A Project Report

On

Network Pathway analysis for Autism Spectrum Disorder and its comorbid conditions.

BY

Aditya Agarwal

2017B1A71075H

Under the supervision of

Dr. Vidya Rajesh

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF

BIO F366: LABORATORY PROJECT

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI (RAJASTHAN)

HYDERABAD CAMPUS

(July 2020)

ACKNOWLEDGMENTS

I wish to express my gratitude towards Prof. Vidya Rajesh, Professor, Department of Biological Sciences for her continuous and enthusiastic support, cooperation and help throughout the duration of the project. I would also like to thank Prof. P. Sankar Ganesh and Prof. K. N. Mohan for their invaluable feedback and contributions and the Academic and Undergraduate Studies Division, BITS Pilani, Hyderabad Campus for giving me the opportunity to register in this project.

Birla Institute of Technology and Science-Pilani,

Hyderabad Campus

Certificate

This is to certify that the project report entitled "**Network Pathway analysis for Autism Spectrum Disorder and its comorbid conditions.**" submitted by Mr/Ms. Aditya Agarwal (ID No. 2017B1A71075H) in partial fulfillment of the requirements of the course BIO F366, Laboratory Project Course, embodies the work done by him under my supervision and guidance.

Date: 08/07/2020

(Dr. Vidya Rajesh)

Professor,

BITS- Pilani, Hyderabad Campus

ABSTRACT

In recent years, several studies have been conducted to study the conditions which are comorbid to Autism Spectrum Disorder (ASD). However, they have been directed towards studying the occurrence of a particular comorbid condition of ASD in isolation. Here, we demonstrate a holistic approach to study the complex network pathways involved in the occurrence of the comorbid conditions along with Autism. By doing a thorough literature survey and searching through relevant databases, we curated a list of genes associated with each comorbid condition. We performed network pathway analysis for each comorbid condition to build a network of protein interactions. We analysed the network pathways to identify important clusters of genes for each comorbid condition. We report protein interactions and connecting genes between ASD and its comorbid conditions that would provide insights useful for explaining the underlying genetic and pathophysiological causes for comorbidity in ASD. We also summarize the available tools for building network pathways and to visualize the protein interactions for future studies.

Keyword(s) : Autism, Autism Spectrum Disorder, Comorbidity, Network pathway analysis, Sleep disorders, Epilepsy-Seizures, Genes, Protein interactions.

Abbreviation(s) : ASD - Autism Spectrum Disorder, ADHD - Attention deficit hyperactivity disorder.

INTRODUCTION

Autism Spectrum or Autism Spectrum disorder (ASD) comprises Autism, Asperger's syndrome and other disorders of neuronal and brain development that are usually recognized at an early age (1-2 years) in children. Children with ASD often exhibit awkward social behavior, repetitive behavior and in general lack social and communication skills [1]. There is no concrete cause identified for Autism but family history and involvement of certain genes are usually considered as the possible causes [2,3,4,5]. Comorbidity is the simultaneous or co-occurrence of multiple disorders with a primary disorder in an individual. A comorbid condition is diagnosed after the primary disorder which offers core symptoms that differ from the comorbid condition [6]. There are many conditions that are comorbid to ASD such as Sleep disorders, Seizures and Epilepsy, ADHD, fragile X syndrome, Anxiety, Bipolar disorder [7]. It is often difficult to analyse the correlation between the prevalence of comorbidities with ASD due to some reasons like unavailability of test groups, non genetic factors contributing to phenotype etc.

There has been research into the comorbid conditions of ASD in the past [7], however each comorbid condition has often been studied in isolation, that is, it has been studied with respect to its co-occurrence with ASD while not taking into consideration other comorbid conditions of ASD. The aim of this study is to analyze and understand the complex network pathways associated with some significant comorbid conditions of ASD that have a possible genetic cause and to identify connecting links between the selected comorbid conditions and ASD and among themselves in terms of protein interactions and the underlying genetic causes. We also look at the various possibilities to model and visualize protein-protein interactions using bioinformatics tools.

METHODS AND RESULTS

1. **Identifying, understanding and selecting comorbid conditions of ASD that are to be studied.**

After filtering out non-significant conditions and conditions with low levels of comorbidity with autism, we selected four commonly occuring conditions that have a profound effect on the affected individual and also have been identified with a genetic basis. These are the four comorbid conditions we are going to discuss :

1) **Sleep Disorders**
2) **Seizures and Epilepsy**
3) **Attention deficit hyperactivity disorder**
4) **Intellectual disability**

After filtering through the comorbid conditions, we did a thorough literature survey of the network pathways, associated genes, protein interactions involved with each comorbid condition.

2. **Thorough literature survey of each comorbid condition to extract an exhaustive list of directly related genes for each comorbid condition.**

To obtain a comprehensive and exhaustive list of genes associated with each comorbid condition, a thorough search of online databases and available literature was conducted. Briefly, we queried the Online Mendelian Inheritance in Man database (OMIM, [8]) and the Human Gene Database (Gene, [9]). The findings were then cross-checked with the results from a thorough literature survey [10,11,12,13,14,15,16,17,18,19,20,21,22,23,24] of the directly associated genes with each comorbid condition and filtered out genes that had no evidence of a direct correlation with the absence or presence of the comorbid condition in order to obtain readable and accurate results from network pathway analysis.

Additionally, recent publications relevant to each comorbid condition were searched on PubMed (http://www.ncbi.nlm.nih.gov/pubmed/) and Google Scholar (http://scholar.google.com). Finally, 89 genes were identified to be associated with epilepsy and seizures while 43 genes were identified to be associated with sleep disorders. A number of protein molecules were also identified for each of the comorbid conditions. The genes used for further analysis for Sleep have been listed out in Table 1 (Please note that genes for Seizures and epilepsy have been made publicly available at: https://rb.gy/eb0rck).

Table 1
Sleep disorders Genes.

Phenotype	Encoded protein function	Genes
Sleep Disorders	Circadian Rhythm	CLOCK, TIMELESS, PER1, PER2 PER3, CRY1, CRY2, RORA, ARNTL, CSNK1D, HCRT, BTBD9
Sleep Disorders	Neurotransmission/ Cell Signalling	ADRA2A, SLC6A11, PTPRD, SLC6A13, SLC6A4, NOS1, ADA, HNMT, CHRND, COMT, FMR1
Sleep Disorders	G-Protein coupled receptors	P2RY11, DRD5, GABBR2
Sleep Disorders	Transcriptional Corepressors	SKOR1, BHLHE41
Sleep Disorders	Transcriptional Activators	TOX3, MEIS1, PPARGC1B
Sleep Disorders	Cell Proliferation/ Cell Immunity	TNFA, NF-kB, IL-5, CIITA, MAP2K5, ETAA1, PDCD2L, ROR1, DMPK
Sleep Disorders	Unknown/Other	CHKB, CPT1B

3. **Feeding the manually curated and cleaned list of genes to network pathway analysis tools.**

We began with a survey of the best already available tools for network pathway analysis. We tested out STRING [25], REACTOME [26], MetaDB [27], iPathwayGuide [28], KEGG pathways [29] to build networks from the available list of genes and protein molecules. We are currently waiting to hear from Qiagen in order to use the QIAGEN network pathway analysis tool [30] which is often used for these studies. In terms of visualization, acceptance from the scientific community, popularity and our expectations, we decided to go ahead with STRING which helps to locate functional links based on following criteria -

1) **Neighborhood in the Genome**
2) **Gene Fusions**
3) **Co-occurrence Across Genomes**
4) **Co-Expression**
5) **Experimental/Biochemical Data**
6) **Association in Curated Databases**
7) **Co-Mentioned in PubMed Abstracts**

(Figure 1 and Figure 2 depict the results for Sleep disorders and Seizures and Epilepsy respectively, obtained after some manual reorientation and K-means clustering applied on the output from STRING)

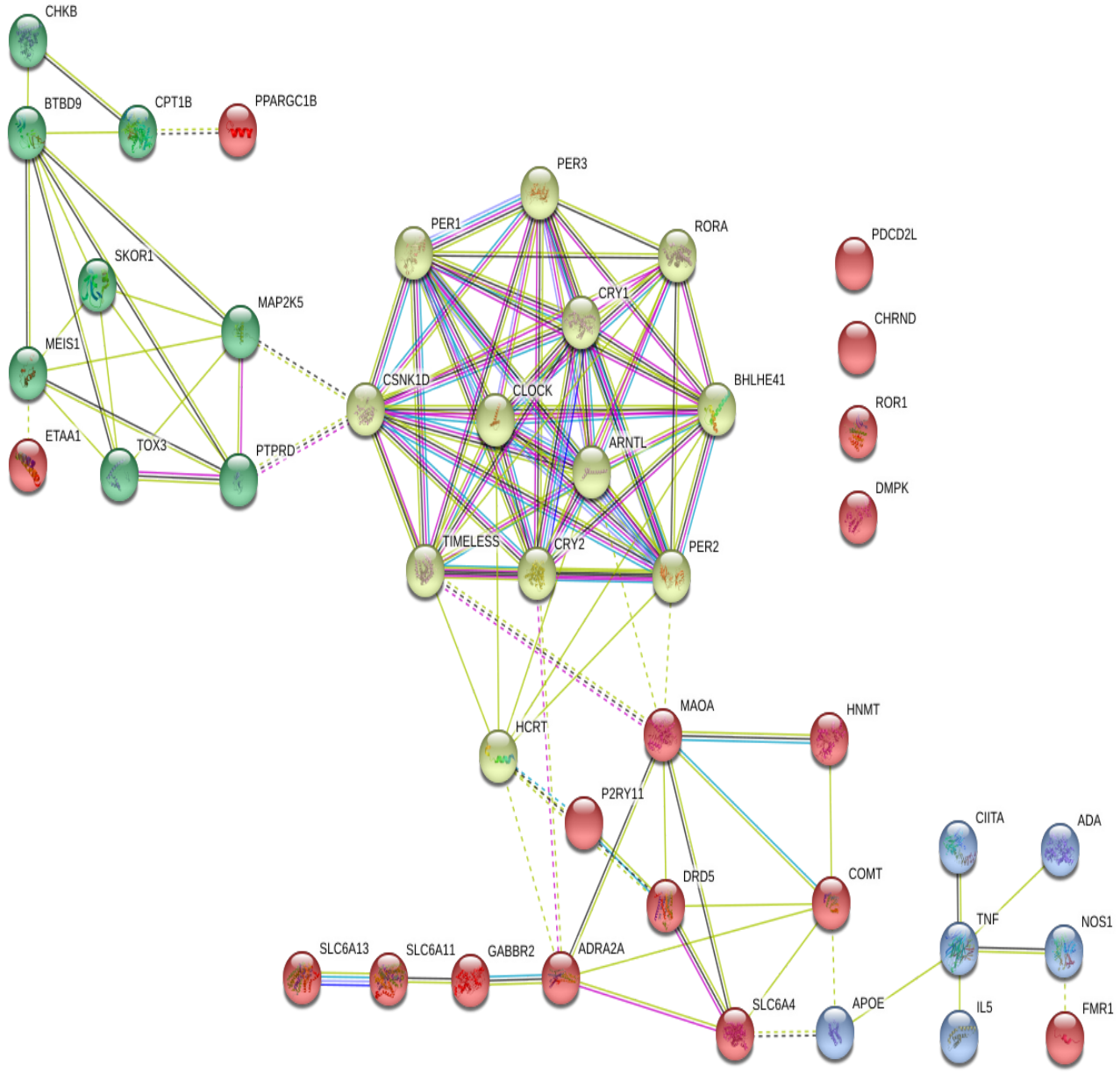

Figure 1. Network of pathways associated with Sleep Disorders. Sleep genes related to circadian rhythm of the human body have really strong functional interactions among themselves which is expected due to their closely knit pathways. K-Means clustering with K=4 has been applied to cluster the genes into four clusters. Blue nodes are genes involved in cell proliferation and neuronal development. Red nodes involve receptors subunits for neurotransmitters and other categories of genes. Yellow nodes are genes related to circadian rhythm while the green nodes represent genes involved in transcriptional activation and repression.

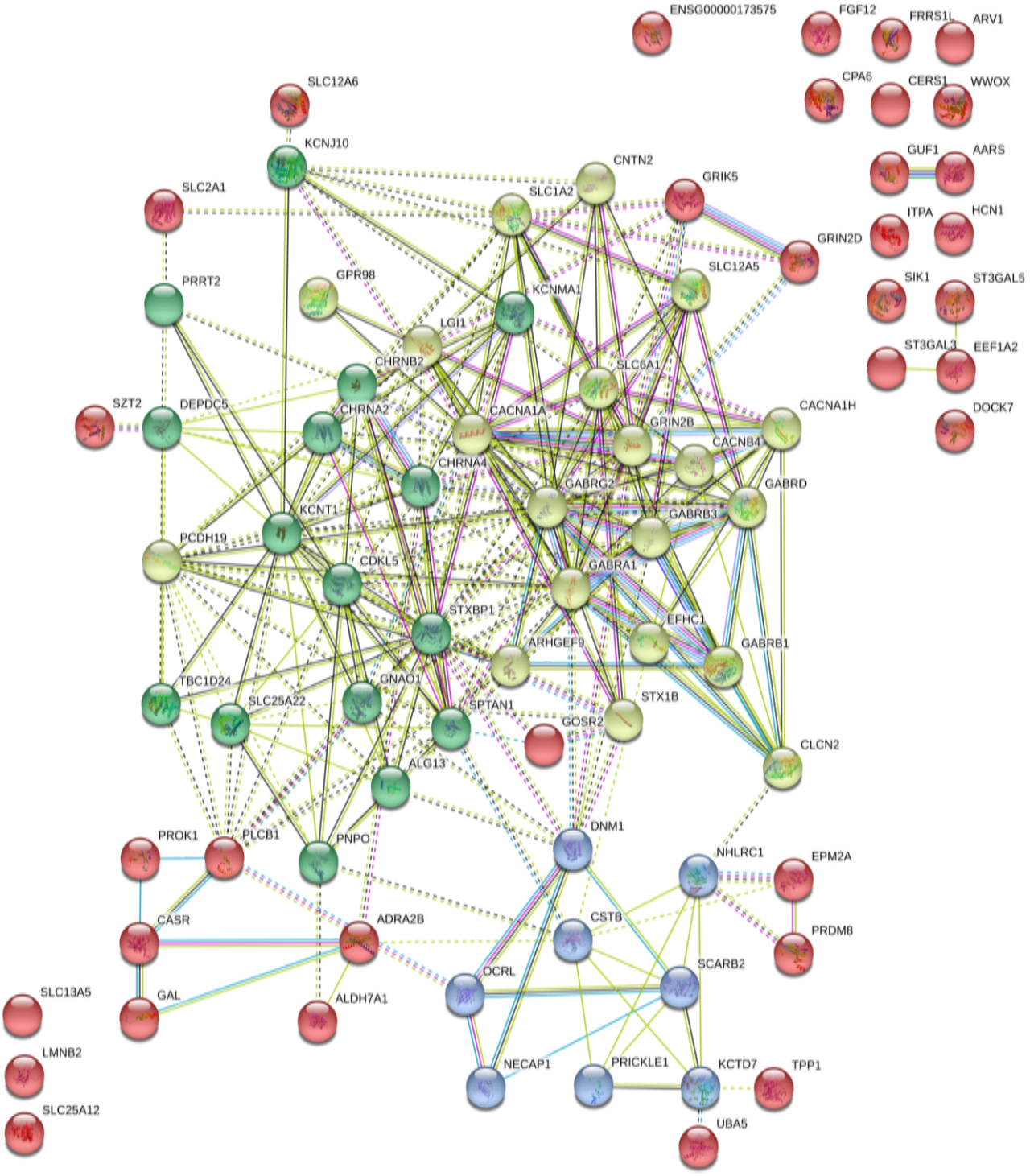

Figure 2. Network of pathways associated with Seizures and Epilepsy. K-Means clustering with K=4 has been applied to cluster the genes into four clusters. The result provides 3 major clusters of - Ion channel genes and transporter genes (Genes depicted as nodes in green), Excitatory and inhibitory hormones producing genes (Genes depicted as nodes in yellow) , Neuro-developmental genes (Genes depicted as nodes in blue). The last cluster of red nodes contains all other categories of genes and genes that did throw any connections to other genes in the network.

4. **Manually identifying clusters of genes to draw inferences from the resulting network pathways.**

Sleep Disorders

It can clearly be seen from the interactome in Figure 1 that sleep genes related to circadian rhythm have really strong functional interactions among themselves which is expected. Another such strong network is between the Genes associated with production, modification and deactivation of neurotransmitters. Important genes that can be seen are MAOA, DRD5, HCRT, CSNK1D, ADRA2A since they connect to multiple genes from different functional groups in the network of pathways. K-means clustering is applied with K=4 to identify four different clusters which can be seen to be associated with Circadian rhythm, Neurotransmitters, Neuronal Development and Signalling Pathways. Also, it can be noted that some genes did not exhibit any protein-protein interactions with any other gene. In no way does this signify their insignificance to Sleep disorders. They may be related to sleep disorders in pathophysiological ways that are not well understood or not reported before.

Seizures and Epilepsy

As is visible from Figure 2, due to the larger number of genes, the interactome looks really complex at first sight. But, with a closer look and some deeper insights, it can be seen that the reason for it being heavy on functional relationships is that the genes involved are ion-channel genes, transporter subunits and genes that are involved in a number of pathways that are clinical to the everyday functioning of our body, hence show connections to a large number of genes in the network. A lot of the genes are also neuro-developmental genes that will definitely be related to ASD as well. K-means clustering with K=4 provides 3 major clusters of - Ion channel genes and transporter genes, Excitatory and inhibitory hormones producing genes, Neuro-developmental genes. The last cluster contains all other categories of genes. Further study is required to understand the complex interaction network obtained between the proteins.

5. **Relating/connecting the individual network pathways to ASD by manually going through the gene interactions available in AutDB for ASD**

Connections found between the Sleep disorders network and Genes associated with ASD found in the Autism Database (AutDB) -

1) **GABA receptor subunits**
2) **Dopamine, Adenosine and Adrenergic receptors**
3) **Monoamine Oxidase**
4) **Period/ Clock genes**
5) **Solute carrier family 6 (neurotransmitter transporter)**
6) **Dystrophia myotonica-protein kinase (DMPK)**
7) **HLA Class II**

DISCUSSION

The study conducted for sleep disorders, seizures and epilepsy can be expanded to several other comorbid conditions of ASD using the same methodology. Deeper analysis of the network pathways is required for understanding the complex pathophysiological pathways involved in each comorbid condition and ASD itself. The study is hence, a drop in the ocean for understanding the complex network of ASD and its comorbid conditions but provides a holistic approach to study the same. We have also looked at another visualization tool, neo4j [31], a graph database for modelling the protein-protein interactions. In further studies we would recommend developing/using better pathway analysis tools for looking at the core protein molecules for each cluster of genes identified in the network of pathways.

REFERENCES

1. American Psychiatric Association. (2020). Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR) (4th ed.). American Psychiatric Association.

2. Steffenburg, S., Gillberg, C., Hellgren, L., Andersson, L., Gillberg, I. C., Jakobsson, G., & Bohman, M. (1989). A Twin Study of Autism in Denmark, Finland, Iceland, Norway and Sweden. *Journal of Child Psychology and Psychiatry*, *30*(3), 405–416. https://doi.org/10.1111/j.1469-7610.1989.tb00254.x

3. Bailey, A., Le Couteur, A., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, E., & Rutter, M. (1995). Autism as a strongly genetic disorder: evidence from a British twin study. *Psychological Medicine*, *25*(1), 63–77. https://doi.org/10.1017/s0033291700028099

4. Couteur, A., Bailey, A., Goode, S., Pickles, A., Gottesman, I., Robertson, S., & Rutter, M. (1996). A Broader Phenotype of Autism: The Clinical Spectrum in Twins. Journal of Child Psychology and Psychiatry, 37(7), 785–801. https://doi.org/10.1111/j.1469-7610.1996.tb01475.x

5. Szatmari, P., Jones, M. B., Zwaigenbaum, L., & MacLean, J. E. (1998). Genetics of Autism: Overview and New Directions. *Journal of Autism and Developmental Disorders*, *28*(5), 351–368. https://doi.org/10.1023/a:1026096203946

6. Angold, A., Costello, E. J., & Erkanli, A. (1999). Comorbidity. Journal of Child Psychology and Psychiatry, 40(1), 57–87. https://doi.org/10.1111/1469-7610.00424

7. Matson, J. L., & Goldin, R. L. (2013). Comorbidity and autism: Trends, topics and future directions. *Research in Autism Spectrum Disorders*, *7*(10), 1228–1233. https://doi.org/10.1016/j.rasd.2013.07.003.

8. Online Mendelian Inheritance in Man Database. http://www.ncbi.nlm.nih.gov/omim/

9. Human Gene Database. https://www.ncbi.nlm.nih.gov/gene

10. Wang, J., Lin, Z.-J., Liu, L., Xu, H.-Q., Shi, Y.-W., Yi, Y.-H., He, N., & Liao, W.-P. (2017). Epilepsy-associated genes. Seizure, 44, 11–20. https://doi.org/10.1016/j.seizure.2016.11.030

11. Barker-Haliski, M., & White, H. S. (2015). Glutamatergic Mechanisms Associated with Seizures and Epilepsy. Cold Spring Harbor Perspectives in Medicine, 5(8), a022863. https://doi.org/10.1101/cshperspect.a022863

12. Da Silva, R. S., Vianna, M. R. M. R., & Bonan, C. D. (2020). Seizures and epilepsy. Behavioral and Neural Genetics of Zebrafish, 413–432. https://doi.org/10.1016/b978-0-12-817528-6.00024-3

13. Galanopoulou, A. S. (2010). Mutations affecting GABAergic signaling in seizures and epilepsy. Pflügers Archiv - European Journal of Physiology, 460(2), 505–523. https://doi.org/10.1007/s00424-010-0816-2

14. Thomas, R. H., & Berkovic, S. F. (2014). The hidden genetics of epilepsy—a clinically important new paradigm. Nature Reviews Neurology, 10(5), 283–292. https://doi.org/10.1038/nrneurol.2014.62

15. Taheri, S., & Mignot, E. (2002). The genetics of sleep disorders. The Lancet Neurology, 1(4), 242–250. https://doi.org/10.1016/s1474-4422(02)00103-5

16. Sehgal, A., & Mignot, E. (2011). Genetics of Sleep and Sleep Disorders. Cell, 146(2), 194–207. https://doi.org/10.1016/j.cell.2011.07.004

17. Ebisawa, T. (2007). Circadian Rhythms in the CNS and Peripheral Clock Disorders: Human Sleep Disorders and Clock Genes. Journal of Pharmacological Sciences, 103(2), 150–154. https://doi.org/10.1254/jphs.fmj06003x5

18. Tafti, M., Maret, S., & Dauvilliers, Y. (2005). Genes for normal sleep and sleep disorders. Annals of Medicine, 37(8), 580–589. https://doi.org/10.1080/07853890500372047

19. Ryan, S., Taylor, C. T., & McNicholas, W. T. (2006). Predictors of Elevated Nuclear Factor-κB–dependent Genes in Obstructive Sleep Apnea Syndrome. American Journal of Respiratory and Critical Care Medicine, 174(7), 824–830. https://doi.org/10.1164/rccm.200601-066oc

20. HLA Class II Genes Associated with REM Sleep Behavior Disorder Carlos H. Schenck, MD,*t Edgar Garcia-Rill, PhD,$ Miriam Segall, PhD,' Harriet Noreen, CHS, and Mark W. Mahowald, MD*$**

21. Raizen, D. M., & Wu, M. N. (2011). Genome-Wide Association Studies of Sleep Disorders. Chest, 139(2), 446–452. https://doi.org/10.1378/chest.10-1313

22. Cortesi, F., Giannotti, F., Ivanenko, A., & Johnson, K. (2010). Sleep in children with autistic spectrum disorder. Sleep Medicine, 11(7), 659–664. https://doi.org/10.1016/j.sleep.2010.01.010

23. Parish, J. M. (2013). Genetic and Immunologic Aspects of Sleep and Sleep Disorders. Chest, 143(5), 1489–1499. https://doi.org/10.1378/chest.12-1219

24. Kripke, D. F., Kline, L. E., Nievergelt, C. M., Murray, S. S., Shadan, F. F., Dawson, A., Poceta, J. S., Cronin, J., Jamil, S. M., Tranah, G. J., Loving, R. T., Grizas, A. P., & Hahn, E. K. (2015). Genetic variants associated with sleep disorders. Sleep Medicine, 16(2), 217–224. https://doi.org/10.1016/j.sleep.2014.11.003

25. Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. von. (2018). STRING

v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Research, 47(D1), D607–D613. https://doi.org/10.1093/nar/gky1131

26. Croft, D., Mundo, A. F., Haw, R., Milacic, M., Weiser, J., Wu, G., Caudy, M., Garapati, P., Gillespie, M., Kamdar, M. R., Jassal, B., Jupe, S., Matthews, L., May, B., Palatnik, S., Rothfels, K., Shamovsky, V., Song, H., Williams, M., … D'Eustachio, P. (2013). The Reactome pathway knowledgebase. Nucleic Acids Research, 42(D1), D472–D477. https://doi.org/10.1093/nar/gkt1102

27. Guo, Z., Tzvetkova, B., Bassik, J. M., Bodziak, T., Wojnar, B. M., Qiao, W., Obaida, M. A., Nelson, S. B., Hu, B. H., & Yu, P. (2015). RNASeqMetaDB: a database and web server for navigating metadata of publicly available mouse RNA-Seq datasets. Bioinformatics, btv503. https://doi.org/10.1093/bioinformatics/btv503

28. Advaita, 2019. https://advaitabio.com/ipathwayguide/

29. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2016). KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Research, 45(D1), D353–D361. https://doi.org/10.1093/nar/gkw1092

30. Krämer, A., Green, J., Pollard, J., & Tugendreich, S. (2013). Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics, 30(4), 523–530. https://doi.org/10.1093/bioinformatics/btt703

31. Webber, J. (2012). A programmatic introduction to Neo4j. Proceedings of the 3rd Annual Conference on Systems, Programming, and Applications: Software for Humanity - SPLASH '12, 217–218. https://doi.org/10.1145/2384716.2384777