BMI 206 Project Data Summary for Paper "Uncovering disease-disease relationships through the incomplete interactome"

Kathleen Keough, Beau Norgeot, Sasha Targ

To obtain the relevant data for this paper, we downloaded the data archive from the online supplement on the Science website (https://www.sciencemag.org/content/347/6224/1257601/suppl/DC1), which contains datasets in the format of tables with tab-separated columns (.tsv files):

- 1) Human interactome dataset listing pairwise interactions of genes and the types of interactions,

```
> head(DataS1_interactome,5)
  gene_ID_1 gene_ID_2              data_source.s.
1      1394      2778                  literature
2 100290337      4214                  literature
3    122704     54460                   complexes
4      4790     79155 binary;literature;signaling
5      2597        70                   signaling
```

containing 141,295 pairs of genes.

```
> nrow(DataS1_interactome)
[1] 141295
```

Here is a summary quantifying the number of instance of each types of interactions and demonstrating how one or more interactions occurs for each gene-gene pair:

```
> table(col_3_split)
col_3_split
            binary  complexes     kinase literature  metabolic regulatory  signaling
    795362      28653      31276       6066      88349       5325       1335      32706
> head(col_3_split,5)
     [,1]         [,2]         [,3]        [,4] [,5] [,6] [,7]
[1,] "literature" ""           ""          ""   ""   ""   ""
[2,] "literature" ""           ""          ""   ""   ""   ""
[3,] "complexes"  ""           ""          ""   ""   ""   ""
[4,] "binary"     "literature" "signaling" ""   ""   ""   ""
[5,] "signaling"  ""           ""          ""   ""   ""   ""
```

- 2) Disease genes dataset.
- 3) Network properties dataset for the diseases considered in the study. Here are summary statistics describing number of genes ascribed to each disease and other parts of the dataset:

```
> dataSummary3
                       disease    number_of_all_genes    LCC_size       LCC_z_score
 abnormalities, multiple   :  1   Min.   :  20.00     Min.   :  1.00   Min.   :-0.500
 actinomycetales infections:  1   1st Qu.:  29.00     1st Qu.:  3.00   1st Qu.: 2.000
 adenocarcinoma            :  1   Median :  48.00     Median :  6.00   Median : 4.900
 adnexal diseases          :  1   Mean   :  99.58     Mean   : 25.89   Mean   : 6.209
 adrenal gland diseases    :  1   3rd Qu.:  99.00     3rd Qu.: 15.50   3rd Qu.: 8.350
 albuminuria               :  1   Max.   :1168.00     Max.   :730.00   Max.   :31.100
 (Other)                   :293
      d_s        d_s_Glass_delta     d_s_p_value
 Min.   :1.050   Min.   :-2.2939   Min.   :0.0000000
 1st Qu.:1.566   1st Qu.:-1.0891   1st Qu.:0.0000000
 Median :1.788   Median :-0.7829   Median :0.0000003
 Mean   :1.823   Mean   :-0.8020   Mean   :0.0251731
 3rd Qu.:2.000   3rd Qu.:-0.5424   3rd Qu.:0.0014005
 Max.   :3.143   Max.   : 0.8095   Max.   :0.4959045
```

- 4) Network relationships for all the pairs of diseases considered in the study, summarized here:

```
> summary(DataS4_disease_pairs)
                  disease_A                  disease_B     s_AB..observed.   d_AB..observed. z..full.rand.
 abnormalities, multiple   :  298   vision disorders :  298   Min.   :-3.1429   Min.   :0.000   Min.   :-50.761
 actinomycetales infections:  297   virus diseases   :  297   1st Qu.: 0.2037   1st Qu.:2.036   1st Qu.:  1.700
 adenocarcinoma            :  296   vasculitis       :  296   Median : 0.3646   Median :2.209   Median :  3.468
 adnexal diseases          :  295   vascular diseases:  295   Mean   : 0.3672   Mean   :2.190   Mean   :  3.520
 adrenal gland diseases    :  294   varicose veins   :  294   3rd Qu.: 0.5437   3rd Qu.:2.386   3rd Qu.:  5.489
 albuminuria               :  293   uveitis, anterior:  293   Max.   : 1.8000   Max.   :3.303   Max.   : 18.072
 (Other)                   :42778   (Other)          :42778
 p..full.rand.     q..full.rand.      z..MeSH.rand.     p..MeSH.rand.     q..MeSH.rand.
 Min.   :0.00000   Min.   :0.0000000   Min.   :-6.247   Min.   :0.00000   Min.   :0.0000000
 1st Qu.:0.00000   1st Qu.:0.0000009   1st Qu.: 2.146   1st Qu.:0.00000   1st Qu.:0.0000004
 Median :0.00000   Median :0.0004255   Median : 3.829   Median :0.00000   Median :0.0002416
 Mean   :0.06091   Mean   :0.0859514   Mean   : 4.252   Mean   :0.07358   Mean   :0.0745713
 3rd Qu.:0.02682   3rd Qu.:0.0516171   3rd Qu.: 5.936   3rd Qu.:0.02000   3rd Qu.:0.0311266
 Max.   :1.00000   Max.   :0.7482555   Max.   :18.903   Max.   :1.00000   Max.   :0.7458767
```