v19_220601_zerofilled

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)
```

```
2022-06-08
```

[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.tail()
```

[6]:
```
                                  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]
```

[7]:
```
#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()
```

[7]:
```
                                    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]
```

[8]:
```
pGroups_proc
```

[8]:
```
                                          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]
```

[9]:
```
#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')
```

[94]:
```
#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)
  #  lfq.to_csv(filename)
    return lfq

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

[94]:
```
(42, 3620)
```

[95]:
```
lfq.head()
```

[95]:
```
 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
```

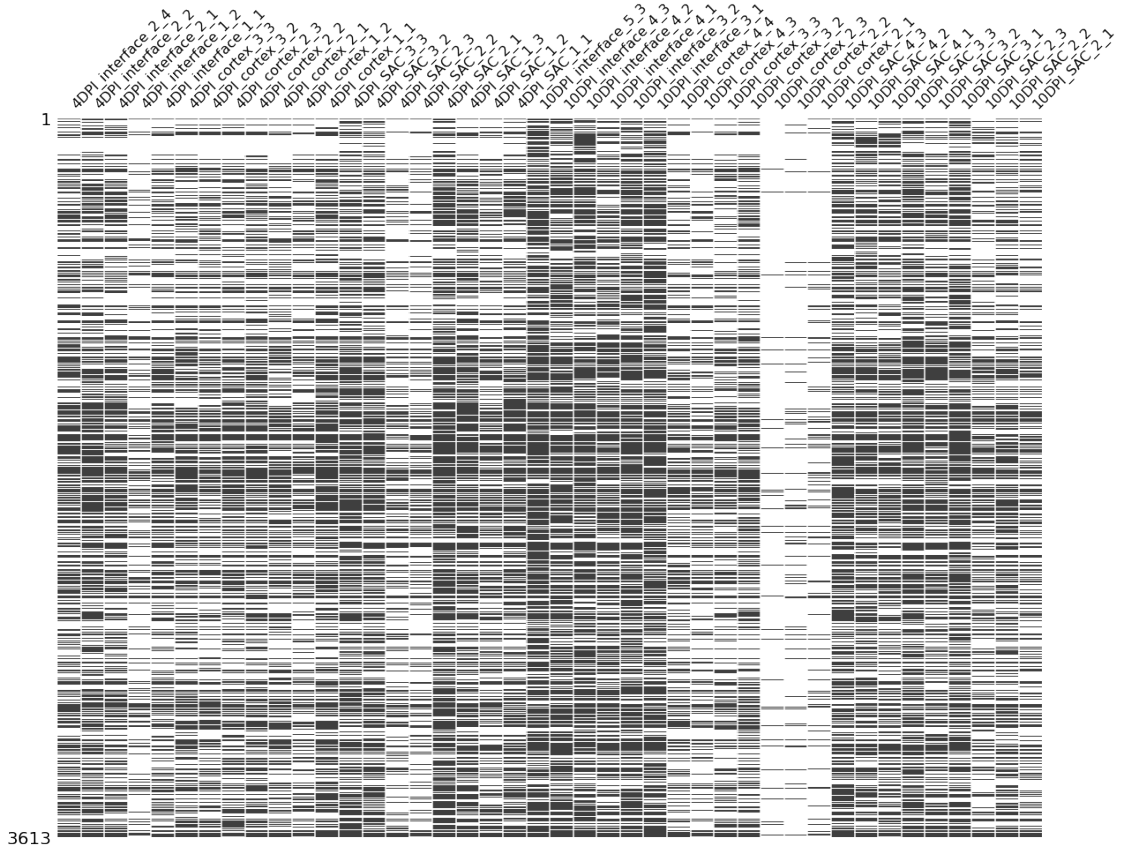
```
 Protein IDs        S4R2J6;S4R165;S4R1X7;S4R2C1;S4R187;S4R2F5  S4R192  \
index
4DPI_interface_2_4                                              0.0     0.0
4DPI_interface_2_2                                              0.0     0.0
4DPI_interface_2_1                                              0.0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

[96]:
```
# 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
```

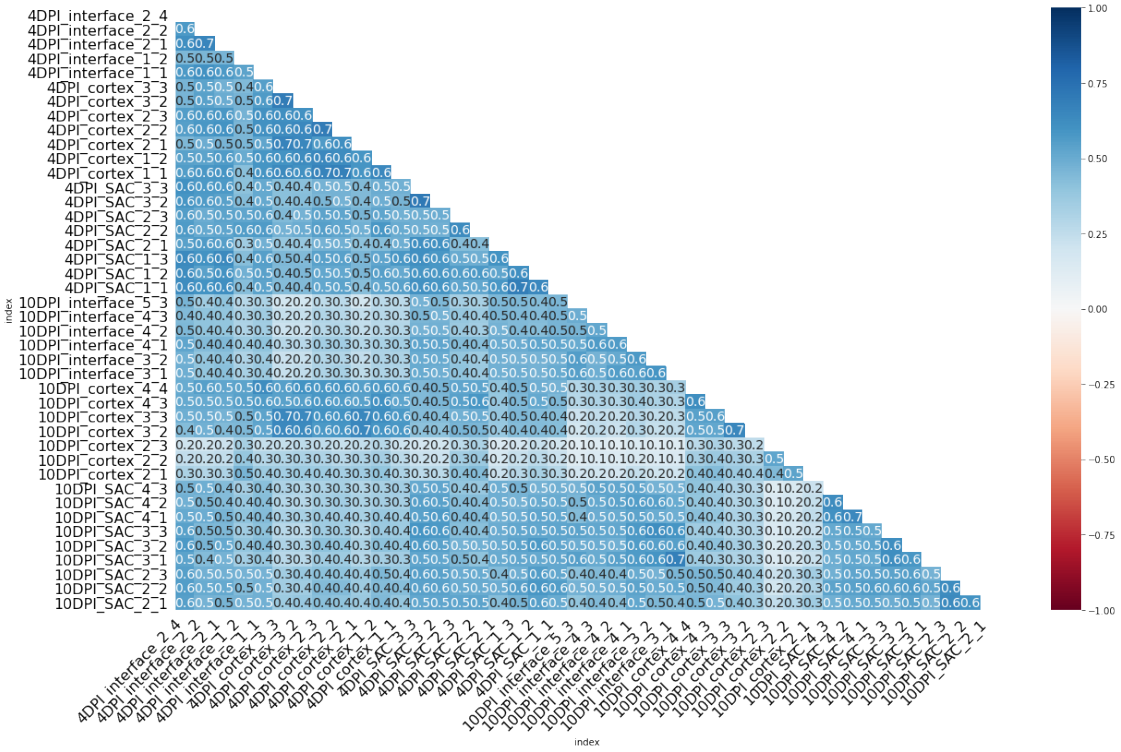
[96]:
```
index
4DPI_interface_2_4     52.809300
4DPI_interface_2_2     47.329089
4DPI_interface_2_1     52.421810
4DPI_interface_1_2     77.359535
4DPI_interface_1_1     59.811791
4DPI_cortex_3_3        56.545807
4DPI_cortex_3_2        60.891226
4DPI_cortex_2_3        57.846665
4DPI_cortex_2_2        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
```

[97]:
```
msno.matrix(lfq_nan.transpose(), sparkline=False, figsize=(20,15))
plt.savefig('outs/' + date + '_missingValues_matrix.png')
```



[98]:
```
msno.heatmap(lfq_nan.transpose())
plt.savefig('outs/' + date + '_missingValues_heatmap.svg')
```



[17]:
```
#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]),)
```

[18]:
```
#z-score calculation to drop outliers
z = np.abs(stats.zscore(percent_missing_values))
z>2
```

[18]:
```
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])
```

[19]:
```
#list of outlier samples
outliers = percent_missing_values.sort_values(ascending=False)[0:3].index
outliers
```

[19]:
```
Index(['10DPI_cortex_2_3', '10DPI_cortex_2_2', '10DPI_cortex_2_1'],
dtype='object', name='index')
```

[20]:
```
#drop outliers
combined_lfq = make_lfq(pGroups_proc, 'combined').drop(outliers, axis=0)
combined_lfq.to_csv('outs/' + date + '_combined_lfq.csv')
```

0.2 Combined analysis

[21]:
```
combined_lfq.head()
```

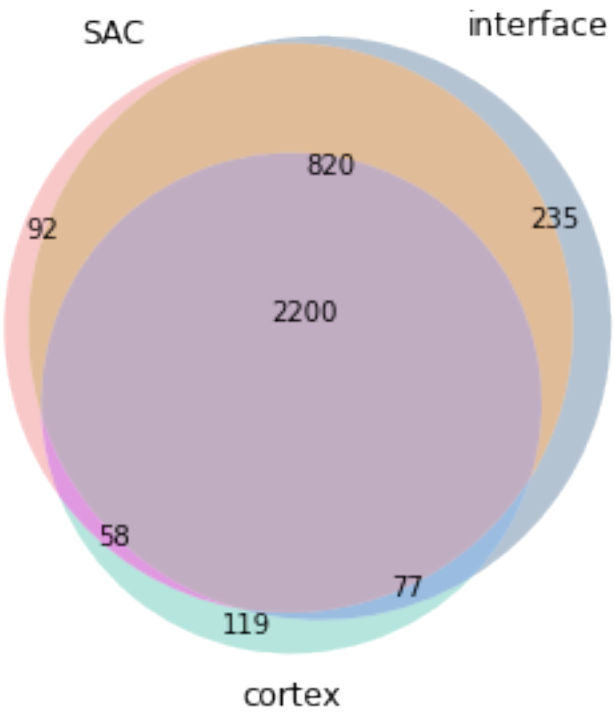
[21]:
```
 Protein IDs        G3UZW7;A0A023T778;Q9CQL1;P61327  O70589;A0A067XG53;F6Y9I5  \
index
4DPI_interface_2_4                                0.0                       0.0
4DPI_interface_2_2                                0.0                       0.0
4DPI_interface_2_1                                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]
```

[22]:
```
#plot venn diagram of protein groups
df = combined_lfq.groupby('Region').sum()

df = df.replace(0, np.nan)
df_missing_values = df.transpose().isnull().sum()
df_present_values = df.shape[1]-df_missing_values

df1 = df.transpose()

df_ab = df1['SAC']
df_ab = df_ab.dropna()
set_ab = df_ab.index.values
set_ab = set(set_ab)

df_le = df1['interface']
df_le = df_le.dropna()
set_le = df_le.index.values
set_le = set(set_le)

df_co = df1['cortex']
df_co = df_co.dropna()
set_co = df_co.index.values
set_co = set(set_co)


df1.head()

df1.dropna(subset=['SAC'])
```

```
#from importlib import reload
#reload(plt)



plt.figure(figsize=(5,5))
vd = venn3([set_ab, set_le, set_co], ('SAC', 'interface', 'cortex'))
vd.get_patch_by_id("100").set_color("#EF767A")
vd.get_patch_by_id('100').set_edgecolor('none')
vd.get_patch_by_id('100').set_alpha(0.4)
vd.get_patch_by_id("001").set_color("#49BEAA")
vd.get_patch_by_id('001').set_edgecolor('none')
vd.get_patch_by_id('001').set_alpha(0.4)
vd.get_patch_by_id("010").set_color("#456990")
vd.get_patch_by_id('010').set_edgecolor('none')
vd.get_patch_by_id('010').set_alpha(0.4)


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



[29]:
```
### 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
```

[30]:
```
X,Y = df_sep('DPI_Region', combined_lfq)
Y
y = pd.factorize(Y)[0]
```

[31]:
```
X.to_csv('/Users/kavyasharman/Documents/Projects/microLESA/outs/' + date +␣
 ↪'_combined_X.csv')
```

0.3 PCA

[32]:
```
pca = PCA()
pca.fit_transform(X)
```

[32]:
```
array([[-1.20613918e+09, -7.35069505e+08,  1.20642660e+09, ...,
        -9.66779606e+06,  3.77862408e+07,  2.94952004e-07],
       [ 5.50887608e+08, -6.44020375e+08,  3.78395103e+07, ...,
         4.48483120e+07,  3.02168233e+07,  2.94952004e-07],
       [ 2.78911665e+08, -1.35151935e+09, -1.10286837e+08, ...,
        -3.92765751e+07, -2.82151875e+07,  2.94952004e-07],
       ...,
       [-2.78629898e+09, -1.91817266e+08,  2.13531506e+08, ...,
        -7.88043640e+06, -1.54576178e+07,  2.94952004e-07],
       [-2.14104411e+09, -9.88192949e+07,  8.11645909e+08, ...,
        -2.20069992e+06, -1.01797990e+07,  2.94952004e-07],
       [-2.01959630e+09, -1.92026814e+08, -3.13821517e+08, ...,
         2.28581628e+06,  2.49944228e+07,  2.94952004e-07]])
```

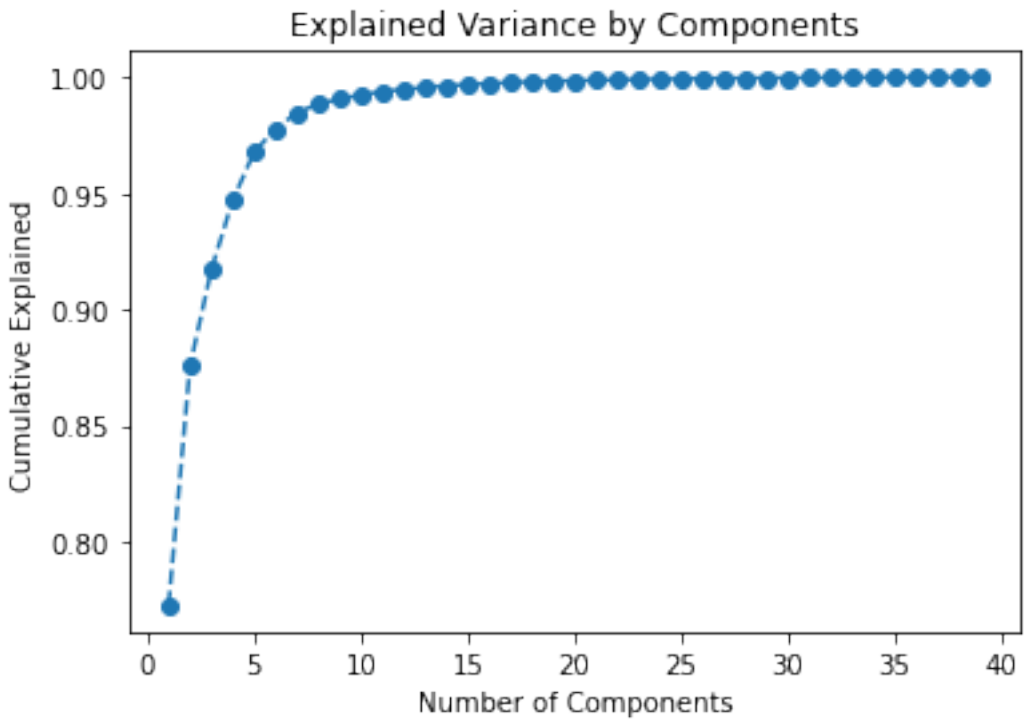
[33]:
```
components = pca.components_
components.shape
```

[33]:
```
(39, 3613)
```

[34]:
```
X.shape
```

[34]:
```
(39, 3613)
```

[35]:
```
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')
```



[36]:
```
pca_expl_var = pca.explained_variance_ratio_
pca_expl_var
```

[36]:
```
array([7.72585327e-01, 1.03009194e-01, 4.14244991e-02, 2.99833868e-02,
       2.06562695e-02, 1.00463271e-02, 6.99901733e-03, 3.76353515e-03,
       2.13194779e-03, 1.71263813e-03, 1.48846065e-03, 9.14578266e-04,
       8.24462531e-04, 6.94519592e-04, 5.66674631e-04, 4.76616717e-04,
       4.02062568e-04, 2.88359638e-04, 2.56272367e-04, 2.26789690e-04,
       2.22757395e-04, 1.81912857e-04, 1.54660794e-04, 1.38718971e-04,
       1.23622374e-04, 1.12507129e-04, 9.61538441e-05, 8.23296203e-05,
       6.69051649e-05, 6.17230689e-05, 5.51396189e-05, 5.09239881e-05,
       4.03391810e-05, 3.97261547e-05, 3.53797674e-05, 3.17609348e-05,
       2.94709459e-05, 2.50296281e-05, 4.93141904e-33])
```

[37]:
```
scores_pca = pca.fit_transform(X)

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

[38]:
```
scores_pca.shape
```

[38]:
```
(39, 39)
```

[39]:
```
pd.DataFrame(scores_pca).to_csv('/Users/kavyasharman/Documents/Projects/
↪microLESA/outs/' + date + '_combined_scores_pca.csv')
```

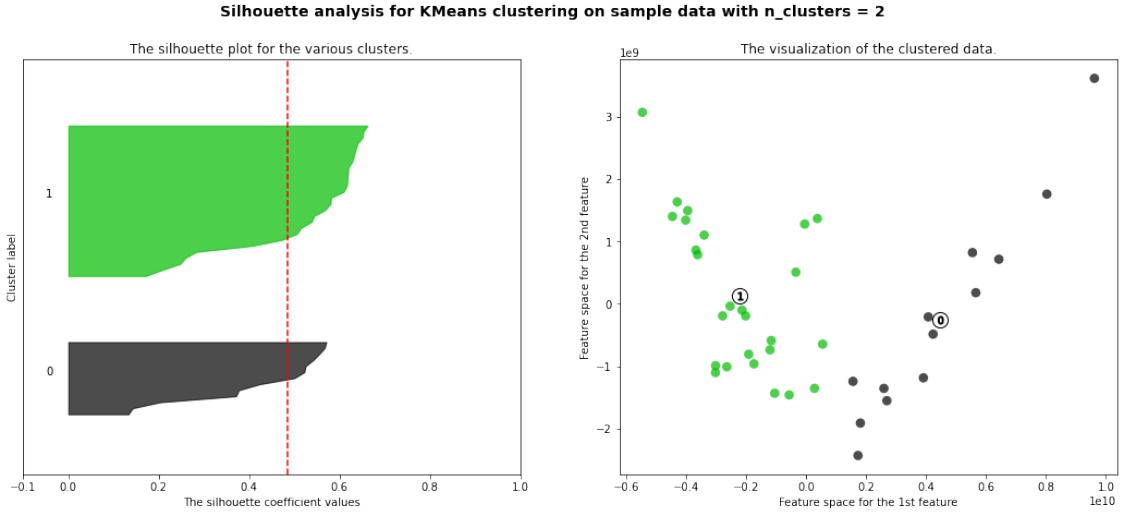
[40]:
```
loadings.to_csv('/Users/kavyasharman/Documents/Projects/microLESA/outs/' + date␣
↪+ '_combined_loadings_pca.csv')
```

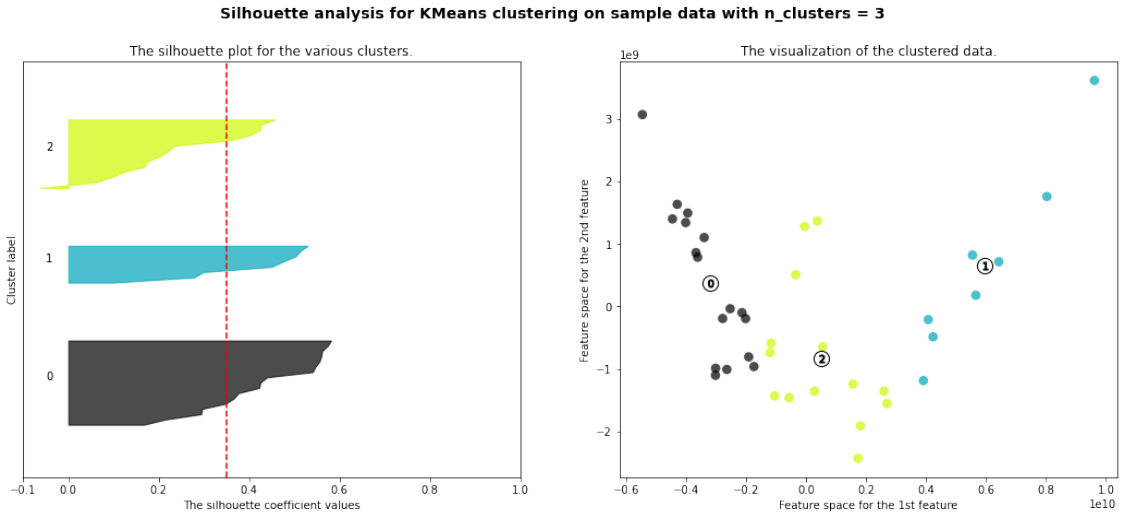
0.3.1 Silhouette scores to determine k

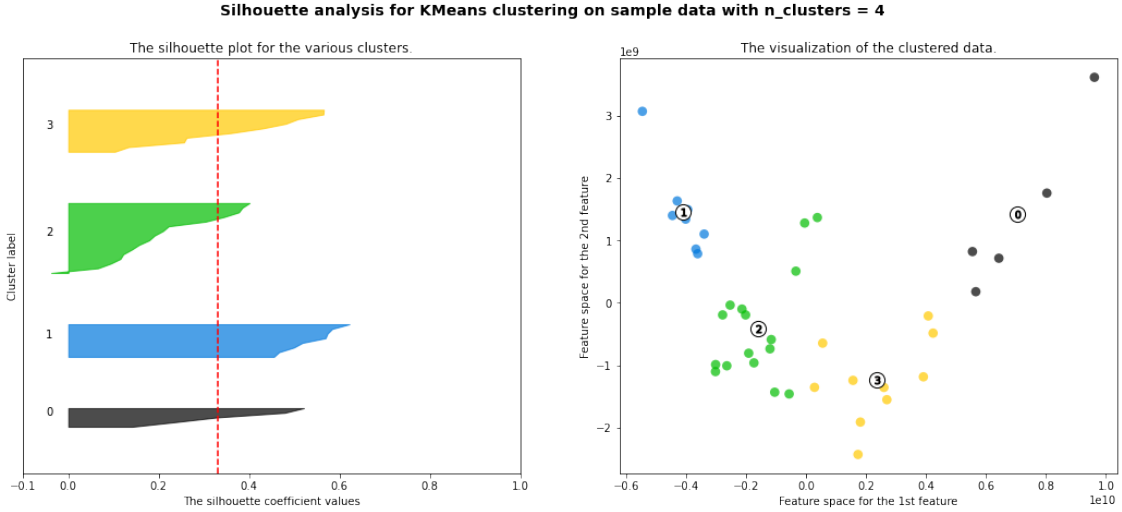
[41]:
```
from sklearn.datasets import make_blobs
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_samples, silhouette_score

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)
```

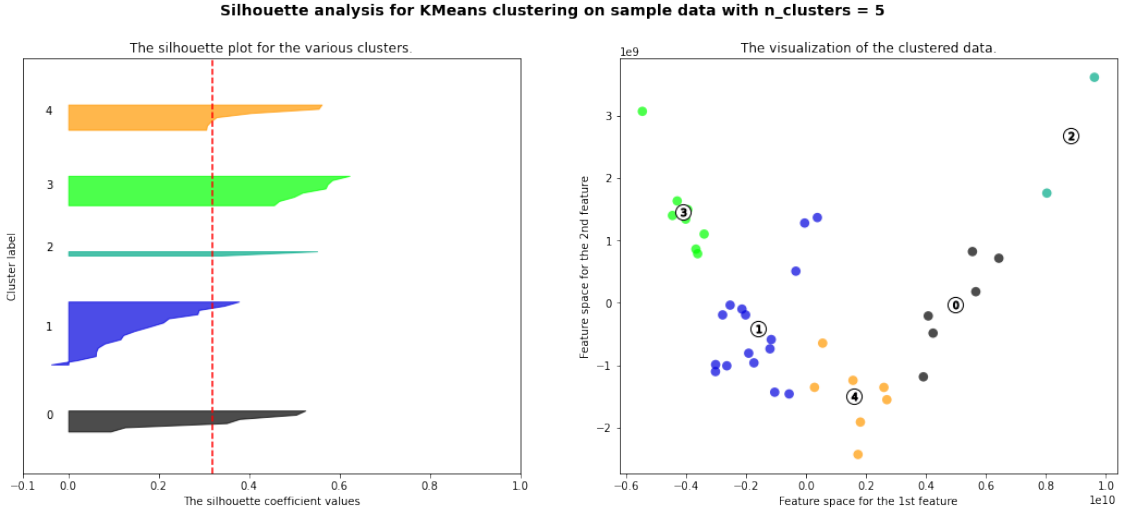
```
    y_lower = 10
    for i in range(n_clusters):
        # Aggregate the silhouette scores for samples belonging to
        # cluster i, and sort them
        ith_cluster_silhouette_values = \
            sample_silhouette_values[cluster_labels == i]

        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')
```

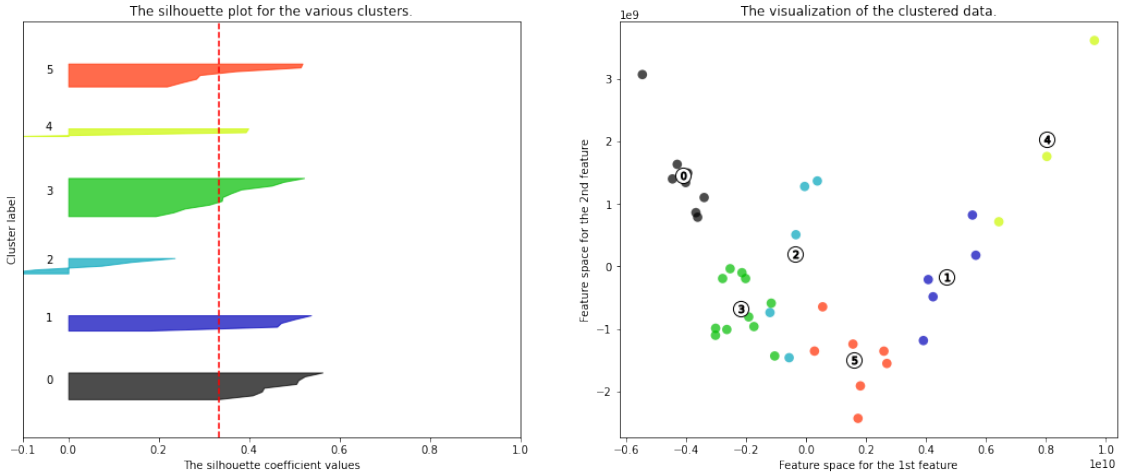




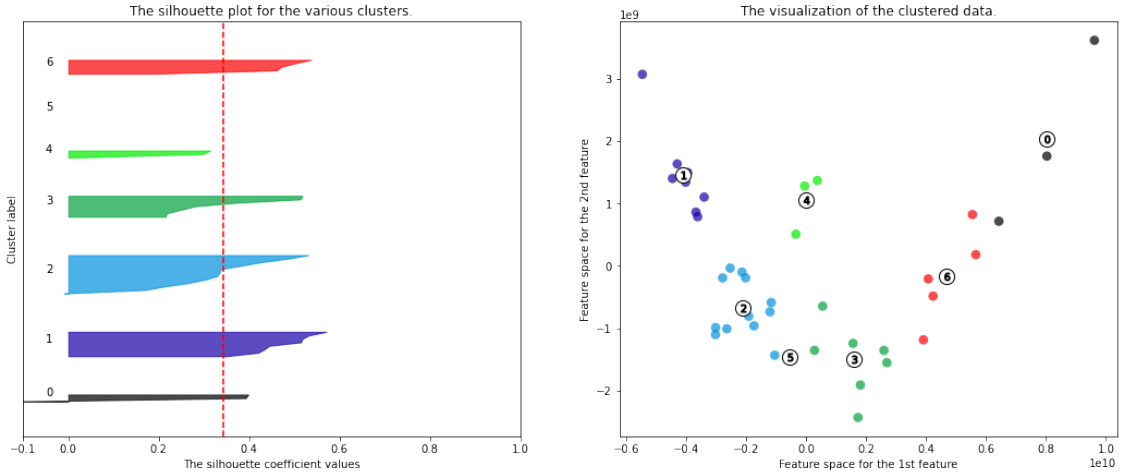

Silhouette analysis for KMeans clustering on sample data with n_clusters = 4
The silhouette plot for the various clusters.
Cluster label
The silhouette coefficient values
The visualization of the clustered data.
Feature space for the 2nd feature
Feature space for the 1st feature
1e9
1e10


Silhouette analysis for KMeans clustering on sample data with n_clusters = 5
The silhouette plot for the various clusters.
Cluster label
The silhouette coefficient values
The visualization of the clustered data.
Feature space for the 2nd feature
Feature space for the 1st feature
1e9
1e10

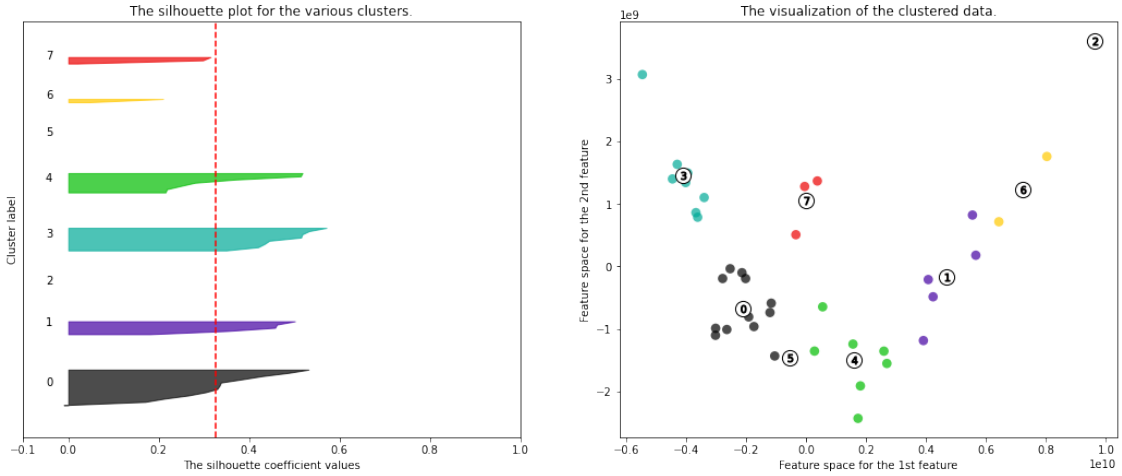
Silhouette analysis for KMeans clustering on sample data with n_clusters = 6



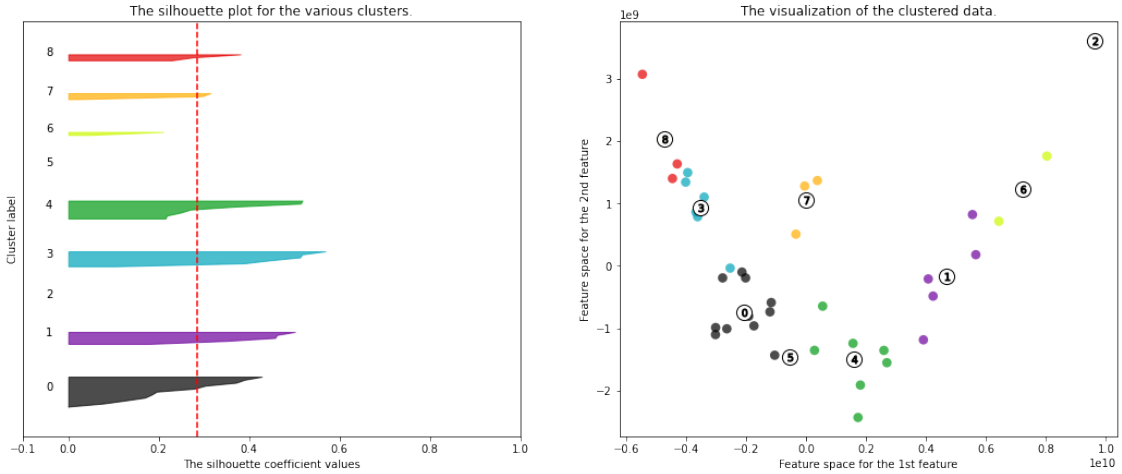
Silhouette analysis for KMeans clustering on sample data with n_clusters = 7



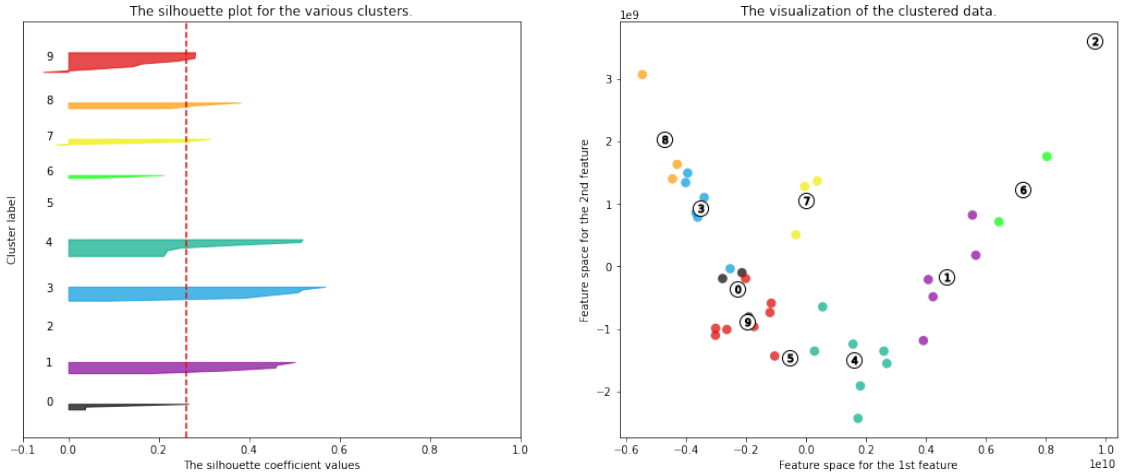
Silhouette analysis for KMeans clustering on sample data with n_clusters = 8



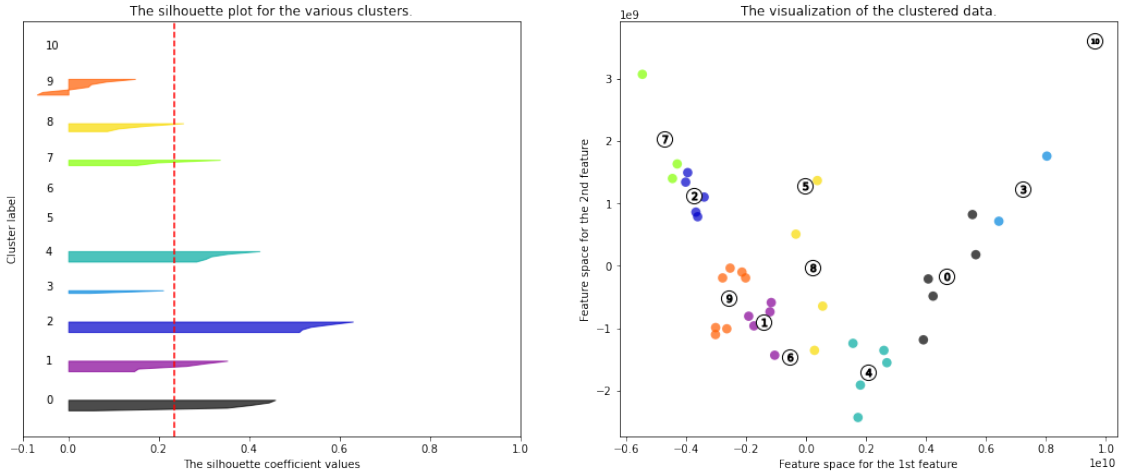
Silhouette analysis for KMeans clustering on sample data with n_clusters = 9



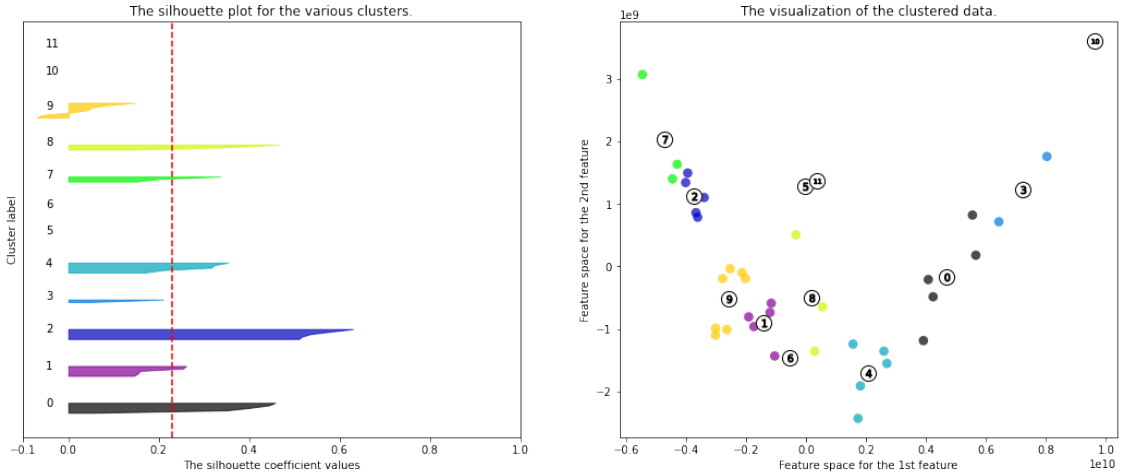
Silhouette analysis for KMeans clustering on sample data with n_clusters = 10



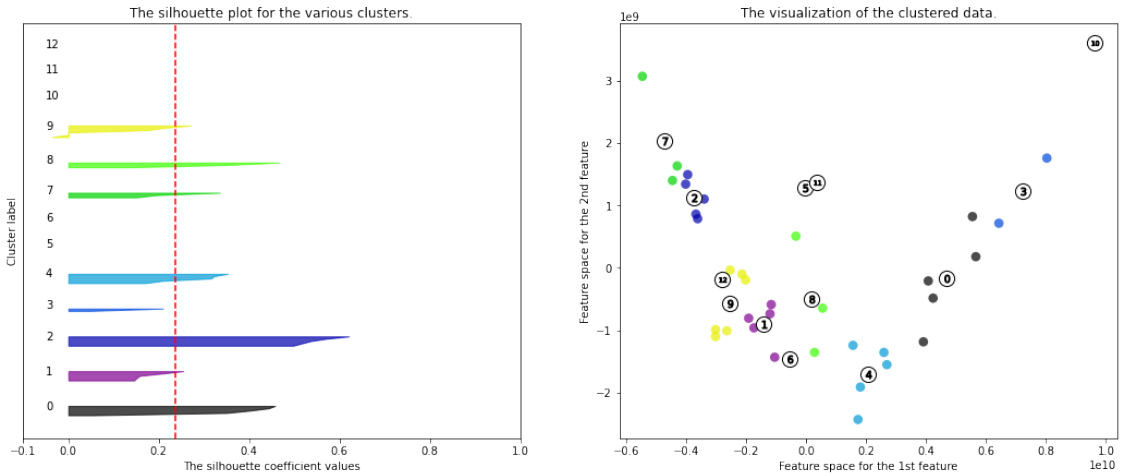
Silhouette analysis for KMeans clustering on sample data with n_clusters = 11

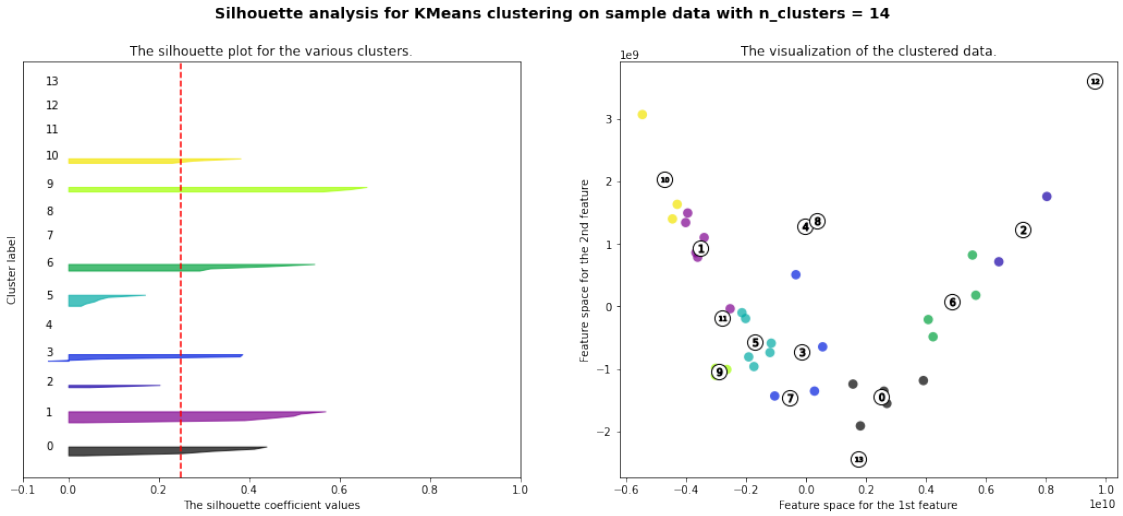


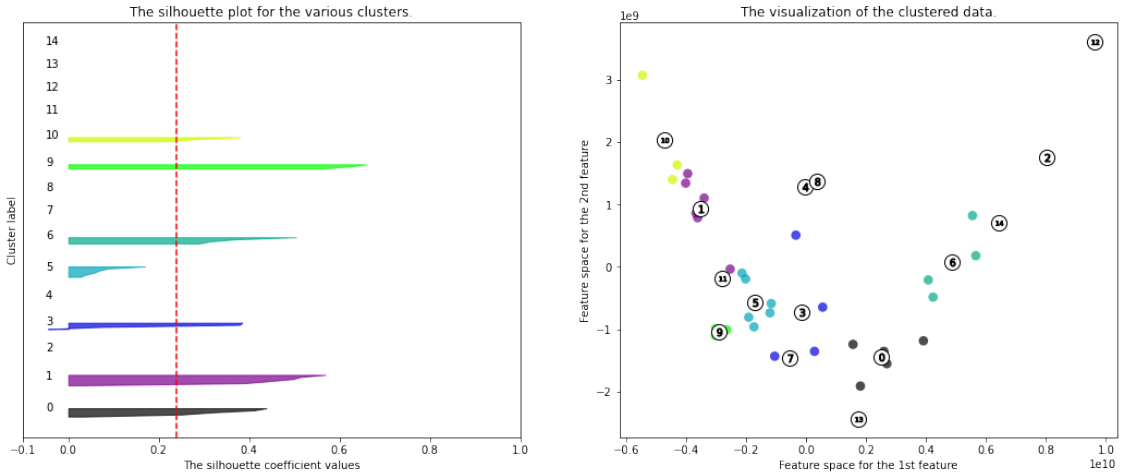
Silhouette analysis for KMeans clustering on sample data with n_clusters = 12



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







[42]:
```
silhouette_scores_euc
```

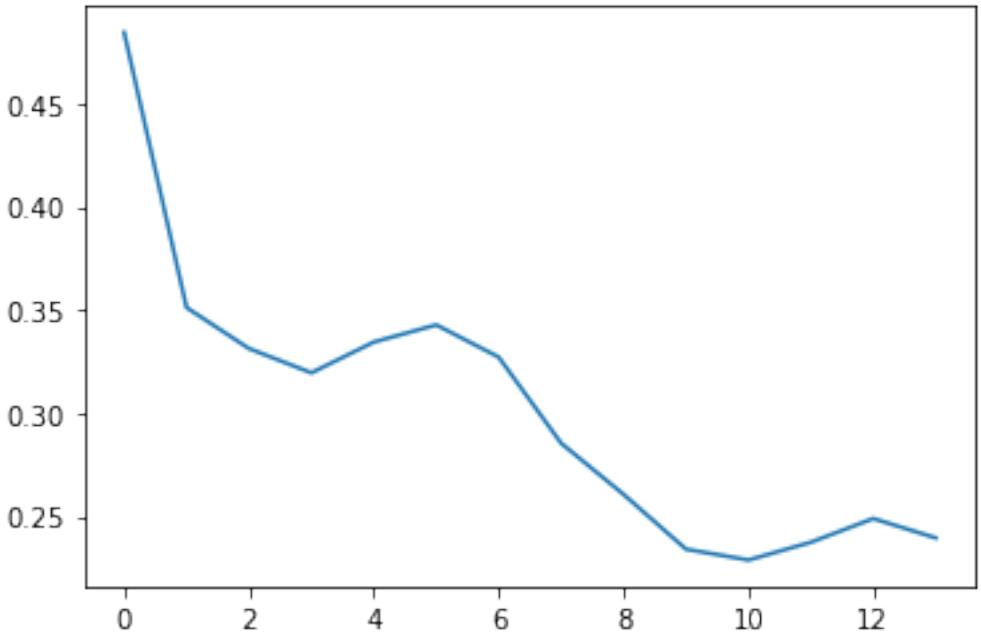
[42]:

	k	score
0	2	0.484378
1	3	0.351101
2	4	0.331259
3	5	0.319485
4	6	0.334392
5	7	0.342704
6	8	0.327104
7	9	0.285408
8	10	0.260599
9	11	0.234079

```
10  12  0.228832
11  13  0.237426
12  14  0.248910
13  15  0.239528
```

[43]:
```
plt.plot(silhouette_scores_euc['score'])
```

[43]:
```
[<matplotlib.lines.Line2D at 0x7facaf0bbfa0>]
```



0.3.2 k=5

[44]:
```
kmeans_pca = KMeans(n_clusters=5, init='k-means++', random_state=42)
kmeans_pca.fit(scores_pca)
kmeans_labels = kmeans_pca.labels_
```

[45]:
```
kmeans_labels_df = pd.DataFrame(kmeans_labels)
```

[46]:
```
cluster_centers = kmeans_pca.cluster_centers_
cluster_centers_df = pd.DataFrame(cluster_centers).to_csv('outs/' + date +␣
 ↪'_combined_kmeans_clusterCenters.csv')
```

[48]:
```
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
```

[48]:

index	G3UZW7;A0A023T778;Q9CQL1;P61327 \
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

                    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

                    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

                   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

                    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

                    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

                    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

                    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

index	A0A075B5S5;A0A0A6YXL5	A0A075B5T3;J3QKO3;A0A0A6YWS9	...	\
			...	
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	...

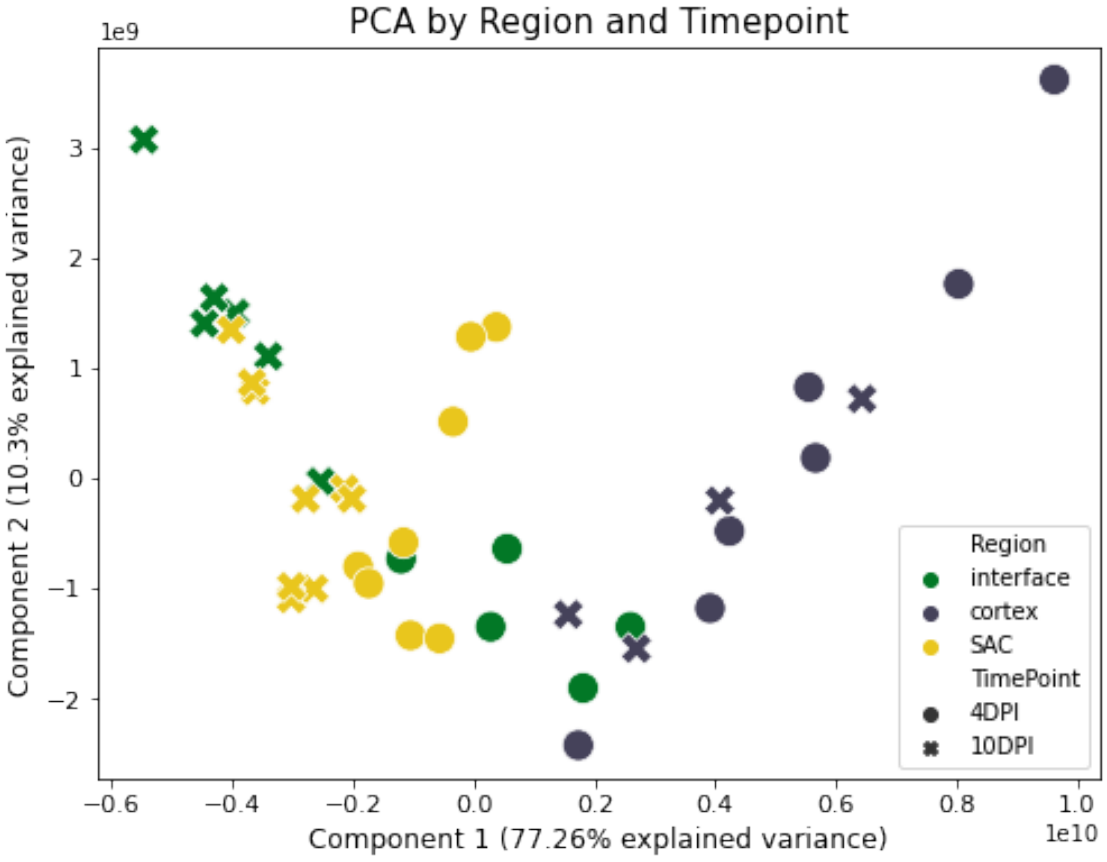
index	36	37	38	\
4DPI_interface_2_4	-9.667796e+06	3.778624e+07	2.949520e-07	
4DPI_interface_2_2	4.484831e+07	3.021682e+07	2.949520e-07	
4DPI_interface_2_1	-3.927658e+07	-2.821519e+07	2.949520e-07	
4DPI_interface_1_2	1.602163e+07	1.157634e+07	2.949520e-07	
4DPI_interface_1_1	-2.771801e+07	-1.684883e+07	2.949520e-07	
4DPI_cortex_3_3	-2.290089e+07	1.351457e+07	2.949520e-07	
4DPI_cortex_3_2	1.978959e+07	-2.313470e+07	2.949520e-07	
4DPI_cortex_2_3	6.588185e+06	-3.016993e+07	2.949520e-07	
4DPI_cortex_2_2	-2.862014e+07	5.637718e+07	2.949520e-07	
4DPI_cortex_2_1	3.341529e+05	2.571811e+05	2.949520e-07	
4DPI_cortex_1_2	-4.056160e+07	-1.501557e+07	2.949520e-07	
4DPI_cortex_1_1	7.268118e+07	-1.174222e+07	2.949520e-07	
4DPI_SAC_3_3	-5.645220e+06	3.757987e+07	2.949520e-07	
4DPI_SAC_3_2	3.419064e+07	-4.386457e+07	2.949520e-07	
4DPI_SAC_2_3	2.145412e+07	1.723141e+06	2.949520e-07	
4DPI_SAC_2_2	-2.183335e+06	-1.551289e+07	2.949520e-07	
4DPI_SAC_2_1	-1.893998e+07	-5.739152e+06	2.949520e-07	
4DPI_SAC_1_3	-2.059662e+07	1.852929e+07	2.949520e-07	
4DPI_SAC_1_2	3.062973e+06	-1.035837e+07	2.949520e-07	
4DPI_SAC_1_1	-1.611010e+06	3.801341e+06	2.949520e-07	
10DPI_interface_5_3	1.634717e+07	-9.292562e+04	2.949520e-07	
10DPI_interface_4_3	-2.843180e+07	-7.313157e+06	2.949520e-07	
10DPI_interface_4_2	7.446369e+06	-2.292092e+06	2.949520e-07	
10DPI_interface_4_1	2.750186e+07	2.790485e+06	2.949520e-07	
10DPI_interface_3_2	1.381678e+07	2.345572e+07	2.949520e-07	
10DPI_interface_3_1	-3.987217e+07	-2.743997e+07	2.949520e-07	
10DPI_cortex_4_4	-7.014139e+06	-1.328333e+07	2.949520e-07	
10DPI_cortex_4_3	5.219814e+06	2.369728e+07	2.949520e-07	
10DPI_cortex_3_3	6.556227e+06	-7.602979e+06	2.949520e-07	
10DPI_cortex_3_2	-4.925120e+06	4.877634e+05	2.949520e-07	
10DPI_SAC_4_3	8.374328e+05	-5.950123e+06	2.949520e-07	
10DPI_SAC_4_2	-9.686167e+06	2.496185e+07	2.949520e-07	
10DPI_SAC_4_1	3.640798e+06	-1.856121e+07	2.949520e-07	
10DPI_SAC_3_3	-2.675078e+06	-1.110516e+07	2.949520e-07	
10DPI_SAC_3_2	-2.930084e+06	-3.858284e+06	2.949520e-07	

```
10DPI_SAC_3_1        2.071382e+07  1.198856e+07  2.949520e-07
10DPI_SAC_2_3       -7.880436e+06 -1.545762e+07  2.949520e-07
10DPI_SAC_2_2       -2.200700e+06 -1.017980e+07  2.949520e-07
10DPI_SAC_2_1        2.285816e+06  2.499442e+07  2.949520e-07

                    Segment K-means PCA  TimePoint  BioRep  TechRep  \
index
4DPI_interface_2_4                    1       4DPI       2        4
4DPI_interface_2_2                    4       4DPI       2        2
4DPI_interface_2_1                    4       4DPI       2        1
4DPI_interface_1_2                    4       4DPI       1        2
4DPI_interface_1_1                    4       4DPI       1        1
4DPI_cortex_3_3                       0       4DPI       3        3
4DPI_cortex_3_2                       0       4DPI       3        2
4DPI_cortex_2_3                       0       4DPI       2        3
4DPI_cortex_2_2                       0       4DPI       2        2
4DPI_cortex_2_1                       4       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                   3      10DPI       5        3
10DPI_interface_4_3                   3      10DPI       4        3
10DPI_interface_4_2                   3      10DPI       4        2
10DPI_interface_4_1                   3      10DPI       4        1
10DPI_interface_3_2                   1      10DPI       3        2
10DPI_interface_3_1                   3      10DPI       3        1
10DPI_cortex_4_4                      4      10DPI       4        4
10DPI_cortex_4_3                      0      10DPI       4        3
10DPI_cortex_3_3                      4      10DPI       3        3
10DPI_cortex_3_2                      0      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                         3      10DPI       3        3
10DPI_SAC_3_2                         3      10DPI       3        2
10DPI_SAC_3_1                         3      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
```

```
                         Region       DPI_Region  BioRep_TechRep
index
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 3659 columns]
```

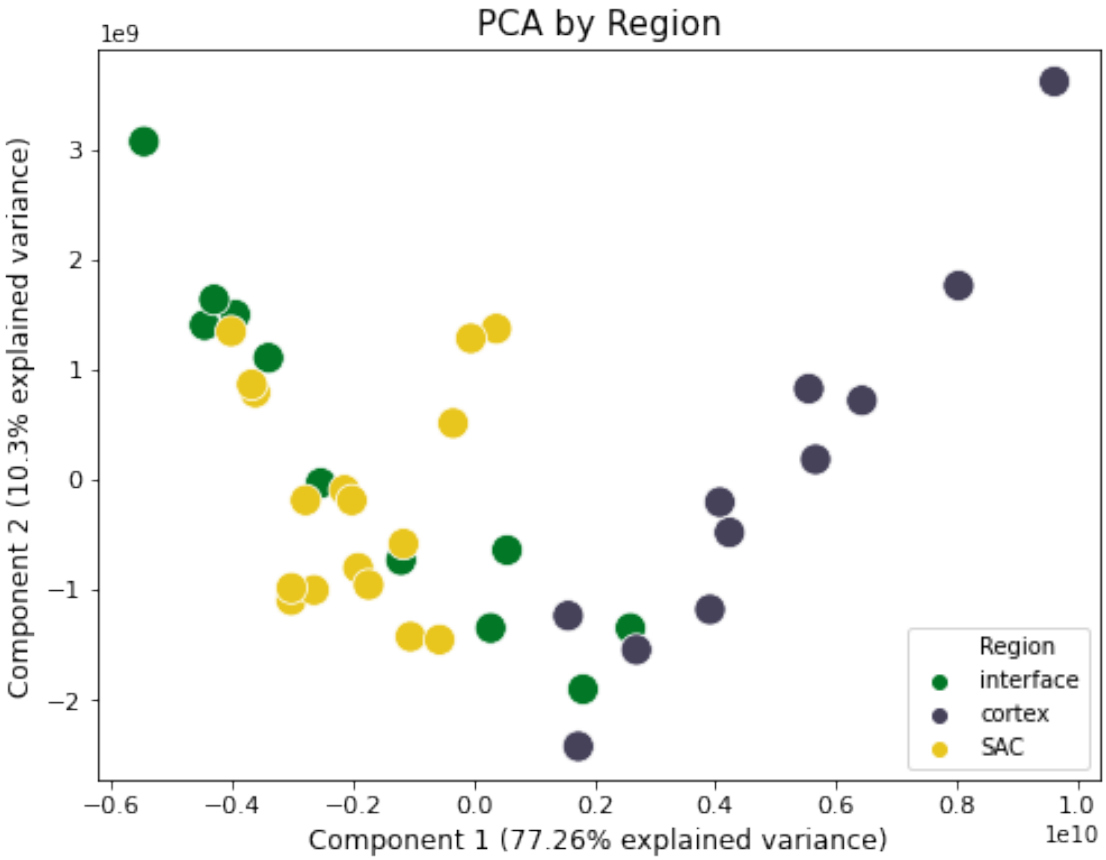
[49]:
```
x_axis = df_segm_pca_kmeans['Component 1']
y_axis = df_segm_pca_kmeans['Component 2']
plt.figure(figsize=(8,6))
```

```
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')
```



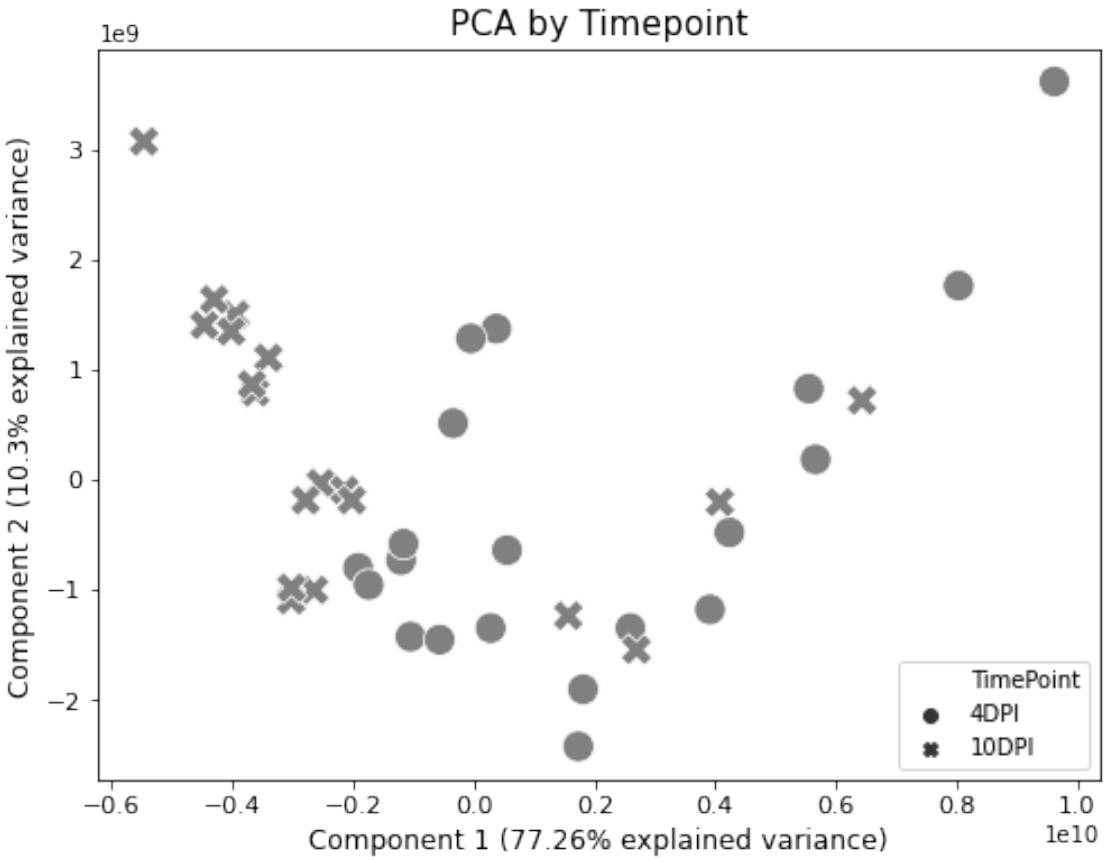
[50]:
```
x_axis = df_segm_pca_kmeans['Component 1']
y_axis = df_segm_pca_kmeans['Component 2']
plt.figure(figsize=(8,6))
sns.scatterplot(x_axis, y_axis,␣
 ↪hue=df_segm_pca_kmeans['Region'],palette=["#027826","#45425A","#E9C61D"],s=200,␣
 ↪legend='full')
plt.title('PCA by Region', fontsize=15)
```

```
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')
```



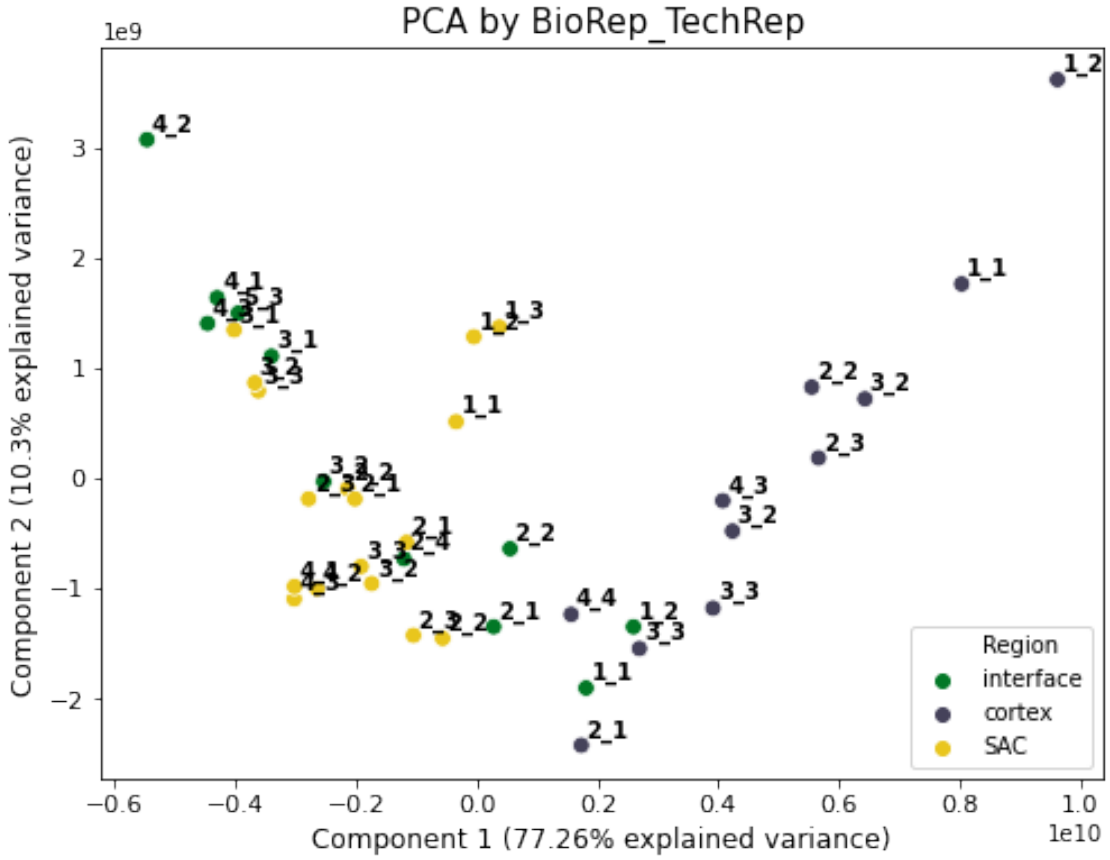
[51]:
```
x_axis = df_segm_pca_kmeans['Component 1']
y_axis = df_segm_pca_kmeans['Component 2']
plt.figure(figsize=(8,6))
sns.scatterplot(x_axis,
                y_axis,
                color='grey',
                style=df_segm_pca_kmeans['TimePoint'],
                s=200,
                legend='full')
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')
```

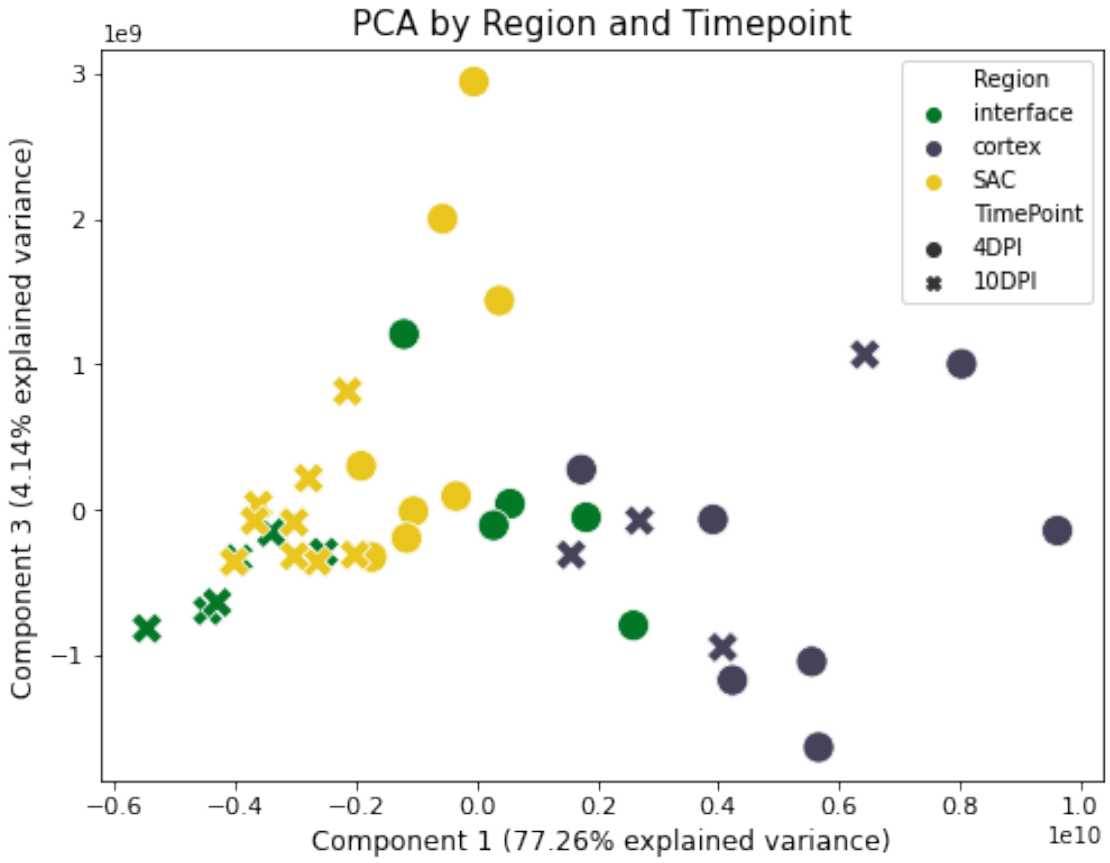


[54]:
```
x_axis = df_segm_pca_kmeans['Component 1']
y_axis = df_segm_pca_kmeans['Component 2']
labels = df_segm_pca_kmeans['BioRep_TechRep']
plt.figure(figsize=(8,6))
sns.scatterplot(x_axis,
                y_axis,
                hue=df_segm_pca_kmeans['Region'],
                palette=["#027826","#45425A","#E9C61D"],
                s=60,
                legend='full')
plt.title('PCA by BioRep_TechRep', fontsize=15)
plt.tick_params(axis='both', which='major', labelsize=11)
```

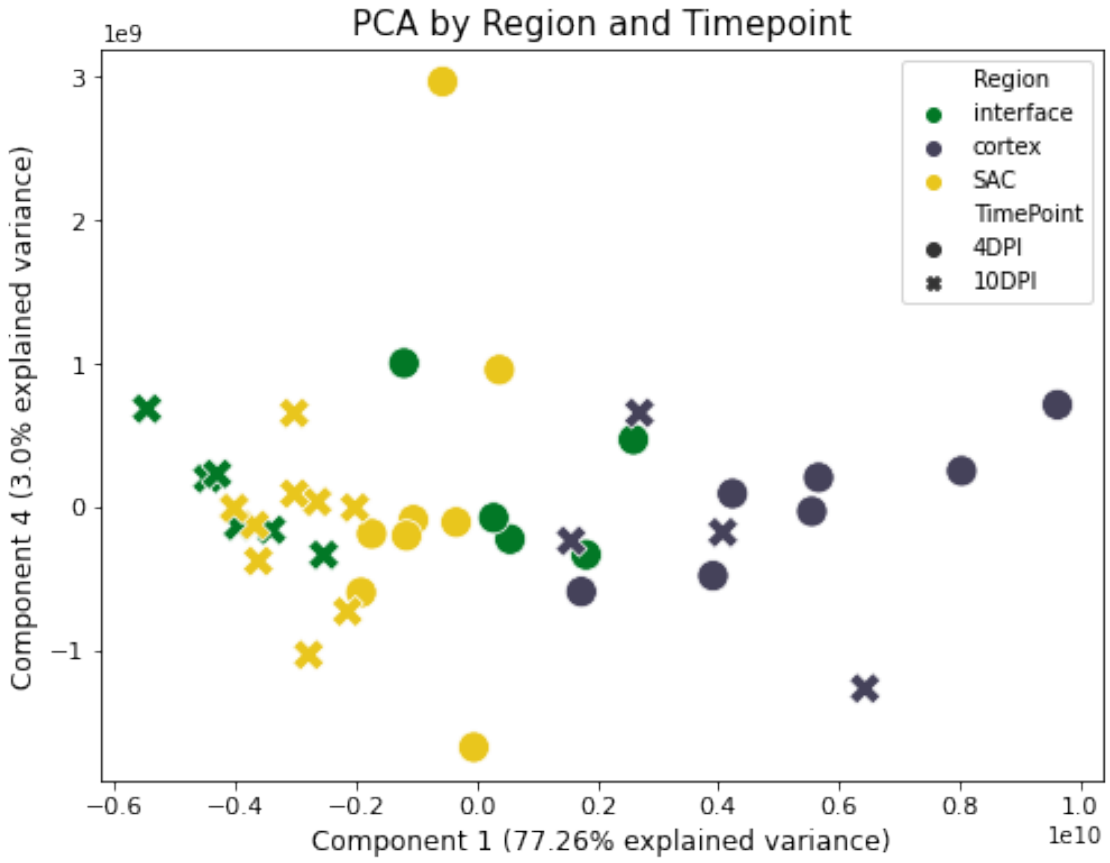
```
plt.tick_params(axis='both', which='minor', labelsize=11)
plt.xlabel('Component 1 (' + str((pca_expl_var[0]*100).round(2)) +'% explained␣
 ↪variance)', fontsize=12)
plt.ylabel('Component 2 (' + str((pca_expl_var[1]*100).round(2)) +'% explained␣
 ↪variance)', fontsize=12)
plt.legend(loc='lower right')

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')
```



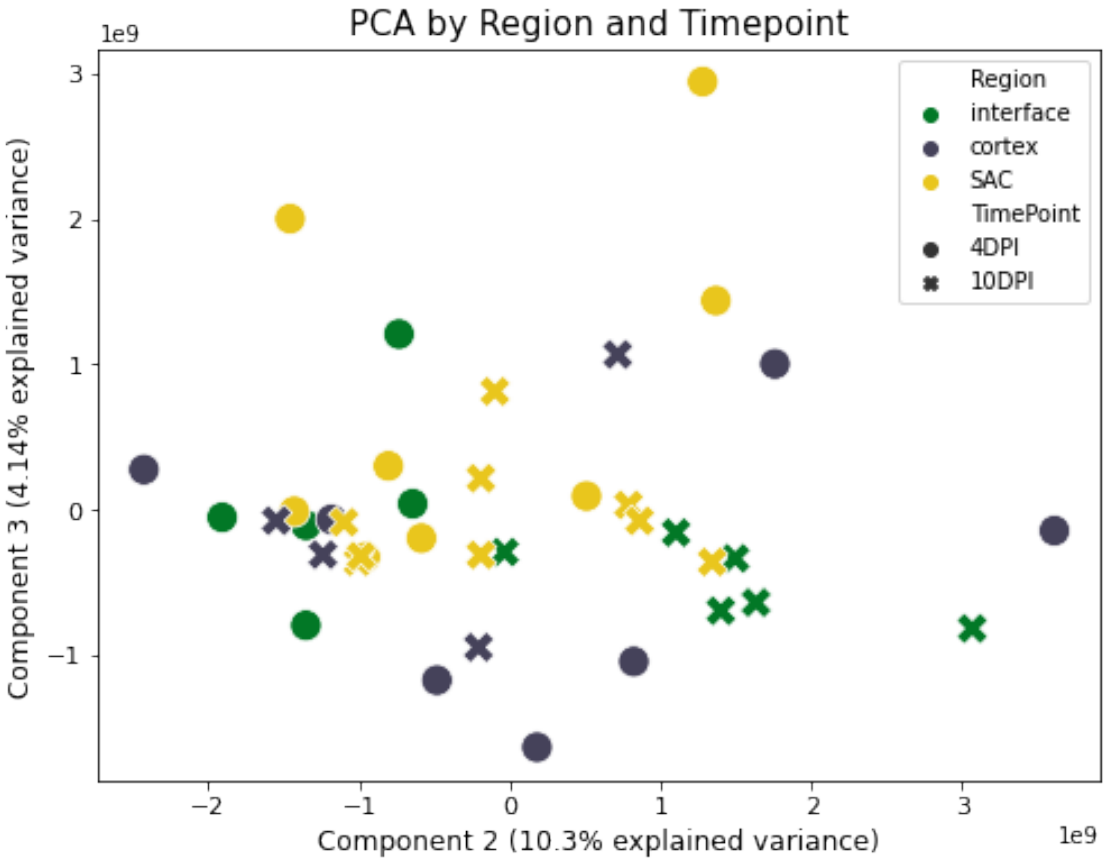
[55]:
```
x_axis = df_segm_pca_kmeans['Component 1']
y_axis = df_segm_pca_kmeans['Component 3']
plt.figure(figsize=(8,6))
sns.scatterplot(x_axis,
                y_axis,
                hue=df_segm_pca_kmeans['Region'],
                palette=["#027826","#45425A","#E9C61D"],
                style=df_segm_pca_kmeans['TimePoint'],
                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')
```



[56]:
```
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')
```

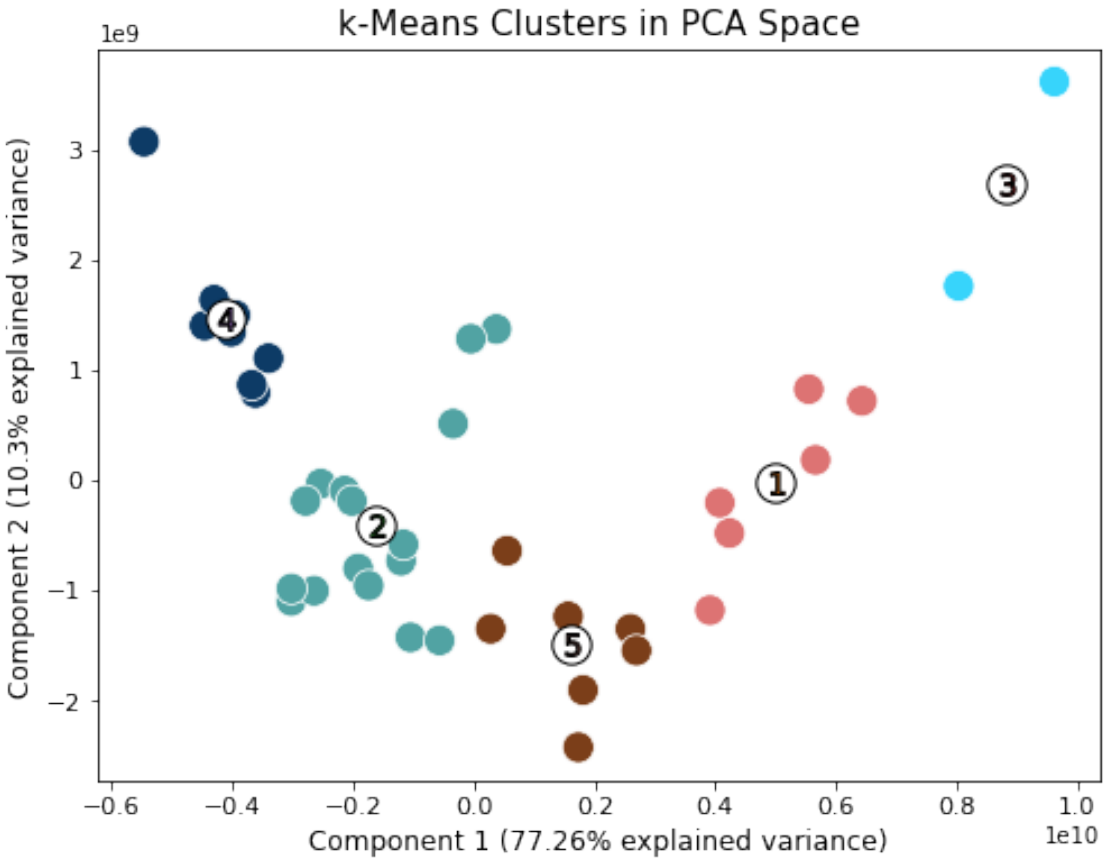


[57]:
```
x_axis = df_segm_pca_kmeans['Component 2']
y_axis = df_segm_pca_kmeans['Component 3']
plt.figure(figsize=(8,6))
sns.scatterplot(x_axis, y_axis,
                hue=df_segm_pca_kmeans['Region'],
                palette=["#027826","#45425A","#E9C61D"],
                style=df_segm_pca_kmeans['TimePoint'],
                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')
```



[58]:
```
palette_kmeans = ["#dd7373","#51a3a3","#37d4fb","#0d3b66","#7b3e19"]
```

[59]:
```
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')
```



[60]:
```
print(cluster_centers)
print(type(cluster_centers))
```

```
[[ 4.97631614e+09 -2.62226294e+07 -6.36404452e+08 -2.77132768e+08
  -9.25103083e+07  3.85743372e+07  2.36238066e+07  1.51952879e+08
  -1.85638030e+08  3.75403214e+06 -3.76877118e+07  4.74203082e+07
  -3.99385925e+07 -1.51797031e+07  9.25132806e+06 -3.72809384e+05
  -2.22019011e+07  1.11803278e+07 -2.58436706e+07 -2.52654634e+07
  -1.57854945e+07  2.28581287e+06  6.09748976e+06 -3.96590426e+06
  -2.99693052e+06  1.38742397e+07  2.62176888e+06 -1.29283084e+07
   9.97017102e+06 -1.04632069e+07  4.54286107e+06 -1.08787849e+07
  -6.15299352e+06  5.28378862e+06  5.87929981e+06  3.46793520e+06
  -4.14142833e+06  6.79536245e+06  2.94952004e-07]
 [-1.61461434e+09 -4.01790671e+08  4.42629517e+08  4.61323893e+07
   1.85518933e+08 -1.18867632e+08 -6.88108370e+07  2.56672292e+05
   6.36768016e+06 -2.46156286e+07  3.84413546e+07 -6.18636555e+06
  -2.09401395e+07 -4.27549157e+07 -4.49396285e+06 -1.24708782e+07
   1.52469667e+07  1.92247824e+07 -1.34558012e+07 -3.10130684e+06
   8.44914232e+06  1.24654478e+07 -7.24884020e+06  1.98426488e+06
   3.98094322e+06 -4.26226131e+06  7.01575538e+06 -2.77713256e+06
```

```
  -4.79593764e+06 -4.11294241e+06  1.05162503e+06 -4.18349027e+06
   1.77869719e+06 -2.19594124e+06 -1.73424287e+06 -2.94816893e+06
   5.48309415e+04  2.95050822e+06  2.94952004e-07]
 [ 8.82376536e+09  2.68506889e+09  4.27130564e+08  4.81770218e+08
   4.83561663e+08 -5.43207091e+07  1.66048357e+07 -1.52276571e+08
   3.08675403e+08  4.16169643e+07  2.99056274e+07 -3.11807489e+07
   7.14139984e+06  1.01589270e+07 -6.83929204e+07 -1.64718554e+07
   1.45543071e+07  8.63889376e+06  6.69433340e+07 -1.73647886e+07
   4.07586421e+07  2.33542144e+07  1.13921544e+07  7.87203352e+06
  -8.97216021e+05 -8.21139372e+06  2.05715124e+07  1.56714453e+07
  -3.72400508e+06  2.08730651e+07 -2.62418375e+06  3.21957118e+05
   1.41653870e+07  2.89186214e+06 -6.73519836e+06  2.10680503e+06
   1.60597906e+07 -1.33788951e+07  2.94952004e-07]
 [-4.11152905e+09  1.46079621e+09 -3.84124687e+08  3.78848633e+07
  -1.50634152e+08  2.13972599e+08  8.58592018e+07  3.08719807e+06
  -4.21398906e+07  2.21866173e+07 -3.56259276e+07  8.99383171e+06
   2.71810883e+07  4.42505564e+07  9.87614677e+06 -8.19533924e+06
  -1.00648921e+07 -1.87364589e+07  1.39875084e+07  8.16244165e+05
   1.56140635e+06 -1.00535558e+07  8.41426936e+06 -3.98084782e+06
  -1.01625394e+07  7.38596103e+06 -2.22751894e+06  6.20357815e+06
   5.74169585e+06  5.20445251e+06 -3.76863355e+05  3.30182448e+06
  -1.77114483e+06  6.38039515e+06  2.68059014e+06  5.08173281e+06
  -2.37488472e+05 -4.66531742e+06  2.94952004e-07]
 [ 1.60294776e+09 -1.49578871e+09 -1.49272741e+08 -4.88487086e+07
  -3.10755883e+08  9.61381523e+06  3.41638953e+07 -9.08526389e+07
   1.04530517e+08  1.57998568e+07 -2.33913196e+07 -2.78755938e+07
   4.89917544e+07  5.72620805e+07  1.05960148e+07  4.28967617e+07
  -8.47564835e+06 -3.45806573e+07  1.77954438e+07  3.27733306e+07
  -1.92116922e+07 -2.56345749e+07 -1.52885128e+06  1.16427184e+06
   5.34017695e+06 -8.24487972e+06 -2.16150818e+07  5.86177940e+06
  -3.08165437e+06  6.45779565e+06 -5.11712751e+06  1.50214348e+07
  -8.14687048e+05 -7.62779391e+06 -2.21460533e+06 -2.64348297e+06
  -8.92628359e+05 -3.41428238e+06  2.94952004e-07]]
<class 'numpy.ndarray'>
```

[61]:
```
cluster_centers.shape
```

[61]:
```
(5, 39)
```

[62]:
```
components.shape
```

[62]:
```
(39, 3613)
```

[63]:
```
centers = cluster_centers.dot(components)
centers.shape
```

[63]:
```
(5, 3613)
```

[64]:
```
centers = pd.DataFrame(centers,
```

```
                        index=['Cluster 1', 'Cluster 2', 'Cluster 3', 'Cluster␣
 ↪4', 'Cluster 5'],
                        columns = X.columns)
centers.to_csv('outs/' + date + '_combined_centers_annotated.csv')
```

[65]:
```
centers_t = centers.T
centers_t.shape
```

[65]:
```
(3613, 5)
```

[66]:
```
pGroups_proc.shape
```

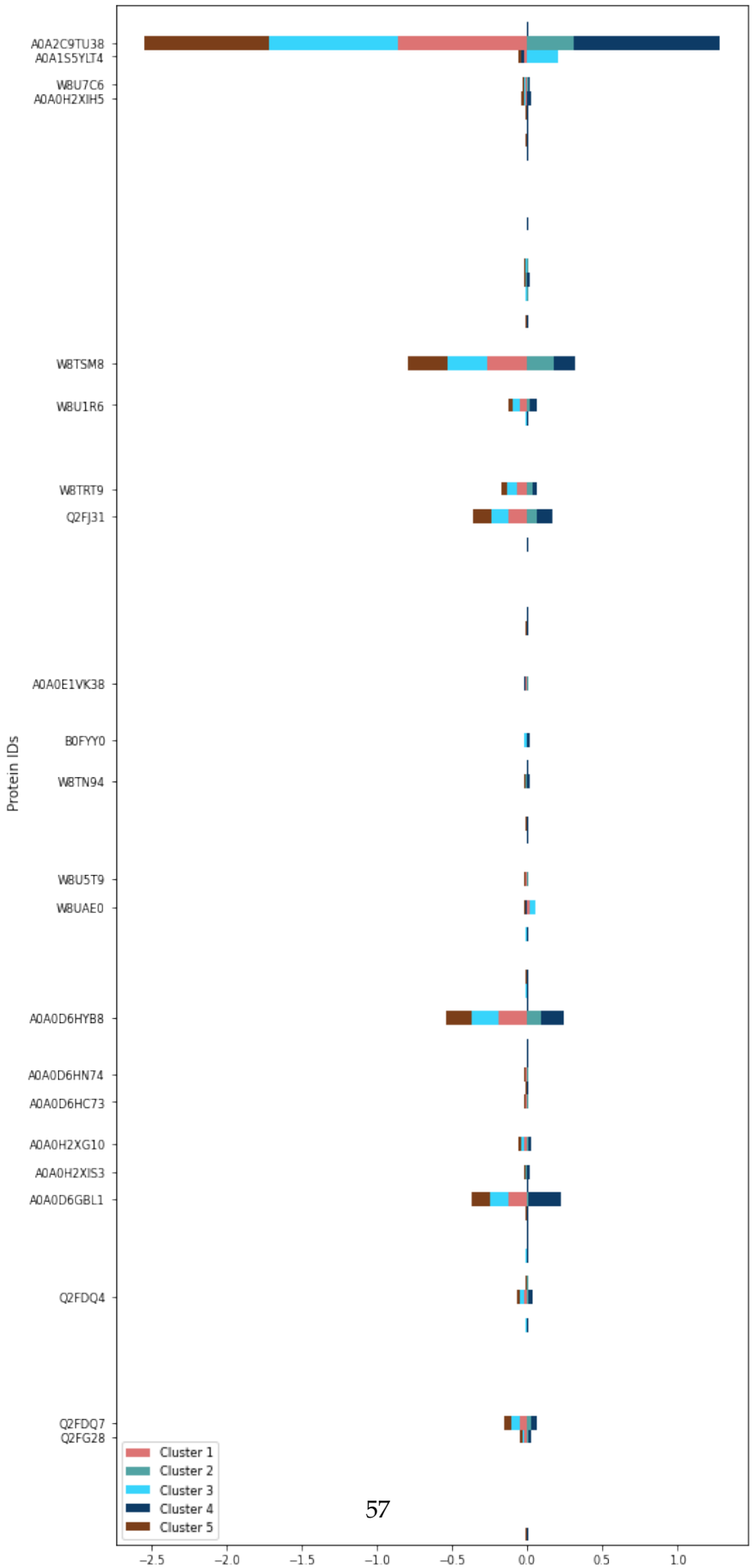
[66]:
```
(3613, 371)
```

[67]:
```
#plot only staph markers
#first pull only staph index then pull centers_t based on that index

staph_idx = pGroups_proc.index[pGroups_proc['Fasta headers'].str.
 ↪contains('aureus', regex=False)].tolist()
centers_t_staph = centers_t.iloc[staph_idx]
centers_t_staph.index = centers_t_staph.index.str.split(';',0).str[0]
centers_t_staph.to_csv('outs/' + date + '_staph_centers_annotated.csv')
centers_t_staph
```

[67]:
```
                Cluster 1     Cluster 2     Cluster 3     Cluster 4  \
 Protein IDs
Q2FER6       -3.003200e+05  1.624625e+05 -3.003200e+05  2.381750e+05
Q2FEQ8       -8.424872e+03 -8.424872e+03 -8.424872e+03  3.264638e+04
Q2FHT6       -4.496051e+04  1.489261e+04 -4.496051e+04  5.451574e+04
Q2FGF0       -2.179487e+04 -2.179487e+04 -2.179487e+04  8.445513e+04
Q2FJA2       -9.495641e+03 -9.495641e+03 -9.495641e+03  3.679561e+04
...                    ...           ...           ...           ...
A0A2C9TKS1   -2.888718e+04 -2.888718e+04 -2.888718e+04  1.119378e+05
A0A1S5YLT4   -2.152005e+06  1.383074e+05  2.077899e+07 -1.974355e+06
A0A2C9TU38   -8.591956e+07  3.073015e+07 -8.591956e+07  9.707224e+07
W8UAB2       -1.310485e+05 -2.405221e+04 -1.310485e+05  2.938203e+05
Q2FH01       -5.815923e+04 -5.815923e+04 -5.815923e+04  2.253670e+05

                Cluster 5
 Protein IDs
Q2FER6       -3.003200e+05
Q2FEQ8       -8.424872e+03
Q2FHT6       -4.496051e+04
Q2FGF0       -2.179487e+04
Q2FJA2       -9.495641e+03
...                    ...
A0A2C9TKS1   -2.888718e+04
A0A1S5YLT4   -2.152005e+06
A0A2C9TU38   -8.298626e+07
W8UAB2       -1.310485e+05
```

```
Q2FH01       -5.815923e+04

[110 rows x 5 columns]
```

[68]:
```
summed = centers_t_staph.abs().sum(axis=1)
sorted_mylist = sorted(((v, i) for i, v in enumerate(summed)), reverse=True)[0:
 ↪22]
final = [lis[1] for lis in sorted_mylist]
finalID = centers_t_staph.iloc[final].index

centers_t_staph.iloc[final].to_csv('outs/' + date + '_finalID_staph_all.csv')

centers_t_staph.plot(kind="barh", stacked=True, width=1, figsize = (8,20),␣
 ↪color=palette_kmeans)
plt.tick_params(axis='both', which='major', labelsize=8)
plt.yticks(final, finalID)
#plt.xlim(-0.5e9, 0.5e9)
plt.legend(fontsize='small')
plt.savefig('outs/' + date + '_staph_kmeansCentroids_all.svg')
```


A0A2C9TU38
A0A1S5YLT4
W8U7C6
A0A0H2XIH5
W8TSM8
W8U1R6
W8TRT9
Q2FJ31
A0A0E1VK38
B0FYY0
W8TN94
W8U5T9
W8UAE0
A0A0D6HYB8
A0A0D6HN74
A0A0D6HC73
A0A0H2XG10
A0A0H2XIS3
A0A0D6GBL1
Q2FDQ4
Q2FDQ7
Q2FG28
Protein IDs
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1e8

[69]:
```
#plot only mouse markers
#first pull only mouse index then pull centers_t based on that index

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

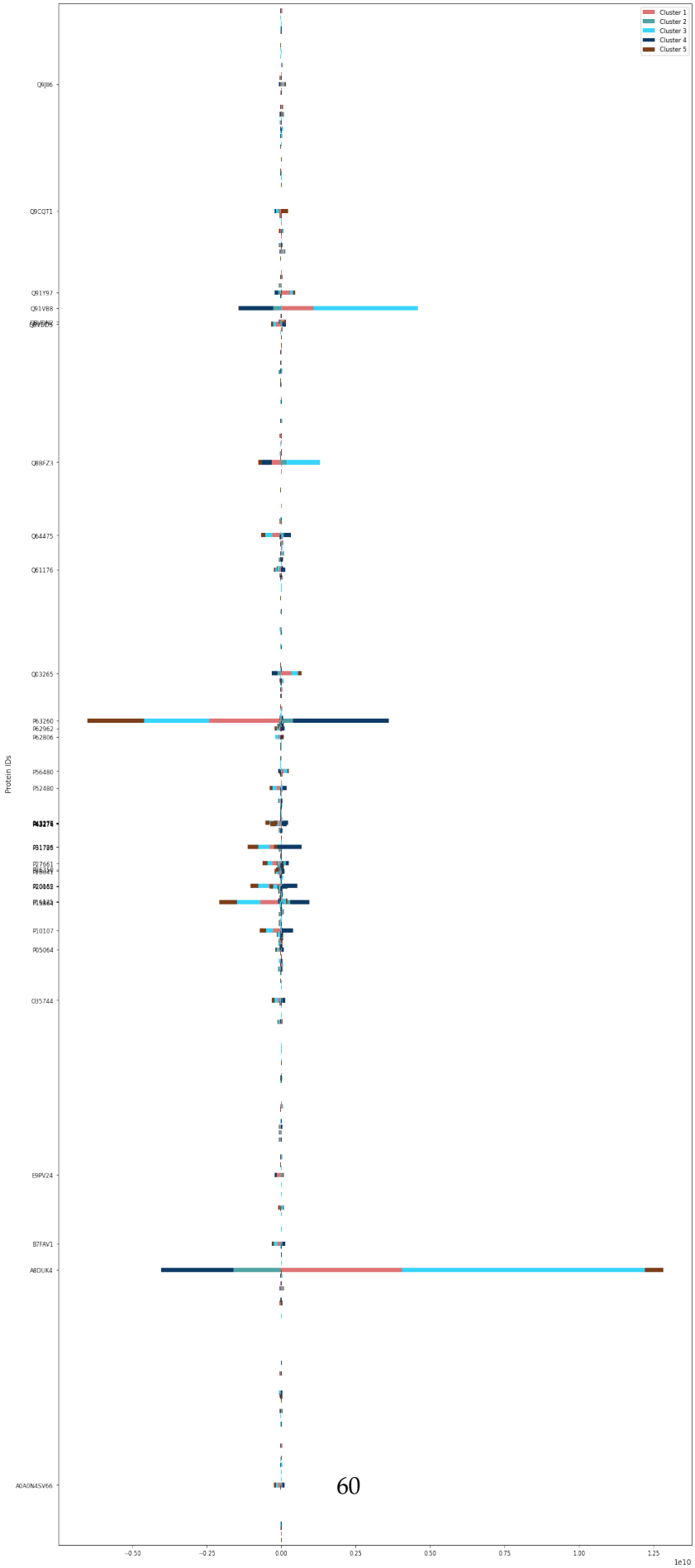
[69]:

Protein IDs	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
G3UZW7	-4.494667e+04	50722.708333	-44946.666664	-17170.416668	-44946.666667
O70589	4.249538e+04	-15230.865385	-29954.615385	-29954.615385	41181.098901
A0A140T8M9	-1.959910e+04	63072.564102	-102757.435898	-102757.435897	19429.706960
A0A075B5M7	-2.005736e+04	10793.891024	-20057.358967	16019.766022	-20057.358973
A0A075B5P3	-2.679115e+05	30169.086538	-267911.538460	441995.961538	-267911.538462
...	...	...	...	...	...
Q9Z2X1	-2.716608e+06	941850.641026	-180424.358974	517575.641026	-364245.787546
Q9Z2Z6	3.939454e+05	-255760.032051	429668.717949	-327846.282051	498845.860806
Q9Z315	-5.958347e+05	189973.381410	-820199.743590	427097.756410	-177278.315018
S4R2J6	9.806026e+03	5933.317308	614597.692308	-84808.557692	-100642.307692
S4R192	-9.051179e+04	1393.205128	-90511.794872	166923.205128	-90511.794872

[3503 rows x 5 columns]

[71]:
```
#plot centroids with top 1% proteins labeled

summed = centers_t_mouse.abs().sum(axis=1)
sorted_mylist = sorted(((v, i) for i, v in enumerate(summed)), reverse=True)[0:
 ↪35]
```

```
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=10, figsize = (16,40),␣
 ↪color=palette_kmeans)
plt.tick_params(axis='both', which='major', labelsize=8)
plt.yticks(final, finalID)
#plt.xlim(-3e11, 3e11)
plt.legend(fontsize='small')
plt.savefig('outs/' + date + '_mouse_kmeansCentroids_all_top35.svg')
```


Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Protein IDs
Q9JII6
Q9CQT1
Q91Y97
Q91VB8
Q88FZ3
Q64475
Q61176
Q03265
P63260
P62962
P62806
P56480
P52480
P31786
P27661
P10107
P05064
O35744
E9PV24
B7FAV1
A8DUK4
A0A0N4SV66
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
1.25
1e10

0.4 Gene Ontology Enrichment Scores

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

[75]:
```
allProt = pd.read_csv('outs/' + date + '_finalID_mouse_all.csv', index_col = 0)
allProt.shape
```

[75]:
```
(3503, 5)
```

[73]:
```
summed = allProt.abs().sum(axis=1)
sorted_mylist = sorted(((v, i) for i, v in enumerate(summed)), reverse=True)[0:
 ↪175]
final = [lis[1] for lis in sorted_mylist]
finalID = allProt.iloc[final].index
finalID
```

[73]:
```
Index(['A8DUK4', 'P63260', 'Q91VB8', 'P15864', 'Q8BFZ3', 'P31725', 'P20152',
       'P10107', 'Q64475', 'Q03265',
       ...
       'Q3UZ39', 'Q8K2B3', 'P54071', 'Q5EBP8', 'Q3UHL6', 'P22599', 'Q64523',
       'F8WIT2', 'Q3THW5', 'P00920'],
      dtype='object', name=' Protein IDs', length=175)
```

[77]:
```
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
```

[77]:
```
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]
```

[78]:
```
lfq.index = lfq.index.str.split(';',0).str[0]
lfq
```

[78]:
```
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]
```

[79]:
```
df = lfq.loc[finalID]
df
```

[79]:
```
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
Q8BFZ3              1.6431e+09             500220         6.6487e+06
...                        ...                ...                ...
P22599              2.9736e+07         2.3138e+07         3.2037e+07
Q64523              4.6711e+07         8.0783e+07          3.532e+07
F8WIT2              2.2663e+07         1.4908e+07         3.6708e+07
Q3THW5              3.9525e+07         3.0173e+07         3.4892e+07
```

```
P00920              2.3748e+07         5.1147e+06         2.2897e+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
Q8BFZ3              4.4522e+08         4.3693e+07               0
...                        ...                ...             ...
P22599              4.2849e+06         2.4446e+07      2.9112e+07
Q64523              5.2353e+07         8.9203e+07      3.3176e+07
F8WIT2              1.3592e+06         4.1288e+06      1.5675e+06
Q3THW5                       0         5.1085e+07      2.1225e+07
P00920                       0         9.2801e+06      1.2551e+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
Q8BFZ3                487500       1.944e+06          511750          404450
...                      ...             ...             ...             ...
P22599            3.2685e+07      2.5324e+07       3.483e+07      1.8534e+07
Q64523            3.0386e+07      1.1554e+08       4.493e+07      1.6208e+07
F8WIT2            4.3106e+06      1.0905e+07      1.2193e+07       2.284e+06
Q3THW5            1.4959e+07      3.1109e+07      2.8054e+07      2.7033e+06
P00920            9.1094e+06      4.3395e+07        3.22e+07      1.4914e+07

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
Q8BFZ3        ...       2.1978e+06     4.975e+08        802690    4.6505e+06
...           ...              ...           ...           ...           ...
P22599        ...       3.1196e+07     2.005e+07    1.6024e+07    1.3344e+07
Q64523        ...       5.4759e+07    6.4041e+07    1.1079e+08     5.712e+07
F8WIT2        ...       1.9329e+06    2.8045e+07    2.4982e+07    2.3535e+07
Q3THW5        ...       1.7866e+07    2.0146e+07    2.7599e+07    2.7402e+07
P00920        ...       1.0463e+07    2.3787e+06    2.5503e+06             0

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
Q8BFZ3           2.016e+07    1.8383e+07    1.1963e+06             0
...                    ...           ...           ...           ...
P22599          1.1475e+07    9.8682e+06    1.4829e+07    1.6949e+06
Q64523          6.8067e+07    6.8305e+07    6.2162e+07    7.2188e+07
F8WIT2           2.866e+07    4.5345e+07    4.2387e+07    1.0955e+07
Q3THW5          4.4057e+07    4.8592e+07    3.2142e+07    3.4779e+07
P00920           3.199e+06    5.1937e+06    5.0533e+06    1.5782e+06

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
Q8BFZ3           5.581e+06    4.5486e+08
...                    ...           ...
P22599          1.2092e+07    6.6191e+06
Q64523          1.1346e+08    5.8851e+07
F8WIT2          1.5971e+07    4.3004e+06
Q3THW5          5.4393e+07    1.1939e+07
P00920          2.9578e+06        989490

[175 rows x 39 columns]
```

[80]:
```
df2 = df.transpose()
df2 = df2.astype('int64')
```

[81]:
```
df3 = StandardScaler().fit_transform(df2)
df3 = pd.DataFrame(df3, columns=df2.columns, index=df2.index)
df3
```

[81]:
```
 Protein IDs           A8DUK4    P63260    Q91VB8    P15864    Q8BFZ3  \
index
4DPI_interface_2_4  -0.540090  0.021508 -0.022581 -0.236924  1.759115
4DPI_interface_2_2   0.019054 -0.337894  0.113558 -0.642223 -0.428211
4DPI_interface_2_1  -0.138874 -0.540108 -0.215675 -0.984486 -0.420023
4DPI_interface_1_2   0.608371 -1.105209 -0.325806 -0.810748  0.163989
4DPI_interface_1_1   0.174370 -1.183182  0.094234 -0.679225 -0.370694
4DPI_cortex_3_3      0.812077 -1.342470  0.739405 -1.104028 -0.428877
4DPI_cortex_3_2      1.171869 -1.082481  0.131356 -1.092429 -0.428228
4DPI_cortex_2_3      1.732487 -1.189867  0.205528 -1.029867 -0.426288
4DPI_cortex_2_2      1.705228 -0.907770  0.786771 -1.080988 -0.428195
4DPI_cortex_2_1      0.017949 -1.231981  0.234514 -0.996560 -0.428338
4DPI_cortex_1_2      3.162245 -0.851188  2.487419 -0.892101  1.517026
4DPI_cortex_1_1      2.284197 -1.246783  2.586436 -1.133441  1.497051
```

```
4DPI_SAC_3_3        -0.653982  0.011098 -0.271394  1.321765 -0.413909
4DPI_SAC_3_2        -0.555428  0.006037 -0.709916  0.867429 -0.428290
4DPI_SAC_2_3        -0.512329 -0.250133 -0.485555 -0.434520 -0.428877
4DPI_SAC_2_2        -0.480516 -0.370315 -0.212042 -0.525994  4.685493
4DPI_SAC_2_1        -0.371781  0.039127 -0.427148  0.498867 -0.426050
4DPI_SAC_1_3         0.162783  0.504720  0.867335 -0.745891  1.806121
4DPI_SAC_1_2        -0.281163  0.401154  2.496209 -0.308182 -0.422143
4DPI_SAC_1_1        -0.043099  0.343665  0.096850 -0.764444 -0.425772
10DPI_interface_5_3 -0.777652  1.419721 -0.753751  0.824938 -0.425504
10DPI_interface_4_3 -0.866100  1.606369 -1.090698  1.305138 -0.426965
10DPI_interface_4_2 -0.873578  2.688393 -1.119343  0.698258 -0.427497
10DPI_interface_4_1 -0.800096  1.695420 -0.998278  1.009681 -0.422012
10DPI_interface_3_2 -0.632316  0.429326 -0.690294  0.874819 -0.427889
10DPI_interface_3_1 -0.719082  1.182173 -0.586279  0.609317 -0.424988
10DPI_cortex_4_4     0.249718 -0.833378 -0.002022 -0.349340 -0.417555
10DPI_cortex_4_3     1.169458 -1.045285  0.320527 -0.289708 -0.428877
10DPI_cortex_3_3     0.492771 -1.152337  0.040476 -1.045689  0.826432
10DPI_cortex_3_2     1.595690 -1.288228  2.579389 -0.947973 -0.425950
10DPI_SAC_4_3       -0.946477  0.383822 -1.059250 -0.383768  0.233606
10DPI_SAC_4_2       -0.805390  0.246927 -0.955010  0.293407 -0.427808
10DPI_SAC_4_1       -0.915585  0.379811 -1.000268  0.264904 -0.422684
10DPI_SAC_3_3       -0.837421  1.047284 -0.539895  1.228074 -0.402031
10DPI_SAC_3_2       -0.835097  1.150373 -0.679543  0.438430 -0.404398
10DPI_SAC_3_1       -0.814382  1.447415 -0.838558  0.712641 -0.427284
10DPI_SAC_2_3       -0.756522  0.301217 -0.319995  3.323850 -0.428877
10DPI_SAC_2_2       -0.702037  0.341660  0.175307  0.376145 -0.421445
10DPI_SAC_2_1       -0.499269  0.311387 -0.652017  1.830863  0.176826

 Protein IDs           P31725    P20152    P10107    Q64475    Q03265  ...  \
index                                                                  ...
4DPI_interface_2_4  -0.206117 -0.124103 -0.339158  0.487461 -0.225207  ...
4DPI_interface_2_2  -0.643921 -0.213255 -0.527045 -0.433295  0.089789  ...
4DPI_interface_2_1  -0.478166 -0.400444 -0.238710 -0.155440 -0.369529  ...
4DPI_interface_1_2  -0.708444 -0.569122 -1.012065  0.125845 -0.241213  ...
4DPI_interface_1_1  -0.682565 -0.517408 -0.894347 -1.043352  0.168928  ...
4DPI_cortex_3_3     -0.716880 -1.326299 -1.073669 -1.645795  1.809980  ...
4DPI_cortex_3_2     -0.716880 -1.250061 -1.073243 -1.052557  1.360204  ...
4DPI_cortex_2_3     -0.716880 -1.103115 -1.059976 -1.384593  2.563874  ...
4DPI_cortex_2_2     -0.706149 -1.122011 -0.999513 -1.012349  1.560213  ...
4DPI_cortex_2_1     -0.716197 -1.377694 -1.067414 -1.569062  2.644484  ...
4DPI_cortex_1_2     -0.715155 -1.028298 -1.053593 -0.699643  0.906860  ...
4DPI_cortex_1_1     -0.716079 -1.165106 -1.051863 -1.281794  0.987024  ...
4DPI_SAC_3_3        -0.209229  0.137459  0.461878  1.972168 -0.525400  ...
4DPI_SAC_3_2        -0.467254  0.003198  0.190899  0.522564 -0.485897  ...
4DPI_SAC_2_3        -0.670474 -0.619947 -0.568406  0.362906 -0.125960  ...
4DPI_SAC_2_2        -0.716077 -0.592846 -1.023940  1.084758 -0.046063  ...
4DPI_SAC_2_1        -0.611808  0.228388  0.345181 -0.117659 -0.187086  ...
```

```
4DPI_SAC_1_3         0.169334 -0.339045 -0.138673 -0.801061 -0.217939  ...
4DPI_SAC_1_2         0.438196 -0.292751  0.361676 -0.780309 -0.500477  ...
4DPI_SAC_1_1         0.151373 -0.312506  0.113881 -0.266356 -0.382281  ...
10DPI_interface_5_3  3.100625  1.692848  1.770771  1.013422 -0.901580  ...
10DPI_interface_4_3  0.123680  1.805398  1.579930  1.870624 -0.827514  ...
10DPI_interface_4_2 -0.433562  1.568746  1.171000  1.245463 -0.759250  ...
10DPI_interface_4_1  0.143423  1.820800  1.449857  0.584194 -0.808740  ...
10DPI_interface_3_2  2.058175  1.296107  1.581119  1.029865 -0.899163  ...
10DPI_interface_3_1  2.024052  1.189421  1.322902  0.444367 -0.726168  ...
10DPI_cortex_4_4    -0.708326 -0.885863 -0.975620 -0.368067  0.463414  ...
10DPI_cortex_4_3    -0.714705 -0.986684 -0.967459 -0.910387  0.495828  ...
10DPI_cortex_3_3    -0.716880 -1.388084 -1.074705 -1.033478  0.805517  ...
10DPI_cortex_3_2    -0.713643 -1.307077 -1.065171 -1.234055  1.836598  ...
10DPI_SAC_4_3       -0.566626  0.631378 -0.248722 -0.012392 -0.887629  ...
10DPI_SAC_4_2       -0.557917  1.028207 -0.059768 -0.264724 -0.928505  ...
10DPI_SAC_4_1       -0.457542  0.442797  0.157376  0.156136 -0.804905  ...
10DPI_SAC_3_3        1.704192  1.220225  1.609022  0.738915 -0.772893  ...
10DPI_SAC_3_2        1.818186  0.917108  1.475871  0.665529 -0.810746  ...
10DPI_SAC_3_1        1.932369  0.795435  2.036087  1.164043 -0.729155  ...
10DPI_SAC_2_3        0.561078  0.472652  0.058406  2.246173 -0.941055  ...
10DPI_SAC_2_2        0.716264  1.252213  0.695106  0.680466 -0.734416  ...
10DPI_SAC_2_1        0.326527  0.419339  0.132100 -0.328529 -0.853945  ...

 Protein IDs           Q3UZ39    Q8K2B3    P54071    Q5EBP8    Q3UHL6  \
index
4DPI_interface_2_4   0.009676 -0.563738 -0.572222  1.176545  0.479523
4DPI_interface_2_2  -0.480505  0.312642  0.108868 -0.112737  1.020549
4DPI_interface_2_1  -0.124023 -0.195956 -0.288919 -0.287406  3.358514
4DPI_interface_1_2  -1.111607 -0.847701 -1.313029  2.526452 -1.398293
4DPI_interface_1_1  -0.794028 -0.183291 -0.396191 -0.633064 -0.763125
4DPI_cortex_3_3     -1.111607  2.389069  2.780540 -1.209991 -1.200441
4DPI_cortex_3_2     -1.075394  2.642528  1.103917 -1.254383 -1.116769
4DPI_cortex_2_3     -1.111607  0.289798  0.747838 -1.480287 -0.601693
4DPI_cortex_2_2     -0.985139  0.927562  2.473702 -1.063374 -0.635564
4DPI_cortex_2_1     -1.111607  2.781043  0.885157 -1.124328 -1.457122
4DPI_cortex_1_2     -1.042013  0.402159  0.387543 -0.610224 -1.050960
4DPI_cortex_1_1     -1.111607  0.372535  1.688444 -1.307895 -0.087518
4DPI_SAC_3_3         0.859422 -0.118965 -0.215371  1.173940 -0.019713
4DPI_SAC_3_2         0.670442 -0.205360  0.332292  0.644875  0.366192
4DPI_SAC_2_3        -0.910960 -0.847701  0.709180 -0.148376  0.388922
4DPI_SAC_2_2        -0.731997 -0.804805 -0.295466  0.099180  0.356018
4DPI_SAC_2_1         0.775021 -0.144981 -0.312418 -0.697214  1.143217
4DPI_SAC_1_3        -0.264832 -0.231101 -0.275555 -0.319366  0.441080
4DPI_SAC_1_2        -0.049937 -0.791810 -0.465452 -0.401526  0.663750
4DPI_SAC_1_1        -0.011629 -0.563415 -0.294121 -0.452111  0.864976
10DPI_interface_5_3  1.366443 -0.589826 -0.831299  2.509054 -0.034459
10DPI_interface_4_3  2.075350 -0.607916 -0.539646  0.123322  0.360525
```

10DPI_interface_4_2	1.317175	-0.534558	-0.955586	-0.056711	0.403089
10DPI_interface_4_1	2.387756	-0.684575	-0.765294	-0.331015	0.798845
10DPI_interface_3_2	1.252235	-0.518486	-0.647940	0.174290	0.137984
10DPI_interface_3_1	0.535851	-0.799975	-0.656345	0.932209	0.511461
10DPI_cortex_4_4	-1.111607	0.561984	0.415706	-0.251920	-1.130942
10DPI_cortex_4_3	-1.111607	0.144019	0.267265	-0.745192	-1.740476
10DPI_cortex_3_3	-0.989523	2.103080	2.198077	-0.986310	-1.499564
10DPI_cortex_3_2	-1.028619	1.714496	1.500717	-1.132283	-1.404205
10DPI_SAC_4_3	0.139248	-0.309391	-0.591542	-0.022605	1.212825
10DPI_SAC_4_2	1.237076	-0.215797	-0.877033	-0.136190	0.888029
10DPI_SAC_4_1	0.181858	-0.695181	-0.943190	0.053271	0.811080
10DPI_SAC_3_3	1.177360	-0.528697	-0.646102	0.962406	0.188661
10DPI_SAC_3_2	0.354553	-0.653498	-0.791834	1.968034	0.611656
10DPI_SAC_3_1	1.073498	-0.685842	-0.365068	1.402105	0.516033
10DPI_SAC_2_3	0.803599	-0.695569	-0.758504	1.101512	-0.351013
10DPI_SAC_2_2	-0.537363	-0.807827	-0.990091	-0.199726	-0.260928
10DPI_SAC_2_1	0.590650	-0.814955	-0.811028	0.117038	-0.770144

Protein IDs	P22599	Q64523	F8WIT2	Q3THW5	P00920
index					
4DPI_interface_2_4	1.060029	-0.806552	0.333282	0.265110	0.825307
4DPI_interface_2_2	0.448626	0.748592	-0.255785	-0.275693	-0.570818
4DPI_interface_2_1	1.273251	-1.326470	1.400135	-0.002805	0.761545
4DPI_interface_1_2	-1.298396	-0.549035	-1.284947	-2.020523	-0.954043
4DPI_interface_1_1	0.569832	1.132906	-1.074569	0.933596	-0.258719
4DPI_cortex_3_3	1.002206	-1.424328	-1.269124	-0.793134	-0.013643
4DPI_cortex_3_2	1.333298	-1.551672	-1.060759	-1.155481	-0.271509
4DPI_cortex_2_3	0.651192	2.335002	-0.559851	-0.221566	2.297385
4DPI_cortex_2_2	1.532065	-0.887842	-0.462015	-0.398230	1.458584
4DPI_cortex_2_1	0.021997	-2.198797	-1.214699	-1.864198	0.163408
4DPI_cortex_1_2	1.854168	-1.613975	-0.514503	-0.225499	0.553250
4DPI_cortex_1_1	2.578901	-0.316210	-0.125894	1.446816	3.128243
4DPI_SAC_3_3	-0.660388	-1.006833	-0.533417	1.745495	-0.678786
4DPI_SAC_3_2	-1.468260	1.232316	-0.343139	-0.897570	-0.639428
4DPI_SAC_2_3	0.612272	-0.370115	-0.942338	0.166398	-0.074633
4DPI_SAC_2_2	-0.560032	0.430142	-0.712401	1.940489	-0.120488
4DPI_SAC_2_1	0.853479	1.009259	-0.392664	1.248641	-0.060097
4DPI_SAC_1_3	1.264170	0.645348	0.203771	0.802443	1.905520
4DPI_SAC_1_2	-0.210963	-0.418359	-0.223654	1.123733	1.281908
4DPI_SAC_1_1	0.141256	0.209230	0.059371	1.953616	2.253478
10DPI_interface_5_3	-0.716358	0.069289	2.108762	-0.058551	0.101444
10DPI_interface_4_3	-0.542148	-0.052531	1.100246	-0.278758	-0.738001
10DPI_interface_4_2	-0.794289	0.402985	0.692647	0.261583	-0.605321
10DPI_interface_4_1	-1.136797	-0.539039	0.663326	0.134073	-0.315896
10DPI_interface_3_2	-0.844300	0.683414	1.966034	0.753521	-0.302776
10DPI_interface_3_1	-0.834747	-0.058054	1.545065	-0.882188	-0.520317
10DPI_cortex_4_4	-0.326331	0.770136	-0.645693	-0.487515	-0.636610

```
10DPI_cortex_4_3    -1.315724 -0.145369 -0.752614  0.355841 -0.561474
10DPI_cortex_3_3    -0.246361 -0.325248 -0.998138 -1.806480 -0.441809
10DPI_cortex_3_2     1.195320 -0.439218 -1.241369 -0.987376 -0.170089
10DPI_SAC_4_3        0.162477 -0.015561  0.742096 -0.855530 -0.775816
10DPI_SAC_4_2       -0.210593  2.118198  0.509432 -0.424541 -0.762959
10DPI_SAC_4_1       -0.458935 -0.331455  0.399519 -0.435933 -0.954043
10DPI_SAC_3_3       -0.632125  0.168198  0.788812  0.527184 -0.714354
10DPI_SAC_3_2       -0.781019  0.179061  2.056198  0.789432 -0.564898
10DPI_SAC_3_1       -0.321327 -0.101324  1.831509 -0.161831 -0.575418
10DPI_SAC_2_3       -1.538398  0.356292 -0.556053 -0.009339 -0.835794
10DPI_SAC_2_2       -0.574951  2.240065 -0.175040  1.124890 -0.732426
10DPI_SAC_2_1       -1.082097 -0.252447 -1.061534 -1.330120 -0.879904

[39 rows x 175 columns]
```

[83]:
```
kmeans = pd.read_csv('outs/' + date + '_combined_df_segm_pca_kmeans.csv',␣
 ↪index_col=0)
kmeans
```

[83]:
```
                     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

                  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

                     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

                    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

                    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

                    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

                    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

                    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

	A0A075B5S5;A0A0A6YXL5	A0A075B5T3;J3QKO3;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	...

index	36	37	38	\
4DPI_interface_2_4	-9.667796e+06	3.778624e+07	2.949520e-07	
4DPI_interface_2_2	4.484831e+07	3.021682e+07	2.949520e-07	
4DPI_interface_2_1	-3.927658e+07	-2.821519e+07	2.949520e-07	
4DPI_interface_1_2	1.602163e+07	1.157634e+07	2.949520e-07	
4DPI_interface_1_1	-2.771801e+07	-1.684883e+07	2.949520e-07	
4DPI_cortex_3_3	-2.290089e+07	1.351457e+07	2.949520e-07	
4DPI_cortex_3_2	1.978959e+07	-2.313470e+07	2.949520e-07	
4DPI_cortex_2_3	6.588185e+06	-3.016993e+07	2.949520e-07	
4DPI_cortex_2_2	-2.862014e+07	5.637718e+07	2.949520e-07	
4DPI_cortex_2_1	3.341529e+05	2.571811e+05	2.949520e-07	
4DPI_cortex_1_2	-4.056160e+07	-1.501557e+07	2.949520e-07	
4DPI_cortex_1_1	7.268118e+07	-1.174222e+07	2.949520e-07	
4DPI_SAC_3_3	-5.645220e+06	3.757987e+07	2.949520e-07	
4DPI_SAC_3_2	3.419064e+07	-4.386457e+07	2.949520e-07	
4DPI_SAC_2_3	2.145412e+07	1.723141e+06	2.949520e-07	
4DPI_SAC_2_2	-2.183335e+06	-1.551289e+07	2.949520e-07	
4DPI_SAC_2_1	-1.893998e+07	-5.739152e+06	2.949520e-07	
4DPI_SAC_1_3	-2.059662e+07	1.852929e+07	2.949520e-07	
4DPI_SAC_1_2	3.062973e+06	-1.035837e+07	2.949520e-07	
4DPI_SAC_1_1	-1.611010e+06	3.801341e+06	2.949520e-07	
10DPI_interface_5_3	1.634717e+07	-9.292562e+04	2.949520e-07	
10DPI_interface_4_3	-2.843180e+07	-7.313157e+06	2.949520e-07	
10DPI_interface_4_2	7.446369e+06	-2.292092e+06	2.949520e-07	
10DPI_interface_4_1	2.750186e+07	2.790485e+06	2.949520e-07	

```
10DPI_interface_3_2  1.381678e+07  2.345572e+07  2.949520e-07
10DPI_interface_3_1 -3.987217e+07 -2.743997e+07  2.949520e-07
10DPI_cortex_4_4    -7.014139e+06 -1.328333e+07  2.949520e-07
10DPI_cortex_4_3     5.219814e+06  2.369728e+07  2.949520e-07
10DPI_cortex_3_3     6.556227e+06 -7.602979e+06  2.949520e-07
10DPI_cortex_3_2    -4.925120e+06  4.877634e+05  2.949520e-07
10DPI_SAC_4_3        8.374328e+05 -5.950123e+06  2.949520e-07
10DPI_SAC_4_2       -9.686167e+06  2.496185e+07  2.949520e-07
10DPI_SAC_4_1        3.640798e+06 -1.856121e+07  2.949520e-07
10DPI_SAC_3_3       -2.675078e+06 -1.110516e+07  2.949520e-07
10DPI_SAC_3_2       -2.930084e+06 -3.858284e+06  2.949520e-07
10DPI_SAC_3_1        2.071382e+07  1.198856e+07  2.949520e-07
10DPI_SAC_2_3       -7.880436e+06 -1.545762e+07  2.949520e-07
10DPI_SAC_2_2       -2.200700e+06 -1.017980e+07  2.949520e-07
10DPI_SAC_2_1        2.285816e+06  2.499442e+07  2.949520e-07

                    Segment K-means PCA  TimePoint  BioRep  TechRep  \
index
4DPI_interface_2_4                    1       4DPI       2        4
4DPI_interface_2_2                    4       4DPI       2        2
4DPI_interface_2_1                    4       4DPI       2        1
4DPI_interface_1_2                    4       4DPI       1        2
4DPI_interface_1_1                    4       4DPI       1        1
4DPI_cortex_3_3                       0       4DPI       3        3
4DPI_cortex_3_2                       0       4DPI       3        2
4DPI_cortex_2_3                       0       4DPI       2        3
4DPI_cortex_2_2                       0       4DPI       2        2
4DPI_cortex_2_1                       4       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                   3      10DPI       5        3
10DPI_interface_4_3                   3      10DPI       4        3
10DPI_interface_4_2                   3      10DPI       4        2
10DPI_interface_4_1                   3      10DPI       4        1
10DPI_interface_3_2                   1      10DPI       3        2
10DPI_interface_3_1                   3      10DPI       3        1
10DPI_cortex_4_4                      4      10DPI       4        4
10DPI_cortex_4_3                      0      10DPI       4        3
10DPI_cortex_3_3                      4      10DPI       3        3
```

10DPI_cortex_3_2	0	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	3	10DPI	3	3
10DPI_SAC_3_2	3	10DPI	3	2
10DPI_SAC_3_1	3	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 3659 columns]
```

[84]:
```
df3['k'] = kmeans['Segment K-means PCA']
df3
```

[84]:

Protein IDs	A8DUK4	P63260	Q91VB8	P15864	Q8BFZ3	\
index						
4DPI_interface_2_4	-0.540090	0.021508	-0.022581	-0.236924	1.759115	
4DPI_interface_2_2	0.019054	-0.337894	0.113558	-0.642223	-0.428211	
4DPI_interface_2_1	-0.138874	-0.540108	-0.215675	-0.984486	-0.420023	
4DPI_interface_1_2	0.608371	-1.105209	-0.325806	-0.810748	0.163989	
4DPI_interface_1_1	0.174370	-1.183182	0.094234	-0.679225	-0.370694	
4DPI_cortex_3_3	0.812077	-1.342470	0.739405	-1.104028	-0.428877	
4DPI_cortex_3_2	1.171869	-1.082481	0.131356	-1.092429	-0.428228	
4DPI_cortex_2_3	1.732487	-1.189867	0.205528	-1.029867	-0.426288	
4DPI_cortex_2_2	1.705228	-0.907770	0.786771	-1.080988	-0.428195	
4DPI_cortex_2_1	0.017949	-1.231981	0.234514	-0.996560	-0.428338	
4DPI_cortex_1_2	3.162245	-0.851188	2.487419	-0.892101	1.517026	
4DPI_cortex_1_1	2.284197	-1.246783	2.586436	-1.133441	1.497051	
4DPI_SAC_3_3	-0.653982	0.011098	-0.271394	1.321765	-0.413909	
4DPI_SAC_3_2	-0.555428	0.006037	-0.709916	0.867429	-0.428290	
4DPI_SAC_2_3	-0.512329	-0.250133	-0.485555	-0.434520	-0.428877	
4DPI_SAC_2_2	-0.480516	-0.370315	-0.212042	-0.525994	4.685493	
4DPI_SAC_2_1	-0.371781	0.039127	-0.427148	0.498867	-0.426050	
4DPI_SAC_1_3	0.162783	0.504720	0.867335	-0.745891	1.806121	
4DPI_SAC_1_2	-0.281163	0.401154	2.496209	-0.308182	-0.422143	
4DPI_SAC_1_1	-0.043099	0.343665	0.096850	-0.764444	-0.425772	
10DPI_interface_5_3	-0.777652	1.419721	-0.753751	0.824938	-0.425504	
10DPI_interface_4_3	-0.866100	1.606369	-1.090698	1.305138	-0.426965	
10DPI_interface_4_2	-0.873578	2.688393	-1.119343	0.698258	-0.427497	
10DPI_interface_4_1	-0.800096	1.695420	-0.998278	1.009681	-0.422012	
10DPI_interface_3_2	-0.632316	0.429326	-0.690294	0.874819	-0.427889	
10DPI_interface_3_1	-0.719082	1.182173	-0.586279	0.609317	-0.424988	
10DPI_cortex_4_4	0.249718	-0.833378	-0.002022	-0.349340	-0.417555	
10DPI_cortex_4_3	1.169458	-1.045285	0.320527	-0.289708	-0.428877	
10DPI_cortex_3_3	0.492771	-1.152337	0.040476	-1.045689	0.826432	
10DPI_cortex_3_2	1.595690	-1.288228	2.579389	-0.947973	-0.425950	
10DPI_SAC_4_3	-0.946477	0.383822	-1.059250	-0.383768	0.233606	
10DPI_SAC_4_2	-0.805390	0.246927	-0.955010	0.293407	-0.427808	
10DPI_SAC_4_1	-0.915585	0.379811	-1.000268	0.264904	-0.422684	
10DPI_SAC_3_3	-0.837421	1.047284	-0.539895	1.228074	-0.402031	
10DPI_SAC_3_2	-0.835097	1.150373	-0.679543	0.438430	-0.404398	

```
10DPI_SAC_3_1       -0.814382  1.447415 -0.838558  0.712641 -0.427284
10DPI_SAC_2_3       -0.756522  0.301217 -0.319995  3.323850 -0.428877
10DPI_SAC_2_2       -0.702037  0.341660  0.175307  0.376145 -0.421445
10DPI_SAC_2_1       -0.499269  0.311387 -0.652017  1.830863  0.176826

 Protein IDs           P31725    P20152    P10107    Q64475    Q03265  ...  \
index                                                                  ...
4DPI_interface_2_4  -0.206117 -0.124103 -0.339158  0.487461 -0.225207  ...
4DPI_interface_2_2  -0.643921 -0.213255 -0.527045 -0.433295  0.089789  ...
4DPI_interface_2_1  -0.478166 -0.400444 -0.238710 -0.155440 -0.369529  ...
4DPI_interface_1_2  -0.708444 -0.569122 -1.012065  0.125845 -0.241213  ...
4DPI_interface_1_1  -0.682565 -0.517408 -0.894347 -1.043352  0.168928  ...
4DPI_cortex_3_3     -0.716880 -1.326299 -1.073669 -1.645795  1.809980  ...
4DPI_cortex_3_2     -0.716880 -1.250061 -1.073243 -1.052557  1.360204  ...
4DPI_cortex_2_3     -0.716880 -1.103115 -1.059976 -1.384593  2.563874  ...
4DPI_cortex_2_2     -0.706149 -1.122011 -0.999513 -1.012349  1.560213  ...
4DPI_cortex_2_1     -0.716197 -1.377694 -1.067414 -1.569062  2.644484  ...
4DPI_cortex_1_2     -0.715155 -1.028298 -1.053593 -0.699643  0.906860  ...
4DPI_cortex_1_1     -0.716079 -1.165106 -1.051863 -1.281794  0.987024  ...
4DPI_SAC_3_3        -0.209229  0.137459  0.461878  1.972168 -0.525400  ...
4DPI_SAC_3_2        -0.467254  0.003198  0.190899  0.522564 -0.485897  ...
4DPI_SAC_2_3        -0.670474 -0.619947 -0.568406  0.362906 -0.125960  ...
4DPI_SAC_2_2        -0.716077 -0.592846 -1.023940  1.084758 -0.046063  ...
4DPI_SAC_2_1        -0.611808  0.228388  0.345181 -0.117659 -0.187086  ...
4DPI_SAC_1_3         0.169334 -0.339045 -0.138673 -0.801061 -0.217939  ...
4DPI_SAC_1_2         0.438196 -0.292751  0.361676 -0.780309 -0.500477  ...
4DPI_SAC_1_1         0.151373 -0.312506  0.113881 -0.266356 -0.382281  ...
10DPI_interface_5_3  3.100625  1.692848  1.770771  1.013422 -0.901580  ...
10DPI_interface_4_3  0.123680  1.805398  1.579930  1.870624 -0.827514  ...
10DPI_interface_4_2 -0.433562  1.568746  1.171000  1.245463 -0.759250  ...
10DPI_interface_4_1  0.143423  1.820800  1.449857  0.584194 -0.808740  ...
10DPI_interface_3_2  2.058175  1.296107  1.581119  1.029865 -0.899163  ...
10DPI_interface_3_1  2.024052  1.189421  1.322902  0.444367 -0.726168  ...
10DPI_cortex_4_4    -0.708326 -0.885863 -0.975620 -0.368067  0.463414  ...
10DPI_cortex_4_3    -0.714705 -0.986684 -0.967459 -0.910387  0.495828  ...
10DPI_cortex_3_3    -0.716880 -1.388084 -1.074705 -1.033478  0.805517  ...
10DPI_cortex_3_2    -0.713643 -1.307077 -1.065171 -1.234055  1.836598  ...
10DPI_SAC_4_3       -0.566626  0.631378 -0.248722 -0.012392 -0.887629  ...
10DPI_SAC_4_2       -0.557917  1.028207 -0.059768 -0.264724 -0.928505  ...
10DPI_SAC_4_1       -0.457542  0.442797  0.157376  0.156136 -0.804905  ...
10DPI_SAC_3_3        1.704192  1.220225  1.609022  0.738915 -0.772893  ...
10DPI_SAC_3_2        1.818186  0.917108  1.475871  0.665529 -0.810746  ...
10DPI_SAC_3_1        1.932369  0.795435  2.036087  1.164043 -0.729155  ...
10DPI_SAC_2_3        0.561078  0.472652  0.058406  2.246173 -0.941055  ...
10DPI_SAC_2_2        0.716264  1.252213  0.695106  0.680466 -0.734416  ...
10DPI_SAC_2_1        0.326527  0.419339  0.132100 -0.328529 -0.853945  ...
```

Protein IDs	Q8K2B3	P54071	Q5EBP8	Q3UHL6	P22599	\
index						
4DPI_interface_2_4	-0.563738	-0.572222	1.176545	0.479523	1.060029	
4DPI_interface_2_2	0.312642	0.108868	-0.112737	1.020549	0.448626	
4DPI_interface_2_1	-0.195956	-0.288919	-0.287406	3.358514	1.273251	
4DPI_interface_1_2	-0.847701	-1.313029	2.526452	-1.398293	-1.298396	
4DPI_interface_1_1	-0.183291	-0.396191	-0.633064	-0.763125	0.569832	
4DPI_cortex_3_3	2.389069	2.780540	-1.209991	-1.200441	1.002206	
4DPI_cortex_3_2	2.642528	1.103917	-1.254383	-1.116769	1.333298	
4DPI_cortex_2_3	0.289798	0.747838	-1.480287	-0.601693	0.651192	
4DPI_cortex_2_2	0.927562	2.473702	-1.063374	-0.635564	1.532065	
4DPI_cortex_2_1	2.781043	0.885157	-1.124328	-1.457122	0.021997	
4DPI_cortex_1_2	0.402159	0.387543	-0.610224	-1.050960	1.854168	
4DPI_cortex_1_1	0.372535	1.688444	-1.307895	-0.087518	2.578901	
4DPI_SAC_3_3	-0.118965	-0.215371	1.173940	-0.019713	-0.660388	
4DPI_SAC_3_2	-0.205360	0.332292	0.644875	0.366192	-1.468260	
4DPI_SAC_2_3	-0.847701	0.709180	-0.148376	0.388922	0.612272	
4DPI_SAC_2_2	-0.804805	-0.295466	0.099180	0.356018	-0.560032	
4DPI_SAC_2_1	-0.144981	-0.312418	-0.697214	1.143217	0.853479	
4DPI_SAC_1_3	-0.231101	-0.275555	-0.319366	0.441080	1.264170	
4DPI_SAC_1_2	-0.791810	-0.465452	-0.401526	0.663750	-0.210963	
4DPI_SAC_1_1	-0.563415	-0.294121	-0.452111	0.864976	0.141256	
10DPI_interface_5_3	-0.589826	-0.831299	2.509054	-0.034459	-0.716358	
10DPI_interface_4_3	-0.607916	-0.539646	0.123322	0.360525	-0.542148	
10DPI_interface_4_2	-0.534558	-0.955586	-0.056711	0.403089	-0.794289	
10DPI_interface_4_1	-0.684575	-0.765294	-0.331015	0.798845	-1.136797	
10DPI_interface_3_2	-0.518486	-0.647940	0.174290	0.137984	-0.844300	
10DPI_interface_3_1	-0.799975	-0.656345	0.932209	0.511461	-0.834747	
10DPI_cortex_4_4	0.561984	0.415706	-0.251920	-1.130942	-0.326331	
10DPI_cortex_4_3	0.144019	0.267265	-0.745192	-1.740476	-1.315724	
10DPI_cortex_3_3	2.103080	2.198077	-0.986310	-1.499564	-0.246361	
10DPI_cortex_3_2	1.714496	1.500717	-1.132283	-1.404205	1.195320	
10DPI_SAC_4_3	-0.309391	-0.591542	-0.022605	1.212825	0.162477	
10DPI_SAC_4_2	-0.215797	-0.877033	-0.136190	0.888029	-0.210593	
10DPI_SAC_4_1	-0.695181	-0.943190	0.053271	0.811080	-0.458935	
10DPI_SAC_3_3	-0.528697	-0.646102	0.962406	0.188661	-0.632125	
10DPI_SAC_3_2	-0.653498	-0.791834	1.968034	0.611656	-0.781019	
10DPI_SAC_3_1	-0.685842	-0.365068	1.402105	0.516033	-0.321327	
10DPI_SAC_2_3	-0.695569	-0.758504	1.101512	-0.351013	-1.538398	
10DPI_SAC_2_2	-0.807827	-0.990091	-0.199726	-0.260928	-0.574951	
10DPI_SAC_2_1	-0.814955	-0.811028	0.117038	-0.770144	-1.082097	

Protein IDs	Q64523	F8WIT2	Q3THW5	P00920	k
index					
4DPI_interface_2_4	-0.806552	0.333282	0.265110	0.825307	1
4DPI_interface_2_2	0.748592	-0.255785	-0.275693	-0.570818	4
4DPI_interface_2_1	-1.326470	1.400135	-0.002805	0.761545	4

```
4DPI_interface_1_2  -0.549035 -1.284947 -2.020523 -0.954043  4
4DPI_interface_1_1   1.132906 -1.074569  0.933596 -0.258719  4
4DPI_cortex_3_3     -1.424328 -1.269124 -0.793134 -0.013643  0
4DPI_cortex_3_2     -1.551672 -1.060759 -1.155481 -0.271509  0
4DPI_cortex_2_3      2.335002 -0.559851 -0.221566  2.297385  0
4DPI_cortex_2_2     -0.887842 -0.462015 -0.398230  1.458584  0
4DPI_cortex_2_1     -2.198797 -1.214699 -1.864198  0.163408  4
4DPI_cortex_1_2     -1.613975 -0.514503 -0.225499  0.553250  2
4DPI_cortex_1_1     -0.316210 -0.125894  1.446816  3.128243  2
4DPI_SAC_3_3        -1.006833 -0.533417  1.745495 -0.678786  1
4DPI_SAC_3_2         1.232316 -0.343139 -0.897570 -0.639428  1
4DPI_SAC_2_3        -0.370115 -0.942338  0.166398 -0.074633  1
4DPI_SAC_2_2         0.430142 -0.712401  1.940489 -0.120488  1
4DPI_SAC_2_1         1.009259 -0.392664  1.248641 -0.060097  1
4DPI_SAC_1_3         0.645348  0.203771  0.802443  1.905520  1
4DPI_SAC_1_2        -0.418359 -0.223654  1.123733  1.281908  1
4DPI_SAC_1_1         0.209230  0.059371  1.953616  2.253478  1
10DPI_interface_5_3  0.069289  2.108762 -0.058551  0.101444  3
10DPI_interface_4_3 -0.052531  1.100246 -0.278758 -0.738001  3
10DPI_interface_4_2  0.402985  0.692647  0.261583 -0.605321  3
10DPI_interface_4_1 -0.539039  0.663326  0.134073 -0.315896  3
10DPI_interface_3_2  0.683414  1.966034  0.753521 -0.302776  1
10DPI_interface_3_1 -0.058054  1.545065 -0.882188 -0.520317  3
10DPI_cortex_4_4     0.770136 -0.645693 -0.487515 -0.636610  4
10DPI_cortex_4_3    -0.145369 -0.752614  0.355841 -0.561474  0
10DPI_cortex_3_3    -0.325248 -0.998138 -1.806480 -0.441809  4
10DPI_cortex_3_2    -0.439218 -1.241369 -0.987376 -0.170089  0
10DPI_SAC_4_3       -0.015561  0.742096 -0.855530 -0.775816  1
10DPI_SAC_4_2        2.118198  0.509432 -0.424541 -0.762959  1
10DPI_SAC_4_1       -0.331455  0.399519 -0.435933 -0.954043  1
10DPI_SAC_3_3        0.168198  0.788812  0.527184 -0.714354  3
10DPI_SAC_3_2        0.179061  2.056198  0.789432 -0.564898  3
10DPI_SAC_3_1       -0.101324  1.831509 -0.161831 -0.575418  3
10DPI_SAC_2_3        0.356292 -0.556053 -0.009339 -0.835794  1
10DPI_SAC_2_2        2.240065 -0.175040  1.124890 -0.732426  1
10DPI_SAC_2_1       -0.252447 -1.061534 -1.330120 -0.879904  1

[39 rows x 176 columns]
```

[85]:
```
df3['k'] = df3['k'].astype('category')
df3.dtypes
```

[85]:
```
 Protein IDs
A8DUK4     float64
P63260     float64
Q91VB8     float64
P15864     float64
Q8BFZ3     float64
```

```
            ...
Q64523     float64
F8WIT2     float64
Q3THW5     float64
P00920     float64
k         category
Length: 176, dtype: object
```

[86]:
```
df4 = df3.groupby(['k']).mean()
df4
```

[86]:
```
 Protein IDs    A8DUK4    P63260    Q91VB8    P15864    Q8BFZ3    P31725  \
k
0              1.364468 -1.142684  0.793830 -0.924165 -0.427736 -0.714189
1             -0.533325  0.193813 -0.198111  0.390770  0.249214 -0.002631
2              2.723221 -1.048986  2.536927 -1.012771  1.507039 -0.715617
3             -0.815426  1.529644 -0.825793  0.853310 -0.420085  1.301621
4              0.203337 -0.912013 -0.008674 -0.786896 -0.153486 -0.664928

 Protein IDs    P20152    P10107    Q64475    Q03265  ...    Q3UZ39    Q8K2B3  \
k                                                    ...
0             -1.182541 -1.039839 -1.206623  1.604449  ... -1.070662  1.351245
1              0.226909  0.107435  0.373217 -0.546620  ...  0.250782 -0.520568
2             -1.096702 -1.052728 -0.990719  0.946942  ... -1.076810  0.387347
3              1.376248  1.551930  0.965820 -0.792006  ...  1.285998 -0.635611
4             -0.764553 -0.827130 -0.639550  0.508770  ... -0.817557  0.647400

 Protein IDs    P54071    Q5EBP8    Q3UHL6    P22599    Q64523    F8WIT2  \
k
0              1.478997 -1.147585 -1.116524  0.733060 -0.352238 -0.890956
1             -0.438029  0.135221  0.396987 -0.219702  0.357684 -0.045421
2              1.037993 -0.959059 -0.569239  2.216534 -0.965092 -0.320199
3             -0.693897  0.938675  0.419476 -0.719851  0.008573  1.348321
4              0.229953 -0.124187 -0.267141  0.063231 -0.249702 -0.581957

 Protein IDs    Q3THW5    P00920
k
0             -0.533324  0.456542
1              0.448206 -0.034434
2              0.610658  1.840746
3              0.041368 -0.491595
4             -0.789088 -0.276721

[5 rows x 175 columns]
```

[87]:
```
df5 = df4[df4>0]
df5
```

[87]:
```
 Protein IDs    A8DUK4    P63260    Q91VB8   P15864    Q8BFZ3    P31725  \
k
0             1.364468       NaN  0.793830      NaN       NaN       NaN
1                  NaN  0.193813       NaN  0.39077  0.249214       NaN
2             2.723221       NaN  2.536927      NaN  1.507039       NaN
3                  NaN  1.529644       NaN  0.85331       NaN  1.301621
4             0.203337       NaN       NaN      NaN       NaN       NaN

 Protein IDs    P20152    P10107    Q64475    Q03265  ...    Q3UZ39    Q8K2B3  \
k                                                     ...
0                  NaN       NaN       NaN  1.604449  ...       NaN  1.351245
1             0.226909  0.107435  0.373217       NaN  ...  0.250782       NaN
2                  NaN       NaN       NaN  0.946942  ...       NaN  0.387347
3             1.376248  1.551930  0.965820       NaN  ...  1.285998       NaN
4                  NaN       NaN       NaN  0.508770  ...       NaN  0.647400

 Protein IDs    P54071    Q5EBP8    Q3UHL6    P22599    Q64523    F8WIT2  \
k
0             1.478997       NaN       NaN  0.733060       NaN       NaN
1                  NaN  0.135221  0.396987       NaN  0.357684       NaN
2             1.037993       NaN       NaN  2.216534       NaN       NaN
3                  NaN  0.938675  0.419476       NaN  0.008573  1.348321
4             0.229953       NaN       NaN  0.063231       NaN       NaN

 Protein IDs    Q3THW5    P00920
k
0                  NaN  0.456542
1             0.448206       NaN
2             0.610658  1.840746
3             0.041368       NaN
4                  NaN       NaN

[5 rows x 175 columns]
```

[88]:
```
proteinIDs = []

for i in range(0,5):
    proteinList = df5.iloc[i][-df5.iloc[i].isnull()].index
    proteinIDs.append(proteinList)

proteinIDs
```

[88]:
```
[Index(['A8DUK4', 'Q91VB8', 'Q03265', 'Q91Y97', 'P31786', 'P56480', 'P16125',
        'Q8VDN2', 'Q9JII6', 'Q99KI0', 'P63038', 'Q9DCW4', 'P13745', 'P14094',
        'P26443', 'A2ARV4', 'Q64433', 'Q61838', 'P99029', 'G5E8R3', 'P08249',
        'Q62468', 'P24270', 'Q64442', 'Q61646', 'O88844', 'Q9EQ20', 'Q9DBM2',
        'Q9D881', 'Q9Z2I8', 'P45952', 'P70441', 'P97450', 'P17751', 'Q8K0L3',
        'Q921I1', 'A3KGU5', 'P38647', 'P51881', 'P56391', 'P14152',
```

```
       'A0A0R4J1V0', 'Q62433', 'P10922', 'Q8BWT1', 'E9PZF0', 'P05202',
       'A0A1L1SV25', 'Q9DB20', 'F8WIP8', 'O08749', 'P48758', 'Q60932',
       'Q9DCX2', 'P16460', 'P07724', 'P17182', 'P56135', 'Q62261', 'Q91X72',
       'Q8VC30', 'Q9QXD6', 'Q8QZT1', 'P52760', 'Q9JLJ2', 'P47199', 'Q8R0Y6',
       'Q8CHT0', 'Q8K2B3', 'P54071', 'P22599', 'P00920'],
      dtype='object', name=' Protein IDs'),
Index(['P63260', 'P15864', 'Q8BFZ3', 'P20152', 'P10107', 'Q64475', 'P27661',
       'P43276', 'P20065', 'P52480', 'P43274', 'Q8VDD5', 'B7FAV1', 'O35744',
       'Q61176', 'A0A0N4SV66', 'P43277', 'P26041', 'P62962', 'E9PV24',
       'P43275', 'P05064', 'P26350', 'P09528', 'Q61233', 'P27773', 'D3Z2H9',
       'P63017', 'P20029', 'P26039', 'O08692', 'P07356', 'P49710', 'P18760',
       'Q99PT1', 'O88569', 'O89053', 'P08071', 'Q99LB4', 'E9Q7Q3', 'Q61599',
       'Q91Z25', 'P40124', 'E9Q3W4', 'P40142', 'P13020', 'P11247', 'P30681',
       'P11499', 'A2AL12', 'E9Q616', 'P14733', 'Q6IRU2', 'Q8K0E8', 'Q3UER8',
       'Q8BH61', 'P19973', 'P05213', 'P17742', 'Q9DCD0', 'P26645', 'P06745',
       'P56391', 'P63101', 'Q9JKF1', 'P24527', 'Q61029', 'A0A1W2P6F6',
       'P61979', 'A0A2R8VHH0', 'O35639', 'A0A0R4J0I1', 'P08113', 'P99024',
       'P09103', 'Q00612', 'Q8BMK4', 'P29351', 'Q9ET01', 'P17182', 'Q9DBJ1',
       'Q6KCD5', 'Q3TML0', 'P62960', 'Q9Z0E6', 'Q8VEK3', 'A0A1B0GR11',
       'Q05816', 'P19324', 'Q3UZ39', 'Q5EBP8', 'Q3UHL6', 'Q64523', 'Q3THW5'],
      dtype='object', name=' Protein IDs'),
Index(['A8DUK4', 'Q91VB8', 'Q8BFZ3', 'Q03265', 'Q91Y97', 'P31786', 'P56480',
       'P16125', 'Q8VDN2', 'Q9JII6', 'P26350', 'Q99KI0', 'D3Z2H9', 'P63038',
       'Q9DCW4', 'Q99PT1', 'P13745', 'P14094', 'P26443', 'A2ARV4', 'P68033',
       'Q64433', 'Q61838', 'P99029', 'A0A1D5RLD8', 'G5E8R3', 'P08249',
       'Q62468', 'P24270', 'Q64442', 'Q61646', 'O88844', 'Q9EQ20', 'Q9DBM2',
       'Q9D881', 'Q9Z2I8', 'P45952', 'P97450', 'P05213', 'P17751', 'P17742',
       'Q8K0L3', 'Q921I1', 'A3KGU5', 'P38647', 'P51881', 'P10126', 'P14152',
       'A0A0R4J1V0', 'Q62433', 'Q8BWT1', 'E9PZF0', 'P05202', 'A0A1L1SV25',
       'P61979', 'Q9DB20', 'A0A0R4J0I1', 'F8WIP8', 'O08749', 'P48758',
       'Q60932', 'Q9DCX2', 'P16460', 'P07724', 'P56135', 'Q62261', 'Q91X72',
       'Q8VC30', 'Q9QXD6', 'Q8QZT1', 'P62960', 'P52760', 'Q9JLJ2', 'P47199',
       'Q8R0Y6', 'Q8CHT0', 'Q8K2B3', 'P54071', 'P22599', 'Q3THW5', 'P00920'],
      dtype='object', name=' Protein IDs'),
Index(['P63260', 'P15864', 'P31725', 'P20152', 'P10107', 'Q64475', 'P27661',
       'P43276', 'P20065', 'P52480', 'P43274', 'Q8VDD5', 'B7FAV1', 'O35744',
       'Q61176', 'A0A0N4SV66', 'P43277', 'P26041', 'P62962', 'P43275',
       'P05064', 'P62806', 'P27005', 'P09528', 'Q61233', 'P27773', 'D3Z2H9',
       'P63017', 'P20029', 'P26039', 'O08692', 'P07356', 'P49710', 'P18760',
       'Q99PT1', 'O88569', 'A0A1W2P768', 'P68033', 'O89053', 'P08071',
       'Q99LB4', 'E9Q7Q3', 'A0A1D5RLD8', 'Q61599', 'Q91Z25', 'P40124',
       'E9Q3W4', 'P40142', 'P13020', 'P11247', 'P30681', 'P11499', 'A2AL12',
       'P14733', 'Q6IRU2', 'Q8BH61', 'P19973', 'P05213', 'P17742', 'Q9DCD0',
       'P26645', 'P06745', 'Q9CQI6', 'P63101', 'Q9JKF1', 'P10126', 'P24527',
       'Q61029', 'A0A1W2P6F6', 'P61979', 'A0A2R8VHH0', 'O35639', 'P08113',
       'P99024', 'P09103', 'Q00612', 'Q8BMK4', 'P29351', 'Q9ET01', 'P17182',
       'Q9DBJ1', 'P48036', 'Q6KCD5', 'Q3TML0', 'P62960', 'Q9Z0E6', 'Q8VEK3',
```

```
       'A0A1B0GR11', 'Q05816', 'P19324', 'Q3UZ39', 'Q5EBP8', 'Q3UHL6',
       'Q64523', 'F8WIT2', 'Q3THW5'],
      dtype='object', name=' Protein IDs'),
 Index(['A8DUK4', 'Q03265', 'Q91Y97', 'Q9CQT1', 'P56480', 'E9PV24', 'P16125',
       'P62806', 'Q8VDN2', 'Q9JII6', 'Q99KI0', 'P63038', 'Q9DCW4', 'P13745',
       'P14094', 'P26443', 'A2ARV4', 'A0A1W2P768', 'Q64433', 'Q61838',
       'G5E8R3', 'P08249', 'Q62468', 'P24270', 'Q64442', 'Q61646', 'O88844',
       'Q9EQ20', 'Q9DBM2', 'Q8K0E8', 'Q3UER8', 'Q9D881', 'Q9Z2I8', 'P45952',
       'P70441', 'Q8K0L3', 'Q921I1', 'A3KGU5', 'P38647', 'P51881', 'P63101',
       'P14152', 'A0A0R4J1V0', 'Q62433', 'P10922', 'Q8BWT1', 'E9PZF0',
       'P05202', 'A0A1L1SV25', 'Q9DB20', 'A0A0R4J0I1', 'F8WIP8', 'O08749',
       'P48758', 'Q60932', 'P09103', 'Q9DCX2', 'P07724', 'P56135', 'Q62261',
       'Q91X72', 'Q8VC30', 'Q9QXD6', 'Q8QZT1', 'P52760', 'Q9JLJ2', 'P47199',
       'Q8R0Y6', 'Q8CHT0', 'Q8K2B3', 'P54071', 'P22599'],
      dtype='object', name=' Protein IDs')]
```

[89]:
```
test = pd.DataFrame(proteinIDs)

test.to_csv('outs/' + date + '_topProteins_175.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

[90]:
```
path = 'outs/'

fileNames = ['/*proteinClass_175.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('_')[0]
        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) + '_175.csv')
```

[91]:
```
#save enrichment scores

geneOntologyTerm = ['ProteinClass']

for go in geneOntologyTerm:
    df = pd.read_csv('outs/panther_combined_gene_ontology_' + str(go) + '_175.
 ↪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) + '_175.
 ↪csv')
```

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

[93]:
```
summed = allProt.abs().sum(axis=1)
sorted_mylist = sorted(((v, i) for i, v in enumerate(summed)), reverse=True)[0:
 ↪35]
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')
```

[ ]:
```

```