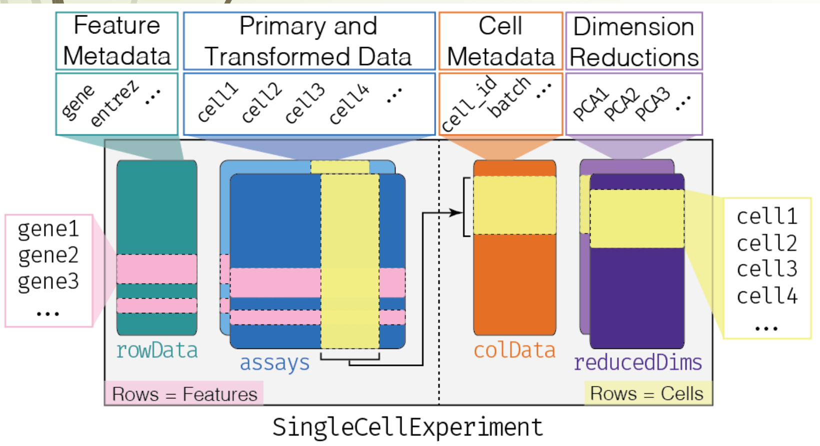
CORTEX seq-FISH: software structure and data integration

Dario Righelli
PostDoc in Davide Risso’s Lab
Department of Statistical Sciences, University of Padova, Italy

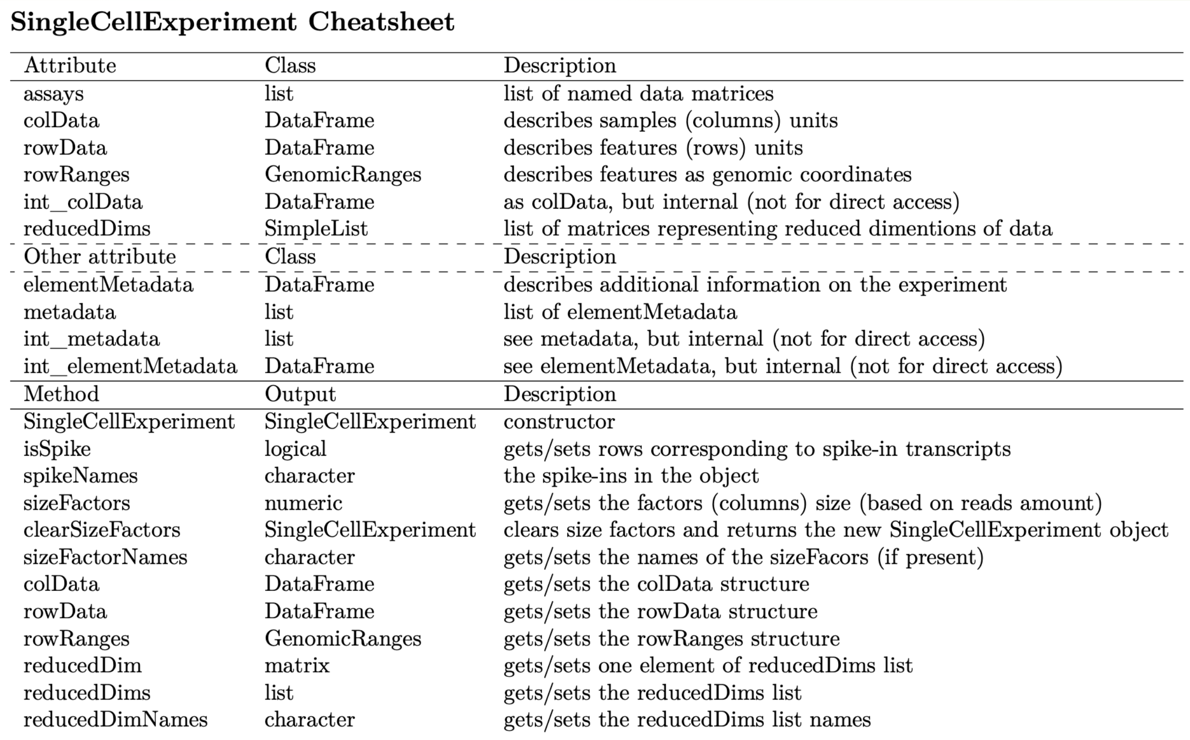
Mathematical Frameworks for Integrative Analysis of
Emerging Biological Data Types – BIRS 2020

SingleCellExperiment class description

- The Bioconductor project
  - Based on the R statistical language
  - Collects tools for genomic data analysis
  - Offers a powerful infrastructure for genomics data
    - Data structures
    - Analysis Software
    - Annotation Databases
    - And More



- **Description**:
  - SingleCellExperiment (BioC >= 3.6 & R >= 3.4)
  - Extends the RangedSummarizedExperiment class
  - Sets the standards for any single cell data storage and analysis in R

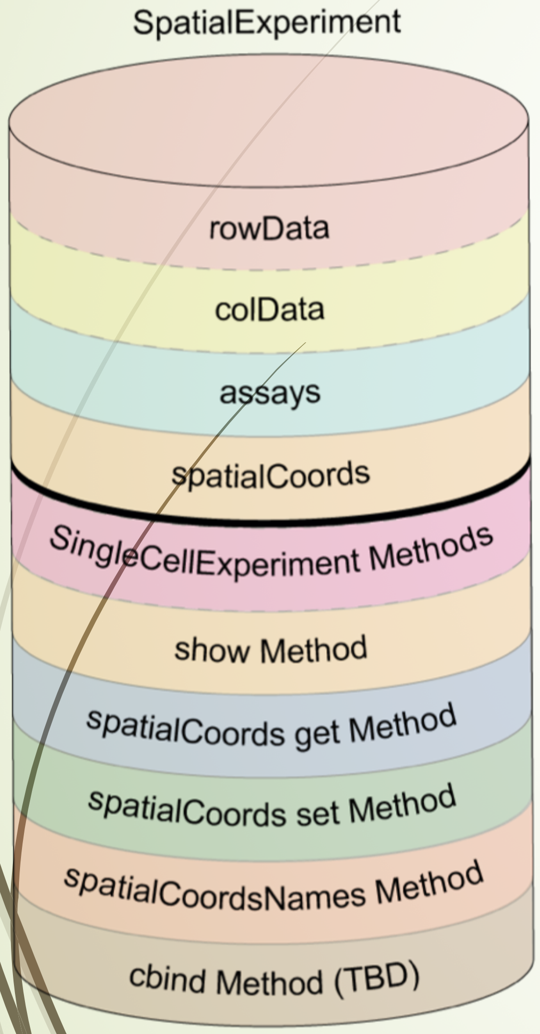


SingleCellExperiment Cheatsheet

Attribute	Class	Description
assays	list	list of named data matrices
colData	DataFrame	describes samples (columns) units
rowData	DataFrame	describes features (rows) units
rowRanges	GenomicRanges	describes features as genomic coordinates
int_colData	DataFrame	as colData, but internal (not for direct access)
reducedDims	SimpleList	list of matrices representing reduced dimentions of data
Other attribute	Class	Description
elementMetadata	DataFrame	describes additional information on the experiment
metadata	list	list of elementMetadata
int_metadata	list	see metadata, but internal (not for direct access)
int_elementMetadata	DataFrame	see elementMetadata, but internal (not for direct access)
Method	Output	Description
SingleCellExperiment	SingleCellExperiment	constructor
isSpike	logical	gets/sets rows corresponding to spike-in transcripts
spikeNames	character	the spike-ins in the object
sizeFactors	numeric	gets/sets the factors (columns) size (based on reads amount)
clearSizeFactors	SingleCellExperiment	clears size factors and returns the new SingleCellExperiment object
sizeFactorNames	character	gets/sets the names of the sizeFacors (if present)
colData	DataFrame	gets/sets the colData structure
rowData	DataFrame	gets/sets the rowData structure
rowRanges	GenomicRanges	gets/sets the rowRanges structure
reducedDim	matrix	gets/sets one element of reducedDims list
reducedDims	list	gets/sets the reducedDims list
reducedDimNames	character	gets/sets the reducedDims list names

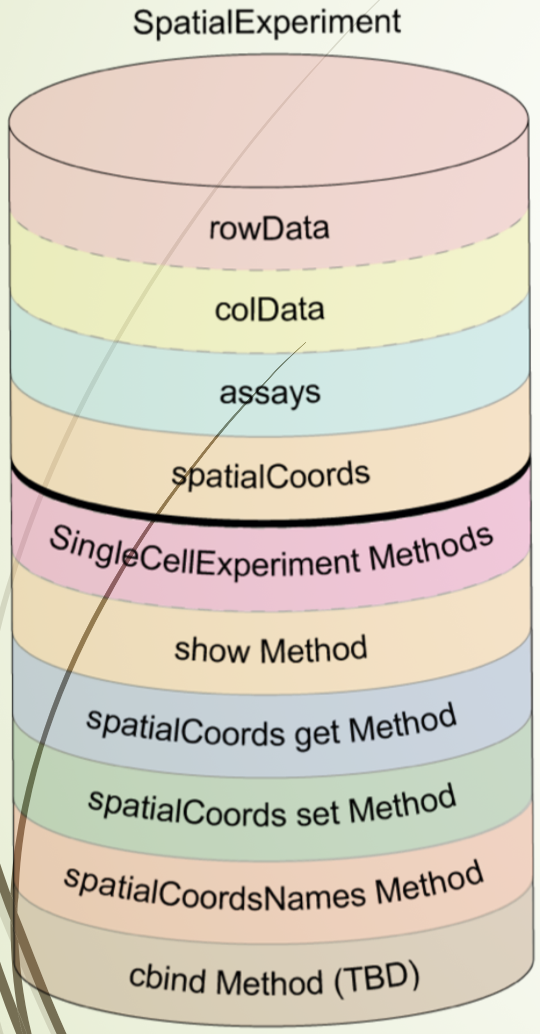
- **Repository**: https://bioconductor.org/packages/SingleCellExperiment/

SpatialExperiment class description



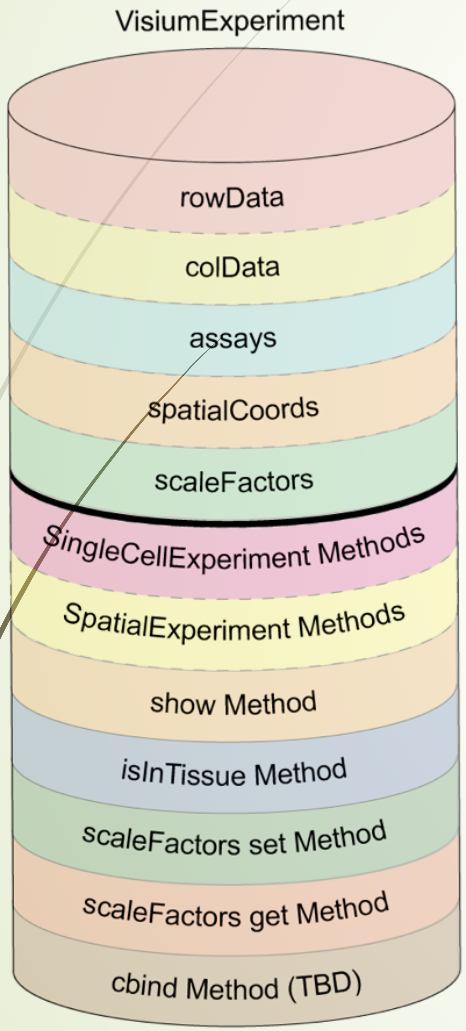
- **Description**:
  - SingleCellExperiment inheritance
  - Puts the basis for spatial experiments analysis
- **Attributes**:
  - Inherited
    - rowData/rowRanges: features specifications (inheritance)
    - colData: samples specifications (inheritance)
    - assays: list of matrices for samples features quantification data (inheritance)
  - spatialCoords: samples spatial coordinates
    - Tipically x and y
    - They are stored inside the int_colData structure
- **Methods**:
  - Inherited
    - SingleCellExperiment methods (inheritance)
    - show: shows a summary of the loaded data
  - spatialCoords: retrieves the x and y coordinates for all the samples
  - spatialCoords <- : assigns new spatial coordinates for all the samples
  - spatialCoordsNames: returns the colnames associated to the spatial coordinates
  - cbind: binds two SpatialExperiment objects (TBD)
- **Repository**: https://github.com/drighelli/SpatialExperiment

SpatialExperiment class usage



- **Description:**
  - We provided a workflow with a dataset of seqFISH and scRNA-seq.
  - Data provided by https://github.com/BIRSBiointegration/Hackathon/tree/master/seqFISH
- Shows how to:
  - Load the data
  - Store scRNA-seq data into a SingleCellExperiment object
  - Install SpatialExperiment package
  - Store seqFISH data into a SpatialExperiment object
  - The show method
  - Get/Set the spatial coordinates
- **Repository:** https://github.com/drighelli/SpatialAnalysis
  - seqFISH_spatialAnalysis HTML workflow

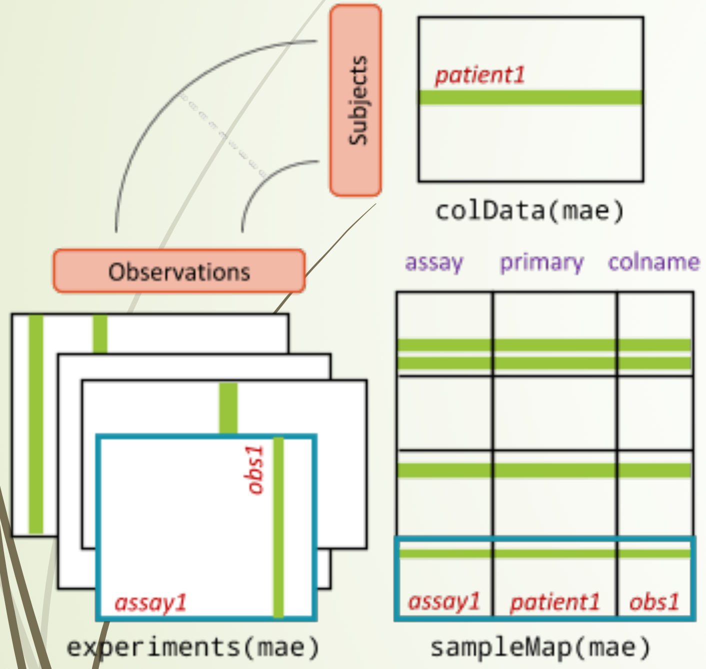
VisiumExperiment class description



- **Description**:
  - SpatialExperiment inheritance
  - Puts the basis for 10x visium spatial gene expression (VSGE) experiments analysis
- **Attributes**:
  - Inherited
    - rowData/rowRanges: features specifications (inheritance)
    - colData: samples specifications (inheritance)
    - assays: list of matrices for samples features quantification data (inheritance)
  - spatialCoords: samples spatial coordinates (re-defined)
    - x, y, inTissue, pxl_col_in_fullres, pxl_row_in_fullres
    - They are stored inside the int_colData structure
  - scaleFactors: factors for provided image different formats conversion
  - imagesPaths: TBD
- **Methods**:
  - Inherited
    - SingleCellExperiment methods (inheritance)
    - SpatialExperiment methods (inheritance)
    - show: shows a summary of the loaded data
  - scaleFactors: retrieves the scale factors list
  - spatialCoords <- : assigns the scale factors list
  - cbind: binds two VisiumExperiment objects (TBD)
- **Repository**: https://github.com/drighelli/SpatialExperiment

SingleCellMultiModal Bioconductor Package

MultiAssayExperiment
Schematic

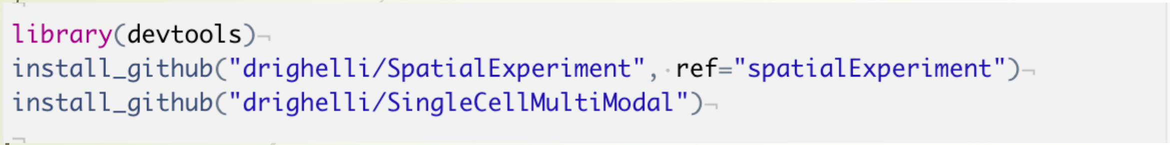


Thanks Marcel!

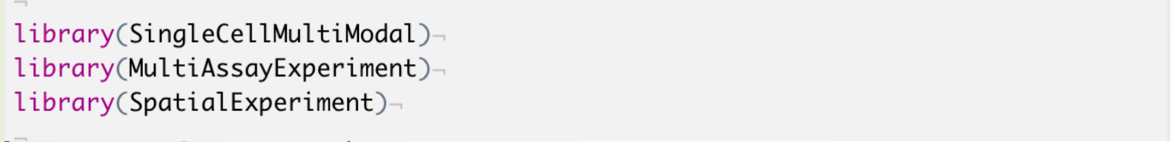
- **Description**:
  - ExperimentHub package
  - Serves multiple datasets of single cell multi-omics
  - Implemented by Marcel Ramos in Levi Waldron's Lab
- **Data available**:
  - scNMT-seq: provided by this workshop (Maint. Marcel)
  - Seq-FISH: provided by this workshop (Maint. Me)
- **Underlying usage idea**:
  - One function call for each type of data available in the package
    - scNMT(dataType="mouse_gastrulation") returns the mouse gastrulation scNMT data.
    - seqFISH(dataType="mouse_visual_cortex") returns the mouse visual cortex seqFISH data.
      - By aid of dataType argument multiple datasets of the same data type can be retrieved
      - Each function returns a MultiAssayExperiment data object within the data.
- **Additional Data**:
  - Any contribution is appreciated:
    - https://github.com/waldronlab/SingleCellMultiModal/wiki/Contributing-Guidelines
  - Multiple functions already present for the easy-to-add other data
- **Repository**: https://bioconductor.org/packages/SingleCellMultiModal/

Seq-FISH Demo

1. Installation & Libraries

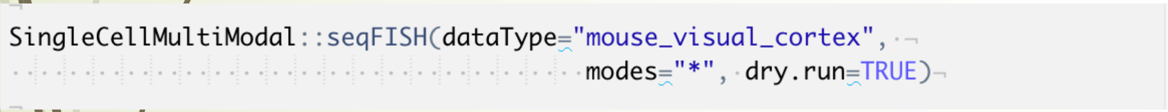


```
library(devtools)
install_github("drighelli/SpatialExperiment", ref="spatialExperiment")
install_github("drighelli/SingleCellMultiModal")
```

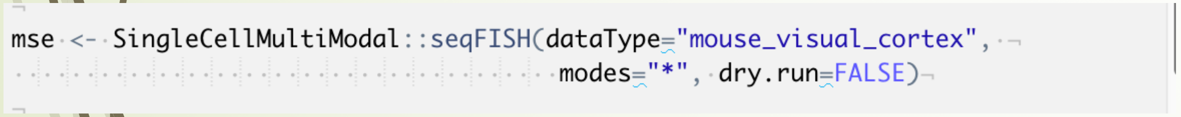


```
library(SingleCellMultiModal)
library(MultiAssayExperiment)
library(SpatialExperiment)
```

2. Data Download

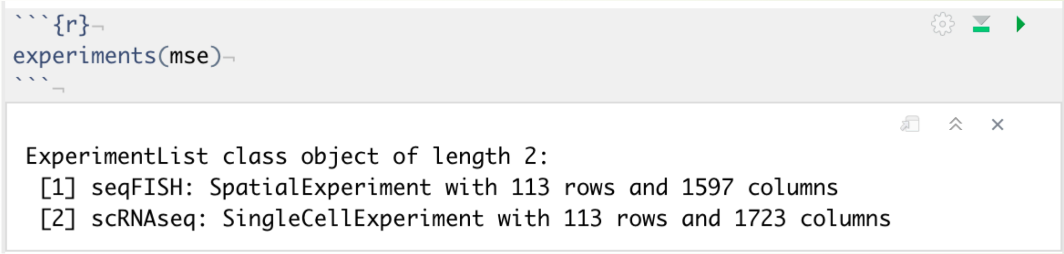


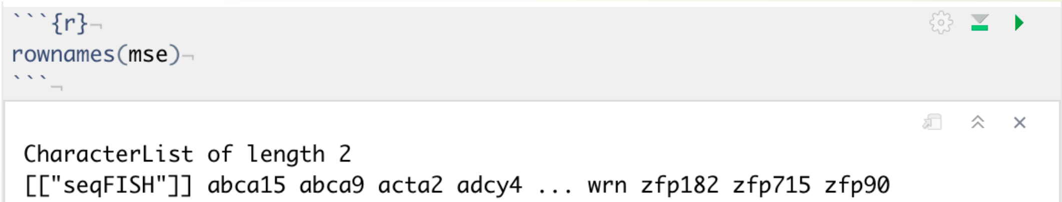
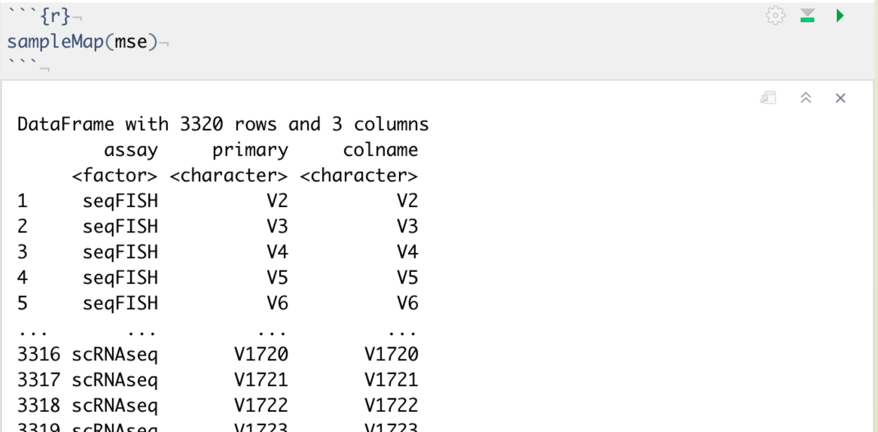
```
SingleCellMultiModal::seqFISH(dataType="mouse_visual_cortex", 
                                        modes="*", dry.run=TRUE)
```



```
mse <- SingleCellMultiModal::seqFISH(dataType="mouse_visual_cortex", 
                                      modes="*", dry.run=FALSE)
```

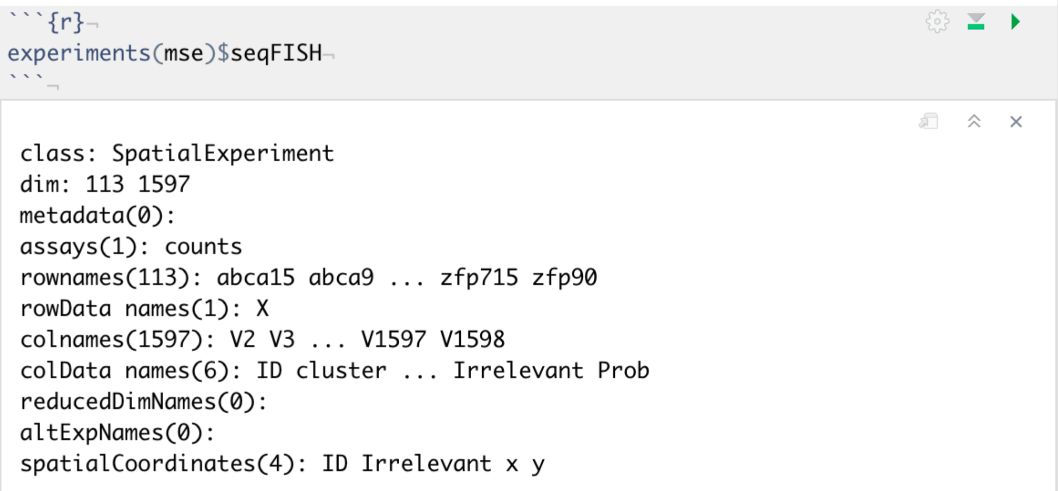
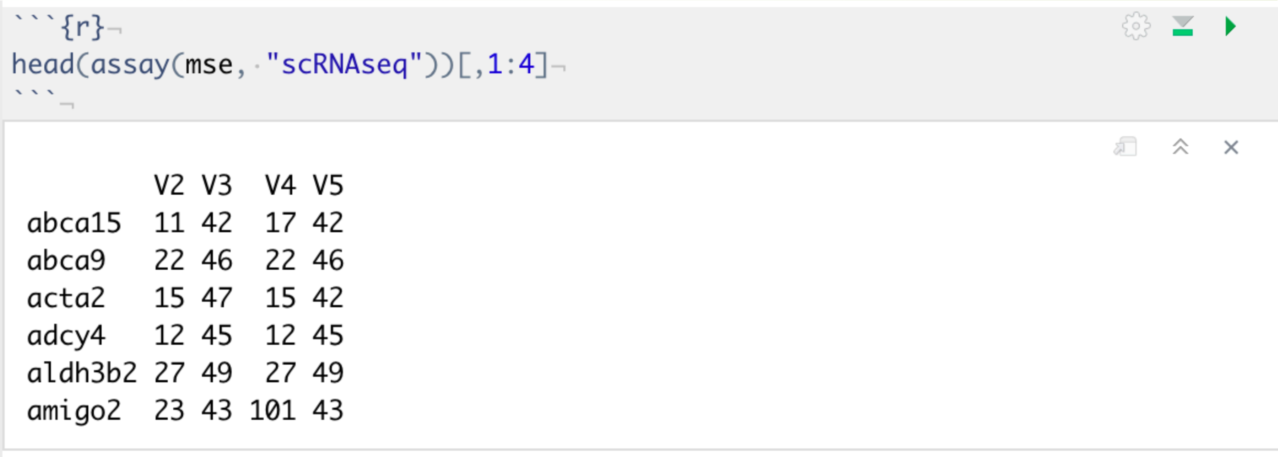
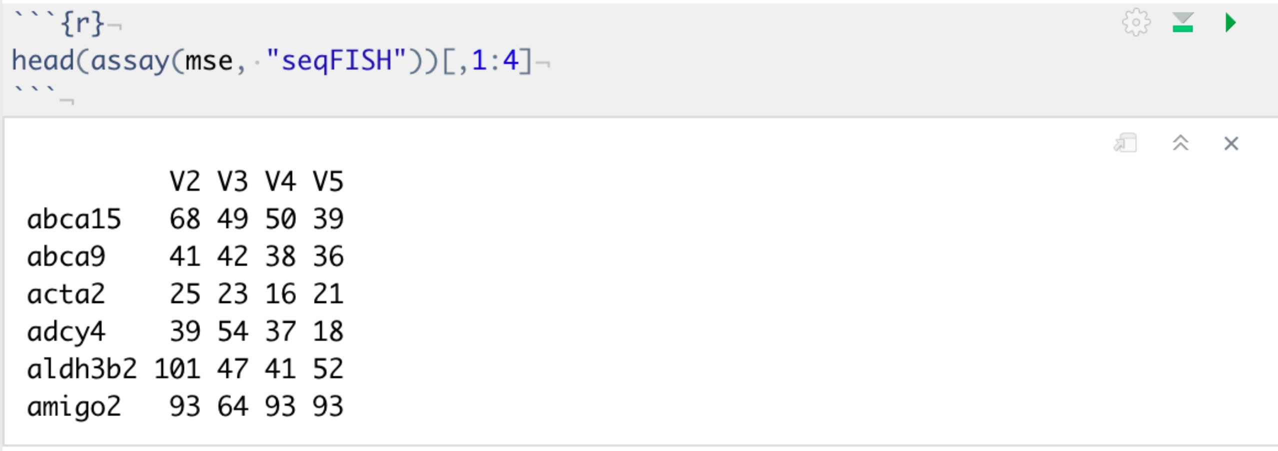
3. Available experiments



~~~
```{r}
experiments(mse)
~~~
```

```
ExperimentList class object of length 2:
 [1] seqFISH: SpatialExperiment with 113 rows and 1597 columns
 [2] scRNAseq: SingleCellExperiment with 113 rows and 1723 columns
```

4. Accessing common data



```
```{r}
rownames(mse)
```
```

```
CharacterList of length 2
[["seqFISH"]] abca15 abca9 acta2 adcy4 ... wrn zfp182 zfp715 zfp90
[["scRNAseq"]] abca15 abca9
```



```
```{r}
sampleMap(mse)
```
```

```
DataFrame with 3320 rows and 3 columns
         assay     primary     colname
      <factor> <character> <character>
1      seqFISH          V2          V2
2      seqFISH          V3          V3
3      seqFISH          V4          V4
4      seqFISH          V5          V5
5      seqFISH          V6          V6
...        ...         ...         ...
3316  scRNAseq       V1720       V1720
3317  scRNAseq       V1721       V1721
3318  scRNAseq       V1722       V1722
3319  scRNAseq       V1723       V1723
```
```

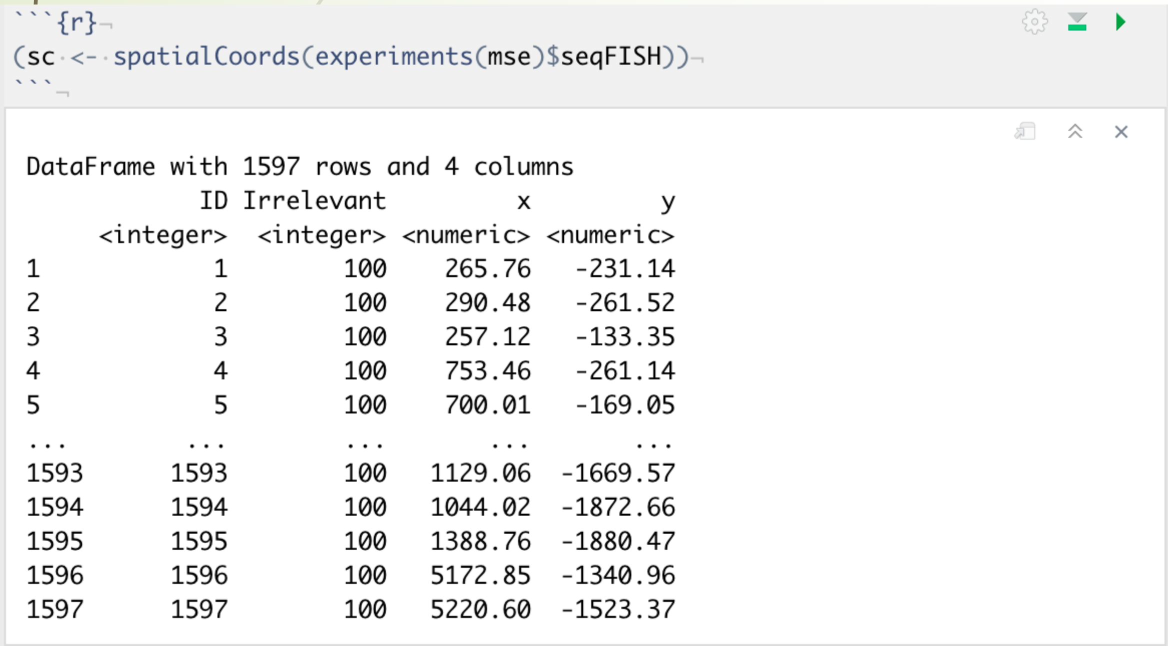
# Seq-FISH Demo

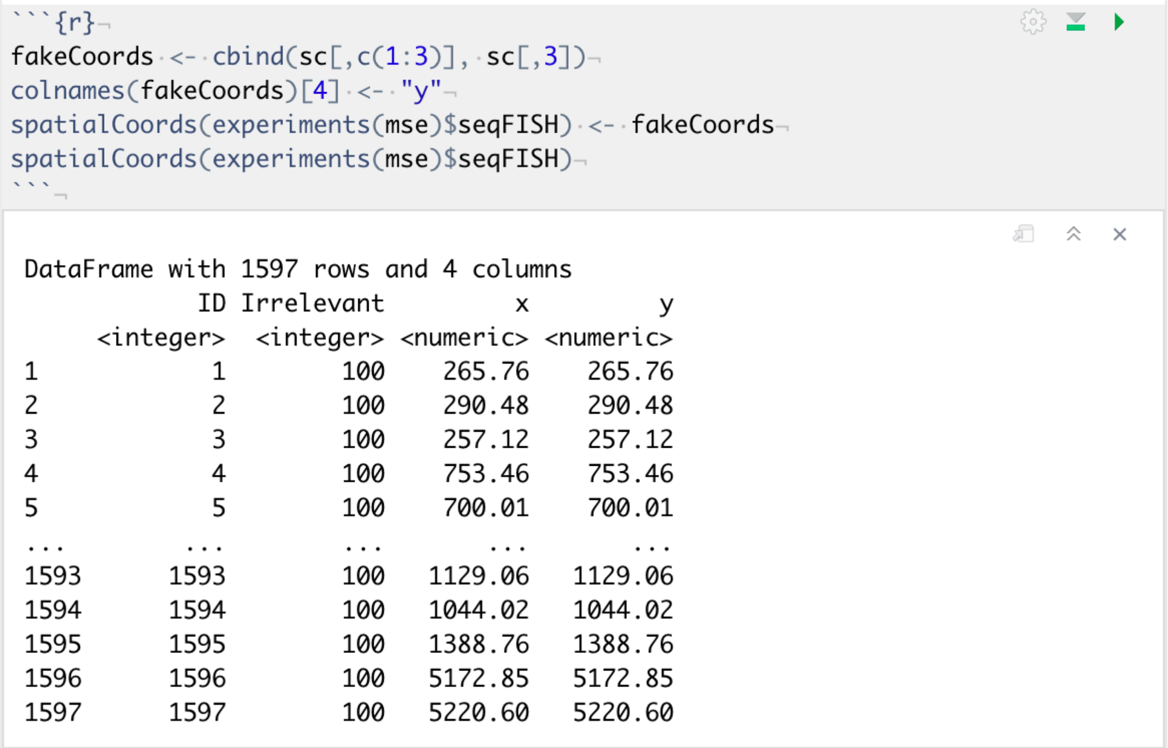
- 5. Retrieving data

~~~
```{r}
experiments(mse)$scRNAseq
~~~
```

```
class: SingleCellExperiment
dim: 113 1723
metadata(0):
assays(1): counts
rownames(113): abca15 abca9 ... zfp715 zfp90
rowData names(1): X
colnames(1723): V2 V3 ... V1723 V1724
colData names(3): V1 V2 V3
reducedDimNames(0):
altExpNames(0):
```



```
```{r}
experiments(mse)$seqFISH
```
```

```
class: SpatialExperiment
dim: 113 1597
metadata(0):
assays(1): counts
rownames(113): abca15 abca9 ... zfp715 zfp90
rowData names(1): X
colnames(1597): V2 V3 ... V1597 V1598
colData names(6): ID cluster ... Irrelevant Prob
reducedDimNames(0):
altExpNames(0):
spatialCoordinates(4): ID Irrelevant x y
```



```
```{r}
head(assay(mse, "scRNAseq"))[,1:4]
```
```

```
        V2 V3  V4 V5
abca15  11 42  17 42
abca9   22 46  22 46
acta2   15 47  15 42
adcy4   12 45  12 45
aldh3b2 27 49  27 49
amigo2  23 43 101 43
```



```
```{r}
head(assay(mse, "seqFISH"))[,1:4]
```
```

```
         V2 V3 V4 V5
abca15   68 49 50 39
abca9   41 42 38 36
acta2   25 23 16 21
adcy4   39 54 37 18
aldh3b2 101 47 41 52
amigo2   93 64 93 93
```
```

Seq-FISH Demo

- 6. Spatial Coordinates



~~~
```{r}
(sc <- spatialCoords(experiments(mse)$seqFISH))
~~~
```

```
DataFrame with 1597 rows and 4 columns
            ID Irrelevant         x         y
     <integer>  <integer> <numeric> <numeric>
1            1        100    265.76   -231.14
2            2        100    290.48   -261.52
3            3        100    257.12   -133.35
4            4        100    753.46   -261.14
5            5        100    700.01   -169.05
...        ...        ...       ...       ...
1593      1593        100   1129.06  -1669.57
1594      1594        100   1044.02  -1872.66
1595      1595        100   1388.76  -1880.47
1596      1596        100   5172.85  -1340.96
1597      1597        100   5220.60  -1523.37
```



```
```{r}
fakeCoords <- cbind(sc[,c(1:3)], sc[,3])
colnames(fakeCoords)[4] <- "y"
spatialCoords(experiments(mse)$seqFISH) <- fakeCoords
spatialCoords(experiments(mse)$seqFISH)
```
```

```
DataFrame with 1597 rows and 4 columns
            ID Irrelevant         x         y
     <integer>  <integer> <numeric> <numeric>
1            1        100    265.76    265.76
2            2        100    290.48    290.48
3            3        100    257.12    257.12
4            4        100    753.46    753.46
5            5        100    700.01    700.01
...        ...        ...       ...       ...
1593      1593        100   1129.06   1129.06
1594      1594        100   1044.02   1044.02
1595      1595        100   1388.76   1388.76
1596      1596        100   5172.85   5172.85
1597      1597        100   5220.60   5220.60
```
```

# Single Cell Multimodal Journal Club

scMultimodal Journal Club

- Started with Levi Waldron’s Lab, leaded by Kelly Eckenrode
- Every Monday on Zoom at 8am EDT/2pm CET
- Communications on: Bioconductor Slack - #education-and-training channel
- Anyone is invited!
- Info available at: https://github.com/waldronlab/data-science-seminar/wiki/Single-cell-multimodal-data

Thanks Kelly!

# Single Cell Multimodal Review

Curated
Single Cell
Multimodal
Benchmarking
Data for
R/Bioconductor

- In collaboration with Levi Waldron's Lab, leaded by Kelly Eckenrode
- Systematic review for multimodal data benchmarking
  - Datasets
  - Methodologies
  - Data Structures
  - Packages
- Info available at: https://github.com/waldronlab/SingleCellMultiModal/wiki/Links-to-resources

Thanks Kelly!

# Acknoledgments

- Davide Risso – Department of Statistics, University of Padua

- Levi Waldron - CUNY Graduate School of Public Health and Health Policy

- Marcel Ramos - CUNY Graduate School of Public Health and Health Policy & Bioconductor core team.

- Kelly Eckenrode – PostDoc, CUNY Graduate School of Public Health and Health Policy

- All of you for the attention!