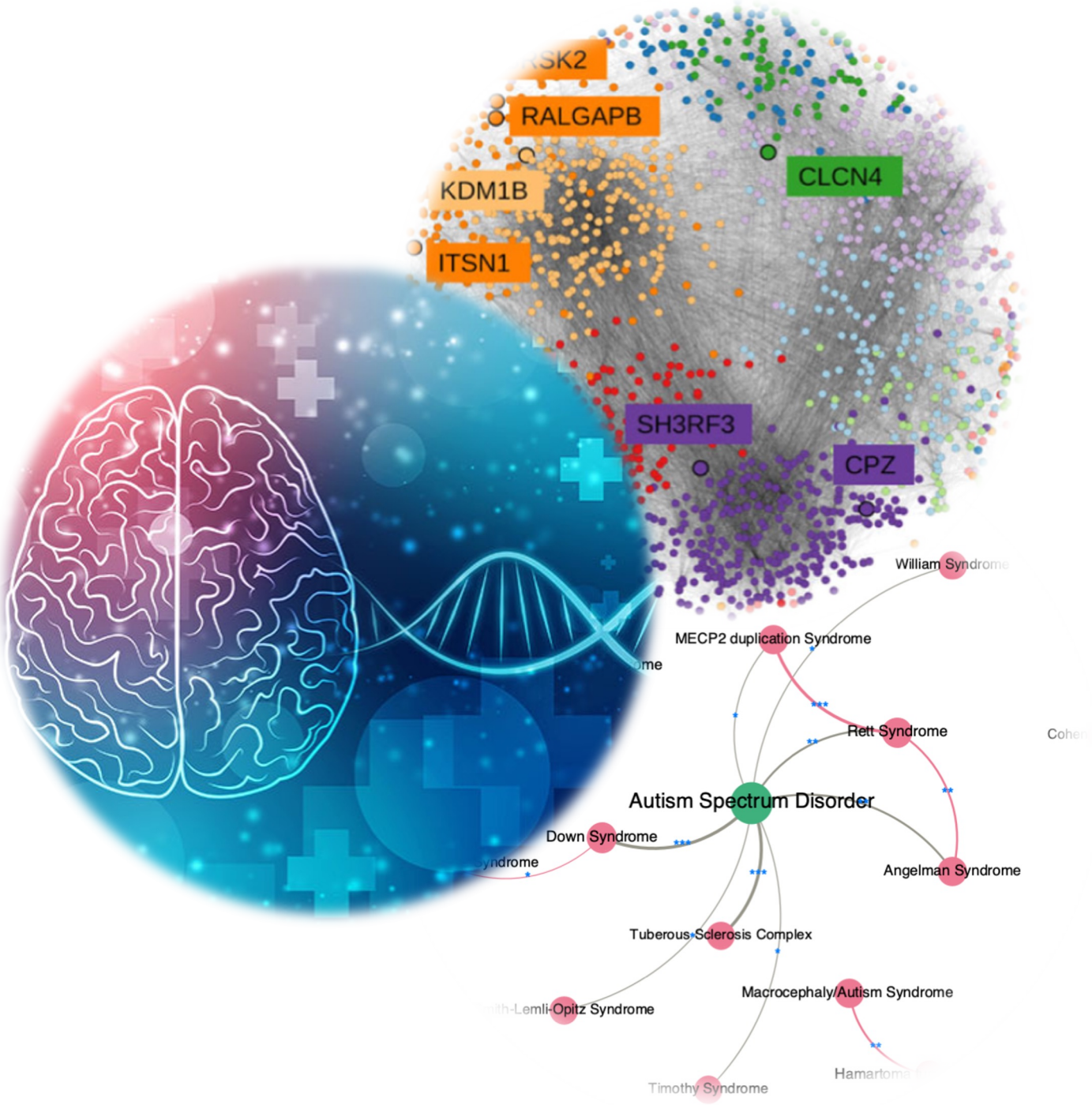


Syndromic vs Non-Syndromic Autism

An Analysis of Disease-to-Disease Network and Disease-to-Gene Network

Presenter: Hao CHEN

Syndromic Autism vs Non-syndromic Autism

- Autism - High Heritability & Genetic Heterogeneity (Tick et al, 2016)
  - 60-90% concordance in monozygotic twins
- Syndromic Autism vs Non-syndromic Autism (Sztainberg & Zoghbi, 2016)

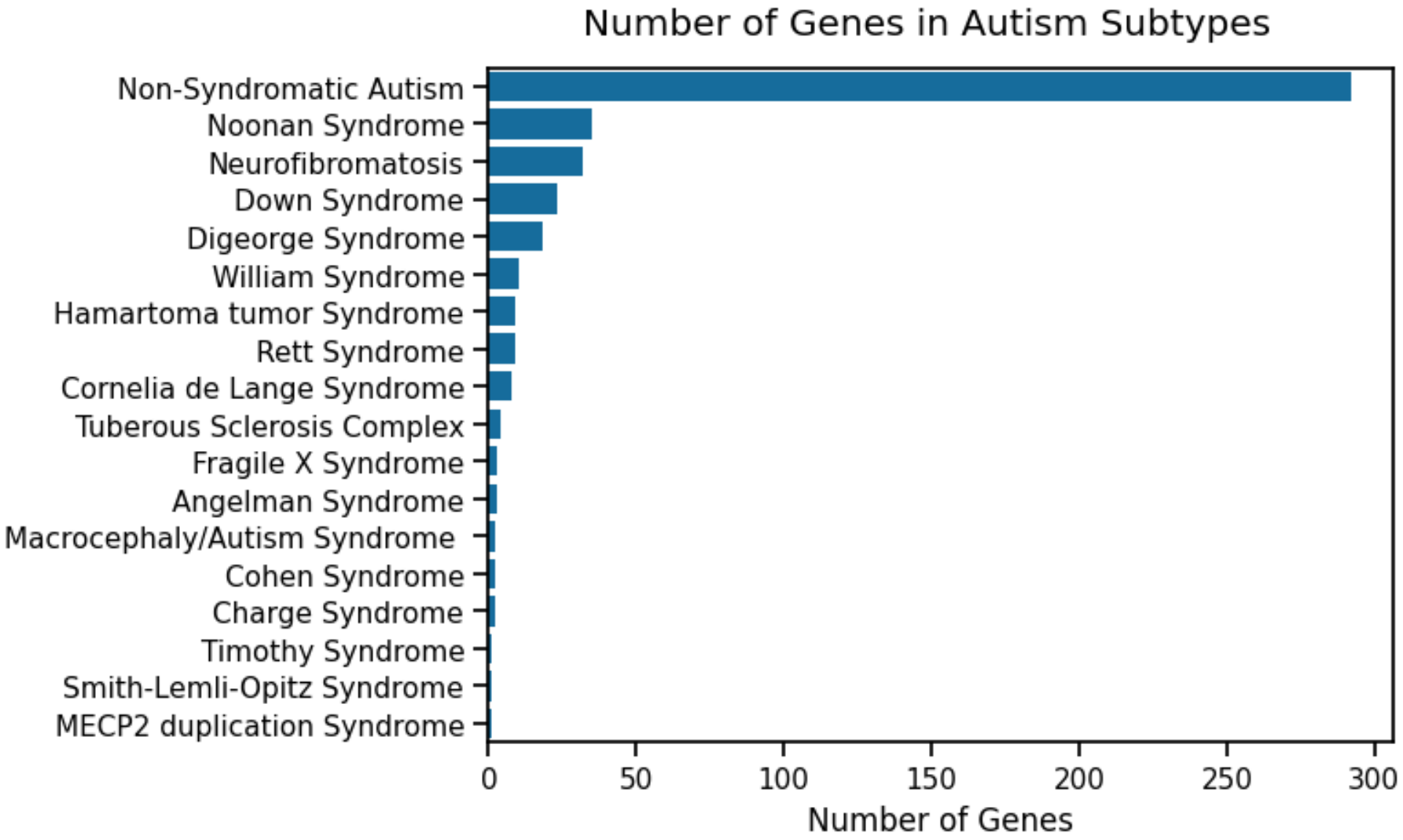
	Syndromic Autism	Non-syndromic Autism
Proportion	5-15% of autism	85-95% of autism
Genetic features	Known genetic causes; Mono-genetic disorders and chromosomal re-arrangement	Unknown genetic causes Genetic heterogeneity GWA (Genome-wide Association Studies) only explain a small portion of heritability
Clinical Presentation	Autism is a comorbidity of underlying complex syndrome	Autism without complex cluster of symptoms
Examples	Fragile X syndrome, Rett syndrome, MECP2 duplication syndrome, tuberous sclerosis complex	Autism Spectrum Disorder (Asperger's syndrome, pervasive developmental disorder, classic autism)

Research Question & Data Source

- How to differentiate syndromic and non-syndromic autism in gene-to-disease network and disease-to-disease networks?

- Data sources: DisGeNet Database (Piñero et al., 2020)
  - Expert Curated Data
    - Gene-disease associations (GDAs) collected from following databases: UniProt, PsyGeNET, Orphanet, the CGI, CTD (human data), ClinGen, and the Genomics England PanelApp
  - 10370 diseases
  - 9413 gene

Data Processing: Regrouped autism subtypes

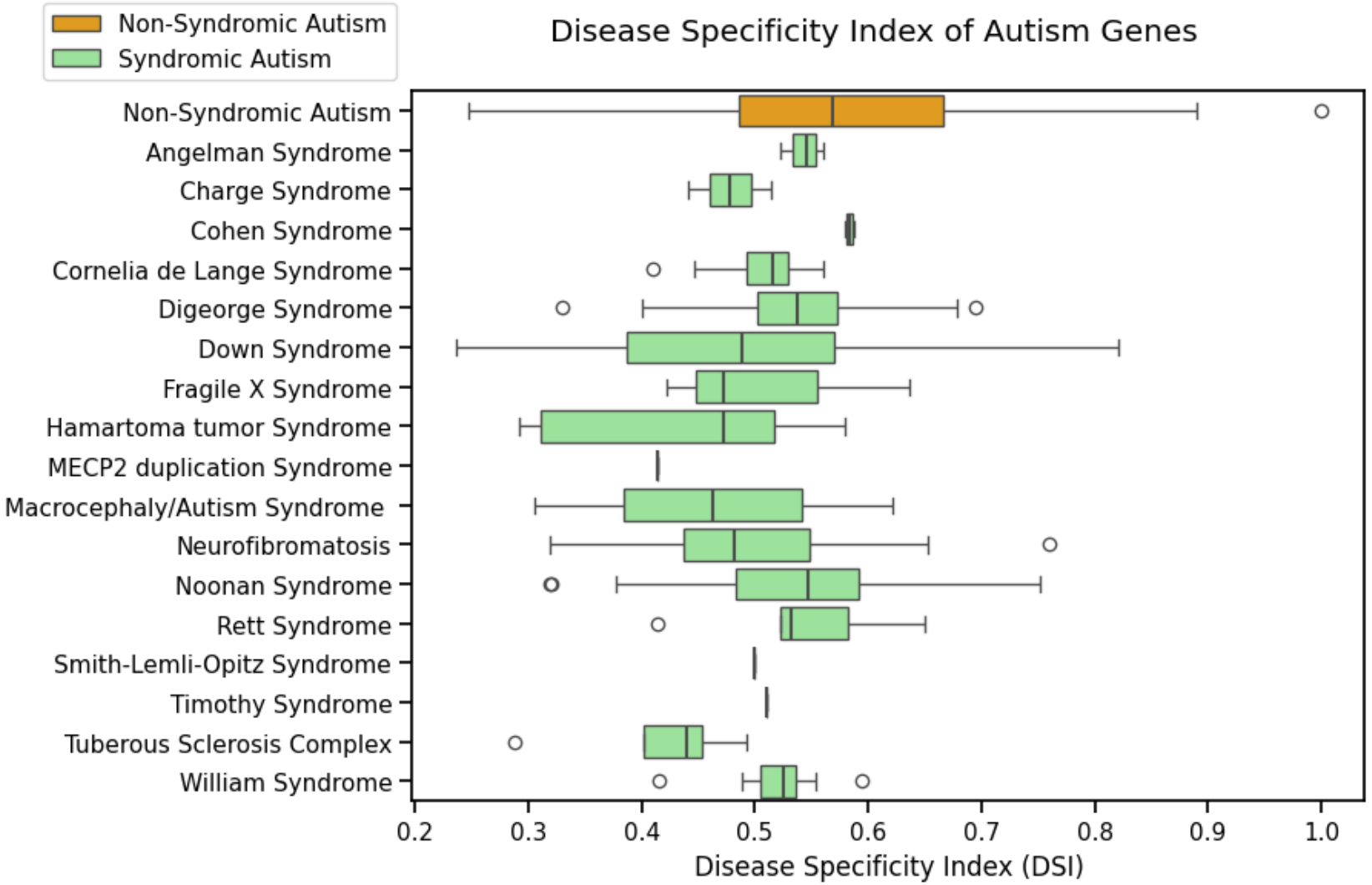


Each subtype were identified by keyword searching.

Ambiguous diagnoses were grouped, e.g., autistic disorder & autistic spectrum disorder.

Different types of a syndrome were grouped, e.g., Smith-Lemli-Opitz Syndrome, Type I and Type II.

Autism Gene Metrics: Disease Specificity Index (DSI)



DSI ranges from 0 to 1.

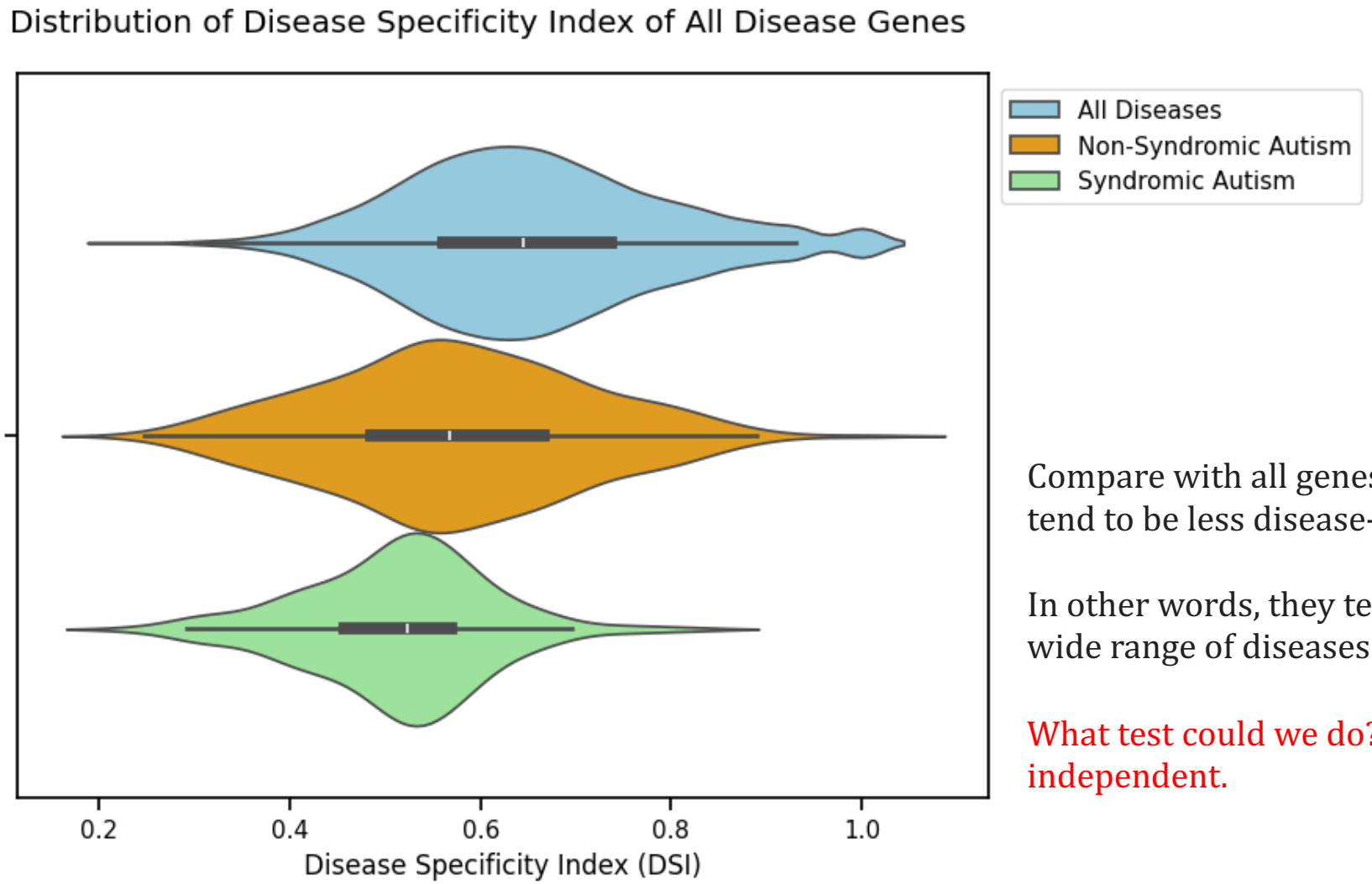
Inversely proportional to the number of diseases associated to a particular gene.

0: gene associated with many diseases.

1: gene associated with only 1 disease.

(Piñero et al., 2020)

Disease Specificity Index (DSI) of All Genes

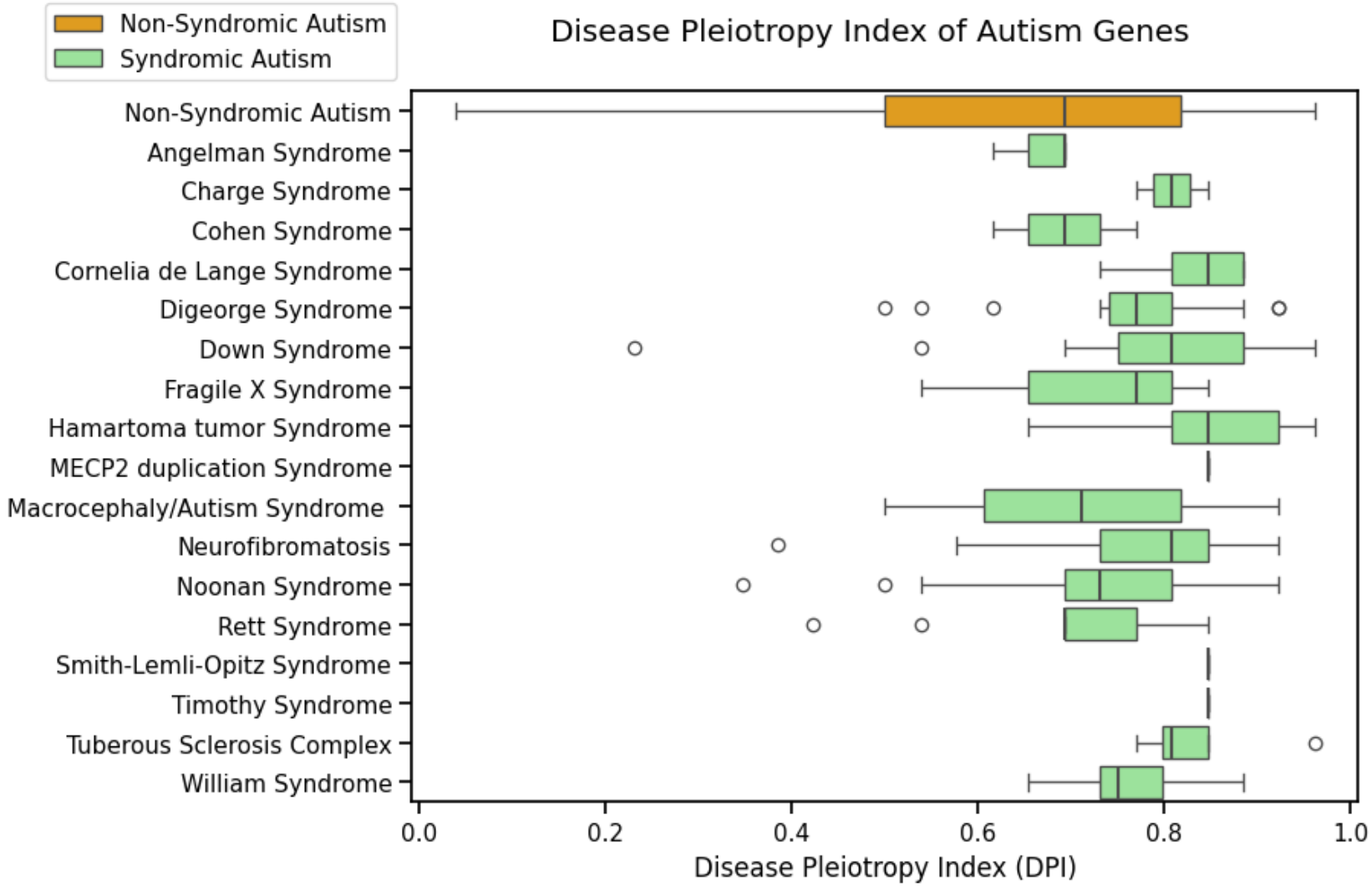


Compare with all genes, genes involved in autism tend to be less disease-specific.

In other words, they tend to be involved in a wide range of diseases.

What test could we do? The samples are not independent.

Autism Gene Metrics: Disease Pleiotropy Index (DPI)



DPI ranges from 0 to 1.

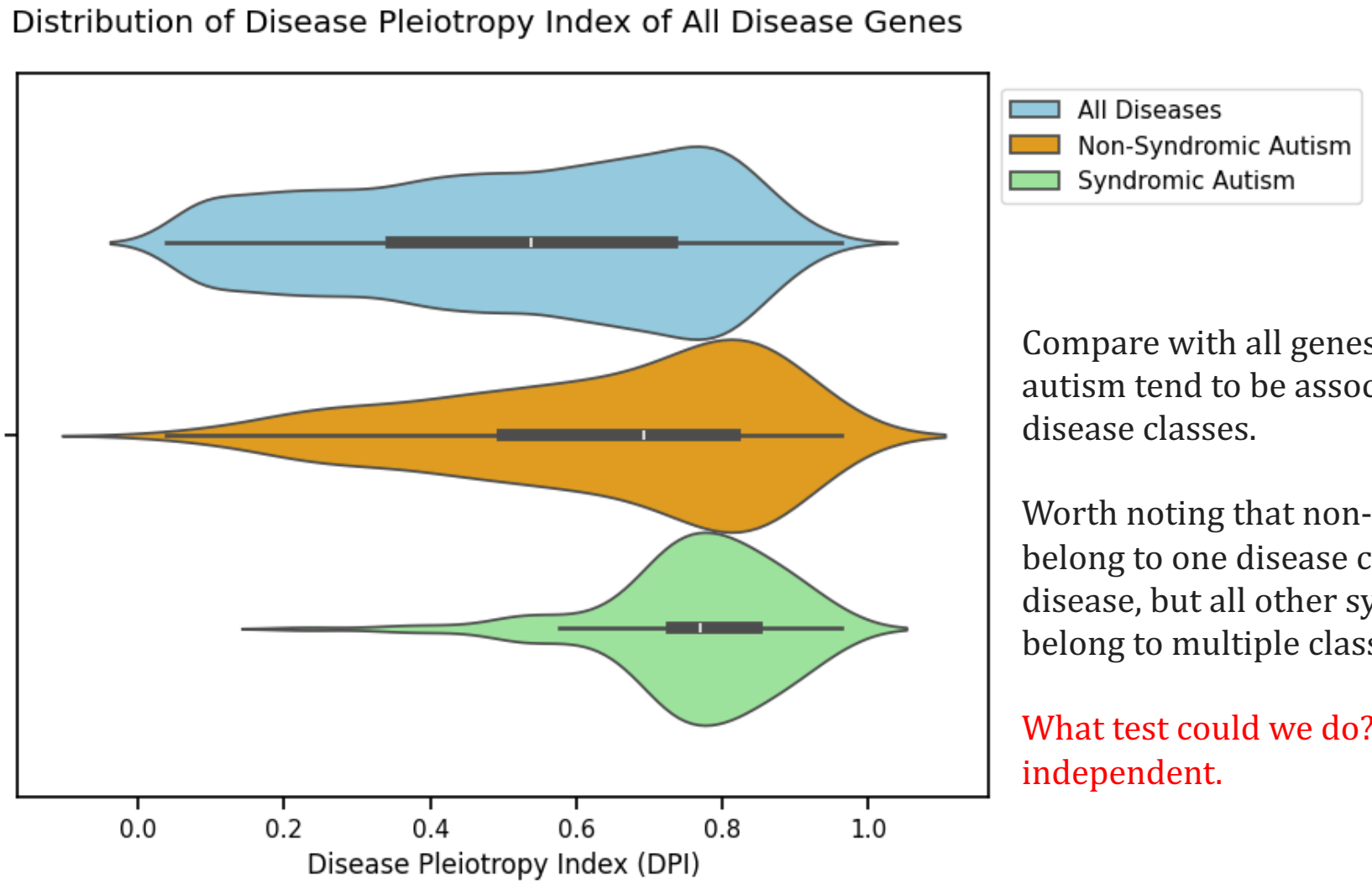
Proportional to the number of different (MeSH) disease classes a gene is associated to.

0: gene associated with 0 class.

1: gene associated with many classes.

(Piñero et al., 2020)

Disease Pleiotropy Index (DPI) of All Genes

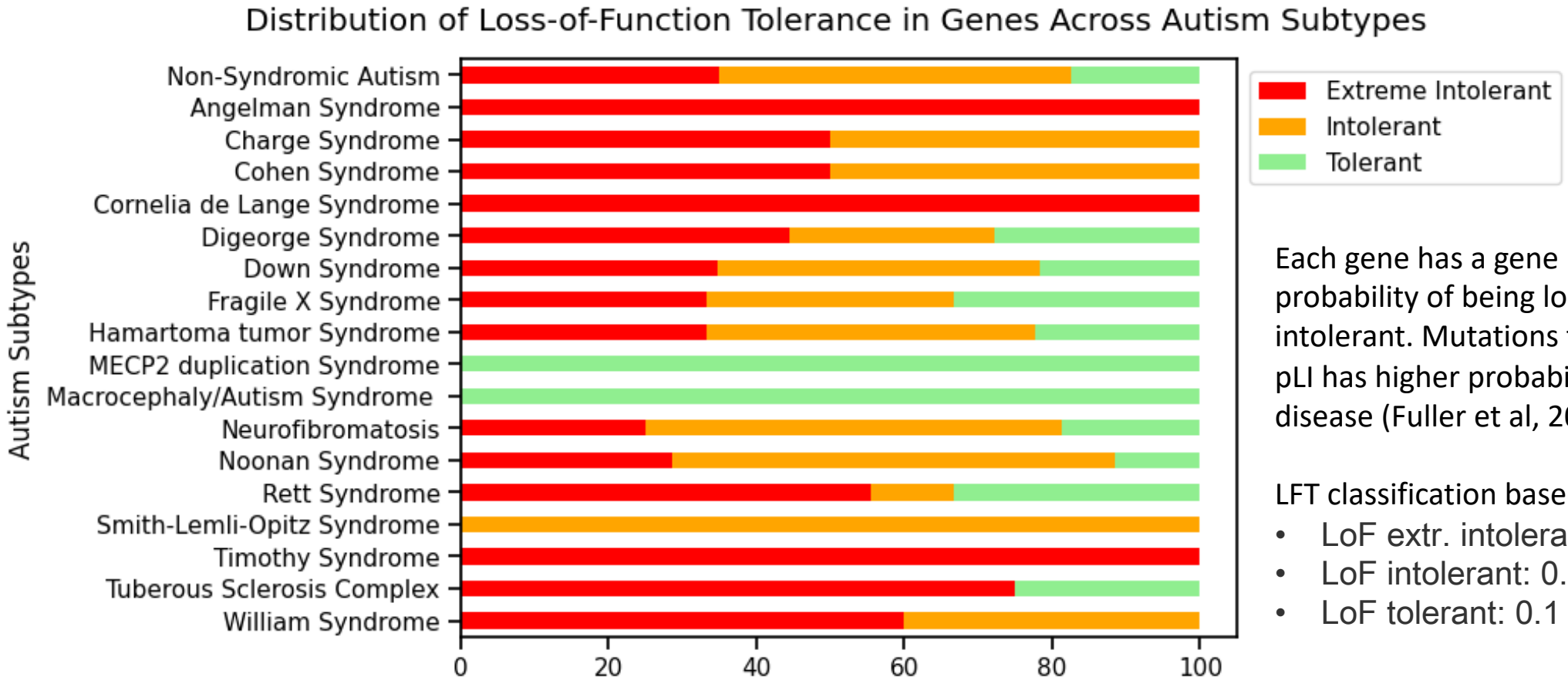


Compare with all genes, genes involved in autism tend to be associated with more disease classes.

Worth noting that non-syndromic autism only belong to one disease class – psychiatric disease, but all other syndromic autism does belong to multiple classes.

What test could we do? The samples are not independent.

Autism Gene Metrics: Loss-of-Function Tolerance(LFT)

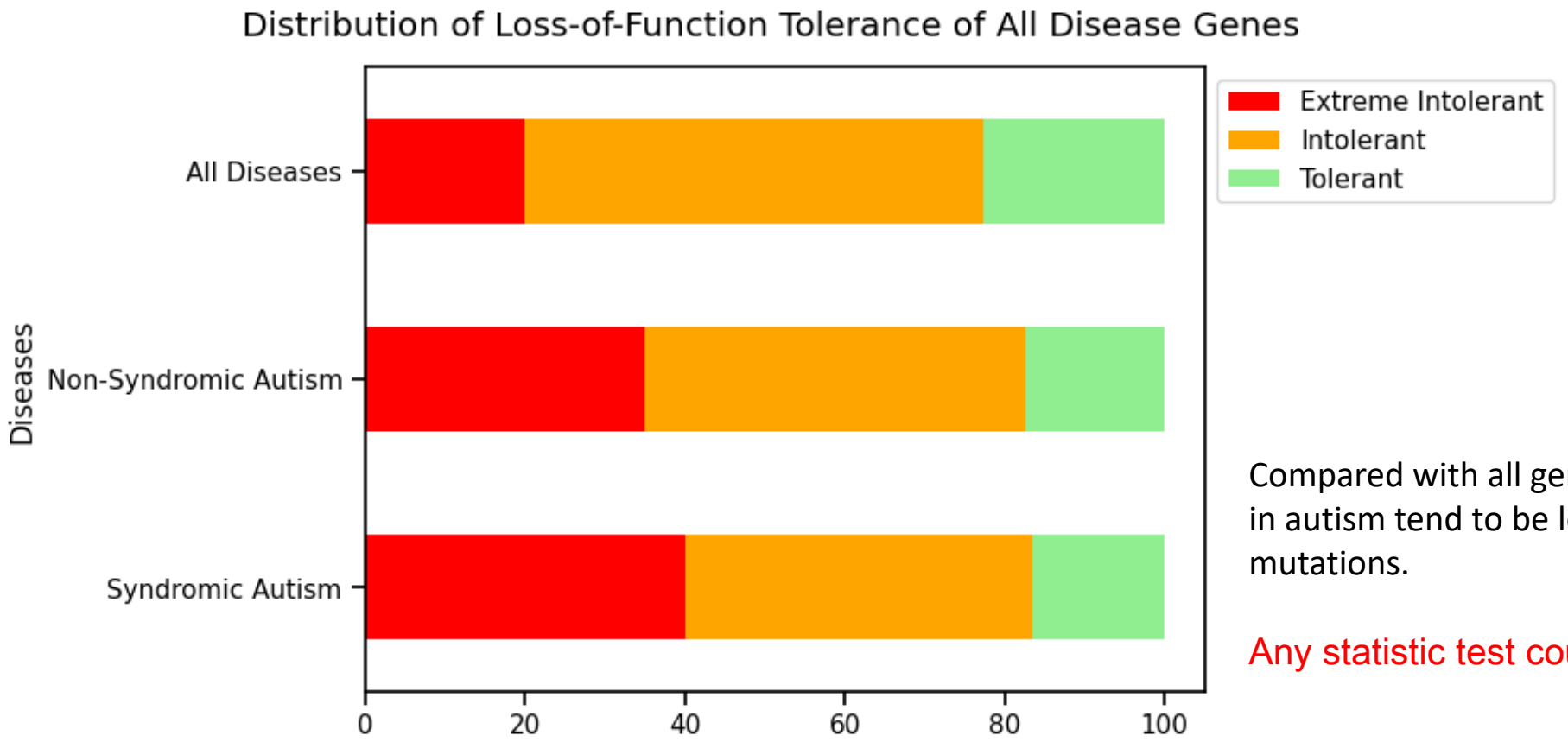


Each gene has a gene metric – pLI, i.e., probability of being loss-of-function intolerant. Mutations to genes with high pLI has higher probability of causing disease (Fuller et al, 2019).

LFT classification based on pLI:

- LoF extr. intolerant: pLI >=0.9.
- LoF intolerant: 0.9 >= pLI >=0.1
- LoF tolerant: 0.1 >= pLI

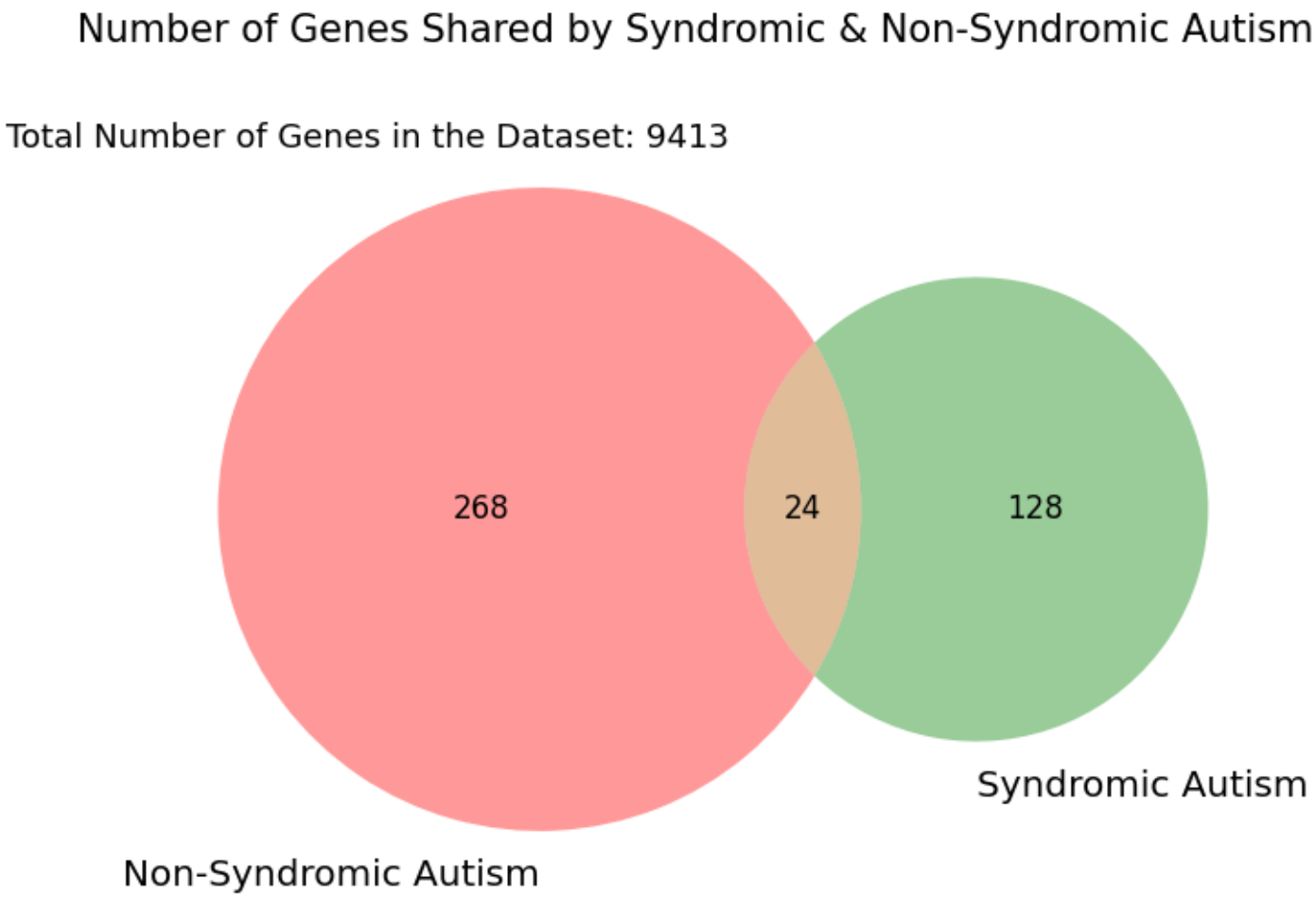
Loss-of-Function Tolerance(LFT) of All Genes

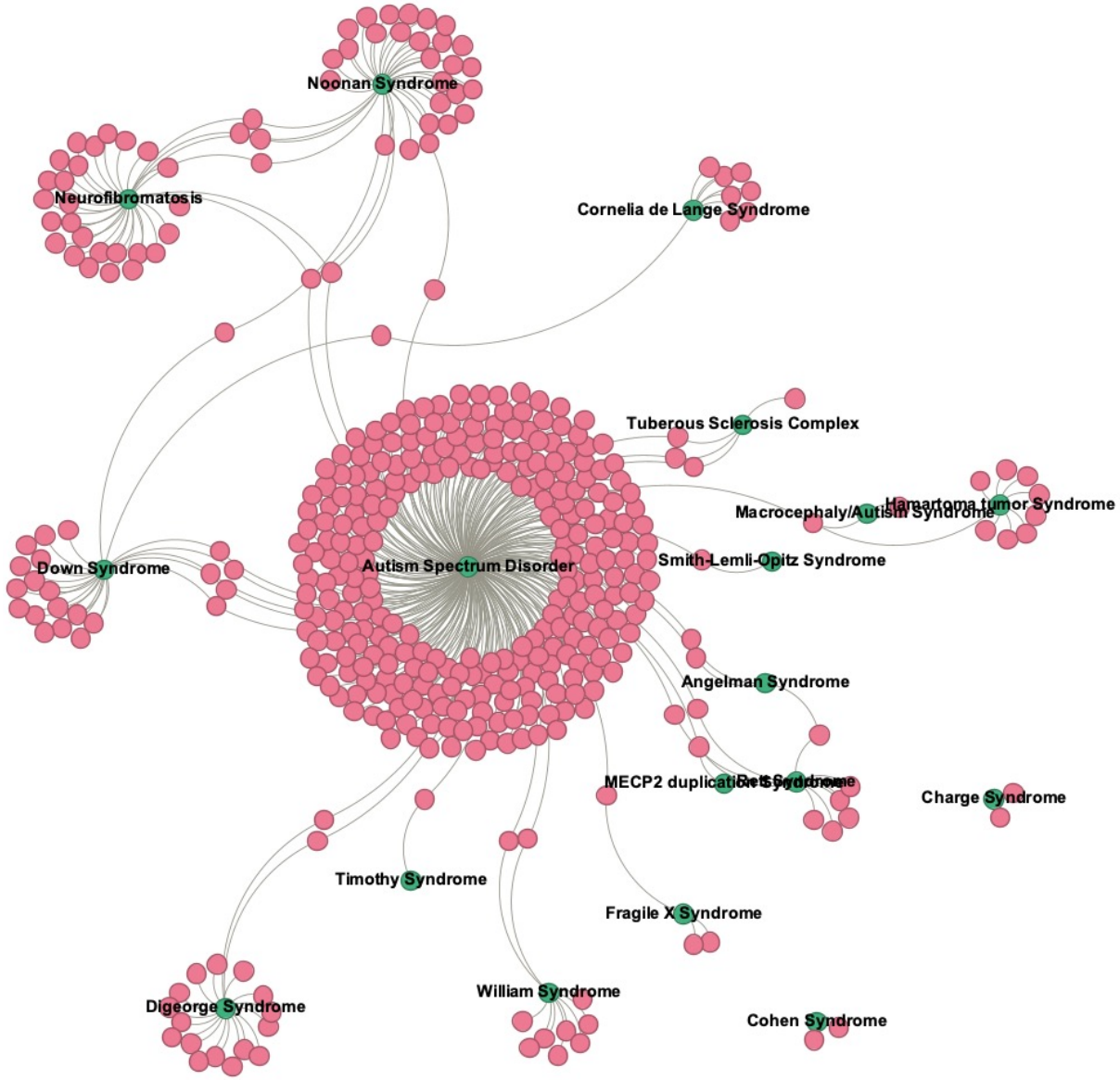


Compared with all genes, genes involved in autism tend to be less tolerant to mutations.

Any statistic test could we do?

Shared Genes





Autism-Gene Bipartite Network

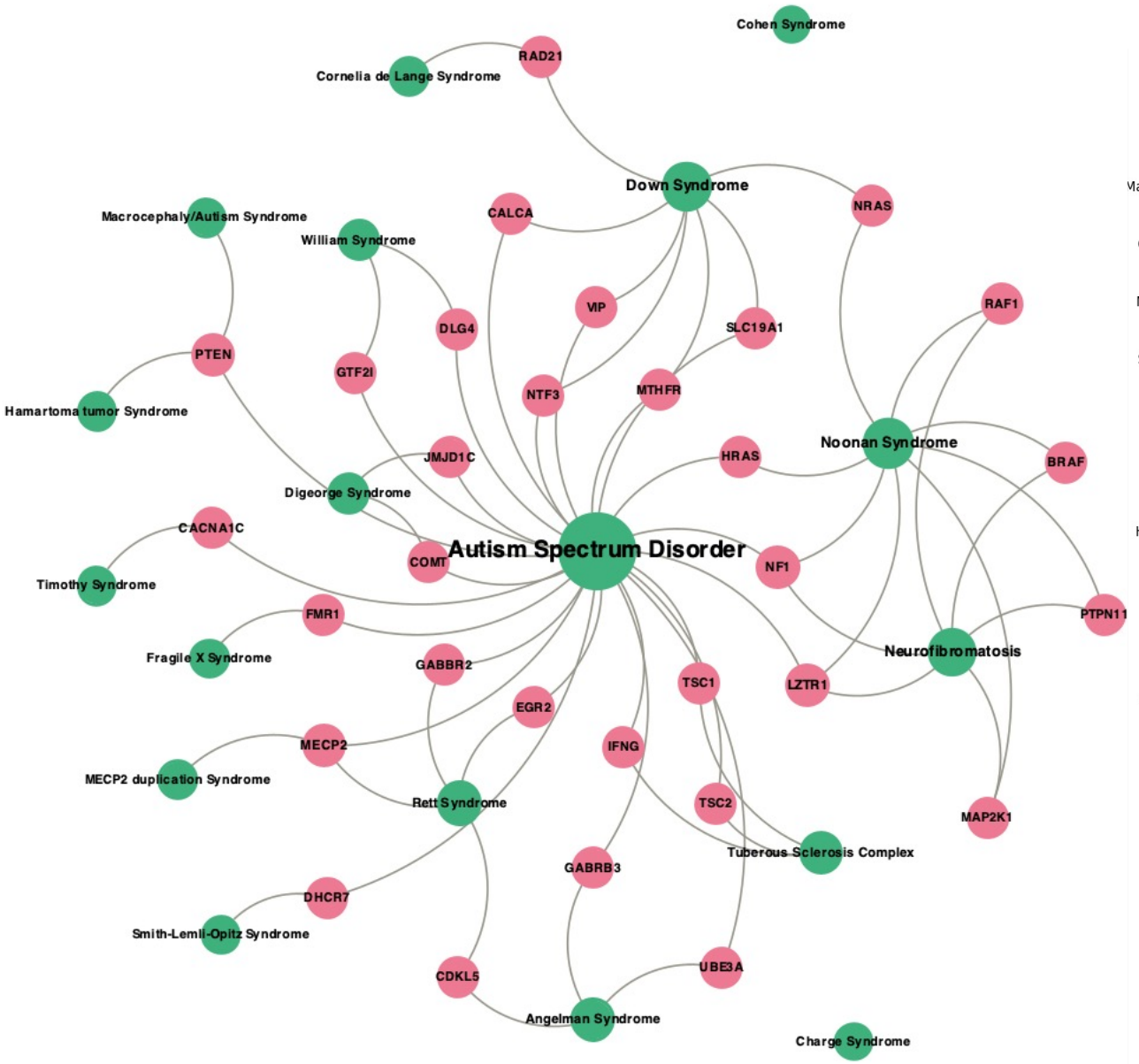
Most genes are associated with non-syndromic autism.

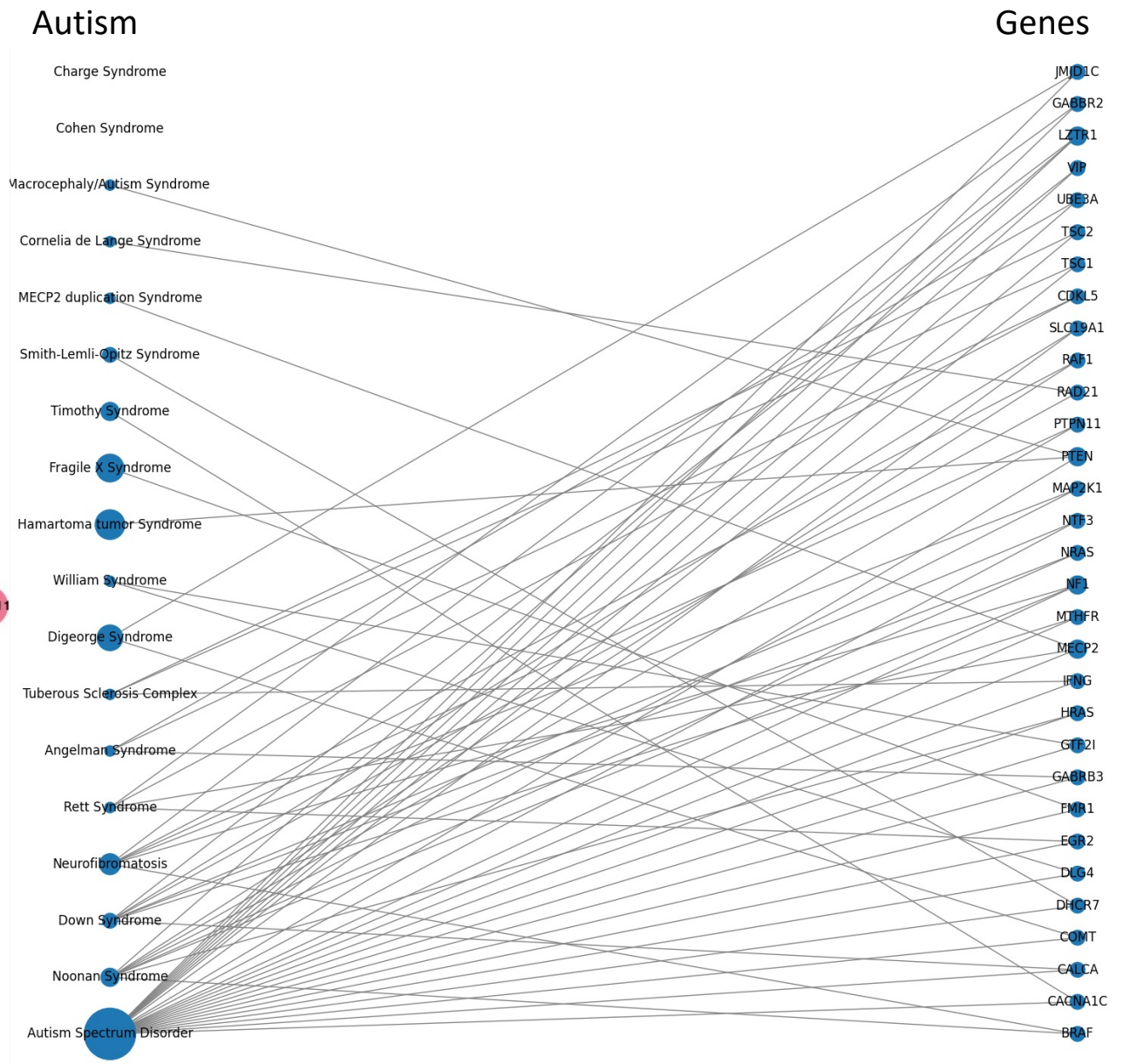
Genes connect

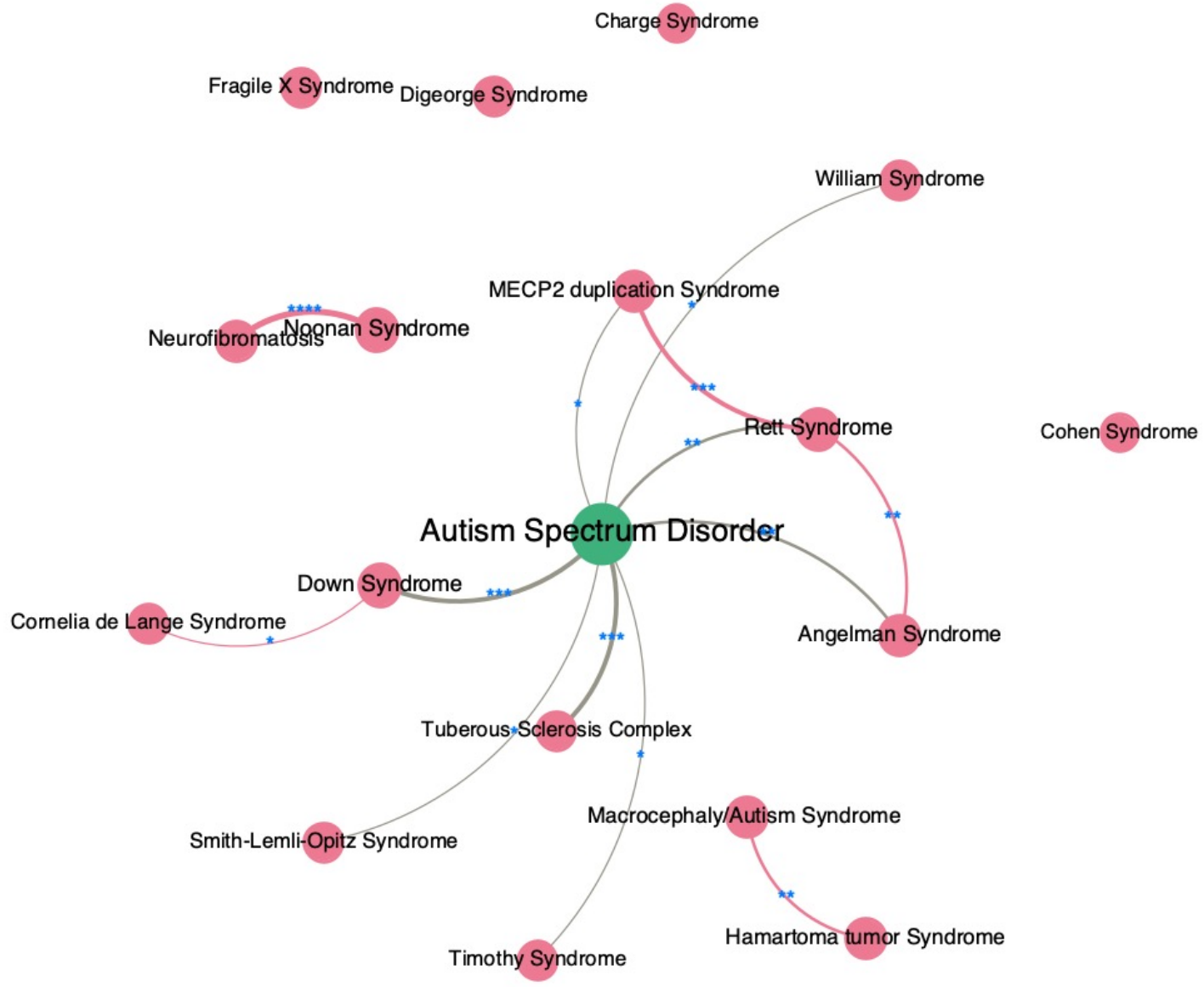
- non-syndromic autism to syndromic autisms
- subtypes of syndromic autism to each other.

Cohen syndrome & Charge syndromes do not share genes with other disorders.

Trimmed Bipartite Network: Identify Potentially Important Genes





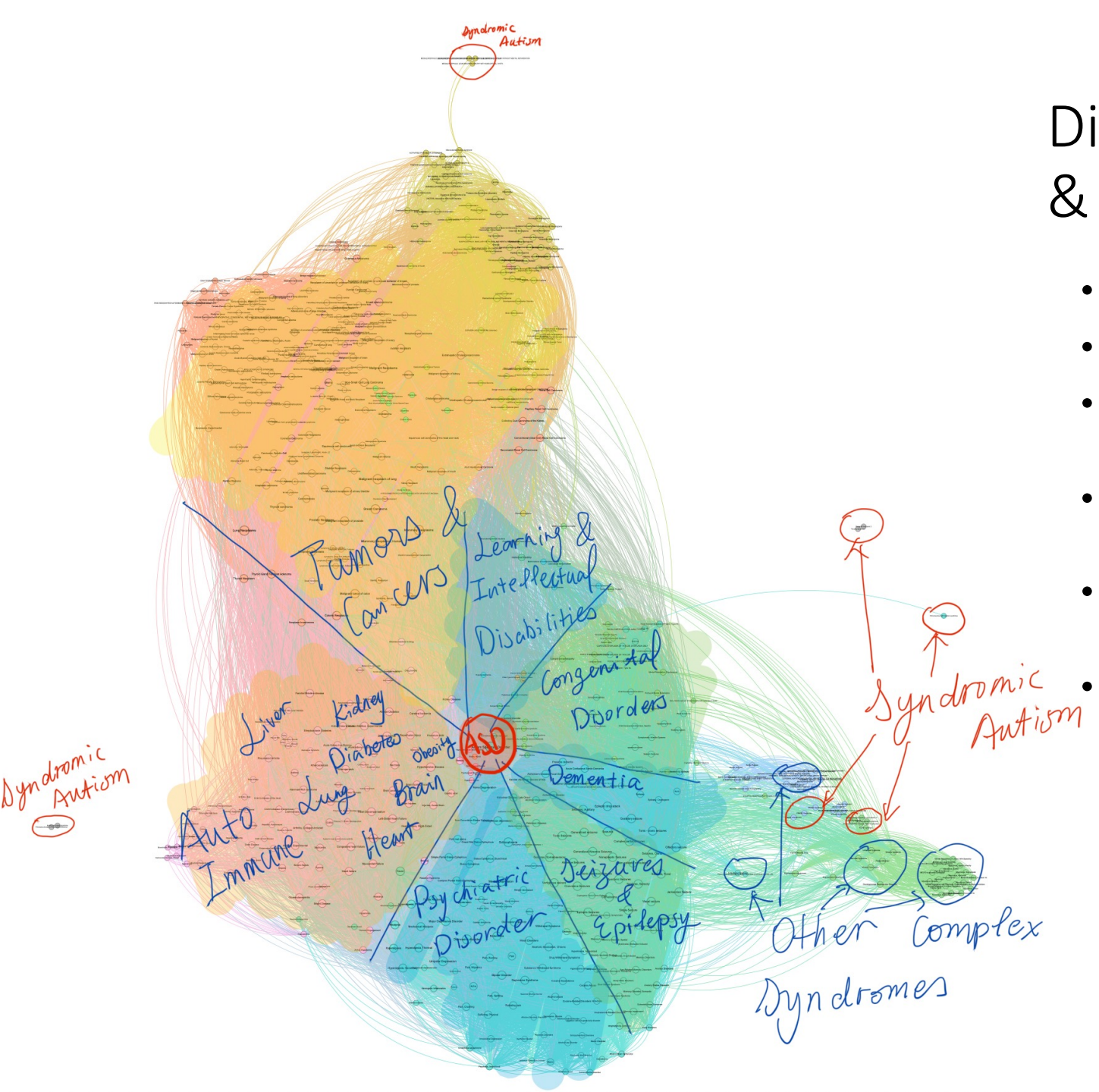


Autism Disease Unipartite Network

- Diseases are nodes. An edge exists between two diseases if they share genes.
- The p-value of edge is calculated using hyper-geometric test.
  - I.e., given that the total number of gene (N=9413), disease A with n_1 of genes, and disease B with n_2 of genes, what is the probability that these two diseases share k genes?
- Edges are kept only if their p-value smaller than 0.05.
- 4 subtypes not connecting to any other subtype.

Note:

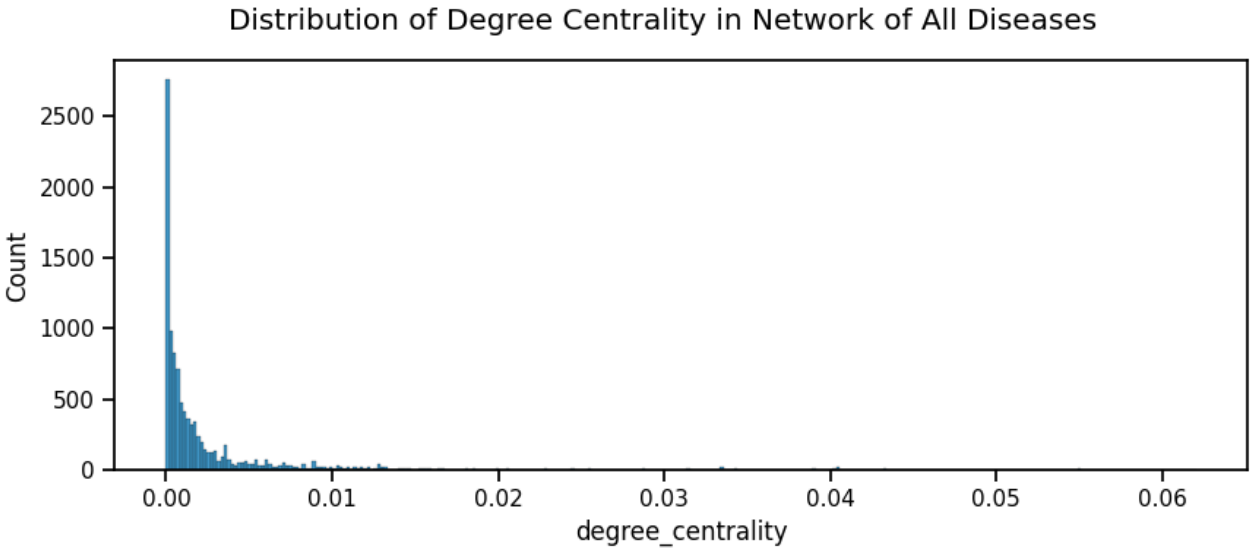
- *: p<.05; **: p<.01; ***: p<.001; ****: p<.0001;

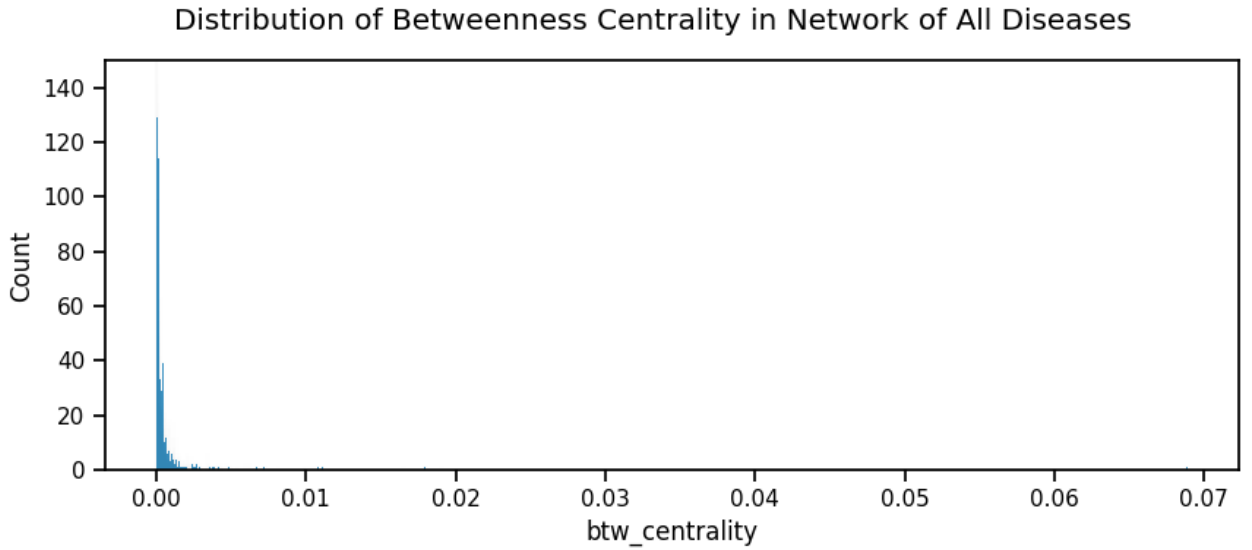


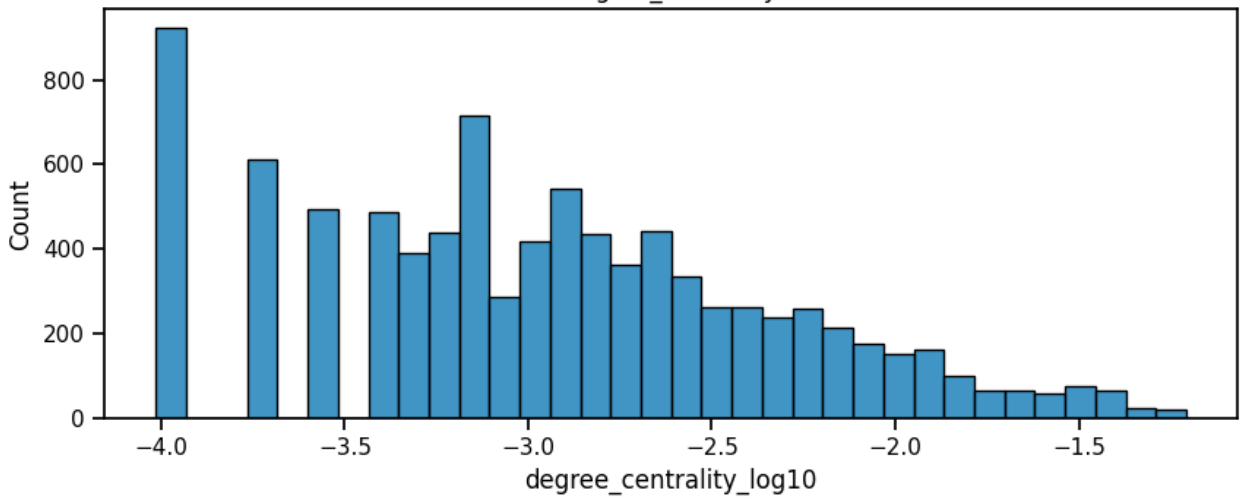
Disease Network for Autism & Their Neighbors

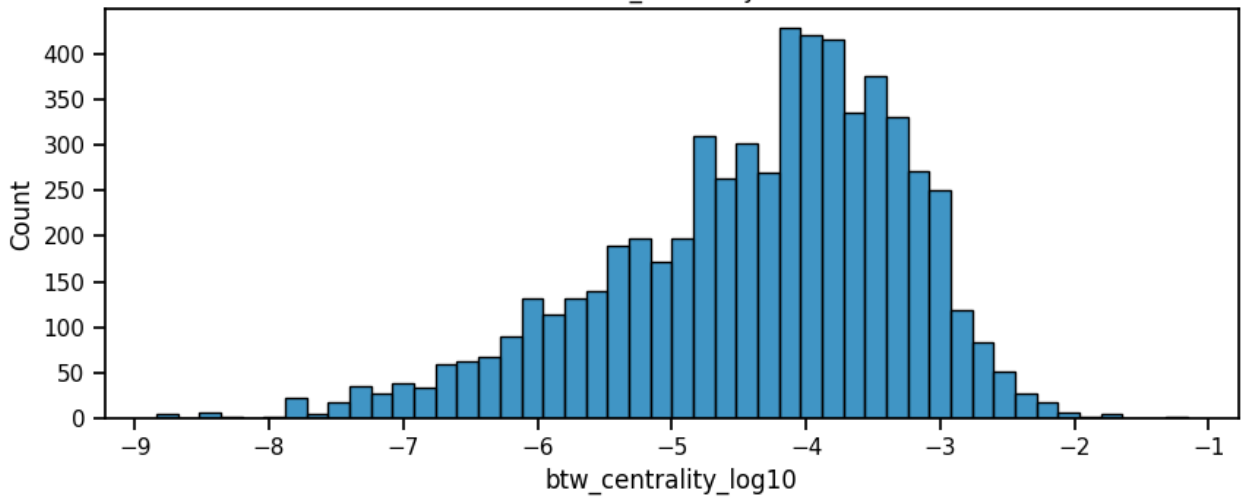
- 728 nodes/diseases, 29495 edges
- Network density: 0.1149
- Edges are kept only if their p-value smaller than 0.0001. Edge weight = -log(p), truncated at 10
- Non-syndromic autism (i.e., autism spectrum disorder) is located at the center of network.
- All syndromic autisms are floating at the peripheral of the network.
- Modularity calculation generated 11 modules. Here are more significant ones:
  - 26.4% + 10.58% + 3.71%: ASD, psychiatric disorders, seizures, dementia, congenital disorders, learning disabilities, autistic and other complex syndromes
  - 28.85% + 13.19%: cancer & tumor
  - 15.66%: heterogeneous disorders relating to different physiological systems, e.g., autoimmune system*, heart, lung, kidney, lung, and diabetes

Network (p<.0001) Centrality Metrics: All Diseases

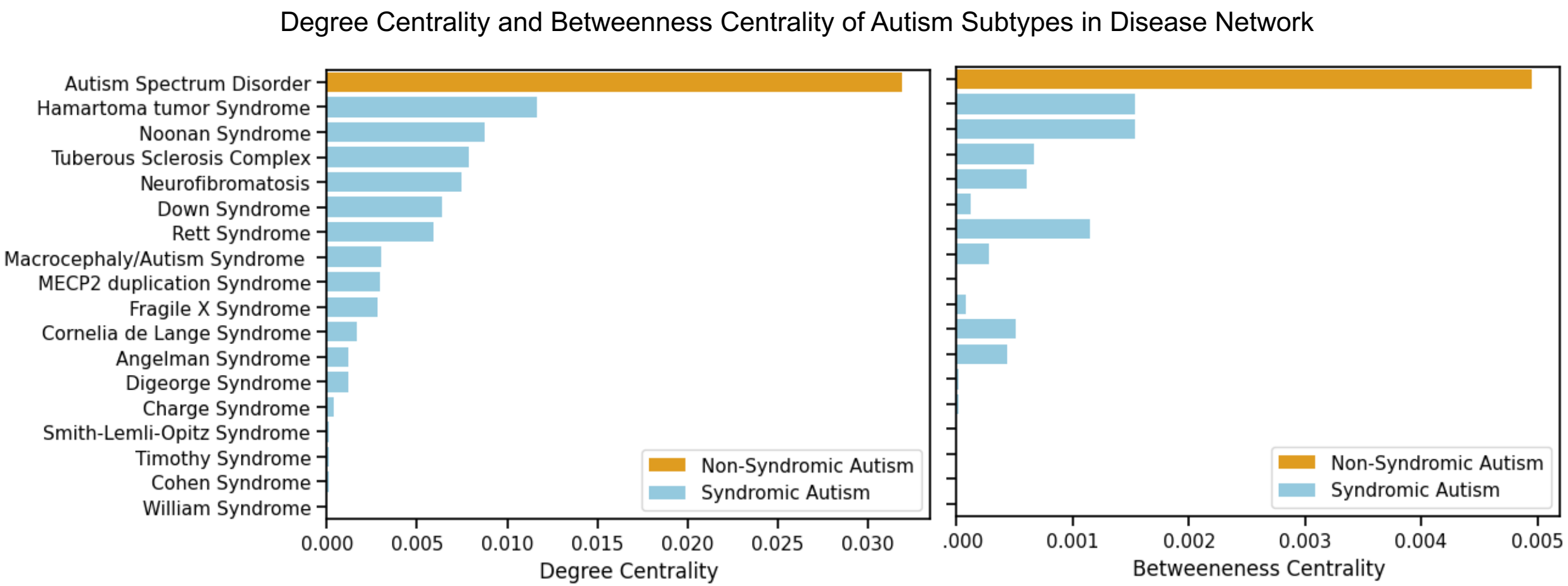




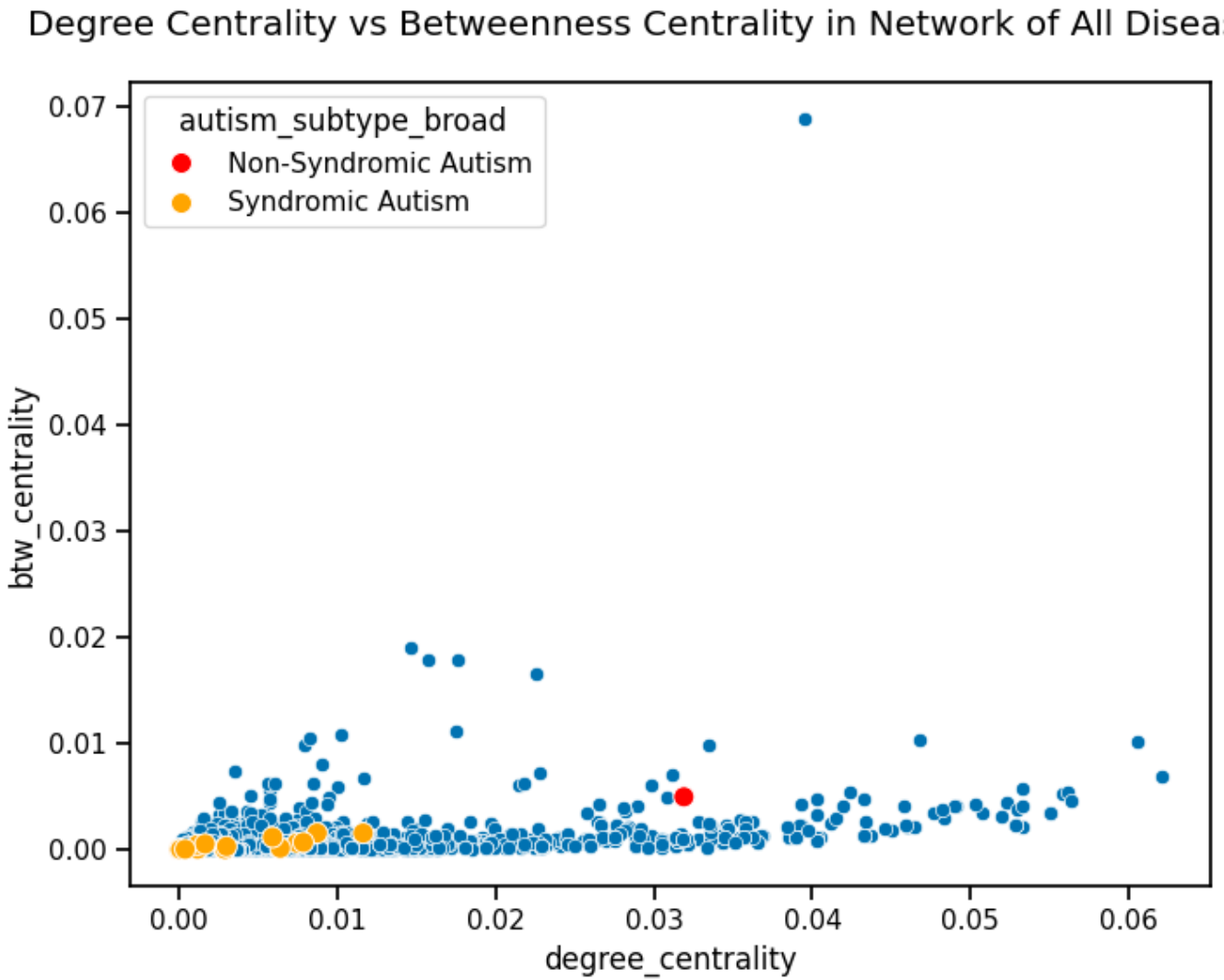




Network (p<.0001) Centrality Metrics: Autism Subtypes

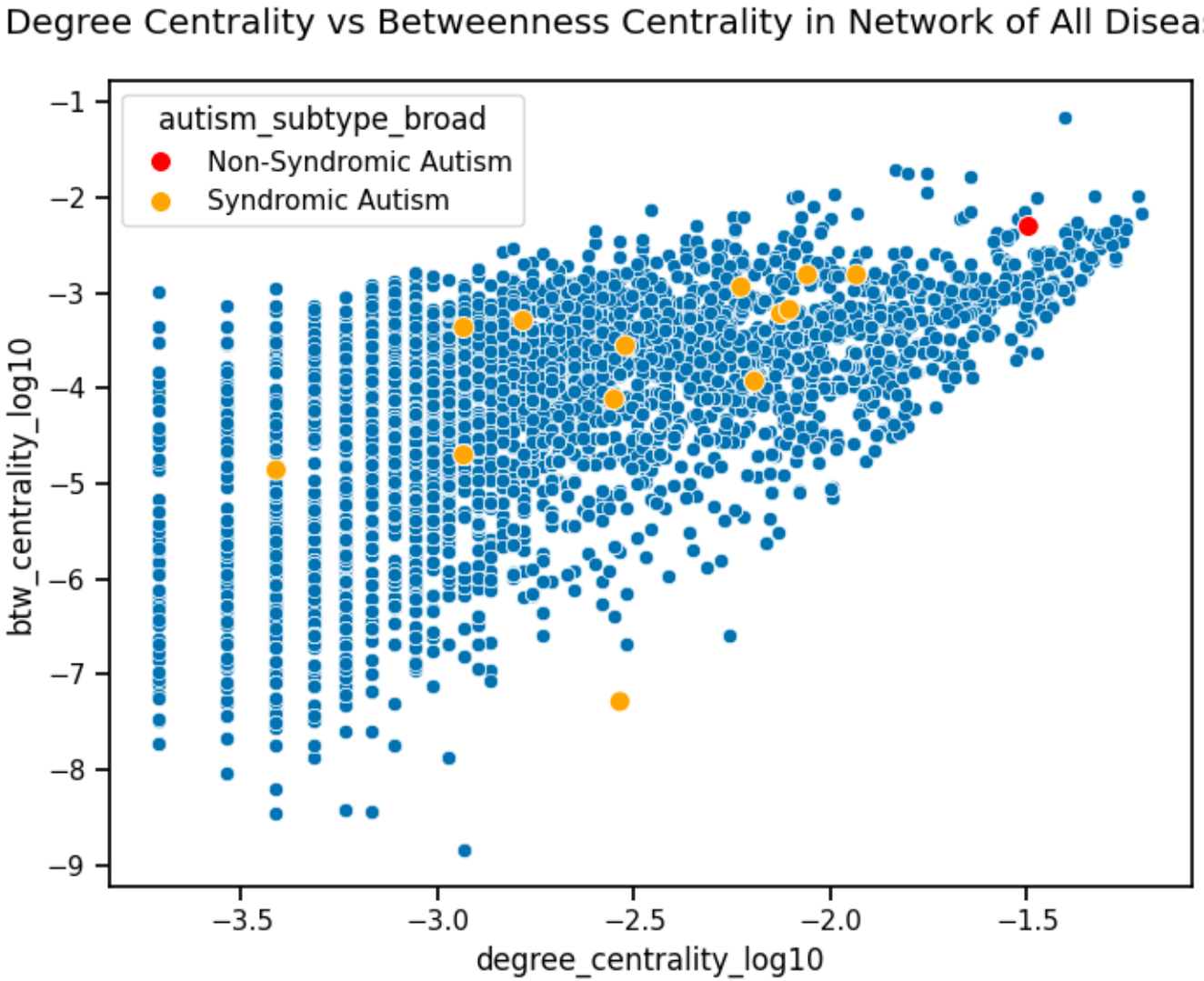


Network Centrality Metrics for Autism



Data points clustering due to skewed distribution.

Does Non-Syndromic Autism Really Behave Differently?

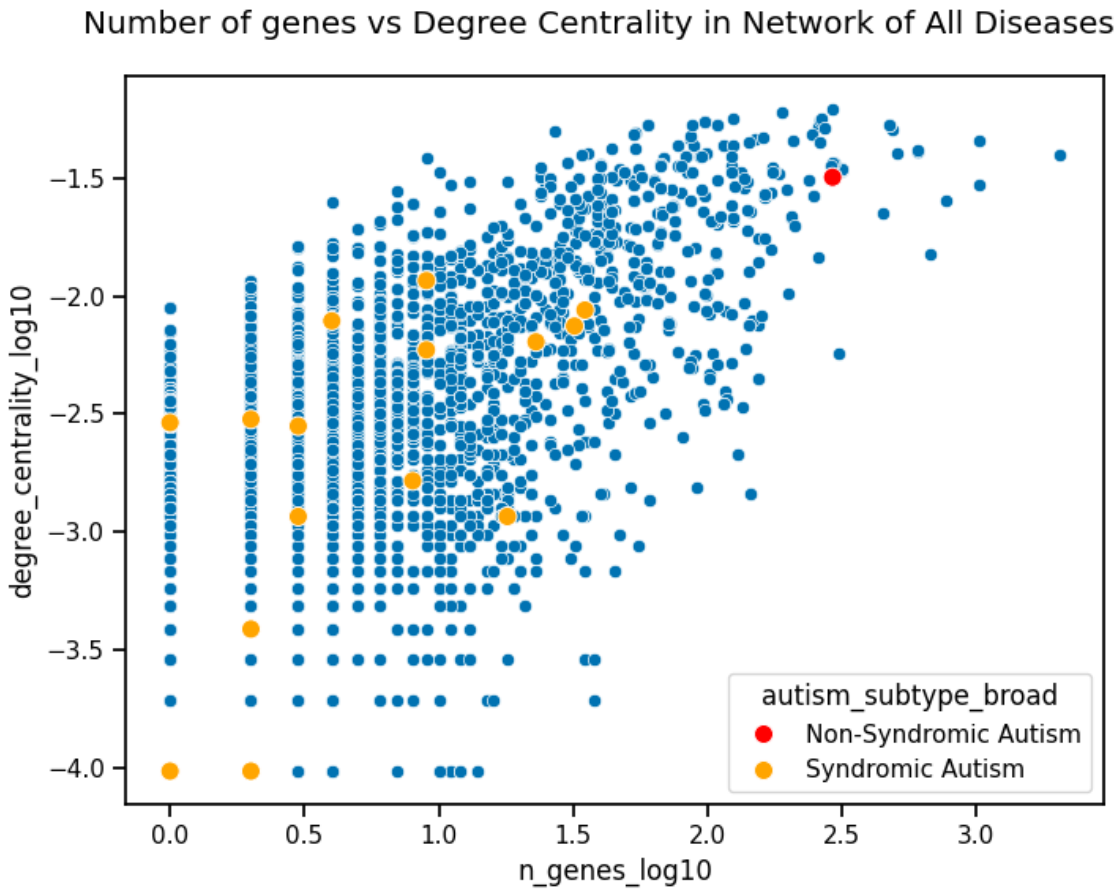


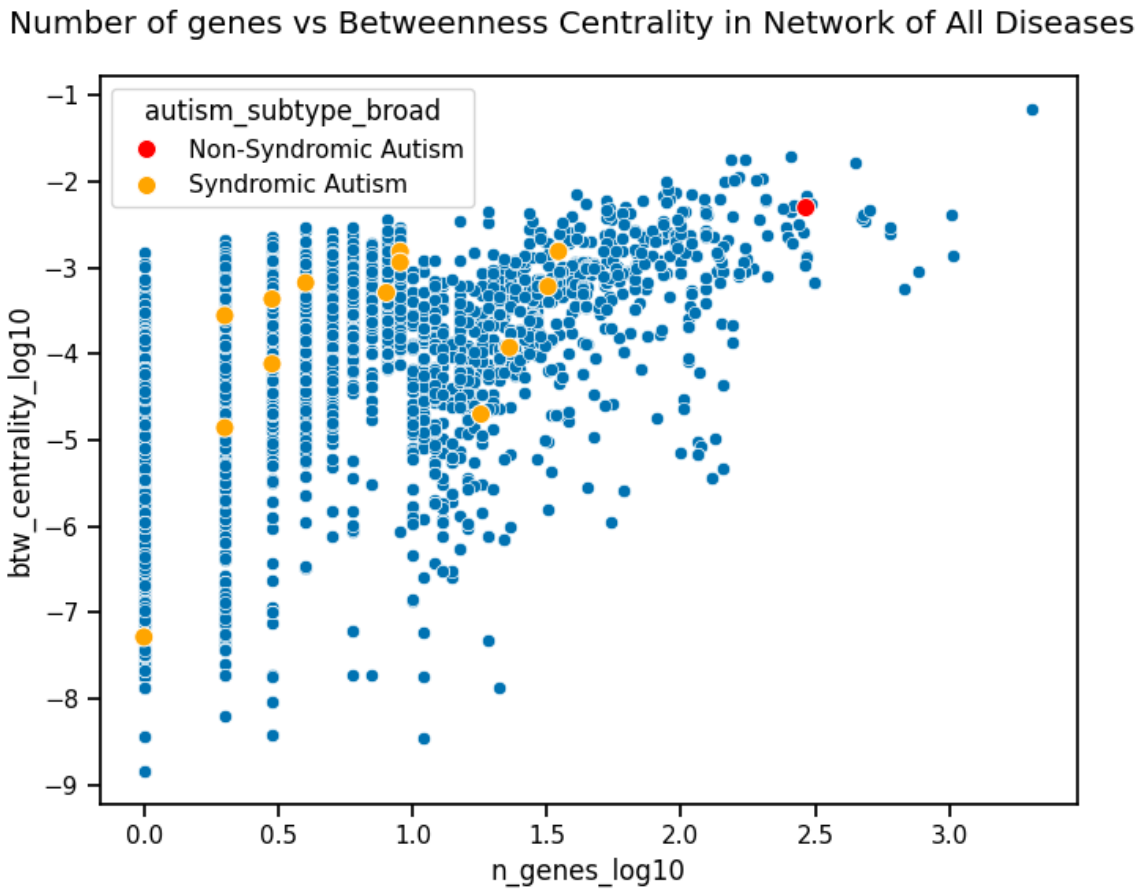
A lot of data are ceiling at 0 and not shown here due to log transformation.

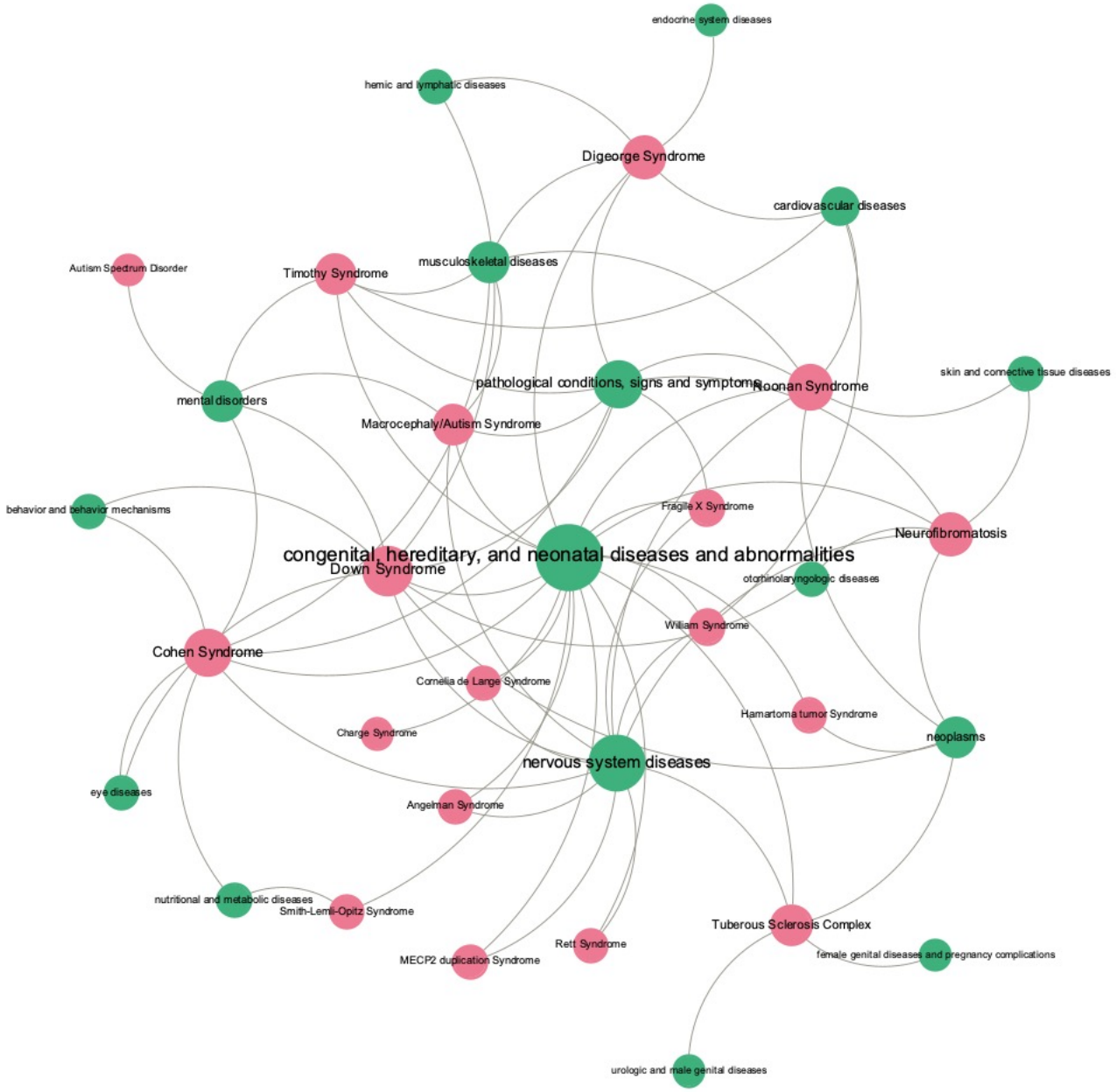
Some purely mono-genetic diseases might not show in the graph, because their centrality measurement is probably 0.

Is it appropriate to use Pearson's or Spearman's correlation analysis here?

Network Centrality Metrics for Autism







Autism and Disease Class Bipartite Network

- By contrast, Syndromic autism dominates over non-syndromic autism in a network of autism and disease classes.
- As syndromic autism diagnosis considered to involve multiple disease diagnosis class.

- Question:
  - Non-syndromic autism dominates in gene-disease network, as non-syndromic autism entails more genes than syndromic autism?
  - Syndromic autism dominates in gene-class network, as syndromic autism involves more disease class?

Limitation

- I did not do much statistical testing, because I was not sure which proper statistical test to use for dependent samples as well as network measurements.
- Hypermetric testing biases against diseases with small number of genes, e.g.,
  - Mono-genetic and mendelian disorders
  - Understudied disorders, e.g., the -log10(p-value) between autism and ADHD was only ~3.3, despite 50-70% comorbidity in clinical population. Because such big datasets often tend to bias toward rare diseases.
- But keeping edges with p-value below .0001 -> too many false positive.
- Further analysis – enrichment analysis?

References

Fuller, Z. L., Berg, J. J., Mostafavi, H., Sella, G., & Przeworski, M. (2019). Measuring intolerance to mutation in human genetics. *Nature Genetics*, *51*(5), 772-776.

Tick, B., Bolton, P., Happé, F., Rutter, M., & Rijsdijk, F. (2016). Heritability of autism spectrum disorders: a meta-analysis of twin studies. *Journal of Child Psychology and Psychiatry*, *57*(5), 585-595.

Sztainberg, Y., & Zoghbi, H. Y. (2016). Lessons learned from studying syndromic autism spectrum disorders. *Nature neuroscience*, *19*(11), 1408-1417.

Piñero, J., Ramírez-Anguita, J. M., Saüch-Pitarch, J., Ronzano, F., Centeno, E., Sanz, F., & Furlong, L. I. (2020). The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic acids research*, *48*(D1), D845-D855.