**Digital Commons – Data Fields**

**Author(s) AND their Email addresses**

Abby Benninghoff, [abby.benninghoff@usu.edu](mailto:abby.benninghoff@usu.edu)

**Title of Your Dataset or Journal article:**

Consumption of the total Western diet promotes colitis and inflammation-associated colorectal cancer in mice. Supplementary File 5 – STRING-db networks and clusters.

**Description (short description of your dataset; indicate it is supporting an article):**

Microsoft Excel document with protein networks obtained from STRING database and clusters predicted by the MCL algorithm. This data set is supporting material for a journal article.

**Comments**:(*anything you need people to know right up front – special software they will need to use your data? Order of download or use? This information appears on the front of the Digital Commons page, with the title, the Journal title, authors, etc. so it’s a good place to show users important information*)

For each data set, there are two sheets as noted by “network” or “clusters” in the sheet name.

|  |  |
| --- | --- |
| **Column descriptions for “network” sheets** | |
| #node1 | node 1 (gene symbol 1) |
| node2 | node 2 (gene symbol 2, connected to node 1) |
| node1\_string\_internal\_id | STRING-db internal ID for node 1 |
| node2\_string\_internal\_id | STRING-db internal ID for node 2 |
| node1\_external\_id | Accession ID for node 1 |
| node2\_external\_id | Accession ID for node 2 |
| neighborhood\_on\_chromosome | Score indicating whether genes are consistently observed in each other’s genome neighborhood |
| gene\_fusion | Score indicating gene fusion events |
| phylogenetic\_cooccurrence | Score indicating degree of similarity of the phylogenetic distribution |
| homology | Score indicating homology for the two proteins |
| coexpression | Score indicating correlation of expression |
| experimentally\_determined\_interaction | Score derived from interaction databases |
| database\_annotated | Score indicating annotation of the database |
| automated\_textmining | Score indicating frequent co-mention of pairs of proteins in scientific literature |
| combined\_score | Score combined from all evidence channels |

|  |  |
| --- | --- |
| **Column descriptions for “clusters” sheets** | |
| #clustering method | method used to identify node clusters |
| cluster number | number assigned to each cluster |
| cluster color | default color assigned to each cluster (does not match final figures in publication, which were re-colored for clarity) |
| gene count | number of genes for each cluster |
| protein name | Official protein name |
| protein identifier | Ensembl protein accession ID |
| protein description | STRING-db description of protein |

**Discipline:** Toxicology

*We usually select based on your department and area*

*You can see a list of disciplines here:* [*http://network.bepress.com/*](http://network.bepress.com/)

*At the bottom of the page, you can click and drill down through the subsequent pages to see additional breakdown*.

**Keywords:**

Western diet; colitis; inflammation; colorectal cancer; calcium; vitamin D; transcript profiling; NanoString

**Journal:** Nutrients

*If applicable, name of journal in which your article is appearing that your dataset is supporting*

**Volume, Issue, Date (year is fine that is all you have):**

*For journal issue*

2020

**Embargo:** 02/18/2020

*Date that you want the data to be made public. If immediate, enter today’s date*