IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

A Meta-analysis of Syndromic Autism Genes

Daria Lunina

European Gymnasium, Moscow, 107113, Russian Federation.; luninadasha@gmail.com
Mentor: Kif Liakath-Ali

ABSTRACT: Autism is a neurodevelopmental disorder characterized by severe impairment of various behavioral functions. Mutations in several genes are known to be associated with autism. In this study, a meta-analysis was performed on 130 genes implicated in syndromic autism by the SFARI database. Results show that several genes are associated with known human pathological conditions, such as delayed speech and language development. Further analysis reveals that many genes are associated with functions related to head development, Rett syndrome (psychoneurological syndrome), and neuron projection morphogenesis. Protein network analysis revealed closely associated phenotypic terms with head development and neuron projection morphogenesis functional groups, suggesting that mutations in these genes significantly affect neuronal formation and, in turn, result in autistic characteristics. This study sheds light on the general role of autism-related genes, their interactions, and how mutations in them may disrupt normal neuronal function.

KEYWORDS: Behavioural and Social Sciences; Neuroscience; ASD; Autism Spectrum Disorder; Genetics; Meta-analysis; Causes of ASD.

Introduction

Autism, or Autism Spectrum Disorder (ASD), is a neurodevelopmental disorder in the category of pervasive developmental disorders characterized by severe and pervasive impairment in reciprocal socialization, qualitative impairment in communication, and repetitive or unusual behavior.[1] The word "autism" was first used in 1911 by German psychiatrist Eugen Bleuler to describe schizophrenia symptoms, such as replacing unsatisfying reality with hallucinations and vivid visions.[2] However, in the 1960s, the meaning radically changed as British child psychologists started using it instead to reflect a lack of imaginative thinking.[3] ASD is a relatively widespread condition, with the Centers for Disease Control stating that 1 in 54 children are affected by autism in the USA.[4] Moreover, estimated rates of autism have increased in the period from the 1960s to the 1980s from 5 to 72 cases per 10 000 children in Europe and the USA.[5,6] The underlying neurobiological pathology of ASD remains unclear. Research findings in this sphere suggest that children with confirmed autism have macrocephaly at 2-3 years old. Brain growth is rapid at 12 months, and the frontal, temporal lobes, and limbic structures enlarge during the first two years of life, correlating with symptom development.[7] Additionally, fMRI studies have found that people with autism have abnormal hypoactivation in the fusiform face area, which can be connected with problems in human perception compared with objects.[8] ASD symptoms usually appear around 18 months, and treatment is most effective when applied as early as possible.[9] Neurochemical studies have identified genetic differences in serotonin transport which are supported by empirical data on ASD-related symptoms.

One of the main problems with identifying the cause of this disorder is that it is unknown whether environmental factors or the genetic heritage is most influential (although the genetic cause is considered most probable), but also the sheer number of genes associated with Autism. As estimated by SFARI (Simons Foundation Autism Research Initiative), a comprehensive database, there are 1003 genes implicated in ASD.[10] Data from this resource shows that 130 of these genes are considered syndromic, meaning they were found to be associated with ASD symptoms and linked to additional characteristics not required for an ASD diagnosis. Analysis of these syndromic genes will further our understanding of the genetic basis of ASD. The study presented here aims to conduct a meta-analysis of these 130 syndromic genes using tools provided by open resources, including Metascape[11] (a gene annotation and analysis resource), and reveal their ontology and biological role in brain development. Results from this study will contribute towards a better understanding of this disorder and will provide a computational analysis of syndromic genes, which will be useful in ASD-related research.

Methods

To perform a meta-analysis of syndromic genes implicated in ASD, a curated list of genes was downloaded from the SFARI Gene database, which classifies genes based on four scores: S (Syndromic), 1 (High Confidence), 2 (Strong Candidate), and 3 (Suggestive Evidence). The syndromic category includes genes with mutations reported to be associated with a substantial degree of increased risk and consistently linked to additional characteristics not required for an ASD diagnosis. It was reasoned that performing a meta-analysis on the Syndromic category of genes would be essential to understand how mutations in these genes may manifest ASD. Therefore, a total of 130 genes in the "Syndromic" category were entered into the Metascape under express and custom analysis (Supplementary Table 1). The genes were analyzed for gene summary, development of disorders, disease and gene association, gene placement

DOI: 10.36838/v5i4.23

on the chromosome, molecular function, RNA tissue category, and gene expression in different tissues.

The characteristics used for the custom analysis are reflected in Supplementary Table 2. Metascape has provided CAME analysis (Id conversion - convert input gene identifiers into Entrez gene[21] IDs of a target species, annotation - extract annotation columns for the gene list, including gene descriptions, functions, and protein classes, membership - flag genes which fall under GO biological process terms that contain "invasion" as a keyword, function enrichment analysis of the gene list - identifying pathways (or complexes, published hit lists, etc.) that have statistically significant p-values and protein network analysis) on the bases of *Homo sapiens* genes. This information was further reflected in graphs and tables, taken from express analysis, which was discussed in this research's "Results" section.

Results

The gene summary provided by Metascape reflected that all genes are protein encoding. Additionally, it was found that all genes are expressed, at least in some tissues. In particular, about 12.3 percent of genes are expressed in brain tissue (16 out of 130 genes) (Supplementary Table 2). Analysis of syndromic genes for known human phenotypes showed pathological conditions associated with ASD (Figure 1). The most frequent phenotypes associated with these genes are delayed speech and language development, autistic behavior, downward slant of the palpebral fissure, strabismus, and neurodevelopmental disorder. Of them, two (downward slant of palpebral fissur-behaviore and strabismus) are the causes of eye development pathologies. As stated by DisGeENT, a database of diseases associated with genes concerning NCBI,[12] autistic behavior is defined by: "Persistent deficits in social interaction and communication and interaction as well as a markedly restricted repertoire of activity and interest as well as repetitive patterns of behavior" while a neurodevelopmental disorder is described as: "behavioral and cognitive disorder with onset during the developmental period that involves impaired or aberrant development of intellectual, motor, or social functions." Thus, it is evident that mutations in these genes are associated both with behavioral abnormalities and eye development pathologies. The lowest p-value negative logarithm is associated with a depressed nasal bridge and is - 23.

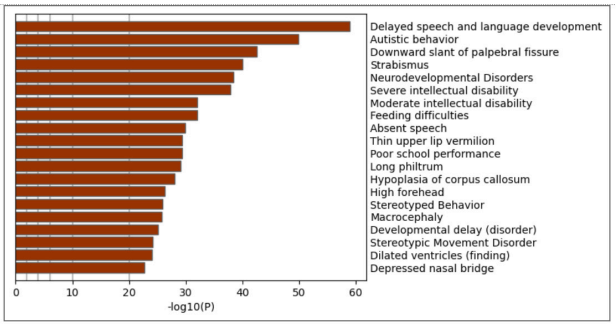


Figure 1: DisGeNET[13] summary. Bar graph shows -log10(P) values of syndromic SFARI genes implicated in known human pathological conditions. Note that top gene enrichment is with delayed speech and language development, which is a clinical phenotype in ASD.

The syndromic gene list was analyzed for enriched terms based on their functional roles. The highest p-values of the functions linked with the syndromic genes are found within the groups the head development, Rett syndrome (psychoneurological syndrome), neuron projection morphogenesis, and covalent chromatin modification (Figure 2). The most significantly enriched genes were found to be associated with head development (22 genes), neuron projection morphogenesis (19 genes), and covalent chromatin modification (16 genes), as indicated in the Metascape express analysis shown in Table 1.

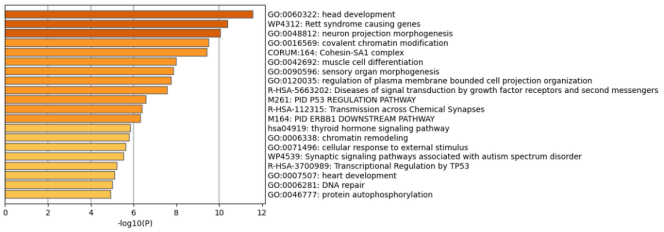


Figure 2: Bar graph of enriched terms across input gene lists, colored by p-values. The bar graph shows that most syndromic genes are associated with head development, Rett syndrome (psychoneurological syndrome), neuron projection morphogenesis, and covalent chromatin modification.

Next, the syndromic genes were analyzed for the network of enriched terms, the relationship between genes, and the functions of the proteins they code for. Gene function could be defined as the encoding of particular molecules and proteins. In contrast, protein function depends on the processes in which they are involved, such as the structural integrity of cells or control over cell division. The nodes represent the similarity within a gene's function, and links between nodes represent similarities between functions of proteins, which are encoded by ASD-related genes (Figure 3). One of the first things that can be seen is the abundance of links between nodes of the same color (cluster ID).

An example is the gene nodes, which influence sensory organ morphogenesis (brown cluster ID). They are strongly linked with each other and, due to the small number of links with genes of different cluster IDs, are separated from the main body of the diagram. In other words, genes of a similar coding function tend to code for proteins with a similar function.

Slightly above the center of the diagram, the group of small nodes that mostly have different gene coding functions, judging from the cluster IDs, appear to have many links between each other. For example, the chromatin remodeling node (magenta cluster ID) is surrounded by many links, which form a uniform circle around it. That means that small groups of genes, which code for proteins of different groups, make up the proteins with similar functions. It is also evident that these links appear to be thick. For example, DNA repair (turquoise cluster ID), which has few associated genes (small node), has many thick links between other nodes (different cluster ID). The thickness of the link represents that the function of one group of genes is strongly related to the function of the other group. This could mean that there are small groups of genes with different coding functions, which are firmly related to one another in two ways: similarity in the functions of proteins, which they code for, and the procedure of executing the gene's

DOI: 10.36838/v5i4.23

coding function. It is also evident that cluster IDs of the smaller nodes represent processes that are not directly connected to ASD symptoms, such as chromatin remodeling mentioned before.

This data suggests two possible explanations for the genetic causes of ASD. The first option is that the interactions between small separate groups of genes, which code for proteins not directly linked to the ASD symptoms, result in the biological processes, which lead to the ASD-related phenotype, such as enlargements of the frontal, temporal lobes, and limbic structures. The second is that interactions between closely related genes of the same function, for example, head development genes, could result in the development of ASD. However, because even closely related groups of nodes have some links with other nodes, it could be stated that a more complex system of interactions between all the clusters results in the formation of ASD.

As shown in Figure 3b, which depicts the network of enriched terms from this analysis, the larger the cluster of genes with closely related functions, the lower the associated p-value. This relationship reflects a higher probability of the stated genes expressing associated proteins. Head development and neuron projection morphogenesis functional groups have the lowest p-values compared to other cluster IDs. Such a conclusion supports the idea that ASD causation might lie in the interactions between closely related genes of the same cluster-ID (function), as they are more likely to code for the associated proteins. In particular, those genes might be in head development and neuron projection functional morphogenesis groups.

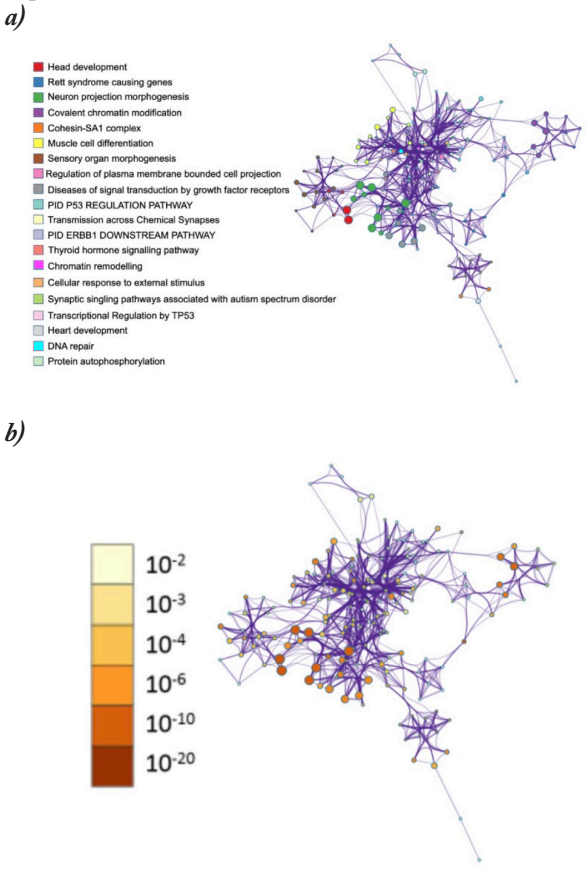
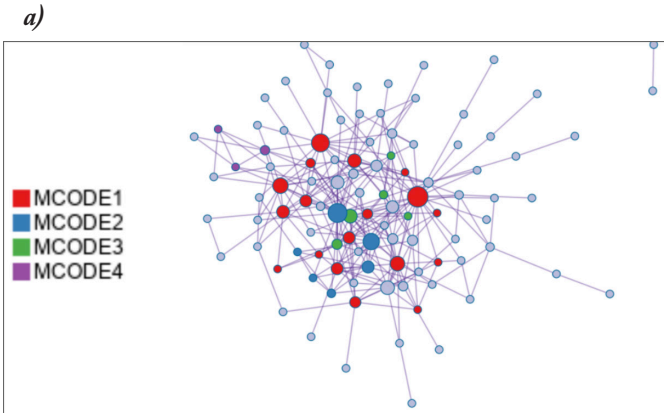


Figure 3: a) Network of enriched terms colored by cluster ID, where nodes that share the same cluster-ID are typically close to each other. The links between the nodes reflect the similarities between the function of the proteins expressed by the genes. The thickness of links represents the genes closely related to their function within the expressed phenotype, reflected in a color; (b) Network of enriched terms (nodes): colored by p-value, where terms containing more genes tend to have a more significant p-value.

For each given gene list, protein-protein interaction enrichment analysis has been carried out with the following databases: STRING,[14] BioGrid,[15] OmniPath, and [16] InWeb_IM.[17] The resultant network contains the subset of proteins that form physical interactions with at least one other member in the list. The Molecular Complex Detection (MCODE) algorithm has been applied to identify densely connected network components. The protein-protein interaction network (Figure 4) reflects a close relationship between four groups of genes encoding for similar proteins: Cohesin-SA1 complex, membrane transport-associated proteins, chromatin-associated genes, and modification-related genes. The fact that the proteins expressed by some ASD syndromic genes are closely related in their functional communication suggests possible mechanisms of ASD causation. However, it should also be stated that proteins depicted in the physical protein-protein interaction network, which do not hold high p-values for the assigned functions, are not included in large groups of nodes, or are interconnected with many nodes of different cluster IDs in the Network of enriched terms. This condition differentiates illustrated findings from the Figure 3a data. Some of the connections within the group of proteins with the same functions depicted in Figure 3 have been deleted. Apart from that, it is evident that many genes stated in Table 4, such as the histone modification gene, are not represented in Figure 3a legend. The possible explanation for these differences lies in the emphasis on the physical interactions between proteins and the fact that only genes with higher p-values were chosen to be depicted, as pointed out in the description of Figure 4. For instance, this condition limits the list of genes that could be included in the interactions between their proteins. Due to that fact, genes with more specific functions had to be chosen, while in Figure 3a, they might have belonged to more general function cluster IDs.



b)

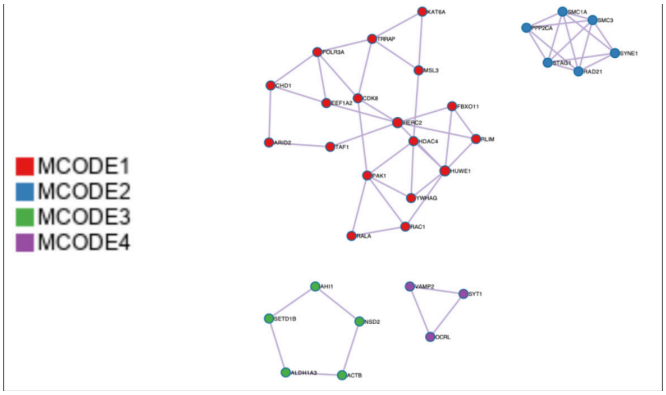


Figure 4: (a) Protein-protein interaction network and MCODE (Molecular Complex Detection) components identified in the gene lists. Only physical interactions in STRING (physical score > 0.132) and BioGrid were used. (b) The MCODE networks were identified for individual gene lists. It also contains pathway and process enrichment analysis applied to each MCODE component independently. The Figure shows interactions within one MCODE group of genes, reflecting the relationship between individual genes within the group.

Table 1: Color codes and descriptions for Figure 4. Three best-scoring terms by p-value as the functional description of the corresponding components, according to network plots within Figure 3.

Color	MCODE	GO	Description	Log10(P)
	MCODE_1	GO:0016569	covalent chromatin modification	-8.1
	MCODE_1	GO:0016570	histone modification	-6.6
	MCODE_1	M76	PID P38 ALPHA BETA PATHWAY	-6.1
	MCODE_2	CORUM:63	Cohesin-SA1 complex	-15.2
	MCODE_2	CORUM:164	Cohesin-SA1 complex	-15.2
	MCODE_2	CORUM:165	Cohesin-SA1 complex	-15.2
	MCODE_3	R-HSA-4839726	Chromatin organization	-5.0
	MCODE_3	R-HSA-3247509	Chromatin modifying enzymes	-5.0
	MCODE_3	GO:0002009	morphogenesis of an epithelium	-4.2
	MCODE_4	R-HSA-8856828	Clathrin-mediated endocytosis	-6.9
	MCODE_4	R-HSA-199991	Membrane Trafficking	-4.9
	MCODE_4	R-HSA-5653656	Vesicle-mediated transport	-4.9

Discussion

The analysis results reflect that most syndromic genes analyzed from the SFARI database, indeed, most frequently, are associated with ASD-related symptoms. The data in Figure 1 showing the phenotypes associated with p-values reflects a similar idea, as, for example, head development may be linked with ASD symptoms. It is characterized by enlargements of the frontal, temporal lobes, and limbic structures, which occur at the age of 2, simultaneously developing other symptoms. Also, there are two trends found within the network of enriched terms. ASD is formed by genes' interactions in many small functional groups not directly linked with ASD symptoms. In a few big functional groups, directly connected with ASD symptoms, such as head development. The latter option also corresponds with smaller p-values, indicating that these genes are more likely to be associated with their function. Protein-protein interactions reflect the enrichment term analysis graph trend, with a few differences. For instance, several links between the gene's functional groups are deleted, compared to Figure 3. Although possible reasons for this and other differences were stated in the "Results" section, there is a lack of certainty.

Further investigations into detailed interactions between proteins of genes may provide more solid explanations for these findings. The genes represented in Figure 4 and Table 1 are responsible for the phenotype, which is not directly linked with ASD symptoms. It should also be stated that the p-values of genes depicted in Table 1 broadly vary even in one indicated group, even though the genes were chosen about their p-values. For instance, in group MCODE_1 covalent chromatin modification gene (GO:0016569) has a Log10(P) of -8.1, while the histone modification gene (GO:0016570) has a Log10(P) of -6.6. It means that the probability that particular gene codes for the stated protein, which in turn interacts with other proteins in the group, is uneven. This identifies the possibility that a different gene, not included in the analysis, may take part in the interaction. The idea also compromises the validity of the gene-to-gene communications reflected both in Figure 3a and Figure 4, showing that the relationship between proteins may not correspond to the assigned genes.

The findings stated above reflect links between genes and ASD-related phenotypes. The links could be explored in further research to understand better genes that may cause ASD. Additionally, relationships between groups of genes and links between groups may shed light on ASD formation, which could result from interactions between these genes rather than an expression of one particular gene.

One study examining the causative role of SFARI genes in ASD reported confounding findings about the database.[18] It combined information about gene expression ands clinical data across 80 samples and built a gene co-expression network analyzed at different granularity. The study found no significant link between SFARI's gene expression and differential expression patterns comparing the group with diagnosed ASD and the control group. Also, no strong correlations between the expression of SFARI genes and ASD diagnosis were found through the gene co-expression network.

DOI: 10.36838/v5i4.23

These findings suggest that genes analyzed in this study might not fully represent the ASD symptoms and causes.

However, the correlation between levels of expression and SFARI's gene scores was statistically significant. It is valuable for the justification of the genes chosen for the study. SFARI genes were also connected with other neurodevelopmental disorders, such as Schizophrenia, Bipolar Disorder, Depressive Disorder, and others.

The study also suggested a new approach to using such databases against biases. Analysis of co-expression networks, including data from the whole network, can provide more meaningful information about genes linked to ASD diagnosis and even suggest new candidates. Incorporating this approach in further studies would be paramount for deeper investigation of ASD-related genes.

Another study investigated the variations of genes linked to ASD. To be more precise, they examined rare CNVs (copy-number variation) and proteins disrupting SNVs (single-nucleotide variants).[19] The study used Infinium PsychArray as a tool for the analysis. It was characterized as an array of SNP (single-nucleotide polymorphism). Using it, researchers performed an integrated analysis of CNVs and SNVs collection of the homogeneous cohort of 127 Italian ASD families genotypes. Researchers found a higher number of rare CNVs, most frequently deletions, in ASD individuals compared to the unaffected group. They also observed a multitude of rare CNVs in the SFARI gene collection used in the study. It was indicated that ASD affected group had an increased transmission of rare SNVs variants from heterozygous parents to probands. These findings support the idea of multi-genetic nature of ASD, which also corresponds with the conclusions of this paper.

Researchers also stated the significant role of the following genes in ASD development: VPS13B, WWOX, CNTNAP2, RBFOX1, MACROD2, APBA2, PARK2, GPHN, and RNF113A. From this list, only one gene (CNTNAP2) was included in this paper's analysis. Hence, further investigation of the abovementioned genes would also provide more insight into ASD gene research.

As this research examined only human genes, it would be insightful to investigate genes of other species, such as mice (*Mus musculus*) or zebrafish (*Danio rerio*). Analyzing ASD genes in other organisms may provide a new understanding of their interactions and a fuller comprehension of ASD causation. Furthermore, as only syndromic genes from the SFARI database were analyzed, further research may examine other score sections from the same database or genes from other databases. Such examination would enrich the conclusions derived from the analysis described above and provide a better understanding of ASD.

It should also be stated that some of the genes used in this research have been marked in Allen Brain Atlas[20] (*actb ahi1 vamp2 syt1 syne1 rps6ka3 prodh phf8 pax6 ntrk2 ntng2 ntng1 nr2f1 mtor hcn1 gabra3 eef1a2 cux2 cep290 slc1a2 cntnap2 camk2a*). Analysis of the ASD-related genes in this database and others would provide an in-depth comprehension of ASD causes.

Acknowledgments

I want to thank Dr. Kif Liakath-Ali for his supervision and mentorship in this project. Without his help and support, this research would not be possible. I would also like to thank the reviewers for their helpful suggestions.

References

1. Levy, S. E.; Mandell, D. S.; Schultz, R. T. Autism. *Lancet* **2009**, *374* (9701), 1627–1638.
2. Whitaker, L. C. *Schizophrenic Disorders: Sense and Nonsense in Conceptualization, Assessment, and Treatment*; Springer, Boston, MA, 1992.
3. Evans, B. How Autism Became Autism: The Radical Transformation of a Central Concept of Child Development in Britain. *Hist. Human Sci*. **2013**, *26* (3), 3–31.
4. for Disease Control, C.; Prevention; Others. Autism and Developmental Disabilities Monitoring (ADDM) Networ k. *Retrieved November* **2012**, *3*, 2012.
5. Sponheim, E.; Skjeldal, O. Autism and Related Disorders: Epidemiological Findings in a Norwegian Study Using IC D-10 Diagnostic Criteria. *J. Autism Dev. Disord*. **1998**, *28* (3), 217–227.
6. Kadesjö, B.; Gillberg, C.; Hagberg, B. Brief Report: Autism and Asperger Syndrome in Seven-Year-Old Children: A To tal Population Study. *J. Autism Dev. Disord.* **1999**, *29* (4), 32 7–331.
7. Minshew, N. J.; Williams, D. L. The New Neurobiology of Autism: Cortex, Connectivity, and Neuronal Organization. *Arch. Neurol*. **2007**, *64* (7), 945–950..
8. DiCicco-Bloom, E.; Lord, C.; Zwaigenbaum, L.; Courchesn e, E.; Dager, S. R.; Schmitz, C.; Schultz, R. T.; Crawley, J.; Young, L. J. The Developmental Neurobiolo gy of Autism Spectrum Disorder. *J. Neurosci*. **2006**, *26* (26), 6897–6906.
9. Volkmar, F. R.; Pauls, D. Autism. *Lancet* **2003**, *362* (9390), 1133–1141.
10. Banerjee-Basu, S.; Packer, A. SFARI Gene: An Evolving Database for the Autism Research Community. *Dis. Model. Mech*. ***2010***, *3* (3-4), 133–135.
11. Zhou, Y.; Zhou, B.; Pache, L.; Chang, M.; Khodabakhshi, A. H.; Tanaseichuk, O.; Benner, C.; Chanda, S. K. Metas cape Provides a Biologist-Oriented Resource for the Anal ysis of Systems-Level Datasets. *Nat. Commun*. **2019**, *10* (1) , 1523.
12. Sherry, S. T.; Ward, M. H.; Kholodov, M.; Baker, J.; Phan, L.; Smigielski, E. M.; Sirotkin, K. dbSNP: The NCBI Data base of Genetic Variation. *Nucleic Acids Res*. **2001**, *29* (1), 308–311.
13. Piñero, J.; Bravo, À.; Queralt-Rosinach, N.; Gutiérrez-Sac ez-Sacristán, A.; Deu-Pons, J.; Centeno, E.; García-García, J.; Sanz, F.; Furlong, L. I. DisGeNET: A Comprehensive Platform Integrating Information on Human Disease-As sociated Genes and Variants. *Nucleic Acids Res*. **2016**, *45* (D1), D833–D839.
14. Szklarczyk, D.; Gable, A. L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N. T.; Mor ris, J. H.; Bork, P.; Jensen, L. J.; Mering, C. von. STRING v11: Protein–protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-W

ide Experimental Datasets. *Nucleic Acids Res*. 2018, 47 (D1), D607–D613.
15. Oughtred, R.; Stark, C.; Breitkreutz, B.-J.; Rust, J.; Boucher, L.; Chang, C.; Kolas, N.; O'Donnell, L.; Leung, G.: McAdam, R.; Zhang, F.; Dolma, S.; Willems, A.; Coulombe-Huntington, J.; Chatr-aryamontri, A.; Dolinski, K.; Tyers, M. The BioGRID Interaction Database: 2019 Update. *Nucleic Acids Res*. **2018**, *47* (D1), D529–D541.
16. Türei, D.; Korcsmáros, T.; Saez-Rodriguez, J. OmniPath: Guidelines and Gateway for Literature-Curated Signaling Pathway Resources. *Nat. Methods* **2016**, *13* (12), 966–967.
17. Li, T.; Wernersson, R.; Hansen, R. B.; Horn, H.; Mercer, J; Slodkowicz, G.; Workman, C. T.; Rigina, O.; Rapacki, K.; Stærfeldt, H. H.; Brunak, S.; Jensen, T. S.; Lage, K. A Scored Human Protein-Protein Interaction Network to Catalyze Genomic Interpretation. *Nat. Methods* **2017**, *14* (1), 61–64.
18. Arpi, M. N.; Simpson, T. I. SFARI Genes and Where to Find Them; Modelling Autism Spectrum Disorder Specific Gene Expression Dysregulation with RNA-Seq Data. *Scientific Reports* **2022**, *12* (1).
19. Bacchelli, E.; Cameli, C.; Viggiano, M.; Igliozzi, R.; Mancini, A.; Tancredi, R.; Battaglia, A.; Maestrini, E. An Integrated Analysis of Rare CNV and Exome Variation in Autism Spectrum Disorder Using the Infinium Psycharray. *Scientific Reports* **2020**, *10*
20. Sunkin, S. M.; Ng, L.; Lau, C.; Dolbeare, T.; Gilbert, T. L.; Thompson, C. L.; Hawrylycz, M.; Dang, C. Allen Brain Atlas: An Integrated Spatio-Temporal Portal for Exploring the Central Nervous System. *Nucleic Acids Res*. **2012**, *41* (D1), D996–D1008.
21. Maglott, D.; Ostell, J.; Pruitt, K. D.; Tatusova, T. Entrez Gene: Gene-Centered Information at NCBI. *Nucleic Acids Res*. **2011**, *39* (Database issue), D52–D57.
22. Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; Harris, M. A.; Hill, D. P.; Issel-Tarver, L.; Kasarskis, A.; Lewis, S.; Matese, J. C.; Richardson, J. E.; Ringwald, M.; Rubin, G. M.; Sherlock, G. Gene Ontology: Tool for the Unification of Biology. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25* (1), 25–29.
23. Thul, P. J.; Lindskog, C. The Human Protein Atlas: A Spatial Map of the Human Proteome. *Protein Sci*. **2018**, *27* (1), 233–244.
24. Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: A Comprehensive Resource for in Silico Drug Discovery and Exploration. *Nucleic Acids Res*. **2006**, *34* (Database issue), D668–D672.
25. Liu, X.; Yu, X.; Zack, D. J.; Zhu, H.; Qian, J. TiGER: A Database for Tissue-Specific Gene Expression and Regulation. *BMC Bioinformatics* **2008**, *9*, 271.

Author

Daria Lunina, the author of this research, is a senior in the IB program in the European gymnasium. She is interested in neuroscience and wants to investigate behavioral disorders further in her career. She also wants to major in neuroscience for her bachelor's degree.

Supplementary materials

Table 1: List of SFARI human syndromic genes used for the following research.

status	gene-symbol	gene-name	ensembl-id	chromosome	genetic-category	gene-score	syndromic	number-of-reports
9	ACTB	actin beta	ENSG00000075624	7	Rare Single Gene Mutation, Syndromic	1	1	4
9	ACTL6B	actin like 6B	ENSG00000077080	7	Rare Single Gene Mutation, Syndromic		1	9
9	ACY1	aminoacylase 1	ENSG00000243989	3	Rare Single Gene Mutation, Syndromic		1	9
9	AHI1	Abelson helper integration site 1	ENSG00000135541	6	Rare Single Gene Mutation, Syndromic, Genetic Association		1	19
9	ALDH1A3	aldehyde dehydrogenase 1 family member A3	ENSG00000184254	15	Rare Single Gene Mutation, Syndromic		1	5
9	ANKS1B	ankyrin repeat and sterile alpha motif domain containing 1B	ENSG00000185046	12	Rare Single Gene Mutation, Syndromic	3	1	4
9	AP1S2	adaptor related protein complex 1 sigma 2 subunit	ENSG00000182287	X	Rare Single Gene Mutation, Syndromic		1	6
9	ARID2	AT-rich interaction domain 2	ENSG00000189079	12	Rare Single Gene Mutation, Syndromic	3	1	9
9	ATP1A1	ATPase Na+/K+ transporting subunit alpha 1	ENSG00000163399	1	Rare Single Gene Mutation, Syndromic	3	1	5
9	ATP1A3	ATPase Na+/K+ transporting subunit alpha 3	ENSG00000105409	19	Rare Single Gene Mutation, Syndromic, Functional	3	1	15

DOI: 10.36838/v5i4.23

9	BCORL1	BCL6 corepressor like 1	ENSG00000085185	X	Rare Single Gene Mutation, Syndromic		1	4
9	ALG6	ALG6, alpha-1,3-glucosyltransferase	ENSG00000088035	1	Syndromic		1	1
9	BRWD3	bromodomain and WD repeat domain containing 3	ENSG00000165288	X	Syndromic		1	3
9	C12orf57	Chromosome 12 open reading frame 57	ENSG00000111678	12	Rare Single Gene Mutation, Syndromic		1	11
9	CACNA1A	Calcium channel, voltage-dependent, P/Q type, alpha 1A subunit	ENSG00000141837	19	Rare Single Gene Mutation, Genetic Association	1	1	19
9	CAMK2A	calcium/calmodulin dependent protein kinase II alpha	ENSG00000070808	5	Rare Single Gene Mutation, Syndromic, Genetic Association, Functional	3	1	8
9	CAMK2B	calcium/calmodulin dependent protein kinase II beta	ENSG00000058404	7	Rare Single Gene Mutation, Syndromic		1	6
9	CCNK	cyclin K	ENSG00000090061	14	Rare Single Gene Mutation, Syndromic		1	3
9	CDK13	cyclin dependent kinase 13	ENSG00000065883	7	Rare Single Gene Mutation, Syndromic		1	12
9	CDK8	cyclin dependent kinase 8	ENSG00000132964	13	Syndromic		1	2
9	CEP290	Centrosomal protein 290kDa	ENSG00000198707	12	Rare Single Gene Mutation, Syndromic	3	1	8
9	CDK19	cyclin dependent kinase 19	ENSG00000155111	6	Rare Single Gene Mutation, Syndromic	3	1	5
9	CHD1	chromodomain helicase DNA binding protein 1	ENSG00000153922	5	Rare Single Gene Mutation	3	1	4
9	CHKB	Choline kinase beta	ENSG00000100288	22	Rare Single Gene Mutation, Syndromic		1	5
9	CLCN4	chloride voltage-gated channel 4	ENSG00000073464	X	Rare Single Gene Mutation, Syndromic	3	1	5
9	CNKSR2	connector enhancer of kinase suppressor of Ras 2	ENSG00000149970	X	Rare Single Gene Mutation, Syndromic	2	1	7
9	CNTNAP2	contactin associated protein-like 2	ENSG00000174469	7	Rare Single Gene Mutation, Syndromic, Genetic Association	2	1	69
9	CTNNA2	catenin alpha 2	ENSG00000066032	2	Syndromic		1	1
9	CUX2	cut like homeobox 2	ENSG00000111249	12	Rare Single Gene Mutation, Syndromic	3	1	8
9	CYP27A1	cytochrome P450 family 27 subfamily A member 1	ENSG00000135929	2	Rare Single Gene Mutation, Syndromic		1	4
9	DEPDC5	DEP domain containing 5	ENSG00000100150	22	Rare Single Gene Mutation, Syndromic		1	13
9	DHX30	DExH-box helicase 30	ENSG00000132153	3	Rare Single Gene Mutation, Syndromic		1	3
9	DLL1	delta like canonical Notch ligand 1	ENSG00000198719	6	Rare Single Gene Mutation, Syndromic	3	1	2
9	DMD	dystrophin (muscular dystrophy, Duchenne and Becker types)	ENSG00000198947	X	Rare Single Gene Mutation, Syndromic, Genetic Association		1	38
9	DOLK	dolichol kinase	ENSG00000175283	9	Syndromic		1	2
9	EEF1A2	Eukaryotic translation elongation factor 1 alpha 2	ENSG00000101210	20	Rare Single Gene Mutation, Syndromic		1	13
9	FBXO11	F-box protein 11	ENSG00000138081	2	Rare Single Gene Mutation, Syndromic	3	1	6
9	FRMPD4	FERM and PDZ domain containing 4	ENSG00000169933	X	Rare Single Gene Mutation, Syndromic		1	6
9	GABBR2	gamma-aminobutyric acid type B receptor subunit 2	ENSG00000136928	9	Rare Single Gene Mutation, Syndromic, Functional	3	1	12

9	GABRA3	Gamma-aminobutyric acid (GABA) A receptor, alpha 3	ENSG00000011677	X	Rare Single Gene Mutation		1	3
9	GATM	Glycine amidinotransferase (L-arginine:glycine amidinotransferase)	ENSG00000171766	15	Syndromic		1	2
9	FBRSL1	fibrosin like 1	ENSG00000112787	12	Syndromic		1	1
9	FGF13	fibroblast growth factor 13	ENSG00000129682	X	Syndromic	3	1	1
9	GALNT2	polypeptide N-acetylgalactosaminyltransferase 2	ENSG00000143641	1	Syndromic		1	2
9	HCFC1	host cell factor C1	ENSG00000172534	X	Rare Single Gene Mutation, Syndromic		1	9
9	HCN1	Hyperpolarization activated cyclic nucleotide-gated potassium channel 1	ENSG00000164588	5	Rare Single Gene Mutation, Genetic Association		1	8
9	HDAC4	histone deacetylase 4	ENSG00000068024	2	Rare Single Gene Mutation, Syndromic, Genetic Association	3	1	16
9	HDAC8	histone deacetylase 8	ENSG00000147099	X	Rare Single Gene Mutation, Syndromic		1	8
9	HEPACAM	hepatic and glial cell adhesion molecule	ENSG00000165478	11	Rare Single Gene Mutation, Syndromic		1	10
9	HERC2	HECT and RLD domain containing E3 ubiquitin protein ligase 2	ENSG00000128731	15	Rare Single Gene Mutation		1	8
9	HOXA1	homeobox A1	ENSG00000105991	7	Rare Single Gene Mutation, Syndromic, Genetic Association		1	15
9	HUWE1	HECT, UBA and WWE domain containing 1, E3 ubiquitin protein ligase	ENSG00000086758	X	Rare Single Gene Mutation, Syndromic		1	19
9	INTS1	integrator complex subunit 1	ENSG00000164880	7	Rare Single Gene Mutation, Syndromic		1	4
9	KAT6A	K(lysine) acetyltransferase 6A	ENSG00000083168	8	Rare Single Gene Mutation, Syndromic	2	1	15
9	KIF5C	Kinesin family member 5C	ENSG00000168280	2	Rare Single Gene Mutation, Syndromic	3	1	10
9	KPTN	kaptin, actin binding protein	ENSG00000118162	19	Rare Single Gene Mutation, Syndromic		1	6
9	LNPK	lunapark, ER junction formation factor	ENSG00000144320	2	Syndromic		1	1
9	MED12L	mediator complex subunit 12L	ENSG00000144893	3	Rare Single Gene Mutation, Syndromic	3	1	4
9	MSL3	MSL complex subunit 3	ENSG00000005302	X	Rare Single Gene Mutation, Syndromic	2	1	6
9	MTOR	Mechanistic target of rapamycin (serine/threonine kinase)	ENSG00000198793	1	Rare Single Gene Mutation, Syndromic, Functional	2	1	18
9	NFIB	nuclear factor I B	ENSG00000147862	9	Rare Single Gene Mutation, Syndromic	3	1	5
9	NFIX	nuclear factor I/X (CCAAT-binding transcription factor)	ENSG00000008441	19	Rare Single Gene Mutation, Syndromic		1	11
9	NOVA2	NOVA alternative splicing regulator 2	ENSG00000104967	19	Syndromic		1	1
9	NR2F1	nuclear receptor subfamily 2 group F member 1	ENSG00000175745	5	Rare Single Gene Mutation, Syndromic, Functional	3	1	19
9	NTNG1	netrin G1	ENSG00000162631	1	Rare Single Gene Mutation, Syndromic, Genetic Association	3	1	8
9	NTNG2	netrin G2	ENSG00000196358	9	Syndromic		1	3
9	NTRK2	neurotrophic receptor tyrosine kinase 2	ENSG00000148053	9	Syndromic		1	3
9	OCRL	oculocerebrorenal syndrome of Lowe	ENSG00000122126	X	Rare Single Gene Mutation, Syndromic		1	7
9	PACS2	phosphofurin acidic cluster sorting protein 2	ENSG00000179364	14	Rare Single Gene Mutation, Syndromic		1	6

DOI: 10.36838/v5i4.23

					, Syndromic, Functional			
9	PCCA	propionyl-CoA carboxylase alpha subunit	ENSG00000175198	13	Rare Single Gene Mutation, Syndromic		1	10
9	PCCB	propionyl-CoA carboxylase beta subunit	ENSG00000114054	3	Rare Single Gene Mutation, Syndromic	1	1	9
9	NSD2	nuclear receptor binding SET domain protein 2	ENSG00000109685	4	Rare Single Gene Mutation, Syndromic	2	1	9
9	PHF8	PHD finger protein 8	ENSG00000172943	X	Rare Single Gene Mutation, Syndromic		1	10
9	PIK3R2	phosphoinositide-3-kinase regulatory subunit 2	ENSG00000105647	19	Rare Single Gene Mutation, Syndromic		1	5
9	POU3F3	POU class 3 homeobox 3	ENSG00000198914	2	Syndromic		1	1
9	PPM1D	protein phosphatase, Mg2+/Mn2+ dependent 1D	ENSG00000170836	17	Rare Single Gene Mutation, Syndromic	3	1	6
9	PPP2CA	protein phosphatase 2 catalytic subunit alpha	ENSG00000113575	5	Syndromic		1	1
9	PRKD1	Protein kinase D1	ENSG00000184304	14	Rare Single Gene Mutation, Syndromic		1	4
9	PRODH	Proline dehydrogenase (oxidase) 1	ENSG00000100033	22	Syndromic, Genetic Association	2	1	6
9	POLR3A	RNA polymerase III subunit A	ENSG00000148606	10	Rare Single Gene Mutation, Syndromic	3	1	8
9	RAC1	Rac family small GTPase 1	ENSG00000136238	7	Syndromic, Functional		1	5
9	RAD21	RAD21cohesin complex component	ENSG00000164754	8	Syndromic		1	4
9	RALA	RAS like proto-oncogene A	ENSG00000006451	7	Syndromic		1	2
9	RHEB	Ras homolog, mTORC1 binding	ENSG00000106615	7	Syndromic		1	1
9	RLIM	Ring finger protein, LIM domain interacting	ENSG00000131263	X	Rare Single Gene Mutation, Syndromic		1	5
9	RNF135	Ring finger protein 135	ENSG00000181481	17	Rare Single Gene Mutation,	3	1	4
					Syndromic, Genetic Association			
9	RORA	RAR-related orphan receptor A	ENSG00000069667	15	Rare Single Gene Mutation, Syndromic, Genetic Association, Functional		1	18
9	RFX4	regulatory factor X4	ENSG00000111783	12	Rare Single Gene Mutation, Syndromic	3	1	2
9	RFX7	regulatory factor X7	ENSG00000181827	15	Rare Single Gene Mutation, Syndromic	3	1	3
9	RIMS2	regulating synaptic membrane exocytosis 2	ENSG00000176406	8	Rare Single Gene Mutation, Syndromic, Genetic Association	3	1	3
9	RPS6KA3	Ribosomal protein S6 kinase, 90kDa, polypeptide 3	ENSG00000177189	X	Rare Single Gene Mutation, Syndromic	3	1	9
9	RSRC1	arginine and serine rich coiled-coil 1	ENSG00000174891	3	Syndromic, Genetic Association		1	7
9	SATB2	SATB homeobox 2	ENSG00000119042	2	Rare Single Gene Mutation, Syndromic, Genetic Association	3	1	33
9	SETD1B	SET domain containing 1B	ENSG00000139718	12	Rare Single Gene Mutation, Syndromic	3	1	12
9	SGSH	N-sulfoglucosamine sulfohydrolase	ENSG00000181523	17	Rare Single Gene Mutation, Syndromic		1	5
9	SIK1	Salt-inducible kinase 1	ENSG00000142178	21	Rare Single Gene Mutation, Syndromic		1	5

9	SLC1A2	Solute carrier family 1 (glial high affinity glutamate transporter), member 2	ENSG00000110436	11	Rare Single Gene Mutation, Genetic Association, Functional		1	6
9	SLC45A1	solute carrier family 45 member 1	ENSG00000162426	1	Rare Single Gene Mutation, Syndromic		1	2
9	SMARCA2	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2	ENSG00000080503	9	Rare Single Gene Mutation, Syndromic, Genetic Association	1	1	18
9	SMC1A	structural maintenance of chromosomes 1A	ENSG00000072501	X	Rare Single Gene Mutation, Syndromic		1	8
9	SMC3	structural maintenance of chromosomes 3	ENSG00000108055	10	Rare Single Gene Mutation, Syndromic	3	1	9
9	SNX14	Sorting nexin 14	ENSG00000135317	6	Rare Single Gene Mutation, Syndromic		1	6
9	STAG1	stromal antigen 1	ENSG00000118007	3	Rare Single Gene Mutation, Syndromic		1	4
9	SOX6	SRY-box transcription factor 6	ENSG00000110693	11	Syndromic		1	1
9	SYNE1	spectrin repeat containing, nuclear envelope 1	ENSG00000131018	6	Rare Single Gene Mutation, Genetic Association	3	1	17
9	SYT1	synaptotagmin 1	ENSG00000067715	12	Rare Single Gene Mutation, Syndromic		1	3
9	TAF1	TATA-box binding protein associated factor 1	ENSG00000147133	X	Rare Single Gene Mutation, Syndromic		1	10
9	TBC1D23	TBC1 domain family member 23	ENSG00000036054	3	Syndromic		1	2
9	TBX1	T-box 1	ENSG00000184058	22	Syndromic, Functional		1	5
9	SUPT16H	SPT16 homolog, facilitates chromatin remodeling subunit	ENSG00000092201	14	Rare Single Gene Mutation	2	1	6
9	TM4SF20	Transmembrane 4 L six family member 20	ENSG00000168955	2	Rare Single Gene Mutation		1	1
9	TET3	tet methylcytosine dioxygenase 3	ENSG00000187605	2	Rare Single Gene Mutation, Syndromic		1	2
9	TRAPPC6B	trafficking protein particle complex 6B	ENSG00000182400	14	Rare Single Gene Mutation, Syndromic		1	5
9	TRRAP	transformation/transcription domain associated protein	ENSG00000196367	7	Rare Single Gene Mutation, Syndromic	3	1	9
9	TTI2	TELO2 interacting protein 2	ENSG00000129696	8	Rare Single Gene Mutation		1	2
9	TTN	titin	ENSG00000155657	2	Rare Single Gene Mutation, Syndromic	3	1	23
9	UNC13A	unc-13 homolog A	ENSG00000130477	19	Rare Single Gene Mutation, Syndromic	3	1	8
9	USP7	Ubiquitin specific peptidase 7 (herpes virus-associated)	ENSG00000187555	16	Rare Single Gene Mutation, Syndromic	2	1	7
9	USP9X	ubiquitin specific peptidase 9 X-linked	ENSG00000124486	X	Rare Single Gene Mutation, Syndromic, Functional	1	1	11
9	VAMP2	vesicle associated membrane protein 2	ENSG00000220205	17	Rare Single Gene Mutation, Syndromic		1	3
9	WASF1	WAS protein family member 1	ENSG00000112290	6	Syndromic		1	1
9	WDR26	WD repeat domain 26	ENSG00000162923	1	Rare Single Gene Mutation, Syndromic		1	6
9	XPC	xeroderma pigmentosum, complementation group C	ENSG00000154767	3	Rare Single Gene Mutation, Syndromic		1	11
9	YY1	YY1transcription factor	ENSG00000100811	14	Rare Single Gene Mutation, Syndromic, Functional		1	5

DOI: 10.36838/v5i4.23

9	ZSWIM 6	zinc finger SWIM-type containing 6	ENSG00000130449	5	Syndromic, Genetic Association		1	2
9	YWHAG	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein gamma	ENSG00000170027	7	Rare Single Gene Mutation, Syndromic	3	1	7
9	ZMIZ1	zinc finger MIZ-type containing 1	ENSG00000108175	10	Rare Single Gene Mutation, Syndromic	2	1	4
9	ZMYM2	zinc finger MYM-type containing 2	ENSG00000121741	13	Rare Single Gene Mutation, Syndromic	2	1	6

Table 2: Characteristics used for custom analysis.

Gene ID
Type of Gene
Description
Gene Summary
Biological Progress (GO)[22]
Molecular Function (GO)
Protein Function (Protein Atlas)[23]
Secreted (UniProt)
Developmental disorder (DDG2P)
Drug (DrugBank)[24]
Disease and gene associations (DisGeNET)
Pubmed Count
Tissue Specificity (TiGER)[25]
GO:0060322 head development
WP4312 Rett syndrome causing genes
GO:0048812 neuron projection morphogenesis
GO:0016569 covalent chromatin modification
CORUM:164 Cohesin-SA1 complex
GO:0042692 muscle cell differentiation
GO:0090596 sensory organ morphogenesis
GO:0120035 regulation of plasma membrane
R-HSA-5663202 Diseases of signal transduction
M261 PID P53 REGULATION PATHWAY
R-HSA-112315 Transmission across Chemical S
M164 PID ERBB1 DOWNSTREAM PATHWAY
hsa04919 thyroid hormone signaling path
GO:0006338 chromatin remodeling
GO:0071496 cellular response to external
WP4539 Synaptic signaling pathways as