

# Bioinformatics Awareness Day (BAD) – Single Cell RNAseq Hackathon

***Valeria Policastro***

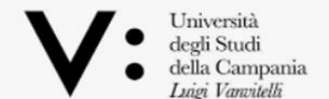
*National Research Council (Italy)*

*Institute for applied mathematics "Mauro Picone" (IAC)*

*Tutor: Annamaria Carissimo, Claudia Angelini*

*valeria.policastro@gmail.com*

*<https://github.com/ValeriaPolicastro>*



- 1. Seurat Object***
- 2. Cell Data Set Object (Monocle3)***
- 3. Data***
- 4. Groups***

## Seurat Object

### Count matrix

```
> tabData[1:6,1:4]
6 x 4 sparse Matrix of class "dgCMatrix"
      10X_P4_0_AAACCTGAGATTACCC 10X_P4_0_AAACCTGAGTGCCAGA 10X_P4_0_AAACCTGCAATCCGT 10X_P4_0_AAACCTGGTAATCGTC
xkr4      .                  .                  .                  .
Rp1       .                  .                  .                  .
Sox17     .                  .                  .                  .
Mrp115    2                  1                  .                  2
Lyp1a1    .                  .                  1                  .
Tcea1     1                  .                  1                  3
```

```
> tabmur <- CreateSeuratObject(counts = tabData, project = "tabulamuris", min.cells = round(dim(tabData)[2]*5/100), min.features = 0, meta.data=info)
> tabmur
An object of class Seurat
8388 features across 58000 samples within 1 assay
Active assay: RNA (8388 features, 0 variable features)
```

8388 genes and 58000 cells

Include features detected in at least this many cells

Include cells where at least this many features are detected

Project name

### Cells information

```
> info[1:5,1:7]
      id tissue subtissue      cell_ontology_class      channel mouse run
10X_P4_0_AAACCTGAGATTACCC 10X_P4_0_AAACCTGAGATTACCC Tongue    <NA> basal cell of epidermis 10X_P4_0 3-M-8 P4
10X_P4_0_AAACCTGAGTGCCAGA 10X_P4_0_AAACCTGAGTGCCAGA Tongue    <NA>      keratinocyte 10X_P4_0 3-M-8 P4
10X_P4_0_AAACCTGCAATCCGT 10X_P4_0_AAACCTGCAATCCGT Tongue    <NA>      keratinocyte 10X_P4_0 3-M-8 P4
10X_P4_0_AAACCTGGTAATCGTC 10X_P4_0_AAACCTGGTAATCGTC Tongue    <NA> basal cell of epidermis 10X_P4_0 3-M-8 P4
10X_P4_0_AAACCTGGTCCAACCT 10X_P4_0_AAACCTGGTCCAACCT Tongue    <NA> basal cell of epidermis 10X_P4_0 3-M-8 P4
```

# SeuratObject

```
> str(tabmur)
Formal class 'Seurat' [package "SeuratObject"] with 13 slots
..@ assays :List of 1
.. ..$ RNA:Formal class 'Assay' [package "SeuratObject"] with 8 slots
.. .. ..@ counts :Formal class 'dgCMatix' [package "Matrix"] with 6 slots
.. .. .. ..@ i :int [1:92061775] 0 2 6 15 17 18 20 21 26 27 ...
.. .. .. ..@ p :int [1:58001] 0 2604 5323 8241 11107 14472 17258 19531 23354 25487 ...
.. .. .. ..@ Dim :int [1:2] 8388 58000
.. .. .. ..@ Dimnames:List of 2
.. .. .. .. ..$ :chr [1:8388] "Mrpl15" "Lyplal" "Tcea1" "Atp6v1h" ...
.. .. .. .. ..$ :chr [1:58000] "10X_P4_0_AAACCTGAGATTACCC" "10X_P4_0_AAACCTGAGTGCCAGA" "10X_P4_0_AAACCTGCAAAATCCGT" "10X_P4_0_AAACCTGGTAATCGTC" ...
.. .. .. ..@ x :num [1:92061775] 2 1 1 1 1 49 1 3 1 1 ...
.. .. .. ..@ factors :list()
.. .. .. ..@ data :Formal class 'dgCMatix' [package "Matrix"] with 6 slots
.. .. .. .. ..@ i :int [1:92061775] 0 2 6 15 17 18 20 21 26 27 ...
.. .. .. .. ..@ p :int [1:58001] 0 2604 5323 8241 11107 14472 17258 19531 23354 25487 ...
.. .. .. .. ..@ Dim :int [1:2] 8388 58000
.. .. .. .. ..@ Dimnames:List of 2
.. .. .. .. .. ..$ :chr [1:8388] "Mrpl15" "Lyplal" "Tcea1" "Atp6v1h" ...
.. .. .. .. .. ..$ :chr [1:58000] "10X_P4_0_AAACCTGAGATTACCC" "10X_P4_0_AAACCTGAGTGCCAGA" "10X_P4_0_AAACCTGCAAAATCCGT" "10X_P4_0_AAACCTGGTAATCGTC" ...
.. .. .. .. ..@ x :num [1:92061775] 2 1 1 1 1 49 1 3 1 1 ...
.. .. .. .. ..@ factors :list()
.. .. .. ..@ scale.data :num[0, 0 ]
.. .. .. ..@ key :chr "rna_"
.. .. .. ..@ assay.orig :NULL
.. .. .. ..@ var.features :logi(0)
.. .. .. ..@ meta.features:'data.frame': 8388 obs. of 0 variables
.. .. .. ..@ misc :list()
.. .. .. ..@ meta.data :'data.frame': 58000 obs. of 10 variables:
.. .. .. .. ..$ orig.ident :Factor w/ 1 level "10X": 1 1 1 1 1 1 1 1 1 1 ...
.. .. .. .. ..$ nCount_RNA :num [1:58000] 9502 16900 20906 16560 19081 ...
.. .. .. .. ..$ nFeature_RNA :int [1:58000] 2604 2719 2918 2866 3365 2786 2273 3823 2133 2432 ...
.. .. .. .. ..$ id :chr [1:58000] "10X_P4_0_AAACCTGAGATTACCC" "10X_P4_0_AAACCTGAGTGCCAGA" "10X_P4_0_AAACCTGCAAAATCCGT" "10X_P4_0_AAACCTGGTAATCGTC" ...
.. .. .. .. ..$ tissue :chr [1:58000] "Tongue" "Tongue" "Tongue" "Tongue" ...
.. .. .. .. ..$ subtissue :chr [1:58000] NA NA NA NA ...
.. .. .. .. ..$ cell_ontology_class :chr [1:58000] "basal cell of epidermis" "keratinocyte" "keratinocyte" "basal cell of epidermis" ...
.. .. .. .. ..$ channel :chr [1:58000] "10X_P4_0" "10X_P4_0" "10X_P4_0" "10X_P4_0" ...
.. .. .. .. ..$ mouse :chr [1:58000] "3-M-8" "3-M-8" "3-M-8" "3-M-8" ...
.. .. .. .. ..$ run :chr [1:58000] "P4" "P4" "P4" "P4" ...
.. .. .. ..@ active.assay: chr "RNA"
.. .. .. ..@ active.ident: Factor w/ 1 level "10X": 1 1 1 1 1 1 1 1 1 1 ...
.. .. .. ..- attr(*, "names")= chr [1:58000] "10X_P4_0_AAACCTGAGATTACCC" "10X_P4_0_AAACCTGAGTGCCAGA" "10X_P4_0_AAACCTGCAAAATCCGT" "10X_P4_0_AAACCTGGTAATCGTC" ...
.. .. .. ..@ graphs :list()
.. .. .. ..@ neighbors :list()
.. .. .. ..@ reductions :list()
.. .. .. ..@ images :list()
.. .. .. ..@ project.name: chr "tabulamuris"
.. .. .. ..@ misc :list()
.. .. .. ..@ version :Classes 'package_version', 'numeric_version' hidden list of 1
.. .. .. ..$ :int [1:3] 4 1 0
.. .. .. ..@ commands :list()
.. .. .. ..@ tools :list()
```

# SeuratObject

## Assays

Formal class 'Seurat' [package "SeuratObject"] with 13 slots

```
..@ assays :List of 1
.. ..$ RNA:Formal class 'Assay' [package "SeuratObject"] with 8 slots
.. .. ..@ counts :Formal class 'dgCMatrix' [package "Matrix"] with 6 slots
.. .. .. ..@ i :int [1:92061775] 0 2 6 15 17 18 20 21 26 27 ...
.. .. .. ..@ p :int [1:58001] 0 2604 5323 8241 11107 14472 17258 19531 23354 25487 ...
.. .. .. ..@ Dim :int [1:2] 8388 58000
.. .. .. ..@ Dimnames:List of 2
.. .. .. .. ..$ :chr [1:8388] "Mrpl15" "Lypla1" "Tcea1" "Atp6v1h" ...
.. .. .. .. ..$ :chr [1:58000] "10X_P4_0_AAACCTGAGATTACCC" "10X_P4_0_AAACCTGAGTGCCAGA" "10X_P4_0_AAACCTGCAAATCCGT" "10X_P4_0_AAACCTGGTAATCGTC" ...
.. .. .. ..@ x :num [1:92061775] 2 1 1 1 1 49 1 3 1 1 ...
.. .. .. ..@ factors :list()
.. .. ..@ data :Formal class 'dgCMatrix' [package "Matrix"] with 6 slots
.. .. .. ..@ i :int [1:92061775] 0 2 6 15 17 18 20 21 26 27 ...
.. .. .. ..@ p :int [1:58001] 0 2604 5323 8241 11107 14472 17258 19531 23354 25487 ...
.. .. .. ..@ Dim :int [1:2] 8388 58000
.. .. .. ..@ Dimnames:List of 2
.. .. .. .. ..$ :chr [1:8388] "Mrpl15" "Lypla1" "Tcea1" "Atp6v1h" ...
.. .. .. .. ..$ :chr [1:58000] "10X_P4_0_AAACCTGAGATTACCC" "10X_P4_0_AAACCTGAGTGCCAGA" "10X_P4_0_AAACCTGCAAATCCGT" "10X_P4_0_AAACCTGGTAATCGTC" ...
.. .. .. ..@ x :num [1:92061775] 2 1 1 1 1 49 1 3 1 1 ...
.. .. .. ..@ factors :list()
.. .. ..@ scale.data :num[0 , 0 ]
.. .. ..@ key :chr "rna_"
.. .. ..@ assay.orig :NULL
.. .. ..@ var.features :logi(0)
.. .. ..@ meta.features:'data.frame': 8388 obs. of 0 variables
```

## Meta data

```
..@ meta.data :data.frame': 58000 obs. of 10 variables:
.. ..$ orig.ident : Factor w/ 1 level "10X": 1 1 1 1 1 1 1 1 1 1 ...
.. ..$ nCount_RNA : num [1:58000] 9502 16900 20906 16560 19081 ...
.. ..$ nFeature_RNA : int [1:58000] 2604 2719 2918 2866 3365 2786 2273 3823 2133 2432 ...
.. ..$ id : chr [1:58000] "10X_P4_0_AAACCTGAGATTACCC" "10X_P4_0_AAACCTGAGTGCCAGA" "10X_P4_0_AAACCTGCAAATCCGT" "10X_P4_0_AAACCTGGTAATCGTC" ...
.. ..$ tissue : chr [1:58000] "Tongue" "Tongue" "Tongue" "Tongue" ...
.. ..$ subtissue : chr [1:58000] NA NA NA NA ...
.. ..$ cell_ontology_class: chr [1:58000] "basal cell of epidermis" "keratinocyte" "keratinocyte" "basal cell of epidermis" ...
.. ..$ channel : chr [1:58000] "10X_P4_0" "10X_P4_0" "10X_P4_0" "10X_P4_0" ...
.. ..$ mouse : chr [1:58000] "3-M-8" "3-M-8" "3-M-8" "3-M-8" ...
.. ..$ run : chr [1:58000] "P4" "P4" "P4" "P4" ...
```

## Analysis

```
..@ active.assay: chr "RNA"
..@ active.ident: Factor w/ 1 level "10X": 1 1 1 1 1 1 1 1 1 1 ...
.. .. attr(*, "names")= chr [1:58000] "10X_P4_0_AAACCTGAGATTACCC" "10X_P4_0_AAACCTGAGTGCCAGA" "10X_P4_0_AAACCTGCAAATCCGT" "10X_P4_0_AAACCTGGTAATCGTC" ...
..@ graphs :list()
..@ neighbors :list()
..@ reductions :list()
..@ images :list()
..@ project.name: chr "tabulamuris"
..@ misc :list()
..@ version :classes 'package_version', 'numeric_version' hidden list of 1
.. ..$ :int [1:3] 4 1 0
..@ commands :list()
..@ tools :list()
```

# SeuratObject

```
> str(tabmur)
Formal class 'Seurat' [package "SeuratObject"] with 13 slots
..@ assays :List of 1
.. ..$ RNA:Formal class 'Assay' [package "SeuratObject"] with 8 slots
.. .. .@ counts :Formal class 'dgCMatrx' [package "Matrix"] with 6 slots
.. .. .@ i : int [1:92061775] 0 2 6 15 17 18 20 21 26 27 ...
.. .. .@ p : int [1:58001] 0 2604 5323 8241 11107 14472 17258 19531 23354 25487 ...
.. .. .@ Dim : int [1:2] 8388 58000
.. .. .@ Dimnames:List of 2
.. .. . . $ : chr [1:8388] "Mrp115" "Lyp1a1" "Tcea1" "Atp6v1h" ...
.. .. . . $ : chr [1:58000] "10X_P4_0_AAACCTGAGATTACCC" "10X_P4_0_AAACCTGAGTGCCAGA" "10X_P4_0_AAACCTGCAAAATCCGT" "10X_P4_0_AAACCTGGTAATCGTC" ...
.. .. .@ x : num [1:92061775] 2 1 1 1 1 49 1 3 1 1 ...
.. .. .@ factors : list()
.. .. .@ data :Formal class 'dgCMatrx' [package "Matrix"] with 6 slots
.. .. .@ i : int [1:92061775] 0 2 6 15 17 18 20 21 26 27 ...
.. .. .@ p : int [1:58001] 0 2604 5323 8241 11107 14472 17258 19531 23354 25487 ...
.. .. .@ Dim : int [1:2] 8388 58000
.. .. .@ Dimnames:List of 2
.. .. . . $ : chr [1:8388] "Mrp115" "Lyp1a1" "Tcea1" "Atp6v1h" ...
.. .. . . $ : chr [1:58000] "10X_P4_0_AAACCTGAGATTACCC" "10X_P4_0_AAACCTGAGTGCCAGA" "10X_P4_0_AAACCTGCAAAATCCGT" "10X_P4_0_AAACCTGGTAATCGTC" ...
.. .. .@ x : num [1:92061775] 2 1 1 1 1 49 1 3 1 1 ...
.. .. .@ factors : list()
.. .. .@ scale.data : num[0 , 0 ]
.. .. .@ key : chr "rna_"
.. .. .@ assay.orig : NULL
.. .. .@ var.features : logi(0)
.. .. .@ meta.features:'data.frame': 8388 obs. of 0 variables
.. .. .@ misc : list()
..@ meta.data : 'data.frame': 58000 obs. of 10 variables:
.. ..$ orig.ident : Factor w/ 1 level "10X": 1 1 1 1 1 1 1 1 1 1 ...
.. ..$ nCount_RNA : num [1:58000] 9502 16900 20906 16560 19081 ...
.. ..$ nFeature_RNA : int [1:58000] 2604 2719 2918 2866 3365 2786 2273 2432 ...
.. ..$ id : chr [1:58000] "10X_P4_0_AAACCTGAGATTACCC" "10X_P4_0_AAACCTGAGTGCCAGA" "10X_P4_0_AAACCTGCAAAATCCGT" "10X_P4_0_AAACCTGGTAATCGTC" ...
.. ..$ tissue : chr [1:58000] "Tongue" "Tongue" "Tongue" "Tongue" ...
.. ..$ subtissue : chr [1:58000] NA NA NA NA ...
.. ..$ cell_ontology_class : chr [1:58000] "basal cell of epidermis" "keratinocyte" "keratinocyte" "basal cell of epidermis" ...
.. ..$ channel : chr [1:58000] "10X_P4_0" "10X_P4_0" "10X_P4_0" "10X_P4_0" ...
.. ..$ mouse : chr [1:58000] "3-M-8" "3-M-8" "3-M-8" "3-M-8" ...
.. ..$ run : chr [1:58000] "P4" "P4" "P4" "P4" ...
..@ active.assay: chr "RNA"
..@ active.ident: Factor w/ 1 level "10X": 1 1 1 1 1 1 1 1 1 1 ...
..@ attr(,"names")= chr [1:58000] "10X_P4_0_AAACCTGAGATTACCC" "10X_P4_0_AAACCTGAGTGCCAGA" "10X_P4_0_AAACCTGCAAAATCCGT" "10X_P4_0_AAACCTGGTAATCGTC" ...
..@ graphs : list()
..@ neighbors : list()
..@ reductions : list()
..@ images : list()
..@ project.name: chr "tabulamuris"
..@ misc : list()
..@ version :Classes 'package_version', 'numeric_version' hidden list of 1
.. ..$ : int [1:3] 4 1 0
..@ commands : list()
..@ tools : list()
```

# Count Data

```
> tabmur@assays$RNA@counts[1:10,1:5]
10 x 5 sparse Matrix of class "dgCMatrx"
      10X_P7_11_TCCCGATCAAAGTGCG 10X_P4_1_CATATTCCAGTAGAGC 10X_P7_6_AACACGTCATGGTCAT 10X_P7_15_CAGCAGCTCCTGTAGA 10X_P4_7_CGAATGTGTATATCCG
Mrpl15 . . . 1 .
Lyplal . . . . .
Tcea1 . . . 1 .
Atp6v1h . . 1 . .
Rb1cc1 . . . . .
Pcmt1 . . . . .
Rrs1 . . . 1 .
Vcpip1 . . 1 . .
Sgk3 . . . . .
Snhg6 1 . 1 . .
```

```
> tabmur[["RNA"]@counts[1:20,1:4]
20 x 4 sparse Matrix of class "dgCMatrx"
      10X_P7_11_TCCCGATCAAAGTGCG 10X_P4_1_CATATTCCAGTAGAGC 10X_P7_6_AACACGTCATGGTCAT 10X_P7_15_CAGCAGCTCCTGTAGA
Mrpl15 . . . 1 .
Lyplal . . . . .
Tcea1 . . . 1 .
Atp6v1h . . 1 . .
Rb1cc1 . . . . .
Pcmt1 . . . . .
Rrs1 . . . 1 .
Vcpip1 . . 1 . .
Sgk3 . . . . .
Snhg6 1 . 1 . .
Cops5 1 . 1 . .
Cspp1 . . . . .
Arfgef1 1 . . . 1
Prex2 . . . . 1
Ncoa2 . . . . .
Tram1 . 2 . 2 .
Lactb2 . . . . .
Terf1 . . . . .
Rpl7 86 . 73 19 3
Rdh10 . . . . .
```

# Normalized and Scaled Data

```
> tabmur[["RNA"]][@data[1:15,1:4]]
15 x 4 sparse Matrix of class "dgCMatix"
      10X_P7_11_TCCCGATCAAAGTGCG 10X_P4_1_CATATTCCAGTAGAGC 10X_P7_6_AACACGTCATGGTCAT 10X_P7_15_CAGCAGCTCCTGTAGA
Mrpl15      .                  .                  0.9590703      .
Lyp1a1      .                  .                  .              .
Tcea1       .                  .                  .              1.429986
Atp6v1h     .                  .                  0.9590703      .
Rb1cc1      .                  .                  .              .
Pcmt1d1     .                  .                  .              .
Rrs1        .                  .                  .              1.429986
Vcp1p1      .                  .                  0.9590703      .
Sgk3        .                  .                  .              .
Snhg6       0.4574755          0.6755725          .              .
Cops5       0.4574755          0.6755725          .              1.429986
Cspp1       .                  .                  .              .
Arfgef1     0.4574755          .                  .              1.429986
Prex2       .                  .                  .              1.429986
Ncoa2       .                  .                  .              .
```

```
> tabmur[["RNA"]][@scale.data[1:8,1:4]]
      10X_P7_11_TCCCGATCAAAGTGCG 10X_P4_1_CATATTCCAGTAGAGC 10X_P7_6_AACACGTCATGGTCAT 10X_P7_15_CAGCAGCTCCTGTAGA
Mrpl15      -0.5345229          -0.5345229          1.2699875          -0.5345229
Lyp1a1      -0.5032524          -0.5032524          -0.5032524         -0.5032524
Tcea1       -0.6460542          -0.6460542          -0.6460542          1.7872001
Atp6v1h     -0.3362647          -0.3362647          2.3106185          -0.3362647
Rb1cc1      -0.4224019          -0.4224019          -0.4224019         -0.4224019
Pcmt1d1     -0.4979030          -0.4979030          -0.4979030         -0.4979030
Rrs1        -0.4539663          -0.4539663          -0.4539663          2.5502712
Vcp1p1      -0.2685347          -0.2685347          2.9813649          -0.2685347
```



# SeuratObject

```
> str(tabmur)
Formal class 'Seurat' [package "SeuratObject"] with 13 slots
..@ assays :List of 1
.. ..$ RNA:Formal class 'Assay' [package "SeuratObject"] with 8 slots
.. .. ..@ counts :Formal class 'dgCMatrx' [package "Matrix"] with 6 slots
.. .. .. ..@ i :int [1:4675089] 9 10 12 15 18 20 21 37 39 48 ...
.. .. .. ..@ p :int [1:2901] 0 3087 5297 6941 8391 9382 10148 12336 13775 15026 ...
.. .. .. ..@ Dim :int [1:2] 8470 2900
.. .. .. ..@ Dimnames:List of 2
.. .. .. .. ..$ : chr [1:8470] "Mrpl15" "Lyplal" "Tcea1" "Atp6v1h" ...
.. .. .. .. ..$ : chr [1:2900] "10X_P7_11_TCCCGATCAAAGTGCG" "10X_P4_1_CATATTCCAGTAGAGC" "10X_P7_6_AACACGTCATGGTCAT" "10X_P7_15_CAGCAGCTCCTGTAGA" ...
.. .. .. ..@ x :num [1:4675089] 1 1 1 2 86 1 3 1 3 1 ...
.. .. .. ..@ factors :list()
.. .. .. ..@ data :Formal class 'dgCMatrx' [package "Matrix"] with 6 slots
.. .. .. .. ..@ i :int [1:4675089] 9 10 12 15 18 20 21 37 39 48 ...
.. .. .. .. ..@ p :int [1:2901] 0 3087 5297 6941 8391 9382 10148 12336 13775 15026 ...
.. .. .. .. ..@ Dim :int [1:2] 8470 2900
.. .. .. .. ..@ Dimnames:List of 2
.. .. .. .. .. ..$ : chr [1:8470] "Mrpl15" "Lyplal" "Tcea1" "Atp6v1h" ...
.. .. .. .. .. ..$ : chr [1:2900] "10X_P7_11_TCCCGATCAAAGTGCG" "10X_P4_1_CATATTCCAGTAGAGC" "10X_P7_6_AACACGTCATGGTCAT" "10X_P7_15_CAGCAGCTCCTGTAGA" ...
.. .. .. .. ..@ x :num [1:4675089] 0.457 0.457 0.457 0.77 3.93 ...
.. .. .. .. ..@ factors :list()
.. .. .. ..@ scale.data :num [1:8470, 1:2900] -0.535 -0.503 -0.646 -0.336 -0.422 ...
.. .. .. ..- attr(,"dimnames")=List of 2
.. .. .. .. ..$ : chr [1:8470] "Mrpl15" "Lyplal" "Tcea1" "Atp6v1h" ...
.. .. .. .. ..$ : chr [1:2900] "10X_P7_11_TCCCGATCAAAGTGCG" "10X_P4_1_CATATTCCAGTAGAGC" "10X_P7_6_AACACGTCATGGTCAT" "10X_P7_15_CAGCAGCTCCTGTAGA" ...
.. .. .. ..@ key : chr "rna_"
.. .. .. ..@ assay.orig : NULL
.. .. .. ..@ var.features :logi(0)
.. .. .. ..@ meta.features:'data.frame': 8470 obs. of 0 variables
.. .. .. ..@ misc :list()
..@ meta.data :'data.frame': 2900 obs. of 10 variables:
.. ..$ orig.ident :Factor w/ 1 level "10X": 1 1 1 1 1 1 1 1 1 1 ...
.. ..$ nCount_RNA :num [1:2900] 17239 10361 6214 3146 2098 ...
.. ..$ nFeature_RNA :int [1:2900] 3087 2210 1644 1450 991 766 2188 1439 1251 1466 ...
.. ..$ id :chr [1:2900] "10X_P7_11_TCCCGATCAAAGTGCG" "10X_P4_1_CATATTCCAGTAGAGC" "10X_P7_6_AACACGTCATGGTCAT" "10X_P7_15_CAGCAGCTCCTGTAGA" ...
.. ..$ tissue :chr [1:2900] "Thymus" "Tongue" "Spleen" "Limb_Muscle" ...
.. ..$ subtissue :chr [1:2900] NA NA NA NA ...
.. ..$ cell_ontology_class:chr [1:2900] "immature T cell" "basal cell of epidermis" "B cell" "endothelial cell" ...
.. ..$ channel :chr [1:2900] "10X_P7_11" "10X_P4_1" "10X_P7_6" "10X_P7_15" ...
.. ..$ mouse :chr [1:2900] "3-F-56" "3-M-9" "3-F-56" "3-F-57" ...
.. ..$ run :chr [1:2900] "P7" "P4" "P7" "P7" ...
..@ active.assay: chr "RNA"
..@ active.ident: Factor w/ 1 level "10X": 1 1 1 1 1 1 1 1 1 1 ...
.. ..- attr(,"names")= chr [1:2900] "10X_P7_11_TCCCGATCAAAGTGCG" "10X_P4_1_CATATTCCAGTAGAGC" "10X_P7_6_AACACGTCATGGTCAT" "10X_P7_15_CAGCAGCTCCTGTAGA" ...
..@ graphs :list()
..@ neighbors :list()
..@ reductions :list()
..@ images :list()
..@ project.name: chr "tabulamuris"
..@ misc :list()
..@ version :Classes 'package_version', 'numeric_version' hidden list of 1
.. ..$ :int [1:3] 4 1 0
..@ commands :List of 2
```

# Meta Data

```
> tabmur@meta.data[1:4,1:8]
```

	orig.ident	nCount_RNA	nFeature_RNA	id	tissue	subtissue	cell_ontology_class	channel
10X_P7_11_TCCCGATCAAAGTGCG	10X	17239	3087	10X_P7_11_TCCCGATCAAAGTGCG	Thymus	<NA>	immature T cell	10X_P7_11
10X_P4_1_CATATTCCAGTAGAGC	10X	10361	2210	10X_P4_1_CATATTCCAGTAGAGC	Tongue	<NA>	basal cell of epidermis	10X_P4_1
10X_P7_6_AACACGTCATGGTCAT	10X	6214	1644	10X_P7_6_AACACGTCATGGTCAT	Spleen	<NA>	B cell	10X_P7_6
10X_P7_15_CAGCAGCTCCTGTAGA	10X	3146	1450	10X_P7_15_CAGCAGCTCCTGTAGA	Limb_Muscle	<NA>	endothelial cell	10X_P7_15

```
> tabmur@meta.data$tissue
```

[1]	"Thymus"	"Tongue"	"Spleen"	"Limb_Muscle"	"Spleen"	"Trachea"	"Marrow"	"Lung"
[9]	"Limb_Muscle"	"Trachea"	"Trachea"	"Trachea"	"Spleen"	"Trachea"	"Mammary_Gland"	"Lung"
[17]	"Kidney"	"Trachea"	"Trachea"	"Spleen"	"Trachea"	"Mammary_Gland"	"Spleen"	"Mammary_Gland"
[25]	"Thymus"	"Tongue"	"Liver"	"Lung"	"Lung"	"Limb_Muscle"	"Marrow"	"Trachea"
[33]	"Limb_Muscle"	"Lung"	"Trachea"	"Liver"	"Lung"	"Bladder"	"Thymus"	"Trachea"
[41]	"Liver"	"Marrow"	"Limb_Muscle"	"Bladder"	"Trachea"	"Limb_Muscle"	"Limb_Muscle"	"Trachea"
[49]	"Trachea"	"Spleen"	"Trachea"	"Tongue"	"Tongue"	"Trachea"	"Trachea"	"Limb_Muscle"
[57]	"Lung"	"Tongue"	"Bladder"	"Trachea"	"Spleen"	"Tongue"	"Tongue"	"Lung"
[65]	"Liver"	"Trachea"	"Tongue"	"Spleen"	"Spleen"	"Trachea"	"Trachea"	"Trachea"

```
> tabmur@meta.data$cell_ontology_class
```

[1]	"immature T cell"	"basal cell of epidermis"
[3]	"B cell"	"endothelial cell"
[5]	"natural killer cell"	"mesenchymal cell"
[7]	"macrophage"	"stromal cell"
[9]	NA	"mesenchymal cell"
[11]	"mesenchymal cell"	"endothelial cell"
[13]	"macrophage"	"mesenchymal cell"
[15]	"endothelial cell"	"stromal cell"
[17]	"kidney capillary endothelial cell"	"mesenchymal cell"
[19]	"mesenchymal cell"	"B cell"
[21]	"blood cell"	"endothelial cell"
[23]	"B cell"	"B cell"
[25]	"immature T cell"	"basal cell of epidermis"
[27]	"hepatocyte"	"stromal cell"
[29]	NA	"B cell"
[31]	"macrophage"	"epithelial cell"

# Monocle3

## Expression matrix

```
> expression_matrix[1:8,1:4]
8 x 4 sparse Matrix of class "dgMatrix"
      AACCTGCAAGACGTG-300.1.1 AACCTGGTGTGAATA-300.1.1 AACCTGTCGGCCGAT-300.1.1 AAAGATGGTTCGTTGA-300.1.1
WBGene00010957      5          10          2          13
WBGene00010958      .           .           .           1
WBGene00010959      3           4           .           5
WBGene00010960      9          10          4           6
WBGene00010961      .           2           1           .
WBGene00000829      .           5           2           2
WBGene00010962      3           8           2           8
WBGene00010963      2           1           1           1
```

data frame containing  
attributes of individual  
cells

data frame containing  
attributes of features  
(e.g. genes)

```
> cds <- new_cell_data_set(expression_data = expression_matrix, cell_metadata = cell_metadata, gene_metadata = gene_annotation)
> cds
class: cell_data_set
dim: 20222 6188
metadata(1): cds_version
assays(1): counts
rownames(20222): WBGene00010957 WBGene00010958 ... WBGene00021594 WBGene00007064
rowData names(3): id gene_short_name num_cells_expressed
colnames(6188): AACCTGCAAGACGTG-300.1.1 AACCTGGTGTGAATA-300.1.1 ... TGCGGGTAGTACTTGC-b02 TTTGTCAAGTACACCT-b02
colData names(19): cell n.umi ... bg.b01.loading bg.b02.loading
reducedDimNames(0):
mainExpName: NULL
altExpNames(0):
```

Cell Data  
Set Object

20222 genes and 6188 cells

# Cell Data Set Object

```
> str(cds)
Formal class 'cell_data_set' [package "monocle3"] with 13 slots
..@ reduce_dim_aux :Formal class 'SimpleList' [package "S4Vectors"] with 4 slots
.. .. ..@ listData : list()
.. .. ..@ elementType : chr "ANY"
.. .. ..@ elementMetadata: NULL
.. .. ..@ metadata : list()
..@ principal_graph_aux:Formal class 'SimpleList' [package "S4Vectors"] with 4 slots
.. .. ..@ listData : list()
.. .. ..@ elementType : chr "ANY"
.. .. ..@ elementMetadata: NULL
.. .. ..@ metadata : list()
..@ principal_graph :Formal class 'SimpleList' [package "S4Vectors"] with 4 slots
.. .. ..@ listData : list()
.. .. ..@ elementType : chr "ANY"
.. .. ..@ elementMetadata: NULL
.. .. ..@ metadata : list()
..@ clusters :Formal class 'SimpleList' [package "S4Vectors"] with 4 slots
.. .. ..@ listData : list()
.. .. ..@ elementType : chr "ANY"
.. .. ..@ elementMetadata: NULL
.. .. ..@ metadata : list()
..@ int_elementMetadata:Formal class 'DFrame' [package "S4Vectors"] with 6 slots
.. .. ..@ rownames : NULL
.. .. ..@ nrow : int 20222
.. .. ..@ listData : List of 1
.. .. .. ..$ rowPairs:Formal class 'DFrame' [package "S4Vectors"] with 6 slots
.. .. .. .. ..@ rownames : NULL
.. .. .. .. ..@ nrow : int 20222
.. .. .. .. ..@ listData : Named list()
.. .. .. .. ..@ elementType : chr "ANY"
.. .. .. .. ..@ elementMetadata: NULL
.. .. .. .. ..@ metadata : list()
.. .. .. ..@ elementType : chr "ANY"
.. .. .. ..@ elementMetadata: NULL
.. .. .. ..@ metadata : list()
..@ int_colData :Formal class 'DFrame' [package "S4Vectors"] with 6 slots
.. .. ..@ rownames : NULL
.. .. ..@ nrow : int 6188
.. .. ..@ listData : List of 3
.. .. .. ..$ reducedDims:Formal class 'DFrame' [package "S4Vectors"] with 6 slots
.. .. .. .. ..@ rownames : NULL
.. .. .. .. ..@ nrow : int 6188
.. .. .. .. ..@ listData : Named list()
.. .. .. .. ..@ elementType : chr "ANY"
.. .. .. .. ..@ elementMetadata: NULL
.. .. .. .. ..@ metadata : list()
.. .. .. ..$ altExps :Formal class 'DFrame' [package "S4Vectors"] with 6 slots
.. .. .. .. ..@ rownames : NULL
.. .. .. .. ..@ nrow : int 6188
.. .. .. .. ..@ listData : Named list()
.. .. .. .. ..@ elementType : chr "ANY"
.. .. .. .. ..@ elementMetadata: NULL
.. .. .. .. ..@ metadata : list()
.. .. .. ..$ colPairs :Formal class 'DFrame' [package "S4Vectors"] with 6 slots
.. .. .. .. ..@ rownames : NULL
.. .. .. .. ..@ nrow : int 6188
.. .. .. .. ..@ listData : Named list()
.. .. .. .. ..@ elementType : chr "ANY"
.. .. .. .. ..@ elementMetadata: NULL
.. .. .. .. ..@ metadata : list()
```

```
..@ colData :Formal class 'DFrame' [package "S4Vectors"] with 6 slots
.. .. ..@ rownames : chr [1:6188] "AAACCTGCAAGACGTG-300.1.1" "AAACCTGGTGTGAATA-300.1.1" "AAACCTGTCGGCCGAT-300.1.1"
.. .. ..@ nrow : int 6188
.. .. ..@ listData : List of 19
.. .. .. ..$ cell : chr [1:6188] "AAACCTGCAAGACGTG-300.1.1" "AAACCTGGTGTGAATA-300.1.1" "AAACCTGTCGGCCGAT-300.1.1"
.. .. .. ..$ n.umi : num [1:6188] 1003 1458 1633 1716 1799 ...
.. .. .. ..$ time.point : Factor w/ 4 levels "300_minutes",...: 1 1 1 1 1 1 1 1 1 1 ...
.. .. .. ..$ batch : Factor w/ 7 levels "waterston_300_minutes",...: 1 1 1 1 1 1 1 1 1 1 ...
.. .. .. ..$ Size_Factor : Named num [1:6188] 0.78 1.13 1.27 1.33 1.4 ...
.. .. .. ..$ attr(*, "names")= chr [1:6188] "AAACCTGCAAGACGTG-300.1.1" "AAACCTGGTGTGAATA-300.1.1" "AAACCTGTCGGCCGAT-300.1.1"
.. .. .. ..$ raw.embryo.time : int [1:6188] 350 190 260 220 340 330 260 170 260 280 ...
.. .. .. ..$ embryo.time : num [1:6188] 350 190 245 225 325 670 260 170 260 300 ...
.. .. .. ..$ raw.embryo.time.bin : Factor w/ 12 levels "< 100","100-130",...: 7 4 5 5 6 12 5 4 5 6 ...
.. .. .. ..$ lineage : chr [1:6188] "ABalpppapav/ABpraaapav" "ABalppppa/ABpraaapa" "ABpxpaaaaa" NA ...
.. .. .. ..$ num_genes_expressed: int [1:6188] 646 857 865 873 1068 1302 849 944 1116 835 ...
.. .. .. ..$ cell.type : chr [1:6188] "AFD" NA NA NA ...
.. .. .. ..$ bg.300.loading : num [1:6188] 0.809 9.221 6.008 7.518 1.819 ...
.. .. .. ..$ bg.400.loading : num [1:6188] 0.232 3.943 2.226 3.039 -0.581 ...
.. .. .. ..$ bg.500.1.loading : num [1:6188] -2 -3.42 -3.63 -3.93 -3.42 ...
.. .. .. ..$ bg.500.2.loading : num [1:6188] -2.43 -3.48 -3.83 -4.29 -3.76 ...
.. .. .. ..$ bg.r17.loading : num [1:6188] -0.544 4.899 1.989 1.911 -1.444 ...
.. .. .. ..$ bg.b01.loading : num [1:6188] -2.285 1.641 -0.137 -0.961 -2.935 ...
.. .. .. ..$ bg.b02.loading : num [1:6188] -2.13 0.153 -0.519 -2.266 -2.614 ...
.. .. ..@ elementType : chr "ANY"
.. .. ..@ elementMetadata: NULL
.. .. ..@ metadata : list()
..@ assays :Formal class 'SimpleAssays' [package "SummarizedExperiment"] with 1 slot
.. .. ..@ data:Formal class 'SimpleList' [package "S4Vectors"] with 4 slots
.. .. .. ..$ counts:Formal class 'dgCMatrix' [package "Matrix"] with 6 slots
.. .. .. .. ..@ i : int [1:4983450] 0 2 3 6 7 8 9 11 45 46 ...
.. .. .. .. ..@ p : int [1:6189] 0 646 1503 2368 3241 4309 5611 6460 7404 8520 ...
.. .. .. .. ..@ Dim : int [1:2] 20222 6188
.. .. .. .. ..@ Dimnames:List of 2
.. .. .. .. .. ..$ : chr [1:20222] "WBGene00010957" "WBGene00010958" "WBGene00010959" "WBGene00010960" ...
.. .. .. .. .. ..$ : chr [1:6188] "AAACCTGCAAGACGTG-300.1.1" "AAACCTGGTGTGAATA-300.1.1" "AAACCTGTCGGCCGAT-300.1.1"
.. .. .. .. ..@ x : num [1:4983450] 5 3 9 3 2 4 8 1 1 1 ...
.. .. .. ..@ factors : list()
.. .. .. ..@ elementType : chr "ANY"
.. .. .. ..@ elementMetadata: NULL
.. .. .. ..@ metadata : list()
..@ NAMES : NULL
..@ elementMetadata :Formal class 'DFrame' [package "S4Vectors"] with 6 slots
.. .. ..@ rownames : NULL
.. .. ..@ nrow : int 20222
.. .. ..@ listData : Named list()
.. .. ..@ elementType : chr "ANY"
.. .. ..@ elementMetadata: NULL
.. .. ..@ metadata : list()
..@ metadata : List of 1
.. ..$ cds_version: chr "1.2.8"
```

# Count Data

```
> cds@assays@data@listData$counts[1:8,1:5]
8 x 5 sparse Matrix of class "dgCMatix"
      AAACCTGCAAGACGTG-300.1.1 AAACCTGGTGTGAATA-300.1.1 AAACCTGTCGGCCGAT-300.1.1 AAAGATGGTTCGTTGA-300.1.1 AACCATGAGAAACCTA-300.1.1
WBGene00010957      5          10          2          13          10
WBGene00010958      .           .           .           1           1
WBGene00010959      3           4           .           5           .
WBGene00010960      9          10          4           6           5
WBGene00010961      .           2           1           .           1
WBGene00000829      .           5           2           2           6
WBGene00010962      3           8           2           8           4
WBGene00010963      2           1           1           1           .
```

```
> exprs(cds)[1:8,1:5]
8 x 5 sparse Matrix of class "dgCMatix"
      AAACCTGCAAGACGTG-300.1.1 AAACCTGGTGTGAATA-300.1.1 AAACCTGTCGGCCGAT-300.1.1 AAAGATGGTTCGTTGA-300.1.1 AACCATGAGAAACCTA-300.1.1
WBGene00010957      5          10          2          13          10
WBGene00010958      .           .           .           1           1
WBGene00010959      3           4           .           5           .
WBGene00010960      9          10          4           6           5
WBGene00010961      .           2           1           .           1
WBGene00000829      .           5           2           2           6
WBGene00010962      3           8           2           8           4
WBGene00010963      2           1           1           1           .
```

# Normalized Data

```
> normalized_counts(cds)[1:8,1:5]
8 x 5 sparse Matrix of class "dgCMatrix"
      AACCTGCAAGACGTG-300.1.1 AACCTGGTGTGAATA-300.1.1 AACCTGTCGGCCGAT-300.1.1 AAAGATGGTTCGTTGA-300.1.1 AACCATGAGAAACCTA-300.1.1
WBGene00010957      0.8700408      0.9923092      0.4109054      1.0312885      0.9112533
WBGene00010958      .      .      .      0.2429813      0.2343096
WBGene00010959      0.6855876      0.6560779      .      0.6765892      .
WBGene00010960      1.0984650      0.9923092      0.6182071      0.7402544      0.6604763
WBGene00010961      .      0.4416785      0.2523384      .      0.2343096
WBGene00000829      .      0.7333767      0.4109054      0.3978606      0.7235441
WBGene00010962      0.6855876      0.9063123      0.4109054      0.8449846      0.5866681
WBGene00010963      0.5521228      0.2747228      0.2523384      0.2429813      .
```

# Meta Data

To access cds colData table

```
> pData(cds)[1:9,1:12]
DataFrame with 9 rows and 12 columns
```

	cell	n.umi	time.point	batch	Size_Factor	raw.embryo.time	embryo.time	embryo.time.bin	raw.embryo.time.bin	lineage	num_genes_expressed	cell.type
	<character>	<numeric>	<factor>	<factor>	<numeric>	<integer>	<numeric>	<factor>	<factor>	<character>	<integer>	<character>
AAACCTGCAAGACGTG-300.1.1	AAACCTGCAAGACGTG-300..	1003	300_minutes	waterston_300_minutes	0.779569	350	350	330-390	330-390	ABalppppav/ABpraaaa..	646	AFD
AAACCTGGTGTGAATA-300.1.1	AAACCTGGTGTGAATA-300..	1458	300_minutes	waterston_300_minutes	1.133212	190	190	170-210	170-210	ABalppppa/ABpraaaapa	857	NA
AAACCTGTCGGCCGAT-300.1.1	AAACCTGTCGGCCGAT-300..	1633	300_minutes	waterston_300_minutes	1.269229	260	245	210-270	210-270	ABpxpaaaaa	865	NA
AAAGATGGTTCGTTGA-300.1.1	AAAGATGGTTCGTTGA-300..	1716	300_minutes	waterston_300_minutes	1.333740	220	225	210-270	210-270	NA	873	NA
AACCATGAGAAACCTA-300.1.1	AACCATGAGAAACCTA-300..	1799	300_minutes	waterston_300_minutes	1.398250	340	325	270-330	330-390	ABalpppppp/ABpraaaa..	1068	ASK_parent
AACCATGAGTTGAGAT-300.1.1	AACCATGAGTTGAGAT-300..	2527	300_minutes	waterston_300_minutes	1.964079	330	670	> 650	330-390	ABalppppppaa/ABpraaa..	1302	ASEL
AACGTTGAGTAAGTAC-300.1.1	AACGTTGAGTAAGTAC-300..	1691	300_minutes	waterston_300_minutes	1.314309	260	260	210-270	210-270	NA	849	NA
AACTCAGTCTTATCTG-300.1.1	AACTCAGTCTTATCTG-300..	1645	300_minutes	waterston_300_minutes	1.278556	170	170	170-210	170-210	ABalppppap/ABpraaaap	944	NA
AACTCCCAGTGACTC-300.1.1	AACTCCCAGTGACTC-300..	2254	300_minutes	waterston_300_minutes	1.751893	260	260	210-270	210-270	NA	1116	NA

To access cds rowData table

```
> fData(cds)[1:8,]
DataFrame with 8 rows and 3 columns
```

	id	gene_short_name	num_cells_expressed
	<character>	<character>	<integer>
WBGene00010957	WBGene00010957	nduo-6	6038
WBGene00010958	WBGene00010958	ndf1-4	1597
WBGene00010959	WBGene00010959	nduo-1	5342
WBGene00010960	WBGene00010960	atp-6	5921
WBGene00010961	WBGene00010961	nduo-2	2686
WBGene00000829	WBGene00000829	ctb-1	5079
WBGene00010962	WBGene00010962	ctc-3	5885
WBGene00010963	WBGene00010963	nduo-4	3661



# Groups

Name	Groups
Bailey Andrew	3
NORAH	3
Amber	3
Arief Gusnanto	2
Euan McDonnell	4
Volodymyr Chapman	3
Adam	4
Morgan Thomas	4
Yousef Alghamdi	1
Sam Llanwarne	4
Sophia Ahmed	2
Sandy Macdonald	2
Lucy Godson	2
Michael Zulcinski	1
Fabio Marcuccio	1
Georgette Tanner	1



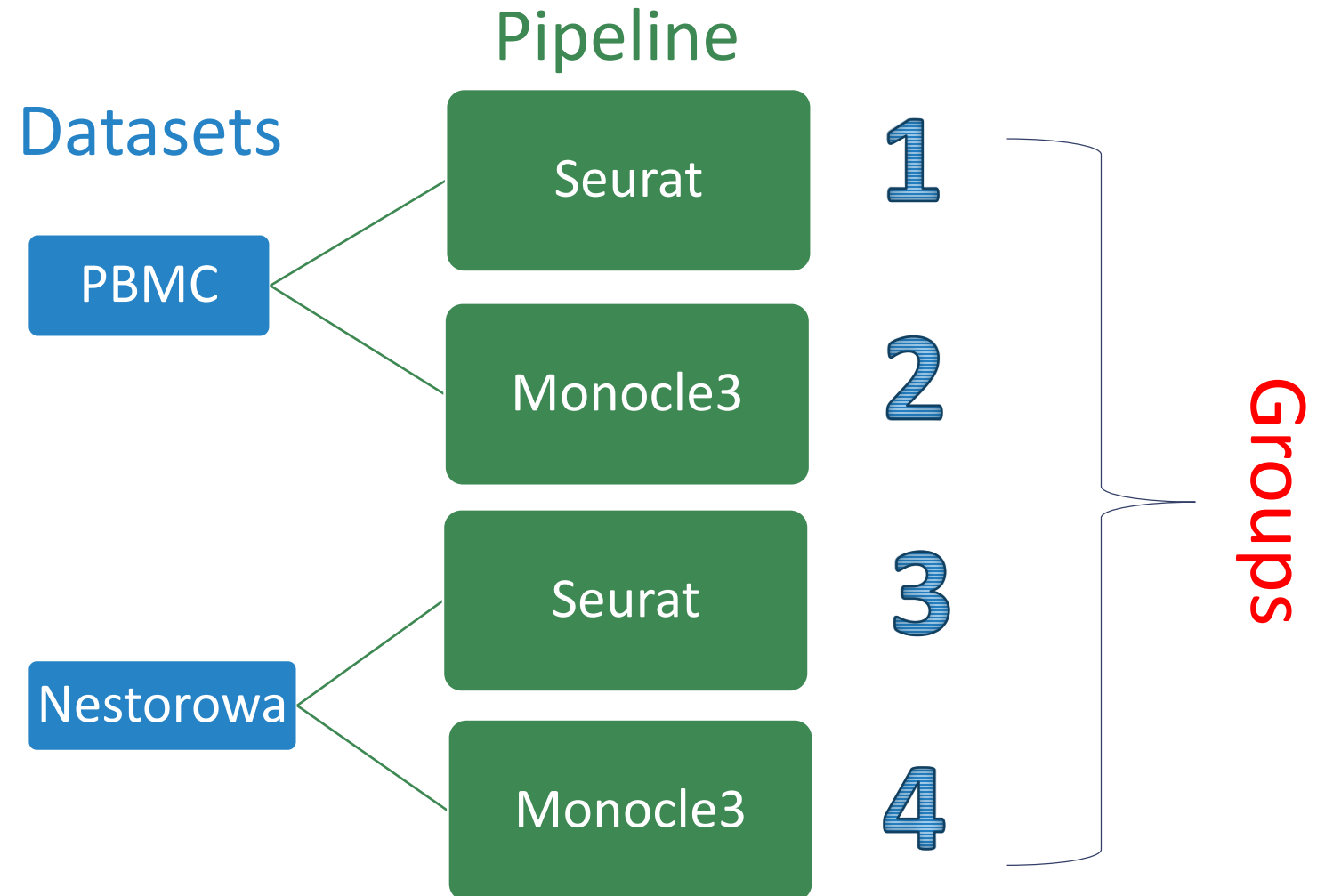
# Hackathon

The dataset is from 10X Genomics on peripheral blood mononuclear cell (Zheng et al. 2017)

- UMI counts
- *pattern* = "*^MT*"

The dataset is generated with SMART-seq2 on Mouse haematopoietic stem cell (HSC) (Nestorowa et al. 2016)

- Read Count
- *pattern* = "*^mt-*"



# Afternoon Session

- 14.00 – Group work on an assigned challenge – Valeria Policastro (Tutors assisted).
- 15.30 – Coffee Break
- 16 – Group work on an assigned challenge – Valeria Policastro (Tutors assisted).
- 17.30 – Wrap up and groups' presentations.
- 18 – Closure.

**This repository contains instructions and material for the scRNAseq Hackathon, School of Mathematics, University of Leeds, 12/05/2022**

---

- <https://conferences.leeds.ac.uk/bad-hackathon/programme/>

## Pipelines

---

During the day we will follow 2 pipelines:

- [Seurat Pipeline \(md format\)](#)
- [Monocle 3 Pipeline \(md format\)](#)

## Data Availability

---

Please download the following folder:

- [Datasets Drive Folder](#)



*Enjoy your  
Hackathon!!!!*