v19_220601_no_imputation

June 20, 2022

0.1 Load Data

[1]:
```
import pandas as pd
import numpy as np
import os
import glob
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
import matplotlib.pyplot as plt
import missingno as msno
from sklearn.cluster import KMeans
import seaborn as sns
from matplotlib_venn import venn3
from scipy import stats
from sklearn.metrics import silhouette_score
%matplotlib inline
import datetime
from chord import Chord
```

[2]:
```
date = str(datetime.date.today())
print(date)
```

```
2022-06-07
```

[3]:
```
#set working directory
os.chdir('/Users/kavyasharman/Documents/Projects/microLESA/')
```

[4]:
```
#load data as df
pGroups = pd.read_csv('/Users/kavyasharman/Documents/Projects/microLESA/data/
 ↪proteinGroups.txt', delimiter = '\t', low_memory = False)
pGroups.shape
```

[4]:
```
(3760, 371)
```

[5]:
```
pGroups.head()
```

[5]:
```
                                          Protein IDs  \
0                    G3UZW7;A0A023T778;Q9CQL1;P61327
1                           O70589;A0A067XG53;F6Y9I5
```

```
2  A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0...
3        A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1
4                      A0A075B5P3;A0A0A6YVP0;P01867

                              Majority protein IDs Peptide counts (all)  \
0                    G3UZW7;A0A023T778;Q9CQL1;P61327              1;1;1;1
1                          O70589;A0A067XG53;F6Y9I5                2;2;1
2  A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0...    1;1;1;1;1;1;1;1;1
3        A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1              2;1;1;1
4                      A0A075B5P3;A0A0A6YVP0;P01867                3;3;3

  Peptide counts (razor+unique) Peptide counts (unique)  \
0                      1;1;1;1                 1;1;1;1
1                        2;2;1                   2;2;1
2            1;1;1;1;1;1;1;1;1       1;1;1;1;1;1;1;1;1
3                      2;1;1;1                 2;1;1;1
4                        3;3;3                   3;3;3

                                     Protein names  \
0  Protein mago nashi homolog 2;Protein mago nash...
1            Peripheral plasma membrane protein CASK
2                  Ig kappa chain V-II region 26-10
3                                               NaN
4                        Ig gamma-2B chain C region

                           Gene names  \
0                        Magohb;Magoh
1                                Cask
2  Igkv1-110;Igkv1-35;Igkv1-99;Igkv1-115
3           Igkv5-39;Igkv5-43;Igkv5-45
4                        Ighg2b;Igh-3

                                     Fasta headers  Number of proteins  \
0  tr|G3UZW7|G3UZW7_MOUSE Protein mago nashi homo...                   4
1  sp|O70589|CSKP_MOUSE Peripheral plasma membran...                   3
2  tr|A0A140T8M9|A0A140T8M9_MOUSE Immunoglobulin ...                   9
3  tr|A0A075B5M7|A0A075B5M7_MOUSE Immunoglobulin ...                   4
4  tr|A0A075B5P3|A0A075B5P3_MOUSE Immunoglobulin ...                   3

   Peptides  ...  Potential contaminant  id      Peptide IDs  \
0         1  ...                   NaN   0             9939
1         2  ...                   NaN   1        5371;7221
2         1  ...                   NaN   2             7144
3         2  ...                   NaN   3       3664;23181
4         3  ...                   NaN   4  2087;3907;22401

   Peptide is razor  Mod. peptide IDs  \
```

```
0              True             10517
1         True;True          5685;7650
2              True              7567
3         True;True    3880;3881;24763
4    True;True;True    2218;4134;23956

                                        Evidence IDs  \
0                       104111;104112;104113;104114
1                       57553;57554;57555;74694;74695
2                       74067;74068;74069;74070;74071
3                 41615;41616;41617;41618;41619;239742
4  24252;24253;24254;24255;24256;24257;24258;4407...

                                             MS/MS IDs          Best MS/MS  \
0                  113199;113200;113201;113202;113203             113200
1                       62421;62422;62423;81010;81011         62422;81011
2           80341;80342;80343;80344;80345;80346;80347              80344
3                45159;45160;45161;45162;45163;260731        45163;260731
4  26297;26298;26299;26300;26301;26302;26303;2630...  26297;47805;252619

   Oxidation (M) site IDs  Oxidation (M) site positions
0                     NaN                           NaN
1                     NaN                           NaN
2                     NaN                           NaN
3                       0                             4
4                     NaN                           NaN

[5 rows x 371 columns]
```

[6]:
```
pGroups['LFQ intensity 4DPI_2_CO_2']
```

[6]:
```
0       0
1       0
2       0
3       0
4       0
       ..
3755    0
3756    0
3757    0
3758    0
3759    0
Name: LFQ intensity 4DPI_2_CO_2, Length: 3760, dtype: int64
```

[7]:
```
pGroups.tail()
```

[7]:
```
                                    Protein IDs  \
3755                                    REV__Q9QZE7
3756  S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5
```

```
3757                                   S4R192
3758                     V9GWV8;Q3V335;V9GXK3
3759           V9GX06;A0A0R4J0X7;S4R2G5;Q64467

                         Majority protein IDs Peptide counts (all)  \
3755                                 REV__Q9QZE7                   1
3756  S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5    16;16;15;15;15;15
3757                                   S4R192                   3
3758                     V9GWV8;Q3V335;V9GXK3               4;3;2
3759                                   V9GX06             4;1;1;1

     Peptide counts (razor+unique) Peptide counts (unique) Protein names  \
3755                             1                      1           NaN
3756                 1;1;1;1;1;1            1;1;1;1;1;1           NaN
3757                             1                      1           NaN
3758                       1;0;0                  1;0;0           NaN
3759                     1;1;1;1                1;1;1;1           NaN

     Gene names                                      Fasta headers  \
3755       NaN  sp|Q9QZE7|TSNAX_MOUSE Translin-associated prot...
3756      Ank3  tr|S4R2J6|S4R2J6_MOUSE Ankyrin-3 OS=Mus muscul...
3757    Sec31a  tr|S4R192|S4R192_MOUSE Protein transport prote...
3758     Nedd4  tr|V9GWV8|V9GWV8_MOUSE E3 ubiquitin-protein li...
3759   Gm11214  tr|V9GX06|V9GX06_MOUSE Predicted gene 11214 (F...

      Number of proteins  Peptides  ...  Potential contaminant    id  \
3755                   1         1  ...                    NaN  3755
3756                   6        16  ...                    NaN  3756
3757                   1         3  ...                    NaN  3757
3758                   3         4  ...                    NaN  3758
3759                   4         4  ...                    NaN  3759

                                        Peptide IDs  \
3755                                          13511
3756  1455;1705;2665;4244;5703;9994;10028;12460;1246...
3757                                 10933;16514;22778
3758                              3286;20769;22815;23095
3759                              1045;15252;22879;22880

                                     Peptide is razor  \
3755                                             True
3756  False;True;False;False;False;False;False;False...
3757                                 False;False;True
3758                              False;True;False;False
3759                              True;False;False;False

                                     Mod. peptide IDs  \
```

```
3755                                                 14418
3756  1533;1796;2831;4501;6036;10573;10608;13162;131...
3757                                  11562;17664;24344
3758                       3485;22231;24385;24386;24676
3759                       1093;16340;24452;24453;24454

                                          Evidence IDs  \
3755                                     136224;136225
3756  16960;16961;16962;16963;16964;16965;16966;1696...
3757  113810;113811;113812;113813;113814;113815;1138...
3758  37878;37879;37880;37881;37882;37883;37884;3788...
3759  12247;12248;154168;154169;154170;154171;154172...

                                             MS/MS IDs  \
3755                                    147817;147818
3756  18407;18408;18409;18410;18411;18412;18413;1841...
3757  123686;123687;123688;123689;123690;123691;1236...
3758  41143;41144;41145;41146;41147;41148;41149;4115...
3759  13285;13286;13287;167332;167333;167334;167335;...

                                           Best MS/MS  \
3755                                          147817
3756  18413;21709;33621;51114;66447;113651;113977;13...
3757                              123694;178974;256766
3758                        41143;230341;257087;259963
3759                        13287;167340;257908;257916

      Oxidation (M) site IDs  Oxidation (M) site positions
3755               1946;1947                       91;92
3756                 552;553                     182;203
3757                     NaN                         NaN
3758                     936                         111
3759                 172;173                     176;179

[5 rows x 371 columns]
```

[8]:
```
#remove proteins identified as "reverse", "only identified by site", and␣
 ↪"potential contaminant"
pGroups_proc = pGroups[pGroups.Reverse != '+']
pGroups_proc = pGroups_proc[pGroups_proc['Only identified by site'] != '+']
pGroups_proc = pGroups_proc[pGroups_proc['Potential contaminant'] != '+']
pGroups_proc = pGroups_proc.reset_index(drop=True)
pGroups_proc.tail()
```

[8]:
```
                                  Protein IDs  \
3608  Q9Z2X1;J3QMT0;J3QM80;J3QP45;J3QMV8;J3QNH2
3609                                     Q9Z2Z6
3610                      Q9Z315;A0A494B9E9
```

```
3611  S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5
3612                                   S4R192

                        Majority protein IDs Peptide counts (all)  \
3608                    Q9Z2X1;J3QMT0;J3QM80         10;7;7;3;2;1
3609                                 Q9Z2Z6                    3
3610                                 Q9Z315                  7;3
3611  S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5    16;16;15;15;15;15
3612                                 S4R192                    3

     Peptide counts (razor+unique) Peptide counts (unique)  \
3608                  8;7;7;3;2;1             8;7;7;3;2;1
3609                            3                       3
3610                         7;3                     7;3
3611                1;1;1;1;1;1             1;1;1;1;1;1
3612                            1                       1

                                      Protein names Gene names  \
3608  Heterogeneous nuclear ribonucleoprotein F;Hete...     Hnrnpf
3609  Mitochondrial carnitine/acylcarnitine carrier ...   Slc25a20
3610           U4/U6.U5 tri-snRNP-associated protein 1      Sart1
3611                                             NaN       Ank3
3612                                             NaN     Sec31a

                                      Fasta headers  Number of proteins  \
3608  sp|Q9Z2X1|HNRPF_MOUSE Heterogeneous nuclear ri...                  6
3609  sp|Q9Z2Z6|MCAT_MOUSE Mitochondrial carnitine/a...                  1
3610  sp|Q9Z315|SNUT1_MOUSE U4/U6.U5 tri-snRNP-assoc...                  2
3611  tr|S4R2J6|S4R2J6_MOUSE Ankyrin-3 OS=Mus muscul...                  6
3612  tr|S4R192|S4R192_MOUSE Protein transport prote...                  1

      Peptides  ...  Potential contaminant    id  \
3608        10  ...                   NaN  3705
3609         3  ...                   NaN  3706
3610         7  ...                   NaN  3707
3611        16  ...                   NaN  3756
3612         3  ...                   NaN  3757

                                        Peptide IDs  \
3608  2655;9429;10400;13472;13473;15976;19341;19692;...
3609                                418;5649;19235
3610               398;2538;2539;4589;4689;13650;15713
3611  1455;1705;2665;4244;5703;9994;10028;12460;1246...
3612                              10933;16514;22778

                                     Peptide is razor  \
3608  False;True;True;True;True;True;True;True;False...
```

```
3609                                    True;True;True
3610                  True;True;True;True;True;True;True
3611  False;True;False;False;False;False;False;False...
3612                                  False;False;True

                                      Mod. peptide IDs  \
3608  2820;9982;11005;14366;14367;14368;17104;20682;...
3609                                    431;5977;20568
3610           411;2693;2694;4863;4970;4971;14610;16826
3611  1533;1796;2831;4501;6036;10573;10608;13162;131...
3612                                  11562;17664;24344

                                         Evidence IDs  \
3608  30975;30976;30977;30978;30979;30980;30981;3098...
3609  5063;5064;5065;5066;5067;5068;5069;5070;5071;5...
3610  4882;4883;4884;4885;4886;4887;29644;29645;2964...
3611  16960;16961;16962;16963;16964;16965;16966;1696...
3612  113810;113811;113812;113813;113814;113815;1138...

                                           MS/MS IDs  \
3608  33513;33514;33515;33516;33517;33518;33519;3352...
3609  5497;5498;5499;5500;5501;5502;5503;5504;5505;5...
3610  5301;5302;5303;5304;5305;5306;32096;32097;3209...
3611  18407;18408;18409;18410;18411;18412;18413;1841...
3612  123686;123687;123688;123689;123690;123691;1236...

                                         Best MS/MS  \
3608  33513;107092;117607;147535;147570;174314;21329...
3609                                  5506;65881;211665
3610        5304;32096;32124;54448;55730;149357;172113
3611  18413;21709;33621;51114;66447;113651;113977;13...
3612                               123694;178974;256766

      Oxidation (M) site IDs  Oxidation (M) site positions
3608               1926;1927                         2;345
3609                     NaN                           NaN
3610                    1928                           159
3611                 552;553                       182;203
3612                     NaN                           NaN

[5 rows x 371 columns]
```

[9]:
```
pGroups_proc
```

[9]:
```
                                         Protein IDs  \
0                       G3UZW7;A0A023T778;Q9CQL1;P61327
1                             O70589;A0A067XG53;F6Y9I5
2     A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0...
```

```
3          A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1
4                      A0A075B5P3;A0A0A6YVP0;P01867
...                                              ...
3608         Q9Z2X1;J3QMT0;J3QM80;J3QP45;J3QMV8;J3QNH2
3609                                           Q9Z2Z6
3610                                Q9Z315;A0A494B9E9
3611         S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5
3612                                           S4R192

                                Majority protein IDs Peptide counts (all)  \
0                     G3UZW7;A0A023T778;Q9CQL1;P61327              1;1;1;1
1                           O70589;A0A067XG53;F6Y9I5                2;2;1
2     A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0...    1;1;1;1;1;1;1;1;1
3          A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1              2;1;1;1
4                      A0A075B5P3;A0A0A6YVP0;P01867                3;3;3
...                                              ...                  ...
3608                          Q9Z2X1;J3QMT0;J3QM80         10;7;7;3;2;1
3609                                           Q9Z2Z6                    3
3610                                           Q9Z315                  7;3
3611         S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5    16;16;15;15;15;15
3612                                           S4R192                    3

     Peptide counts (razor+unique) Peptide counts (unique)  \
0                         1;1;1;1                1;1;1;1
1                           2;2;1                  2;2;1
2              1;1;1;1;1;1;1;1;1      1;1;1;1;1;1;1;1;1
3                         2;1;1;1                2;1;1;1
4                           3;3;3                  3;3;3
...                           ...                    ...
3608                8;7;7;3;2;1            8;7;7;3;2;1
3609                           3                      3
3610                         7;3                    7;3
3611                1;1;1;1;1;1            1;1;1;1;1;1
3612                           1                      1

                                       Protein names  \
0     Protein mago nashi homolog 2;Protein mago nash...
1               Peripheral plasma membrane protein CASK
2                     Ig kappa chain V-II region 26-10
3                                                  NaN
4                           Ig gamma-2B chain C region
...                                               ...
3608  Heterogeneous nuclear ribonucleoprotein F;Hete...
3609  Mitochondrial carnitine/acylcarnitine carrier ...
3610            U4/U6.U5 tri-snRNP-associated protein 1
3611                                               NaN
3612                                               NaN
```

```
                                Gene names  \
0                              Magohb;Magoh
1                                      Cask
2     Igkv1-110;Igkv1-35;Igkv1-99;Igkv1-115
3                Igkv5-39;Igkv5-43;Igkv5-45
4                              Ighg2b;Igh-3
...                                     ...
3608                                 Hnrnpf
3609                              Slc25a20
3610                                 Sart1
3611                                  Ank3
3612                                Sec31a

                                          Fasta headers  Number of proteins  \
0     tr|G3UZW7|G3UZW7_MOUSE Protein mago nashi homo...                    4
1     sp|O70589|CSKP_MOUSE Peripheral plasma membran...                    3
2     tr|A0A140T8M9|A0A140T8M9_MOUSE Immunoglobulin ...                    9
3     tr|A0A075B5M7|A0A075B5M7_MOUSE Immunoglobulin ...                    4
4     tr|A0A075B5P3|A0A075B5P3_MOUSE Immunoglobulin ...                    3
...                                                 ...                  ...
3608  sp|Q9Z2X1|HNRPF_MOUSE Heterogeneous nuclear ri...                    6
3609  sp|Q9Z2Z6|MCAT_MOUSE Mitochondrial carnitine/a...                    1
3610  sp|Q9Z315|SNUT1_MOUSE U4/U6.U5 tri-snRNP-assoc...                    2
3611  tr|S4R2J6|S4R2J6_MOUSE Ankyrin-3 OS=Mus muscul...                    6
3612  tr|S4R192|S4R192_MOUSE Protein transport prote...                    1

      Peptides  ...  Potential contaminant    id  \
0            1  ...                   NaN     0
1            2  ...                   NaN     1
2            1  ...                   NaN     2
3            2  ...                   NaN     3
4            3  ...                   NaN     4
...        ...  ...                   ...   ...
3608        10  ...                   NaN  3705
3609         3  ...                   NaN  3706
3610         7  ...                   NaN  3707
3611        16  ...                   NaN  3756
3612         3  ...                   NaN  3757

                                      Peptide IDs  \
0                                            9939
1                                       5371;7221
2                                            7144
3                                       3664;23181
4                                   2087;3907;22401
...                                             ...
```

```
3608  2655;9429;10400;13472;13473;15976;19341;19692;...
3609                                     418;5649;19235
3610                 398;2538;2539;4589;4689;13650;15713
3611  1455;1705;2665;4244;5703;9994;10028;12460;1246...
3612                                  10933;16514;22778

                                      Peptide is razor  \
0                                                 True
1                                            True;True
2                                                 True
3                                            True;True
4                                      True;True;True
...                                                 ...
3608  False;True;True;True;True;True;True;True;False...
3609                                   True;True;True
3610                 True;True;True;True;True;True;True
3611  False;True;False;False;False;False;False;False...
3612                                  False;False;True

                                      Mod. peptide IDs  \
0                                                10517
1                                             5685;7650
2                                                 7567
3                                      3880;3881;24763
4                                      2218;4134;23956
...                                                 ...
3608  2820;9982;11005;14366;14367;14368;17104;20682;...
3609                                     431;5977;20568
3610          411;2693;2694;4863;4970;4971;14610;16826
3611  1533;1796;2831;4501;6036;10573;10608;13162;131...
3612                                  11562;17664;24344

                                        Evidence IDs  \
0                          104111;104112;104113;104114
1                        57553;57554;57555;74694;74695
2                        74067;74068;74069;74070;74071
3                 41615;41616;41617;41618;41619;239742
4     24252;24253;24254;24255;24256;24257;24258;4407...
...                                                 ...
3608  30975;30976;30977;30978;30979;30980;30981;3098...
3609  5063;5064;5065;5066;5067;5068;5069;5070;5071;5...
3610  4882;4883;4884;4885;4886;4887;29644;29645;2964...
3611  16960;16961;16962;16963;16964;16965;16966;1696...
3612  113810;113811;113812;113813;113814;113815;1138...

                                                MS/MS IDs  \
0                    113199;113200;113201;113202;113203
```

```
1                         62421;62422;62423;81010;81011
2              80341;80342;80343;80344;80345;80346;80347
3                   45159;45160;45161;45162;45163;260731
4       26297;26298;26299;26300;26301;26302;26303;2630...
...                                                   ...
3608    33513;33514;33515;33516;33517;33518;33519;3352...
3609    5497;5498;5499;5500;5501;5502;5503;5504;5505;5...
3610    5301;5302;5303;5304;5305;5306;32096;32097;3209...
3611    18407;18408;18409;18410;18411;18412;18413;1841...
3612    123686;123687;123688;123689;123690;123691;1236...

                                            Best MS/MS  \
0                                               113200
1                                            62422;81011
2                                                  80344
3                                             45163;260731
4                                        26297;47805;252619
...                                                   ...
3608    33513;107092;117607;147535;147570;174314;21329...
3609                                      5506;65881;211665
3610           5304;32096;32124;54448;55730;149357;172113
3611    18413;21709;33621;51114;66447;113651;113977;13...
3612                                  123694;178974;256766

      Oxidation (M) site IDs  Oxidation (M) site positions
0                        NaN                           NaN
1                        NaN                           NaN
2                        NaN                           NaN
3                          0                             4
4                        NaN                           NaN
...                      ...                           ...
3608               1926;1927                         2;345
3609                     NaN                           NaN
3610                    1928                           159
3611                 552;553                       182;203
3612                     NaN                           NaN

[3613 rows x 371 columns]
```

[10]:
```
#create database of all proteins in protein groups
proteins = pGroups_proc[' Protein IDs'].str.split(";", expand=True)
proteins = proteins.stack()
proteins = list(set(proteins))
proteins = pd.DataFrame(proteins)
proteins.to_csv('outs/' + date + '_allProteinsFromPGroups.csv')
```

[215]:
```
#extract LFQ intensity and create new df with expt classifiers
#instead of group by region or timepoint, add region+timepoint as factors
```

```
def make_lfq(dataframe, species):
    temp = dataframe.columns
    lfq = pd.DataFrame()
    filename = str('/Users/kavyasharman/Documents/Projects/microLESA/outs/' +␣
 ↪date + '_'+ species + '_lfq.csv')
    string = "LFQ"
    for i in temp:
        if string in i:
            lfq = lfq.append(dataframe[i])
    lfq.columns = dataframe[' Protein IDs']
    lfq = lfq.reset_index()
    lfq['index'] = lfq['index'].str.strip("LFQ intensity ")
    lfq[['TimePoint', 'BioRep', 'Region', 'TechRep']] = lfq['index'].str.
 ↪split("_", expand=True)
    lfq['Region'] = lfq['Region'].replace({'LE': 'interface', 'AB': 'SAC', 'CO':
 ↪ 'cortex'})
    lfq['DPI_Region_BioRep'] = lfq['TimePoint'] + "_" + lfq['Region'] + "_" +␣
 ↪lfq['BioRep']
    lfq['BioRep_TechRep'] = lfq['BioRep'] + "_" + lfq['TechRep']
    lfq['DPI_Region'] = lfq['TimePoint'] + "_" + lfq['Region']
    lfq['index'] = lfq['TimePoint'] + "_" + lfq['Region'] + "_" + lfq['BioRep']␣
 ↪+ "_" + lfq['TechRep']
    lfq = lfq.set_index('index')
    lfq = lfq.sort_values(by =['TimePoint', 'Region', 'BioRep', 'TechRep'],␣
 ↪ascending=False)
    return lfq

#extract LFQ intensity for all proteins and create new df 'lfq' with expt␣
 ↪classifiers
lfq = make_lfq(pGroups_proc, 'combined')
lfq.shape
```

[215]:
```
(42, 3620)
```

[222]:
```
lfq.head()
```

[222]:
```
 Protein IDs        G3UZW7;A0A023T778;Q9CQL1;P61327  O70589;A0A067XG53;F6Y9I5  \
index
4DPI_interface_2_4                                0.0                       0.0
4DPI_interface_2_2                                0.0                       0.0
4DPI_interface_2_1                                0.0                       0.0
4DPI_interface_1_2                                0.0                       0.0
4DPI_interface_1_1                                0.0                       0.0

 Protein IDs        A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0A075B5K8;F6XWB
2;A0A140T8N1;A0A0G2JDE5;P01631  \
index
```

```
4DPI_interface_2_4                                                  0.0
4DPI_interface_2_2                                                  0.0
4DPI_interface_2_1                                                  0.0
4DPI_interface_1_2                                                  0.0
4DPI_interface_1_1                                                  0.0

 Protein IDs        A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1  \
index
4DPI_interface_2_4                                          0.0
4DPI_interface_2_2                                          0.0
4DPI_interface_2_1                                          0.0
4DPI_interface_1_2                                          0.0
4DPI_interface_1_1                                          0.0

 Protein IDs        A0A075B5P3;A0A0A6YVP0;P01867  \
index
4DPI_interface_2_4                         0.0
4DPI_interface_2_2                         0.0
4DPI_interface_2_1                         0.0
4DPI_interface_1_2                         0.0
4DPI_interface_1_1                         0.0

 Protein IDs        A0A075B5P5;A0A1Y7VJN6;P03987  \
index
4DPI_interface_2_4                         0.0
4DPI_interface_2_2                         0.0
4DPI_interface_2_1                         0.0
4DPI_interface_1_2                         0.0
4DPI_interface_1_1                         0.0

 Protein IDs        A0A075B6A0;A0A075B5P6;P01872  \
index
4DPI_interface_2_4                  11850000.0
4DPI_interface_2_2                  13198000.0
4DPI_interface_2_1                  14041000.0
4DPI_interface_1_2                   8327300.0
4DPI_interface_1_1                   9088300.0

 Protein IDs        A0A075B5R2;A0A075B5R3;A0A0A6YX91;P01786  \
index
4DPI_interface_2_4                                489990.0
4DPI_interface_2_2                                    0.0
4DPI_interface_2_1                                1280300.0
4DPI_interface_1_2                                    0.0
4DPI_interface_1_1                                    0.0

 Protein IDs        A0A075B5S5;A0A0A6YXL5  A0A075B5T3;J3QK03;A0A0A6YWS9  ...  \
```

```
index                                                                  ...
4DPI_interface_2_4                    0.0                          0.0  ...
4DPI_interface_2_2                    0.0                          0.0  ...
4DPI_interface_2_1                    0.0                          0.0  ...
4DPI_interface_1_2                    0.0                          0.0  ...
4DPI_interface_1_1                    0.0                          0.0  ...

 Protein IDs        Q9Z315;A0A494B9E9  \
index
4DPI_interface_2_4           954220.0
4DPI_interface_2_2          1377400.0
4DPI_interface_2_1           922440.0
4DPI_interface_1_2                0.0
4DPI_interface_1_1          1117600.0

 Protein IDs        S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5  S4R192  \
index
4DPI_interface_2_4                                        0.0     0.0
4DPI_interface_2_2                                        0.0     0.0
4DPI_interface_2_1                                        0.0     0.0
4DPI_interface_1_2                                        0.0     0.0
4DPI_interface_1_1                                        0.0     0.0

 Protein IDs        TimePoint  BioRep     Region  TechRep  DPI_Region_BioRep  \
index
4DPI_interface_2_4       4DPI       2  interface        4   4DPI_interface_2
4DPI_interface_2_2       4DPI       2  interface        2   4DPI_interface_2
4DPI_interface_2_1       4DPI       2  interface        1   4DPI_interface_2
4DPI_interface_1_2       4DPI       1  interface        2   4DPI_interface_1
4DPI_interface_1_1       4DPI       1  interface        1   4DPI_interface_1

 Protein IDs        BioRep_TechRep      DPI_Region
index
4DPI_interface_2_4             2_4  4DPI_interface
4DPI_interface_2_2             2_2  4DPI_interface
4DPI_interface_2_1             2_1  4DPI_interface
4DPI_interface_1_2             1_2  4DPI_interface
4DPI_interface_1_1             1_1  4DPI_interface

[5 rows x 3620 columns]
```

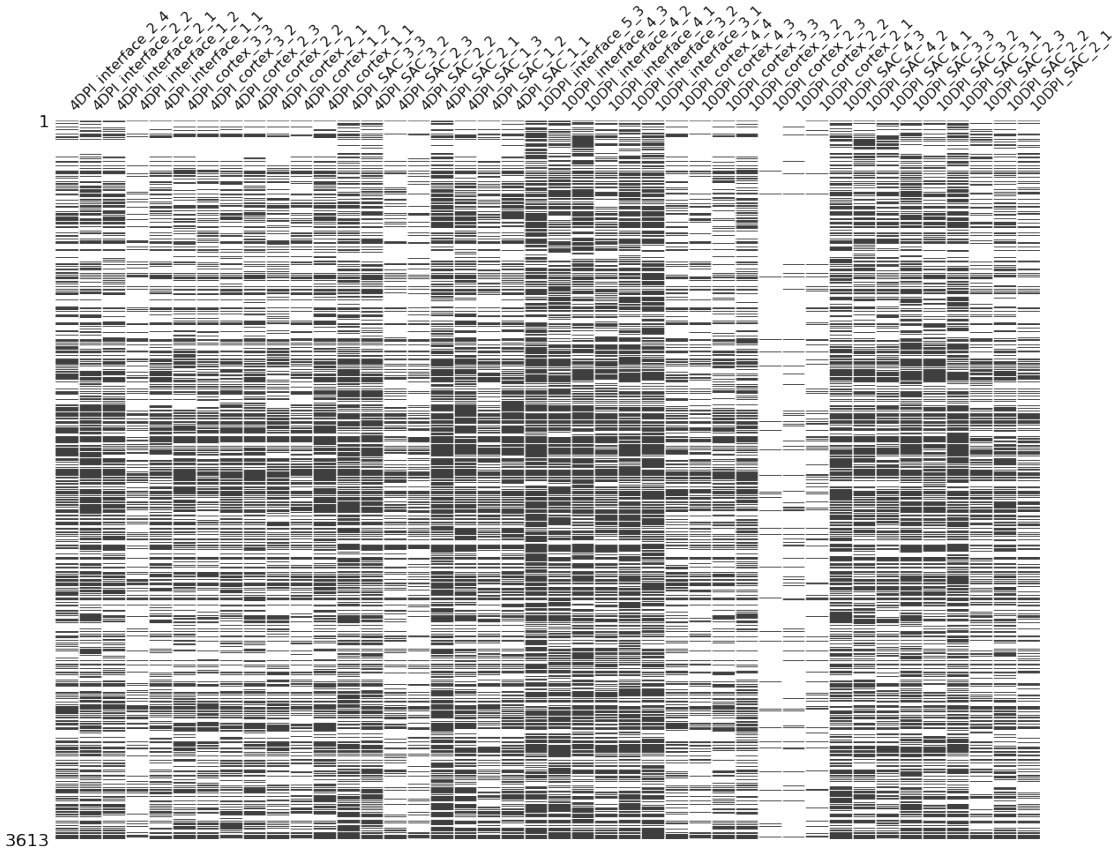
[219]:
```
# calculate missing values (i.e. LFQ = 0)
lfq_nan = lfq.drop(['TimePoint', 'BioRep', 'Region', 'TechRep',␣
 ↪'DPI_Region_BioRep', 'BioRep_TechRep', 'DPI_Region'], axis=1)
lfq_nan = lfq_nan.replace(0, np.nan)
number_missing_values = lfq_nan.transpose().isnull().sum()
percent_missing_values = lfq_nan.transpose().isnull().mean() * 100
```

```
percent_missing_values
```

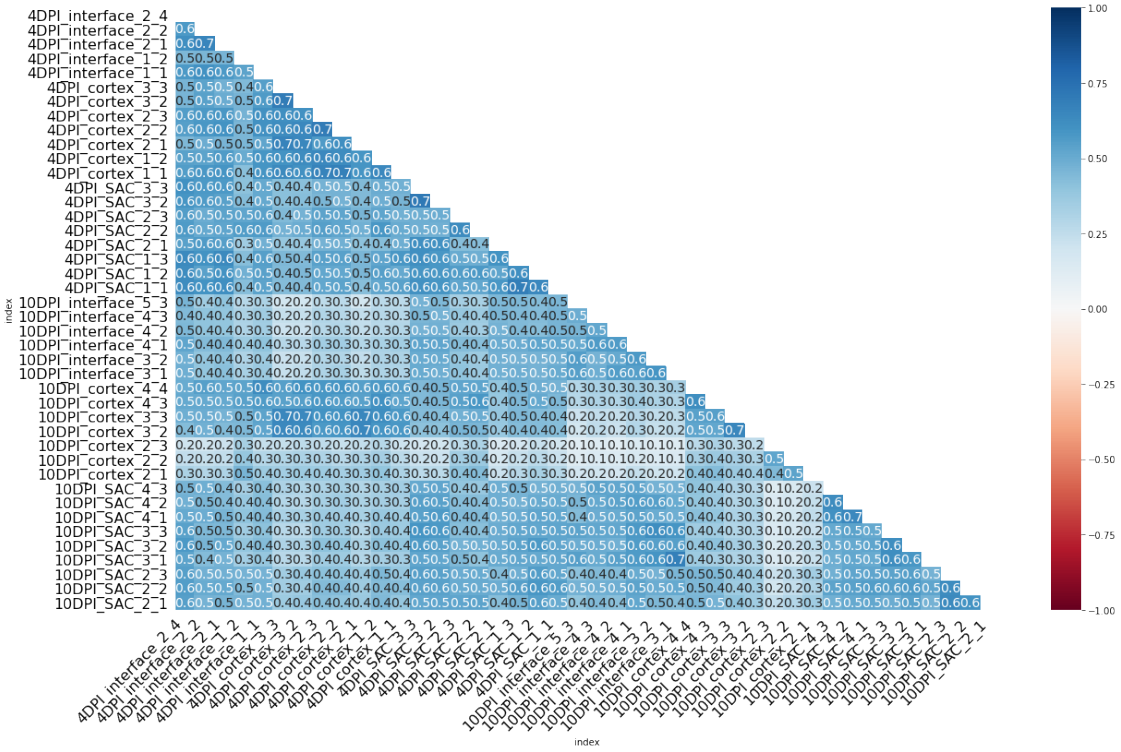
[219]:
```
index
4DPI_interface_2_4     52.809300
4DPI_interface_2_2     47.329089
4DPI_interface_2_1     52.421810
4DPI_interface_1_2     77.359535
4DPI_interface_1_1     59.811791
4DPI_cortex_3_3        56.545807
4DPI_cortex_3_2        60.891226
4DPI_cortex_2_3        57.846665
4DPI_cortex_2_2        53.390534
4DPI_cortex_2_1        58.815389
4DPI_cortex_1_2        65.292001
4DPI_cortex_1_1        50.429006
4DPI_SAC_3_3           46.277332
4DPI_SAC_3_2           48.879048
4DPI_SAC_2_3           71.851647
4DPI_SAC_2_2           73.124827
4DPI_SAC_2_1           37.032937
4DPI_SAC_1_3           46.000554
4DPI_SAC_1_2           59.562690
4DPI_SAC_1_1           48.851370
10DPI_interface_5_3    31.082203
10DPI_interface_4_3    38.998063
10DPI_interface_4_2    36.424024
10DPI_interface_4_1    46.360365
10DPI_interface_3_2    36.534736
10DPI_interface_3_1    37.337393
10DPI_cortex_4_4       64.738445
10DPI_cortex_4_3       72.349848
10DPI_cortex_3_3       63.963465
10DPI_cortex_3_2       59.009134
10DPI_cortex_2_3       95.599225
10DPI_cortex_2_2       93.052865
10DPI_cortex_2_1       89.593136
10DPI_SAC_4_3          41.821201
10DPI_SAC_4_2          47.993357
10DPI_SAC_4_1          51.868254
10DPI_SAC_3_3          42.402436
10DPI_SAC_3_2          52.920011
10DPI_SAC_3_1          41.129255
10DPI_SAC_2_3          63.520620
10DPI_SAC_2_2          55.438694
10DPI_SAC_2_1          66.343759
dtype: float64
```

[220]:
```
pd.DataFrame(percent_missing_values).to_csv('outs/' + date + '_missingValues.
↪csv')
```

[221]:
```
msno.matrix(lfq_nan.transpose(), sparkline=False, figsize=(20,15))
plt.savefig('outs/' + date + '_missingValues_matrix.png')
```



[223]:
```
msno.heatmap(lfq_nan.transpose())
plt.savefig('outs/' + date + '_missingValues_heatmap.svg')
```



[45]:
```
#z-score calculation to drop outliers

z = np.abs(stats.zscore(percent_missing_values))
print(z)
print(np.where(z > 2))
```

```
[0.21484596 0.58111471 0.24074375 1.42596406 0.25316412 0.03488274
 0.32530796 0.12182532 0.17599927 0.1865698  0.61943287 0.37393238
 0.65140871 0.47752355 1.05784547 1.14293821 1.269256   0.66990713
 0.23651554 0.47937339 1.66697207 1.13791721 1.30995253 0.64585919
 1.30255316 1.24890774 0.58243603 1.09114263 0.53064044 0.19951869
 2.64501006 2.47482458 2.24359431 0.9492333  0.5367185  0.27774059
 0.91038662 0.20744659 0.99547936 0.50104297 0.03911095 0.68972687]
(array([30, 31, 32]),)
```

[46]:
```
#z-score calculation to drop outliers
z = np.abs(stats.zscore(percent_missing_values))
z>2
```

[46]:
```
array([False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False,  True,  True,  True, False, False, False,
       False, False, False, False, False, False])
```

[47]:
```
#list of outlier samples
outliers = percent_missing_values.sort_values(ascending=False)[0:3].index
outliers
```

[47]:
```
Index(['10DPI_cortex_2_3', '10DPI_cortex_2_2', '10DPI_cortex_2_1'],
dtype='object', name='index')
```

[48]:
```
#drop outliers
combined_lfq = make_lfq(pGroups_proc, 'combined').drop(outliers, axis=0)
combined_lfq.to_csv('outs/' + date + '_combined_lfq.csv')
```

0.1.1 Remove proteins with missing values

[49]:
```
combined_lfq
```

[49]:

Protein IDs	G3UZW7;A0A023T778;Q9CQL1;P61327 \
index	
4DPI_interface_2_4	0.0
4DPI_interface_2_2	0.0
4DPI_interface_2_1	0.0
4DPI_interface_1_2	0.0
4DPI_interface_1_1	0.0
4DPI_cortex_3_3	0.0
4DPI_cortex_3_2	0.0
4DPI_cortex_2_3	0.0
4DPI_cortex_2_2	0.0
4DPI_cortex_2_1	0.0
4DPI_cortex_1_2	0.0
4DPI_cortex_1_1	0.0
4DPI_SAC_3_3	0.0
4DPI_SAC_3_2	0.0
4DPI_SAC_2_3	0.0
4DPI_SAC_2_2	0.0
4DPI_SAC_2_1	189280.0
4DPI_SAC_1_3	561840.0
4DPI_SAC_1_2	779590.0
4DPI_SAC_1_1	0.0
10DPI_interface_5_3	0.0
10DPI_interface_4_3	0.0
10DPI_interface_4_2	222210.0
10DPI_interface_4_1	0.0
10DPI_interface_3_2	0.0
10DPI_interface_3_1	0.0
10DPI_cortex_4_4	0.0
10DPI_cortex_4_3	0.0
10DPI_cortex_3_3	0.0
10DPI_cortex_3_2	0.0
10DPI_SAC_4_3	0.0

```
10DPI_SAC_4_2                                    0.0
10DPI_SAC_4_1                                    0.0
10DPI_SAC_3_3                                    0.0
10DPI_SAC_3_2                                    0.0
10DPI_SAC_3_1                                    0.0
10DPI_SAC_2_3                                    0.0
10DPI_SAC_2_2                                    0.0
10DPI_SAC_2_1                                    0.0

 Protein IDs         O70589;A0A067XG53;F6Y9I5  \
index
4DPI_interface_2_4                          0.0
4DPI_interface_2_2                          0.0
4DPI_interface_2_1                          0.0
4DPI_interface_1_2                          0.0
4DPI_interface_1_1                          0.0
4DPI_cortex_3_3                             0.0
4DPI_cortex_3_2                             0.0
4DPI_cortex_2_3                             0.0
4DPI_cortex_2_2                             0.0
4DPI_cortex_2_1                        497950.0
4DPI_cortex_1_2                             0.0
4DPI_cortex_1_1                             0.0
4DPI_SAC_3_3                                0.0
4DPI_SAC_3_2                                0.0
4DPI_SAC_2_3                                0.0
4DPI_SAC_2_2                                0.0
4DPI_SAC_2_1                           235580.0
4DPI_SAC_1_3                                0.0
4DPI_SAC_1_2                                0.0
4DPI_SAC_1_1                                0.0
10DPI_interface_5_3                         0.0
10DPI_interface_4_3                         0.0
10DPI_interface_4_2                         0.0
10DPI_interface_4_1                         0.0
10DPI_interface_3_2                         0.0
10DPI_interface_3_1                         0.0
10DPI_cortex_4_4                            0.0
10DPI_cortex_4_3                            0.0
10DPI_cortex_3_3                            0.0
10DPI_cortex_3_2                       434700.0
10DPI_SAC_4_3                               0.0
10DPI_SAC_4_2                               0.0
10DPI_SAC_4_1                               0.0
10DPI_SAC_3_3                               0.0
10DPI_SAC_3_2                               0.0
10DPI_SAC_3_1                               0.0
```

```
10DPI_SAC_2_3                              0.0
10DPI_SAC_2_2                              0.0
10DPI_SAC_2_1                              0.0

 Protein IDs         A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0A075B5K8;F6XW
B2;A0A140T8N1;A0A0G2JDE5;P01631  \
index
4DPI_interface_2_4                                                      0.0
4DPI_interface_2_2                                                      0.0
4DPI_interface_2_1                                                      0.0
4DPI_interface_1_2                                                      0.0
4DPI_interface_1_1                                                      0.0
4DPI_cortex_3_3                                                         0.0
4DPI_cortex_3_2                                                         0.0
4DPI_cortex_2_3                                                         0.0
4DPI_cortex_2_2                                                         0.0
4DPI_cortex_2_1                                                         0.0
4DPI_cortex_1_2                                                         0.0
4DPI_cortex_1_1                                                         0.0
4DPI_SAC_3_3                                                            0.0
4DPI_SAC_3_2                                                            0.0
4DPI_SAC_2_3                                                            0.0
4DPI_SAC_2_2                                                            0.0
4DPI_SAC_2_1                                                            0.0
4DPI_SAC_1_3                                                            0.0
4DPI_SAC_1_2                                                            0.0
4DPI_SAC_1_1                                                            0.0
10DPI_interface_5_3                                                     0.0
10DPI_interface_4_3                                                     0.0
10DPI_interface_4_2                                                     0.0
10DPI_interface_4_1                                                     0.0
10DPI_interface_3_2                                                     0.0
10DPI_interface_3_1                                                     0.0
10DPI_cortex_4_4                                                        0.0
10DPI_cortex_4_3                                                        0.0
10DPI_cortex_3_3                                                   855310.0
10DPI_cortex_3_2                                                   498950.0
10DPI_SAC_4_3                                                     1674100.0
10DPI_SAC_4_2                                                      431080.0
10DPI_SAC_4_1                                                      548100.0
10DPI_SAC_3_3                                                           0.0
10DPI_SAC_3_2                                                           0.0
10DPI_SAC_3_1                                                           0.0
10DPI_SAC_2_3                                                           0.0
10DPI_SAC_2_2                                                           0.0
10DPI_SAC_2_1                                                           0.0
```

	Protein IDs	A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1 \
index		
4DPI_interface_2_4		0.0
4DPI_interface_2_2		0.0
4DPI_interface_2_1		0.0
4DPI_interface_1_2		0.0
4DPI_interface_1_1		0.0
4DPI_cortex_3_3		0.0
4DPI_cortex_3_2		0.0
4DPI_cortex_2_3		0.0
4DPI_cortex_2_2		0.0
4DPI_cortex_2_1		0.0
4DPI_cortex_1_2		0.0
4DPI_cortex_1_1		0.0
4DPI_SAC_3_3		0.0
4DPI_SAC_3_2		0.0
4DPI_SAC_2_3		0.0
4DPI_SAC_2_2		0.0
4DPI_SAC_2_1		0.0
4DPI_SAC_1_3		0.0
4DPI_SAC_1_2		0.0
4DPI_SAC_1_1		0.0
10DPI_interface_5_3		81477.0
10DPI_interface_4_3		207140.0
10DPI_interface_4_2		0.0
10DPI_interface_4_1		0.0
10DPI_interface_3_2		0.0
10DPI_interface_3_1		0.0
10DPI_cortex_4_4		0.0
10DPI_cortex_4_3		0.0
10DPI_cortex_3_3		0.0
10DPI_cortex_3_2		0.0
10DPI_SAC_4_3		202930.0
10DPI_SAC_4_2		290690.0
10DPI_SAC_4_1		0.0
10DPI_SAC_3_3		0.0
10DPI_SAC_3_2		0.0
10DPI_SAC_3_1		0.0
10DPI_SAC_2_3		0.0
10DPI_SAC_2_2		0.0
10DPI_SAC_2_1		0.0

	Protein IDs	A0A075B5P3;A0A0A6YVP0;P01867 \
index		
4DPI_interface_2_4		0.0
4DPI_interface_2_2		0.0
4DPI_interface_2_1		0.0

```
4DPI_interface_1_2                          0.0
4DPI_interface_1_1                          0.0
4DPI_cortex_3_3                             0.0
4DPI_cortex_3_2                             0.0
4DPI_cortex_2_3                             0.0
4DPI_cortex_2_2                             0.0
4DPI_cortex_2_1                             0.0
4DPI_cortex_1_2                             0.0
4DPI_cortex_1_1                             0.0
4DPI_SAC_3_3                                0.0
4DPI_SAC_3_2                                0.0
4DPI_SAC_2_3                                0.0
4DPI_SAC_2_2                                0.0
4DPI_SAC_2_1                                0.0
4DPI_SAC_1_3                                0.0
4DPI_SAC_1_2                                0.0
4DPI_SAC_1_1                                0.0
10DPI_interface_5_3                         0.0
10DPI_interface_4_3                   1183700.0
10DPI_interface_4_2                   2295600.0
10DPI_interface_4_1                   1475700.0
10DPI_interface_3_2                    869990.0
10DPI_interface_3_1                         0.0
10DPI_cortex_4_4                            0.0
10DPI_cortex_4_3                            0.0
10DPI_cortex_3_3                            0.0
10DPI_cortex_3_2                            0.0
10DPI_SAC_4_3                         1207600.0
10DPI_SAC_4_2                         1526100.0
10DPI_SAC_4_1                         1165600.0
10DPI_SAC_3_3                               0.0
10DPI_SAC_3_2                               0.0
10DPI_SAC_3_1                          724260.0
10DPI_SAC_2_3                               0.0
10DPI_SAC_2_2                               0.0
10DPI_SAC_2_1                               0.0

 Protein IDs         A0A075B5P5;A0A1Y7VJN6;P03987  \
index
4DPI_interface_2_4                          0.0
4DPI_interface_2_2                          0.0
4DPI_interface_2_1                          0.0
4DPI_interface_1_2                          0.0
4DPI_interface_1_1                          0.0
4DPI_cortex_3_3                             0.0
4DPI_cortex_3_2                             0.0
4DPI_cortex_2_3                             0.0
```

```
4DPI_cortex_2_2                              0.0
4DPI_cortex_2_1                              0.0
4DPI_cortex_1_2                              0.0
4DPI_cortex_1_1                              0.0
4DPI_SAC_3_3                                 0.0
4DPI_SAC_3_2                                 0.0
4DPI_SAC_2_3                                 0.0
4DPI_SAC_2_2                                 0.0
4DPI_SAC_2_1                                 0.0
4DPI_SAC_1_3                                 0.0
4DPI_SAC_1_2                                 0.0
4DPI_SAC_1_1                                 0.0
10DPI_interface_5_3                    1109600.0
10DPI_interface_4_3                   12624000.0
10DPI_interface_4_2                   14386000.0
10DPI_interface_4_1                   19824000.0
10DPI_interface_3_2                    3005900.0
10DPI_interface_3_1                     447980.0
10DPI_cortex_4_4                      21601000.0
10DPI_cortex_4_3                        526710.0
10DPI_cortex_3_3                       1102700.0
10DPI_cortex_3_2                        654170.0
10DPI_SAC_4_3                         29513000.0
10DPI_SAC_4_2                         13752000.0
10DPI_SAC_4_1                         17690000.0
10DPI_SAC_3_3                           848950.0
10DPI_SAC_3_2                           997060.0
10DPI_SAC_3_1                           295760.0
10DPI_SAC_2_3                                0.0
10DPI_SAC_2_2                          1941400.0
10DPI_SAC_2_1                                0.0

 Protein IDs         A0A075B6A0;A0A075B5P6;P01872  \
index
4DPI_interface_2_4                    11850000.0
4DPI_interface_2_2                    13198000.0
4DPI_interface_2_1                    14041000.0
4DPI_interface_1_2                     8327300.0
4DPI_interface_1_1                     9088300.0
4DPI_cortex_3_3                       28479000.0
4DPI_cortex_3_2                        9475600.0
4DPI_cortex_2_3                       12548000.0
4DPI_cortex_2_2                       17613000.0
4DPI_cortex_2_1                       18376000.0
4DPI_cortex_1_2                       13749000.0
4DPI_cortex_1_1                       19070000.0
4DPI_SAC_3_3                           4038600.0
```

4DPI_SAC_3_2	1051500.0
4DPI_SAC_2_3	10500000.0
4DPI_SAC_2_2	4068800.0
4DPI_SAC_2_1	6212200.0
4DPI_SAC_1_3	7820100.0
4DPI_SAC_1_2	3859700.0
4DPI_SAC_1_1	4501600.0
10DPI_interface_5_3	7111900.0
10DPI_interface_4_3	15394000.0
10DPI_interface_4_2	20355000.0
10DPI_interface_4_1	24453000.0
10DPI_interface_3_2	6101300.0
10DPI_interface_3_1	9250300.0
10DPI_cortex_4_4	49938000.0
10DPI_cortex_4_3	16417000.0
10DPI_cortex_3_3	16583000.0
10DPI_cortex_3_2	14249000.0
10DPI_SAC_4_3	36836000.0
10DPI_SAC_4_2	21015000.0
10DPI_SAC_4_1	20078000.0
10DPI_SAC_3_3	5357600.0
10DPI_SAC_3_2	2988700.0
10DPI_SAC_3_1	6025900.0
10DPI_SAC_2_3	0.0
10DPI_SAC_2_2	6863300.0
10DPI_SAC_2_1	2568900.0

Protein IDs	A0A075B5R2;A0A075B5R3;A0A0A6YX91;P01786 \
index	
4DPI_interface_2_4	489990.0
4DPI_interface_2_2	0.0
4DPI_interface_2_1	1280300.0
4DPI_interface_1_2	0.0
4DPI_interface_1_1	0.0
4DPI_cortex_3_3	0.0
4DPI_cortex_3_2	0.0
4DPI_cortex_2_3	537080.0
4DPI_cortex_2_2	1492000.0
4DPI_cortex_2_1	247720.0
4DPI_cortex_1_2	0.0
4DPI_cortex_1_1	0.0
4DPI_SAC_3_3	0.0
4DPI_SAC_3_2	0.0
4DPI_SAC_2_3	0.0
4DPI_SAC_2_2	0.0
4DPI_SAC_2_1	0.0
4DPI_SAC_1_3	0.0

4DPI_SAC_1_2	0.0
4DPI_SAC_1_1	0.0
10DPI_interface_5_3	923480.0
10DPI_interface_4_3	2742000.0
10DPI_interface_4_2	2199900.0
10DPI_interface_4_1	2193900.0
10DPI_interface_3_2	875140.0
10DPI_interface_3_1	1436900.0
10DPI_cortex_4_4	1717800.0
10DPI_cortex_4_3	1395700.0
10DPI_cortex_3_3	880010.0
10DPI_cortex_3_2	708140.0
10DPI_SAC_4_3	1401500.0
10DPI_SAC_4_2	0.0
10DPI_SAC_4_1	1453700.0
10DPI_SAC_3_3	536050.0
10DPI_SAC_3_2	2143600.0
10DPI_SAC_3_1	944670.0
10DPI_SAC_2_3	0.0
10DPI_SAC_2_2	533920.0
10DPI_SAC_2_1	0.0

Protein IDs	A0A075B5S5;A0A0A6YXL5	A0A075B5T3;J3QK03;A0A0A6YWS9	...	\
index			...	
4DPI_interface_2_4	0.0	0.0	...	
4DPI_interface_2_2	0.0	0.0	...	
4DPI_interface_2_1	0.0	0.0	...	
4DPI_interface_1_2	0.0	0.0	...	
4DPI_interface_1_1	0.0	0.0	...	
4DPI_cortex_3_3	0.0	0.0	...	
4DPI_cortex_3_2	0.0	0.0	...	
4DPI_cortex_2_3	0.0	0.0	...	
4DPI_cortex_2_2	0.0	0.0	...	
4DPI_cortex_2_1	0.0	0.0	...	
4DPI_cortex_1_2	0.0	0.0	...	
4DPI_cortex_1_1	0.0	0.0	...	
4DPI_SAC_3_3	0.0	0.0	...	
4DPI_SAC_3_2	0.0	0.0	...	
4DPI_SAC_2_3	0.0	0.0	...	
4DPI_SAC_2_2	0.0	0.0	...	
4DPI_SAC_2_1	0.0	0.0	...	
4DPI_SAC_1_3	0.0	0.0	...	
4DPI_SAC_1_2	0.0	0.0	...	
4DPI_SAC_1_1	0.0	0.0	...	
10DPI_interface_5_3	582000.0	624130.0	...	
10DPI_interface_4_3	10474000.0	1520300.0	...	
10DPI_interface_4_2	4267600.0	3319600.0	...	

10DPI_interface_4_1	1899800.0	1112100.0	...
10DPI_interface_3_2	521140.0	826650.0	...
10DPI_interface_3_1	554440.0	0.0	...
10DPI_cortex_4_4	3546300.0	2818100.0	...
10DPI_cortex_4_3	0.0	2965700.0	...
10DPI_cortex_3_3	0.0	992900.0	...
10DPI_cortex_3_2	739180.0	0.0	...
10DPI_SAC_4_3	2447600.0	5502600.0	...
10DPI_SAC_4_2	3972000.0	2063300.0	...
10DPI_SAC_4_1	8814000.0	2556300.0	...
10DPI_SAC_3_3	0.0	1867700.0	...
10DPI_SAC_3_2	0.0	766080.0	...
10DPI_SAC_3_1	0.0	0.0	...
10DPI_SAC_2_3	0.0	0.0	...
10DPI_SAC_2_2	0.0	0.0	...
10DPI_SAC_2_1	0.0	0.0	...

Protein IDs	Q9Z315;A0A494B9E9 \
index	
4DPI_interface_2_4	954220.0
4DPI_interface_2_2	1377400.0
4DPI_interface_2_1	922440.0
4DPI_interface_1_2	0.0
4DPI_interface_1_1	1117600.0
4DPI_cortex_3_3	638210.0
4DPI_cortex_3_2	661230.0
4DPI_cortex_2_3	0.0
4DPI_cortex_2_2	494860.0
4DPI_cortex_2_1	674540.0
4DPI_cortex_1_2	0.0
4DPI_cortex_1_1	334760.0
4DPI_SAC_3_3	1396800.0
4DPI_SAC_3_2	1242700.0
4DPI_SAC_2_3	1401200.0
4DPI_SAC_2_2	0.0
4DPI_SAC_2_1	458240.0
4DPI_SAC_1_3	780230.0
4DPI_SAC_1_2	1219000.0
4DPI_SAC_1_1	600830.0
10DPI_interface_5_3	1281000.0
10DPI_interface_4_3	2218300.0
10DPI_interface_4_2	1495700.0
10DPI_interface_4_1	927020.0
10DPI_interface_3_2	1865800.0
10DPI_interface_3_1	1160000.0
10DPI_cortex_4_4	946380.0
10DPI_cortex_4_3	0.0

10DPI_cortex_3_3	633750.0
10DPI_cortex_3_2	556170.0
10DPI_SAC_4_3	1137800.0
10DPI_SAC_4_2	1302200.0
10DPI_SAC_4_1	0.0
10DPI_SAC_3_3	1731200.0
10DPI_SAC_3_2	1444700.0
10DPI_SAC_3_1	1059500.0
10DPI_SAC_2_3	3343300.0
10DPI_SAC_2_2	800330.0
10DPI_SAC_2_1	2338200.0

Protein IDs	S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5	S4R192	\
index			
4DPI_interface_2_4	0.0	0.0	
4DPI_interface_2_2	0.0	0.0	
4DPI_interface_2_1	0.0	0.0	
4DPI_interface_1_2	0.0	0.0	
4DPI_interface_1_1	0.0	0.0	
4DPI_cortex_3_3	0.0	0.0	
4DPI_cortex_3_2	0.0	0.0	
4DPI_cortex_2_3	662690.0	0.0	
4DPI_cortex_2_2	0.0	0.0	
4DPI_cortex_2_1	0.0	0.0	
4DPI_cortex_1_2	714480.0	0.0	
4DPI_cortex_1_1	716000.0	0.0	
4DPI_SAC_3_3	407450.0	0.0	
4DPI_SAC_3_2	0.0	357930.0	
4DPI_SAC_2_3	0.0	0.0	
4DPI_SAC_2_2	0.0	0.0	
4DPI_SAC_2_1	360500.0	169050.0	
4DPI_SAC_1_3	610920.0	0.0	
4DPI_SAC_1_2	326340.0	0.0	
4DPI_SAC_1_1	0.0	251670.0	
10DPI_interface_5_3	126670.0	0.0	
10DPI_interface_4_3	0.0	305120.0	
10DPI_interface_4_2	0.0	549550.0	
10DPI_interface_4_1	0.0	0.0	
10DPI_interface_3_2	0.0	364880.0	
10DPI_interface_3_1	0.0	781480.0	
10DPI_cortex_4_4	0.0	0.0	
10DPI_cortex_4_3	0.0	0.0	
10DPI_cortex_3_3	0.0	0.0	
10DPI_cortex_3_2	0.0	0.0	
10DPI_SAC_4_3	0.0	0.0	
10DPI_SAC_4_2	0.0	0.0	
10DPI_SAC_4_1	0.0	326950.0	

```
10DPI_SAC_3_3                                          0.0  423330.0
10DPI_SAC_3_2                                          0.0       0.0
10DPI_SAC_3_1                                          0.0       0.0
10DPI_SAC_2_3                                          0.0       0.0
10DPI_SAC_2_2                                          0.0       0.0
10DPI_SAC_2_1                                          0.0       0.0

 Protein IDs         TimePoint  BioRep     Region  TechRep  DPI_Region_BioRep  \
index
4DPI_interface_2_4        4DPI       2  interface        4   4DPI_interface_2
4DPI_interface_2_2        4DPI       2  interface        2   4DPI_interface_2
4DPI_interface_2_1        4DPI       2  interface        1   4DPI_interface_2
4DPI_interface_1_2        4DPI       1  interface        2   4DPI_interface_1
4DPI_interface_1_1        4DPI       1  interface        1   4DPI_interface_1
4DPI_cortex_3_3           4DPI       3     cortex        3      4DPI_cortex_3
4DPI_cortex_3_2           4DPI       3     cortex        2      4DPI_cortex_3
4DPI_cortex_2_3           4DPI       2     cortex        3      4DPI_cortex_2
4DPI_cortex_2_2           4DPI       2     cortex        2      4DPI_cortex_2
4DPI_cortex_2_1           4DPI       2     cortex        1      4DPI_cortex_2
4DPI_cortex_1_2           4DPI       1     cortex        2      4DPI_cortex_1
4DPI_cortex_1_1           4DPI       1     cortex        1      4DPI_cortex_1
4DPI_SAC_3_3              4DPI       3        SAC        3         4DPI_SAC_3
4DPI_SAC_3_2              4DPI       3        SAC        2         4DPI_SAC_3
4DPI_SAC_2_3              4DPI       2        SAC        3         4DPI_SAC_2
4DPI_SAC_2_2              4DPI       2        SAC        2         4DPI_SAC_2
4DPI_SAC_2_1              4DPI       2        SAC        1         4DPI_SAC_2
4DPI_SAC_1_3              4DPI       1        SAC        3         4DPI_SAC_1
4DPI_SAC_1_2              4DPI       1        SAC        2         4DPI_SAC_1
4DPI_SAC_1_1              4DPI       1        SAC        1         4DPI_SAC_1
10DPI_interface_5_3      10DPI       5  interface        3  10DPI_interface_5
10DPI_interface_4_3      10DPI       4  interface        3  10DPI_interface_4
10DPI_interface_4_2      10DPI       4  interface        2  10DPI_interface_4
10DPI_interface_4_1      10DPI       4  interface        1  10DPI_interface_4
10DPI_interface_3_2      10DPI       3  interface        2  10DPI_interface_3
10DPI_interface_3_1      10DPI       3  interface        1  10DPI_interface_3
10DPI_cortex_4_4         10DPI       4     cortex        4     10DPI_cortex_4
10DPI_cortex_4_3         10DPI       4     cortex        3     10DPI_cortex_4
10DPI_cortex_3_3         10DPI       3     cortex        3     10DPI_cortex_3
10DPI_cortex_3_2         10DPI       3     cortex        2     10DPI_cortex_3
10DPI_SAC_4_3            10DPI       4        SAC        3        10DPI_SAC_4
10DPI_SAC_4_2            10DPI       4        SAC        2        10DPI_SAC_4
10DPI_SAC_4_1            10DPI       4        SAC        1        10DPI_SAC_4
10DPI_SAC_3_3            10DPI       3        SAC        3        10DPI_SAC_3
10DPI_SAC_3_2            10DPI       3        SAC        2        10DPI_SAC_3
10DPI_SAC_3_1            10DPI       3        SAC        1        10DPI_SAC_3
10DPI_SAC_2_3            10DPI       2        SAC        3        10DPI_SAC_2
10DPI_SAC_2_2            10DPI       2        SAC        2        10DPI_SAC_2
```

```
10DPI_SAC_2_1            10DPI       2        SAC        1        10DPI_SAC_2

 Protein IDs         BioRep_TechRep       DPI_Region
index
4DPI_interface_2_4              2_4   4DPI_interface
4DPI_interface_2_2              2_2   4DPI_interface
4DPI_interface_2_1              2_1   4DPI_interface
4DPI_interface_1_2              1_2   4DPI_interface
4DPI_interface_1_1              1_1   4DPI_interface
4DPI_cortex_3_3                 3_3      4DPI_cortex
4DPI_cortex_3_2                 3_2      4DPI_cortex
4DPI_cortex_2_3                 2_3      4DPI_cortex
4DPI_cortex_2_2                 2_2      4DPI_cortex
4DPI_cortex_2_1                 2_1      4DPI_cortex
4DPI_cortex_1_2                 1_2      4DPI_cortex
4DPI_cortex_1_1                 1_1      4DPI_cortex
4DPI_SAC_3_3                    3_3         4DPI_SAC
4DPI_SAC_3_2                    3_2         4DPI_SAC
4DPI_SAC_2_3                    2_3         4DPI_SAC
4DPI_SAC_2_2                    2_2         4DPI_SAC
4DPI_SAC_2_1                    2_1         4DPI_SAC
4DPI_SAC_1_3                    1_3         4DPI_SAC
4DPI_SAC_1_2                    1_2         4DPI_SAC
4DPI_SAC_1_1                    1_1         4DPI_SAC
10DPI_interface_5_3             5_3  10DPI_interface
10DPI_interface_4_3             4_3  10DPI_interface
10DPI_interface_4_2             4_2  10DPI_interface
10DPI_interface_4_1             4_1  10DPI_interface
10DPI_interface_3_2             3_2  10DPI_interface
10DPI_interface_3_1             3_1  10DPI_interface
10DPI_cortex_4_4                4_4     10DPI_cortex
10DPI_cortex_4_3                4_3     10DPI_cortex
10DPI_cortex_3_3                3_3     10DPI_cortex
10DPI_cortex_3_2                3_2     10DPI_cortex
10DPI_SAC_4_3                   4_3        10DPI_SAC
10DPI_SAC_4_2                   4_2        10DPI_SAC
10DPI_SAC_4_1                   4_1        10DPI_SAC
10DPI_SAC_3_3                   3_3        10DPI_SAC
10DPI_SAC_3_2                   3_2        10DPI_SAC
10DPI_SAC_3_1                   3_1        10DPI_SAC
10DPI_SAC_2_3                   2_3        10DPI_SAC
10DPI_SAC_2_2                   2_2        10DPI_SAC
10DPI_SAC_2_1                   2_1        10DPI_SAC

[39 rows x 3620 columns]
```

[50]:
```
combined_lfq = combined_lfq.replace(0, np.nan)
combined_lfq
```

[50]:
```
 Protein IDs         G3UZW7;A0A023T778;Q9CQL1;P61327  \
index
4DPI_interface_2_4                                NaN
4DPI_interface_2_2                                NaN
4DPI_interface_2_1                                NaN
4DPI_interface_1_2                                NaN
4DPI_interface_1_1                                NaN
4DPI_cortex_3_3                                   NaN
4DPI_cortex_3_2                                   NaN
4DPI_cortex_2_3                                   NaN
4DPI_cortex_2_2                                   NaN
4DPI_cortex_2_1                                   NaN
4DPI_cortex_1_2                                   NaN
4DPI_cortex_1_1                                   NaN
4DPI_SAC_3_3                                      NaN
4DPI_SAC_3_2                                      NaN
4DPI_SAC_2_3                                      NaN
4DPI_SAC_2_2                                      NaN
4DPI_SAC_2_1                                 189280.0
4DPI_SAC_1_3                                 561840.0
4DPI_SAC_1_2                                 779590.0
4DPI_SAC_1_1                                      NaN
10DPI_interface_5_3                               NaN
10DPI_interface_4_3                               NaN
10DPI_interface_4_2                          222210.0
10DPI_interface_4_1                               NaN
10DPI_interface_3_2                               NaN
10DPI_interface_3_1                               NaN
10DPI_cortex_4_4                                  NaN
10DPI_cortex_4_3                                  NaN
10DPI_cortex_3_3                                  NaN
10DPI_cortex_3_2                                  NaN
10DPI_SAC_4_3                                     NaN
10DPI_SAC_4_2                                     NaN
10DPI_SAC_4_1                                     NaN
10DPI_SAC_3_3                                     NaN
10DPI_SAC_3_2                                     NaN
10DPI_SAC_3_1                                     NaN
10DPI_SAC_2_3                                     NaN
10DPI_SAC_2_2                                     NaN
10DPI_SAC_2_1                                     NaN

 Protein IDs         O70589;A0A067XG53;F6Y9I5  \
index
```

```
4DPI_interface_2_4                      NaN
4DPI_interface_2_2                      NaN
4DPI_interface_2_1                      NaN
4DPI_interface_1_2                      NaN
4DPI_interface_1_1                      NaN
4DPI_cortex_3_3                         NaN
4DPI_cortex_3_2                         NaN
4DPI_cortex_2_3                         NaN
4DPI_cortex_2_2                         NaN
4DPI_cortex_2_1                    497950.0
4DPI_cortex_1_2                         NaN
4DPI_cortex_1_1                         NaN
4DPI_SAC_3_3                            NaN
4DPI_SAC_3_2                            NaN
4DPI_SAC_2_3                            NaN
4DPI_SAC_2_2                            NaN
4DPI_SAC_2_1                       235580.0
4DPI_SAC_1_3                            NaN
4DPI_SAC_1_2                            NaN
4DPI_SAC_1_1                            NaN
10DPI_interface_5_3                     NaN
10DPI_interface_4_3                     NaN
10DPI_interface_4_2                     NaN
10DPI_interface_4_1                     NaN
10DPI_interface_3_2                     NaN
10DPI_interface_3_1                     NaN
10DPI_cortex_4_4                        NaN
10DPI_cortex_4_3                        NaN
10DPI_cortex_3_3                        NaN
10DPI_cortex_3_2                   434700.0
10DPI_SAC_4_3                           NaN
10DPI_SAC_4_2                           NaN
10DPI_SAC_4_1                           NaN
10DPI_SAC_3_3                           NaN
10DPI_SAC_3_2                           NaN
10DPI_SAC_3_1                           NaN
10DPI_SAC_2_3                           NaN
10DPI_SAC_2_2                           NaN
10DPI_SAC_2_1                           NaN

 Protein IDs         A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0A075B5K8;F6XW
B2;A0A140T8N1;A0A0G2JDE5;P01631  \
index
4DPI_interface_2_4                                                      NaN
4DPI_interface_2_2                                                      NaN
4DPI_interface_2_1                                                      NaN
4DPI_interface_1_2                                                      NaN
```

```
4DPI_interface_1_1                                                    NaN
4DPI_cortex_3_3                                                       NaN
4DPI_cortex_3_2                                                       NaN
4DPI_cortex_2_3                                                       NaN
4DPI_cortex_2_2                                                       NaN
4DPI_cortex_2_1                                                       NaN
4DPI_cortex_1_2                                                       NaN
4DPI_cortex_1_1                                                       NaN
4DPI_SAC_3_3                                                          NaN
4DPI_SAC_3_2                                                          NaN
4DPI_SAC_2_3                                                          NaN
4DPI_SAC_2_2                                                          NaN
4DPI_SAC_2_1                                                          NaN
4DPI_SAC_1_3                                                          NaN
4DPI_SAC_1_2                                                          NaN
4DPI_SAC_1_1                                                          NaN
10DPI_interface_5_3                                                   NaN
10DPI_interface_4_3                                                   NaN
10DPI_interface_4_2                                                   NaN
10DPI_interface_4_1                                                   NaN
10DPI_interface_3_2                                                   NaN
10DPI_interface_3_1                                                   NaN
10DPI_cortex_4_4                                                      NaN
10DPI_cortex_4_3                                                      NaN
10DPI_cortex_3_3                                                 855310.0
10DPI_cortex_3_2                                                 498950.0
10DPI_SAC_4_3                                                   1674100.0
10DPI_SAC_4_2                                                    431080.0
10DPI_SAC_4_1                                                    548100.0
10DPI_SAC_3_3                                                         NaN
10DPI_SAC_3_2                                                         NaN
10DPI_SAC_3_1                                                         NaN
10DPI_SAC_2_3                                                         NaN
10DPI_SAC_2_2                                                         NaN
10DPI_SAC_2_1                                                         NaN

 Protein IDs         A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1  \
index
4DPI_interface_2_4                                                  NaN
4DPI_interface_2_2                                                  NaN
4DPI_interface_2_1                                                  NaN
4DPI_interface_1_2                                                  NaN
4DPI_interface_1_1                                                  NaN
4DPI_cortex_3_3                                                     NaN
4DPI_cortex_3_2                                                     NaN
4DPI_cortex_2_3                                                     NaN
4DPI_cortex_2_2                                                     NaN
```

```
4DPI_cortex_2_1                                          NaN
4DPI_cortex_1_2                                          NaN
4DPI_cortex_1_1                                          NaN
4DPI_SAC_3_3                                             NaN
4DPI_SAC_3_2                                             NaN
4DPI_SAC_2_3                                             NaN
4DPI_SAC_2_2                                             NaN
4DPI_SAC_2_1                                             NaN
4DPI_SAC_1_3                                             NaN
4DPI_SAC_1_2                                             NaN
4DPI_SAC_1_1                                             NaN
10DPI_interface_5_3                                  81477.0
10DPI_interface_4_3                                 207140.0
10DPI_interface_4_2                                      NaN
10DPI_interface_4_1                                      NaN
10DPI_interface_3_2                                      NaN
10DPI_interface_3_1                                      NaN
10DPI_cortex_4_4                                         NaN
10DPI_cortex_4_3                                         NaN
10DPI_cortex_3_3                                         NaN
10DPI_cortex_3_2                                         NaN
10DPI_SAC_4_3                                       202930.0
10DPI_SAC_4_2                                       290690.0
10DPI_SAC_4_1                                            NaN
10DPI_SAC_3_3                                            NaN
10DPI_SAC_3_2                                            NaN
10DPI_SAC_3_1                                            NaN
10DPI_SAC_2_3                                            NaN
10DPI_SAC_2_2                                            NaN
10DPI_SAC_2_1                                            NaN

 Protein IDs         A0A075B5P3;A0A0A6YVP0;P01867  \
index
4DPI_interface_2_4                              NaN
4DPI_interface_2_2                              NaN
4DPI_interface_2_1                              NaN
4DPI_interface_1_2                              NaN
4DPI_interface_1_1                              NaN
4DPI_cortex_3_3                                 NaN
4DPI_cortex_3_2                                 NaN
4DPI_cortex_2_3                                 NaN
4DPI_cortex_2_2                                 NaN
4DPI_cortex_2_1                                 NaN
4DPI_cortex_1_2                                 NaN
4DPI_cortex_1_1                                 NaN
4DPI_SAC_3_3                                    NaN
4DPI_SAC_3_2                                    NaN
```

```
4DPI_SAC_2_3                                NaN
4DPI_SAC_2_2                                NaN
4DPI_SAC_2_1                                NaN
4DPI_SAC_1_3                                NaN
4DPI_SAC_1_2                                NaN
4DPI_SAC_1_1                                NaN
10DPI_interface_5_3                         NaN
10DPI_interface_4_3                   1183700.0
10DPI_interface_4_2                   2295600.0
10DPI_interface_4_1                   1475700.0
10DPI_interface_3_2                    869990.0
10DPI_interface_3_1                         NaN
10DPI_cortex_4_4                            NaN
10DPI_cortex_4_3                            NaN
10DPI_cortex_3_3                            NaN
10DPI_cortex_3_2                            NaN
10DPI_SAC_4_3                         1207600.0
10DPI_SAC_4_2                         1526100.0
10DPI_SAC_4_1                         1165600.0
10DPI_SAC_3_3                               NaN
10DPI_SAC_3_2                               NaN
10DPI_SAC_3_1                          724260.0
10DPI_SAC_2_3                               NaN
10DPI_SAC_2_2                               NaN
10DPI_SAC_2_1                               NaN

 Protein IDs         A0A075B5P5;A0A1Y7VJN6;P03987  \
index
4DPI_interface_2_4                          NaN
4DPI_interface_2_2                          NaN
4DPI_interface_2_1                          NaN
4DPI_interface_1_2                          NaN
4DPI_interface_1_1                          NaN
4DPI_cortex_3_3                             NaN
4DPI_cortex_3_2                             NaN
4DPI_cortex_2_3                             NaN
4DPI_cortex_2_2                             NaN
4DPI_cortex_2_1                             NaN
4DPI_cortex_1_2                             NaN
4DPI_cortex_1_1                             NaN
4DPI_SAC_3_3                                NaN
4DPI_SAC_3_2                                NaN
4DPI_SAC_2_3                                NaN
4DPI_SAC_2_2                                NaN
4DPI_SAC_2_1                                NaN
4DPI_SAC_1_3                                NaN
4DPI_SAC_1_2                                NaN
```

```
4DPI_SAC_1_1                                  NaN
10DPI_interface_5_3                     1109600.0
10DPI_interface_4_3                    12624000.0
10DPI_interface_4_2                    14386000.0
10DPI_interface_4_1                    19824000.0
10DPI_interface_3_2                     3005900.0
10DPI_interface_3_1                      447980.0
10DPI_cortex_4_4                       21601000.0
10DPI_cortex_4_3                         526710.0
10DPI_cortex_3_3                        1102700.0
10DPI_cortex_3_2                         654170.0
10DPI_SAC_4_3                          29513000.0
10DPI_SAC_4_2                          13752000.0
10DPI_SAC_4_1                          17690000.0
10DPI_SAC_3_3                            848950.0
10DPI_SAC_3_2                            997060.0
10DPI_SAC_3_1                            295760.0
10DPI_SAC_2_3                                 NaN
10DPI_SAC_2_2                           1941400.0
10DPI_SAC_2_1                                 NaN

 Protein IDs         A0A075B6A0;A0A075B5P6;P01872  \
index
4DPI_interface_2_4                     11850000.0
4DPI_interface_2_2                     13198000.0
4DPI_interface_2_1                     14041000.0
4DPI_interface_1_2                      8327300.0
4DPI_interface_1_1                      9088300.0
4DPI_cortex_3_3                        28479000.0
4DPI_cortex_3_2                         9475600.0
4DPI_cortex_2_3                        12548000.0
4DPI_cortex_2_2                        17613000.0
4DPI_cortex_2_1                        18376000.0
4DPI_cortex_1_2                        13749000.0
4DPI_cortex_1_1                        19070000.0
4DPI_SAC_3_3                            4038600.0
4DPI_SAC_3_2                            1051500.0
4DPI_SAC_2_3                           10500000.0
4DPI_SAC_2_2                            4068800.0
4DPI_SAC_2_1                            6212200.0
4DPI_SAC_1_3                            7820100.0
4DPI_SAC_1_2                            3859700.0
4DPI_SAC_1_1                            4501600.0
10DPI_interface_5_3                     7111900.0
10DPI_interface_4_3                    15394000.0
10DPI_interface_4_2                    20355000.0
10DPI_interface_4_1                    24453000.0
```

10DPI_interface_3_2	6101300.0
10DPI_interface_3_1	9250300.0
10DPI_cortex_4_4	49938000.0
10DPI_cortex_4_3	16417000.0
10DPI_cortex_3_3	16583000.0
10DPI_cortex_3_2	14249000.0
10DPI_SAC_4_3	36836000.0
10DPI_SAC_4_2	21015000.0
10DPI_SAC_4_1	20078000.0
10DPI_SAC_3_3	5357600.0
10DPI_SAC_3_2	2988700.0
10DPI_SAC_3_1	6025900.0
10DPI_SAC_2_3	NaN
10DPI_SAC_2_2	6863300.0
10DPI_SAC_2_1	2568900.0

Protein IDs	A0A075B5R2;A0A075B5R3;A0A0A6YX91;P01786 \
index	
4DPI_interface_2_4	489990.0
4DPI_interface_2_2	NaN
4DPI_interface_2_1	1280300.0
4DPI_interface_1_2	NaN
4DPI_interface_1_1	NaN
4DPI_cortex_3_3	NaN
4DPI_cortex_3_2	NaN
4DPI_cortex_2_3	537080.0
4DPI_cortex_2_2	1492000.0
4DPI_cortex_2_1	247720.0
4DPI_cortex_1_2	NaN
4DPI_cortex_1_1	NaN
4DPI_SAC_3_3	NaN
4DPI_SAC_3_2	NaN
4DPI_SAC_2_3	NaN
4DPI_SAC_2_2	NaN
4DPI_SAC_2_1	NaN
4DPI_SAC_1_3	NaN
4DPI_SAC_1_2	NaN
4DPI_SAC_1_1	NaN
10DPI_interface_5_3	923480.0
10DPI_interface_4_3	2742000.0
10DPI_interface_4_2	2199900.0
10DPI_interface_4_1	2193900.0
10DPI_interface_3_2	875140.0
10DPI_interface_3_1	1436900.0
10DPI_cortex_4_4	1717800.0
10DPI_cortex_4_3	1395700.0
10DPI_cortex_3_3	880010.0

10DPI_cortex_3_2	708140.0
10DPI_SAC_4_3	1401500.0
10DPI_SAC_4_2	NaN
10DPI_SAC_4_1	1453700.0
10DPI_SAC_3_3	536050.0
10DPI_SAC_3_2	2143600.0
10DPI_SAC_3_1	944670.0
10DPI_SAC_2_3	NaN
10DPI_SAC_2_2	533920.0
10DPI_SAC_2_1	NaN

Protein IDs	A0A075B5S5;A0A0A6YXL5	A0A075B5T3;J3QK03;A0A0A6YWS9	...	\
index			...	
4DPI_interface_2_4	NaN	NaN	...	
4DPI_interface_2_2	NaN	NaN	...	
4DPI_interface_2_1	NaN	NaN	...	
4DPI_interface_1_2	NaN	NaN	...	
4DPI_interface_1_1	NaN	NaN	...	
4DPI_cortex_3_3	NaN	NaN	...	
4DPI_cortex_3_2	NaN	NaN	...	
4DPI_cortex_2_3	NaN	NaN	...	
4DPI_cortex_2_2	NaN	NaN	...	
4DPI_cortex_2_1	NaN	NaN	...	
4DPI_cortex_1_2	NaN	NaN	...	
4DPI_cortex_1_1	NaN	NaN	...	
4DPI_SAC_3_3	NaN	NaN	...	
4DPI_SAC_3_2	NaN	NaN	...	
4DPI_SAC_2_3	NaN	NaN	...	
4DPI_SAC_2_2	NaN	NaN	...	
4DPI_SAC_2_1	NaN	NaN	...	
4DPI_SAC_1_3	NaN	NaN	...	
4DPI_SAC_1_2	NaN	NaN	...	
4DPI_SAC_1_1	NaN	NaN	...	
10DPI_interface_5_3	582000.0	624130.0	...	
10DPI_interface_4_3	10474000.0	1520300.0	...	
10DPI_interface_4_2	4267600.0	3319600.0	...	
10DPI_interface_4_1	1899800.0	1112100.0	...	
10DPI_interface_3_2	521140.0	826650.0	...	
10DPI_interface_3_1	554440.0	NaN	...	
10DPI_cortex_4_4	3546300.0	2818100.0	...	
10DPI_cortex_4_3	NaN	2965700.0	...	
10DPI_cortex_3_3	NaN	992900.0	...	
10DPI_cortex_3_2	739180.0	NaN	...	
10DPI_SAC_4_3	2447600.0	5502600.0	...	
10DPI_SAC_4_2	3972000.0	2063300.0	...	
10DPI_SAC_4_1	8814000.0	2556300.0	...	
10DPI_SAC_3_3	NaN	1867700.0	...	

```
10DPI_SAC_3_2                          NaN                     766080.0  ...
10DPI_SAC_3_1                          NaN                          NaN  ...
10DPI_SAC_2_3                          NaN                          NaN  ...
10DPI_SAC_2_2                          NaN                          NaN  ...
10DPI_SAC_2_1                          NaN                          NaN  ...

 Protein IDs         Q9Z315;A0A494B9E9  \
index
4DPI_interface_2_4           954220.0
4DPI_interface_2_2          1377400.0
4DPI_interface_2_1           922440.0
4DPI_interface_1_2                NaN
4DPI_interface_1_1          1117600.0
4DPI_cortex_3_3              638210.0
4DPI_cortex_3_2              661230.0
4DPI_cortex_2_3                   NaN
4DPI_cortex_2_2              494860.0
4DPI_cortex_2_1              674540.0
4DPI_cortex_1_2                   NaN
4DPI_cortex_1_1              334760.0
4DPI_SAC_3_3                1396800.0
4DPI_SAC_3_2                1242700.0
4DPI_SAC_2_3                1401200.0
4DPI_SAC_2_2                      NaN
4DPI_SAC_2_1                 458240.0
4DPI_SAC_1_3                 780230.0
4DPI_SAC_1_2                1219000.0
4DPI_SAC_1_1                 600830.0
10DPI_interface_5_3         1281000.0
10DPI_interface_4_3         2218300.0
10DPI_interface_4_2         1495700.0
10DPI_interface_4_1          927020.0
10DPI_interface_3_2         1865800.0
10DPI_interface_3_1         1160000.0
10DPI_cortex_4_4             946380.0
10DPI_cortex_4_3                  NaN
10DPI_cortex_3_3             633750.0
10DPI_cortex_3_2             556170.0
10DPI_SAC_4_3               1137800.0
10DPI_SAC_4_2               1302200.0
10DPI_SAC_4_1                     NaN
10DPI_SAC_3_3               1731200.0
10DPI_SAC_3_2               1444700.0
10DPI_SAC_3_1               1059500.0
10DPI_SAC_2_3               3343300.0
10DPI_SAC_2_2                800330.0
10DPI_SAC_2_1               2338200.0
```

Protein IDs	S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5	S4R192
index		
4DPI_interface_2_4	NaN	NaN
4DPI_interface_2_2	NaN	NaN
4DPI_interface_2_1	NaN	NaN
4DPI_interface_1_2	NaN	NaN
4DPI_interface_1_1	NaN	NaN
4DPI_cortex_3_3	NaN	NaN
4DPI_cortex_3_2	NaN	NaN
4DPI_cortex_2_3	662690.0	NaN
4DPI_cortex_2_2	NaN	NaN
4DPI_cortex_2_1	NaN	NaN
4DPI_cortex_1_2	714480.0	NaN
4DPI_cortex_1_1	716000.0	NaN
4DPI_SAC_3_3	407450.0	NaN
4DPI_SAC_3_2	NaN	357930.0
4DPI_SAC_2_3	NaN	NaN
4DPI_SAC_2_2	NaN	NaN
4DPI_SAC_2_1	360500.0	169050.0
4DPI_SAC_1_3	610920.0	NaN
4DPI_SAC_1_2	326340.0	NaN
4DPI_SAC_1_1	NaN	251670.0
10DPI_interface_5_3	126670.0	NaN
10DPI_interface_4_3	NaN	305120.0
10DPI_interface_4_2	NaN	549550.0
10DPI_interface_4_1	NaN	NaN
10DPI_interface_3_2	NaN	364880.0
10DPI_interface_3_1	NaN	781480.0
10DPI_cortex_4_4	NaN	NaN
10DPI_cortex_4_3	NaN	NaN
10DPI_cortex_3_3	NaN	NaN
10DPI_cortex_3_2	NaN	NaN
10DPI_SAC_4_3	NaN	NaN
10DPI_SAC_4_2	NaN	NaN
10DPI_SAC_4_1	NaN	326950.0
10DPI_SAC_3_3	NaN	423330.0
10DPI_SAC_3_2	NaN	NaN
10DPI_SAC_3_1	NaN	NaN
10DPI_SAC_2_3	NaN	NaN
10DPI_SAC_2_2	NaN	NaN
10DPI_SAC_2_1	NaN	NaN

Protein IDs	TimePoint	BioRep	Region	TechRep	DPI_Region_BioRep
index					
4DPI_interface_2_4	4DPI	2	interface	4	4DPI_interface_2
4DPI_interface_2_2	4DPI	2	interface	2	4DPI_interface_2

4DPI_interface_2_1	4DPI	2	interface	1	4DPI_interface_2
4DPI_interface_1_2	4DPI	1	interface	2	4DPI_interface_1
4DPI_interface_1_1	4DPI	1	interface	1	4DPI_interface_1
4DPI_cortex_3_3	4DPI	3	cortex	3	4DPI_cortex_3
4DPI_cortex_3_2	4DPI	3	cortex	2	4DPI_cortex_3
4DPI_cortex_2_3	4DPI	2	cortex	3	4DPI_cortex_2
4DPI_cortex_2_2	4DPI	2	cortex	2	4DPI_cortex_2
4DPI_cortex_2_1	4DPI	2	cortex	1	4DPI_cortex_2
4DPI_cortex_1_2	4DPI	1	cortex	2	4DPI_cortex_1
4DPI_cortex_1_1	4DPI	1	cortex	1	4DPI_cortex_1
4DPI_SAC_3_3	4DPI	3	SAC	3	4DPI_SAC_3
4DPI_SAC_3_2	4DPI	3	SAC	2	4DPI_SAC_3
4DPI_SAC_2_3	4DPI	2	SAC	3	4DPI_SAC_2
4DPI_SAC_2_2	4DPI	2	SAC	2	4DPI_SAC_2
4DPI_SAC_2_1	4DPI	2	SAC	1	4DPI_SAC_2
4DPI_SAC_1_3	4DPI	1	SAC	3	4DPI_SAC_1
4DPI_SAC_1_2	4DPI	1	SAC	2	4DPI_SAC_1
4DPI_SAC_1_1	4DPI	1	SAC	1	4DPI_SAC_1
10DPI_interface_5_3	10DPI	5	interface	3	10DPI_interface_5
10DPI_interface_4_3	10DPI	4	interface	3	10DPI_interface_4
10DPI_interface_4_2	10DPI	4	interface	2	10DPI_interface_4
10DPI_interface_4_1	10DPI	4	interface	1	10DPI_interface_4
10DPI_interface_3_2	10DPI	3	interface	2	10DPI_interface_3
10DPI_interface_3_1	10DPI	3	interface	1	10DPI_interface_3
10DPI_cortex_4_4	10DPI	4	cortex	4	10DPI_cortex_4
10DPI_cortex_4_3	10DPI	4	cortex	3	10DPI_cortex_4
10DPI_cortex_3_3	10DPI	3	cortex	3	10DPI_cortex_3
10DPI_cortex_3_2	10DPI	3	cortex	2	10DPI_cortex_3
10DPI_SAC_4_3	10DPI	4	SAC	3	10DPI_SAC_4
10DPI_SAC_4_2	10DPI	4	SAC	2	10DPI_SAC_4
10DPI_SAC_4_1	10DPI	4	SAC	1	10DPI_SAC_4
10DPI_SAC_3_3	10DPI	3	SAC	3	10DPI_SAC_3
10DPI_SAC_3_2	10DPI	3	SAC	2	10DPI_SAC_3
10DPI_SAC_3_1	10DPI	3	SAC	1	10DPI_SAC_3
10DPI_SAC_2_3	10DPI	2	SAC	3	10DPI_SAC_2
10DPI_SAC_2_2	10DPI	2	SAC	2	10DPI_SAC_2
10DPI_SAC_2_1	10DPI	2	SAC	1	10DPI_SAC_2

Protein IDs	BioRep_TechRep	DPI_Region
index		
4DPI_interface_2_4	2_4	4DPI_interface
4DPI_interface_2_2	2_2	4DPI_interface
4DPI_interface_2_1	2_1	4DPI_interface
4DPI_interface_1_2	1_2	4DPI_interface
4DPI_interface_1_1	1_1	4DPI_interface
4DPI_cortex_3_3	3_3	4DPI_cortex
4DPI_cortex_3_2	3_2	4DPI_cortex

```
4DPI_cortex_2_3                2_3      4DPI_cortex
4DPI_cortex_2_2                2_2      4DPI_cortex
4DPI_cortex_2_1                2_1      4DPI_cortex
4DPI_cortex_1_2                1_2      4DPI_cortex
4DPI_cortex_1_1                1_1      4DPI_cortex
4DPI_SAC_3_3                   3_3         4DPI_SAC
4DPI_SAC_3_2                   3_2         4DPI_SAC
4DPI_SAC_2_3                   2_3         4DPI_SAC
4DPI_SAC_2_2                   2_2         4DPI_SAC
4DPI_SAC_2_1                   2_1         4DPI_SAC
4DPI_SAC_1_3                   1_3         4DPI_SAC
4DPI_SAC_1_2                   1_2         4DPI_SAC
4DPI_SAC_1_1                   1_1         4DPI_SAC
10DPI_interface_5_3            5_3  10DPI_interface
10DPI_interface_4_3            4_3  10DPI_interface
10DPI_interface_4_2            4_2  10DPI_interface
10DPI_interface_4_1            4_1  10DPI_interface
10DPI_interface_3_2            3_2  10DPI_interface
10DPI_interface_3_1            3_1  10DPI_interface
10DPI_cortex_4_4               4_4     10DPI_cortex
10DPI_cortex_4_3               4_3     10DPI_cortex
10DPI_cortex_3_3               3_3     10DPI_cortex
10DPI_cortex_3_2               3_2     10DPI_cortex
10DPI_SAC_4_3                  4_3        10DPI_SAC
10DPI_SAC_4_2                  4_2        10DPI_SAC
10DPI_SAC_4_1                  4_1        10DPI_SAC
10DPI_SAC_3_3                  3_3        10DPI_SAC
10DPI_SAC_3_2                  3_2        10DPI_SAC
10DPI_SAC_3_1                  3_1        10DPI_SAC
10DPI_SAC_2_3                  2_3        10DPI_SAC
10DPI_SAC_2_2                  2_2        10DPI_SAC
10DPI_SAC_2_1                  2_1        10DPI_SAC

[39 rows x 3620 columns]
```

[51]:
```
combined_lfq = combined_lfq.dropna(axis=1)
combined_lfq.shape
```

[51]:
```
(39, 294)
```

0.2 Combined analysis

[52]:
```
combined_lfq.head()
```

[52]:
```
 Protein IDs        A0A0J9YTY0;A0A0J9YUL3;Q8C1B7;A0A0J9YUV6;A0A0J9YVA6  \
index
4DPI_interface_2_4                                                8884400.0
4DPI_interface_2_2                                                4805100.0
```

```
4DPI_interface_2_1                                          4679400.0
4DPI_interface_1_2                                          2752000.0
4DPI_interface_1_1                                          6446300.0

 Protein IDs        A0A0J9YUZ4;P63158;A0A0J9YUD8;D3YVC6;D3YZ18  \
index
4DPI_interface_2_4                                  13795000.0
4DPI_interface_2_2                                  16043000.0
4DPI_interface_2_1                                  14321000.0
4DPI_interface_1_2                                   2559000.0
4DPI_interface_1_1                                  18740000.0

 Protein IDs        A0A0R4J093;Q9DBP5  A0A0R4J0I1;P07759;E9Q499;Q80X76  \
index
4DPI_interface_2_4         5749000.0                     73676000.0
4DPI_interface_2_2        10405000.0                     50598000.0
4DPI_interface_2_1         9228800.0                    105190000.0
4DPI_interface_1_2         6292300.0                     22129000.0
4DPI_interface_1_1        18973000.0                     38062000.0

 Protein IDs        O70404;A0A0R4J0R1;A0A0U1RPE8  A0A0R4J0Z1;P08003  \
index
4DPI_interface_2_4                    4863000.0         20661000.0
4DPI_interface_2_2                    5048700.0         20709000.0
4DPI_interface_2_1                    4421600.0         17844000.0
4DPI_interface_1_2                    1316400.0         22085000.0
4DPI_interface_1_1                    2682100.0         24030000.0

 Protein IDs
E9QL31;A0A0R4J104;P98078;E9PX84;Q3TRE6;F6TQN9;Q9DCE6;E0CXT5;E0CZ53;E0CYJ2  \
index
4DPI_interface_2_4                                         10187000.0
4DPI_interface_2_2                                         10975000.0
4DPI_interface_2_1                                          8142400.0
4DPI_interface_1_2                                           704720.0
4DPI_interface_1_1                                          4963900.0

 Protein IDs        A0A0R4J195;Q9CSU0;F6YB25;F7DAY5;Q8VDS4  \
index
4DPI_interface_2_4                                 1394900.0
4DPI_interface_2_2                                  959390.0
4DPI_interface_2_1                                  549190.0
4DPI_interface_1_2                                 1506600.0
4DPI_interface_1_1                                 1400400.0

 Protein IDs
A0A1B0GQU8;P35980;A0A1B0GSS8;A0A1B0GSF7;A0A1B0GSA8;A0A1B0GS28;A0A1B0GRZ3  \
```

```
index
4DPI_interface_2_4                                          11981000.0
4DPI_interface_2_2                                          22897000.0
4DPI_interface_2_1                                          20121000.0
4DPI_interface_1_2                                          36610000.0
4DPI_interface_1_1                                          23111000.0

 Protein IDs        P06151;Q564E2;A0A1B0GSR9;A0A1B0GSX0;D3YZQ9;A0A1B0GSL7;A0A1B0
GT41;A0A1B0GQX5;A0A1B0GS79;A0A1B0GRW9;A0A1B0GRC1;A0A1B0GRS2;D3YVR7;D3YZE4;A0A1B0
GRE9;A0A1B0GSR2;P00342  \
index
4DPI_interface_2_4                                          48765000.0
4DPI_interface_2_2                                          84281000.0
4DPI_interface_2_1                                          62405000.0
4DPI_interface_1_2                                          61114000.0
4DPI_interface_1_1                                          36589000.0

 Protein IDs        ...      Q9Z1Q5     Q9Z2D6  Q9Z2I8;A0A0N4SWD1  TimePoint  \
index               ...
4DPI_interface_2_4  ...  15999000.0  3568300.0         12254000.0       4DPI
4DPI_interface_2_2  ...  21039000.0  1709200.0         17425000.0       4DPI
4DPI_interface_2_1  ...  27600000.0  1083700.0         14862000.0       4DPI
4DPI_interface_1_2  ...  22611000.0  2757000.0          7584200.0       4DPI
4DPI_interface_1_1  ...  23281000.0  5819700.0         16155000.0       4DPI

 Protein IDs        BioRep     Region  TechRep  DPI_Region_BioRep  \
index
4DPI_interface_2_4       2  interface        4   4DPI_interface_2
4DPI_interface_2_2       2  interface        2   4DPI_interface_2
4DPI_interface_2_1       2  interface        1   4DPI_interface_2
4DPI_interface_1_2       1  interface        2   4DPI_interface_1
4DPI_interface_1_1       1  interface        1   4DPI_interface_1

 Protein IDs        BioRep_TechRep      DPI_Region
index
4DPI_interface_2_4             2_4  4DPI_interface
4DPI_interface_2_2             2_2  4DPI_interface
4DPI_interface_2_1             2_1  4DPI_interface
4DPI_interface_1_2             1_2  4DPI_interface
4DPI_interface_1_1             1_1  4DPI_interface

[5 rows x 294 columns]
```

[60]:
```
### Function to separate data and labels
def df_sep(label, dataframe):
    X = dataframe.drop(['TimePoint', 'TechRep', 'BioRep', 'Region',␣
 ↪'DPI_Region_BioRep', 'DPI_Region', 'BioRep_TechRep'], axis=1)
```

```
    y = dataframe[label]
    return X,y
```

[61]:
```
X,Y = df_sep('DPI_Region', combined_lfq)
Y
y = pd.factorize(Y)[0]
```

[62]:
```
X.to_csv('/Users/kavyasharman/Documents/Projects/microLESA/outs/' + date +␣
↪'_combined_X.csv')
```

0.3 PCA

[63]:
```
pca = PCA()
pca.fit_transform(X)
```

[63]:
```
array([[-1.28844218e+09, -7.11460826e+08,  6.31811418e+08, ...,
         1.03164582e+07, -1.17846874e+07,  2.92942842e-07],
       [ 5.28450418e+08, -5.99037940e+08,  1.94400410e+08, ...,
         2.28226796e+07, -5.41262642e+04,  2.92942842e-07],
       [ 2.67146200e+08, -1.33471004e+09, -8.64781484e+06, ...,
        -2.71817551e+07,  2.19963656e+06,  2.92942842e-07],
       ...,
       [-2.75694213e+09, -1.98017577e+08,  4.88713329e+08, ...,
        -5.89531193e+06, -4.21840665e+06,  2.92942842e-07],
       [-2.09023427e+09, -1.40884868e+08,  1.02851665e+09, ...,
        -3.57478427e+06,  3.89483396e+06,  2.92942842e-07],
       [-2.01507561e+09, -1.92089267e+08, -3.76911129e+08, ...,
         9.41837935e+06, -6.70903517e+06,  2.92942842e-07]])
```

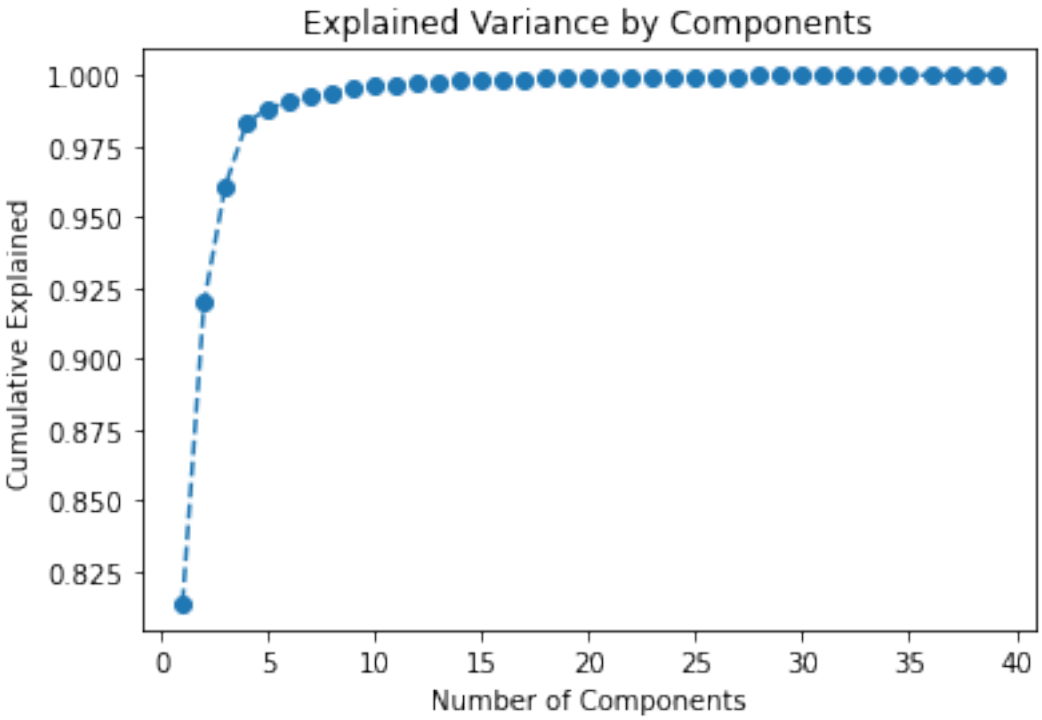
[64]:
```
components = pca.components_
components.shape
```

[64]:
```
(39, 287)
```

[65]:
```
X.shape
```

[65]:
```
(39, 287)
```

[66]:
```
plt.figure()
plt.plot(range(1,40), pca.explained_variance_ratio_.cumsum(), marker = 'o',␣
↪linestyle = '--')
plt.title('Explained Variance by Components')
plt.xlabel('Number of Components')
plt.ylabel('Cumulative Explained')
plt.savefig('outs/' + date + '_combined_PCA_explainedVariance.svg')
```



[67]:
```
pca_expl_var = pca.explained_variance_ratio_
pca_expl_var
```

[67]:
```
array([8.13496140e-01, 1.06088298e-01, 4.13704385e-02, 2.22253702e-02,
       5.13907045e-03, 2.23608178e-03, 1.84227069e-03, 1.69922876e-03,
       1.17444745e-03, 8.60060940e-04, 7.28495022e-04, 4.43397100e-04,
       3.70086472e-04, 3.25457317e-04, 3.03346162e-04, 2.37529489e-04,
       2.18973177e-04, 1.92780927e-04, 1.65274961e-04, 1.14038533e-04,
       9.70533536e-05, 9.03254288e-05, 7.95687166e-05, 7.23250045e-05,
       5.98391318e-05, 5.79711467e-05, 5.17241307e-05, 4.12563898e-05,
       3.54151933e-05, 3.43962530e-05, 2.82922579e-05, 2.28373465e-05,
       2.19429014e-05, 1.94594840e-05, 1.89470482e-05, 1.44635243e-05,
       1.38293277e-05, 9.56707717e-06, 5.19960053e-33])
```

[68]:
```
scores_pca = pca.fit_transform(X)

#pca loadings
loadings = pd.DataFrame(pca.components_.T, index=X.columns)
```

[69]:
```
scores_pca.shape
```

[69]:
```
(39, 39)
```

[70]:
```
pd.DataFrame(scores_pca).to_csv('/Users/kavyasharman/Documents/Projects/
 ↪microLESA/outs/' + date + '_combined_scores_pca.csv')
```

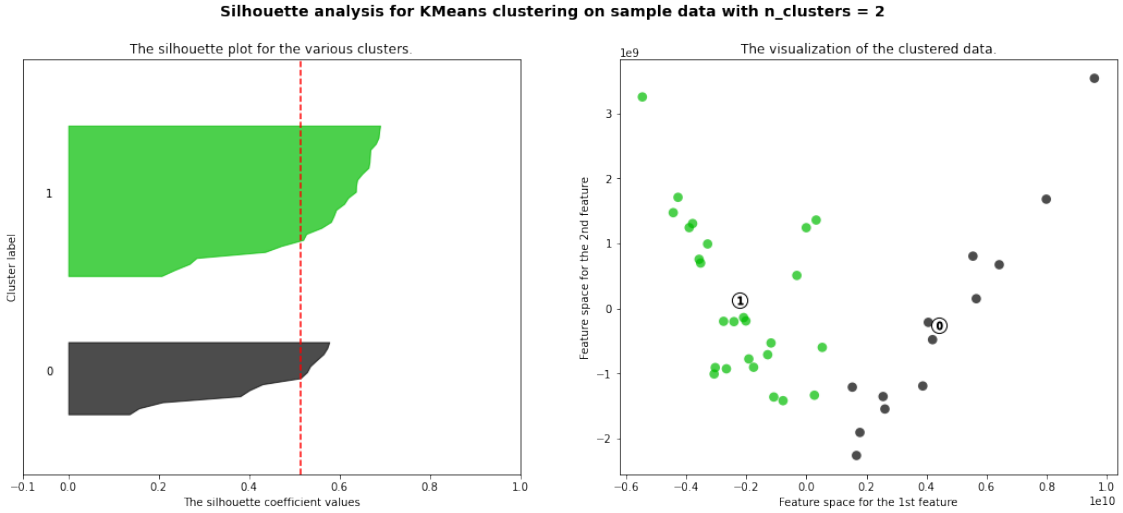
[71]:
```
loadings.to_csv('/Users/kavyasharman/Documents/Projects/microLESA/outs/' + date␣
↪+ '_combined_loadings_pca.csv')
```

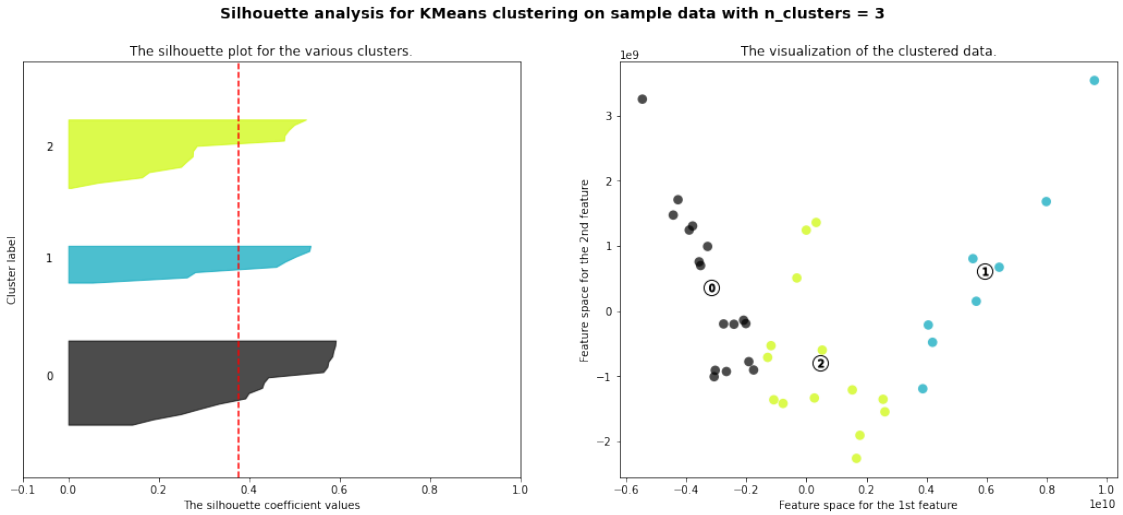
0.3.1 Silhouette scores to determine k

[72]:
```
from sklearn.datasets import make_blobs
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_samples, silhouette_score

import matplotlib.pyplot as plt
import matplotlib.cm as cm
import numpy as np

range_n_clusters = [2, 3, 4, 5, 6, 7, 8, 9, 10,11,12,13,14,15]
silhouette_scores_euc = pd.DataFrame(columns = ['k', 'score'])

for n_clusters in range_n_clusters:
    # Create a subplot with 1 row and 2 columns
    fig, (ax1, ax2) = plt.subplots(1, 2)
    fig.set_size_inches(18, 7)

    # The 1st subplot is the silhouette plot
    # The silhouette coefficient can range from -1, 1 but in this example all
    # lie within [-0.1, 1]
    ax1.set_xlim([-0.1, 1])
    # The (n_clusters+1)*10 is for inserting blank space between silhouette
    # plots of individual clusters, to demarcate them clearly.
    ax1.set_ylim([0, len(scores_pca) + (n_clusters + 1) * 10])


    # Initialize the clusterer with n_clusters value and a random generator
    # seed of 10 for reproducibility.
    clusterer = KMeans(n_clusters=n_clusters, random_state=42)
    cluster_labels = clusterer.fit_predict(scores_pca)

    # The silhouette_score gives the average value for all the samples.
    # This gives a perspective into the density and separation of the formed
    # clusters
    silhouette_avg = silhouette_score(scores_pca, cluster_labels)
  #  print("For n_clusters =", n_clusters,
  #        "The average silhouette_score is :", silhouette_avg)
    silhouette_scores_euc = silhouette_scores_euc.append(pd.DataFrame({'k':␣
↪n_clusters, 'score': silhouette_avg}, index=[0]), ignore_index=True)

    # Compute the silhouette scores for each sample
    sample_silhouette_values = silhouette_samples(scores_pca, cluster_labels)
```

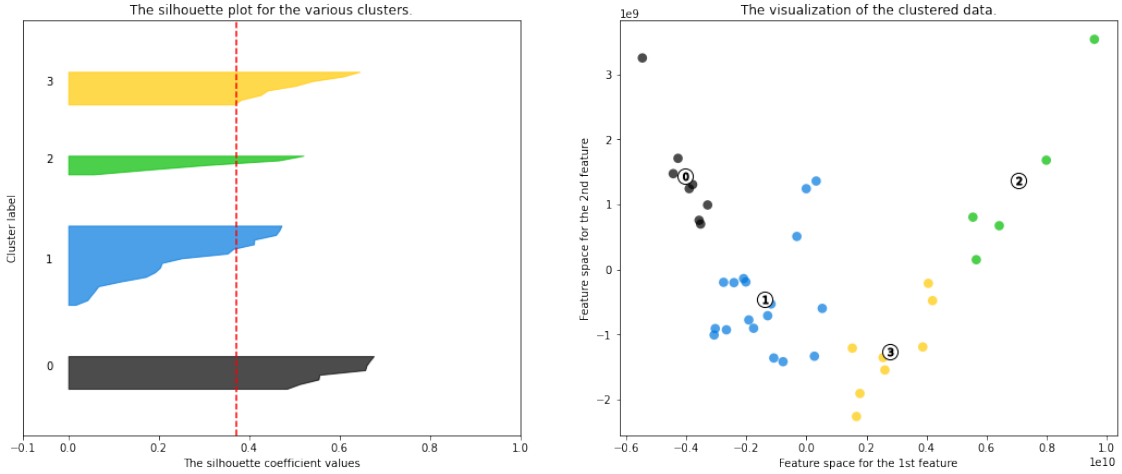
```
    y_lower = 10
    for i in range(n_clusters):
        # Aggregate the silhouette scores for samples belonging to
        # cluster i, and sort them
        ith_cluster_silhouette_values = \
            sample_silhouette_values[cluster_labels == i]

        ith_cluster_silhouette_values.sort()

        size_cluster_i = ith_cluster_silhouette_values.shape[0]
        y_upper = y_lower + size_cluster_i

        color = cm.nipy_spectral(float(i) / n_clusters)
        ax1.fill_betweenx(np.arange(y_lower, y_upper),
                          0, ith_cluster_silhouette_values,
                          facecolor=color, edgecolor=color, alpha=0.7)

        # Label the silhouette plots with their cluster numbers at the middle
        ax1.text(-0.05, y_lower + 0.5 * size_cluster_i, str(i))

        # Compute the new y_lower for next plot
        y_lower = y_upper + 10  # 10 for the 0 samples

    ax1.set_title("The silhouette plot for the various clusters.")
    ax1.set_xlabel("The silhouette coefficient values")
    ax1.set_ylabel("Cluster label")

    # The vertical line for average silhouette score of all the values
    ax1.axvline(x=silhouette_avg, color="red", linestyle="--")

    ax1.set_yticks([])  # Clear the yaxis labels / ticks
    ax1.set_xticks([-0.1, 0, 0.2, 0.4, 0.6, 0.8, 1])

    # 2nd Plot showing the actual clusters formed
    colors = cm.nipy_spectral(cluster_labels.astype(float) / n_clusters)
    ax2.scatter(scores_pca[:, 0], scores_pca[:, 1], marker='.', s=300, lw=0,␣
↪alpha=0.7,
                c=colors, edgecolor='k')

    # Labeling the clusters
    centers = clusterer.cluster_centers_
    # Draw white circles at cluster centers
    ax2.scatter(centers[:, 0], centers[:, 1], marker='o',
                c="white", alpha=1, s=200, edgecolor='k')

    for i, c in enumerate(centers):
        ax2.scatter(c[0], c[1], marker='$%d$' % i, alpha=1,
```

```
                    s=50, edgecolor='k')

    ax2.set_title("The visualization of the clustered data.")
    ax2.set_xlabel("Feature space for the 1st feature")
    ax2.set_ylabel("Feature space for the 2nd feature")

    plt.suptitle(("Silhouette analysis for KMeans clustering on sample data "
                  "with n_clusters = %d" % n_clusters),
                 fontsize=14, fontweight='bold')

    plt.savefig('outs/' + date + '_combined_silhouette_scores_euc_' +␣
↪str(n_clusters) + '.svg')

silhouette_scores_euc.to_csv('outs/' + date + '_combined_silhouette_scores_euc.
↪csv')
```

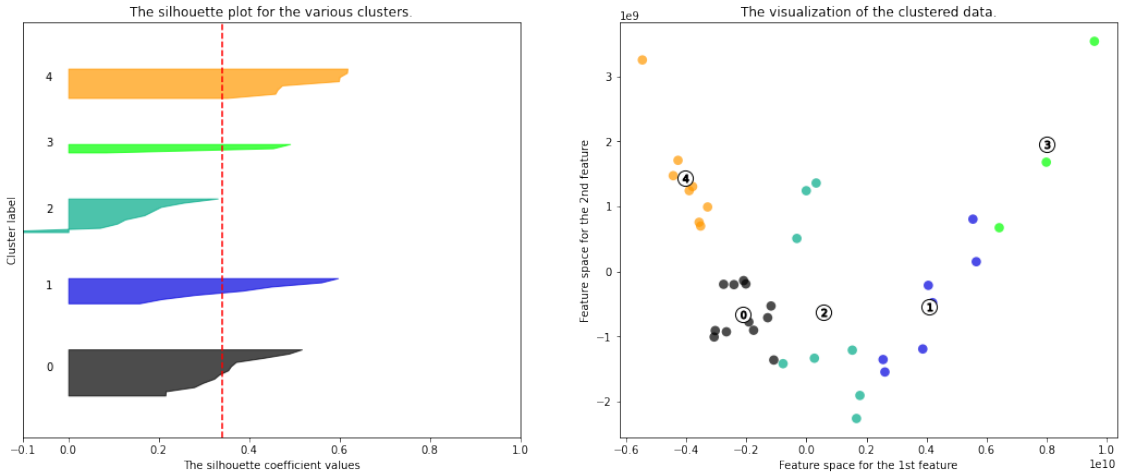




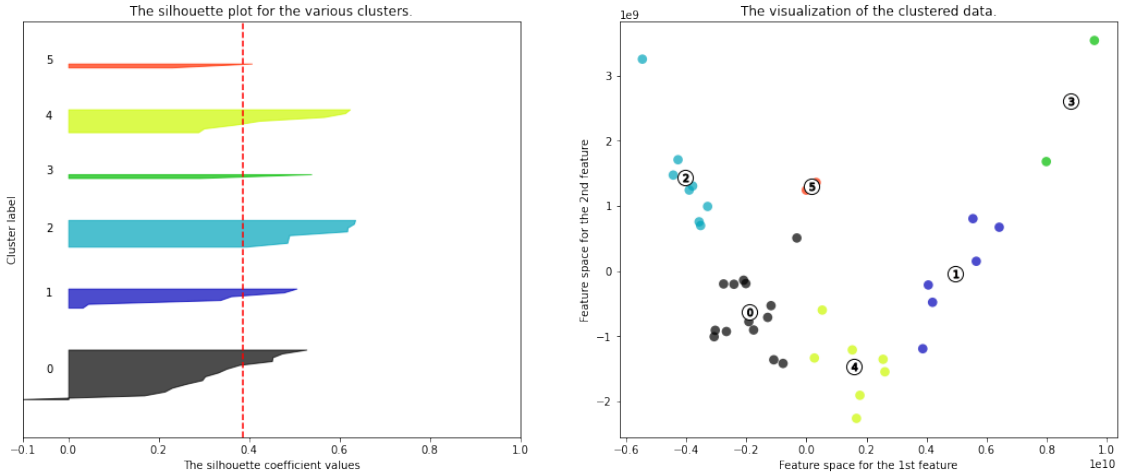
Silhouette analysis for KMeans clustering on sample data with n_clusters = 4



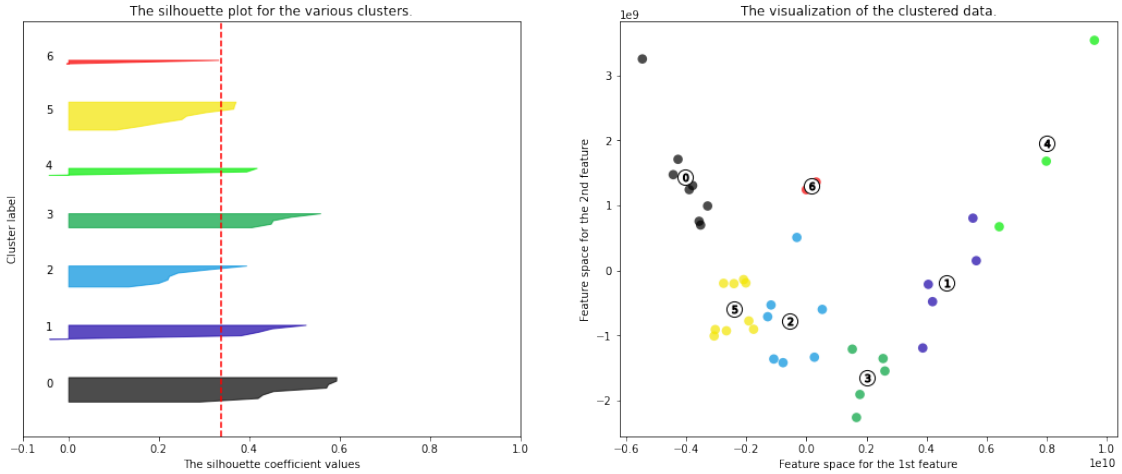
Silhouette analysis for KMeans clustering on sample data with n_clusters = 5



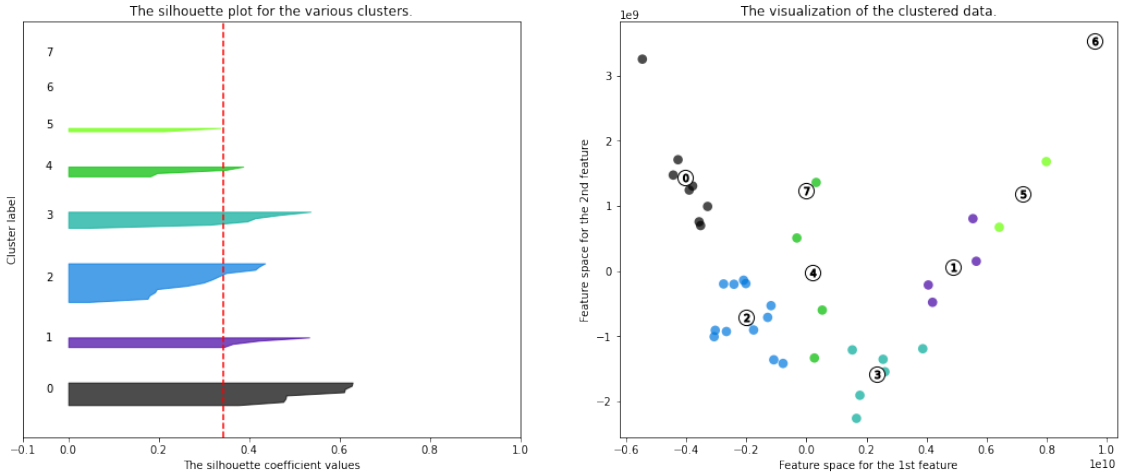
Silhouette analysis for KMeans clustering on sample data with n_clusters = 6



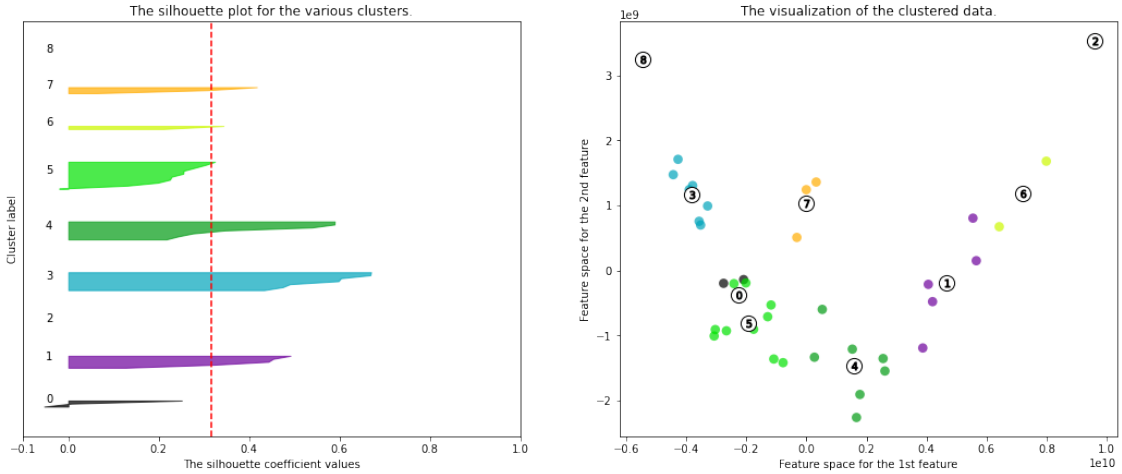
Silhouette analysis for KMeans clustering on sample data with n_clusters = 7



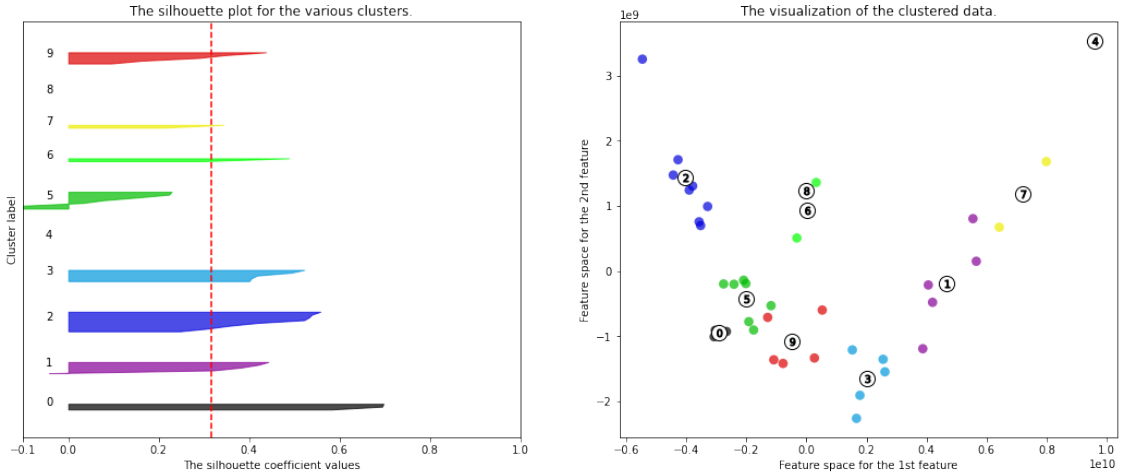
Silhouette analysis for KMeans clustering on sample data with n_clusters = 8



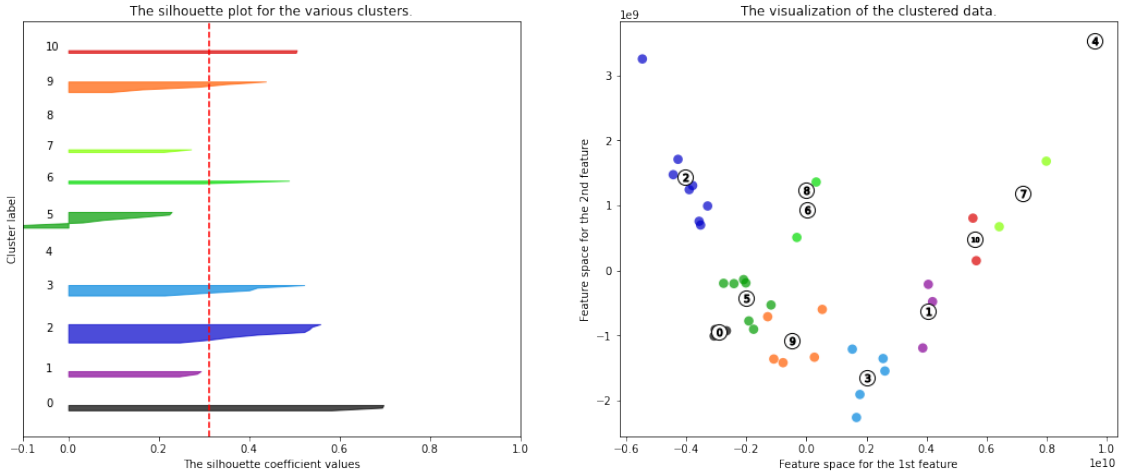
Silhouette analysis for KMeans clustering on sample data with n_clusters = 9



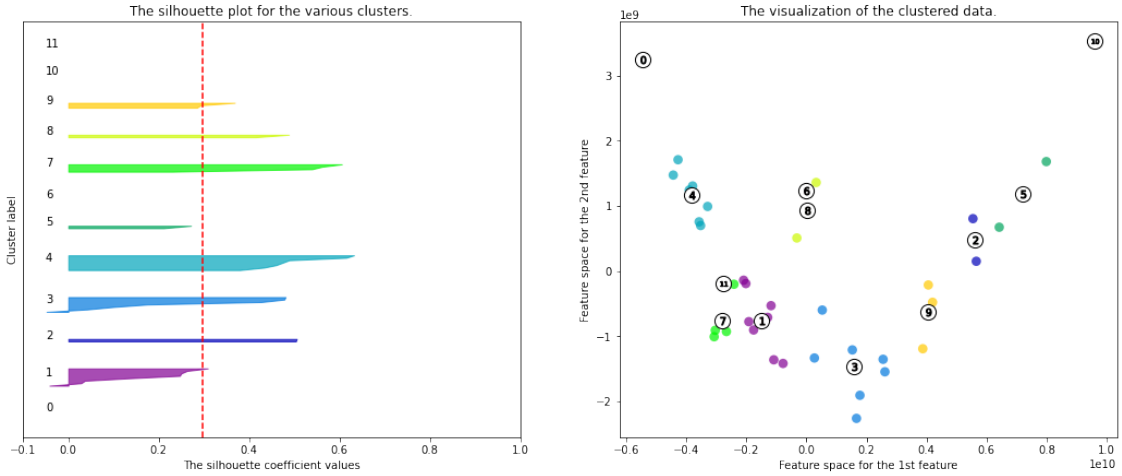
Silhouette analysis for KMeans clustering on sample data with n_clusters = 10



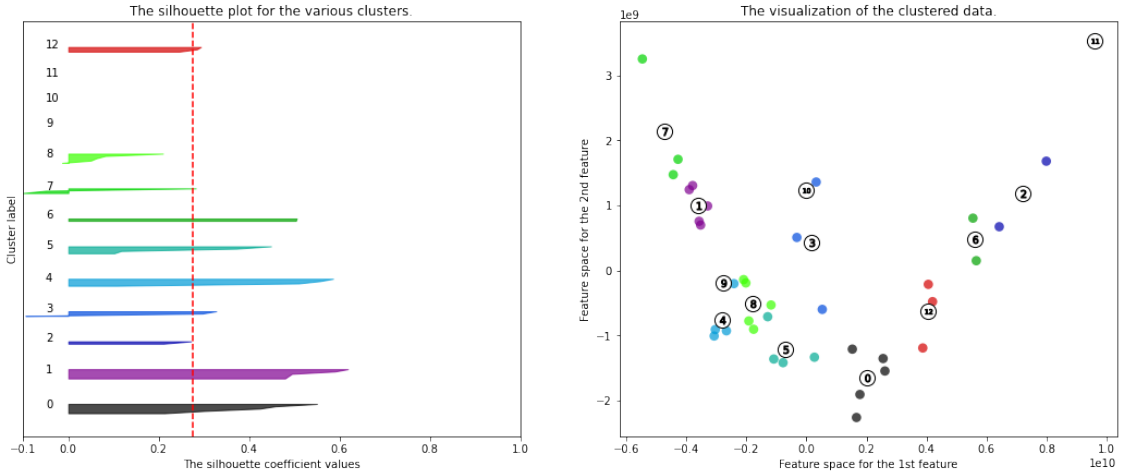
Silhouette analysis for KMeans clustering on sample data with n_clusters = 11

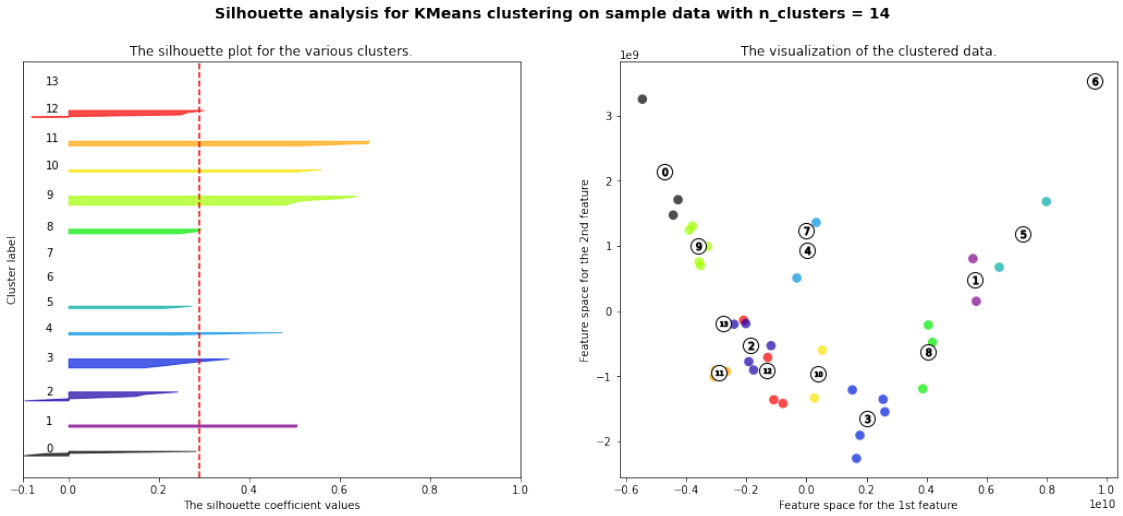


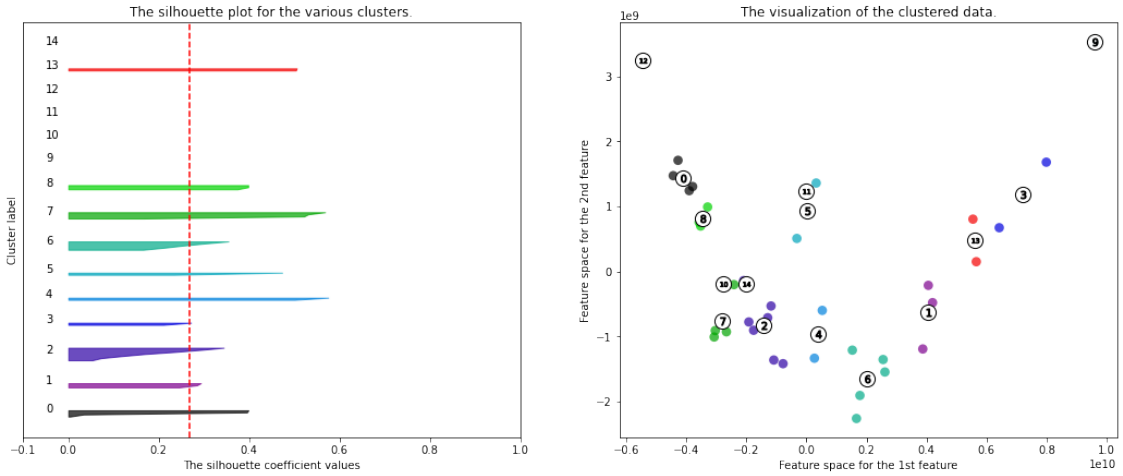
Silhouette analysis for KMeans clustering on sample data with n_clusters = 12



Silhouette analysis for KMeans clustering on sample data with n_clusters = 13







[73]:
```
silhouette_scores_euc
```

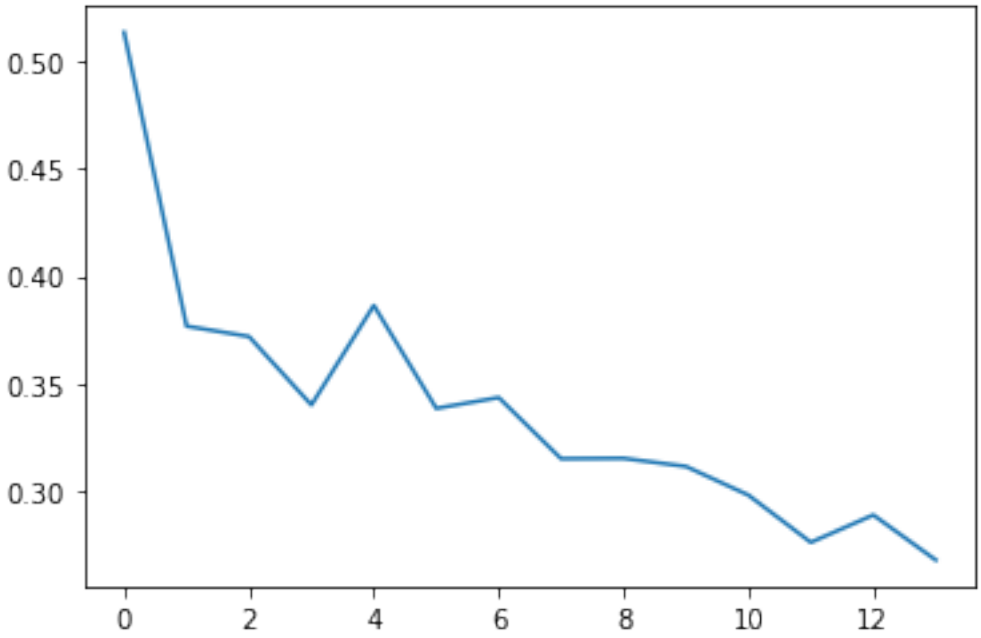
[73]:

	k	score
0	2	0.513197
1	3	0.376735
2	4	0.371904
3	5	0.340121
4	6	0.386350
5	7	0.338519
6	8	0.343562
7	9	0.315147
8	10	0.315310
9	11	0.311548

```
10  12  0.298222
11  13  0.276279
12  14  0.289012
13  15  0.268083
```

[74]:
```
plt.plot(silhouette_scores_euc['score'])
```

[74]:
```
[<matplotlib.lines.Line2D at 0x7fdc0f199940>]
```



0.3.2 k=4

[77]:
```
kmeans_pca = KMeans(n_clusters=4, init='k-means++', random_state=42)
kmeans_pca.fit(scores_pca)
kmeans_labels = kmeans_pca.labels_
```

[78]:
```
kmeans_labels_df = pd.DataFrame(kmeans_labels)
```

[79]:
```
cluster_centers = kmeans_pca.cluster_centers_
cluster_centers_df = pd.DataFrame(cluster_centers).to_csv('outs/' + date +␣
 ↪'_combined_kmeans_clusterCenters.csv')
```

[81]:
```
cluster_centers = kmeans_pca.cluster_centers_
df_segm_pca_kmeans = pd.concat([X.reset_index(drop=True), pd.
 ↪DataFrame(scores_pca)], axis=1)
df_segm_pca_kmeans.columns.values[-34]
df_segm_pca_kmeans.columns.values[-39:-34 ] = ['Component 1', 'Component 2',␣
 ↪'Component 3', 'Component 4', 'Component 5']
df_segm_pca_kmeans['Segment K-means PCA'] = kmeans_pca.labels_
```

```
df_segm_pca_kmeans.index = X.index
df_segm_pca_kmeans[['TimePoint', 'BioRep', 'TechRep', 'Region', 'DPI_Region',␣
 ↪'BioRep_TechRep']] = combined_lfq[['TimePoint', 'BioRep', 'TechRep',␣
 ↪'Region', 'DPI_Region', 'BioRep_TechRep']]
df_segm_pca_kmeans.to_csv('outs/' + date + '_combined_df_segm_pca_kmeans.csv')
df_segm_pca_kmeans
```

[81]:

index	A0A0J9YTY0;A0A0J9YUL3;Q8C1B7;A0A0J9YUV6;A0A0J9YVA6 \
4DPI_interface_2_4	8884400.0
4DPI_interface_2_2	4805100.0
4DPI_interface_2_1	4679400.0
4DPI_interface_1_2	2752000.0
4DPI_interface_1_1	6446300.0
4DPI_cortex_3_3	3966100.0
4DPI_cortex_3_2	3174000.0
4DPI_cortex_2_3	3505600.0
4DPI_cortex_2_2	3623400.0
4DPI_cortex_2_1	5408800.0
4DPI_cortex_1_2	3500500.0
4DPI_cortex_1_1	4027400.0
4DPI_SAC_3_3	10933000.0
4DPI_SAC_3_2	9909200.0
4DPI_SAC_2_3	7513200.0
4DPI_SAC_2_2	5452600.0
4DPI_SAC_2_1	11385000.0
4DPI_SAC_1_3	8059300.0
4DPI_SAC_1_2	5600200.0
4DPI_SAC_1_1	6698000.0
10DPI_interface_5_3	14004000.0
10DPI_interface_4_3	11083000.0
10DPI_interface_4_2	10662000.0
10DPI_interface_4_1	10168000.0
10DPI_interface_3_2	10654000.0
10DPI_interface_3_1	13143000.0
10DPI_cortex_4_4	4613200.0
10DPI_cortex_4_3	7330400.0
10DPI_cortex_3_3	2525300.0
10DPI_cortex_3_2	4286400.0
10DPI_SAC_4_3	9362000.0
10DPI_SAC_4_2	10750000.0
10DPI_SAC_4_1	10526000.0
10DPI_SAC_3_3	22060000.0
10DPI_SAC_3_2	16201000.0
10DPI_SAC_3_1	11879000.0
10DPI_SAC_2_3	10519000.0
10DPI_SAC_2_2	11053000.0

```
10DPI_SAC_2_1                                          12673000.0

                    A0A0J9YUZ4;P63158;A0A0J9YUD8;D3YVC6;D3YZ18  \
index
4DPI_interface_2_4                                   13795000.0
4DPI_interface_2_2                                   16043000.0
4DPI_interface_2_1                                   14321000.0
4DPI_interface_1_2                                    2559000.0
4DPI_interface_1_1                                   18740000.0
4DPI_cortex_3_3                                      10538000.0
4DPI_cortex_3_2                                       5800300.0
4DPI_cortex_2_3                                       7019000.0
4DPI_cortex_2_2                                       9811200.0
4DPI_cortex_2_1                                       7080900.0
4DPI_cortex_1_2                                      13352000.0
4DPI_cortex_1_1                                      15997000.0
4DPI_SAC_3_3                                         24605000.0
4DPI_SAC_3_2                                         18164000.0
4DPI_SAC_2_3                                         11153000.0
4DPI_SAC_2_2                                          7476200.0
4DPI_SAC_2_1                                         24354000.0
4DPI_SAC_1_3                                         15659000.0
4DPI_SAC_1_2                                         18231000.0
4DPI_SAC_1_1                                         17122000.0
10DPI_interface_5_3                                  19714000.0
10DPI_interface_4_3                                  41455000.0
10DPI_interface_4_2                                  18782000.0
10DPI_interface_4_1                                  26015000.0
10DPI_interface_3_2                                  28269000.0
10DPI_interface_3_1                                  18962000.0
10DPI_cortex_4_4                                      6571500.0
10DPI_cortex_4_3                                     12236000.0
10DPI_cortex_3_3                                      2883700.0
10DPI_cortex_3_2                                      3060800.0
10DPI_SAC_4_3                                        19296000.0
10DPI_SAC_4_2                                        26852000.0
10DPI_SAC_4_1                                        17329000.0
10DPI_SAC_3_3                                        20678000.0
10DPI_SAC_3_2                                        18474000.0
10DPI_SAC_3_1                                        22727000.0
10DPI_SAC_2_3                                        21473000.0
10DPI_SAC_2_2                                        33776000.0
10DPI_SAC_2_1                                        13833000.0

                    A0A0R4J093;Q9DBP5  A0A0R4J0I1;P07759;E9Q499;Q80X76  \
index
4DPI_interface_2_4          5749000.0                      73676000.0
```

```
4DPI_interface_2_2          10405000.0                    50598000.0
4DPI_interface_2_1           9228800.0                   105190000.0
4DPI_interface_1_2           6292300.0                    22129000.0
4DPI_interface_1_1          18973000.0                    38062000.0
4DPI_cortex_3_3             14700000.0                    32753000.0
4DPI_cortex_3_2             12382000.0                    24879000.0
4DPI_cortex_2_3             15484000.0                    63258000.0
4DPI_cortex_2_2             11188000.0                    47764000.0
4DPI_cortex_2_1              7899100.0                    44351000.0
4DPI_cortex_1_2             15600000.0                    60586000.0
4DPI_cortex_1_1             14481000.0                    84161000.0
4DPI_SAC_3_3                10172000.0                    33907000.0
4DPI_SAC_3_2                11552000.0                    28193000.0
4DPI_SAC_2_3                13766000.0                    56852000.0
4DPI_SAC_2_2                 9574300.0                    37292000.0
4DPI_SAC_2_1                12936000.0                    60285000.0
4DPI_SAC_1_3                 6236600.0                    79677000.0
4DPI_SAC_1_2                 4541100.0                    63271000.0
4DPI_SAC_1_1                 5351400.0                    53311000.0
10DPI_interface_5_3         12167000.0                    16442000.0
10DPI_interface_4_3         11927000.0                    20993000.0
10DPI_interface_4_2         12228000.0                    37156000.0
10DPI_interface_4_1          9103100.0                    37811000.0
10DPI_interface_3_2         12785000.0                    17304000.0
10DPI_interface_3_1         13679000.0                    23913000.0
10DPI_cortex_4_4            15905000.0                    21923000.0
10DPI_cortex_4_3             3591100.0                    28778000.0
10DPI_cortex_3_3             2846000.0                    17751000.0
10DPI_cortex_3_2             8662600.0                    16765000.0
10DPI_SAC_4_3               12920000.0                    26820000.0
10DPI_SAC_4_2               12945000.0                    18110000.0
10DPI_SAC_4_1               11294000.0                    17675000.0
10DPI_SAC_3_3               11293000.0                    17459000.0
10DPI_SAC_3_2               10115000.0                    16184000.0
10DPI_SAC_3_1                9660300.0                    20254000.0
10DPI_SAC_2_3                7141400.0                    11286000.0
10DPI_SAC_2_2               11842000.0                    14575000.0
10DPI_SAC_2_1                8280200.0                    18280000.0

                    O70404;A0A0R4J0R1;A0A0U1RPE8  A0A0R4J0Z1;P08003  \
index
4DPI_interface_2_4                     4863000.0        20661000.0
4DPI_interface_2_2                     5048700.0        20709000.0
4DPI_interface_2_1                     4421600.0        17844000.0
4DPI_interface_1_2                     1316400.0        22085000.0
4DPI_interface_1_1                     2682100.0        24030000.0
4DPI_cortex_3_3                        2543300.0         7914900.0
```

```
4DPI_cortex_3_2                          2405800.0         10756000.0
4DPI_cortex_2_3                          3551100.0          9577000.0
4DPI_cortex_2_2                          3017800.0         14510000.0
4DPI_cortex_2_1                          3218600.0          9714800.0
4DPI_cortex_1_2                          4396200.0          5588800.0
4DPI_cortex_1_1                           559210.0         12618000.0
4DPI_SAC_3_3                             6073100.0         24566000.0
4DPI_SAC_3_2                             5000700.0         20284000.0
4DPI_SAC_2_3                             7744500.0         14750000.0
4DPI_SAC_2_2                             3349900.0         14702000.0
4DPI_SAC_2_1                             7488800.0         22398000.0
4DPI_SAC_1_3                             4729600.0         22668000.0
4DPI_SAC_1_2                             2086200.0         16305000.0
4DPI_SAC_1_1                             3328200.0         20802000.0
10DPI_interface_5_3                      7822400.0         28551000.0
10DPI_interface_4_3                      5549700.0         25417000.0
10DPI_interface_4_2                      3494600.0         20562000.0
10DPI_interface_4_1                      3959500.0         14898000.0
10DPI_interface_3_2                     10313000.0         21001000.0
10DPI_interface_3_1                      5659300.0         17979000.0
10DPI_cortex_4_4                         6568100.0         16769000.0
10DPI_cortex_4_3                         3419900.0         11998000.0
10DPI_cortex_3_3                         2486500.0         10442000.0
10DPI_cortex_3_2                         3308800.0          5336900.0
10DPI_SAC_4_3                            6796900.0         23906000.0
10DPI_SAC_4_2                            7350200.0         19011000.0
10DPI_SAC_4_1                            4744200.0         19508000.0
10DPI_SAC_3_3                           12653000.0         20440000.0
10DPI_SAC_3_2                           11506000.0         19529000.0
10DPI_SAC_3_1                            9135600.0         22306000.0
10DPI_SAC_2_3                           12044000.0         13540000.0
10DPI_SAC_2_2                           11489000.0         15259000.0
10DPI_SAC_2_1                            6878200.0         18782000.0

E9QL31;A0A0R4J104;P98078;E9PX84;Q3TRE6;F6TQN9;Q9DCE6;E0CXT5;E0CZ53;E0CYJ2  \
index
4DPI_interface_2_4                                        10187000.0
4DPI_interface_2_2                                        10975000.0
4DPI_interface_2_1                                         8142400.0
4DPI_interface_1_2                                          704720.0
4DPI_interface_1_1                                         4963900.0
4DPI_cortex_3_3                                           26064000.0
4DPI_cortex_3_2                                           19354000.0
4DPI_cortex_2_3                                           18882000.0
4DPI_cortex_2_2                                           14360000.0
4DPI_cortex_2_1                                           29494000.0
4DPI_cortex_1_2                                            8878300.0
```

4DPI_cortex_1_1	10476000.0
4DPI_SAC_3_3	7031500.0
4DPI_SAC_3_2	5700800.0
4DPI_SAC_2_3	5403700.0
4DPI_SAC_2_2	4984100.0
4DPI_SAC_2_1	7089200.0
4DPI_SAC_1_3	9872000.0
4DPI_SAC_1_2	7354500.0
4DPI_SAC_1_1	9433100.0
10DPI_interface_5_3	8816700.0
10DPI_interface_4_3	6936800.0
10DPI_interface_4_2	3421900.0
10DPI_interface_4_1	3713000.0
10DPI_interface_3_2	5232300.0
10DPI_interface_3_1	6887000.0
10DPI_cortex_4_4	7816500.0
10DPI_cortex_4_3	14825000.0
10DPI_cortex_3_3	22599000.0
10DPI_cortex_3_2	25867000.0
10DPI_SAC_4_3	5934200.0
10DPI_SAC_4_2	2999000.0
10DPI_SAC_4_1	4308500.0
10DPI_SAC_3_3	7713200.0
10DPI_SAC_3_2	9466900.0
10DPI_SAC_3_1	8250600.0
10DPI_SAC_2_3	6240900.0
10DPI_SAC_2_2	3403200.0
10DPI_SAC_2_1	8101900.0

index	A0A0R4J195;Q9CSU0;F6YB25;F7DAY5;Q8VDS4 \
4DPI_interface_2_4	1394900.0
4DPI_interface_2_2	959390.0
4DPI_interface_2_1	549190.0
4DPI_interface_1_2	1506600.0
4DPI_interface_1_1	1400400.0
4DPI_cortex_3_3	734590.0
4DPI_cortex_3_2	544140.0
4DPI_cortex_2_3	1263200.0
4DPI_cortex_2_2	911490.0
4DPI_cortex_2_1	1163700.0
4DPI_cortex_1_2	857580.0
4DPI_cortex_1_1	813620.0
4DPI_SAC_3_3	2025800.0
4DPI_SAC_3_2	1299600.0
4DPI_SAC_2_3	1140400.0
4DPI_SAC_2_2	1866900.0

```
4DPI_SAC_2_1                                    1723900.0
4DPI_SAC_1_3                                    1446200.0
4DPI_SAC_1_2                                    2133800.0
4DPI_SAC_1_1                                    1344000.0
10DPI_interface_5_3                             1129300.0
10DPI_interface_4_3                             1920700.0
10DPI_interface_4_2                             1817400.0
10DPI_interface_4_1                             1819300.0
10DPI_interface_3_2                             1001900.0
10DPI_interface_3_1                             2526900.0
10DPI_cortex_4_4                                1237700.0
10DPI_cortex_4_3                                2119000.0
10DPI_cortex_3_3                                 881230.0
10DPI_cortex_3_2                                 797640.0
10DPI_SAC_4_3                                   1421000.0
10DPI_SAC_4_2                                   3697900.0
10DPI_SAC_4_1                                   2148600.0
10DPI_SAC_3_3                                   1626400.0
10DPI_SAC_3_2                                   1196300.0
10DPI_SAC_3_1                                    819090.0
10DPI_SAC_2_3                                   2583700.0
10DPI_SAC_2_2                                   1870900.0
10DPI_SAC_2_1                                   2521700.0

A0A1B0GQU8;P35980;A0A1B0GSS8;A0A1B0GSF7;A0A1B0GSA8;A0A1B0GS28;A0A1B0GRZ3  \
index
4DPI_interface_2_4                                      11981000.0
4DPI_interface_2_2                                      22897000.0
4DPI_interface_2_1                                      20121000.0
4DPI_interface_1_2                                      36610000.0
4DPI_interface_1_1                                      23111000.0
4DPI_cortex_3_3                                         16177000.0
4DPI_cortex_3_2                                         17586000.0
4DPI_cortex_2_3                                         21878000.0
4DPI_cortex_2_2                                         16561000.0
4DPI_cortex_2_1                                         23720000.0
4DPI_cortex_1_2                                         22475000.0
4DPI_cortex_1_1                                         20512000.0
4DPI_SAC_3_3                                            16860000.0
4DPI_SAC_3_2                                            22906000.0
4DPI_SAC_2_3                                            22223000.0
4DPI_SAC_2_2                                            26184000.0
4DPI_SAC_2_1                                            24626000.0
4DPI_SAC_1_3                                            18432000.0
4DPI_SAC_1_2                                            15517000.0
4DPI_SAC_1_1                                            15370000.0
10DPI_interface_5_3                                     11978000.0
```

```
10DPI_interface_4_3                                                 10966000.0
10DPI_interface_4_2                                                 11930000.0
10DPI_interface_4_1                                                 15307000.0
10DPI_interface_3_2                                                 23524000.0
10DPI_interface_3_1                                                 14105000.0
10DPI_cortex_4_4                                                    19951000.0
10DPI_cortex_4_3                                                    10131000.0
10DPI_cortex_3_3                                                    20335000.0
10DPI_cortex_3_2                                                    15715000.0
10DPI_SAC_4_3                                                       16438000.0
10DPI_SAC_4_2                                                       17678000.0
10DPI_SAC_4_1                                                       14233000.0
10DPI_SAC_3_3                                                       13098000.0
10DPI_SAC_3_2                                                        7518100.0
10DPI_SAC_3_1                                                       11445000.0
10DPI_SAC_2_3                                                       13901000.0
10DPI_SAC_2_2                                                       12134000.0
10DPI_SAC_2_1                                                       12099000.0

                     P06151;Q564E2;A0A1B0GSR9;A0A1B0GSX0;D3YZQ9;A0A1B0GSL7;A0A1B
0GT41;A0A1B0GQX5;A0A1B0GS79;A0A1B0GRW9;A0A1B0GRC1;A0A1B0GRS2;D3YVR7;D3YZE4;A0A1B
0GRE9;A0A1B0GSR2;P00342  \
index
4DPI_interface_2_4                                                  48765000.0
4DPI_interface_2_2                                                  84281000.0
4DPI_interface_2_1                                                  62405000.0
4DPI_interface_1_2                                                  61114000.0
4DPI_interface_1_1                                                  36589000.0
4DPI_cortex_3_3                                                     33949000.0
4DPI_cortex_3_2                                                     29980000.0
4DPI_cortex_2_3                                                     62446000.0
4DPI_cortex_2_2                                                     63940000.0
4DPI_cortex_2_1                                                     25525000.0
4DPI_cortex_1_2                                                     45436000.0
4DPI_cortex_1_1                                                     61932000.0
4DPI_SAC_3_3                                                        58681000.0
4DPI_SAC_3_2                                                        56833000.0
4DPI_SAC_2_3                                                        55551000.0
4DPI_SAC_2_2                                                        50517000.0
4DPI_SAC_2_1                                                        84878000.0
4DPI_SAC_1_3                                                        81791000.0
4DPI_SAC_1_2                                                        81333000.0
4DPI_SAC_1_1                                                        59981000.0
10DPI_interface_5_3                                                 38969000.0
10DPI_interface_4_3                                                 61117000.0
10DPI_interface_4_2                                                 41654000.0
10DPI_interface_4_1                                                 71086000.0
```

```
10DPI_interface_3_2                                                51215000.0
10DPI_interface_3_1                                                34041000.0
10DPI_cortex_4_4                                                   51373000.0
10DPI_cortex_4_3                                                   36612000.0
10DPI_cortex_3_3                                                   38540000.0
10DPI_cortex_3_2                                                   19880000.0
10DPI_SAC_4_3                                                      79881000.0
10DPI_SAC_4_2                                                      92821000.0
10DPI_SAC_4_1                                                      91489000.0
10DPI_SAC_3_3                                                      47923000.0
10DPI_SAC_3_2                                                      87510000.0
10DPI_SAC_3_1                                                      39045000.0
10DPI_SAC_2_3                                                      27205000.0
10DPI_SAC_2_2                                                      50918000.0
10DPI_SAC_2_1                                                      16776000.0

                     ...            36            37            38  \
index                ...
4DPI_interface_2_4   ...  1.031646e+07 -1.178469e+07  2.929428e-07
4DPI_interface_2_2   ...  2.282268e+07 -5.412626e+04  2.929428e-07
4DPI_interface_2_1   ... -2.718176e+07  2.199637e+06  2.929428e-07
4DPI_interface_1_2   ...  1.065881e+07 -7.706425e+05  2.929428e-07
4DPI_interface_1_1   ... -1.414153e+07  1.404734e+06  2.929428e-07
4DPI_cortex_3_3      ... -3.510812e+06 -1.258512e+07  2.929428e-07
4DPI_cortex_3_2      ... -1.934336e+07  6.848204e+06  2.929428e-07
4DPI_cortex_2_3      ... -2.514064e+07  9.140492e+06  2.929428e-07
4DPI_cortex_2_2      ...  4.206068e+07 -1.488583e+07  2.929428e-07
4DPI_cortex_2_1      ...  9.972975e+06  1.967906e+06  2.929428e-07
4DPI_cortex_1_2      ... -1.634307e+07 -2.895322e+06  2.929428e-07
4DPI_cortex_1_1      ...  6.039433e+06  1.772419e+07  2.929428e-07
4DPI_SAC_3_3         ...  1.320228e+07  4.387337e+04  2.929428e-07
4DPI_SAC_3_2         ... -3.905192e+06  2.317435e+07  2.929428e-07
4DPI_SAC_2_3         ...  1.960526e+06 -8.147903e+05  2.929428e-07
4DPI_SAC_2_2         ... -5.308582e+06  3.966172e+06  2.929428e-07
4DPI_SAC_2_1         ... -2.123865e+06 -4.008947e+06  2.929428e-07
4DPI_SAC_1_3         ...  9.412037e+06 -1.014204e+07  2.929428e-07
4DPI_SAC_1_2         ... -8.900910e+06  6.486090e+06  2.929428e-07
4DPI_SAC_1_1         ...  3.946169e+06 -1.159411e+07  2.929428e-07
10DPI_interface_5_3  ...  1.110906e+07  9.848782e+06  2.929428e-07
10DPI_interface_4_3  ... -1.415164e+07 -2.113884e+06  2.929428e-07
10DPI_interface_4_2  ... -1.030126e+07 -2.357560e+07  2.929428e-07
10DPI_interface_4_1  ...  2.432251e+07  2.044187e+07  2.929428e-07
10DPI_interface_3_2  ...  1.381270e+07 -3.013517e+07  2.929428e-07
10DPI_interface_3_1  ... -2.592833e+07  5.400819e+06  2.929428e-07
10DPI_cortex_4_4     ... -4.740456e+06 -7.302414e+06  2.929428e-07
10DPI_cortex_4_3     ...  1.500560e+07  2.581753e+06  2.929428e-07
10DPI_cortex_3_3     ... -8.358536e+06  3.333076e+06  2.929428e-07
```

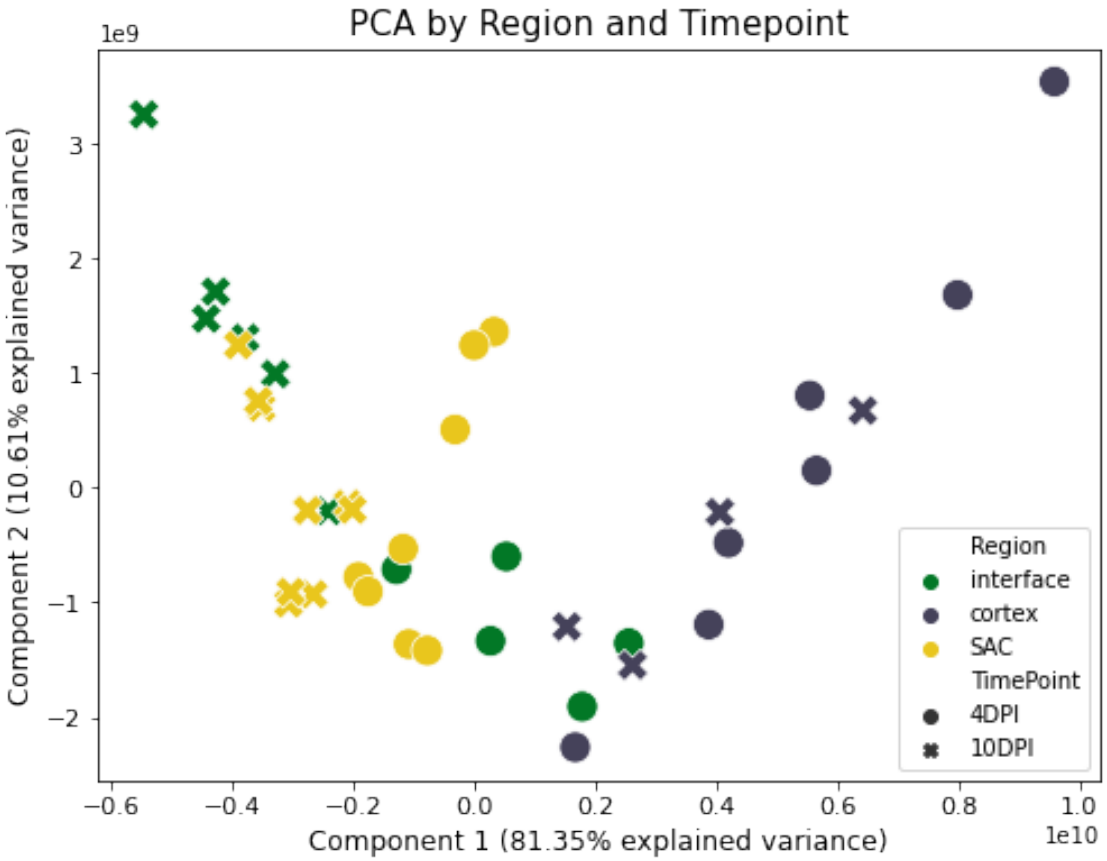
```
10DPI_cortex_3_2     ...  4.094822e+06 -3.480078e+06  2.929428e-07
10DPI_SAC_4_3        ...  6.421815e+06  1.283613e+07  2.929428e-07
10DPI_SAC_4_2        ...  4.099798e+06 -4.146932e+06  2.929428e-07
10DPI_SAC_4_1        ... -1.157357e+07 -3.995583e+06  2.929428e-07
10DPI_SAC_3_3        ... -2.697074e+07 -3.500647e+06  2.929428e-07
10DPI_SAC_3_2        ...  8.463252e+04 -1.427429e+07  2.929428e-07
10DPI_SAC_3_1        ...  1.863299e+07  4.169473e+07  2.929428e-07
10DPI_SAC_2_3        ... -5.895312e+06 -4.218407e+06  2.929428e-07
10DPI_SAC_2_2        ... -3.574784e+06  3.894834e+06  2.929428e-07
10DPI_SAC_2_1        ...  9.418379e+06 -6.709035e+06  2.929428e-07

                   Segment K-means PCA  TimePoint  BioRep  TechRep  \
index
4DPI_interface_2_4                   1       4DPI       2        4
4DPI_interface_2_2                   1       4DPI       2        2
4DPI_interface_2_1                   1       4DPI       2        1
4DPI_interface_1_2                   3       4DPI       1        2
4DPI_interface_1_1                   3       4DPI       1        1
4DPI_cortex_3_3                      3       4DPI       3        3
4DPI_cortex_3_2                      3       4DPI       3        2
4DPI_cortex_2_3                      2       4DPI       2        3
4DPI_cortex_2_2                      2       4DPI       2        2
4DPI_cortex_2_1                      3       4DPI       2        1
4DPI_cortex_1_2                      2       4DPI       1        2
4DPI_cortex_1_1                      2       4DPI       1        1
4DPI_SAC_3_3                         1       4DPI       3        3
4DPI_SAC_3_2                         1       4DPI       3        2
4DPI_SAC_2_3                         1       4DPI       2        3
4DPI_SAC_2_2                         1       4DPI       2        2
4DPI_SAC_2_1                         1       4DPI       2        1
4DPI_SAC_1_3                         1       4DPI       1        3
4DPI_SAC_1_2                         1       4DPI       1        2
4DPI_SAC_1_1                         1       4DPI       1        1
10DPI_interface_5_3                  0      10DPI       5        3
10DPI_interface_4_3                  0      10DPI       4        3
10DPI_interface_4_2                  0      10DPI       4        2
10DPI_interface_4_1                  0      10DPI       4        1
10DPI_interface_3_2                  1      10DPI       3        2
10DPI_interface_3_1                  0      10DPI       3        1
10DPI_cortex_4_4                     3      10DPI       4        4
10DPI_cortex_4_3                     3      10DPI       4        3
10DPI_cortex_3_3                     3      10DPI       3        3
10DPI_cortex_3_2                     2      10DPI       3        2
10DPI_SAC_4_3                        1      10DPI       4        3
10DPI_SAC_4_2                        1      10DPI       4        2
10DPI_SAC_4_1                        1      10DPI       4        1
10DPI_SAC_3_3                        0      10DPI       3        3
```

10DPI_SAC_3_2	0	10DPI	3	2
10DPI_SAC_3_1	0	10DPI	3	1
10DPI_SAC_2_3	1	10DPI	2	3
10DPI_SAC_2_2	1	10DPI	2	2
10DPI_SAC_2_1	1	10DPI	2	1

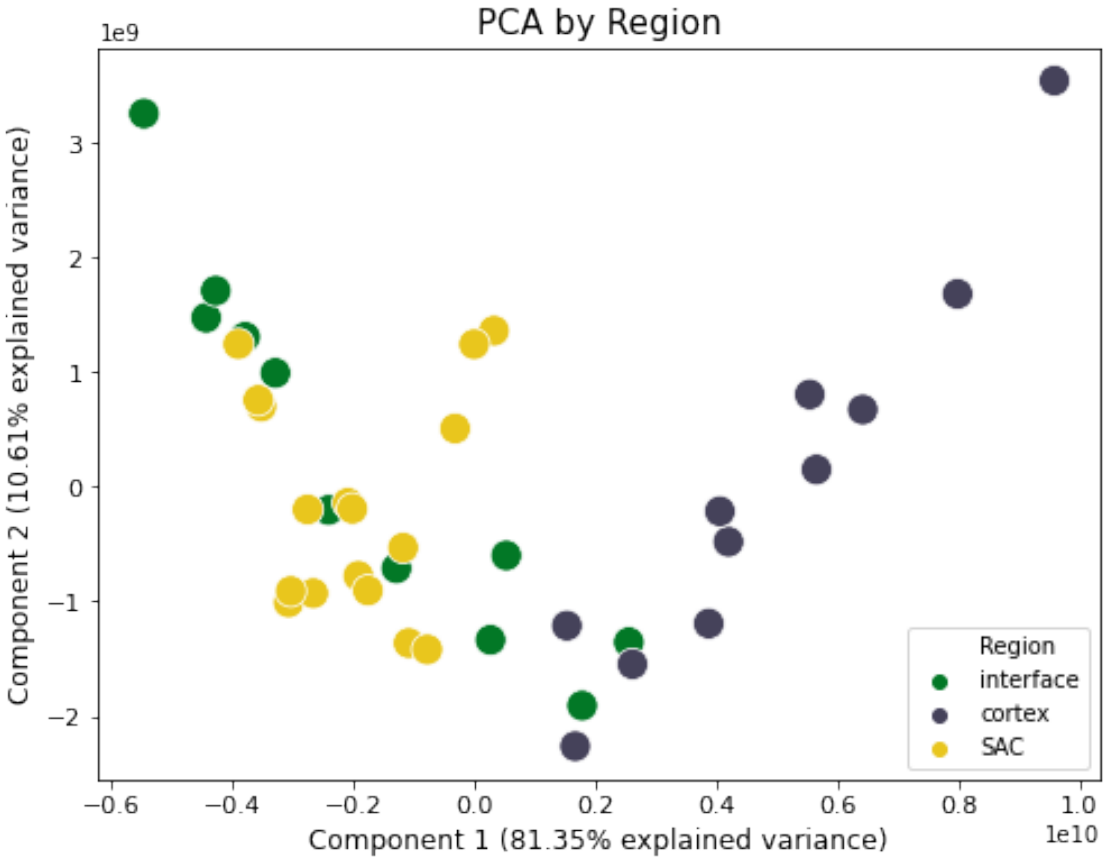
index	Region	DPI_Region	BioRep_TechRep
4DPI_interface_2_4	interface	4DPI_interface	2_4
4DPI_interface_2_2	interface	4DPI_interface	2_2
4DPI_interface_2_1	interface	4DPI_interface	2_1
4DPI_interface_1_2	interface	4DPI_interface	1_2
4DPI_interface_1_1	interface	4DPI_interface	1_1
4DPI_cortex_3_3	cortex	4DPI_cortex	3_3
4DPI_cortex_3_2	cortex	4DPI_cortex	3_2
4DPI_cortex_2_3	cortex	4DPI_cortex	2_3
4DPI_cortex_2_2	cortex	4DPI_cortex	2_2
4DPI_cortex_2_1	cortex	4DPI_cortex	2_1
4DPI_cortex_1_2	cortex	4DPI_cortex	1_2
4DPI_cortex_1_1	cortex	4DPI_cortex	1_1
4DPI_SAC_3_3	SAC	4DPI_SAC	3_3
4DPI_SAC_3_2	SAC	4DPI_SAC	3_2
4DPI_SAC_2_3	SAC	4DPI_SAC	2_3
4DPI_SAC_2_2	SAC	4DPI_SAC	2_2
4DPI_SAC_2_1	SAC	4DPI_SAC	2_1
4DPI_SAC_1_3	SAC	4DPI_SAC	1_3
4DPI_SAC_1_2	SAC	4DPI_SAC	1_2
4DPI_SAC_1_1	SAC	4DPI_SAC	1_1
10DPI_interface_5_3	interface	10DPI_interface	5_3
10DPI_interface_4_3	interface	10DPI_interface	4_3
10DPI_interface_4_2	interface	10DPI_interface	4_2
10DPI_interface_4_1	interface	10DPI_interface	4_1
10DPI_interface_3_2	interface	10DPI_interface	3_2
10DPI_interface_3_1	interface	10DPI_interface	3_1
10DPI_cortex_4_4	cortex	10DPI_cortex	4_4
10DPI_cortex_4_3	cortex	10DPI_cortex	4_3
10DPI_cortex_3_3	cortex	10DPI_cortex	3_3
10DPI_cortex_3_2	cortex	10DPI_cortex	3_2
10DPI_SAC_4_3	SAC	10DPI_SAC	4_3
10DPI_SAC_4_2	SAC	10DPI_SAC	4_2
10DPI_SAC_4_1	SAC	10DPI_SAC	4_1
10DPI_SAC_3_3	SAC	10DPI_SAC	3_3
10DPI_SAC_3_2	SAC	10DPI_SAC	3_2
10DPI_SAC_3_1	SAC	10DPI_SAC	3_1
10DPI_SAC_2_3	SAC	10DPI_SAC	2_3
10DPI_SAC_2_2	SAC	10DPI_SAC	2_2
10DPI_SAC_2_1	SAC	10DPI_SAC	2_1

```
[39 rows x 333 columns]
```

[82]:
```
x_axis = df_segm_pca_kmeans['Component 1']
y_axis = df_segm_pca_kmeans['Component 2']
plt.figure(figsize=(8,6))
sns.scatterplot(x_axis, y_axis,␣
 ↪hue=df_segm_pca_kmeans['Region'],palette=["#027826","#45425A","#E9C61D"],␣
 ↪style=df_segm_pca_kmeans['TimePoint'], s=200, legend='full')
plt.title('PCA by Region and Timepoint', fontsize=15)
plt.tick_params(axis='both', which='major', labelsize=11)
plt.tick_params(axis='both', which='minor', labelsize=11)
plt.xlabel('Component 1 (' + str((pca_expl_var[0]*100).round(2)) +'% explained␣
 ↪variance)', fontsize=12)
plt.ylabel('Component 2 (' + str((pca_expl_var[1]*100).round(2)) +'% explained␣
 ↪variance)', fontsize=12)
plt.legend(loc='lower right')
plt.savefig('outs/' + date + '_combined_PCAwithRegionsandTimepoints.svg')
```

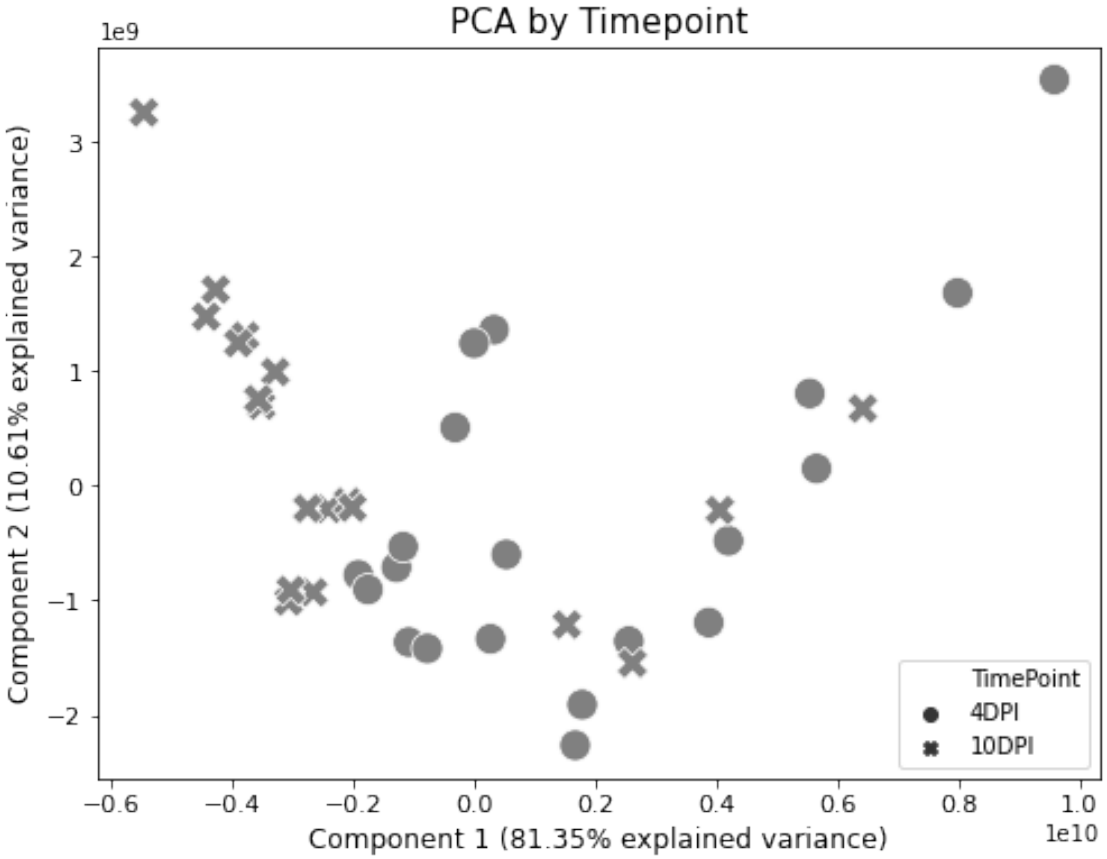


[83]:
```
x_axis = df_segm_pca_kmeans['Component 1']
y_axis = df_segm_pca_kmeans['Component 2']
plt.figure(figsize=(8,6))
sns.scatterplot(x_axis, y_axis,␣
 ↪hue=df_segm_pca_kmeans['Region'],palette=["#027826","#45425A","#E9C61D"],s=200,␣
 ↪legend='full')
plt.title('PCA by Region', fontsize=15)
plt.tick_params(axis='both', which='major', labelsize=11)
plt.tick_params(axis='both', which='minor', labelsize=11)
plt.xlabel('Component 1 (' + str((pca_expl_var[0]*100).round(2)) +'% explained␣
 ↪variance)', fontsize=12)
plt.ylabel('Component 2 (' + str((pca_expl_var[1]*100).round(2)) +'% explained␣
 ↪variance)', fontsize=12)
plt.legend(loc='lower right')
plt.savefig('outs/' + date + '_combined_PCAbyRegion.svg')
```



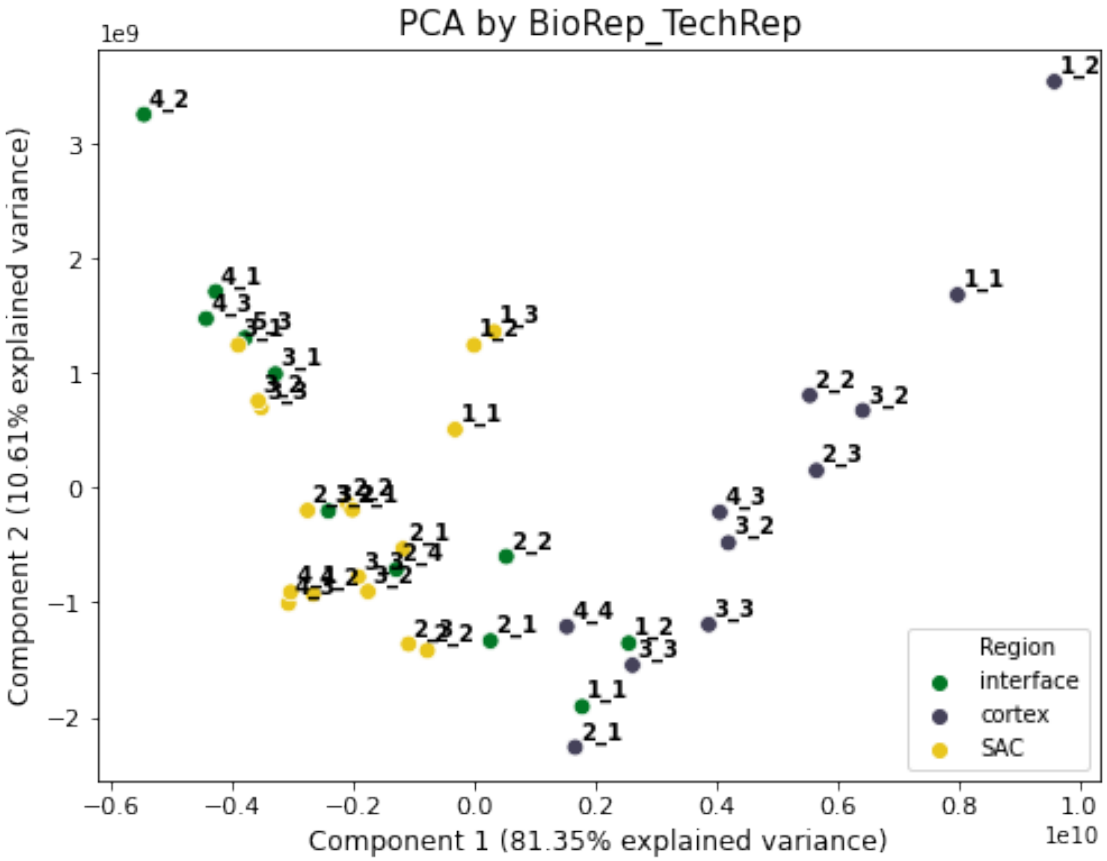
[84]:
```
x_axis = df_segm_pca_kmeans['Component 1']
y_axis = df_segm_pca_kmeans['Component 2']
plt.figure(figsize=(8,6))
sns.scatterplot(x_axis,
```

```
                y_axis,
                color='grey',
                style=df_segm_pca_kmeans['TimePoint'],
                s=200,
                legend='full')
plt.title('PCA by Timepoint', fontsize=15)
plt.tick_params(axis='both', which='major', labelsize=11)
plt.tick_params(axis='both', which='minor', labelsize=11)
plt.xlabel('Component 1 (' + str((pca_expl_var[0]*100).round(2)) +'% explained␣
 ↪variance)', fontsize=12)
plt.ylabel('Component 2 (' + str((pca_expl_var[1]*100).round(2)) +'% explained␣
 ↪variance)', fontsize=12)
plt.legend(loc='lower right')
plt.savefig('outs/' + date + '_combined_PCAbyTimepoint.svg')
```

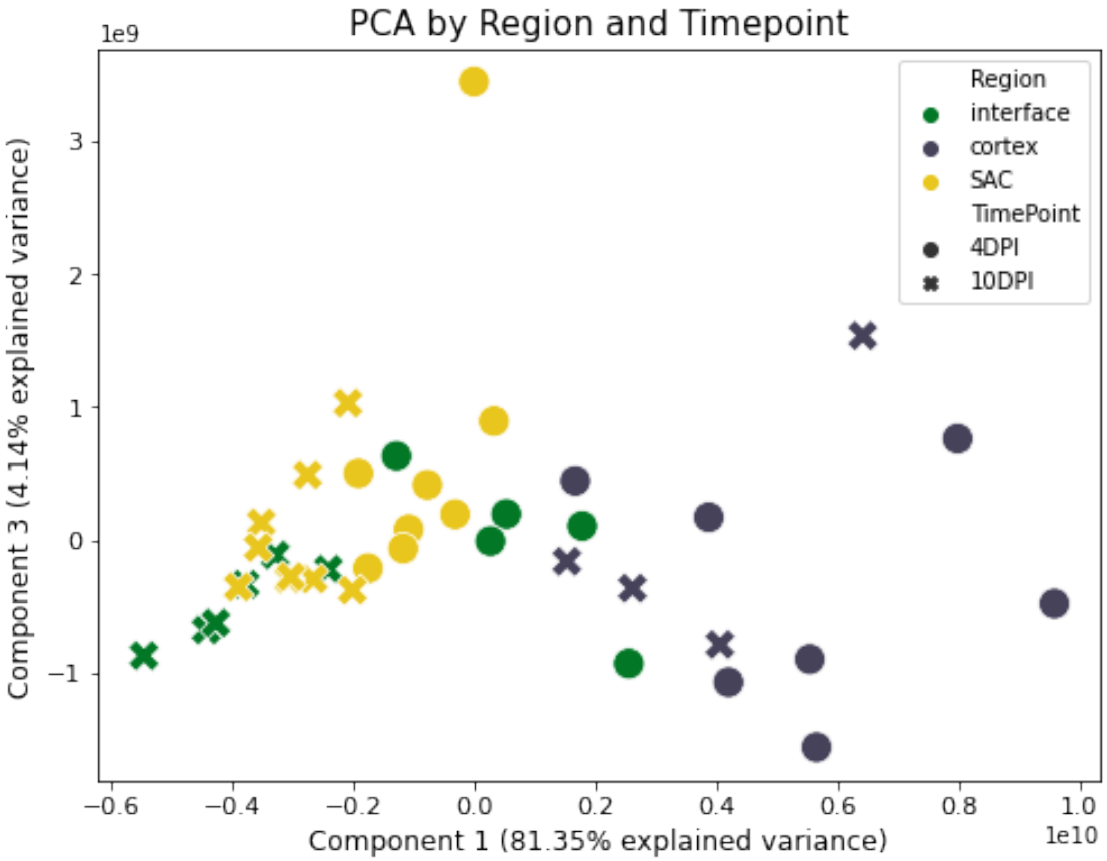


[174]:
```
x_axis = df_segm_pca_kmeans['Component 1']
y_axis = df_segm_pca_kmeans['Component 2']
labels = df_segm_pca_kmeans['BioRep_TechRep']
plt.figure(figsize=(8,6))
sns.scatterplot(x_axis,
```

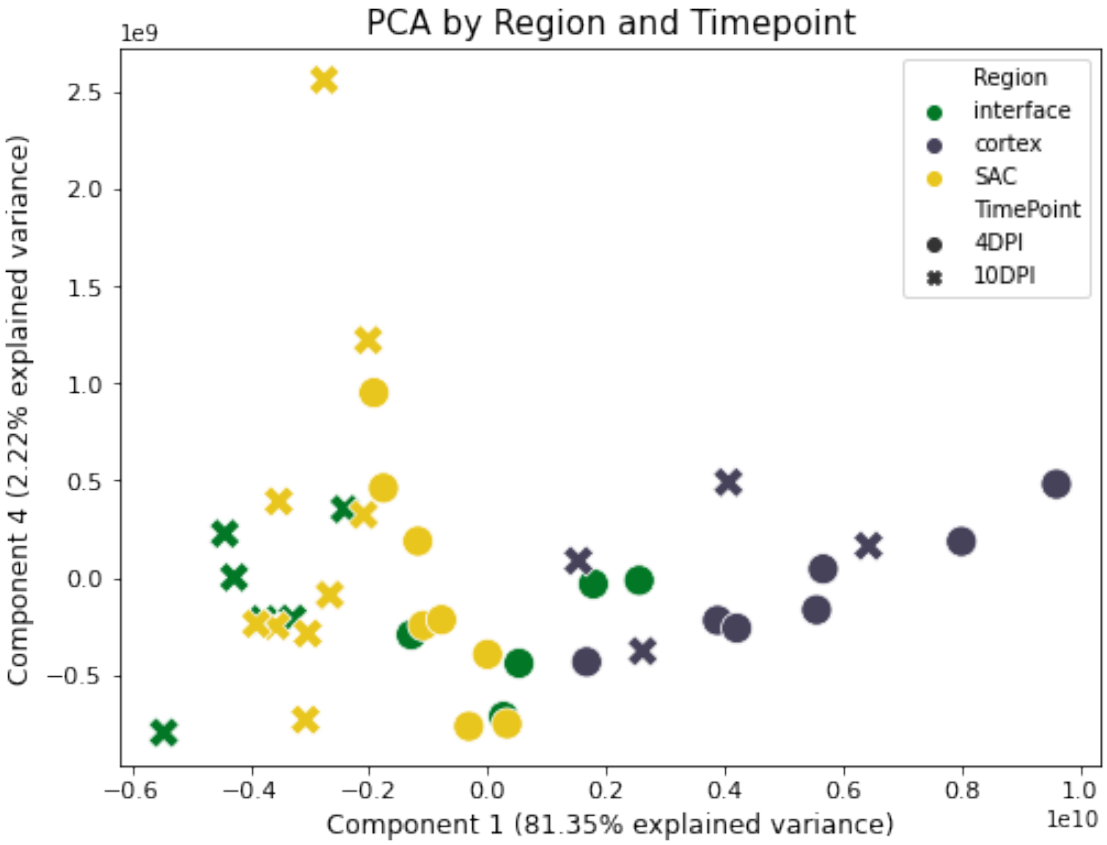
```
                y_axis,
                hue=df_segm_pca_kmeans['Region'],
                palette=["#027826","#45425A","#E9C61D"],
                s=60,
                legend='full')
plt.title('PCA by BioRep_TechRep', fontsize=15)
plt.tick_params(axis='both', which='major', labelsize=11)
plt.tick_params(axis='both', which='minor', labelsize=11)
plt.xlabel('Component 1 (' + str((pca_expl_var[0]*100).round(2)) +'% explained␣
 ↪variance)', fontsize=12)
plt.ylabel('Component 2 (' + str((pca_expl_var[1]*100).round(2)) +'% explained␣
 ↪variance)', fontsize=12)
plt.legend(loc='lower right')

for line in range(0,df_segm_pca_kmeans.shape[0]):
     plt.annotate(xy = (x_axis[line], y_axis[line]),
               text = labels[line],
               horizontalalignment='left',
               verticalalignment='bottom',
               size='medium',
               color='black',
               weight='semibold',
               xytext=(2, 2),
               textcoords='offset points')

plt.savefig('outs/' + date + '_combined_PCAbyBioRep_TechRep.svg')
```



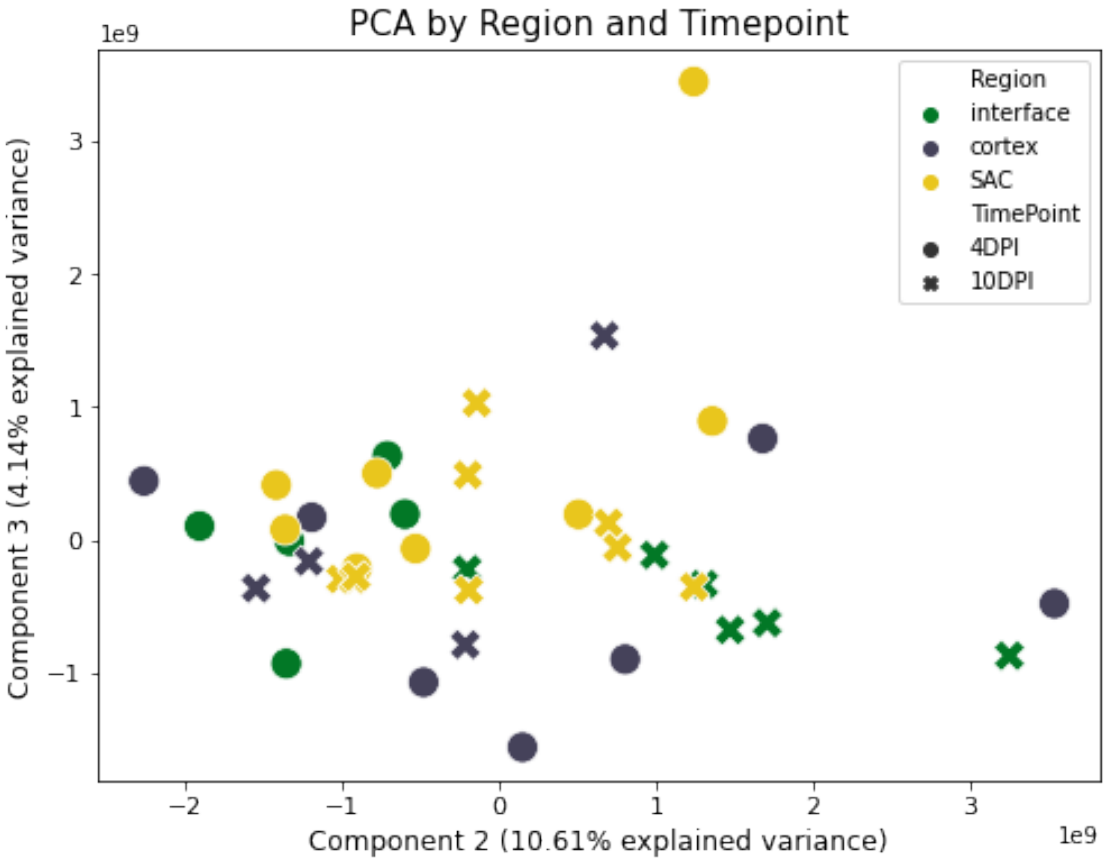
[175]:
```
x_axis = df_segm_pca_kmeans['Component 1']
y_axis = df_segm_pca_kmeans['Component 3']
plt.figure(figsize=(8,6))
sns.scatterplot(x_axis,
                y_axis,
                hue=df_segm_pca_kmeans['Region'],
                palette=["#027826","#45425A","#E9C61D"],
                style=df_segm_pca_kmeans['TimePoint'],
                s=200,
                legend='full')
plt.title('PCA by Region and Timepoint', fontsize=15)
plt.tick_params(axis='both', which='major', labelsize=11)
plt.tick_params(axis='both', which='minor', labelsize=11)
plt.xlabel('Component 1 (' + str((pca_expl_var[0]*100).round(2)) +'% explained␣
 ↪variance)', fontsize=12)
plt.ylabel('Component 3 (' + str((pca_expl_var[2]*100).round(2)) +'% explained␣
 ↪variance)', fontsize=12)
plt.legend(loc='upper right')
plt.savefig('outs/' + date + '_combined_PCAwithRegionsandTimepoints_PC13.svg')
```



[176]:
```
x_axis = df_segm_pca_kmeans['Component 1']
y_axis = df_segm_pca_kmeans['Component 4']
plt.figure(figsize=(8,6))
sns.scatterplot(x_axis,
                y_axis,
                hue=df_segm_pca_kmeans['Region'],
                palette=["#027826","#45425A","#E9C61D"],
                style=df_segm_pca_kmeans['TimePoint'],
                s=200,
                legend='full')
plt.title('PCA by Region and Timepoint', fontsize=15)
plt.tick_params(axis='both', which='major', labelsize=11)
plt.tick_params(axis='both', which='minor', labelsize=11)
plt.xlabel('Component 1 (' + str((pca_expl_var[0]*100).round(2)) +'% explained␣
 ↪variance)', fontsize=12)
plt.ylabel('Component 4 (' + str((pca_expl_var[3]*100).round(2)) +'% explained␣
 ↪variance)', fontsize=12)
plt.legend(loc='upper right')
plt.savefig('outs/' + date + '_combined_PCAwithRegionsandTimepoints_PC14.svg')
```



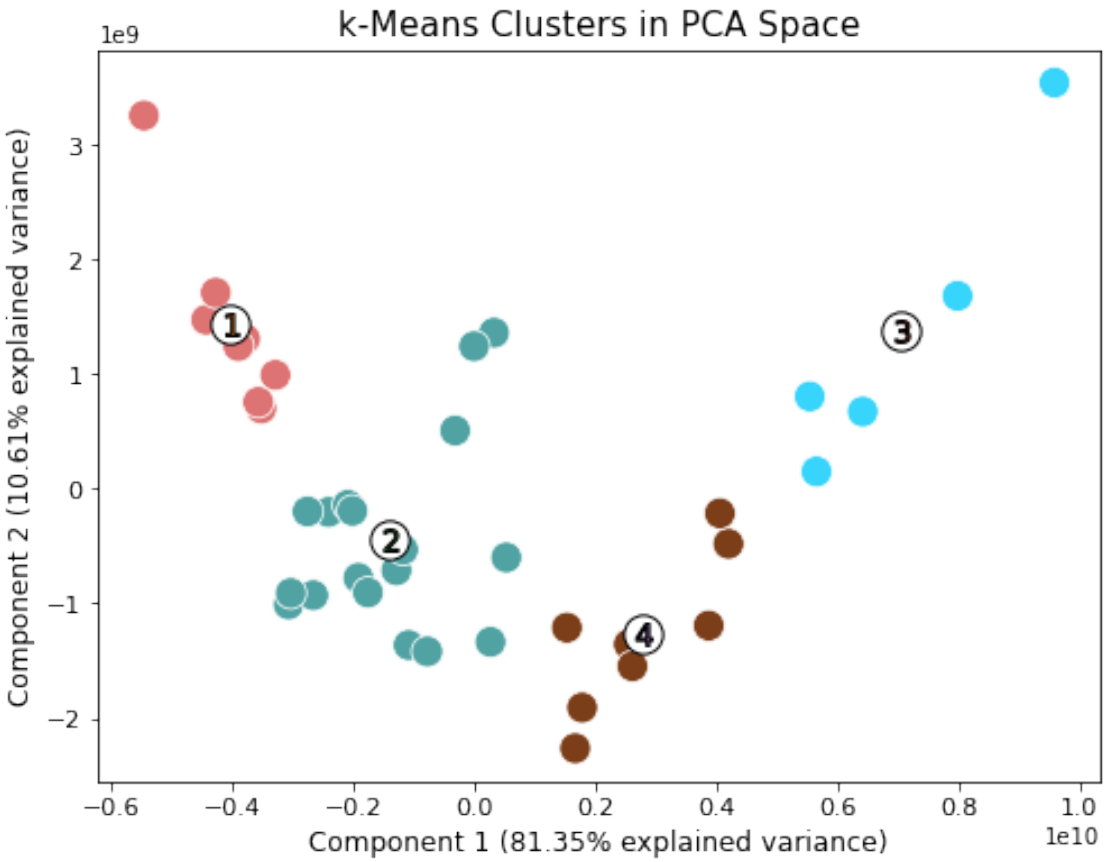
[177]:
```
x_axis = df_segm_pca_kmeans['Component 2']
y_axis = df_segm_pca_kmeans['Component 3']
plt.figure(figsize=(8,6))
sns.scatterplot(x_axis, y_axis,
                hue=df_segm_pca_kmeans['Region'],
                palette=["#027826","#45425A","#E9C61D"],
                style=df_segm_pca_kmeans['TimePoint'],
                s=200,
                legend='full')
plt.title('PCA by Region and Timepoint', fontsize=15)
plt.tick_params(axis='both', which='major', labelsize=11)
plt.tick_params(axis='both', which='minor', labelsize=11)
plt.xlabel('Component 2 (' + str((pca_expl_var[1]*100).round(2)) +'% explained␣
 ↪variance)', fontsize=12)
plt.ylabel('Component 3 (' + str((pca_expl_var[2]*100).round(2)) +'% explained␣
 ↪variance)', fontsize=12)
plt.legend(loc='upper right')
plt.savefig('outs/' + date + '_combined_PCAwithRegionsandTimepoints_PC23.svg')
```



[178]:
```
palette_kmeans = ["#dd7373","#51a3a3","#37d4fb","#7b3e19"]
```

[179]:
```
x_axis = df_segm_pca_kmeans['Component 1']
y_axis = df_segm_pca_kmeans['Component 2']
plt.figure(figsize=(8,6))

sns.scatterplot(x_axis, y_axis, hue=df_segm_pca_kmeans['Segment K-means␣
 ↪PCA'],palette=palette_kmeans,s=200, legend=None)

plt.title('k-Means Clusters in PCA Space', fontsize=15)
plt.tick_params(axis='both', which='major', labelsize=11)
plt.tick_params(axis='both', which='minor', labelsize=11)
plt.xlabel('Component 1 (' + str((pca_expl_var[0]*100).round(2)) +'% explained␣
 ↪variance)', fontsize=12)
plt.ylabel('Component 2 (' + str((pca_expl_var[1]*100).round(2)) +'% explained␣
 ↪variance)', fontsize=12)

plt.scatter(cluster_centers[:, 0], cluster_centers[:, 1], marker='o',
                c="white", alpha=1, s=300, edgecolor='k')
```

```
for i, c in enumerate(cluster_centers):
    i = i+1
    plt.scatter(c[0], c[1], marker='$%d$' % i, alpha=1,
                    s=100, edgecolor='k')

plt.savefig('outs/' + date + '_combined_kClustersinPCAspace.svg')
```



[180]:
```
print(cluster_centers)
print(type(cluster_centers))
```

```
[[-4.03019735e+09  1.42861860e+09 -3.62243311e+08 -1.35817336e+08
   1.06984274e+08  3.29454473e+07 -1.88805686e+07 -6.93786757e+07
  -3.30601347e+07  3.75564111e+06  2.05380309e+07  8.19662582e+06
   1.36503173e+07  2.08327123e+07  1.83816557e+07 -1.63459977e+07
   1.45807440e+07  1.69444361e+07  7.66743612e+06 -3.92161608e+06
  -8.04736787e+05  4.48151238e+06  3.27507424e+06  6.61627058e+06
   2.44923765e+06 -1.99081818e+06 -2.88351258e+06 -2.04113397e+06
  -1.47470614e+06 -4.03186553e+06 -3.00987622e+06  1.32354349e+06
  -1.13636306e+06  1.02620642e+06  5.05175036e+06  6.24579750e+05
  -2.90034849e+06  4.24022242e+06  2.92942842e-07]
```

```
[-1.40253631e+09 -4.50482097e+08  3.40041833e+08  6.28997019e+07
 -9.79711332e+07 -2.71040892e+07  2.27752926e+07  8.64032150e+07
  2.36974976e+06  9.84461085e+06 -7.81950527e+06 -4.57226219e+06
 -6.15663700e+06 -7.18565738e+06 -1.29394263e+07  1.33325889e+07
 -1.08417688e+07 -1.06147884e+07 -4.85661124e+06  3.16629298e+06
  1.29649164e+06 -1.78728383e+06  4.86709905e+05 -3.64280238e+06
 -1.35243966e+06  2.53569738e+05  4.50953066e+06  6.36056732e+05
  1.11333230e+06  2.42739822e+06  2.59040179e+06 -9.23552395e+05
  1.58763058e+05  1.05391964e+05 -1.96455976e+06  9.01510395e+04
  1.49715915e+06 -1.94459615e+06  2.92942842e-07]
[ 7.04060932e+09  1.36934796e+09 -1.26842256e+08  1.42314384e+08
  4.52560521e+07  1.64069463e+07  1.71095864e+07  1.88903315e+07
 -2.58743434e+07  6.08497911e+07  3.80342308e+07 -3.96927884e+07
 -8.62387196e+06 -7.87102830e+06  6.65039646e+06  1.33382000e+07
  1.99196938e+07 -1.28346415e+07 -2.25619600e+06 -2.22523794e+06
  2.80125176e+06 -2.31503553e+06  1.72611708e+07 -1.51500516e+07
 -7.97151832e+06 -1.46370109e+07 -2.86138679e+06  5.88732688e+06
  1.87291287e+06 -4.07563098e+06  3.69996399e+06  3.01382455e+06
 -3.31673591e+06 -5.40107936e+06  1.03092320e+05 -1.89425282e+06
  2.14224554e+06  1.12069078e+06  2.92942842e-07]
[ 2.78552322e+09 -1.27087635e+09 -3.23574402e+08 -9.46534829e+07
  8.51657433e+07  1.77844118e+07 -4.30573311e+07 -1.36835015e+08
  4.38996623e+07 -6.39371350e+07 -2.67155383e+07  2.68989569e+07
  5.59203597e+06  2.54409469e+05  6.57555582e+06 -2.19887024e+07
 -2.63657271e+06  1.49604887e+07  4.67006167e+06 -1.81176940e+06
 -3.86315175e+06  9.86773451e+05 -1.51584033e+07  1.10488170e+07
  5.57595053e+06  1.05684181e+07 -5.47456466e+06 -3.06957298e+06
 -2.20086208e+06  1.11748890e+06 -5.13100531e+06 -1.12919094e+06
  2.85210612e+06  2.11233626e+06 -6.95923611e+05  3.56488424e+05
 -1.80716306e+06 -5.65312807e+05  2.92942842e-07]]
<class 'numpy.ndarray'>
```

[181]:
```
cluster_centers.shape
```

[181]:
```
(4, 39)
```

[182]:
```
components.shape
```

[182]:
```
(39, 287)
```

[183]:
```
centers = cluster_centers.dot(components)
centers.shape
```

[183]:
```
(4, 287)
```

[184]:
```
centers = pd.DataFrame(centers,
                       index=['Cluster 1', 'Cluster 2', 'Cluster 3', 'Cluster␣
 ↪4'],
                       columns = X.columns)
centers.to_csv('outs/' + date + '_combined_centers_annotated.csv')
```

[185]:
```
centers_t = centers.T
centers_t.shape
```

[185]:
```
(287, 4)
```

[186]:
```
centers
```

[186]:
```
 Protein IDs  A0A0J9YTY0;A0A0J9YUL3;Q8C1B7;A0A0J9YUV6;A0A0J9YVA6  \
Cluster 1                                          5.347031e+06
Cluster 2                                          5.557197e+05
Cluster 3                                         -4.514309e+06
Cluster 4                                         -3.775957e+06

 Protein IDs  A0A0J9YUZ4;P63158;A0A0J9YUD8;D3YVC6;D3YZ18  A0A0R4J093;Q9DBP5  \
Cluster 1                                    6.832731e+06       6.331833e+05
Cluster 2                                    2.468034e+06      -8.206000e+05
Cluster 3                                   -6.670144e+06       2.444753e+06
Cluster 4                                   -8.216969e+06      -3.148042e+05

 Protein IDs  A0A0R4J0I1;P07759;E9Q499;Q80X76  O70404;A0A0R4J0R1;A0A0U1RPE8  \
Cluster 1                      -1.365104e+07                 2.010879e+06
Cluster 2                       5.144795e+06                 8.577997e+05
Cluster 3                       1.707926e+07                -2.495012e+06
Cluster 4                      -8.599288e+06                -2.381546e+06

 Protein IDs  A0A0R4J0Z1;P08003  \
Cluster 1          3.832855e+06
Cluster 2          1.883494e+06
Cluster 3         -7.851255e+06
Cluster 4         -3.163682e+06

 Protein IDs
E9QL31;A0A0R4J104;P98078;E9PX84;Q3TRE6;F6TQN9;Q9DCE6;E0CXT5;E0CZ53;E0CYJ2  \
Cluster 1                                          -2.891130e+06
Cluster 2                                          -2.992265e+06
Cluster 3                                           5.900767e+06
Cluster 4                                           5.935747e+06

 Protein IDs  A0A0R4J195;Q9CSU0;F6YB25;F7DAY5;Q8VDS4  \
Cluster 1                              114204.262820
Cluster 2                              236712.735043
Cluster 3                             -564013.487179
Cluster 4                             -294299.487179

 Protein IDs
A0A1B0GQU8;P35980;A0A1B0GSS8;A0A1B0GSF7;A0A1B0GSA8;A0A1B0GS28;A0A1B0GRZ3  \
Cluster 1                                          -5.603615e+06
Cluster 2                                           5.265530e+05
```

```
Cluster 3                                                1.781197e+06
Cluster 4                                                3.305622e+06

 Protein IDs  P06151;Q564E2;A0A1B0GSR9;A0A1B0GSX0;D3YZQ9;A0A1B0GSL7;A0A1B0GT41;A
0A1B0GQX5;A0A1B0GS79;A0A1B0GRW9;A0A1B0GRC1;A0A1B0GRS2;D3YVR7;D3YZE4;A0A1B0GRE9;A
0A1B0GSR2;P00342  \
Cluster 1                                           -1.792952e+06
Cluster 2                                            8.612312e+06
Cluster 3                                           -3.734277e+06
Cluster 4                                           -1.525083e+07

 Protein IDs  ...  Q9JKF1;A0A0U1RNG5;A0A0U1RPI2;A0A0U1RPU3  Q9JLJ2;Q3U367  \
Cluster 1     ...                            1.680077e+07  -6.899002e+06
Cluster 2     ...                            5.252364e+06  -4.775488e+06
Cluster 3     ...                           -1.551040e+07   1.191792e+07
Cluster 4     ...                           -1.892459e+07   1.019515e+07

 Protein IDs        Q9QZQ8  Q9R1P4;A0A1B0GS70        Q9WU78  \
Cluster 1     8.566707e+06       3.967160e+06  4.199472e+06
Cluster 2     4.460265e+05       2.049018e+06  9.130855e+04
Cluster 3    -6.123298e+06      -5.780895e+06 -1.380523e+06
Cluster 4    -5.743205e+06      -4.964390e+06 -3.542090e+06

 Protein IDs  Q9Z0P4;Q3UZP7;A0A1W2P879;A0A1W2P831        Q9Z1D1        Q9Z1Q5  \
Cluster 1                          -882132.500000  4.513414e+05  2.564748e+06
Cluster 2                           -98701.666667  2.961831e+05  1.874555e+06
Cluster 3                          909390.000000 -1.005512e+06 -6.734190e+06
Cluster 4                          535842.500000 -4.893086e+05 -2.573627e+06

 Protein IDs        Q9Z2D6  Q9Z2I8;A0A0N4SWD1
Cluster 1    -1.676840e+06      -1.577885e+07
Cluster 2    -4.070078e+05      -8.384045e+06
Cluster 3     2.729420e+06       2.848534e+07
Cluster 4     8.867200e+05       1.683961e+07

[4 rows x 287 columns]
```

[187]:
```
centers_t
```

[187]:
```
                                                     Cluster 1  \
 Protein IDs
A0A0J9YTY0;A0A0J9YUL3;Q8C1B7;A0A0J9YUV6;A0A0J9YVA6  5.347031e+06
A0A0J9YUZ4;P63158;A0A0J9YUD8;D3YVC6;D3YZ18           6.832731e+06
A0A0R4J093;Q9DBP5                                    6.331833e+05
A0A0R4J0I1;P07759;E9Q499;Q80X76                     -1.365104e+07
O70404;A0A0R4J0R1;A0A0U1RPE8                         2.010879e+06
...                                                           ...
Q9Z0P4;Q3UZP7;A0A1W2P879;A0A1W2P831                 -8.821325e+05
```

```
Q9Z1D1                                                 4.513414e+05
Q9Z1Q5                                                 2.564748e+06
Q9Z2D6                                                -1.676840e+06
Q9Z2I8;A0A0N4SWD1                                     -1.577885e+07

                                                          Cluster 2  \
 Protein IDs
A0A0J9YTY0;A0A0J9YUL3;Q8C1B7;A0A0J9YUV6;A0A0J9YVA6     5.557197e+05
A0A0J9YUZ4;P63158;A0A0J9YUD8;D3YVC6;D3YZ18             2.468034e+06
A0A0R4J093;Q9DBP5                                     -8.206000e+05
A0A0R4J0I1;P07759;E9Q499;Q80X76                        5.144795e+06
O70404;A0A0R4J0R1;A0A0U1RPE8                           8.577997e+05
...                                                             ...
Q9Z0P4;Q3UZP7;A0A1W2P879;A0A1W2P831                   -9.870167e+04
Q9Z1D1                                                 2.961831e+05
Q9Z1Q5                                                 1.874555e+06
Q9Z2D6                                                -4.070078e+05
Q9Z2I8;A0A0N4SWD1                                     -8.384045e+06

                                                          Cluster 3     Cluster 4
 Protein IDs
A0A0J9YTY0;A0A0J9YUL3;Q8C1B7;A0A0J9YUV6;A0A0J9YVA6 -4.514309e+06 -3.775957e+06
A0A0J9YUZ4;P63158;A0A0J9YUD8;D3YVC6;D3YZ18         -6.670144e+06 -8.216969e+06
A0A0R4J093;Q9DBP5                                   2.444753e+06 -3.148042e+05
A0A0R4J0I1;P07759;E9Q499;Q80X76                     1.707926e+07 -8.599288e+06
O70404;A0A0R4J0R1;A0A0U1RPE8                       -2.495012e+06 -2.381546e+06
...                                                          ...           ...
Q9Z0P4;Q3UZP7;A0A1W2P879;A0A1W2P831                 9.093900e+05  5.358425e+05
Q9Z1D1                                             -1.005512e+06 -4.893086e+05
Q9Z1Q5                                             -6.734190e+06 -2.573627e+06
Q9Z2D6                                              2.729420e+06  8.867200e+05
Q9Z2I8;A0A0N4SWD1                                   2.848534e+07  1.683961e+07

[287 rows x 4 columns]
```

[188]:
```
#make temp df of pGroups_proc with only proteins present in all samples
temp = pGroups_proc.copy()
temp = temp.set_index(' Protein IDs')
temp = temp.loc[centers_t.index]

#plot only mouse markers; should be all proteins since this is the dataset with␣
 ↪proteins present in all amples
#first pull only mouse index then pull centers_t based on that index

mouse_idx = temp.index[temp['Fasta headers'].str.contains('musculus',␣
 ↪regex=False, na=False)].tolist()
centers_t_mouse = centers_t.loc[mouse_idx]
```

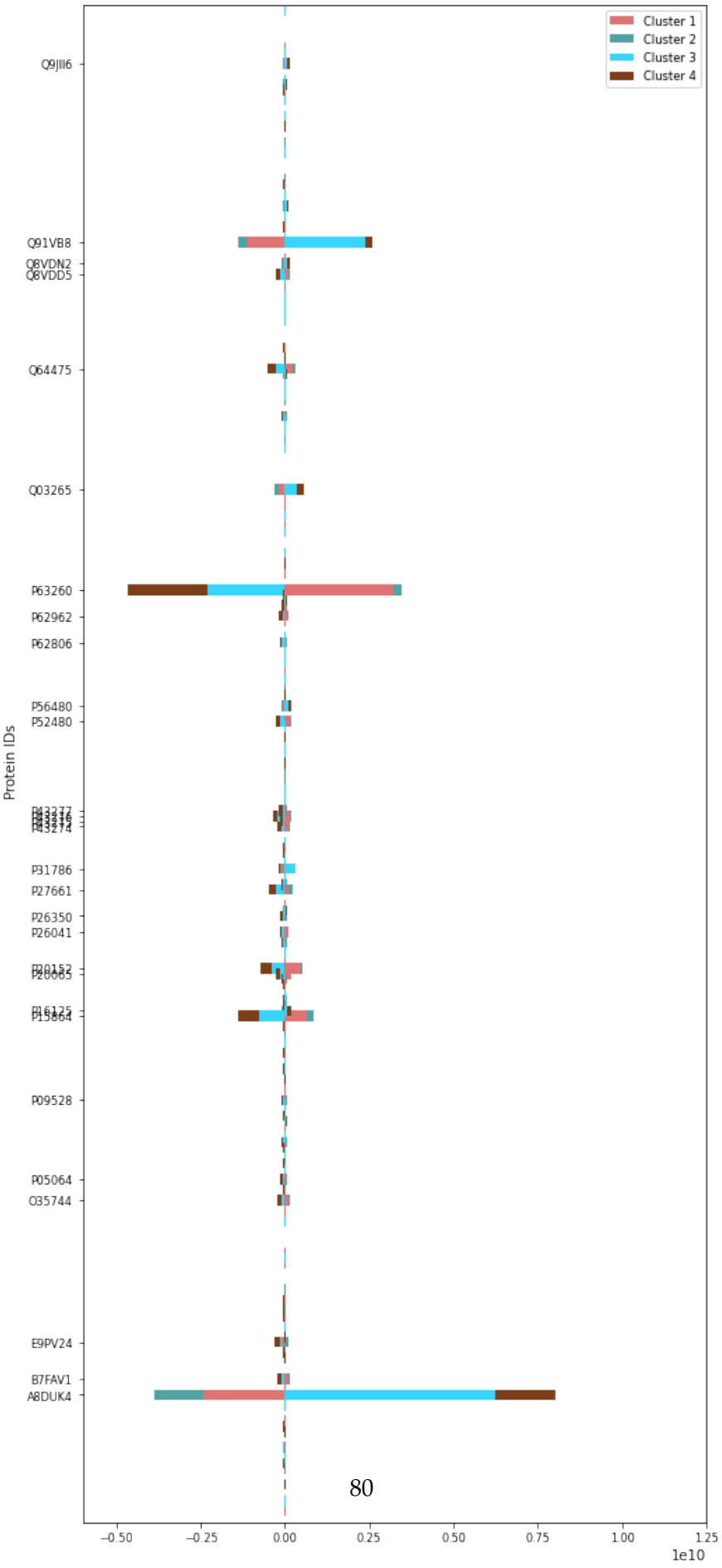
```
centers_t_mouse
centers_t_mouse.index = centers_t_mouse.index.str.split(';',0).str[0]
centers_t_mouse.to_csv('outs/' + date + '_mouse_centers_annotated.csv')
centers_t_mouse
```

[188]:

Protein IDs	Cluster 1	Cluster 2	Cluster 3	Cluster 4
A0A0J9YTY0	5.347031e+06	5.557197e+05	-4.514309e+06	-3.775957e+06
A0A0J9YUZ4	6.832731e+06	2.468034e+06	-6.670144e+06	-8.216969e+06
A0A0R4J093	6.331833e+05	-8.206000e+05	2.444753e+06	-3.148042e+05
A0A0R4J0I1	-1.365104e+07	5.144795e+06	1.707926e+07	-8.599288e+06
O70404	2.010879e+06	8.577997e+05	-2.495012e+06	-2.381546e+06
...	...	...	...	...
Q9Z0P4	-8.821325e+05	-9.870167e+04	9.093900e+05	5.358425e+05
Q9Z1D1	4.513414e+05	2.961831e+05	-1.005512e+06	-4.893086e+05
Q9Z1Q5	2.564748e+06	1.874555e+06	-6.734190e+06	-2.573627e+06
Q9Z2D6	-1.676840e+06	-4.070078e+05	2.729420e+06	8.867200e+05
Q9Z2I8	-1.577885e+07	-8.384045e+06	2.848534e+07	1.683961e+07

[287 rows x 4 columns]

[189]:
```
#plot centroids with top 10% proteins labeled

summed = centers_t_mouse.abs().sum(axis=1)
sorted_mylist = sorted(((v, i) for i, v in enumerate(summed)), reverse=True)[0:
 ↪29]
final = [lis[1] for lis in sorted_mylist]
finalID = centers_t_mouse.iloc[final].index

centers_t_mouse.to_csv('outs/' + date + '_finalID_mouse_all.csv')

centers_t_mouse.plot(kind="barh", stacked=True, width=2, figsize = (8,20),␣
 ↪color=palette_kmeans)
plt.tick_params(axis='both', which='major', labelsize=8)
plt.yticks(final, finalID)
plt.xlim(-0.60e10, 1.25e10)
plt.legend(fontsize='small')
plt.savefig('outs/' + date + '_mouse_kmeansCentroids_all_top29.svg')
```


Cluster 1
Cluster 2
Cluster 3
Cluster 4
Protein IDs
Q9JII6
Q91VB8
Q8VDN2
Q8VDD5
Q64475
Q03265
P63260
P62962
P62806
P56480
P52480
P43277
P43276
P43275
P43274
P31786
P27661
P26350
P26041
P20152
P20065
P16125
P15864
P09528
P05064
O35744
E9PV24
B7FAV1
A8DUK4
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
1.25
1e10

0.4 Gene Ontology Enrichment Analysis

0.4.1 Extract top 100 proteins driving the clustering and run gene ontology enrichment analysis using their LFQ intensity values after normalization

[190]:
```
allProt = pd.read_csv('outs/' + date + '_finalID_mouse_all.csv', index_col = 0)
allProt.shape
```

[190]:
```
(287, 4)
```

[191]:
```
summed = allProt.abs().sum(axis=1)
sorted_mylist = sorted(((v, i) for i, v in enumerate(summed)), reverse=True)[0:
 ↪100]
final = [lis[1] for lis in sorted_mylist]
finalID = allProt.iloc[final].index
finalID
```

[191]:
```
Index(['A8DUK4', 'P63260', 'Q91VB8', 'P15864', 'P20152', 'Q03265', 'Q64475',
       'P27661', 'P43276', 'P31786', 'P20065', 'P52480', 'Q8VDD5', 'E9PV24',
       'B7FAV1', 'P43274', 'O35744', 'P62962', 'P56480', 'P16125', 'P26041',
       'P43277', 'Q8VDN2', 'P62806', 'P05064', 'P43275', 'Q9JII6', 'P26350',
       'P09528', 'P27773', 'Q61233', 'Q99KI0', 'P20029', 'P63017', 'P07356',
       'P63038', 'P26039', 'Q9DCW4', 'Q64433', 'P18760', 'D3Z2H9', 'P26443',
       'O88569', 'P17742', 'A0A1D5RLD8', 'A0A1W2P768', 'P11499', 'Q99PT1',
       'P08249', 'P40124', 'E9Q7Q3', 'P40142', 'E9Q3W4', 'Q91Z25', 'Q6IRU2',
       'P13020', 'A2AL12', 'Q9Z2I8', 'E9Q616', 'Q61838', 'P63101', 'P70441',
       'P24270', 'P14733', 'P06745', 'P17182', 'Q9DCD0', 'P05213', 'P38647',
       'P26645', 'Q8BWT1', 'P45952', 'Q9JKF1', 'P99029', 'A3KGU5', 'P08113',
       'P10922', 'A0A1W2P6F6', 'O08749', 'P09103', 'Q61029', 'P17751',
       'P05202', 'Q921I1', 'Q62433', 'P99024', 'P14152', 'O35639',
       'A0A0R4J0I1', 'Q60932', 'P61979', 'Q9DCX2', 'A0A1L1SV25', 'F8WIP8',
       'P10126', 'P48036', 'Q8QZT1', 'E9PZF0', 'F8WIT2', 'P09405'],
      dtype='object', name=' Protein IDs')
```

[192]:
```
#import lfq intensity data
lfq = pd.read_csv('outs/' + date + '_combined_lfq.csv', index_col=0)
lfq = lfq.transpose()
lfq = lfq.drop(['TimePoint', 'TechRep', 'BioRep', 'Region',␣
 ↪'DPI_Region_BioRep', 'DPI_Region', 'BioRep_TechRep'], axis=0)
lfq
```

[192]:
```
index                                                 4DPI_interface_2_4  \
G3UZW7;A0A023T778;Q9CQL1;P61327                                        0
O70589;A0A067XG53;F6Y9I5                                               0
A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0A...                     0
A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1                            0
A0A075B5P3;A0A0A6YVP0;P01867                                           0
```

```
...                                                          ...
Q9Z2X1;J3QMT0;J3QM80;J3QP45;J3QMV8;J3QNH2                5.246e+06
Q9Z2Z6                                                           0
Q9Z315;A0A494B9E9                                           954220
S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5                        0
S4R192                                                           0

index                                              4DPI_interface_2_2  \
G3UZW7;A0A023T778;Q9CQL1;P61327                                   0
O70589;A0A067XG53;F6Y9I5                                         0
A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0A...               0
A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1                      0
A0A075B5P3;A0A0A6YVP0;P01867                                     0
...                                                            ...
Q9Z2X1;J3QMT0;J3QM80;J3QP45;J3QMV8;J3QNH2              3.0364e+06
Q9Z2Z6                                                    668000
Q9Z315;A0A494B9E9                                      1.3774e+06
S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5                         0
S4R192                                                           0

index                                              4DPI_interface_2_1  \
G3UZW7;A0A023T778;Q9CQL1;P61327                                   0
O70589;A0A067XG53;F6Y9I5                                         0
A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0A...               0
A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1                      0
A0A075B5P3;A0A0A6YVP0;P01867                                     0
...                                                            ...
Q9Z2X1;J3QMT0;J3QM80;J3QP45;J3QMV8;J3QNH2              3.9714e+06
Q9Z2Z6                                                           0
Q9Z315;A0A494B9E9                                           922440
S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5                         0
S4R192                                                           0

index                                              4DPI_interface_1_2  \
G3UZW7;A0A023T778;Q9CQL1;P61327                                   0
O70589;A0A067XG53;F6Y9I5                                         0
A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0A...               0
A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1                      0
A0A075B5P3;A0A0A6YVP0;P01867                                     0
...                                                            ...
Q9Z2X1;J3QMT0;J3QM80;J3QP45;J3QMV8;J3QNH2              9.2853e+06
Q9Z2Z6                                                 1.4414e+06
Q9Z315;A0A494B9E9                                                0
S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5                         0
S4R192                                                           0

index                                              4DPI_interface_1_1  \
```

```
G3UZW7;A0A023T778;Q9CQL1;P61327                                    0
O70589;A0A067XG53;F6Y9I5                                           0
A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0A...                 0
A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1                        0
A0A075B5P3;A0A0A6YVP0;P01867                                       0
...                                                              ...
Q9Z2X1;J3QMT0;J3QM80;J3QP45;J3QMV8;J3QNH2                  6.0399e+06
Q9Z2Z6                                                     1.1892e+06
Q9Z315;A0A494B9E9                                          1.1176e+06
S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5                           0
S4R192                                                             0

index                                              4DPI_cortex_3_3  \
G3UZW7;A0A023T778;Q9CQL1;P61327                                   0
O70589;A0A067XG53;F6Y9I5                                          0
A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0A...                0
A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1                       0
A0A075B5P3;A0A0A6YVP0;P01867                                      0
...                                                             ...
Q9Z2X1;J3QMT0;J3QM80;J3QP45;J3QMV8;J3QNH2              2.6449e+06
Q9Z2Z6                                                 1.3931e+06
Q9Z315;A0A494B9E9                                          638210
S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5                      0
S4R192                                                            0

index                                              4DPI_cortex_3_2  \
G3UZW7;A0A023T778;Q9CQL1;P61327                                   0
O70589;A0A067XG53;F6Y9I5                                          0
A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0A...                0
A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1                       0
A0A075B5P3;A0A0A6YVP0;P01867                                      0
...                                                             ...
Q9Z2X1;J3QMT0;J3QM80;J3QP45;J3QMV8;J3QNH2                         0
Q9Z2Z6                                                    950880
Q9Z315;A0A494B9E9                                         661230
S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5                      0
S4R192                                                            0

index                                              4DPI_cortex_2_3  \
G3UZW7;A0A023T778;Q9CQL1;P61327                                   0
O70589;A0A067XG53;F6Y9I5                                          0
A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0A...                0
A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1                       0
A0A075B5P3;A0A0A6YVP0;P01867                                      0
...                                                             ...
Q9Z2X1;J3QMT0;J3QM80;J3QP45;J3QMV8;J3QNH2              2.4288e+06
Q9Z2Z6                                                 1.0941e+06
```

```
Q9Z315;A0A494B9E9                                                    0
S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5                        662690
S4R192                                                               0

index                                                  4DPI_cortex_2_2  \
G3UZW7;A0A023T778;Q9CQL1;P61327                                      0
O70589;A0A067XG53;F6Y9I5                                             0
A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0A...                   0
A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1                          0
A0A075B5P3;A0A0A6YVP0;P01867                                         0
...                                                                ...
Q9Z2X1;J3QMT0;J3QM80;J3QP45;J3QMV8;J3QNH2                    2.6804e+06
Q9Z2Z6                                                       1.1699e+06
Q9Z315;A0A494B9E9                                                494860
S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5                             0
S4R192                                                               0

index                                                  4DPI_cortex_2_1  ...  \
G3UZW7;A0A023T778;Q9CQL1;P61327                                      0  ...
O70589;A0A067XG53;F6Y9I5                                        497950  ...
A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0A...                   0  ...
A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1                          0  ...
A0A075B5P3;A0A0A6YVP0;P01867                                         0  ...
...                                                                ...  ...
Q9Z2X1;J3QMT0;J3QM80;J3QP45;J3QMV8;J3QNH2                    2.0776e+06  ...
Q9Z2Z6                                                       1.1146e+06  ...
Q9Z315;A0A494B9E9                                                674540  ...
S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5                             0  ...
S4R192                                                               0  ...

index                                                 10DPI_cortex_3_2  \
G3UZW7;A0A023T778;Q9CQL1;P61327                                      0
O70589;A0A067XG53;F6Y9I5                                        434700
A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0A...              498950
A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1                          0
A0A075B5P3;A0A0A6YVP0;P01867                                         0
...                                                                ...
Q9Z2X1;J3QMT0;J3QM80;J3QP45;J3QMV8;J3QNH2                    2.1112e+06
Q9Z2Z6                                                               0
Q9Z315;A0A494B9E9                                                556170
S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5                             0
S4R192                                                               0

index                                                  10DPI_SAC_4_3  \
G3UZW7;A0A023T778;Q9CQL1;P61327                                      0
O70589;A0A067XG53;F6Y9I5                                             0
A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0A...           1.6741e+06
```

```
A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1              202930
A0A075B5P3;A0A0A6YVP0;P01867                          1.2076e+06
...                                                          ...
Q9Z2X1;J3QMT0;J3QM80;J3QP45;J3QMV8;J3QNH2             6.5735e+06
Q9Z2Z6                                                   338860
Q9Z315;A0A494B9E9                                     1.1378e+06
S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5                      0
S4R192                                                        0

index                                                10DPI_SAC_4_2  \
G3UZW7;A0A023T778;Q9CQL1;P61327                               0
O70589;A0A067XG53;F6Y9I5                                      0
A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0A...        431080
A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1               290690
A0A075B5P3;A0A0A6YVP0;P01867                          1.5261e+06
...                                                          ...
Q9Z2X1;J3QMT0;J3QM80;J3QP45;J3QMV8;J3QNH2              6.941e+06
Q9Z2Z6                                                        0
Q9Z315;A0A494B9E9                                     1.3022e+06
S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5                      0
S4R192                                                        0

index                                                10DPI_SAC_4_1  \
G3UZW7;A0A023T778;Q9CQL1;P61327                               0
O70589;A0A067XG53;F6Y9I5                                      0
A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0A...        548100
A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1                   0
A0A075B5P3;A0A0A6YVP0;P01867                          1.1656e+06
...                                                          ...
Q9Z2X1;J3QMT0;J3QM80;J3QP45;J3QMV8;J3QNH2             6.2944e+06
Q9Z2Z6                                                        0
Q9Z315;A0A494B9E9                                             0
S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5                      0
S4R192                                                   326950

index                                                10DPI_SAC_3_3  \
G3UZW7;A0A023T778;Q9CQL1;P61327                               0
O70589;A0A067XG53;F6Y9I5                                      0
A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0A...            0
A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1                   0
A0A075B5P3;A0A0A6YVP0;P01867                                  0
...                                                          ...
Q9Z2X1;J3QMT0;J3QM80;J3QP45;J3QMV8;J3QNH2             4.6115e+06
Q9Z2Z6                                                        0
Q9Z315;A0A494B9E9                                     1.7312e+06
S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5                      0
S4R192                                                   423330
```

```
index                                                 10DPI_SAC_3_2  \
G3UZW7;A0A023T778;Q9CQL1;P61327                                   0
O70589;A0A067XG53;F6Y9I5                                          0
A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0A...                0
A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1                       0
A0A075B5P3;A0A0A6YVP0;P01867                                      0
...                                                             ...
Q9Z2X1;J3QMT0;J3QM80;J3QP45;J3QMV8;J3QNH2                 6.5767e+06
Q9Z2Z6                                                            0
Q9Z315;A0A494B9E9                                         1.4447e+06
S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5                          0
S4R192                                                            0

index                                                 10DPI_SAC_3_1  \
G3UZW7;A0A023T778;Q9CQL1;P61327                                   0
O70589;A0A067XG53;F6Y9I5                                          0
A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0A...                0
A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1                       0
A0A075B5P3;A0A0A6YVP0;P01867                                 724260
...                                                             ...
Q9Z2X1;J3QMT0;J3QM80;J3QP45;J3QMV8;J3QNH2                 5.9742e+06
Q9Z2Z6                                                            0
Q9Z315;A0A494B9E9                                         1.0595e+06
S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5                          0
S4R192                                                            0

index                                                 10DPI_SAC_2_3  \
G3UZW7;A0A023T778;Q9CQL1;P61327                                   0
O70589;A0A067XG53;F6Y9I5                                          0
A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0A...                0
A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1                       0
A0A075B5P3;A0A0A6YVP0;P01867                                      0
...                                                             ...
Q9Z2X1;J3QMT0;J3QM80;J3QP45;J3QMV8;J3QNH2                 2.8952e+06
Q9Z2Z6                                                            0
Q9Z315;A0A494B9E9                                         3.3433e+06
S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5                          0
S4R192                                                            0

index                                                 10DPI_SAC_2_2 10DPI_SAC_2_1
G3UZW7;A0A023T778;Q9CQL1;P61327                                   0             0
O70589;A0A067XG53;F6Y9I5                                          0             0
A0A140T8M9;A0A140T8M0;A0A0B4J1I0;A0A075B5N0;A0A...                0             0
A0A075B5M7;A0A0G2JDV4;A0A0B4J1J2;A0A0B4J1J1                       0             0
A0A075B5P3;A0A0A6YVP0;P01867                                      0             0
...                                                             ...           ...
```

```
Q9Z2X1;J3QMT0;J3QM80;J3QP45;J3QMV8;J3QNH2              3.8473e+06    6.8296e+06
Q9Z2Z6                                                         0             0
Q9Z315;A0A494B9E9                                         800330    2.3382e+06
S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5                       0             0
S4R192                                                         0             0

[3613 rows x 39 columns]
```

[193]:
```
#for ID's with multiple proteins per protein group, select first protein in␣
 ↪protein group for analysis
lfq.index = lfq.index.str.split(';',0).str[0]
lfq
```

[193]:
```
index      4DPI_interface_2_4 4DPI_interface_2_2 4DPI_interface_2_1  \
G3UZW7                      0                  0                  0
O70589                      0                  0                  0
A0A140T8M9                  0                  0                  0
A0A075B5M7                  0                  0                  0
A0A075B5P3                  0                  0                  0
...                       ...                ...                ...
Q9Z2X1              5.246e+06          3.0364e+06         3.9714e+06
Q9Z2Z6                      0              668000                  0
Q9Z315                 954220          1.3774e+06             922440
S4R2J6                      0                  0                  0
S4R192                      0                  0                  0

index      4DPI_interface_1_2 4DPI_interface_1_1 4DPI_cortex_3_3  \
G3UZW7                      0                  0               0
O70589                      0                  0               0
A0A140T8M9                  0                  0               0
A0A075B5M7                  0                  0               0
A0A075B5P3                  0                  0               0
...                       ...                ...             ...
Q9Z2X1             9.2853e+06         6.0399e+06      2.6449e+06
Q9Z2Z6             1.4414e+06         1.1892e+06      1.3931e+06
Q9Z315                      0         1.1176e+06          638210
S4R2J6                      0                  0               0
S4R192                      0                  0               0

index      4DPI_cortex_3_2 4DPI_cortex_2_3 4DPI_cortex_2_2 4DPI_cortex_2_1  \
G3UZW7                   0               0               0               0
O70589                   0               0               0          497950
A0A140T8M9               0               0               0               0
A0A075B5M7               0               0               0               0
A0A075B5P3               0               0               0               0
...                    ...             ...             ...             ...
Q9Z2X1                   0       2.4288e+06      2.6804e+06      2.0776e+06
Q9Z2Z6              950880       1.0941e+06      1.1699e+06      1.1146e+06
```

```
Q9Z315              661230               0          494860          674540
S4R2J6                   0          662690               0               0
S4R192                   0               0               0               0

index       ... 10DPI_cortex_3_2 10DPI_SAC_4_3 10DPI_SAC_4_2 10DPI_SAC_4_1  \
G3UZW7      ...                0            0             0             0
O70589      ...           434700            0             0             0
A0A140T8M9  ...           498950    1.6741e+06        431080        548100
A0A075B5M7  ...                0       202930        290690             0
A0A075B5P3  ...                0    1.2076e+06    1.5261e+06    1.1656e+06
...         ...              ...          ...           ...           ...
Q9Z2X1      ...       2.1112e+06    6.5735e+06     6.941e+06    6.2944e+06
Q9Z2Z6      ...                0       338860             0             0
Q9Z315      ...           556170    1.1378e+06    1.3022e+06             0
S4R2J6      ...                0            0             0             0
S4R192      ...                0            0             0        326950

index      10DPI_SAC_3_3 10DPI_SAC_3_2 10DPI_SAC_3_1 10DPI_SAC_2_3  \
G3UZW7                 0             0             0             0
O70589                 0             0             0             0
A0A140T8M9             0             0             0             0
A0A075B5M7             0             0             0             0
A0A075B5P3             0             0        724260             0
...                  ...           ...           ...           ...
Q9Z2X1        4.6115e+06    6.5767e+06    5.9742e+06    2.8952e+06
Q9Z2Z6                 0             0             0             0
Q9Z315        1.7312e+06    1.4447e+06    1.0595e+06    3.3433e+06
S4R2J6                 0             0             0             0
S4R192            423330             0             0             0

index      10DPI_SAC_2_2 10DPI_SAC_2_1
G3UZW7                 0             0
O70589                 0             0
A0A140T8M9             0             0
A0A075B5M7             0             0
A0A075B5P3             0             0
...                  ...           ...
Q9Z2X1        3.8473e+06    6.8296e+06
Q9Z2Z6                 0             0
Q9Z315            800330    2.3382e+06
S4R2J6                 0             0
S4R192                 0             0

[3613 rows x 39 columns]
```

[194]:
```
#select top 100 proteins based on highest k-means centroid values
df = lfq.loc[finalID]
```

```
df
```

[194]:
```
index        4DPI_interface_2_4 4DPI_interface_2_2 4DPI_interface_2_1  \
 Protein IDs
A8DUK4               1.7442e+09         3.4139e+09         2.9423e+09
P63260               4.1307e+09          3.378e+09         2.9545e+09
Q91VB8               1.7297e+09         1.9171e+09         1.4639e+09
P15864               9.1421e+08         6.0707e+08          3.477e+08
P20152               4.5815e+08         4.2805e+08         3.6485e+08
...                         ...                ...                ...
P48036               5.4002e+07         4.5385e+07         6.3113e+07
Q8QZT1               1.1984e+07         1.7382e+07         1.5672e+07
E9PZF0               3.8056e+07         5.7943e+07         3.3903e+07
F8WIT2               2.2663e+07         1.4908e+07         3.6708e+07
P09405               6.0954e+07         4.9944e+07         5.7182e+07

index        4DPI_interface_1_2 4DPI_interface_1_1 4DPI_cortex_3_3  \
 Protein IDs
A8DUK4               5.1737e+09         3.8777e+09       5.782e+09
P63260                1.771e+09         1.6077e+09      1.2741e+09
Q91VB8               1.3123e+09         1.8905e+09      2.7786e+09
P15864               4.7936e+08         5.7903e+08      2.5711e+08
P20152                3.079e+08         3.2536e+08      5.2257e+07
...                         ...                ...             ...
P48036               5.3553e+07         5.4752e+07       1.362e+07
Q8QZT1                 8.09e+06         1.1016e+07      5.4671e+07
E9PZF0               2.0192e+07         3.1333e+07      6.8917e+07
F8WIT2               1.3592e+06         4.1288e+06      1.5675e+06
P09405               3.2236e+07         4.5901e+07      1.6173e+07

index        4DPI_cortex_3_2 4DPI_cortex_2_3 4DPI_cortex_2_2 4DPI_cortex_2_1  \
 Protein IDs
A8DUK4            6.8564e+09      8.5305e+09      8.4491e+09      3.4106e+09
P63260            1.8186e+09      1.5937e+09      2.1845e+09      1.5055e+09
Q91VB8            1.9416e+09      2.0437e+09      2.8438e+09      2.0836e+09
P15864             2.659e+08      3.1331e+08      2.7457e+08      3.3855e+08
P20152            7.7997e+07      1.2761e+08      1.2123e+08      3.4905e+07
...                      ...             ...             ...             ...
P48036            2.8752e+07      4.4965e+07      4.3174e+07      2.4549e+07
Q8QZT1            3.6034e+07      3.6496e+07      4.7283e+07      5.1152e+07
E9PZF0             6.386e+07        3.55e+07      5.2153e+07      5.4563e+07
F8WIT2            4.3106e+06      1.0905e+07      1.2193e+07       2.284e+06
P09405            1.7219e+07      3.9117e+07      3.0935e+07      7.6649e+06

index         ... 10DPI_cortex_3_2 10DPI_SAC_4_3 10DPI_SAC_4_2 10DPI_SAC_4_1  \
 Protein IDs  ...
A8DUK4        ...        8.122e+09    5.3066e+08    9.5197e+08    6.2291e+08
```

```
P63260        ...       1.3877e+09    4.8895e+09    4.6028e+09    4.8811e+09
Q91VB8        ...       5.3114e+09    3.0269e+08    4.4618e+08    3.8388e+08
P15864        ...       3.7537e+08    8.0293e+08    1.3161e+09    1.2945e+09
P20152        ...       5.8747e+07    7.1322e+08     8.472e+08    6.4955e+08
...           ...              ...           ...           ...           ...
P48036        ...       2.8541e+07    7.5051e+07    7.6192e+07    6.2647e+07
Q8QZT1        ...        2.434e+07    6.1046e+06    3.5123e+06    7.2743e+06
E9PZF0        ...       4.9142e+07    4.5855e+07    3.8078e+07    3.7067e+07
F8WIT2        ...       1.9329e+06    2.8045e+07    2.4982e+07    2.3535e+07
P09405        ...        2.367e+07    3.7192e+07     4.091e+07    4.1109e+07

index        10DPI_SAC_3_3 10DPI_SAC_3_2 10DPI_SAC_3_1 10DPI_SAC_2_3  \
 Protein IDs
A8DUK4          8.5632e+08    8.6326e+08    9.2512e+08    1.0979e+09
P63260           6.279e+09    6.4949e+09     7.117e+09    4.7165e+09
Q91VB8          1.0176e+09    8.2537e+08    6.0648e+08    1.3203e+09
P15864          2.0244e+09     1.426e+09    1.6338e+09    3.6126e+09
P20152          9.1203e+08    8.0969e+08    7.6861e+08    6.5963e+08
...                    ...           ...           ...           ...
P48036          6.6472e+07    9.4073e+07    7.7432e+07     5.382e+07
Q8QZT1          5.6125e+06        385710    9.5847e+06    6.7139e+06
E9PZF0          2.3838e+07    1.8537e+07    2.1536e+07    1.8987e+07
F8WIT2           2.866e+07    4.5345e+07    4.2387e+07    1.0955e+07
P09405          4.5695e+07    3.9475e+07    5.1029e+07    3.6213e+07

index        10DPI_SAC_2_2 10DPI_SAC_2_1
 Protein IDs
A8DUK4          1.2606e+09    1.8661e+09
P63260          4.8012e+09    4.7378e+09
Q91VB8          2.0021e+09    8.6326e+08
P15864          1.3788e+09    2.4812e+09
P20152          9.2283e+08    6.4163e+08
...                    ...           ...
P48036          4.0377e+07    5.0462e+07
Q8QZT1          1.0046e+07     2.324e+06
E9PZF0          2.9525e+07    2.8315e+07
F8WIT2          1.5971e+07    4.3004e+06
P09405          5.2582e+07    4.8543e+07

[100 rows x 39 columns]
```

[195]:
```
#transpose and change datatype to int64
df2 = df.transpose()
df2 = df2.astype('int64')
```

[196]:
```
#standardize values by removing the mean and scaling to unit variance
df3 = StandardScaler().fit_transform(df2)
df3 = pd.DataFrame(df3, columns=df2.columns, index=df2.index)
```

```
df3
```

[196]:

Protein IDs	A8DUK4	P63260	Q91VB8	P15864	P20152	\
index						
4DPI_interface_2_4	-0.540090	0.021508	-0.022581	-0.236924	-0.124103	
4DPI_interface_2_2	0.019054	-0.337894	0.113558	-0.642223	-0.213255	
4DPI_interface_2_1	-0.138874	-0.540108	-0.215675	-0.984486	-0.400444	
4DPI_interface_1_2	0.608371	-1.105209	-0.325806	-0.810748	-0.569122	
4DPI_interface_1_1	0.174370	-1.183182	0.094234	-0.679225	-0.517408	
4DPI_cortex_3_3	0.812077	-1.342470	0.739405	-1.104028	-1.326299	
4DPI_cortex_3_2	1.171869	-1.082481	0.131356	-1.092429	-1.250061	
4DPI_cortex_2_3	1.732487	-1.189867	0.205528	-1.029867	-1.103115	
4DPI_cortex_2_2	1.705228	-0.907770	0.786771	-1.080988	-1.122011	
4DPI_cortex_2_1	0.017949	-1.231981	0.234514	-0.996560	-1.377694	
4DPI_cortex_1_2	3.162245	-0.851188	2.487419	-0.892101	-1.028298	
4DPI_cortex_1_1	2.284197	-1.246783	2.586436	-1.133441	-1.165106	
4DPI_SAC_3_3	-0.653982	0.011098	-0.271394	1.321765	0.137459	
4DPI_SAC_3_2	-0.555428	0.006037	-0.709916	0.867429	0.003198	
4DPI_SAC_2_3	-0.512329	-0.250133	-0.485555	-0.434520	-0.619947	
4DPI_SAC_2_2	-0.480516	-0.370315	-0.212042	-0.525994	-0.592846	
4DPI_SAC_2_1	-0.371781	0.039127	-0.427148	0.498867	0.228388	
4DPI_SAC_1_3	0.162783	0.504720	0.867335	-0.745891	-0.339045	
4DPI_SAC_1_2	-0.281163	0.401154	2.496209	-0.308182	-0.292751	
4DPI_SAC_1_1	-0.043099	0.343665	0.096850	-0.764444	-0.312506	
10DPI_interface_5_3	-0.777652	1.419721	-0.753751	0.824938	1.692848	
10DPI_interface_4_3	-0.866100	1.606369	-1.090698	1.305138	1.805398	
10DPI_interface_4_2	-0.873578	2.688393	-1.119343	0.698258	1.568746	
10DPI_interface_4_1	-0.800096	1.695420	-0.998278	1.009681	1.820800	
10DPI_interface_3_2	-0.632316	0.429326	-0.690294	0.874819	1.296107	
10DPI_interface_3_1	-0.719082	1.182173	-0.586279	0.609317	1.189421	
10DPI_cortex_4_4	0.249718	-0.833378	-0.002022	-0.349340	-0.885863	
10DPI_cortex_4_3	1.169458	-1.045285	0.320527	-0.289708	-0.986684	
10DPI_cortex_3_3	0.492771	-1.152337	0.040476	-1.045689	-1.388084	
10DPI_cortex_3_2	1.595690	-1.288228	2.579389	-0.947973	-1.307077	
10DPI_SAC_4_3	-0.946477	0.383822	-1.059250	-0.383768	0.631378	
10DPI_SAC_4_2	-0.805390	0.246927	-0.955010	0.293407	1.028207	
10DPI_SAC_4_1	-0.915585	0.379811	-1.000268	0.264904	0.442797	
10DPI_SAC_3_3	-0.837421	1.047284	-0.539895	1.228074	1.220225	
10DPI_SAC_3_2	-0.835097	1.150373	-0.679543	0.438430	0.917108	
10DPI_SAC_3_1	-0.814382	1.447415	-0.838558	0.712641	0.795435	
10DPI_SAC_2_3	-0.756522	0.301217	-0.319995	3.323850	0.472652	
10DPI_SAC_2_2	-0.702037	0.341660	0.175307	0.376145	1.252213	
10DPI_SAC_2_1	-0.499269	0.311387	-0.652017	1.830863	0.419339	

Protein IDs	Q03265	Q64475	P27661	P43276	P31786	...	\
index						...	
4DPI_interface_2_4	-0.225207	0.487461	0.304723	-0.028911	0.134433	...	

```
4DPI_interface_2_2   0.089789 -0.433295 -0.113263 -0.522751 -0.504536  ...
4DPI_interface_2_1  -0.369529 -0.155440 -0.818871 -1.047645 -1.302654  ...
4DPI_interface_1_2  -0.241213  0.125845 -0.007414  0.002511 -1.015394  ...
4DPI_interface_1_1   0.168928 -1.043352 -0.340969 -0.393190 -0.514812  ...
4DPI_cortex_3_3      1.809980 -1.645795 -1.310597 -1.351087  0.551226  ...
4DPI_cortex_3_2      1.360204 -1.052557 -1.656457 -1.324128  0.710533  ...
4DPI_cortex_2_3      2.563874 -1.384593 -0.422396 -1.128870  2.363262  ...
4DPI_cortex_2_2      1.560213 -1.012349 -1.323757 -0.949139  1.326104  ...
4DPI_cortex_2_1      2.644484 -1.569062 -1.415339 -1.385625 -0.472758  ...
4DPI_cortex_1_2      0.906860 -0.699643 -0.763131 -0.753384  1.490523  ...
4DPI_cortex_1_1      0.987024 -1.281794 -0.206395 -1.246260  0.739623  ...
4DPI_SAC_3_3        -0.525400  1.972168  2.332589  0.926510 -0.777197  ...
4DPI_SAC_3_2        -0.485897  0.522564  0.658175  0.951647 -0.857984  ...
4DPI_SAC_2_3        -0.125960  0.362906 -0.387693  0.358201  1.684032  ...
4DPI_SAC_2_2        -0.046063  1.084758  0.075848  0.243042  1.603192  ...
4DPI_SAC_2_1        -0.187086 -0.117659  0.580671  0.241366 -0.301172  ...
4DPI_SAC_1_3        -0.217939 -0.801061  1.767597 -0.877708  0.019972  ...
4DPI_SAC_1_2        -0.500477 -0.780309  1.001093 -0.349095  0.899931  ...
4DPI_SAC_1_1        -0.382281 -0.266356  0.663553 -0.823244  0.018918  ...
10DPI_interface_5_3 -0.901580  1.013422 -0.382157  0.127778 -0.526985  ...
10DPI_interface_4_3 -0.827514  1.870624  0.372136  1.542423 -0.595809  ...
10DPI_interface_4_2 -0.759250  1.245463  0.559159  0.560974 -0.991206  ...
10DPI_interface_4_1 -0.808740  0.584194  0.445623  1.316398 -0.703630  ...
10DPI_interface_3_2 -0.899163  1.029865  0.112986  0.569039 -0.266707  ...
10DPI_interface_3_1 -0.726168  0.444367  0.102009  0.726513 -0.385806  ...
10DPI_cortex_4_4     0.463414 -0.368067  0.051774 -0.178477 -0.821938  ...
10DPI_cortex_4_3     0.495828 -0.910387  0.338825 -0.099504  0.317561  ...
10DPI_cortex_3_3     0.805517 -1.033478 -1.298924 -1.414690 -0.097914  ...
10DPI_cortex_3_2     1.836598 -1.234055 -1.243058 -1.222343  2.401152  ...
10DPI_SAC_4_3       -0.887629 -0.012392 -0.042908 -0.335688 -1.490924  ...
10DPI_SAC_4_2       -0.928505 -0.264724  0.373718  0.088606 -1.390982  ...
10DPI_SAC_4_1       -0.804905  0.156136 -0.493826 -0.204923 -1.472079  ...
10DPI_SAC_3_3       -0.772893  0.738915  0.963448  1.472510 -0.798119  ...
10DPI_SAC_3_2       -0.810746  0.665529  0.529107  0.355426 -0.462956  ...
10DPI_SAC_3_1       -0.729155  1.164043  0.115897  0.340239 -0.021449  ...
10DPI_SAC_2_3       -0.941055  2.246173  2.216491  2.602791  0.750162  ...
10DPI_SAC_2_2       -0.734416  0.680466  1.143132  0.986525  0.257537  ...
10DPI_SAC_2_1       -0.853945 -0.328529 -2.481400  2.224163  0.504851  ...

 Protein IDs           P61979    Q9DCX2  A0A1L1SV25    F8WIP8    P10126  \
index
4DPI_interface_2_4   0.573323 -0.334925    0.018748  0.107914  1.212955
4DPI_interface_2_2  -0.566791  0.321537    1.474015  0.968526 -0.854949
4DPI_interface_2_1  -0.193876  0.310361    0.307097 -0.023676 -0.441144
4DPI_interface_1_2  -1.254921 -0.812839    0.199176  1.497482  1.173070
4DPI_interface_1_1  -0.503401  0.292111    1.774868  0.328190 -1.647993
4DPI_cortex_3_3     -1.516123  2.398939    0.720059 -0.229218 -0.556935
```

4DPI_cortex_3_2	-0.977258	1.324569	0.731844	0.454642	0.938331
4DPI_cortex_2_3	-0.687040	2.349517	1.431181	2.344207	-0.239975
4DPI_cortex_2_2	-0.905276	2.138293	2.043236	3.298078	-0.636912
4DPI_cortex_2_1	-0.815326	0.759878	-0.269183	-0.294759	-1.810814
4DPI_cortex_1_2	0.065149	1.148796	-0.013677	2.515370	0.541643
4DPI_cortex_1_1	0.772643	1.394336	2.132254	0.542312	0.812029
4DPI_SAC_3_3	0.751048	-0.350119	-0.709544	0.106389	-0.141509
4DPI_SAC_3_2	-0.764715	-0.412290	-0.747891	0.087704	-1.318569
4DPI_SAC_2_3	-0.978095	0.036965	-1.984384	0.284609	-0.330962
4DPI_SAC_2_2	0.503350	-0.281836	-1.038753	-0.296284	0.070711
4DPI_SAC_2_1	0.015152	-0.096370	0.178716	0.518885	-0.455685
4DPI_SAC_1_3	-0.344091	-0.479700	0.296030	0.670741	0.446418
4DPI_SAC_1_2	-0.576277	-0.367495	-0.237596	-0.092830	0.151312
4DPI_SAC_1_1	0.586827	-0.592533	-0.200805	0.315036	1.908031
10DPI_interface_5_3	0.013980	-0.907178	-0.050588	-0.081766	2.083774
10DPI_interface_4_3	0.920569	-1.111181	0.140489	-1.168785	0.730058
10DPI_interface_4_2	1.032170	-0.882729	0.063615	-1.169293	0.702471
10DPI_interface_4_1	0.202920	-0.728393	-0.880639	-1.214748	-0.650622
10DPI_interface_3_2	0.118885	-0.491401	-1.009320	-0.610496	0.617134
10DPI_interface_3_1	-0.119885	-0.893251	-0.182499	-0.552520	1.172655
10DPI_cortex_4_4	-1.469306	0.694825	1.663895	-0.550092	0.606789
10DPI_cortex_4_3	0.562498	-0.194255	0.216405	-0.609141	-1.383216
10DPI_cortex_3_3	-2.290354	1.912836	0.919092	-1.265154	-1.059899
10DPI_cortex_3_2	-1.737204	0.966998	-0.815013	-1.216605	-1.192724
10DPI_SAC_4_3	1.407205	-0.724979	0.791728	-0.444244	-1.382800
10DPI_SAC_4_2	2.253920	-0.875848	-0.514998	-0.913915	-0.916771
10DPI_SAC_4_1	2.103817	-0.870530	-0.130691	-0.379494	-0.759516
10DPI_SAC_3_3	1.343202	-0.625610	-1.035522	-0.322308	0.767741
10DPI_SAC_3_2	0.129320	-1.125903	-0.712475	-0.495334	1.261565
10DPI_SAC_3_1	0.764775	-0.839165	-0.457866	-0.619811	1.450602
10DPI_SAC_2_3	0.705905	-0.551101	-1.993058	-0.614956	-0.233784
10DPI_SAC_2_2	0.323393	-0.616799	-0.699494	-0.551729	0.151312
10DPI_SAC_2_1	0.549887	-0.883532	-1.418453	-0.322929	-0.783821

Protein IDs	P48036	Q8QZT1	E9PZF0	F8WIT2	P09405
index					
4DPI_interface_2_4	-0.024608	-0.362371	-0.030764	0.333282	1.597443
4DPI_interface_2_2	-0.503046	0.048106	1.395860	-0.255785	0.741501
4DPI_interface_2_1	0.481258	-0.081927	-0.328686	1.400135	1.304200
4DPI_interface_1_2	-0.049537	-0.658481	-1.312265	-1.284947	-0.635158
4DPI_interface_1_1	0.017034	-0.435981	-0.513049	-1.074569	0.427190
4DPI_cortex_3_3	-2.266719	2.883655	2.183097	-1.269124	-1.883931
4DPI_cortex_3_2	-1.426552	1.466451	1.820325	-1.060759	-1.802613
4DPI_cortex_2_3	-0.526365	1.501583	-0.214122	-0.559851	-0.100214
4DPI_cortex_2_2	-0.625806	2.321853	0.980506	-0.462015	-0.736300
4DPI_cortex_2_1	-1.659913	2.616062	1.153391	-1.214699	-2.545370
4DPI_cortex_1_2	-1.030787	0.707471	1.691200	-0.514503	-0.741976

```
4DPI_cortex_1_1     -0.595879  0.885410  2.194431 -0.125894 -0.369901
4DPI_SAC_3_3        -1.048333 -0.082839 -0.089157 -0.533417  1.397490
4DPI_SAC_3_2        -0.713698 -0.405868 -0.592748 -0.343139  0.263853
4DPI_SAC_2_3         0.411022 -0.084284  0.319239 -0.942338 -1.200577
4DPI_SAC_2_2        -0.711533  0.207415 -0.308169 -0.712401  0.020676
4DPI_SAC_2_1        -0.634745 -0.365261  0.010556 -0.392664  1.233921
4DPI_SAC_1_3        -0.324874  0.654773  0.692699  0.203771  0.307000
4DPI_SAC_1_2        -0.207055 -0.169680 -0.464053 -0.223654 -0.654438
4DPI_SAC_1_1        -0.554571  0.140422  0.212710  0.059371  0.638882
10DPI_interface_5_3  0.992343 -0.835014 -0.679406  2.108762  1.231277
10DPI_interface_4_3  0.235904 -0.576910 -0.186289  1.100246  1.105335
10DPI_interface_4_2  0.341730 -0.686738 -0.772089  0.692647  0.991598
10DPI_interface_4_1  0.772141 -0.715809 -1.494046  0.663326  1.152602
10DPI_interface_3_2  1.074627 -0.620855 -1.031775  1.966034 -0.021150
10DPI_interface_3_1  1.524249 -0.677940 -1.059968  1.545065  0.638104
10DPI_cortex_4_4     1.000782 -0.770248 -0.080477 -0.645693 -0.976057
10DPI_cortex_4_3     2.065094  0.627398  0.954609 -0.752614 -1.150666
10DPI_cortex_3_3    -0.488110  0.434782  0.854680 -0.998138 -1.193425
10DPI_cortex_3_2    -1.438267  0.577210  0.764507 -1.241369 -1.301098
10DPI_SAC_4_3        1.144086 -0.809456  0.528709  0.742096 -0.249867
10DPI_SAC_4_2        1.207437 -1.006581 -0.029186  0.509432  0.039178
10DPI_SAC_4_1        0.455384 -0.720509 -0.101711  0.399519  0.054649
10DPI_SAC_3_3        0.667758 -0.846876 -1.050714  0.788812  0.411175
10DPI_SAC_3_2        2.200236 -1.244334 -1.430989  2.056198 -0.072382
10DPI_SAC_3_1        1.276285 -0.544820 -1.215851  1.831509  0.825852
10DPI_SAC_2_3       -0.034713 -0.763123 -1.398708 -0.556053 -0.325977
10DPI_SAC_2_2       -0.781103 -0.509742 -0.642748 -0.175040  0.946585
10DPI_SAC_2_1       -0.221158 -1.096942 -0.729549 -1.061534  0.632585

[39 rows x 100 columns]
```

[197]:
```
#load in kmeans results
kmeans = pd.read_csv('outs/' + date + '_combined_df_segm_pca_kmeans.csv',␣
 ↪index_col=0)
kmeans
```

[197]:
```
                    A0A0J9YTY0;A0A0J9YUL3;Q8C1B7;A0A0J9YUV6;A0A0J9YVA6  \
index
4DPI_interface_2_4                                                8884400.0
4DPI_interface_2_2                                                4805100.0
4DPI_interface_2_1                                                4679400.0
4DPI_interface_1_2                                                2752000.0
4DPI_interface_1_1                                                6446300.0
4DPI_cortex_3_3                                                   3966100.0
4DPI_cortex_3_2                                                   3174000.0
4DPI_cortex_2_3                                                   3505600.0
4DPI_cortex_2_2                                                   3623400.0
4DPI_cortex_2_1                                                   5408800.0
```

```
4DPI_cortex_1_2                                                  3500500.0
4DPI_cortex_1_1                                                  4027400.0
4DPI_SAC_3_3                                                    10933000.0
4DPI_SAC_3_2                                                     9909200.0
4DPI_SAC_2_3                                                     7513200.0
4DPI_SAC_2_2                                                     5452600.0
4DPI_SAC_2_1                                                    11385000.0
4DPI_SAC_1_3                                                     8059300.0
4DPI_SAC_1_2                                                     5600200.0
4DPI_SAC_1_1                                                     6698000.0
10DPI_interface_5_3                                             14004000.0
10DPI_interface_4_3                                             11083000.0
10DPI_interface_4_2                                             10662000.0
10DPI_interface_4_1                                             10168000.0
10DPI_interface_3_2                                             10654000.0
10DPI_interface_3_1                                             13143000.0
10DPI_cortex_4_4                                                 4613200.0
10DPI_cortex_4_3                                                 7330400.0
10DPI_cortex_3_3                                                 2525300.0
10DPI_cortex_3_2                                                 4286400.0
10DPI_SAC_4_3                                                    9362000.0
10DPI_SAC_4_2                                                   10750000.0
10DPI_SAC_4_1                                                   10526000.0
10DPI_SAC_3_3                                                   22060000.0
10DPI_SAC_3_2                                                   16201000.0
10DPI_SAC_3_1                                                   11879000.0
10DPI_SAC_2_3                                                   10519000.0
10DPI_SAC_2_2                                                   11053000.0
10DPI_SAC_2_1                                                   12673000.0

                    A0A0J9YUZ4;P63158;A0A0J9YUD8;D3YVC6;D3YZ18  \
index
4DPI_interface_2_4                                     13795000.0
4DPI_interface_2_2                                     16043000.0
4DPI_interface_2_1                                     14321000.0
4DPI_interface_1_2                                      2559000.0
4DPI_interface_1_1                                     18740000.0
4DPI_cortex_3_3                                        10538000.0
4DPI_cortex_3_2                                         5800300.0
4DPI_cortex_2_3                                         7019000.0
4DPI_cortex_2_2                                         9811200.0
4DPI_cortex_2_1                                         7080900.0
4DPI_cortex_1_2                                        13352000.0
4DPI_cortex_1_1                                        15997000.0
4DPI_SAC_3_3                                           24605000.0
4DPI_SAC_3_2                                           18164000.0
4DPI_SAC_2_3                                           11153000.0
```

4DPI_SAC_2_2	7476200.0
4DPI_SAC_2_1	24354000.0
4DPI_SAC_1_3	15659000.0
4DPI_SAC_1_2	18231000.0
4DPI_SAC_1_1	17122000.0
10DPI_interface_5_3	19714000.0
10DPI_interface_4_3	41455000.0
10DPI_interface_4_2	18782000.0
10DPI_interface_4_1	26015000.0
10DPI_interface_3_2	28269000.0
10DPI_interface_3_1	18962000.0
10DPI_cortex_4_4	6571500.0
10DPI_cortex_4_3	12236000.0
10DPI_cortex_3_3	2883700.0
10DPI_cortex_3_2	3060800.0
10DPI_SAC_4_3	19296000.0
10DPI_SAC_4_2	26852000.0
10DPI_SAC_4_1	17329000.0
10DPI_SAC_3_3	20678000.0
10DPI_SAC_3_2	18474000.0
10DPI_SAC_3_1	22727000.0
10DPI_SAC_2_3	21473000.0
10DPI_SAC_2_2	33776000.0
10DPI_SAC_2_1	13833000.0

index	A0A0R4J093;Q9DBP5	A0A0R4J0I1;P07759;E9Q499;Q80X76	\
4DPI_interface_2_4	5749000.0	73676000.0	
4DPI_interface_2_2	10405000.0	50598000.0	
4DPI_interface_2_1	9228800.0	105190000.0	
4DPI_interface_1_2	6292300.0	22129000.0	
4DPI_interface_1_1	18973000.0	38062000.0	
4DPI_cortex_3_3	14700000.0	32753000.0	
4DPI_cortex_3_2	12382000.0	24879000.0	
4DPI_cortex_2_3	15484000.0	63258000.0	
4DPI_cortex_2_2	11188000.0	47764000.0	
4DPI_cortex_2_1	7899100.0	44351000.0	
4DPI_cortex_1_2	15600000.0	60586000.0	
4DPI_cortex_1_1	14481000.0	84161000.0	
4DPI_SAC_3_3	10172000.0	33907000.0	
4DPI_SAC_3_2	11552000.0	28193000.0	
4DPI_SAC_2_3	13766000.0	56852000.0	
4DPI_SAC_2_2	9574300.0	37292000.0	
4DPI_SAC_2_1	12936000.0	60285000.0	
4DPI_SAC_1_3	6236600.0	79677000.0	
4DPI_SAC_1_2	4541100.0	63271000.0	
4DPI_SAC_1_1	5351400.0	53311000.0	

10DPI_interface_5_3	12167000.0	16442000.0
10DPI_interface_4_3	11927000.0	20993000.0
10DPI_interface_4_2	12228000.0	37156000.0
10DPI_interface_4_1	9103100.0	37811000.0
10DPI_interface_3_2	12785000.0	17304000.0
10DPI_interface_3_1	13679000.0	23913000.0
10DPI_cortex_4_4	15905000.0	21923000.0
10DPI_cortex_4_3	3591100.0	28778000.0
10DPI_cortex_3_3	2846000.0	17751000.0
10DPI_cortex_3_2	8662600.0	16765000.0
10DPI_SAC_4_3	12920000.0	26820000.0
10DPI_SAC_4_2	12945000.0	18110000.0
10DPI_SAC_4_1	11294000.0	17675000.0
10DPI_SAC_3_3	11293000.0	17459000.0
10DPI_SAC_3_2	10115000.0	16184000.0
10DPI_SAC_3_1	9660300.0	20254000.0
10DPI_SAC_2_3	7141400.0	11286000.0
10DPI_SAC_2_2	11842000.0	14575000.0
10DPI_SAC_2_1	8280200.0	18280000.0

index	O70404;A0A0R4J0R1;A0A0U1RPE8	A0A0R4J0Z1;P08003 \
4DPI_interface_2_4	4863000.0	20661000.0
4DPI_interface_2_2	5048700.0	20709000.0
4DPI_interface_2_1	4421600.0	17844000.0
4DPI_interface_1_2	1316400.0	22085000.0
4DPI_interface_1_1	2682100.0	24030000.0
4DPI_cortex_3_3	2543300.0	7914900.0
4DPI_cortex_3_2	2405800.0	10756000.0
4DPI_cortex_2_3	3551100.0	9577000.0
4DPI_cortex_2_2	3017800.0	14510000.0
4DPI_cortex_2_1	3218600.0	9714800.0
4DPI_cortex_1_2	4396200.0	5588800.0
4DPI_cortex_1_1	559210.0	12618000.0
4DPI_SAC_3_3	6073100.0	24566000.0
4DPI_SAC_3_2	5000700.0	20284000.0
4DPI_SAC_2_3	7744500.0	14750000.0
4DPI_SAC_2_2	3349900.0	14702000.0
4DPI_SAC_2_1	7488800.0	22398000.0
4DPI_SAC_1_3	4729600.0	22668000.0
4DPI_SAC_1_2	2086200.0	16305000.0
4DPI_SAC_1_1	3328200.0	20802000.0
10DPI_interface_5_3	7822400.0	28551000.0
10DPI_interface_4_3	5549700.0	25417000.0
10DPI_interface_4_2	3494600.0	20562000.0
10DPI_interface_4_1	3959500.0	14898000.0
10DPI_interface_3_2	10313000.0	21001000.0

```
10DPI_interface_3_1                    5659300.0        17979000.0
10DPI_cortex_4_4                       6568100.0        16769000.0
10DPI_cortex_4_3                       3419900.0        11998000.0
10DPI_cortex_3_3                       2486500.0        10442000.0
10DPI_cortex_3_2                       3308800.0         5336900.0
10DPI_SAC_4_3                          6796900.0        23906000.0
10DPI_SAC_4_2                          7350200.0        19011000.0
10DPI_SAC_4_1                          4744200.0        19508000.0
10DPI_SAC_3_3                         12653000.0        20440000.0
10DPI_SAC_3_2                         11506000.0        19529000.0
10DPI_SAC_3_1                          9135600.0        22306000.0
10DPI_SAC_2_3                         12044000.0        13540000.0
10DPI_SAC_2_2                         11489000.0        15259000.0
10DPI_SAC_2_1                          6878200.0        18782000.0

E9QL31;A0A0R4J104;P98078;E9PX84;Q3TRE6;F6TQN9;Q9DCE6;E0CXT5;E0CZ53;E0CYJ2  \
index
4DPI_interface_2_4                                      10187000.0
4DPI_interface_2_2                                      10975000.0
4DPI_interface_2_1                                       8142400.0
4DPI_interface_1_2                                        704720.0
4DPI_interface_1_1                                       4963900.0
4DPI_cortex_3_3                                         26064000.0
4DPI_cortex_3_2                                         19354000.0
4DPI_cortex_2_3                                         18882000.0
4DPI_cortex_2_2                                         14360000.0
4DPI_cortex_2_1                                         29494000.0
4DPI_cortex_1_2                                          8878300.0
4DPI_cortex_1_1                                         10476000.0
4DPI_SAC_3_3                                             7031500.0
4DPI_SAC_3_2                                             5700800.0
4DPI_SAC_2_3                                             5403700.0
4DPI_SAC_2_2                                             4984100.0
4DPI_SAC_2_1                                             7089200.0
4DPI_SAC_1_3                                             9872000.0
4DPI_SAC_1_2                                             7354500.0
4DPI_SAC_1_1                                             9433100.0
10DPI_interface_5_3                                      8816700.0
10DPI_interface_4_3                                      6936800.0
10DPI_interface_4_2                                      3421900.0
10DPI_interface_4_1                                      3713000.0
10DPI_interface_3_2                                      5232300.0
10DPI_interface_3_1                                      6887000.0
10DPI_cortex_4_4                                         7816500.0
10DPI_cortex_4_3                                        14825000.0
10DPI_cortex_3_3                                        22599000.0
10DPI_cortex_3_2                                        25867000.0
```

```
10DPI_SAC_4_3                                                 5934200.0
10DPI_SAC_4_2                                                 2999000.0
10DPI_SAC_4_1                                                 4308500.0
10DPI_SAC_3_3                                                 7713200.0
10DPI_SAC_3_2                                                 9466900.0
10DPI_SAC_3_1                                                 8250600.0
10DPI_SAC_2_3                                                 6240900.0
10DPI_SAC_2_2                                                 3403200.0
10DPI_SAC_2_1                                                 8101900.0

                    A0A0R4J195;Q9CSU0;F6YB25;F7DAY5;Q8VDS4  \
index
4DPI_interface_2_4                                1394900.0
4DPI_interface_2_2                                 959390.0
4DPI_interface_2_1                                 549190.0
4DPI_interface_1_2                                1506600.0
4DPI_interface_1_1                                1400400.0
4DPI_cortex_3_3                                    734590.0
4DPI_cortex_3_2                                    544140.0
4DPI_cortex_2_3                                   1263200.0
4DPI_cortex_2_2                                    911490.0
4DPI_cortex_2_1                                   1163700.0
4DPI_cortex_1_2                                    857580.0
4DPI_cortex_1_1                                    813620.0
4DPI_SAC_3_3                                      2025800.0
4DPI_SAC_3_2                                      1299600.0
4DPI_SAC_2_3                                      1140400.0
4DPI_SAC_2_2                                      1866900.0
4DPI_SAC_2_1                                      1723900.0
4DPI_SAC_1_3                                      1446200.0
4DPI_SAC_1_2                                      2133800.0
4DPI_SAC_1_1                                      1344000.0
10DPI_interface_5_3                               1129300.0
10DPI_interface_4_3                               1920700.0
10DPI_interface_4_2                               1817400.0
10DPI_interface_4_1                               1819300.0
10DPI_interface_3_2                               1001900.0
10DPI_interface_3_1                               2526900.0
10DPI_cortex_4_4                                  1237700.0
10DPI_cortex_4_3                                  2119000.0
10DPI_cortex_3_3                                   881230.0
10DPI_cortex_3_2                                   797640.0
10DPI_SAC_4_3                                     1421000.0
10DPI_SAC_4_2                                     3697900.0
10DPI_SAC_4_1                                     2148600.0
10DPI_SAC_3_3                                     1626400.0
10DPI_SAC_3_2                                     1196300.0
```

```
10DPI_SAC_3_1                                          819090.0
10DPI_SAC_2_3                                         2583700.0
10DPI_SAC_2_2                                         1870900.0
10DPI_SAC_2_1                                         2521700.0

A0A1B0GQU8;P35980;A0A1B0GSS8;A0A1B0GSF7;A0A1B0GSA8;A0A1B0GS28;A0A1B0GRZ3  \
index
4DPI_interface_2_4                                              11981000.0
4DPI_interface_2_2                                              22897000.0
4DPI_interface_2_1                                              20121000.0
4DPI_interface_1_2                                              36610000.0
4DPI_interface_1_1                                              23111000.0
4DPI_cortex_3_3                                                 16177000.0
4DPI_cortex_3_2                                                 17586000.0
4DPI_cortex_2_3                                                 21878000.0
4DPI_cortex_2_2                                                 16561000.0
4DPI_cortex_2_1                                                 23720000.0
4DPI_cortex_1_2                                                 22475000.0
4DPI_cortex_1_1                                                 20512000.0
4DPI_SAC_3_3                                                    16860000.0
4DPI_SAC_3_2                                                    22906000.0
4DPI_SAC_2_3                                                    22223000.0
4DPI_SAC_2_2                                                    26184000.0
4DPI_SAC_2_1                                                    24626000.0
4DPI_SAC_1_3                                                    18432000.0
4DPI_SAC_1_2                                                    15517000.0
4DPI_SAC_1_1                                                    15370000.0
10DPI_interface_5_3                                             11978000.0
10DPI_interface_4_3                                             10966000.0
10DPI_interface_4_2                                             11930000.0
10DPI_interface_4_1                                             15307000.0
10DPI_interface_3_2                                             23524000.0
10DPI_interface_3_1                                             14105000.0
10DPI_cortex_4_4                                                19951000.0
10DPI_cortex_4_3                                                10131000.0
10DPI_cortex_3_3                                                20335000.0
10DPI_cortex_3_2                                                15715000.0
10DPI_SAC_4_3                                                   16438000.0
10DPI_SAC_4_2                                                   17678000.0
10DPI_SAC_4_1                                                   14233000.0
10DPI_SAC_3_3                                                   13098000.0
10DPI_SAC_3_2                                                    7518100.0
10DPI_SAC_3_1                                                   11445000.0
10DPI_SAC_2_3                                                   13901000.0
10DPI_SAC_2_2                                                   12134000.0
10DPI_SAC_2_1                                                   12099000.0
```

```
                    P06151;Q564E2;A0A1B0GSR9;A0A1B0GSX0;D3YZQ9;A0A1B0GSL7;A0A1B
0GT41;A0A1B0GQX5;A0A1B0GS79;A0A1B0GRW9;A0A1B0GRC1;A0A1B0GRS2;D3YVR7;D3YZE4;A0A1B
0GRE9;A0A1B0GSR2;P00342  \
index
4DPI_interface_2_4                                          48765000.0
4DPI_interface_2_2                                          84281000.0
4DPI_interface_2_1                                          62405000.0
4DPI_interface_1_2                                          61114000.0
4DPI_interface_1_1                                          36589000.0
4DPI_cortex_3_3                                             33949000.0
4DPI_cortex_3_2                                             29980000.0
4DPI_cortex_2_3                                             62446000.0
4DPI_cortex_2_2                                             63940000.0
4DPI_cortex_2_1                                             25525000.0
4DPI_cortex_1_2                                             45436000.0
4DPI_cortex_1_1                                             61932000.0
4DPI_SAC_3_3                                                58681000.0
4DPI_SAC_3_2                                                56833000.0
4DPI_SAC_2_3                                                55551000.0
4DPI_SAC_2_2                                                50517000.0
4DPI_SAC_2_1                                                84878000.0
4DPI_SAC_1_3                                                81791000.0
4DPI_SAC_1_2                                                81333000.0
4DPI_SAC_1_1                                                59981000.0
10DPI_interface_5_3                                         38969000.0
10DPI_interface_4_3                                         61117000.0
10DPI_interface_4_2                                         41654000.0
10DPI_interface_4_1                                         71086000.0
10DPI_interface_3_2                                         51215000.0
10DPI_interface_3_1                                         34041000.0
10DPI_cortex_4_4                                            51373000.0
10DPI_cortex_4_3                                            36612000.0
10DPI_cortex_3_3                                            38540000.0
10DPI_cortex_3_2                                            19880000.0
10DPI_SAC_4_3                                               79881000.0
10DPI_SAC_4_2                                               92821000.0
10DPI_SAC_4_1                                               91489000.0
10DPI_SAC_3_3                                               47923000.0
10DPI_SAC_3_2                                               87510000.0
10DPI_SAC_3_1                                               39045000.0
10DPI_SAC_2_3                                               27205000.0
10DPI_SAC_2_2                                               50918000.0
10DPI_SAC_2_1                                               16776000.0

                    ...            36            37            38  \
index               ...
4DPI_interface_2_4  ...  1.031646e+07 -1.178469e+07  2.929428e-07
```

```
4DPI_interface_2_2   ...  2.282268e+07 -5.412626e+04  2.929428e-07
4DPI_interface_2_1   ... -2.718176e+07  2.199637e+06  2.929428e-07
4DPI_interface_1_2   ...  1.065881e+07 -7.706425e+05  2.929428e-07
4DPI_interface_1_1   ... -1.414153e+07  1.404734e+06  2.929428e-07
4DPI_cortex_3_3      ... -3.510812e+06 -1.258512e+07  2.929428e-07
4DPI_cortex_3_2      ... -1.934336e+07  6.848204e+06  2.929428e-07
4DPI_cortex_2_3      ... -2.514064e+07  9.140492e+06  2.929428e-07
4DPI_cortex_2_2      ...  4.206068e+07 -1.488583e+07  2.929428e-07
4DPI_cortex_2_1      ...  9.972975e+06  1.967906e+06  2.929428e-07
4DPI_cortex_1_2      ... -1.634307e+07 -2.895322e+06  2.929428e-07
4DPI_cortex_1_1      ...  6.039433e+06  1.772419e+07  2.929428e-07
4DPI_SAC_3_3         ...  1.320228e+07  4.387337e+04  2.929428e-07
4DPI_SAC_3_2         ... -3.905192e+06  2.317435e+07  2.929428e-07
4DPI_SAC_2_3         ...  1.960526e+06 -8.147903e+05  2.929428e-07
4DPI_SAC_2_2         ... -5.308582e+06  3.966172e+06  2.929428e-07
4DPI_SAC_2_1         ... -2.123865e+06 -4.008947e+06  2.929428e-07
4DPI_SAC_1_3         ...  9.412037e+06 -1.014204e+07  2.929428e-07
4DPI_SAC_1_2         ... -8.900910e+06  6.486090e+06  2.929428e-07
4DPI_SAC_1_1         ...  3.946169e+06 -1.159411e+07  2.929428e-07
10DPI_interface_5_3  ...  1.110906e+07  9.848782e+06  2.929428e-07
10DPI_interface_4_3  ... -1.415164e+07 -2.113884e+06  2.929428e-07
10DPI_interface_4_2  ... -1.030126e+07 -2.357560e+07  2.929428e-07
10DPI_interface_4_1  ...  2.432251e+07  2.044187e+07  2.929428e-07
10DPI_interface_3_2  ...  1.381270e+07 -3.013517e+07  2.929428e-07
10DPI_interface_3_1  ... -2.592833e+07  5.400819e+06  2.929428e-07
10DPI_cortex_4_4     ... -4.740456e+06 -7.302414e+06  2.929428e-07
10DPI_cortex_4_3     ...  1.500560e+07  2.581753e+06  2.929428e-07
10DPI_cortex_3_3     ... -8.358536e+06  3.333076e+06  2.929428e-07
10DPI_cortex_3_2     ...  4.094822e+06 -3.480078e+06  2.929428e-07
10DPI_SAC_4_3        ...  6.421815e+06  1.283613e+07  2.929428e-07
10DPI_SAC_4_2        ...  4.099798e+06 -4.146932e+06  2.929428e-07
10DPI_SAC_4_1        ... -1.157357e+07 -3.995583e+06  2.929428e-07
10DPI_SAC_3_3        ... -2.697074e+07 -3.500647e+06  2.929428e-07
10DPI_SAC_3_2        ...  8.463252e+04 -1.427429e+07  2.929428e-07
10DPI_SAC_3_1        ...  1.863299e+07  4.169473e+07  2.929428e-07
10DPI_SAC_2_3        ... -5.895312e+06 -4.218407e+06  2.929428e-07
10DPI_SAC_2_2        ... -3.574784e+06  3.894834e+06  2.929428e-07
10DPI_SAC_2_1        ...  9.418379e+06 -6.709035e+06  2.929428e-07

                    Segment K-means PCA  TimePoint  BioRep  TechRep  \
index
4DPI_interface_2_4                    1       4DPI       2        4
4DPI_interface_2_2                    1       4DPI       2        2
4DPI_interface_2_1                    1       4DPI       2        1
4DPI_interface_1_2                    3       4DPI       1        2
4DPI_interface_1_1                    3       4DPI       1        1
4DPI_cortex_3_3                       3       4DPI       3        3
```

4DPI_cortex_3_2	3	4DPI	3	2
4DPI_cortex_2_3	2	4DPI	2	3
4DPI_cortex_2_2	2	4DPI	2	2
4DPI_cortex_2_1	3	4DPI	2	1
4DPI_cortex_1_2	2	4DPI	1	2
4DPI_cortex_1_1	2	4DPI	1	1
4DPI_SAC_3_3	1	4DPI	3	3
4DPI_SAC_3_2	1	4DPI	3	2
4DPI_SAC_2_3	1	4DPI	2	3
4DPI_SAC_2_2	1	4DPI	2	2
4DPI_SAC_2_1	1	4DPI	2	1
4DPI_SAC_1_3	1	4DPI	1	3
4DPI_SAC_1_2	1	4DPI	1	2
4DPI_SAC_1_1	1	4DPI	1	1
10DPI_interface_5_3	0	10DPI	5	3
10DPI_interface_4_3	0	10DPI	4	3
10DPI_interface_4_2	0	10DPI	4	2
10DPI_interface_4_1	0	10DPI	4	1
10DPI_interface_3_2	1	10DPI	3	2
10DPI_interface_3_1	0	10DPI	3	1
10DPI_cortex_4_4	3	10DPI	4	4
10DPI_cortex_4_3	3	10DPI	4	3
10DPI_cortex_3_3	3	10DPI	3	3
10DPI_cortex_3_2	2	10DPI	3	2
10DPI_SAC_4_3	1	10DPI	4	3
10DPI_SAC_4_2	1	10DPI	4	2
10DPI_SAC_4_1	1	10DPI	4	1
10DPI_SAC_3_3	0	10DPI	3	3
10DPI_SAC_3_2	0	10DPI	3	2
10DPI_SAC_3_1	0	10DPI	3	1
10DPI_SAC_2_3	1	10DPI	2	3
10DPI_SAC_2_2	1	10DPI	2	2
10DPI_SAC_2_1	1	10DPI	2	1

index	Region	DPI_Region	BioRep_TechRep
4DPI_interface_2_4	interface	4DPI_interface	2_4
4DPI_interface_2_2	interface	4DPI_interface	2_2
4DPI_interface_2_1	interface	4DPI_interface	2_1
4DPI_interface_1_2	interface	4DPI_interface	1_2
4DPI_interface_1_1	interface	4DPI_interface	1_1
4DPI_cortex_3_3	cortex	4DPI_cortex	3_3
4DPI_cortex_3_2	cortex	4DPI_cortex	3_2
4DPI_cortex_2_3	cortex	4DPI_cortex	2_3
4DPI_cortex_2_2	cortex	4DPI_cortex	2_2
4DPI_cortex_2_1	cortex	4DPI_cortex	2_1
4DPI_cortex_1_2	cortex	4DPI_cortex	1_2

```
4DPI_cortex_1_1         cortex      4DPI_cortex            1_1
4DPI_SAC_3_3               SAC         4DPI_SAC            3_3
4DPI_SAC_3_2               SAC         4DPI_SAC            3_2
4DPI_SAC_2_3               SAC         4DPI_SAC            2_3
4DPI_SAC_2_2               SAC         4DPI_SAC            2_2
4DPI_SAC_2_1               SAC         4DPI_SAC            2_1
4DPI_SAC_1_3               SAC         4DPI_SAC            1_3
4DPI_SAC_1_2               SAC         4DPI_SAC            1_2
4DPI_SAC_1_1               SAC         4DPI_SAC            1_1
10DPI_interface_5_3  interface  10DPI_interface            5_3
10DPI_interface_4_3  interface  10DPI_interface            4_3
10DPI_interface_4_2  interface  10DPI_interface            4_2
10DPI_interface_4_1  interface  10DPI_interface            4_1
10DPI_interface_3_2  interface  10DPI_interface            3_2
10DPI_interface_3_1  interface  10DPI_interface            3_1
10DPI_cortex_4_4        cortex     10DPI_cortex            4_4
10DPI_cortex_4_3        cortex     10DPI_cortex            4_3
10DPI_cortex_3_3        cortex     10DPI_cortex            3_3
10DPI_cortex_3_2        cortex     10DPI_cortex            3_2
10DPI_SAC_4_3              SAC        10DPI_SAC            4_3
10DPI_SAC_4_2              SAC        10DPI_SAC            4_2
10DPI_SAC_4_1              SAC        10DPI_SAC            4_1
10DPI_SAC_3_3              SAC        10DPI_SAC            3_3
10DPI_SAC_3_2              SAC        10DPI_SAC            3_2
10DPI_SAC_3_1              SAC        10DPI_SAC            3_1
10DPI_SAC_2_3              SAC        10DPI_SAC            2_3
10DPI_SAC_2_2              SAC        10DPI_SAC            2_2
10DPI_SAC_2_1              SAC        10DPI_SAC            2_1

[39 rows x 333 columns]
```

[198]:
```
#append k cluster ID to each sample
df3['k'] = kmeans['Segment K-means PCA']
df3
```

[198]:
```
 Protein IDs           A8DUK4    P63260    Q91VB8    P15864    P20152  \
index
4DPI_interface_2_4  -0.540090  0.021508 -0.022581 -0.236924 -0.124103
4DPI_interface_2_2   0.019054 -0.337894  0.113558 -0.642223 -0.213255
4DPI_interface_2_1  -0.138874 -0.540108 -0.215675 -0.984486 -0.400444
4DPI_interface_1_2   0.608371 -1.105209 -0.325806 -0.810748 -0.569122
4DPI_interface_1_1   0.174370 -1.183182  0.094234 -0.679225 -0.517408
4DPI_cortex_3_3      0.812077 -1.342470  0.739405 -1.104028 -1.326299
4DPI_cortex_3_2      1.171869 -1.082481  0.131356 -1.092429 -1.250061
4DPI_cortex_2_3      1.732487 -1.189867  0.205528 -1.029867 -1.103115
4DPI_cortex_2_2      1.705228 -0.907770  0.786771 -1.080988 -1.122011
4DPI_cortex_2_1      0.017949 -1.231981  0.234514 -0.996560 -1.377694
4DPI_cortex_1_2      3.162245 -0.851188  2.487419 -0.892101 -1.028298
```

```
4DPI_cortex_1_1      2.284197 -1.246783  2.586436 -1.133441 -1.165106
4DPI_SAC_3_3        -0.653982  0.011098 -0.271394  1.321765  0.137459
4DPI_SAC_3_2        -0.555428  0.006037 -0.709916  0.867429  0.003198
4DPI_SAC_2_3        -0.512329 -0.250133 -0.485555 -0.434520 -0.619947
4DPI_SAC_2_2        -0.480516 -0.370315 -0.212042 -0.525994 -0.592846
4DPI_SAC_2_1        -0.371781  0.039127 -0.427148  0.498867  0.228388
4DPI_SAC_1_3         0.162783  0.504720  0.867335 -0.745891 -0.339045
4DPI_SAC_1_2        -0.281163  0.401154  2.496209 -0.308182 -0.292751
4DPI_SAC_1_1        -0.043099  0.343665  0.096850 -0.764444 -0.312506
10DPI_interface_5_3 -0.777652  1.419721 -0.753751  0.824938  1.692848
10DPI_interface_4_3 -0.866100  1.606369 -1.090698  1.305138  1.805398
10DPI_interface_4_2 -0.873578  2.688393 -1.119343  0.698258  1.568746
10DPI_interface_4_1 -0.800096  1.695420 -0.998278  1.009681  1.820800
10DPI_interface_3_2 -0.632316  0.429326 -0.690294  0.874819  1.296107
10DPI_interface_3_1 -0.719082  1.182173 -0.586279  0.609317  1.189421
10DPI_cortex_4_4     0.249718 -0.833378 -0.002022 -0.349340 -0.885863
10DPI_cortex_4_3     1.169458 -1.045285  0.320527 -0.289708 -0.986684
10DPI_cortex_3_3     0.492771 -1.152337  0.040476 -1.045689 -1.388084
10DPI_cortex_3_2     1.595690 -1.288228  2.579389 -0.947973 -1.307077
10DPI_SAC_4_3       -0.946477  0.383822 -1.059250 -0.383768  0.631378
10DPI_SAC_4_2       -0.805390  0.246927 -0.955010  0.293407  1.028207
10DPI_SAC_4_1       -0.915585  0.379811 -1.000268  0.264904  0.442797
10DPI_SAC_3_3       -0.837421  1.047284 -0.539895  1.228074  1.220225
10DPI_SAC_3_2       -0.835097  1.150373 -0.679543  0.438430  0.917108
10DPI_SAC_3_1       -0.814382  1.447415 -0.838558  0.712641  0.795435
10DPI_SAC_2_3       -0.756522  0.301217 -0.319995  3.323850  0.472652
10DPI_SAC_2_2       -0.702037  0.341660  0.175307  0.376145  1.252213
10DPI_SAC_2_1       -0.499269  0.311387 -0.652017  1.830863  0.419339

 Protein IDs           Q03265    Q64475    P27661    P43276    P31786  ...  \
index                                                                  ...
4DPI_interface_2_4  -0.225207  0.487461  0.304723 -0.028911  0.134433  ...
4DPI_interface_2_2   0.089789 -0.433295 -0.113263 -0.522751 -0.504536  ...
4DPI_interface_2_1  -0.369529 -0.155440 -0.818871 -1.047645 -1.302654  ...
4DPI_interface_1_2  -0.241213  0.125845 -0.007414  0.002511 -1.015394  ...
4DPI_interface_1_1   0.168928 -1.043352 -0.340969 -0.393190 -0.514812  ...
4DPI_cortex_3_3      1.809980 -1.645795 -1.310597 -1.351087  0.551226  ...
4DPI_cortex_3_2      1.360204 -1.052557 -1.656457 -1.324128  0.710533  ...
4DPI_cortex_2_3      2.563874 -1.384593 -0.422396 -1.128870  2.363262  ...
4DPI_cortex_2_2      1.560213 -1.012349 -1.323757 -0.949139  1.326104  ...
4DPI_cortex_2_1      2.644484 -1.569062 -1.415339 -1.385625 -0.472758  ...
4DPI_cortex_1_2      0.906860 -0.699643 -0.763131 -0.753384  1.490523  ...
4DPI_cortex_1_1      0.987024 -1.281794 -0.206395 -1.246260  0.739623  ...
4DPI_SAC_3_3        -0.525400  1.972168  2.332589  0.926510 -0.777197  ...
4DPI_SAC_3_2        -0.485897  0.522564  0.658175  0.951647 -0.857984  ...
4DPI_SAC_2_3        -0.125960  0.362906 -0.387693  0.358201  1.684032  ...
4DPI_SAC_2_2        -0.046063  1.084758  0.075848  0.243042  1.603192  ...
```

```
4DPI_SAC_2_1        -0.187086 -0.117659  0.580671  0.241366 -0.301172  ...
4DPI_SAC_1_3        -0.217939 -0.801061  1.767597 -0.877708  0.019972  ...
4DPI_SAC_1_2        -0.500477 -0.780309  1.001093 -0.349095  0.899931  ...
4DPI_SAC_1_1        -0.382281 -0.266356  0.663553 -0.823244  0.018918  ...
10DPI_interface_5_3 -0.901580  1.013422 -0.382157  0.127778 -0.526985  ...
10DPI_interface_4_3 -0.827514  1.870624  0.372136  1.542423 -0.595809  ...
10DPI_interface_4_2 -0.759250  1.245463  0.559159  0.560974 -0.991206  ...
10DPI_interface_4_1 -0.808740  0.584194  0.445623  1.316398 -0.703630  ...
10DPI_interface_3_2 -0.899163  1.029865  0.112986  0.569039 -0.266707  ...
10DPI_interface_3_1 -0.726168  0.444367  0.102009  0.726513 -0.385806  ...
10DPI_cortex_4_4     0.463414 -0.368067  0.051774 -0.178477 -0.821938  ...
10DPI_cortex_4_3     0.495828 -0.910387  0.338825 -0.099504  0.317561  ...
10DPI_cortex_3_3     0.805517 -1.033478 -1.298924 -1.414690 -0.097914  ...
10DPI_cortex_3_2     1.836598 -1.234055 -1.243058 -1.222343  2.401152  ...
10DPI_SAC_4_3       -0.887629 -0.012392 -0.042908 -0.335688 -1.490924  ...
10DPI_SAC_4_2       -0.928505 -0.264724  0.373718  0.088606 -1.390982  ...
10DPI_SAC_4_1       -0.804905  0.156136 -0.493826 -0.204923 -1.472079  ...
10DPI_SAC_3_3       -0.772893  0.738915  0.963448  1.472510 -0.798119  ...
10DPI_SAC_3_2       -0.810746  0.665529  0.529107  0.355426 -0.462956  ...
10DPI_SAC_3_1       -0.729155  1.164043  0.115897  0.340239 -0.021449  ...
10DPI_SAC_2_3       -0.941055  2.246173  2.216491  2.602791  0.750162  ...
10DPI_SAC_2_2       -0.734416  0.680466  1.143132  0.986525  0.257537  ...
10DPI_SAC_2_1       -0.853945 -0.328529 -2.481400  2.224163  0.504851  ...

 Protein IDs           Q9DCX2  A0A1L1SV25    F8WIP8    P10126    P48036  \
index
4DPI_interface_2_4  -0.334925    0.018748  0.107914  1.212955 -0.024608
4DPI_interface_2_2   0.321537    1.474015  0.968526 -0.854949 -0.503046
4DPI_interface_2_1   0.310361    0.307097 -0.023676 -0.441144  0.481258
4DPI_interface_1_2  -0.812839    0.199176  1.497482  1.173070 -0.049537
4DPI_interface_1_1   0.292111    1.774868  0.328190 -1.647993  0.017034
4DPI_cortex_3_3      2.398939    0.720059 -0.229218 -0.556935 -2.266719
4DPI_cortex_3_2      1.324569    0.731844  0.454642  0.938331 -1.426552
4DPI_cortex_2_3      2.349517    1.431181  2.344207 -0.239975 -0.526365
4DPI_cortex_2_2      2.138293    2.043236  3.298078 -0.636912 -0.625806
4DPI_cortex_2_1      0.759878   -0.269183 -0.294759 -1.810814 -1.659913
4DPI_cortex_1_2      1.148796   -0.013677  2.515370  0.541643 -1.030787
4DPI_cortex_1_1      1.394336    2.132254  0.542312  0.812029 -0.595879
4DPI_SAC_3_3        -0.350119   -0.709544  0.106389 -0.141509 -1.048333
4DPI_SAC_3_2        -0.412290   -0.747891  0.087704 -1.318569 -0.713698
4DPI_SAC_2_3         0.036965   -1.984384  0.284609 -0.330962  0.411022
4DPI_SAC_2_2        -0.281836   -1.038753 -0.296284  0.070711 -0.711533
4DPI_SAC_2_1        -0.096370    0.178716  0.518885 -0.455685 -0.634745
4DPI_SAC_1_3        -0.479700    0.296030  0.670741  0.446418 -0.324874
4DPI_SAC_1_2        -0.367495   -0.237596 -0.092830  0.151312 -0.207055
4DPI_SAC_1_1        -0.592533   -0.200805  0.315036  1.908031 -0.554571
10DPI_interface_5_3 -0.907178   -0.050588 -0.081766  2.083774  0.992343
```

10DPI_interface_4_3	-1.111181	0.140489	-1.168785	0.730058	0.235904
10DPI_interface_4_2	-0.882729	0.063615	-1.169293	0.702471	0.341730
10DPI_interface_4_1	-0.728393	-0.880639	-1.214748	-0.650622	0.772141
10DPI_interface_3_2	-0.491401	-1.009320	-0.610496	0.617134	1.074627
10DPI_interface_3_1	-0.893251	-0.182499	-0.552520	1.172655	1.524249
10DPI_cortex_4_4	0.694825	1.663895	-0.550092	0.606789	1.000782
10DPI_cortex_4_3	-0.194255	0.216405	-0.609141	-1.383216	2.065094
10DPI_cortex_3_3	1.912836	0.919092	-1.265154	-1.059899	-0.488110
10DPI_cortex_3_2	0.966998	-0.815013	-1.216605	-1.192724	-1.438267
10DPI_SAC_4_3	-0.724979	0.791728	-0.444244	-1.382800	1.144086
10DPI_SAC_4_2	-0.875848	-0.514998	-0.913915	-0.916771	1.207437
10DPI_SAC_4_1	-0.870530	-0.130691	-0.379494	-0.759516	0.455384
10DPI_SAC_3_3	-0.625610	-1.035522	-0.322308	0.767741	0.667758
10DPI_SAC_3_2	-1.125903	-0.712475	-0.495334	1.261565	2.200236
10DPI_SAC_3_1	-0.839165	-0.457866	-0.619811	1.450602	1.276285
10DPI_SAC_2_3	-0.551101	-1.993058	-0.614956	-0.233784	-0.034713
10DPI_SAC_2_2	-0.616799	-0.699494	-0.551729	0.151312	-0.781103
10DPI_SAC_2_1	-0.883532	-1.418453	-0.322929	-0.783821	-0.221158

Protein IDs	Q8QZT1	E9PZF0	F8WIT2	P09405	k
index					
4DPI_interface_2_4	-0.362371	-0.030764	0.333282	1.597443	1
4DPI_interface_2_2	0.048106	1.395860	-0.255785	0.741501	1
4DPI_interface_2_1	-0.081927	-0.328686	1.400135	1.304200	1
4DPI_interface_1_2	-0.658481	-1.312265	-1.284947	-0.635158	3
4DPI_interface_1_1	-0.435981	-0.513049	-1.074569	0.427190	3
4DPI_cortex_3_3	2.883655	2.183097	-1.269124	-1.883931	3
4DPI_cortex_3_2	1.466451	1.820325	-1.060759	-1.802613	3
4DPI_cortex_2_3	1.501583	-0.214122	-0.559851	-0.100214	2
4DPI_cortex_2_2	2.321853	0.980506	-0.462015	-0.736300	2
4DPI_cortex_2_1	2.616062	1.153391	-1.214699	-2.545370	3
4DPI_cortex_1_2	0.707471	1.691200	-0.514503	-0.741976	2
4DPI_cortex_1_1	0.885410	2.194431	-0.125894	-0.369901	2
4DPI_SAC_3_3	-0.082839	-0.089157	-0.533417	1.397490	1
4DPI_SAC_3_2	-0.405868	-0.592748	-0.343139	0.263853	1
4DPI_SAC_2_3	-0.084284	0.319239	-0.942338	-1.200577	1
4DPI_SAC_2_2	0.207415	-0.308169	-0.712401	0.020676	1
4DPI_SAC_2_1	-0.365261	0.010556	-0.392664	1.233921	1
4DPI_SAC_1_3	0.654773	0.692699	0.203771	0.307000	1
4DPI_SAC_1_2	-0.169680	-0.464053	-0.223654	-0.654438	1
4DPI_SAC_1_1	0.140422	0.212710	0.059371	0.638882	1
10DPI_interface_5_3	-0.835014	-0.679406	2.108762	1.231277	0
10DPI_interface_4_3	-0.576910	-0.186289	1.100246	1.105335	0
10DPI_interface_4_2	-0.686738	-0.772089	0.692647	0.991598	0
10DPI_interface_4_1	-0.715809	-1.494046	0.663326	1.152602	0
10DPI_interface_3_2	-0.620855	-1.031775	1.966034	-0.021150	1
10DPI_interface_3_1	-0.677940	-1.059968	1.545065	0.638104	0

```
10DPI_cortex_4_4    -0.770248 -0.080477 -0.645693 -0.976057  3
10DPI_cortex_4_3     0.627398  0.954609 -0.752614 -1.150666  3
10DPI_cortex_3_3     0.434782  0.854680 -0.998138 -1.193425  3
10DPI_cortex_3_2     0.577210  0.764507 -1.241369 -1.301098  2
10DPI_SAC_4_3       -0.809456  0.528709  0.742096 -0.249867  1
10DPI_SAC_4_2       -1.006581 -0.029186  0.509432  0.039178  1
10DPI_SAC_4_1       -0.720509 -0.101711  0.399519  0.054649  1
10DPI_SAC_3_3       -0.846876 -1.050714  0.788812  0.411175  0
10DPI_SAC_3_2       -1.244334 -1.430989  2.056198 -0.072382  0
10DPI_SAC_3_1       -0.544820 -1.215851  1.831509  0.825852  0
10DPI_SAC_2_3       -0.763123 -1.398708 -0.556053 -0.325977  1
10DPI_SAC_2_2       -0.509742 -0.642748 -0.175040  0.946585  1
10DPI_SAC_2_1       -1.096942 -0.729549 -1.061534  0.632585  1

[39 rows x 101 columns]
```

[199]:
```
#change k to categorical value
df3['k'] = df3['k'].astype('category')
df3.dtypes
```

[199]:
```
 Protein IDs
A8DUK4     float64
P63260     float64
Q91VB8     float64
P15864     float64
P20152     float64
            ...
Q8QZT1     float64
E9PZF0     float64
F8WIT2     float64
P09405     float64
k         category
Length: 101, dtype: object
```

[200]:
```
#calculate mean for each cluster
df4 = df3.groupby(['k']).mean()
df4
```

[200]:
```
 Protein IDs    A8DUK4    P63260    Q91VB8    P15864    P20152    Q03265  \
k
0            -0.815426  1.529644 -0.825793  0.853310  1.376248 -0.792006
1            -0.480723  0.123501 -0.181771  0.256979  0.167602 -0.501426
2             2.095969 -1.096767  1.729109 -1.016874 -1.145122  1.570914
3             0.587073 -1.122040  0.154086 -0.795966 -1.037652  0.938393

 Protein IDs    Q64475    P27661    P43276    P31786  ...    P61979    Q9DCX2  \
k                                                    ...
0             0.965820  0.338153  0.805282 -0.560745  ...  0.535881 -0.889176
1             0.299041  0.382923  0.277885 -0.138400  ...  0.359382 -0.403366
```

```
2           -1.122487 -0.791747 -1.059999  1.664133  ... -0.498346  1.599588
3           -0.937107 -0.704888 -0.768024 -0.167937  ... -1.033024  0.797008

 Protein IDs  A0A1L1SV25    F8WIP8    P10126    P48036    Q8QZT1    E9PZF0  \
k
0             -0.389436 -0.703070  0.939780  1.001331 -0.766055 -0.986169
1             -0.423258 -0.066153 -0.170091 -0.054757 -0.334929 -0.143749
2              0.955596  1.496673 -0.143188 -0.843421  1.198705  1.083304
3              0.744519 -0.083506 -0.467583 -0.350990  0.770455  0.632539

 Protein IDs    F8WIT2    P09405
k
0             1.348321  0.785445
1             0.023201  0.373664
2            -0.580727 -0.649898
3            -1.037568 -1.220004

[4 rows x 100 columns]
```

[201]:
```
#select proteins with values greater than 0 for each cluster; these will be␣
 ↪used for GO analysis
df5 = df4[df4>0]
df5
```

[201]:
```
 Protein IDs    A8DUK4    P63260    Q91VB8    P15864    P20152    Q03265  \
k
0                 NaN  1.529644       NaN  0.853310  1.376248       NaN
1                 NaN  0.123501       NaN  0.256979  0.167602       NaN
2            2.095969       NaN  1.729109       NaN       NaN  1.570914
3            0.587073       NaN  0.154086       NaN       NaN  0.938393

 Protein IDs    Q64475    P27661    P43276    P31786  ...    P61979    Q9DCX2  \
k                                                    ...
0            0.965820  0.338153  0.805282       NaN  ...  0.535881       NaN
1            0.299041  0.382923  0.277885       NaN  ...  0.359382       NaN
2                 NaN       NaN       NaN  1.664133  ...       NaN  1.599588
3                 NaN       NaN       NaN       NaN  ...       NaN  0.797008

 Protein IDs  A0A1L1SV25    F8WIP8   P10126    P48036    Q8QZT1    E9PZF0  \
k
0                 NaN       NaN  0.93978  1.001331       NaN       NaN
1                 NaN       NaN      NaN       NaN       NaN       NaN
2            0.955596  1.496673      NaN       NaN  1.198705  1.083304
3            0.744519       NaN      NaN       NaN  0.770455  0.632539

 Protein IDs    F8WIT2    P09405
k
0            1.348321  0.785445
```

```
1             0.023201  0.373664
2                  NaN       NaN
3                  NaN       NaN

[4 rows x 100 columns]
```

[202]:
```
#make list of proteins with values greater than 0 for each cluster
proteinIDs = []

for i in range(0,4):
    proteinList = df5.iloc[i][-df5.iloc[i].isnull()].index
    proteinIDs.append(proteinList)

proteinIDs
```

[202]:
```
[Index(['P63260', 'P15864', 'P20152', 'Q64475', 'P27661', 'P43276', 'P20065',
        'P52480', 'Q8VDD5', 'B7FAV1', 'P43274', 'O35744', 'P62962', 'P26041',
        'P43277', 'P62806', 'P05064', 'P43275', 'P09528', 'P27773', 'Q61233',
        'P20029', 'P63017', 'P07356', 'P26039', 'P18760', 'D3Z2H9', 'O88569',
        'P17742', 'A0A1D5RLD8', 'A0A1W2P768', 'P11499', 'Q99PT1', 'P40124',
        'E9Q7Q3', 'P40142', 'E9Q3W4', 'Q91Z25', 'Q6IRU2', 'P13020', 'A2AL12',
        'P63101', 'P14733', 'P06745', 'P17182', 'Q9DCD0', 'P05213', 'P26645',
        'Q9JKF1', 'P08113', 'A0A1W2P6F6', 'P09103', 'Q61029', 'P99024',
        'O35639', 'P61979', 'P10126', 'P48036', 'F8WIT2', 'P09405'],
       dtype='object', name=' Protein IDs'),
 Index(['P63260', 'P15864', 'P20152', 'Q64475', 'P27661', 'P43276', 'P20065',
        'P52480', 'Q8VDD5', 'E9PV24', 'B7FAV1', 'P43274', 'O35744', 'P62962',
        'P26041', 'P43277', 'P62806', 'P05064', 'P43275', 'P26350', 'P09528',
        'P27773', 'Q61233', 'P20029', 'P63017', 'P07356', 'P26039', 'P18760',
        'D3Z2H9', 'O88569', 'P17742', 'A0A1D5RLD8', 'P11499', 'Q99PT1',
        'P40124', 'E9Q7Q3', 'P40142', 'E9Q3W4', 'Q91Z25', 'Q6IRU2', 'P13020',
        'A2AL12', 'E9Q616', 'P63101', 'P14733', 'P06745', 'P17182', 'Q9DCD0',
        'P05213', 'P26645', 'Q9JKF1', 'P08113', 'A0A1W2P6F6', 'P09103',
        'Q61029', 'Q921I1', 'P99024', 'O35639', 'A0A0R4J0I1', 'P61979',
        'F8WIT2', 'P09405'],
       dtype='object', name=' Protein IDs'),
 Index(['A8DUK4', 'Q91VB8', 'Q03265', 'P31786', 'P56480', 'P16125', 'Q8VDN2',
        'Q9JII6', 'Q99KI0', 'P63038', 'Q9DCW4', 'Q64433', 'P26443', 'P17742',
        'P08249', 'Q9Z2I8', 'Q61838', 'P70441', 'P24270', 'P05213', 'P38647',
        'Q8BWT1', 'P45952', 'P99029', 'A3KGU5', 'O08749', 'P17751', 'P05202',
        'Q921I1', 'Q62433', 'P14152', 'A0A0R4J0I1', 'Q60932', 'Q9DCX2',
        'A0A1L1SV25', 'F8WIP8', 'Q8QZT1', 'E9PZF0'],
       dtype='object', name=' Protein IDs'),
 Index(['A8DUK4', 'Q91VB8', 'Q03265', 'P56480', 'P16125', 'Q8VDN2', 'Q9JII6',
        'Q99KI0', 'P63038', 'Q9DCW4', 'Q64433', 'P26443', 'P08249', 'Q9Z2I8',
        'Q61838', 'P70441', 'P24270', 'P38647', 'Q8BWT1', 'P45952', 'A3KGU5',
        'P10922', 'O08749', 'P17751', 'P05202', 'Q921I1', 'Q62433', 'P14152',
        'Q60932', 'Q9DCX2', 'A0A1L1SV25', 'Q8QZT1', 'E9PZF0'],
```

```
       dtype='object', name=' Protein IDs')]
```

[203]:
```
#save protein list as df
proteinIDs = pd.DataFrame(proteinIDs)
proteinIDs.to_csv('outs/' + date + '_topProteins_100.csv')
```

0.4.2 Run proteins from line above in PANTHER and generate gene ontology classifications, then proceed with code below to combine GO information per cluster into one dataframe

[205]:
```
#combine individial GO results from PANTHER into a single df

path = 'outs/'

fileNames = ['/*proteinClass_100_noImputation.txt']
geneOntologyTerm = ['ProteinClass']

for file, go in zip(fileNames, geneOntologyTerm):

    files = glob.glob(path + file)

    li = []
    clusterID = []

    for f in files:
        c = f.split('/')[1].split('_')[1]
        clusterID.append(c)

    for filename, c in zip(files, clusterID):
        df = pd.read_csv(filename, index_col = None, header = None, sep = '\t')
        df['cluster'] = str(c)
        li.append(df)

    finalDf = pd.concat(li, axis=0, ignore_index=True).sort_values('cluster')
    finalDf.columns = ['idx','Category name (Accession)','# genes', 'Percent of␣
↪gene hit against total # genes', 'Percent of gene hit against total #␣
↪Protein Class hits', 'cluster']

    finalDf.to_csv('outs/' + date + '_panther_combined_gene_ontology_' +␣
↪str(go) + '_100_noImputation.csv')
```

[206]:
```
#extract and save gene counts in a single df

geneOntologyTerm = ['ProteinClass']

for go in geneOntologyTerm:
    df = pd.read_csv('outs/panther_combined_gene_ontology_' + str(go) +␣
↪'_100_noImputation.csv', index_col = 0)
```

```
    df = df[['Category name (Accession)', '# genes', 'cluster']]
    df = df.pivot(index = 'Category name (Accession)', columns = 'cluster')
    df['# genes'].to_csv('outs/' + date + '_numberGenes_' + str(go) +␣
↪'_100_noImputation.csv')
```

0.5 Extract LFQ intensity data for top proteins based on highest k-means centroid values

[208]:
```
summed = allProt.abs().sum(axis=1)
sorted_mylist = sorted(((v, i) for i, v in enumerate(summed)), reverse=True)[0:
↪29]
final = [lis[1] for lis in sorted_mylist]
finalID = allProt.iloc[final].index
finalID

#get LFQ intensity data for top proteins based on highest k-means centroid␣
↪values
df = lfq.loc[finalID]
df.to_csv('outs/' + date + '_topProteinLFQ.csv')
```

[ ]:
```
```