

CSAMA 2023, Lecture 1a: Introduction to Bioconductor and R with single-cell RNA-seq

Vincent Carey

Road map

- RStudio in brief
 - panes, global settings
 - help, plots, zoom
 - history and debugging
- R in brief: vectors, lists, functions, packages
- Collecting and annotating experimental data: SummarizedExperiment
- HumanPrimaryCellAtlas
- TENxPBMCDData in SingleCellExperiment
- Annotating cells with SingleR, exploring the outcome

Three pane view: Console, environment/history/git, plots

Activities RStudio May 15 05:36

Recent
Starred
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
stvj
DSM
teac
YES
glmm
cityhal
Upcom
Appoint

https://docs.google.com/presentation/d/1doswBT3TOU54LDTFO7Z5mr8TaTzCajOGxlQmXBuDM/edit?slide=id.g2445f54b8fa_0_5

Sign in to your... How to Use Gi... Bioconductor... GA4GH Driver... The Trouble w... python - no m... The Spatial Co... Research strat... SelectorGadget

MIND2022-paper - master - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Console Terminal Background Jobs

```
R 4.3.0 ~ /CSAMA/lecture/1-monday/
1: boxplot(split(as.numeric(hd["S100A4", ]), hd$label.main))
2: split(as.numeric(hd["S100A4", ]), hd$label.main)
3: as.numeric(hd["S100A4", ])
4: as.numeric(hd["S100A4", ])
5: .local(x, ...)
6: as.vector(x, mode = "numeric")
7: .handleSimpleError(function (cond)
8: h(simpleError(msg, call))
9: .handleSimpleError(function (cond)
10: h(simpleError(msg, call))

Selection: 0
> boxplot(split(as.numeric(hd["S100A4", ]), hd$label.main))
Error in h(simpleError(msg, call)) :
  error in evaluating the argument 'x' in selecting a method for function 'boxplot':
  argument "f" is missing, with no default

Enter a frame number, or 0 to exit

1: boxplot(split(as.numeric(hd["S100A4", ]), hd$label.main))
2: split(as.numeric(hd["S100A4", ]), hd$label.main)
3: split(as.numeric(hd["S100A4", ]), hd$label.main)
4: split.default(as.numeric(hd["S100A4", ]), hd$label.main)
5: .handleSimpleError(function (cond)
6: h(simpleError(msg, call))

Selection: 0
> boxplot(split(as.numeric(hd["S100A4", ]), hd$label.main))
> par(mar=c(10,2,2,2), las=2)
> boxplot(split(as.numeric(hd["S100A4", ]), hd$label.main))
> boxplot(split(as.numeric(hd["ORMDL3", ]), hd$label.main))
```

Environment History Connections Git Tutorial

Search results: boxpl

```
boxplot(split(as.numeric(hd["S100A4", ]), hd$label.main))
boxplot(split(as.numeric(hd["HLA-DRA", ]), hd$label.main))
geom_boxplot(alpha = 0.4)
example(geom_boxplot)
geom_boxplot(alpha = 0.4)
geom_boxplot(alpha = 0.4)
geom_boxplot(alpha = 0.4)
geom_boxplot(alpha = 0.4)
geom_boxplot(alpha = 0.4)
```

Files Plots Packages Help Viewer Presentation

Zoom Export Publish

is of mammal
genes associat
to be co-or
combined a large
microarray datasets
and analysed larg
data. Results: Us
Layout Express3D
ations of genes exp
neages. We discuss
a number of these
the coexpression
h the genes that
ons: We consider t
man primary cells
mononuclear phag

Global options under "Tools"

Activities RStudio May 15 05:36

Recent

https://docs.google.com/presentation/d/1doswBT3TOU54LDTFO7Z5mr8TaTzcajOQxiQmXbuDTM/edit?slide=id.g2445f54b8fa_0_5

Sign in to your... How to Use Gl... Bioconductor... GA4GH Driver... The Trouble w... python - no m... The Spatial Co... Research strat... SelectorGadget

MIND2022-paper - master - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Console Terminal Background Jobs

```
R 4.3.0 ~ /CSAMA/lecture1-monday/
1: boxplot(split(as.numeric(hd["S100A4", ]), hd$label.main))
2: split(as.numeric(hd["S100A4", ]), hd$label.main)
3: as.numeric(hd["S100A4", ])
4: as.numeric(hd["S100A4", ])
5: .local(x, ...)
6: as.vector(x, mode = "numeric")
7: .handleSimpleError(function (cond)
  .Internal(C_tryCatchHelper(addr, 1, cond)
8: h(simpleError(msg, call))
9: .handleSimpleError(function (cond)
  .Internal(C_tryCatchHelper(addr, 1, cond)
10: h(simpleError(msg, call))

Selection: 0
> boxplot(split(as.numeric(assay(hd["S100A4", ]), hd$label
+ )
Error in h(simpleError(msg, call)) :
  error in evaluating the argument 'x' in selecting a method
for function 'boxplot': argument "f" is missing, with no default

Enter a frame number, or 0 to exit

1: boxplot(split(as.numeric(assay(hd["S100A4", ]), hd$label
2: split(as.numeric(assay(hd["S100A4", ]), hd$label.main)
3: split(as.numeric(assay(hd["S100A4", ]), hd$label.main)
4: split.default(as.numeric(assay(hd["S100A4", ]), hd$label
5: .handleSimpleError(function (cond)
  .Internal(C_tryCatchHelper(addr, 1, cond)
6: h(simpleError(msg, call))

Selection: 0
> boxplot(split(as.numeric(assay(hd["S100A4", ]), hd$label.main))
> par(mar=c(10,2,2,2), las=2)
> boxplot(split(as.numeric(assay(hd["S100A4", ]), hd$label.main))
> boxplot(split(as.numeric(assay(hd["ORMDL3", ]), hd$label.main))
>
```

Environment History Connections Git Tutorial

Search results: boxpl

boxplot(split(as.numeric(assay(hd["S100A4",]), hd\$label.main))

LA-DRA",]), hd\$label.main)

Options

General Code Console Appearance Pane Layout Packages R Markdown Python Sweave Spelling Git/SVN Publishing Terminal Accessibility

Basic Graphics Advanced

R Sessions

Default working directory (when not in a project):

~ Browse...

☒ Restore most recently opened project at startup

☒ Restore previously open source documents at startup

Workspace

☒ Restore .Rdata into workspace at startup

Save workspace to .Rdata on exit: Ask

History

☒ Always save history (even when not saving .Rdata)

☐ Remove duplicate entries in history

Other

☒ Wrap around when navigating to previous/next tab

☒ Automatically notify me of updates to RStudio

☒ Send automated crash reports to RStudio

OK Cancel Apply

Embryonic_stem
BM
Chondro
Endothelial
Epithelial
Erythroid
Fibroblast
Gastrointestinal
Hepatic
HSC
HSC_CD34+
IPS_cells
Keratinocytes
Macrophage
MEP
Monocyte
MSC
Myelocyte
Myeloid
Neuron
Neutrophil
NK
Osteoblast
Platelet
Pre-B_cell_CD34+
Pro-B_cell_CD34+
Pro-Myelocyte
Smooth_muscle_cells
T_cells
Tissue_stem_cells

isation of mammal
es genes associat
ctions to be co-o
combined a large
craarray datasets
s and analysed larg
se data.Results: Us
Layout Express3D
ations of genes exp
neages. We discuss
a number of these
r the coexpression
h the genes that th
man primary cells
mononuclear phag

Getting help

The screenshot shows the RStudio environment with the following components:

- Console:** Displays R code and an error message. The code attempts to create a boxplot using `boxplot(split(as.numeric(hd["S100A4"]), hd$label.main))`. An error occurs: `Error in h(simpleError(msg, call)) : error in evaluating the argument 'x' in selecting a method for function 'boxplot': argument "f" is missing, with no default`. The user then enters a frame number (0) and runs the code again with `par(mar=c(10,2,2,2), las=2)`. The console also shows the output of `?HumanPrimaryCellAtlasData`.
- Environment:** Shows the current environment with variables like `boxplot`, `split`, `as.numeric`, `assay`, `chk`, `label.main`, `geom_boxplot`, `example`, `alpha`, `cellont`, and `ensembl`.
- History:** Shows the history of executed commands.
- Connections:** Shows the connections to the R environment.
- Git:** Shows the Git status.
- Tutorial:** Shows the tutorial progress.
- Files:** Shows the files in the current project.
- Plots:** Shows the plots in the current project.
- Packages:** Shows the packages loaded in the current session.
- Help:** Shows the help for the `boxplot` function. The help text includes a description of the function, its arguments, and details about the `HumanPrimaryCellAtlasData` object.

```
2: split(as.numeric(hd["S100A4"]), hd$label.main)
3: as.numeric(hd["S100A4"])
4: as.numeric(hd["S100A4"])
5: .local(x, ...)
6: as.vector(x, mode = "numeric")
7: .handleSimpleError(function (cond)
.Internal(C_tryCatchHelper(addr, 1, cond)
8: h(simpleError(msg, call))
9: .handleSimpleError(function (cond)
.Internal(C_tryCatchHelper(addr, 1, cond)
10: h(simpleError(msg, call))

Selection: 0
> boxplot(split(as.numeric(assay(hd["S100A4"]), hd$label.main))
+ )
Error in h(simpleError(msg, call)) :
  error in evaluating the argument 'x' in selecting a method for function 'boxplot':
  argument "f" is missing, with no default

Enter a frame number, or 0 to exit

1: boxplot(split(as.numeric(assay(hd["S100A4"]), hd$label.main)))
2: split(as.numeric(assay(hd["S100A4"]), hd$label.main))
3: split(as.numeric(assay(hd["S100A4"]), hd$label.main))
4: split.default(as.numeric(assay(hd["S100A4"]), hd$label.main))
5: .handleSimpleError(function (cond)
.Internal(C_tryCatchHelper(addr, 1, cond)
6: h(simpleError(msg, call))

Selection: 0
> boxplot(split(as.numeric(assay(hd["S100A4"]), hd$label.main))
> par(mar=c(10,2,2,2), las=2)
> boxplot(split(as.numeric(assay(hd["S100A4"]), hd$label.main))
> boxplot(split(as.numeric(assay(hd["ORMDL3"]), hd$label.main))
> ?HumanPrimaryCellAtlasData
> |
```

boxplot

Search results: boxpl

boxplot(split(as.numeric(assay(chk["S100A4"],2)), chk\$label.main))

boxplot(split(as.numeric(assay(chk["HLA-DRA"],2)), chk\$label.main))

geom_boxplot(alpha = 0.4)

example(geom_boxplot)

geom_boxplot(alpha = 0.4)

geom_boxplot(alpha = 0.4)

geom_boxplot(alpha = 0.4)

geom_boxplot(alpha = 0.4)

geom_boxplot(alpha = 0.4)

Files Plots Packages Help Viewer Presentation

R: Obtain the HPCA data Find in Topic

HumanPrimaryCellAtlasData {cellidx}

R Documentation

Obtain the HPCA data

Description

Download and cache the normalized expression values of the data stored in the Human Primary Cell Atlas. The data will be downloaded from ExperimentHub, returning a [SummarizedExperiment](#) object for further use.

Usage

```
HumanPrimaryCellAtlasData(
  ensembl = FALSE,
  cell.ont = c("all", "nonna", "none")
)
```

Arguments

ensembl Logical scalar indicating whether to convert row names to Ensembl IDs. Genes without a mapping to a non-duplicated Ensembl ID are discarded.

cell.ont String specifying whether Cell Ontology terms should be included in the `colData`. If "nonna", all samples without a valid term are discarded; if "all", all samples are returned with (possibly NA) terms; if "none", terms are not added.

Details

This function provides normalized expression values for 713 microarray samples from the Human Primary Cell Atlas (HPCA) (Mabbott

R in brief

- Console: "Read-Eval-Print" loop – always asking for input
- Make your own data: `x <- c(2.3, 4, 7, 3.2, 5.2, 12)`
 - `c` is the "combine" operator that produces vectors; `"#"` used for commenting
- logic: `x[3] == 7`
- matrix, array: structure for homogeneous (typically numeric) data
 - `y <- matrix(x, ncol=2)` # comment: `y = matrix(x, ncol=2)` also works
 - `y[3,1]` # element
 - `y[,1]` # column; explore `dimnames` so that `y["c", "B"] == y[3,1]`
- `data.frame`: structure for heterogeneous tabular data
 - rows are "records", columns are "attributes"
- Import data: `read.csv`, `read.delim`, `readLines`
- Call functions: `median(x)`; `table(cut(x,3))` # exercise: explain
- Acquire families of functions: `library(MASS)`; `help(package="MASS")`
- `search()`; `installed.packages()`; `install.packages()`

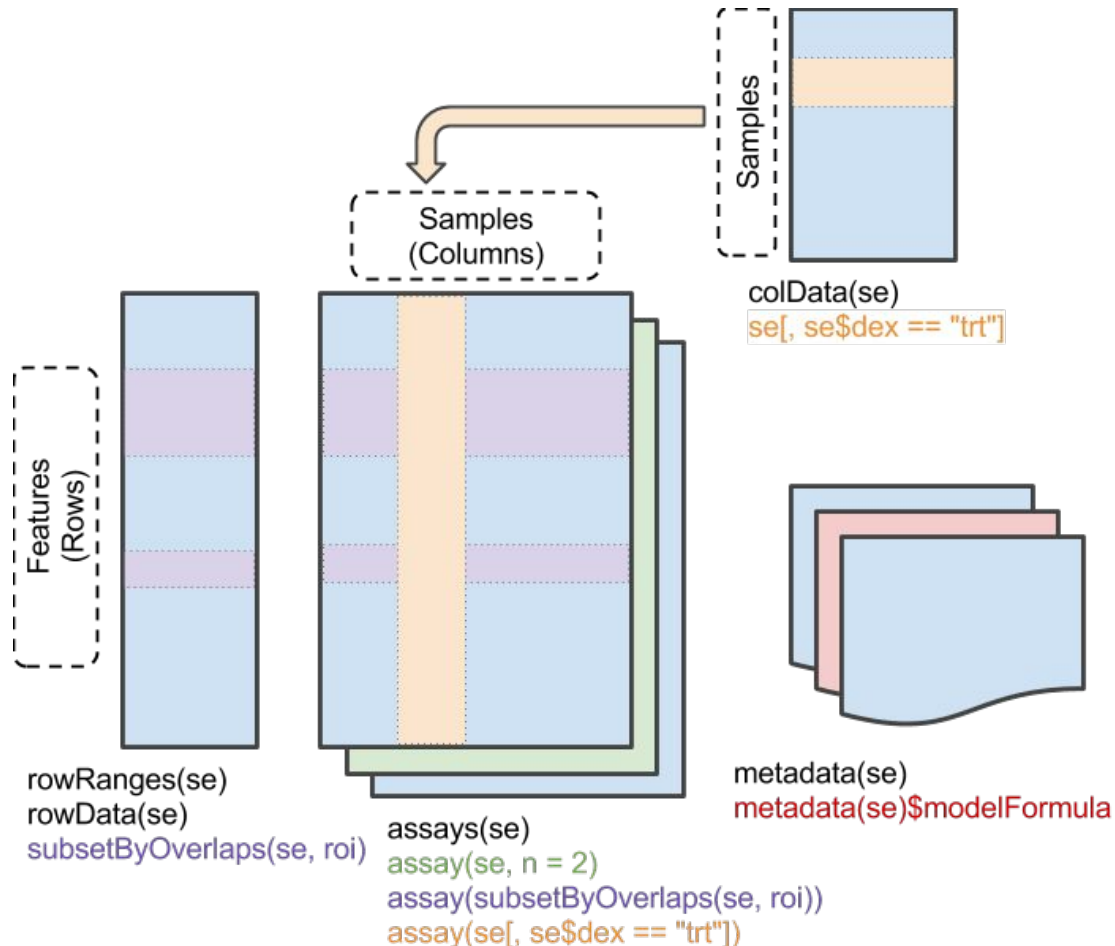
Bioconductor in brief: SummarizedExperiment

Used to organize information on multiple samples.

If `SE` is a `SummarizedExperiment`
`g` is a vector of feature identifiers and `s` is a vector of sample identifiers then

`SE[g, s]` is a new `SummarizedExperiment` restricted to genes in `g` and samples in `s`; quantifications via `assay()`

If `cv` is a variable in `colData(SE)`, then `SE$cv` is a vector with its values



RESEARCH ARTICLE

Open Access

An expression atlas of human primary cells: inference of gene function from coexpression networks

Neil A Mabbott^{††}, J Kenneth Baillie, Helen Brown, Tom C Freeman^{††} and David A Hume^{††}

Abstract

Background: The specialisation of mammalian cells in time and space requires genes associated with specific pathways and functions to be co-ordinately expressed. Here we have combined a large number of publically available microarray datasets derived from human primary cells and analysed large correlation graphs of these data.

Results: Using the network analysis tool BioLayout *Express*^{3D} we identify robust co-associations of genes expressed in a wide variety of cell lineages. We discuss the biological significance of a number of these associations, in particular the coexpression of key transcription factors with the genes that they are likely to control.

Conclusions: We consider the regulation of genes in human primary cells and specifically in the human mononuclear phagocyte system. Of particular note is the fact that these data do not support the identity of putative markers of antigen-presenting dendritic cells, nor classification of M1 and M2 activation states, a current subject of debate within immunological field. We have provided this data resource on the BioGPS web site (<http://biogps.org/dataset/2429/primary-cell-atlas/>) and on macrophages.com (<http://www.macrophages.com/hu-cell-atlas>).

Keywords: Clustering, Meta-analysis, Human, Primary cells, Dendritic cell, Macrophage, Microarray, Transcriptomics

An SE
based on
a paper

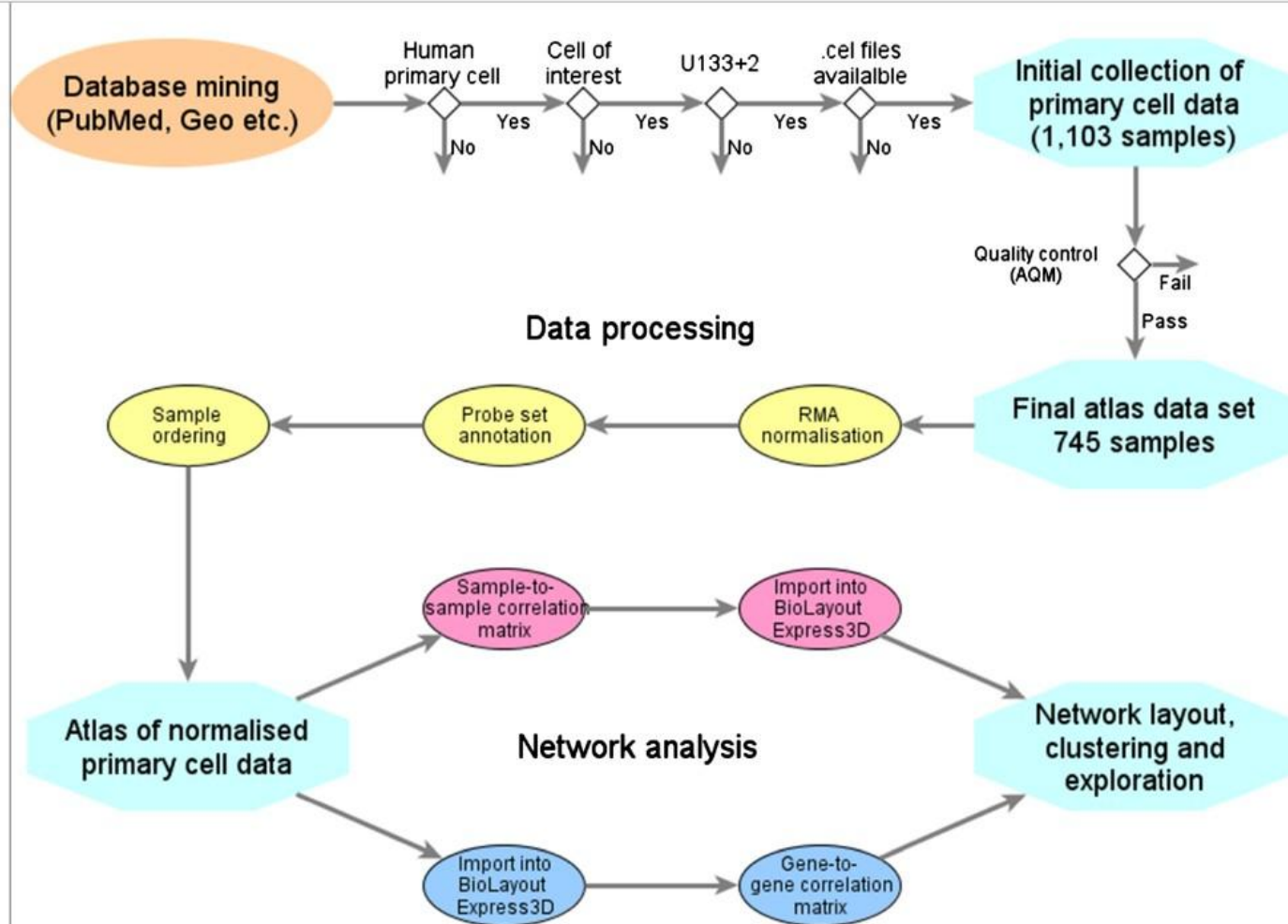


Figure 1 Data analysis workflow. Data analysis pipeline, from the selection of microarray data, through to normalisation, annotation and network analysis.

Three steps to data:

- library
- help or ls [optional]
- call a function and assign the result

```
> library(celldex)
0/0 packages newly attached/loaded, see sessionInfo() for details.
> ?HumanPrimaryCellAtlasData
> hd = HumanPrimaryCellAtlasData()
snapshotDate(): 2023-04-24
see ?celldex and browseVignettes('celldex') for documentation
loading from cache
see ?celldex and browseVignettes('celldex') for documentation
loading from cache
> hd
class: SummarizedExperiment
dim: 19363 713
metadata(0):
assays(1): logcounts
rownames(19363): A1BG A1BG-AS1 ... ZZEF1 ZZZ3
rowData names(0):
colnames(713): GSM112490 GSM112491 ... GSM92233 GSM92234
colData names(3): label.main label.fine label.ont
>
```

Environment History Connections Git Tutorial

To Console To Source

Search results: boxpl

```
boxplot(split(as.numeric(assay(chk["S100A4"],2)), chk$label.main))
boxplot(split(as.numeric(assay(chk["HLA-DRA"],2)), chk$label.main))
geom_boxplot(alpha = 0.4)
example(geom_boxplot)
geom_boxplot(alpha = 0.4)
geom_boxplot(alpha = 0.4)
geom_boxplot(alpha = 0.4)
```

Files Plots Packages Help Viewer Presentation

R: Obtain the HPCA data Find in Topic

HumanPrimaryCellAtlasData {celldex}

R Documentation

Obtain the HPCA data

Description

Download and cache the normalized expression values of the data stored in the Human Primary Cell Atlas. The data will be downloaded from ExperimentHub, returning a [SummarizedExperiment](#) object for further use.

Usage

```
HumanPrimaryCellAtlasData(
  ensembl = FALSE,
  cell.ont = c("all", "nonna", "none")
)
```

Arguments

```
Console Terminal Background Jobs
R 4.3.0 ~ /CSAMA/lecture/1-monday/

> assay(hd[c("S100A4", "LYZ"), 1:4])
      GSM112490 GSM112491 GSM112540 GSM112541
S100A4 13.0099 12.9919 12.9057 11.9467
LYZ    11.6434 11.4303 12.9031 11.5178
> table(hd$label.main)
```

Astrocyte	B_cell	BM
2	26	7
BM & Prog.	Chondrocytes	CMP
1	8	2
DC	Embryonic_stem_cells	Endothelial_cells
88	17	64
Epithelial_cells	Erythroblast	Fibroblasts
16	8	10
Gametocytes	GMP	Hepatocytes
5	2	3
HSC_-G-CSF	HSC_CD34+	iPS_cells
10	6	42
Keratinocytes	Macrophage	MEP
25	90	2
Monocyte	MSC	Myelocyte
60	9	2
Neuroepithelial_cell	Neurons	Neutrophils
1	16	21
NK_cell	Osteoblasts	Platelets
5	15	5
Pre-B_cell_CD34-	Pro-B_cell_CD34+	Pro-Myelocyte
2	2	2

Informal notations ...

Environment History Connections Git Tutorial

Search results: boxplot

boxplot(split(as.numeric(assay(chk["S100A4",],2)), chk\$label.main))
boxplot(split(as.numeric(assay(chk["HLA-DRA",],2)), chk\$label.main)
geom_boxplot(alpha = 0.4)
example(geom_boxplot)
geom_boxplot(alpha = 0.4)
geom_boxplot(alpha = 0.4)
geom_boxplot(alpha = 0.4)

Files Plots Packages Help Viewer Presentation

R: Obtain the HPCA data Find in Topic

HumanPrimaryCellAtlasData {cellIndex} R Documentation

Obtain the HPCA data

Description

Download and cache the normalized expression values of the data stored in the Human Primary Cell Atlas. The data will be downloaded from ExperimentHub, returning a [SummarizedExperiment](#) object for further use.

Usage

HumanPrimaryCellAtlasData(
 ensembl = FALSE,
 cell.ont = c("all", "nonna", "none")
)

Arguments

Logical scalar indicating whether to convert gene names to Ensembl IDs. Genes

... consider label.ont

Paradigms of syntax: classic R and tidyverse

- `data.frame x[,], x$`
- R 4.3 introduces a "pipe" operator, `|>`
- tidyverse/dplyr: `x |> select(...) |> filter(...)`
- a matter of taste? efficiencies (human and mechanical)?
- `table(hd$label.main)` gave us the frequencies of different cell types in the SummarizedExperiment `hd`
- Consider how to get frequencies of *subtypes* of "Macrophage" in the primary cell atlas, found in `hd$label.fine`
 - What functions come to mind to help carry this out?

```
Console Terminal Background Jobs
R 4.3.0 · ~/CSAMA/lecture/1-monday/
> table(hd[,hd$label.main=="Macrophage"]$label.fine)

      Macrophage:Alveolar
      4
Macrophage:Alveolar:B._anthracis_spores
      3
      Macrophage:monocyte-derived
      26
      Macrophage:monocyte-derived:IFNa
      9
      Macrophage:monocyte-derived:IL-4/cntrl
      5
Macrophage:monocyte-derived:IL-4/Dex/cntrl
      5
      Macrophage:monocyte-derived:IL-4/Dex/TGFb
      10
      Macrophage:monocyte-derived:IL-4/TGFb
      5
      Macrophage:monocyte-derived:M-CSF
      2
      Macrophage:monocyte-derived:M-CSF/IFNg
      2
Macrophage:monocyte-derived:M-CSF/IFNg/Pam3Cys
      2
      Macrophage:monocyte-derived:M-CSF/Pam3Cys
      2
      Macrophage:monocyte-derived:S._aureus
      15

>
```

Environment History Connections Git Tutorial

Search results: boxpl

boxplot(split(as.numeric(assay(chk["S100A4"],
chk\$label.main))

boxplot(split(as.numeric(assay(chk["HLA-DRA"...
geom_boxplot(alpha = 0.4)
example(geom_boxplot)
geom_boxplot(alpha = 0.4)
geom_boxplot(alpha = 0.4)

Files Plots Packages Help Viewer Presentation

R: Obtain the HPCA data Find in Topic

HumanPrimaryCellAtlasData {cellDex} R Documentation

Obtain the HPCA data

Description

Download and cache the normalized expression values of the data stored in the Human Primary Cell Atlas. The data will be downloaded from ExperimentHub, returning a [SummarizedExperiment](#) object for further use.

Usage

HumanPrimaryCellAtlasData(
 ensembl = FALSE,
 cell.ont = c("all", "nonna", "none")
)

Arguments


```

Console Terminal Background Jobs
R 4.3.0 ~ /CSAMA/lecture/1-monday/

Macrophage:monocyte-derived:M-CSF
2
Macrophage:monocyte-derived:M-CSF/IFNg
2
Macrophage:monocyte-derived:M-CSF/IFNg/Pam3Cys
2
Macrophage:monocyte-derived:M-CSF/Pam3Cys
2
Macrophage:monocyte-derived:S._aureus
15

> as.data.frame(colData(hd)) |> filter(label.main=="Macrophage") |> group_by(label.fine) |> summarise(n=n())
# A tibble: 13 x 2
  label.fine          n
  <chr>          <int>
1 Macrophage:Alveolar          4
2 Macrophage:Alveolar:B._anthracis_spores      3
3 Macrophage:monocyte-derived          26
4 Macrophage:monocyte-derived:IFNa             9
5 Macrophage:monocyte-derived:IL-4/Dex/TGFb     10
6 Macrophage:monocyte-derived:IL-4/Dex/cntrl      5
7 Macrophage:monocyte-derived:IL-4/TGFb          5
8 Macrophage:monocyte-derived:IL-4/cntrl          5
9 Macrophage:monocyte-derived:M-CSF             2
10 Macrophage:monocyte-derived:M-CSF/IFNg         2
11 Macrophage:monocyte-derived:M-CSF/IFNg/Pam3Cys  2
12 Macrophage:monocyte-derived:M-CSF/Pam3Cys      2
13 Macrophage:monocyte-derived:S._aureus          15
>

```

Conversion from
DataFrame, use dplyr

EnvironmentHistoryConnectionsG

To Console

To Source

Search results: boxpl

boxplot(split(as.numeric(a

chk\$label.main))

boxplot(split(as.numeric(a

geom_boxplot(alpha = 0.4)

example(geom_boxplot)

geom_boxplot(alpha = 0.4)

geom_boxplot(alpha = 0.4)

FilesPlotsPackagesHelpViewer

←

→

🏠

🔍

R: Obtain the HPCA data ▾ Find in Topic

HumanPrimaryCellAtlasData {ce

Obtain the HPCA data

Description

Download and cache the normal the data stored in the Human Pri will be downloaded from Experim [SummarizedExperiment](#) object fr

Usage

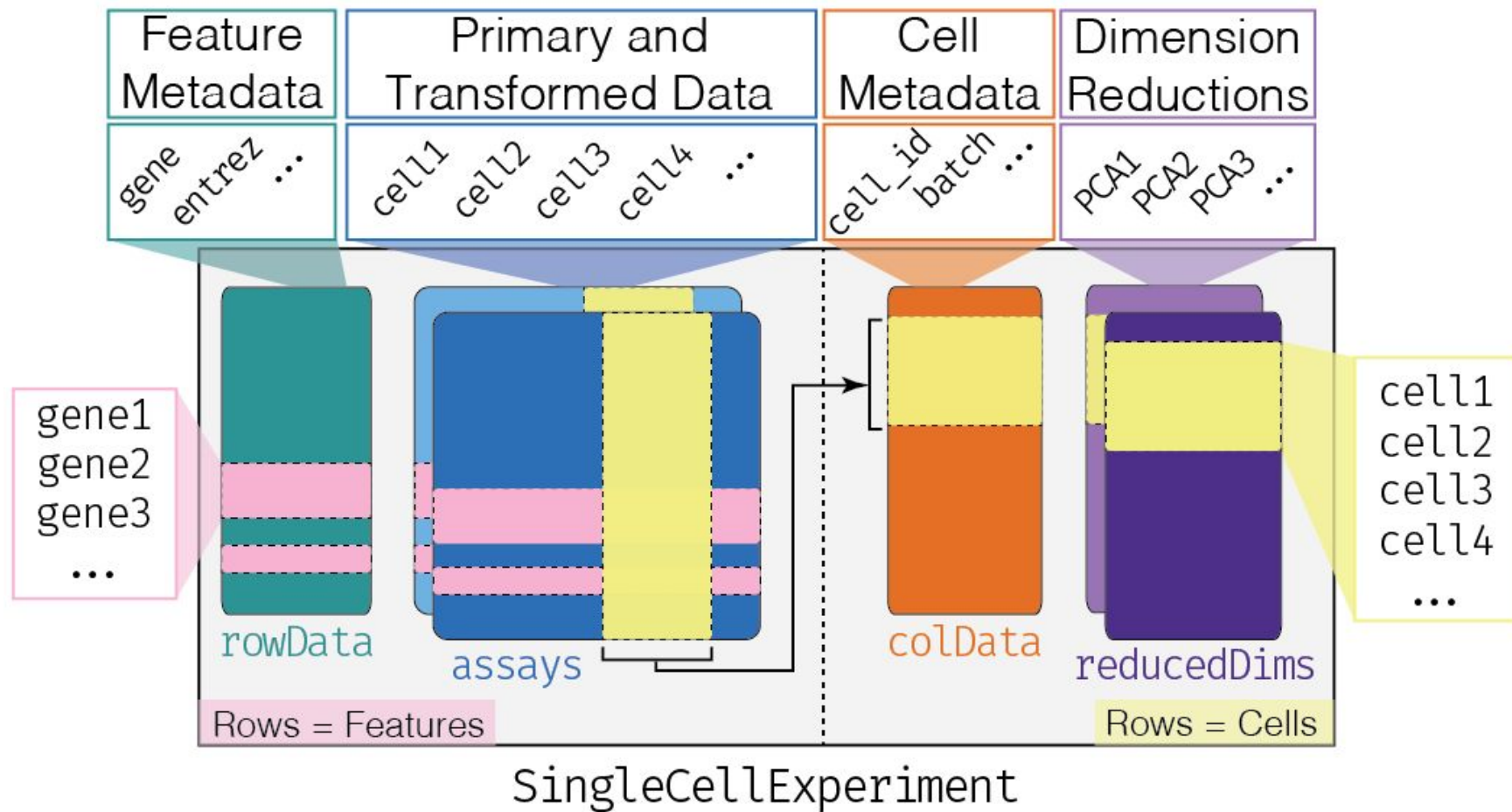
HumanPrimaryCellAtlasDat
ensembl = FALSE,
cell.ont = c("all", "n
)

Arguments

Finale

- We are going to present code that uses HumanPrimaryCellAtlasData to "annotate" a collection of PBMCs assayed by TENx
- The relevant package is SingleR – which "trains" on the atlas data to form criteria for cell labeling on a test set
- We will project the data using principal components and produce interactive visualization of cells suggesting clustering and genes relevant to differentiation

SingleCellExperiment extends SummarizedExperiment

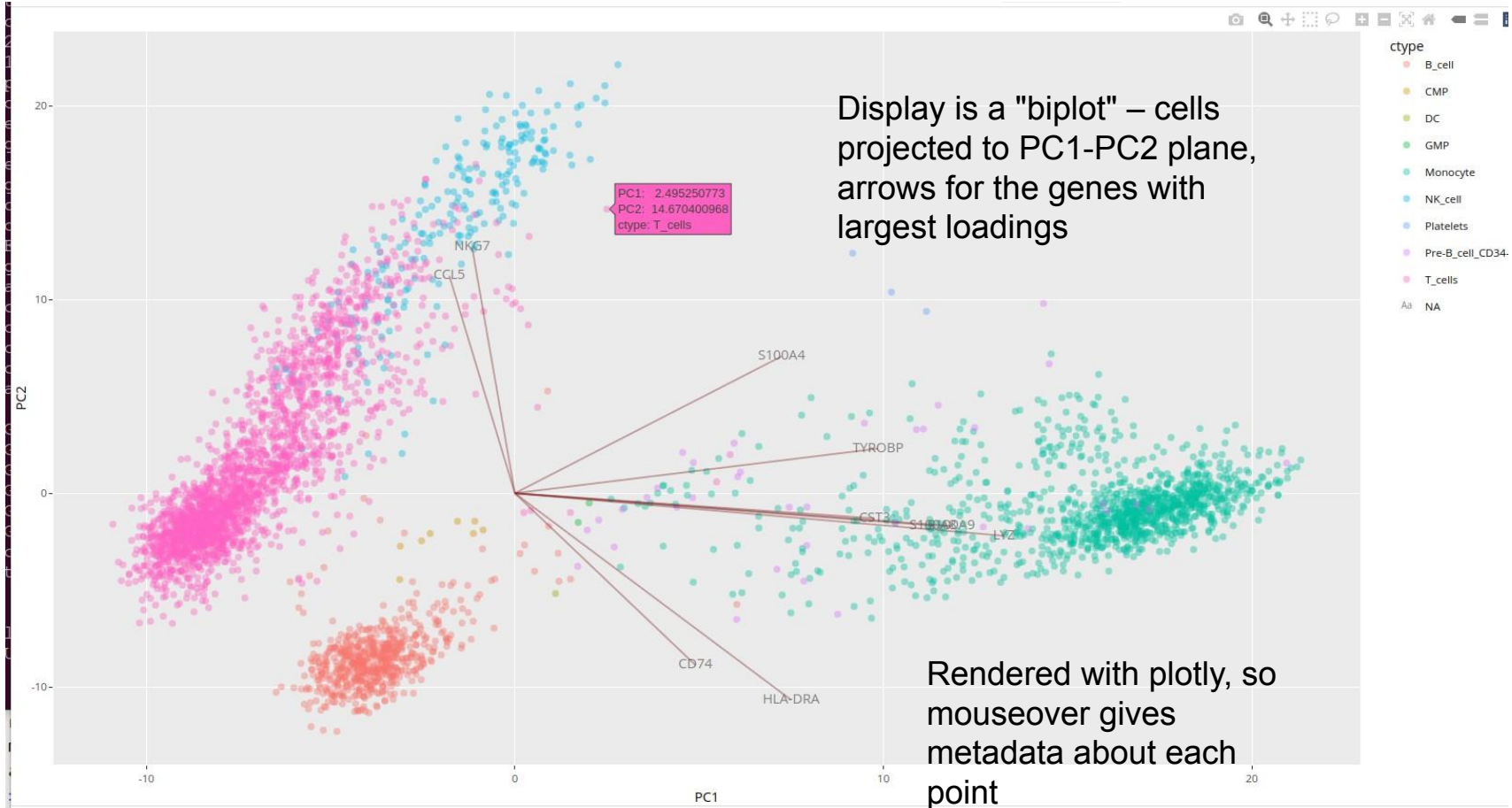


"Desk check"

```
library(TENxPBMCDData) # data package
vp = TENxPBMCDData("pbmc4k") # 4340 cells SCE
vp
hassym = which(!is.na(rowData(vp)$Symbol)) # some genes have name NA
vps = vp[hassym,] # remove them
rownames(vps) = rowData(vps)$Symbol # compatible with 'hd'
library(scater) # analysis tools
vps = logNormCounts(vps)
vps # adds an assay, normalized
library(SingleR)
vsing = SingleR(vps, hd, hd$label.main) # scores 'similarity'
# between atlas and PBMC

vps$label.main = vsing$label.main
vpssds = rowSds(assay(vps))
kp = which(vpssds > quantile(vpssds, .8)) # what does it do?
vpslim = vps[kp,]
mat = t(as.matrix(assay(vpslim,2)))
library(irlba) # approximate PCA
apca = prcomp_irlba(mat, 4)
```

The 4340 PBMCs were labeled using HPCA reference



Summary

- Briefly examined RStudio and basic R syntax
- Reviewed SummarizedExperiment and SingleCellExperiment
- Acquired the HumanPrimaryCellAtlasData SummarizedExperiment from celldex
- Acquired the 4K PBMC dataset from TENxPBMCDData
- Used HPCA to label 4K PBMCs and showed how expression patterns lead to coherent groupings in PCA space
- Used plotly to make an interactive display of the biplot which embellishes the projective rendering with directional information about important genes
- Many questions arise: How good is the labeling? Can cell subtypes be discriminated? Can we use more modern reference data? Can other omics datatypes be used to improve cell type inference?
 - if you followed most of this talk, you can tackle all these questions directly

Gists

to produce the interactive biplot,

- use R 4.3, BiocManager version 3.17
- be sure devtools is installed
- `devtools::source_url("https://gist.githubusercontent.com/vjcitn/b7a44c748c193663525ec227812f0af3/raw/fba74d7d2df667082a11ddc6626aa793f6da1704/pbmcDemo.R", sha1="fb630f3197f60fc89cecfac53644f92e51db9937")$value`
- depending on how much software and data need to be installed in your R, the plot will be produced in a few minutes
- biplot code [here](#)
- overall script [here](#)