PEM data overview

SURF64

25/06/2018

Available datasets

- A total of 6 files:
  - Covariate data: environmental measurements (e.g. humidity, temperature), individual information (e.g. age, BMI, gender) and technical covariates (with respect to an OMIC platform)
  - Exposures: measured and background PM2.5 (Particulate Matter 2.5) and UFP (Ultra Fine Particles)
  - Inflammatory proteins (p=13)
  - Metabolites (p=11,217)
  - Transcripts (p=23,557)
  - DNA methylation (p=485,512)

Covariates

```r
covar=read.table('Covariates.txt', sep='\t')
dim(covar)
```

```
## [1] 526  19
```

```r
head(covar)
```

```
##      subjectidp      temp   relhum gender  age session  city   bmi season
## 001A       001A  7.916000 81.16000    Men 58.5       A Turin 22.60 winter
## 001B       001B 16.091667 73.66667    Men 58.5       B Turin 22.92 spring
## 001C       001C 11.770833 34.20833    Men 58.5       C Turin 22.92 Autumn
## 002A       002A  7.630435 82.60870  Women 58.6       A Turin 25.48 winter
## 002B       002B 16.025000 74.37500  Women 58.6       B Turin 25.48 spring
## 002C       002C 14.795833 91.58333  Women 58.6       C Turin 25.48 Autumn
##      physical_act id       date education_adj          chip position plate
## 001A     1.795932  1  1/21/2014             2 200053730042   R03C02     6
## 001B     1.514928  1 05/06/2014             2 200053730042   R01C02     6
## 001C     1.436621  1 10/23/2014             2 200053730042   R01C01    NA
## 002A     1.677509  2  1/21/2014             2 101231030012   R03C02     6
## 002B     1.738323  2 05/06/2014             2           NA     <NA>     6
## 002C     1.738323  2       <NA>             2 101231030012   R03C01    NA
##      isolation labeling hybridization
## 001A         1        1             1
## 001B         1        1             1
## 001C        NA       NA            NA
## 002A         1        1             1
## 002B         1        1             1
## 002C        NA       NA            NA
```

Exposures

```
expo=read.table('Exposures.txt', sep='\t')
dim(expo) # not the same size
```

```
## [1] 521  10
```

```
head(expo)
```

```
##      subjectidp id pm25_adj_p pm25abs_adj_p pm25_adj_o pm25abs_adj_o
## 001A       001A  1      22.78         4.004      19.75         5.398
## 001B       001B  1      10.26         1.690      11.10         1.615
## 001C       001C  1      16.32         1.661       6.37         1.571
## 002A       002A  2      13.63         3.578      19.47         4.279
## 002B       002B  2      11.61         1.795      12.19         2.072
## 002C       002C  2      15.69         2.228      25.75         3.298
##      pncmedian modeledpnc modeledpm25abs modeledpm25
## 001A     12042   13185.75       3.297130    30.95109
## 001B     12707   13185.75       3.297130    30.95109
## 001C     11164   13185.75       3.297130    30.95109
## 002A     12717   13122.62       3.327551    30.95240
## 002B     11308   13122.62       3.327551    30.95240
## 002C      9176   13122.62       3.327551    30.95240
```

Proteins

```
proteins=read.table('Proteins.txt', sep='\t')
dim(proteins) # not the same number of observations
```

```
## [1] 336  13
```

```
range(proteins)
```

```
## [1]      0.07 224562.10
```

$\Rightarrow$ Only positive values: log-transformation of the data?

Proteins

```
head(proteins[,1:10])
```

```
##       IL.8.C_imputed EGF.2_imputed MPO.5_imputed VEGF_imputed IL.17_imputed
## 201A          4.61         48.32      25353.53        38.10          3.84
## 201B          6.47         17.82      22707.26        27.68          4.07
## 202A          7.35         29.97      19043.52        69.94          6.79
## 202B          4.28          2.96       9966.67        28.98          5.08
## 203A          6.03         86.09      32180.47        81.81          7.14
## 203B          2.97          3.39      16404.33        31.33          6.10
##       MDC.CC_imputed G.CSF_imputed Eotaxi_imputed CRP_imputed IP.10_imputed
## 201A         379.51          1.91          94.94    3669.236         21.22
## 201B         355.46          3.63          82.13    4041.735         26.67
## 202A         780.20          4.20         167.66    1479.613         48.73
## 202B         620.31          3.63          93.14    1862.627         65.05
## 203A         687.28          3.63          85.83    1390.502         13.98
## 203B         372.38          1.02          78.39    1321.342         17.92
```

$\Rightarrow$ The proteins are on different scales: rescaling of the data?

Metabolites

```
metabolites=read.table('Metabolites.txt', sep='\t')
dim(metabolites) # more variables
```

```
## [1]   400 11217
```

```
range(metabolites, na.rm=TRUE)
```

```
## [1]    10000 145932752
```