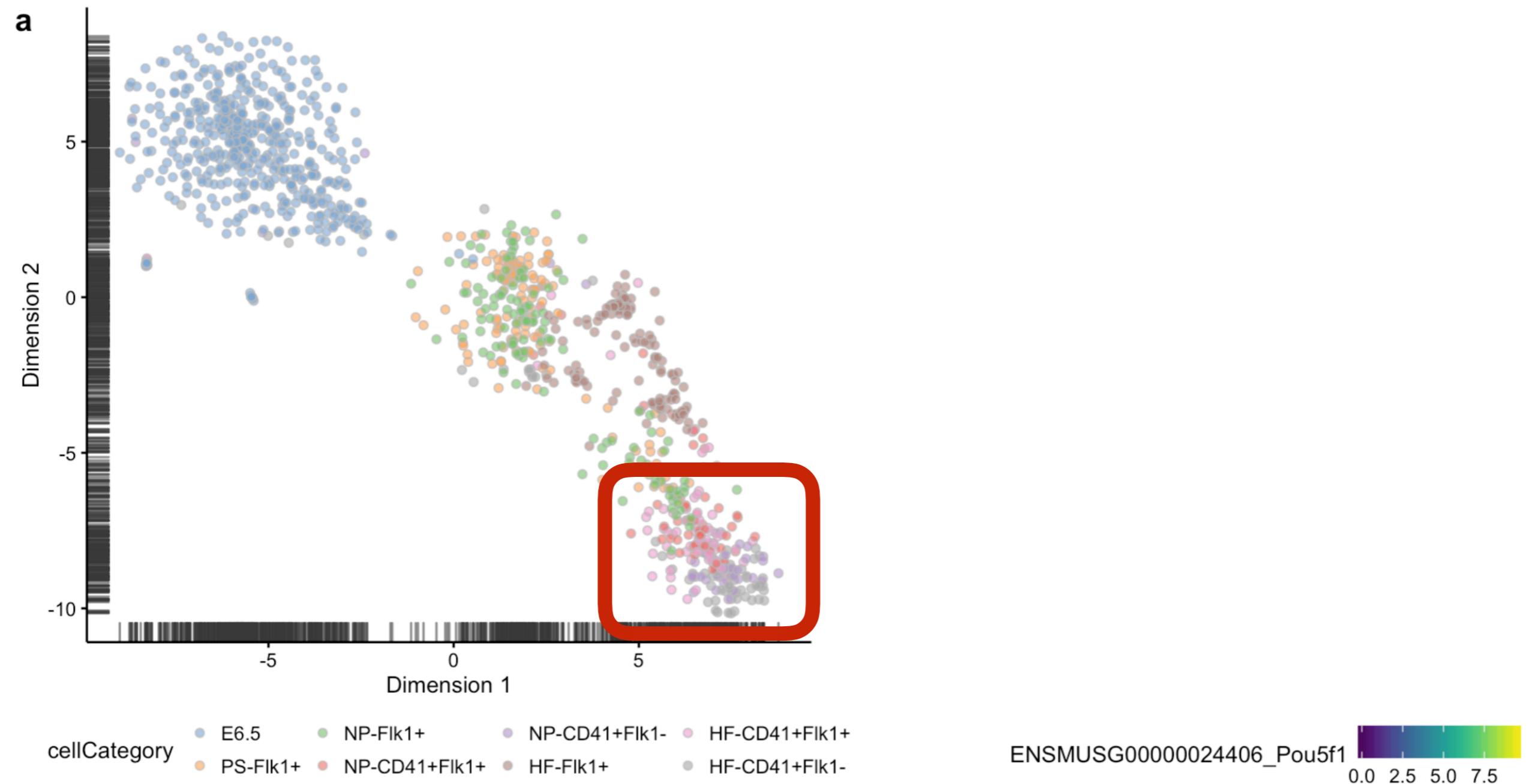
Differentiation to mouse mesoderm (Scialdone et al, *Nature*, 2016)



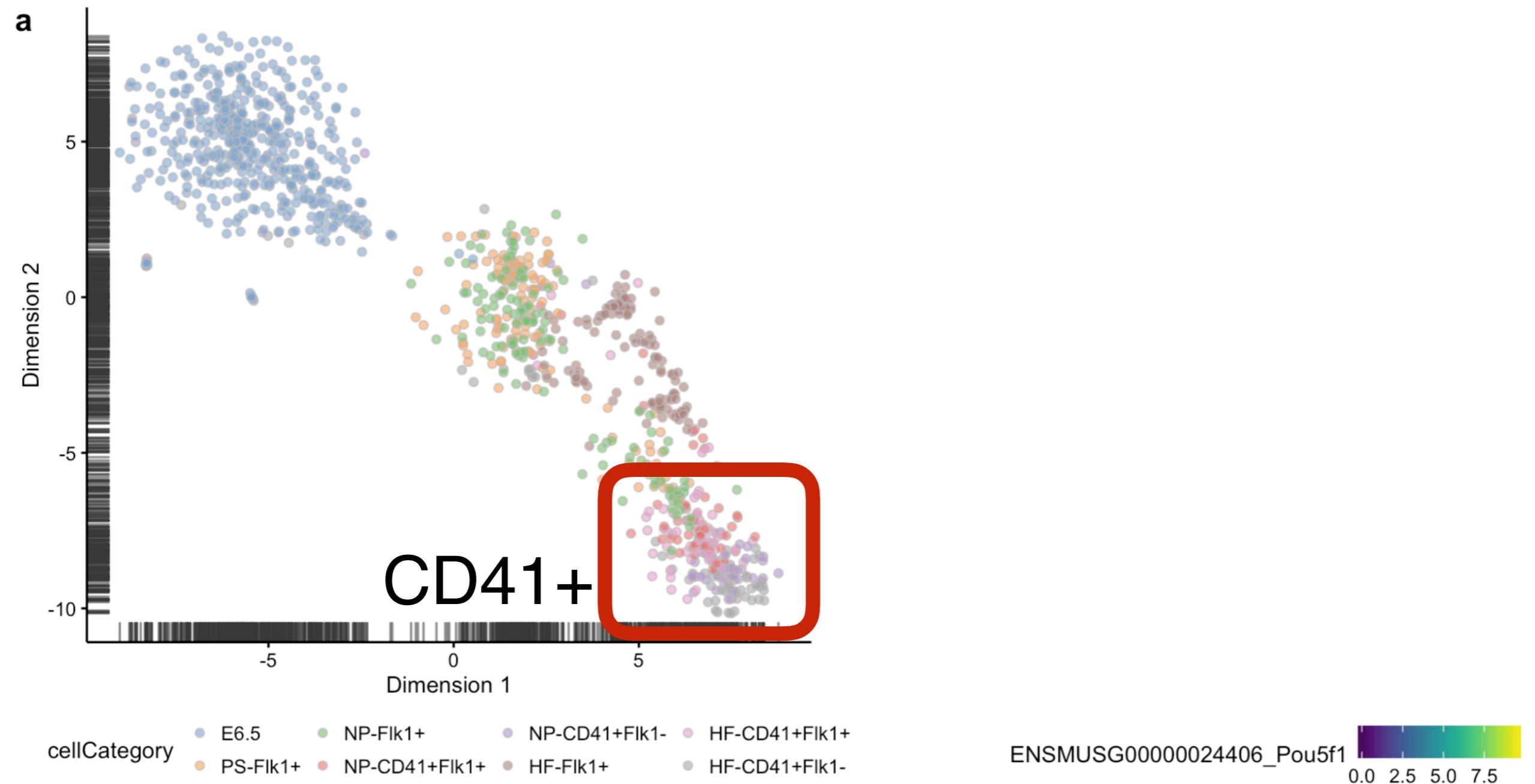
Differentiation to mouse mesoderm (Scialdone et al, *Nature*, 2016)



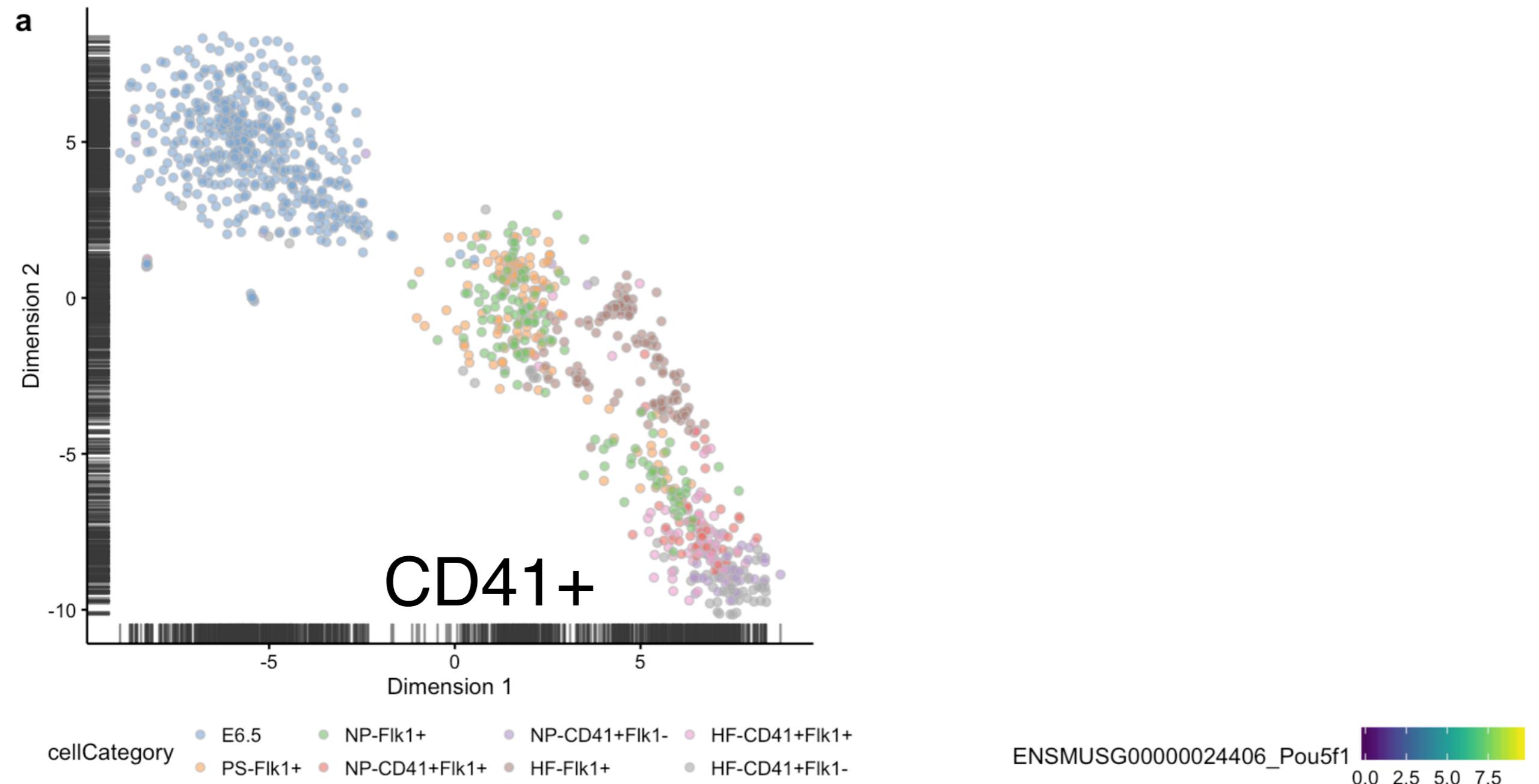
Differentiation to mouse mesoderm (Scialdone et al, *Nature*, 2016)



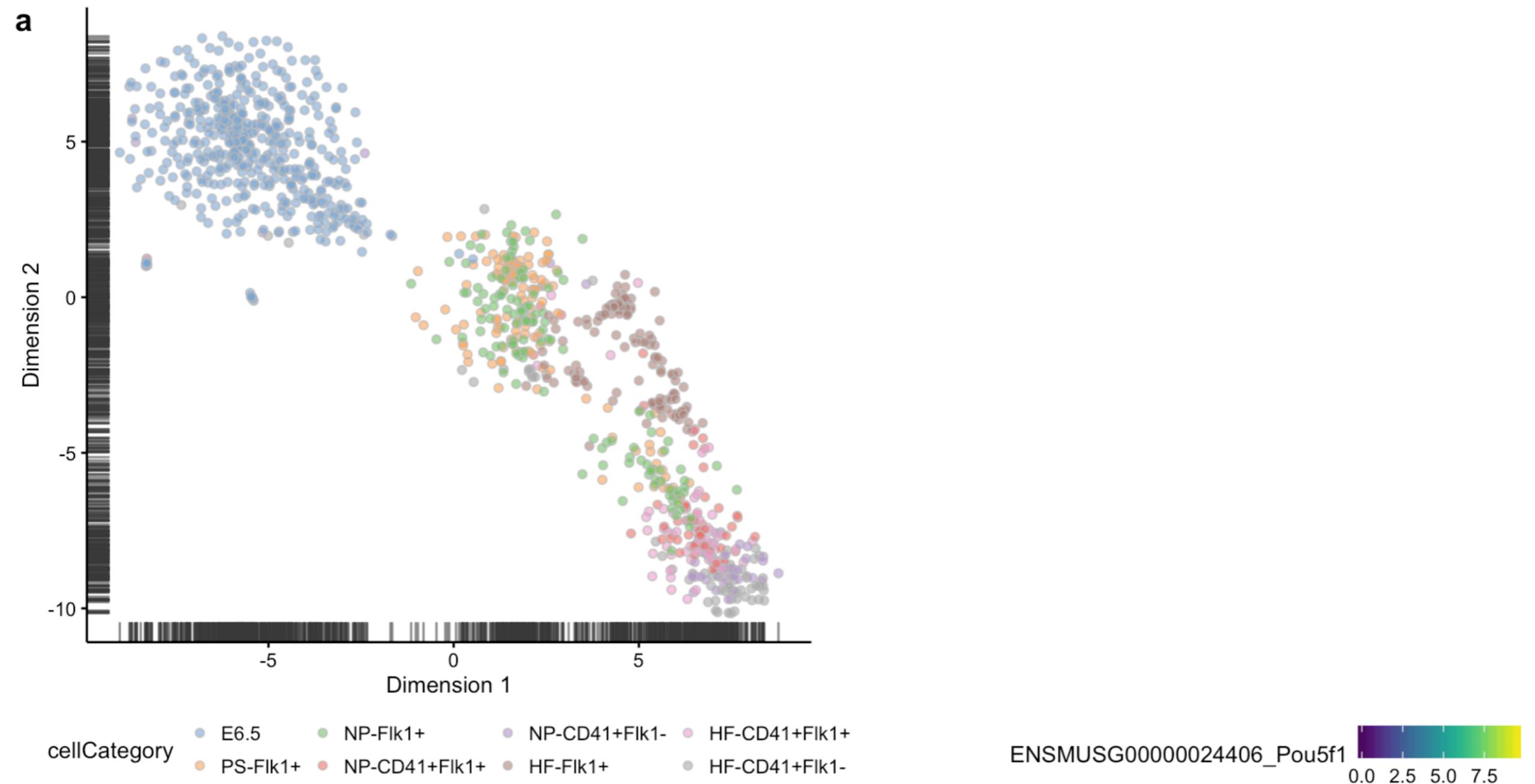
Differentiation to mouse mesoderm (Scialdone et al, *Nature*, 2016)



Differentiation to mouse mesoderm (Scialdone et al, *Nature*, 2016)

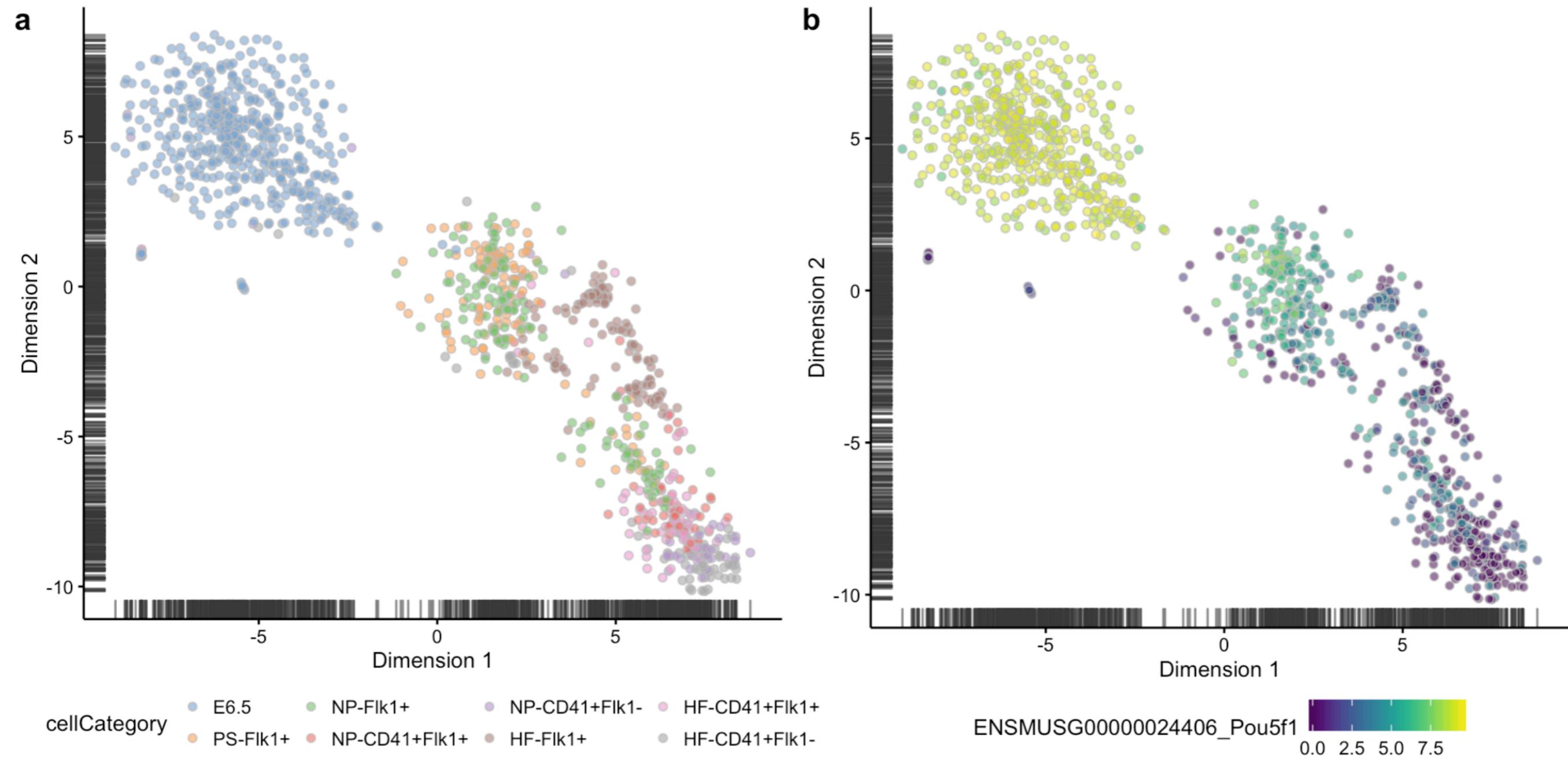


Differentiation to mouse mesoderm (Scialdone et al, *Nature*, 2016)

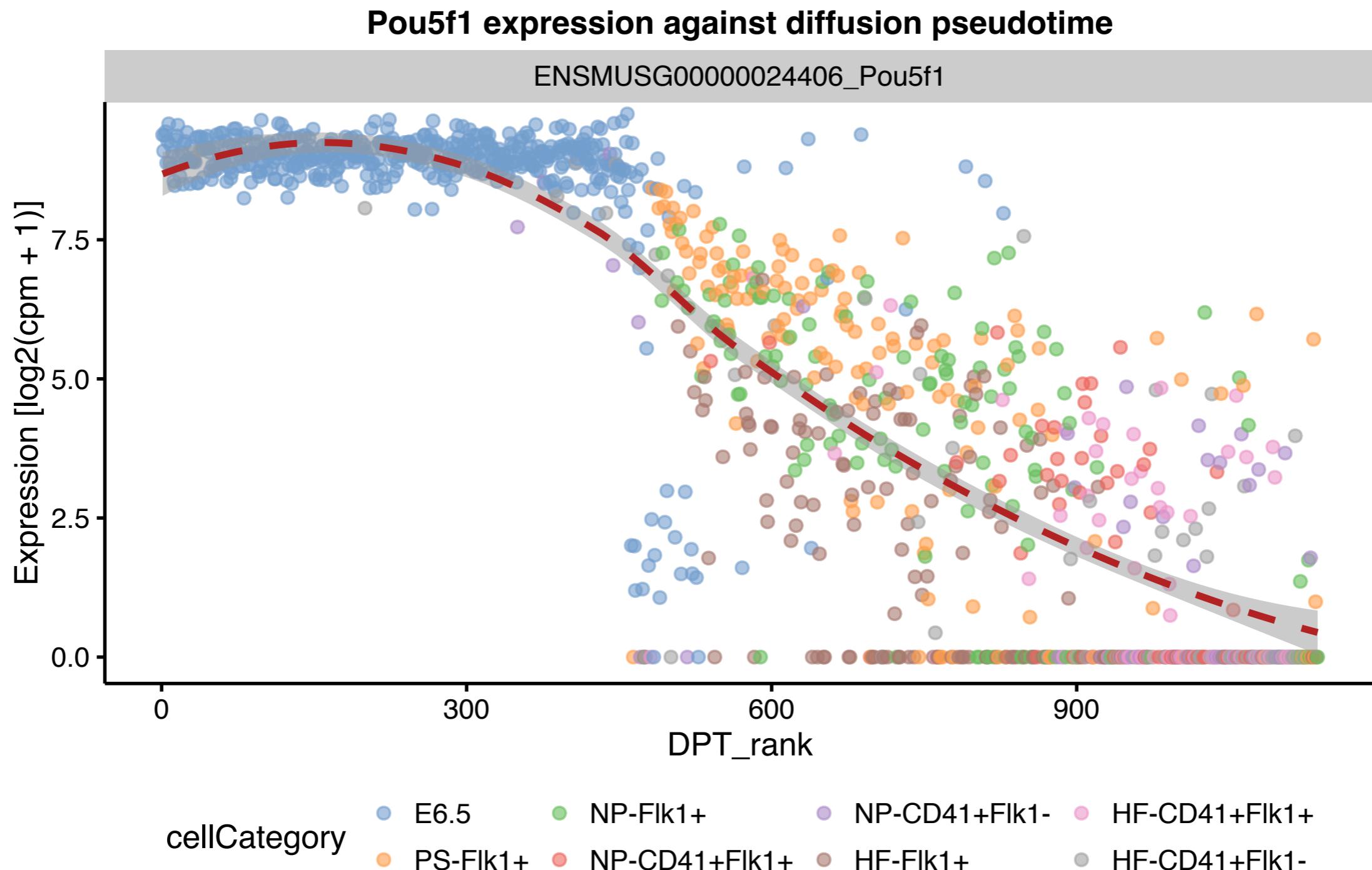


Differentiation to mouse mesoderm (Scialdone et al, *Nature*, 2016)

Pou5f1 (Oct4) Expression



scRNA-seq data is high-resolution and high-dimensional. What how do we make the most of it?



How do we make the most of single-cell RNA-seq data?

scater package for pre-processing, QC,
normalisation and visualisation

<http://bioconductor.org/packages/scater/>

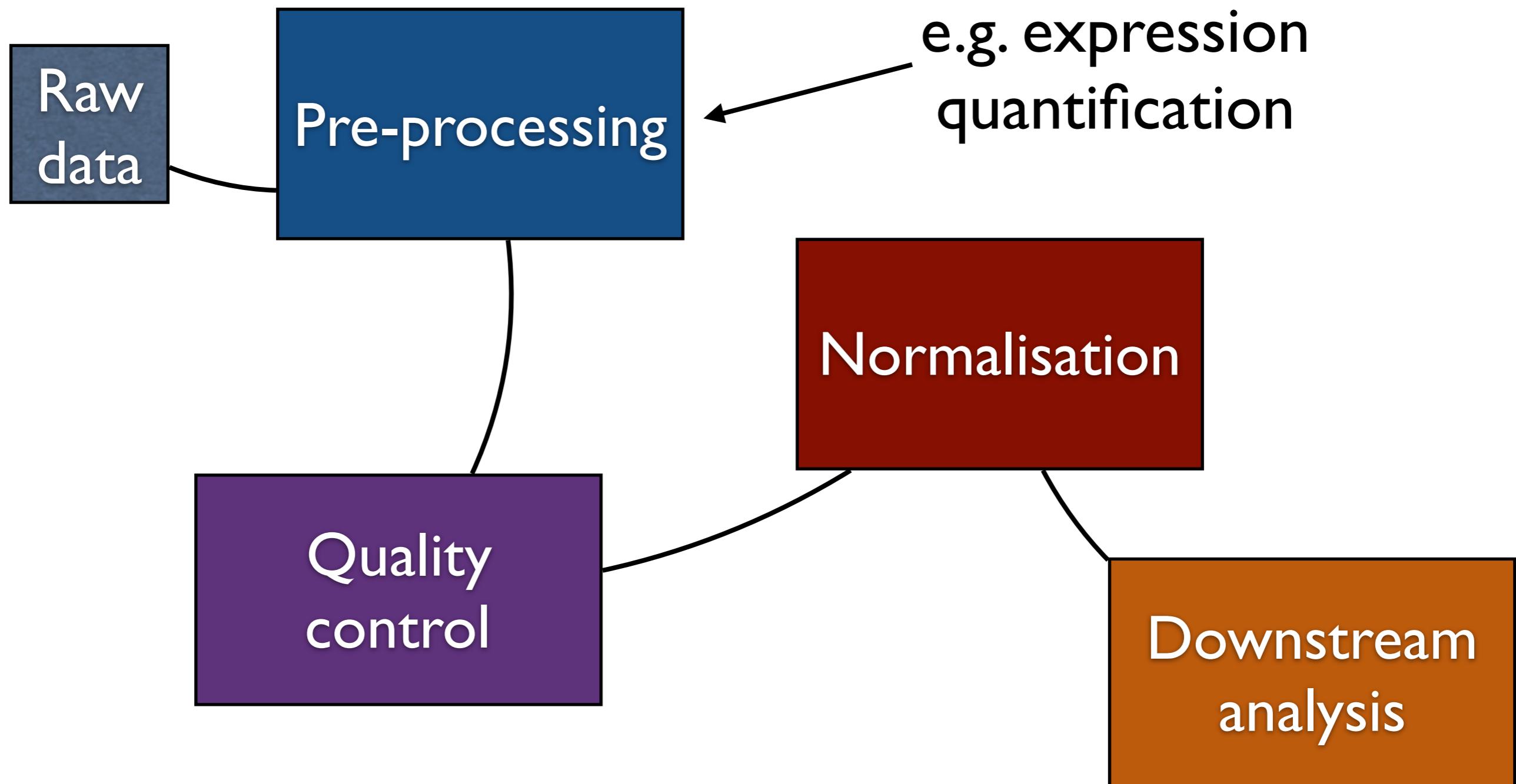
So we'll just assume that all the cells worked well?

So we'll just assume that all the cells worked well?



**Garbage in,
garbage out.**

We need to conduct careful pre-processing, QC and normalisation before getting to any downstream analyses



How do we QC our data?

How do we QC our data?

- This depends on context, so is hands on

How do we QC our data?

- This depends on context, so is hands on
- Visualise it

How do we QC our data?

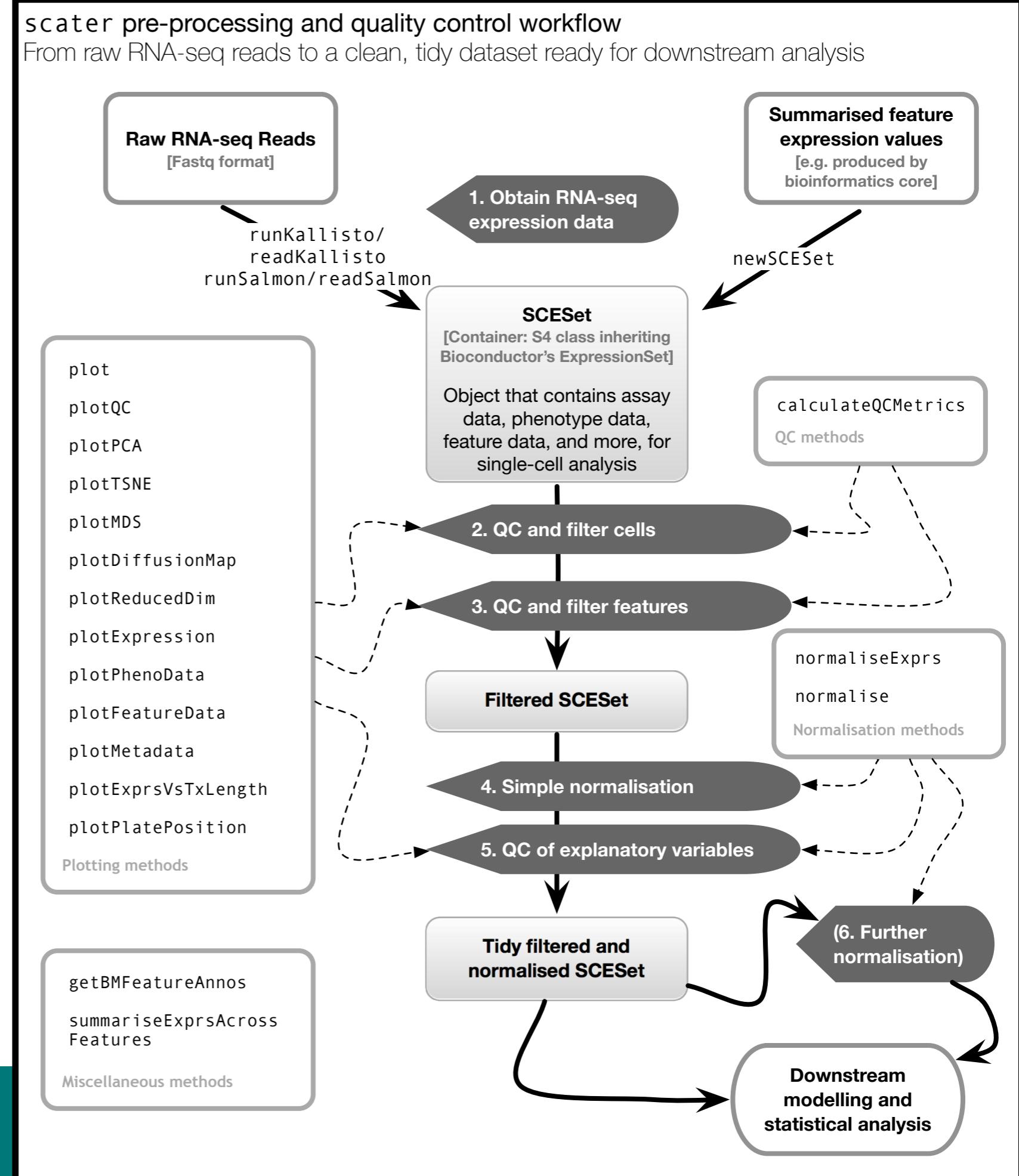
- This depends on context, so is hands on
- Visualise it
- Calculate gene-level and cell-level QC metrics

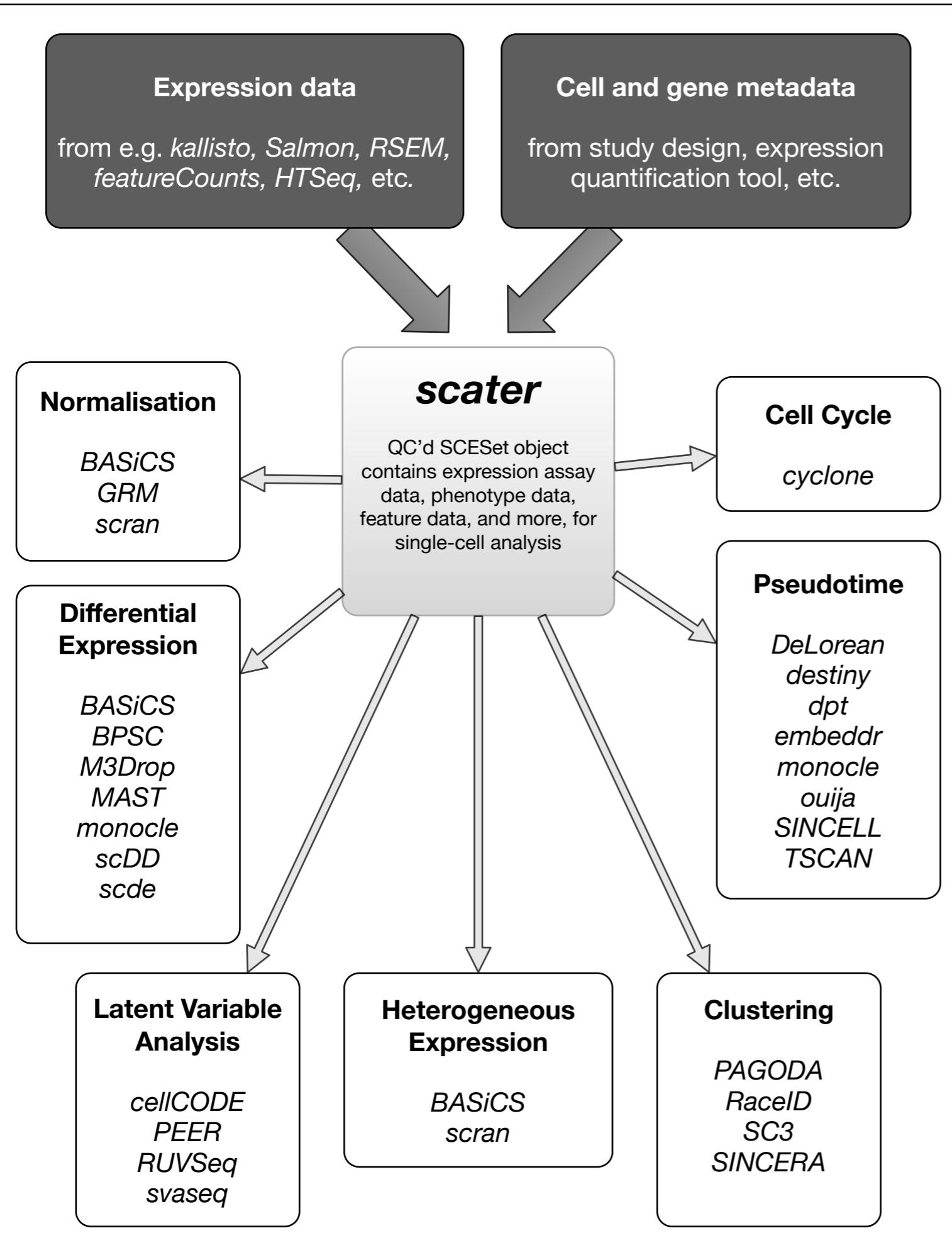
How do we QC our data?

- This depends on context, so is hands on
- Visualise it
- Calculate gene-level and cell-level QC metrics
- Use QC metrics to filter out potentially problematic genes and cells

scater workflow: powerful and flexible

- Integrated with *Salmon* + *kallisto*
- Automatic calculation of QC metrics
- QC diagnostic plots
- Cell and gene filtering
- Simple normalisation
- Sophisticated data structure for single-cell data
- Beautiful plots
- Shiny GUI



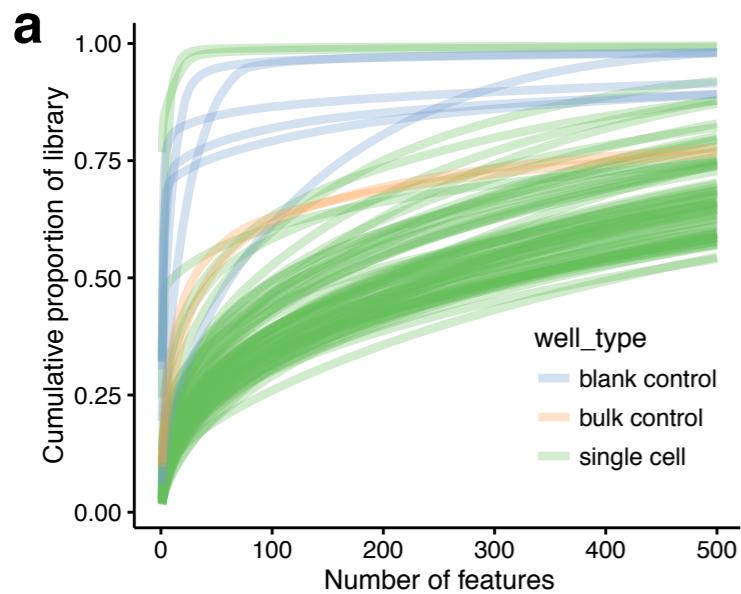


Little example

# cells	C1 Machine 1	C1 Machine 2
Patient A	11	13
Patient B	25	24

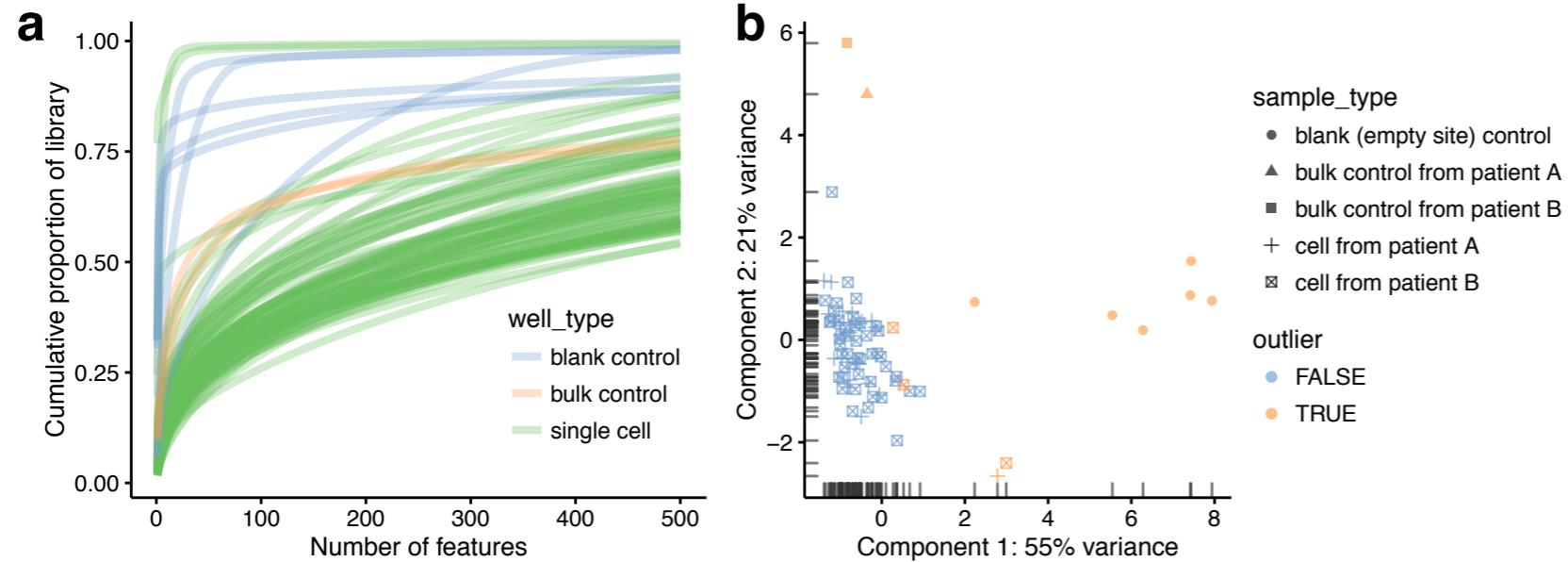
For detailed case study see *scater* preprint on bioRxiv
doi: <http://dx.doi.org/10.1101/069633>

QC diagnostic plots in scater



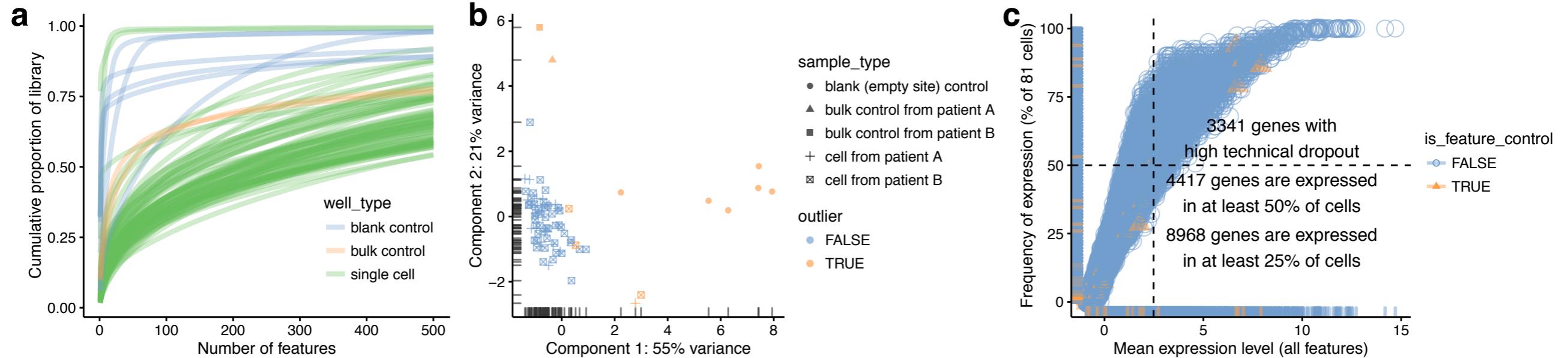
Then filter problematic genes and cells

QC diagnostic plots in scater



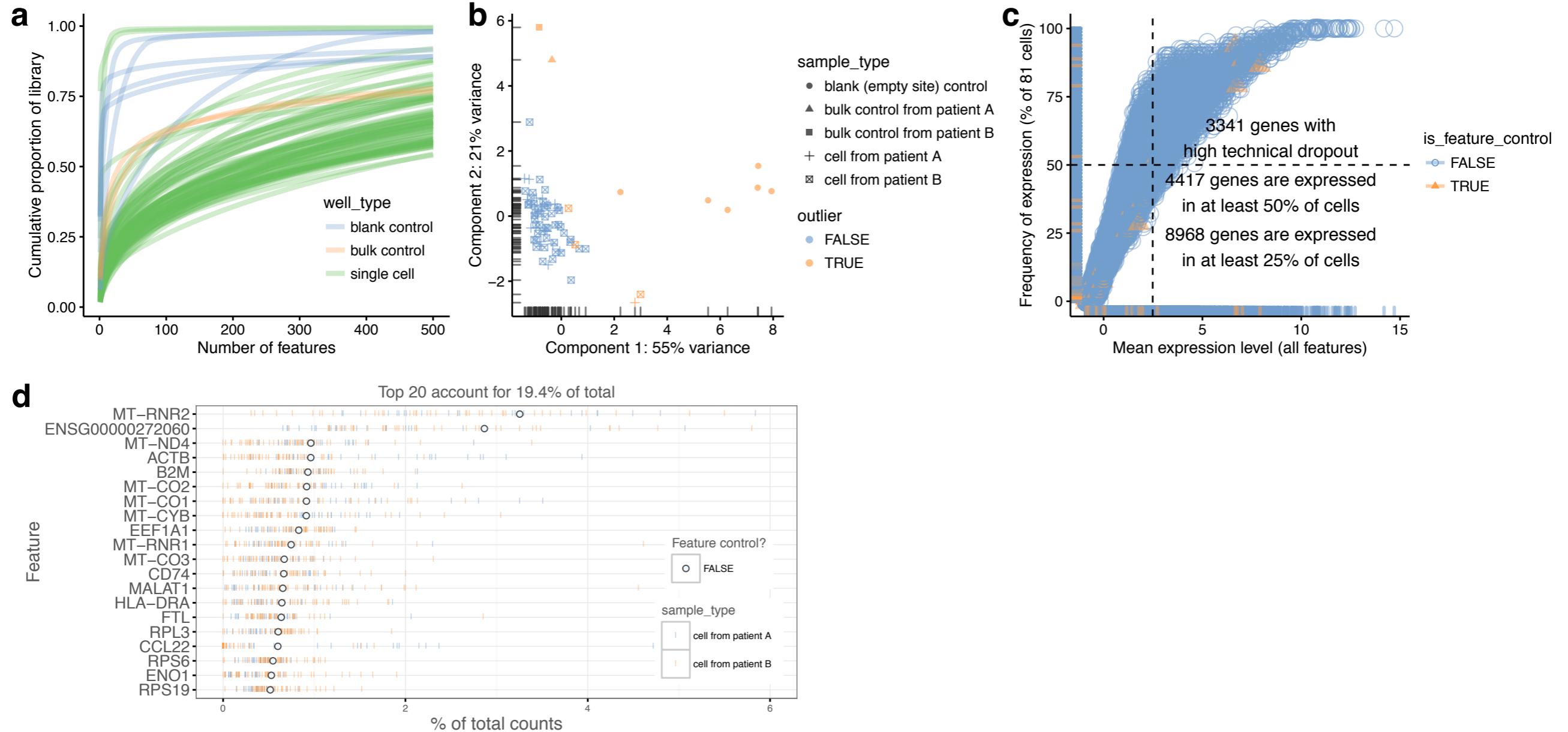
Then filter problematic genes and cells

QC diagnostic plots in scater



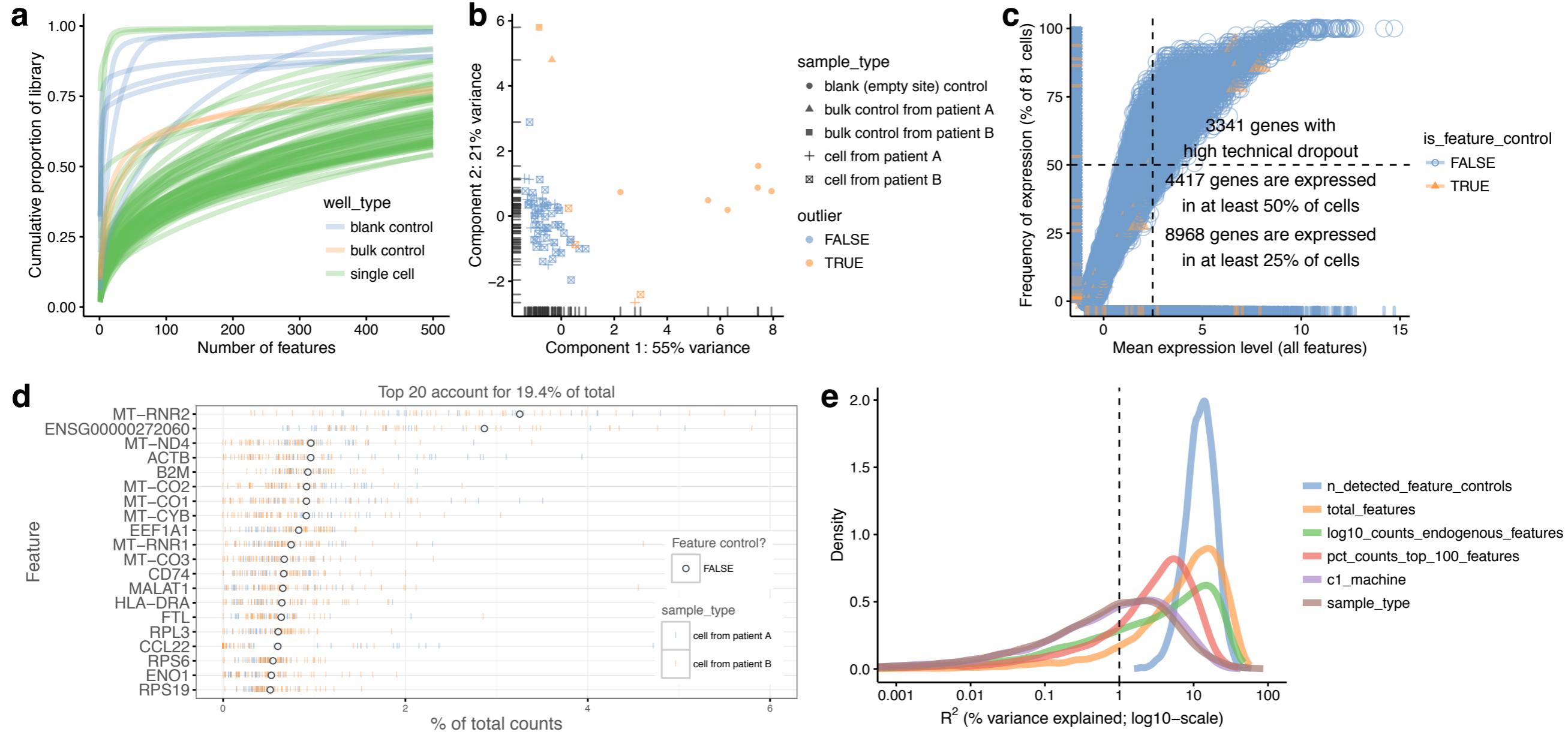
Then filter problematic genes and cells

QC diagnostic plots in scater



Then filter problematic genes and cells

QC diagnostic plots in scater



Then filter problematic genes and cells

How do we normalise scRNA-seq data?

How do we normalise scRNA-seq data?

- Still an open research question...

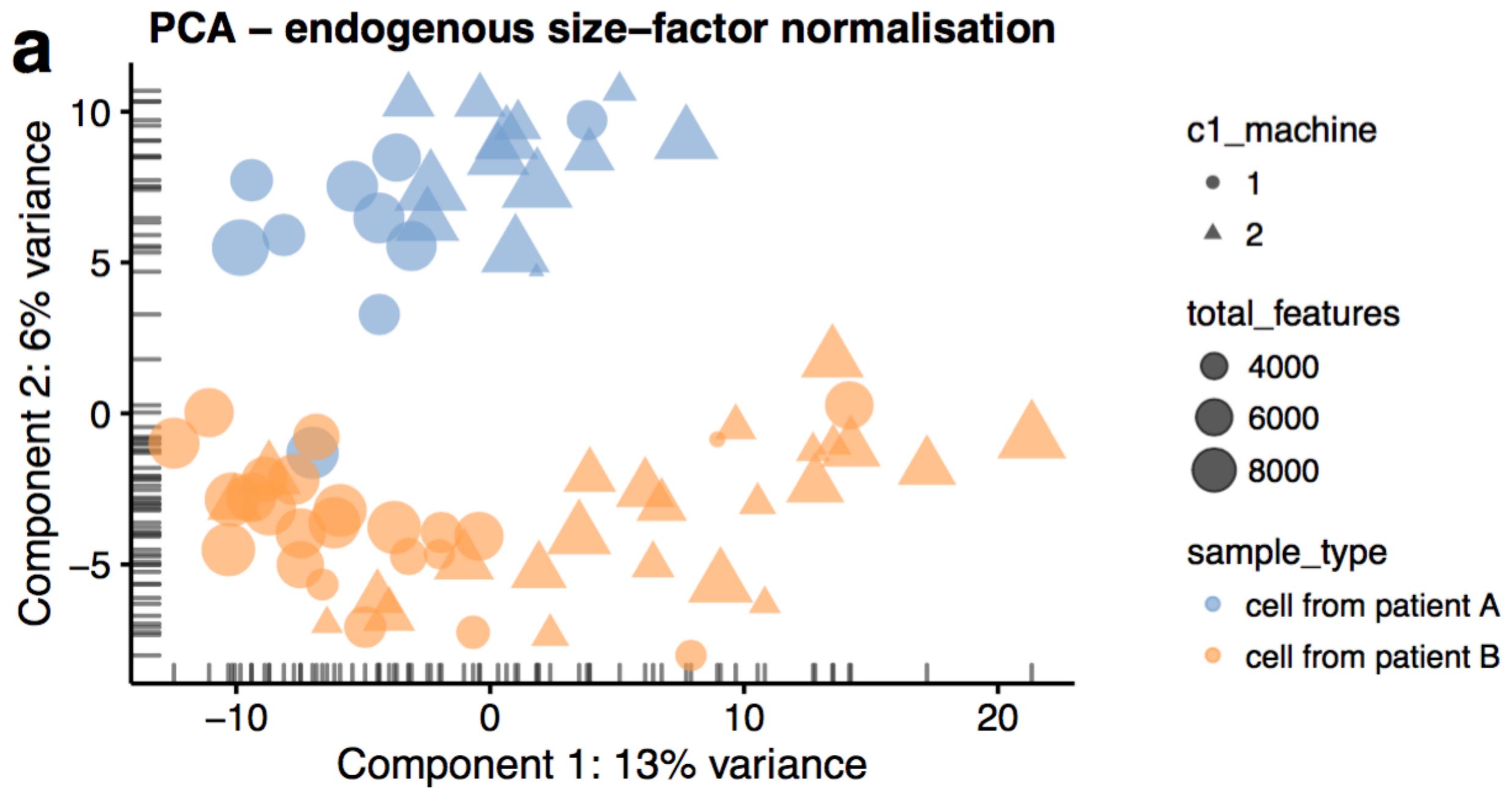
How do we normalise scRNA-seq data?

- Still an open research question...
- It depends on the nature of your study and the scientific questions you wish to answer

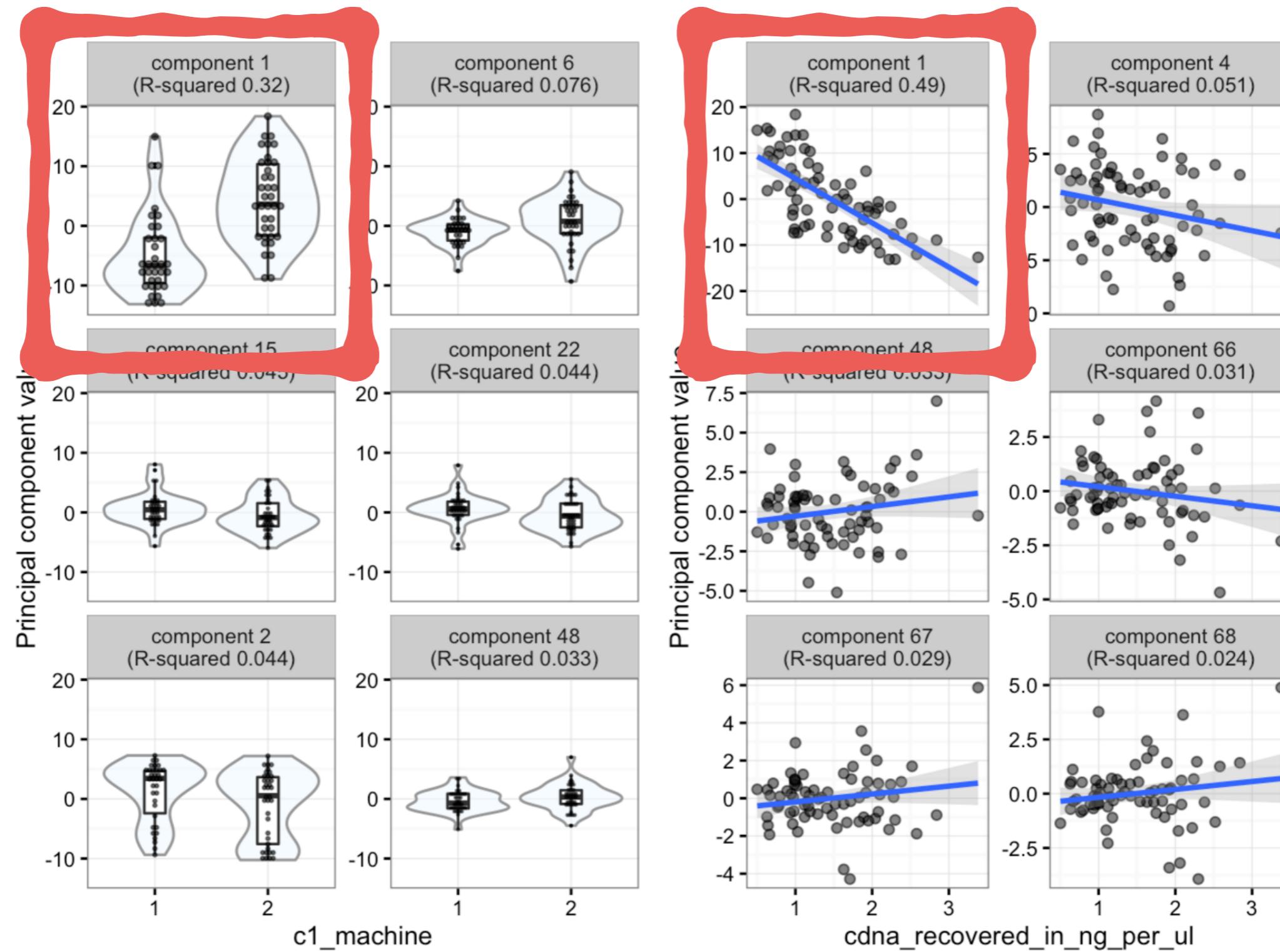
How do we normalise scRNA-seq data?

- Still an open research question...
- It depends on the nature of your study and the scientific questions you wish to answer
- One good option is the single-cell specific method for computing scaling normalisation factors from the *scran* package* - tightly integrated with *scater*

But scaling normalisation does not remove batch effects



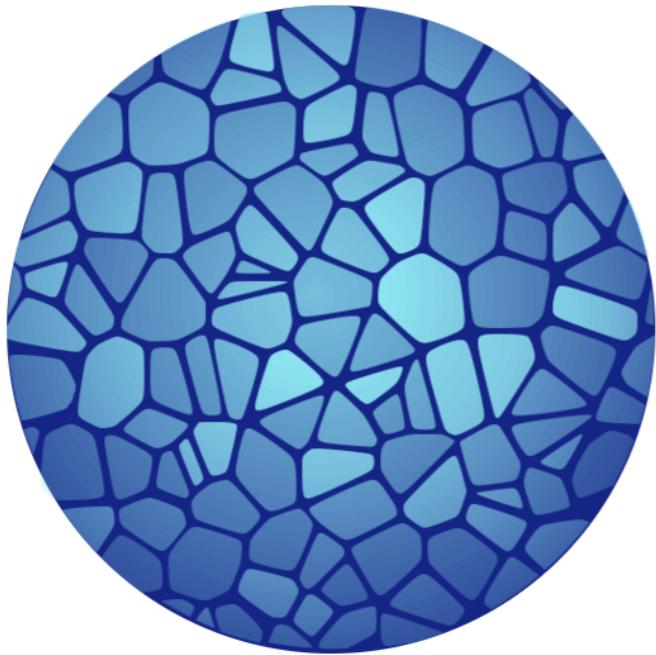
Big effect from C1 machine used



Plots from plotQC in scater - see pre-print

EMBL-EBI





HUMAN
CELL
ATLAS

www.humancellatlas.org

A Google Maps for the Human Body

A group of scientists has taken the first important steps towards creating the Human Cell Atlas—a complete inventory of our staggeringly diverse cells.

ED YONG | OCT 14, 2016 | SCIENCE

coverage in *The Atlantic*

So what do we need to make the most of scRNA-seq data in statistical/computational analyses?

<http://bioconductor.org/packages/scater/>

So what do we need to make the most of scRNA-seq data in statistical/computational analyses?

- Carefully designed studies...***can be done #believe***

<http://bioconductor.org/packages/scater/>

So what do we need to make the most of scRNA-seq data in statistical/computational analyses?

- Carefully designed studies...***can be done #believe***
- Rigorous and flexible QC of cells and genes...***scater***

<http://bioconductor.org/packages/scater/>

So what do we need to make the most of scRNA-seq data in statistical/computational analyses?

- Carefully designed studies...***can be done #believe***
- Rigorous and flexible QC of cells and genes...***scater***
- Appropriate handling of batch effects...***more needed***

<http://bioconductor.org/packages/scater/>

So what do we need to make the most of scRNA-seq data in statistical/computational analyses?

- Carefully designed studies...***can be done #believe***
- Rigorous and flexible QC of cells and genes...***scater***
- Appropriate handling of batch effects...***more needed***
- General statistical methods and software tools that take into account the challenges of single-cell data, particularly covariates and batch effects....***lots of work to do!***

<http://bioconductor.org/packages/scater/>

Acknowledgements

Stegle Lab (EMBL-EBI):

Oliver Stegle
Christof Angermueller
Damian Arnol
Amelie Baud
Florian Buettner
Na Cai
Francesco Paolo Casale
Danilo Horta
Kate Howell
Will Jones
Elo Madissoon
Bogdan Mirauta
Raghd Rostom
Roland Schwarz
Lara Urban
Verena Zuber

scater co-authors:

Kieran Campbell (Oxford)
Aaron Lun (EMBL-EBI/CRUK Cambridge)
Quin Wills (Oxford)

scater: R/Bioconductor package for pre-processing, QC, normalisation and visualisation

- **<http://bioconductor.org/packages/scater/>**
- bioRxiv pre-print: **<http://dx.doi.org/10.1101/069633>***
- “A step-by-step workflow for low-level analysis of single-cell RNA-seq data”:
<http://f1000research.com/articles/5-2122/v1>

Thanks!



#RCatLadies #ResearchParasites

EMBL-EBI

Thanks!



#RCatLadies #ResearchParasites

EMBL-EBI