Broliden_5325

09 October, 2020

Contents

#Load libraries and other scripts

#Defining some variables for the analysis

Loading data and metadata

```
## $ASV_tissue_B2.csv
## [1] 506  95
##
## $ASV_tissue_B1.csv
## [1] 506   1
##
## $ASV_CVL_V2_B1.csv
## [1] 506  27
##
## $ASV_CVL_V2_B2.csv
## [1] 506  49
##
## $ASV_CVL_V2_C.csv
## [1] 389  62
##
## $ASV_CVL_V3.csv
## [1] 506 111
##
## $ASV_tissue_B2.csv
##  [1] "P089" "P094" "P100" "P051" "P118" "P098" "P085" "P114" "P108" "P121"
## [11] "P112" "P106" "P105" "P111" "P102" "P110" "P032" "P107" "P103" "P063"
## [21] "P120" "P119" "P116" "P117" "P109" "P115" "P113" "P021" "P036" "P045"
## [31] "P044" "P014" "P002" "P019" "P034" "P066" "P046" "P024" "P029" "P031"
## [41] "P050" "P041" "P003" "P008" "P059" "P018" "P052" "P039" "P042" "P057"
## [51] "P043" "P053" "P064" "P068" "P081" "P078" "P071" "P065" "P069" "P040"
## [61] "P061" "P027" "P086" "P087" "P080" "P073" "P083" "P079" "P074" "P082"
## [71] "P070" "P077" "P025" "P075" "P062" "P006" "P038" "P023" "P009" "P007"
## [81] "P047" "P016" "P011" "P026" "P048" "P028" "P060" "P010" "P049" "P037"
```

```
## [91] "P013" "P099" "P093" "P091" "P020"
##
## $ASV_tissue_B1.csv
## [1] "P001"
##
## $ASV_CVL_V2_B1.csv
##  [1] "P059" "P081" "P075" "P044" "P043" "P084" "P051" "P061" "P040" "P065"
## [11] "P078" "P052" "P068" "P094" "P089" "P100" "P098" "P054" "P008" "P045"
## [21] "P057" "P006" "P062" "P027" "P036" "P003" "P046"
##
## $ASV_CVL_V2_B2.csv
##  [1] "P007" "P009" "P010" "P011" "P012" "P013" "P014" "P015" "P018" "P023"
## [11] "P028" "P029" "P032" "P033" "P035" "P037" "P038" "P047" "P049" "P055"
## [21] "P060" "P063" "P064" "P067" "P082" "P085" "P087" "P088" "P091" "P093"
## [31] "P097" "P103" "P104" "P105" "P106" "P107" "P108" "P110" "P111" "P112"
## [41] "P113" "P114" "P115" "P116" "P117" "P118" "P119" "P120" "P121"
##
## $ASV_CVL_V2_C.csv
##  [1] "P053" "P061" "P073" "P054" "P062" "P058" "P036" "P003" "P046" "P059"
## [11] "P081" "P075" "P044" "P043" "P084" "P076" "P086" "P051" "P072" "P040"
## [21] "P065" "P080" "P078" "P052" "P056" "P079" "P068" "P099" "P094" "P089"
## [31] "P100" "P095" "P098" "P096" "P092" "P090" "P031" "P048" "P039" "P008"
## [41] "P045" "P022" "P057" "P004" "P002" "P020" "P025" "P021" "P006" "P026"
## [51] "P027" "P041" "P024" "P042" "P001" "P005" "P050" "P071" "P069" "P066"
## [61] "P074" "P034"
##
## $ASV_CVL_V3.csv
##   [1] "P001" "P002" "P003" "P004" "P005" "P006" "P007" "P008" "P009" "P010"
##  [11] "P011" "P012" "P013" "P014" "P017" "P018" "P019" "P020" "P021" "P022"
##  [21] "P023" "P024" "P025" "P026" "P027" "P028" "P029" "P030" "P031" "P032"
##  [31] "P033" "P034" "P035" "P036" "P037" "P038" "P039" "P040" "P041" "P042"
##  [41] "P043" "P044" "P045" "P047" "P048" "P049" "P050" "P051" "P052" "P053"
##  [51] "P057" "P059" "P060" "P061" "P062" "P063" "P064" "P066" "P067" "P068"
##  [61] "P069" "P070" "P071" "P072" "P073" "P074" "P075" "P076" "P077" "P078"
##  [71] "P079" "P080" "P081" "P082" "P083" "P085" "P086" "P087" "P088" "P089"
##  [81] "P090" "P091" "P092" "P093" "P094" "P095" "P096" "P097" "P098" "P099"
##  [91] "P100" "P101" "P102" "P109" "P103" "P105" "P106" "P107" "P108" "P110"
## [101] "P111" "P112" "P113" "P114" "P115" "P116" "P117" "P118" "P119" "P120"
## [111] "P121"
##
## $ASV_tissue_B2.csv
## [1] 767  95
##
## $ASV_tissue_B1.csv
## [1] 767   1
##
## $ASV_CVL_V2_B1.csv
## [1] 767  27
##
## $ASV_CVL_V2_B2.csv
## [1] 767  49
##
## $ASV_CVL_V2_C.csv
## [1] 767  62
```

```
##
## $ASV_CVL_V3.csv
## [1] 767 111
```

Merging, renaming and correcting batch effects on datasets

```
## $ASV_CVL_V2_B1.csv
## [1] 767  27
##
## $ASV_CVL_V2_B2.csv
## [1] 767  49
##
## $ASV_CVL_V2_C.csv
## [1] 767  62
##
## $ASV_CVL_V3.csv
## [1] 767 111
##
## $ASV_tissue_V3
## [1] 767  96
```

Batch correction with different approaches

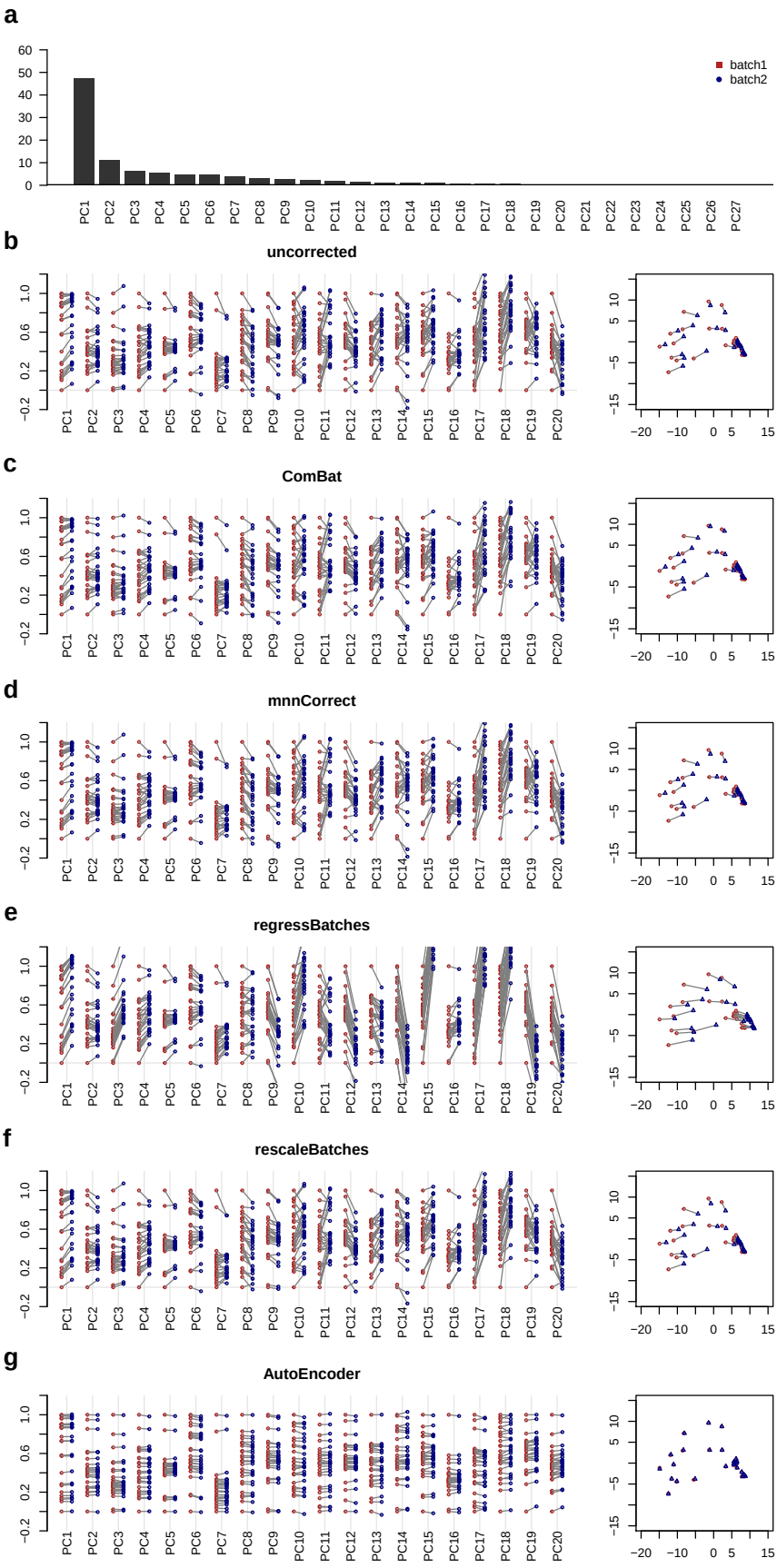
```
## Standardizing Data across genes
```

Batch correction with Autoencoder

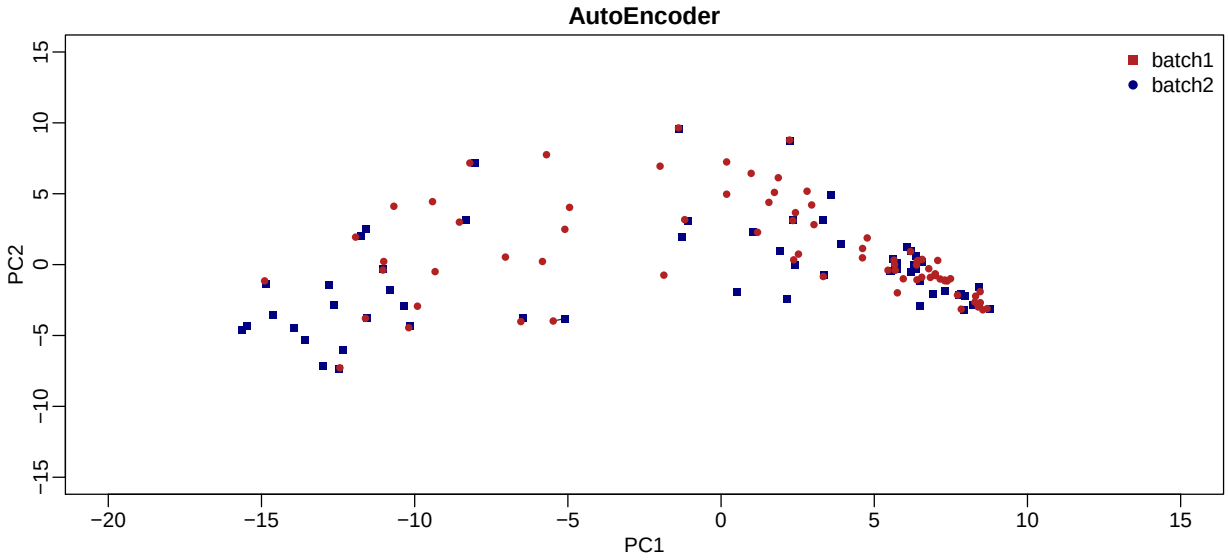
```
## [1]  27 552
## [1] 552
## [1]  27 552
## [1]  27 552
## Model: "functional_1"
## _________________________________________________________________________
## Layer (type)                 Output Shape              Param #
## =========================================================================
## input_1 (InputLayer)         [(None, 552)]             0
## _________________________________________________________________________
## dense (Dense)                (None, 200)               110600
## _________________________________________________________________________
## leaky_re_lu (LeakyReLU)      (None, 200)               0
## _________________________________________________________________________
## dense_1 (Dense)              (None, 10)                2010
## _________________________________________________________________________
## leaky_re_lu_1 (LeakyReLU)    (None, 10)                0
## =========================================================================
## Total params: 112,610
## Trainable params: 112,610
## Non-trainable params: 0
## _________________________________________________________________________
## Model: "functional_3"
## _________________________________________________________________________
## Layer (type)                 Output Shape              Param #
```

```
## ==================================================================================
## input_2 (InputLayer)                [(None, 10)]                  0
## _________________________________________________________________________________
## dense_2 (Dense)                     (None, 200)                   2200
## _________________________________________________________________________________
## leaky_re_lu_2 (LeakyReLU)           (None, 200)                   0
## _________________________________________________________________________________
## dense_3 (Dense)                     (None, 552)                   110952
## _________________________________________________________________________________
## leaky_re_lu_3 (LeakyReLU)           (None, 552)                   0
## ==================================================================================
## Total params: 113,152
## Trainable params: 113,152
## Non-trainable params: 0
## _________________________________________________________________________________
## Model: "functional_5"
## _________________________________________________________________________________
## Layer (type)                        Output Shape                  Param #
## ==================================================================================
## input_3 (InputLayer)                [(None, 552)]                 0
## _________________________________________________________________________________
## functional_1 (Functional)           (None, 10)                    112610
## _________________________________________________________________________________
## functional_3 (Functional)           (None, 552)                   113152
## ==================================================================================
## Total params: 225,762
## Trainable params: 225,762
## Non-trainable params: 0
## _________________________________________________________________________________
```

Plotting the batch-corrected data for comparisson



As we can observe above, when using enough epochs (between 1000 and 2000), the autoencoder fits the data and learns the structure of the data.




```
## $ASV_CVL_V3.csv
## [1] 767 111
##
## $ASV_tissue_V3
## [1] 767  96
##
## $ASV_CVL_V2
## [1] 552 111
##
## $ASV_CVL_V3.csv
##   [1] "P001" "P002" "P003" "P004" "P005" "P006" "P007" "P008" "P009" "P010"
##  [11] "P011" "P012" "P013" "P014" "P017" "P018" "P019" "P020" "P021" "P022"
##  [21] "P023" "P024" "P025" "P026" "P027" "P028" "P029" "P030" "P031" "P032"
##  [31] "P033" "P034" "P035" "P036" "P037" "P038" "P039" "P040" "P041" "P042"
##  [41] "P043" "P044" "P045" "P047" "P048" "P049" "P050" "P051" "P052" "P053"
##  [51] "P057" "P059" "P060" "P061" "P062" "P063" "P064" "P066" "P067" "P068"
##  [61] "P069" "P070" "P071" "P072" "P073" "P074" "P075" "P076" "P077" "P078"
##  [71] "P079" "P080" "P081" "P082" "P083" "P085" "P086" "P087" "P088" "P089"
##  [81] "P090" "P091" "P092" "P093" "P094" "P095" "P096" "P097" "P098" "P099"
##  [91] "P100" "P101" "P102" "P109" "P103" "P105" "P106" "P107" "P108" "P110"
## [101] "P111" "P112" "P113" "P114" "P115" "P116" "P117" "P118" "P119" "P120"
## [111] "P121"
##
## $ASV_tissue_V3
##  [1] "P089" "P094" "P100" "P051" "P118" "P098" "P085" "P114" "P108" "P121"
## [11] "P112" "P106" "P105" "P111" "P102" "P110" "P032" "P107" "P103" "P063"
## [21] "P120" "P119" "P116" "P117" "P109" "P115" "P113" "P021" "P036" "P045"
## [31] "P044" "P014" "P002" "P019" "P034" "P066" "P046" "P024" "P029" "P031"
## [41] "P050" "P041" "P003" "P008" "P059" "P018" "P052" "P039" "P042" "P057"
## [51] "P043" "P053" "P064" "P068" "P081" "P078" "P071" "P065" "P069" "P040"
## [61] "P061" "P027" "P086" "P087" "P080" "P073" "P083" "P079" "P074" "P082"
## [71] "P070" "P077" "P025" "P075" "P062" "P006" "P038" "P023" "P009" "P007"
## [81] "P047" "P016" "P011" "P026" "P048" "P028" "P060" "P010" "P049" "P037"
```

```
## [91] "P013" "P099" "P093" "P091" "P020" "P001"
##
## $ASV_CVL_V2
##   [1] "P059" "P081" "P075" "P044" "P043" "P084" "P051" "P061" "P040" "P065"
##  [11] "P078" "P052" "P068" "P094" "P089" "P100" "P098" "P054" "P008" "P045"
##  [21] "P057" "P006" "P062" "P027" "P036" "P003" "P046" "P007" "P009" "P010"
##  [31] "P011" "P012" "P013" "P014" "P015" "P018" "P023" "P028" "P029" "P032"
##  [41] "P033" "P035" "P037" "P038" "P047" "P049" "P055" "P060" "P063" "P064"
##  [51] "P067" "P082" "P085" "P087" "P088" "P091" "P093" "P097" "P103" "P104"
##  [61] "P105" "P106" "P107" "P108" "P110" "P111" "P112" "P113" "P114" "P115"
##  [71] "P116" "P117" "P118" "P119" "P120" "P121" "P001" "P002" "P004" "P005"
##  [81] "P020" "P021" "P022" "P024" "P025" "P026" "P031" "P034" "P039" "P041"
##  [91] "P042" "P048" "P050" "P053" "P056" "P058" "P066" "P069" "P071" "P072"
## [101] "P073" "P074" "P076" "P079" "P080" "P086" "P090" "P092" "P095" "P096"
## [111] "P099"
##
## $ASV_CVL_V3.csv
## [1] 767 111
##
## $ASV_tissue_V3
## [1] 767  96
##
## $ASV_CVL_V2
## [1] 767 111
```