Report

Motivating Problem From Domain

Fanconi anemia (FA) is a rare disease that affects how DNA is translated and repaired. This leads to abnormalities in bone marrow and skeletal structure as well as an increased risk for cancer. Due to causing low red and white blood cells, patients are at an increased risk for anemia, infections, and excessive bleeding. There are 22 known genes that will lead to FA if mutated. Through visualizing the gene network, new insights on the potential mechanism of the disease may be revealed.[1]

Computational Problem Formulation

Using a list of FA gene connections, visualize networks using all genes via nodes and edges.

Specific Approach

Using a list of the known genes associated with FA, a list of known gene interactions can be made with just the FA genes, shortened from a list of all known disease gene interactions . This specified list can then be used to visualize the types and strengths of the correlations between the FA genes.

Specific Implementation of Approach

Starting with a set of FA genes from OMIM[2] and a string input of all gene linkages, all occurrences of direct FA gene connections can be identified. If an FA gene is not directly connected to another FA gene, another connection to other FA genes through a non-FA gene can also be determined. In visualization through cytoscape[3], the size of the node can correlate to its degree of connections and the thickness of the edges can correlate to the number of shortests paths, implemented from Dijkstra's algorithm, between the nodes.

Pseudocode

Input = associated_gene_file
Geneconnections = STRING1.txt

open(input):
        Take each gene from each loci and append it to a master FA gene list

open(geneconnections)
        Make list of all gene connections

For FA_gene in master_FA_list:
        Make list of all direct connections (check in all gene connection list if both genes are in
        master_FA_list)
        If there are direct connections:
                Append connection to final_list if not already
        If there are no direct connections:
                Check for indirect connections
                Make list of all FA_gene connections with non FA_gene
                See if non FA_gene has connection to different FA_gene
                If indirect connection found:
                        Append indirect connections to final_list if not already

Write final_list to new input file

Results and Discussion

The code took about 10-15 minutes to run. Print checks throughout code showed it worked as expected. A suitable output file was produced. Assumptions made while creating this code is that FA genes will either be directly connected to each other or a gene may have no direct connections to other FA genes and in that case may be indirectly connected to FA genes through non-FA genes. A limitation is that if there are two domains of connections that are not connected to each other directly, an indirect non FA gene connection would not be identified. Another limitation is the parsing of the input file is specific for that input file and may need to be edited if applying to a different set of genes. It was also taken into consideration if there would be duplicate entries of gene, which were not included in the final list.

Overall, the code was successful, producing an output file allowing for visualization in cytoscape, which can be seen below in Figure 1.
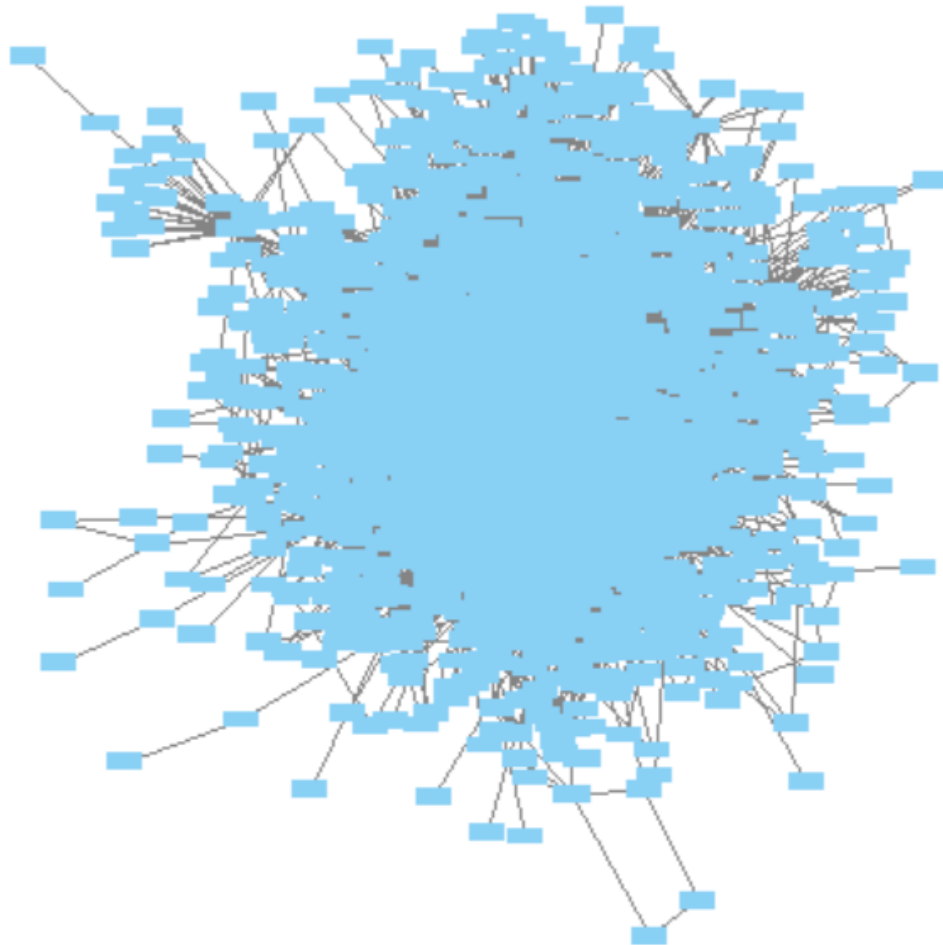

Figure 1. Visualizations of FA gene connections from cytospace.

There were 15,317 total interactions found. This resulted in a dense visualization. To make it more readable, in the future separate maps can be created for each FA gene. It should also be noted there are no apparent separated domains of connections.

Sources

1. National Center for Advancing Translational Sciences |.

    rarediseasesinfonihgov.

    https://rarediseases.info.nih.gov/diseases/6425/fanconi-anemia.

2. OMIM - Online Mendelian Inheritance in Man. omimorg. https://omim.org/.

3. Shannon P. 2003. Cytoscape: A Software Environment for Integrated Models of

    Biomolecular Interaction Networks. Genome Research. 13(11):2498–2504.

    doi:https://doi.org/10.1101/gr.1239303.

    https://www.ncbi.nlm.nih.gov/pmc/articles/PMC403769/.