Searching for narcolepsy disease modules using a protein-protein interaction network

Dylan Steinecke

Disease background

Narcolepsy is a rare disease affecting the brain, often causing patients to involuntarily fall asleep after experiencing strong emotions, to suddenly lose muscle strength, and sometimes to hallucinate[1,2]. Narcolepsy's causes can be a combination of genetics, certain infections, and traumatic brain injury. It is believed to be an autoimmune disease in which the immune system attacks neurons that produce a signaling molecule, orexin. Orexin is involved in multiple molecular networks in the brain, regulating a range of neurological functions including wakefulness, appetite, olfaction, cognition, and locomotion[3]. Thus, dysfunctional orexin signaling results in narcolepsy's neurological symptoms. However, evidence suggests that orexin deficiency does not always correlate with narcolepsy; either can exist without the other[4]. Although some is known about narcolepsy's pathology, it is incompletely understood. Similarly, some treatments exist, but they are far from perfect. If the disease mechanism can be better understood, treatment development can be better informed. More specifically, if the molecules in the various immune system pathways and neurological pathways involved in narcolepsy[4] can be better understood, narcolepsy can be better understood, paving the way for better treatments and outcomes for those suffering from narcolepsy and perhaps other related autoimmune neurological conditions that often correlate with narcolepsy.

Data sources

STRING's human protein-protein interaction network was chosen as the interaction network. Because narcolepsy has relatively few known proteins associated with it and is a complex disease, STRING's large comprehensive network could be useful for adding more background interactome information. Narcolepsy's associated proteins were obtained from DisGeNET's[5] gene-disease associations for narcolepsy (proteins were mapped to genes via HGNC). DisGeNET was chosen because it is one of the largest publicly accessible databases of gene-disease associations, integrating data from numerous manually curated sources and other sources of computationally predicted associations[5]. To improve data quality, only associations found from curated sources were used; text mined sources were excluded because they depended on an imperfect named entity recognition system (precision unknown) and then depended on a relation extraction module with arguably insufficient precision (~75% at best)[6].

Disease module analysis

To identify a disease module, first, a networkx graph of STRING protein-protein interactions and DisGeNET protein-narcolepsy associations was constructed. Next, a subgraph of the narcolepsy-associated proteins was identified via the networkx.subgraph() command. This subgraph's largest connected component was comprised of 4 proteins, the second largest were comprised of 2 proteins (eleven components). (Notably, when not filtering the narcolepsy proteins by high quality data sources, the largest component had 8 proteins.) The interactome's largest connected component had 16,090 nodes.

To test how robust the connected component size was to varying the input list, a random subset of 90% of the narcolepsy-associated proteins were supplied to the networkx.subgraph() function 100 times. The largest component's size tended to still be 4 (64% of the time), but was 3 (18%) and 2 (18%) at times (Fig 1). To compare how the narcolepsy proteins' connectivity compared to a random set of proteins, the connected component process described above was repeated on a random subset of interactome proteins/genes, the same amount as 90% of the narcolepsy protein list. Surprisingly, these largest connected components identified tended to be larger, averaging a size of 8.1 (Fig 2). It would be interesting to compare not with random subsets of the interactome but with random subsets of the neurological interactome.

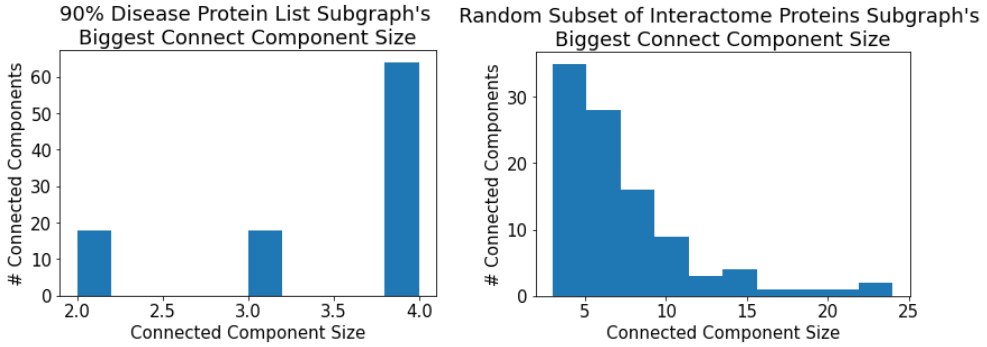

Fig 1. The connected component sizes of the narcolepsy proteins' subgraph when a random subset of 90% of the proteins were chosen as the input list. **Fig 2.** Similar to Fig 1, except that the protein list are random interactome proteins, the same amount of proteins as the Fig 1 protein subset.

This suggests that currently known narcolepsy-associated proteins and protein-protein interactions do not reveal a disease module. This may be due to multiple non-mutually-exclusive reasons: narcolepsy-associated proteins are insufficiently identified, narcolepsy-associated proteins' interacting partners are insufficiently identified, other non-protein biomolecules such as hormones and neurotransmitters are part of the disease module, or that no large disease module exists. To determine if the last reason is true, the first three reasons must be addressed.

The proteins in the largest narcolepsy-associated module are transducin beta-like 1X-related protein 1 (UniProt: Q9BZK7), centromere protein C (UniProt: Q03188), histone deacetylase 9 (UniProt: Q9UKV0), and histone H2A type 1-C (UniProt: Q93077). These proteins affect DNA transcription. Transducin beta-like 1X-related protein 1 is encoded by TBLR1 and is reported to be a cofactor required to modulate transcription by nuclear receptors[7]. Centromere protein C is encoded by CENPC. It plays a key role in forming kinetochores[8], protein complexes to which microtubules attach and which are associated with chromosomal centromeres during mitosis. Histone H2A type 1-C is encoded by H2AC6. It is a part of histones, proteins around which DNA wraps and which plays a key role in regulating transcription, repairing DNA, replicating DNA, and stabilizing chromosomes[9]. Histone deacetylase 9 is encoded by HDAC9. It deacetylates part of histone cores[10], repressing transcription and developmental processes such as growing skeletal muscle. Without it, neuronal apoptosis would happen because MAPK10 would phosphorylate JUN and separately HDAC1 would be recruited to the JUN promoter and JUN would be transcribed[11]. A cursory search of these molecules' roles in narcolepsy did not reveal much insight. To find out reported associations between narcolepsy

and these molecules, one could further investigate the sources of the DisGeNET gene-disease associations.

(Interestingly, the orexin precursor protein was not among DisGeNET's high quality gene-disease associations. When added, it combined with 2 other proteins to make a connected component. It appears that there is likely missing information here connecting orexin to other proteins involved in narcolepsy.)

References

1. Wikipedia contributors. Narcolepsy [Internet]. Wikipedia, The Free Encyclopedia; 2022 Oct 12, 15:40 UTC [cited 2022 Oct 24]. Available from: https://en.wikipedia.org/w/index.php?title=Narcolepsy&oldid=1115664480.

2. Bassetti CLA, Adamantidis A, Burdakov D, Han F, Gay S, Kallweit U, Khatami R, Koning F, Kornum BR, Lammers GJ, Liblau RS, Luppi PH, Mayer G, Pollmächer T, Sakurai T, Sallusto F, Scammell TE, Tafti M, Dauvilliers Y. Narcolepsy - clinical spectrum, aetiopathophysiology, diagnosis and treatment. Nat Rev Neurol. 2019 Sep;15(9):519-539. doi: 10.1038/s41582-019-0226-9. Epub 2019 Jul 19. PMID: 31324898.

3. Li J, Hu Z, de Lecea L. The hypocretins/orexins: integrators of multiple physiological functions. Br J Pharmacol. 2014 Jan;171(2):332-50. doi: 10.1111/bph.12415. PMID: 24102345; PMCID: PMC3904255.

4. https://www.nature.com/articles/s41582-019-0226-9&sa=D&source=docs&ust=1666639491384801&usg=AOvVaw1uJO1XStpG5LfA35PobdHL

5. Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, García-García J, Sanz F, Furlong LI. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res. 2017 Jan 4;45(D1):D833-D839. doi: 10.1093/nar/gkw943. Epub 2016 Oct 19. PMID: 27924018; PMCID: PMC5210640.

6. Bravo Serrano, Àlex. BeFree : a text mining system for the extraction of biomedical information from literature. 2016

7. Perissi V, Aggarwal A, Glass CK, Rose DW, Rosenfeld MG. A corepressor/coactivator exchange complex required for transcriptional activation by nuclear receptors and other regulated transcription factors. Cell. 2004 Feb 20;116(4):511-26. doi: 10.1016/s0092-8674(04)00133-3. PMID: 14980219.

8. Gascoigne KE, Takeuchi K, Suzuki A, et al. Induced ectopic kinetochore assembly bypasses the requirement for CENP-A nucleosomes. Cell. 2011 Apr;145(3):410-422. DOI: 10.1016/j.cell.2011.03.031. PMID: 21529714; PMCID: PMC3085131.

9. https://www.uniprot.org/uniprotkb/Q93077/entry

10. Zhou X, Marks PA, Rifkind RA, Richon VM. Cloning and characterization of a histone deacetylase, HDAC9. Proceedings of the National Academy of Sciences of the United States of America. 2001 Sep;98(19):10572-10577. DOI: 10.1073/pnas.191375098. PMID: 11535832; PMCID: PMC58507.

11. www.uniprot.org/uniprotkb/Q9UKV0/entry