2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

A PheWAS Model of Autism Spectrum Disorder

John Matta, Daniel Dobrino, Swade Howard, Dacosta Yeboah, Jonathan Kopel,
Yasser EL-Manzalawy, and Tayo Obafemi-Ajayi

Abstract— **Children with Autism Spectrum Disorder (ASD) exhibit a wide diversity in type, number, and severity of social deficits as well as communicative and cognitive difficulties. It is a challenge to categorize the phenotypes of a particular ASD patient with their unique genetic variants. There is a need for a better understanding of the connections between genotype information and the phenotypes to sort out the heterogeneity of ASD. In this study, single nucleotide polymorphism (SNP) and phenotype data obtained from a simplex ASD sample are combined using a PheWAS-inspired approach to construct a phenotype-phenotype network. The network is clustered, yielding groups of etiologically related phenotypes. These clusters are analyzed to identify relevant genes associated with each set of phenotypes. The results identified multiple discriminant SNPs associated with varied phenotype clusters such as ASD aberrant behavior (self-injury, compulsiveness and hyperactivity), as well as IQ and language skills. Overall, these SNPs were linked to 22 significant genes. An extensive literature search revealed that eight of these are known to have strong evidence of association with ASD. The others have been linked to related disorders such as mental conditions, cognition, and social functioning.**

Clinical relevance— **This study further informs on connections between certain groups of ASD phenotypes and their unique genetic variants. Such insight regarding the heterogeneity of ASD would support clinicians to advance more tailored interventions and improve outcomes for ASD patients.**

I. Introduction

Some specific susceptibility genes have been identified in association with ASD [1]. However, detection of the etilogical basis of subgroups [2] that have more clearly defined characteristics still remains a challenge. In general, ASD is broken down into social, behavior, and communication language related deficits. There is need for clinically relevant subgrouping to stratify ASD patients across the spectrum based on severity and types of symptoms. These subgroups could then be linked with genetic variants detected in these patients to provide a meaningful association between the genotypes and phenotypes observed in ASD patients. Given the increase in ASD prevalence and the corresponding increasing associated economic burden, there is a need for autonomous machine learning models to provide a better understanding of the connections between genotype information and varied phenotypes associated with ASD. This information would provide clinicians greater insight into the pathophysiology and potential pharmacological or social interventions for ASD patients based on genotype and phenotype correlations. The study used a well-known dataset to explore subgroups of phenotypes driven by etiological characteristics.

The phenome-wide association study (PheWAS) has contributed substantially to the effort in discovering etiological links between genes and diseases. PheWAS studies can uncover associations between genetic variants, as well as pleiotropic relationships between one variant and several phenotypes [3]. Pleiotropy implies a single gene influencing two or more distinct phenotypic traits. Thus, these studies can provide a more complete understanding of the complex relationships among the genetic architecture and functions of biological systems [4]. PheWAS employs regression techniques on genetic information (specifically, single nucleotide polymorphisms (SNPs)) of a given sample population of probands to derive an association between observed phenotypes and SNPs. The traditional PheWAS output yields a plot of the statistical significance power of association of multiple diseases (for each one) to a single SNP. In the context of ASD, there is only one disorder. However, given its heterogeneity, it can be viewed as an aggregate disorder with multiple subgroups within the spectrum. This provides a basis for deriving a PheWAS model to identify novel ASD genetic and cross-phenotype associations.

The PheWAS-inspired model employed in this study identifies genotype (quantified by SNPs) associations among ASD phenotypes. Constructing a phenotype-phenotype network is done by linking phenotypes that have common associated SNPs. The resulting phenotype-phenotype network is a useful structure that can be clustered to detect commonalities among traits. Prior studies on ASD have focused on construction of networks of ASD subgroup patients [5]. A genome-wide association analysis of the Simons Simplex Collection (SSC) data found that "reducing phenotypic heterogeneity has at most a modest impact on genetic homogeneity" [6]. However, a classifier using SNPs was able to predict ASD among an SSC sample with over 70% accuracy [7].

Clustering has been used successfully with phenotype information [8]–[11] and also genetic information [12]. Specifically, clustering analysis allows for groupings of similar objects, such as phenotypic information, to provide insight into potential correlations or associations between two groups. In [13], the clustering is focused on linking genotype combinations to ASD phenotype subgroups. In contrast to prior

John Matta, Daniel Dobrino, and Swade Howard are with the Dept. of Computer Science, Southern Illinois University Edwardsville, Edwardsville, IL. Corresponding email:jmatta@siue.edu

Dacosta Yeboah (Dept. of Computer Science) and Tayo Obafemi-Ajayi (Engineering Program) are with Missouri State University, Springfield, MO.

Jonathan Kopel is with Texas Tech University Health Sciences Center School of Medicine, Lubbock, TX.

Yasser EL-Manzalawy is with Dept. of Translational Data Science and Informatics, Geisinger Health System, Danville, PA.

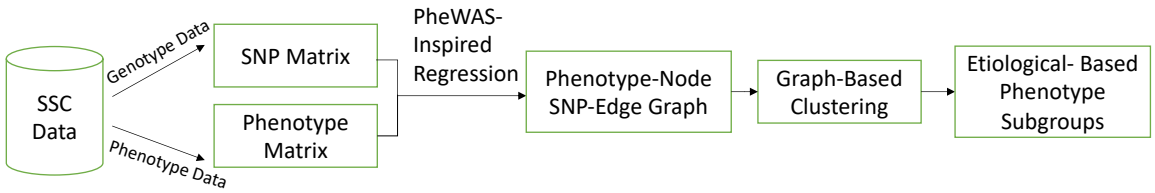

Fig. 1. Overall framework for deriving etiological-based phenotype subgroups.

clustering analysis, this current work explores identifying subgroups of phenotypes driven by etilogical characteristics. The hypothesis is that effective unsupervised learning modalities applied to ASD phenotype and genotype data will yield clinically relevant groupings of ASD phenotypes which could potentially improve diagnoses and assist with prognosis and tailored intervention methods.

II. Methods

A. Data

The phenotype and genetic data are obtained from the SSC simplex collection [14], supported by the Simons Foundation for Autism Research Initiative. Simplex implies only one member of the immediate family has ASD. (This research has been conducted under the guidelines and approval of the Institutional Review Boards at both Southern Illinois University Edwardsville and Missouri State University.) We utilize a subset of 51 phenotypes that spanned ASD specific core-symptom measures, cognitive and adaptive functioning, behavioral problems, and neurological indicators. This included component scores from Autism Diagnostic Interview - Revised (ADI-R), Autism Diagnostic Observation Schedule (ADOS), Repetitive Behavior Scale (RBS), Social Responsiveness Scale (SRS), Aberrant Behavior Checklist (ABC), Child Behavior Checklist (CBCL), IQ, Vineland adaptive measures, dysmorphology examination, and Broader Autism Phenotype Questionnaire (BAPQ) for the parents. Similar to the work done in [13], the selected phenotype subset includes the dsymorphology measure used to distinguish complex autism (dysmorphic and/or microcephalic) [15] from essential autism (non-dysmorphic and not microcephalic). Not all the SSC sites conducted dsymorphology exams, hence this study sample is limited to 560 probands, out of a total of 2759 SSC probands.

The corresponding SSC genotype data are quantified by SNPs. The DNA specimens were genotyped using 3 different Illumina SNP genotyping chip arrays: mv1 (2.44 million), mv3 (18.05 billion) and omni2.5 (12.86 billion). Spencer et al. [16] conducted a genome-wide SNP prioritization analysis which carried out a preliminary GWAS analysis on the entire SNP database by ranking SNPs based on the strength of their primary association (as indicated by increasing values of p-values, the smaller the value, the stronger the association) with ASD as a whole. Using this ranking, the top most significant SNPs were selected based on a parameter threshold for the p-value. In addition, all SNPs that had the same allele representation across all probands were filtered out, as these SNPs contain no discriminant information. Using a p-value threshold of < 0.1 yielded a set of 14,564 SNPs which we will utilize for subsequent analysis.

B. Construction of the Phenotype-Phenotype Network

The overall framework for transforming the data into etilogical-based phenotype subgroups is shown in Fig. 1. The genotype data is encoded into a high dimensional SNP matrix, where entries identify the nature of the variant representing the proband's risk allele to the reference allele. Likewise, the phenotype information for each proband is encoded in an additional matrix. The SNP matrix is size $m \times n$ where m represents the number of probands and n represents the number of SNPs, while the phenotype matrix is size $m \times p$ where p represents the number of phenotypes. A multi-trait mixed regression model (LIMIX [17]) is applied to both SNP and phenotype matrices to emulate the PheWAS model. This yields a matrix of scores quantifying the association between phenotypes and SNPs, as well as a corresponding matrix of p-values, representing the strength of the associations. The matrix of p-values are subsequently filtered based on a given threshold. We utilized the threshold value ($< 1 \times 10^{-3}$) that yielded a sparse but connected network.

A phenotype-phenotype graph was constructed using the shared SNP associations identified by the PheWAS-inspired analysis. The graph was clustered using the Louvain method [18] which has been shown to be successful in clustering phenotype-phenotype networks [3]. Louvain is a widely used clustering method that uses modularity, a well-known metric for assessing the relative goodness of a set of clusters based on network properties. It consists of two repeated phases. In the first phase, nodes are evaluated and merged into clusters with neighboring nodes, based on which merging results in the largest gain in modularity. The first phase ends when no node can be moved to increase modularity. In the second phase, a new graph is created in which the communities found during the first phase are converted to nodes. The phases are repeated on the new graph until no further increase in modularity is possible.

III. Results and Discussion

The network analysis and clustering yielded 6 phenotype-based clusters. The phenotypes for each cluster are enumerated in Table I. To characterize the clusters, the underlying SNPs were compared to understand the shared properties of their phenotype members and the relevance of the important

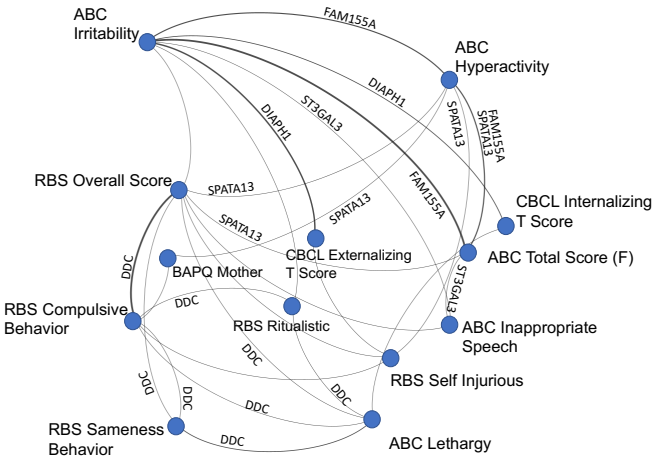

Fig. 2. Cluster 2. Genes link phenotypes concerning aberrant behavior.

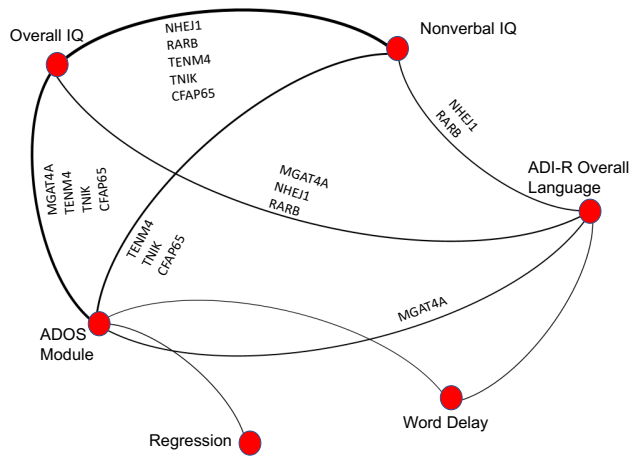

Fig. 3. Cluster 6. SNPs link phenotypes concerning language and IQ.

genes represented. If a cluster contained more than 3 nodes with the same associated SNP, that SNP is noted in the *SNPs* column of Table I. The exact number of occurrences of the SNP is shown in parenthesis. For each SNP, gene information was obtained and is listed in Table I as *Important Genes*.

Cluster 1 yielded phenotypes focused on communication and socialization, such as ADI-R Nonverbal Communication and ADOS Social Affect. There are 4 key genes associated with this cluster, two of which are known to have a strong association with ASD: XCL1 [19] and SEMA3E [20]. Evidence weakly linking LRBA to autism is shown in [21]. CACNG4, which has been linked to sociability in mice [22] and rats, is also relevant to this socialization-focused cluster.

Cluster 2 is visualized in Fig. 2. It is dominated by behavioral phenotypes, mainly relating to aberrant behavior such as irritability, inappropriate speech, compulsiveness and self-injury. It is interesting to note how different the behavioral associations of genes in this cluster are from the previous cluster. As shown in Fig. 2, the DCC gene links the RBS behavior attributes, while the ABC component scores of irritability, hyperactivity and total score are linked by FAM155A. SPATA13 is a bridge in the graph that connects the RBS to ABC scores. DCC has strong evidence for ASD association [23] as well as risky behaviors [24]. SPATA13 also presents strong evidence for ASD association [25] while FAM155A is a candidate gene for impulsive behavior [26], schizophrenia and ADHD [27]. DIAPH1 has strong evidence of association with ASD [28]. ST3GAL3 is linked with ASD, intellectual disability, and ADHD [29].

Cluster 3 consists of the SRS parent components scores, BAPQ average score for father, the ABC stereotype and ADI-R repeated behavior score. The OBSCN gene is related to this cluster. It is known to be linked with ASD [30] and bipolar disorder [31]. KRT26 also shows weak evidence of ASD association [32]. Cluster 4 consists mainly of the SRS teacher components scores along with ADOS repetitive and stereotype behavior. The corresponding genes are SGCD (linked with ASD [33]) and EPHB2 (presents strong evidence of association with ASD [34] and anxiety disorders [35]). Cluster 5 is made up of mainly vineland adaptive skill phenotypes. Of its genes, LOC105369506 is associated with childhood antisocial behavior [36], while ZDHHC7 is linked with ASD [37] and schizophrenia [38].

Cluster 6 is visualized in Fig. 3. This is a tightly constructed cluster with a 4-node clique in which 4 of 6 edges represent at least 3 common important genes. The nonverbal IQ and overall IQ phenotypes have 16 SNPs and 5 important genes in common, making this the heaviest weight edge in the entire network. This cluster also consists of phenotypes relating to language, such as ADI-R overall language and word delay. Important genes show connections to cognition and brain disorders. There is evidence that NHEJ1 is involved in brain development, cognition [39], and microcephaly [40]. MGAT4A is linked with schizophrenia [41], PTSD [42], and bipolar disorder [43]. RARB is a strong ASD evidence candidate gene [44]. TENM4 is associated with ASD, mental disorders, and intelligence [45]. TNIK is linked with ASD [46] and psychiatric disorders [47]. CFAP65 is a candidate gene for epilepsy [48].

IV. Conclusion

This study suggests that a PheWAS-inspired approach using SNPs with phenotype data has potential in identifying genes associated with ASD. The clustering of the ASD phenotype-phenotype graph yielded discriminant genes for each cluster. The extensive literature review identified 8 genes with strong previous evidence of association to ASD, as well as 14 genes with weaker previous evidence of links to ASD and other related conditions.

Overall, the PheWAS-inspired methodology was successful at finding genes associated with ASD: from over 14,000 SNPS, 22 important genes were isolated. The clustering was successful in matching related phenotypes, and in differentiating different aspects of ASD. In particular, cluster 2 effectively isolated aberrant behavior phenotypes, and cluster 6 effectively differentiated cognition and brain-related phenotypes. All clusters included genes with associations both to ASD generally, as well as to individual phenotype

TABLE I

DESCRIPTION OF OUTCOME CLUSTERS BY PHENOTYPES, SNPS AND IMPORTANT GENES.

#	Phenotypes	SNPs (Occurrences)	Important Genes
1	ADI-R scores (nonverbal communication, socialization, abnormality evidence), ADOS scores (communication, communication & social, CSS, reciprocal social, social affect), RBS restricted behavior score, SRS-P motivation score, Dysmorphic	rs8084578(5), rs1927636(4), rs3753938(4), rs16860907(3), rs1933099(3), rs1469064(3), rs2535370(3), rs12505042(3), rs1599167(3), rs1051774(3), rs6534010(3), rs1927621(3)	SEMA3E, LRBA, CACNG4
2	ABC scores (irritability, lethargy, hyperactivity, inappropriate speech, overall total), RBS scores (self injurious, compulsive behavior, ritualistic behavior, sameness behavior, overall total), CBCL externalizing & internalizing T scores, BAPQ (mother)	rs9945776(5), rs7335101(5), rs6707140(4), rs10098925(4), rs9959803(4), rs7326004(3), rs250791(3), rs16996444(3), rs3828139(3)	DCC, SPATA13, FAM155A, DIAPH1, ST3GAL3
3	ABC Stereotype, ADI-R Restricted & Repetitive Behavior, BAPQ Father, SRS-P Awareness, Communication, Mannerisms, Overall T Score	rs719867(6), rs2109217(4), rs35765056(4), rs4653939(3), rs4653942(3)	OBSCN, KRT26
4	ADOS Restricted Repetitive, RBS Stereotyped Behavior, SRS-T Scores (awareness, cognition, communication, mannerisms, motivation, T score)	rs4705000(4), rs9566309(4), rs7982105(3), rs655089(3), rs6687487(3), rs6810871(3), rs716897(3), rs12306561(3), rs2303492(3)	SGCD, EPHB2, RASGRF2, ARFIP2
5	Phrase delay, SRS-P cognition score, Vineland scores (communication, daily living, socialization, overall)	rs1522026(6), rs855017(6), rs1563119(6), rs700874(6), rs1718031(6), rs700873(6), rs855025(6), rs7982105(4), rs11215264(3), rs9566309(3), rs7174994(3), rs7192876(3)	LOC105369506, ZDHHC7
6	ADI-R Overall Language, ADOS Module, Word Delay, Regression, Overall IQ, Nonverbal IQ	rs4871046(3), rs10173578(3), rs16859536(3), rs6709739(3), rs6765578(3), rs948396(3), rs1792136(3), rs9827202(3), rs6781167(3), rs12465007(3), rs10799754(3)	MGAT4A, NHEJ1, RARB, TENM4, TNIK, CFAP65

ADI-R: Autism Diagnostic Interview – Revised; ADOS: Autism Diagnostic Observation Schedule; RBS: Repetitive Behavior Scale; SRS: Social Responsiveness Scale; ABC: Aberrant Behavior Checklist; BAPQ: Broader Autism Phenotype Questionnaire; CBCL: Child Behavior Checklist; CSS: Calibrated Severity score;

manifestations. This PheWAS cluster analysis provided new avenues for further clinical investigation into biological targets involved in ASD as well as developing targeted therapies directed at these gene targets. Larger sample sizes and broader genetic screening is needed to validate the associations shown in this study.

REFERENCES

[1] B. S. Abrahams and D. H. Geschwind, "Advances in autism genetics: on the threshold of a new neurobiology," *Nature reviews genetics*, vol. 9, no. 5, pp. 341–355, 2008.

[2] J. Miles, "Autism subgroups from a medical genetics perspective," *Autism Spectrum Disorders*, pp. 705–721, 2011.

[3] A. Verma, L. Bang, J. E. Miller, Y. Zhang, M. T. M. Lee, Y. Zhang, M. Byrska-Bishop, D. J. Carey, M. D. Ritchie, S. A. Pendergrass *et al.*, "Human-disease phenotype map derived from phewas across 38,682 individuals," *The American Journal of Human Genetics*, vol. 104, no. 1, pp. 55–64, 2019.

[4] A. L. Tyler, D. C. Crawford, and S. A. Pendergrass, "The detection and characterization of pleiotropy: discovery, progress, and promise," *Briefings in bioinformatics*, vol. 17, no. 1, pp. 13–22, 2016.

[5] H. Bruining, L. De Sonneville, H. Swaab, M. De Jonge, M. Kas, H. van Engeland, and J. Vorstman, "Dissecting the clinical heterogeneity of autism spectrum disorders through defined genotypes," *PLoS One*, vol. 5, no. 5, p. e10887, 2010.

[6] P. Chaste, L. Klei, S. J. Sanders, V. Hus, M. T. Murtha, J. K. Lowe, A. J. Willsey, D. Moreno-De-Luca, W. Y. Timothy, E. Fombonne *et al.*, "A genome-wide association study of autism using the simons simplex collection: Does reducing phenotypic heterogeneity in autism increase genetic homogeneity?" *Bio. Psy.*, vol. 77, no. 9, pp. 775–784, 2015.

[7] E. Skafidas, R. Testa, D. Zantomio, G. Chana, I. P. Everall, and C. Pantelis, "Predicting the diagnosis of autism spectrum disorder using gene pathway analysis," *Mol. Psych.*, vol. 19, no. 4, 2014.

[8] J. Kramer, L. Boone, T. Clifford, J. Bruce, and J. D. Matta, "Analysis of medical data using community detection on inferred networks," *IEEE Journal of Biomedical and Health Informatics*, 2020.

[9] T. Clifford, J. Bruce, and J. Matta, "Using node-based resilience clustering to predict and analyze medical data," in *2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. IEEE, 2019, pp. 1–8.

[10] J. Matta, J. Zhao, G. Ercal, and T. Obafemi-Ajayi, "Applications of node-based resilience graph theoretic framework to clustering autism spectrum disorders phenotypes," *Applied Network Science*, vol. 3, no. 1, p. 38, 2018.

[11] J. Matta, T. Obafemi-Ajayi, J. Borwey, K. Sinha, D. Wunsch, and G. Ercal, "Node-based resilience measure clustering with applications to noisy and overlapping communities in complex networks," *Applied Sciences*, vol. 8, no. 8, p. 1307, 2018.

[12] J. Dale, J. Matta, S. Howard, G. Ercal, W. Qiu, and T. Obafemi-Ajayi, "Analysis of grapevine gene expression data using node-based resilience clustering," in *2018 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. IEEE, 2018, pp. 1–8.

[13] J. Zhao, T. Nguyen, J. Kopel, P. B. Koob, D. A. Adieroh, and T. Obafemi-Ajayi, "Genotype combinations linked to phenotype subgroups in autism spectrum disorders," in *2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. IEEE, 2019, pp. 1–8.

[14] G. D. Fischbach and C. Lord, "The simons simplex collection: a resource for identification of autism genetic risk factors," *Neuron*, vol. 68, no. 2, pp. 192–195, 2010.

[15] J. H. Miles, "Complex autism spectrum disorders and cutting-edge molecular diagnostic tests," *JAMA*, vol. 314, no. 9, pp. 879–880, 2015.

[16] M. Spencer, N. Takahashi, S. Chakraborty, J. Miles, and C.-R. Shyu, "Heritable genotype contrast mining reveals novel gene associations specific to autism subgroups," *Journal of Biomedical Informatics*, vol. 77, pp. 50–61, 2018.

[17] C. Lippert, F. P. Casale, B. Rakitsch, and O. Stegle, "Limix: genetic analysis of multiple traits," *BioRxiv*, 2014.

[18] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, p. P10008, 2008.

[19] S. Nazeen, N. P. Palmer, B. Berger, and I. S. Kohane, "Integrative analysis of genetic data sets reveals a shared innate immune component in autism spectrum disorder and its co-morbidities," *Genome biology*, vol. 17, no. 1, p. 228, 2016.

[20] C.-Y. Lin, K.-W. Chang, C.-Y. Lin, J.-Y. Wu, H. Coon, P.-H. Huang, H.-N. Ho, S. Akbarian, S. S.-F. Gau, and H.-S. Huang, "Allele-specific expression in a family quartet with autism reveals mono-to-biallelic switch and novel transcriptional processes of autism susceptibility genes," *Scientific reports*, vol. 8, no. 1, pp. 1–15, 2018.

[21] L. Medrihan, A. Rohlmann, R. Fairless, J. Andrae, M. Döring, M. Missler, W. Zhang, and M. W. Kilimann, "Neurobeachin, a protein implicated in membrane protein traffic and autism, is required for the formation and functioning of central synapses," *The Journal of physiology*, vol. 587, no. 21, pp. 5095–5106, 2009.

[22] M. Lackinger, A. Ö. Sungur, R. Daswani, M. Soutschek, S. Bicker, L. Stemmler, T. Wüst, R. Fiore, C. Dieterich, R. K. Schwarting *et al.*, "A placental mammal-specific micro rna cluster acts as a natural brake for sociability in mice," *EMBO reports*, vol. 20, no. 2, p. e46429, 2019.

[23] Y. Li, S. Qiu, W. Zhong, Y. Li, Y. Liu, Y. Cheng, and Y. Liu, "Association between dcc polymorphisms and susceptibility to autism spectrum disorder," *Jrnl of autism and devel. disorders*, 2020.

[24] J. Ye, L. Liu, X. Xu, Y. Wen, P. Li, B. Cheng, S. Cheng, L. Zhang, M. Ma, X. Qi *et al.*, "A genome-wide multiphenotypic association analysis identified candidate genes and gene ontology shared by four common risky behaviors," *Aging*, vol. 12, no. 4, p. 3287, 2020.

[25] S. W. Kong, C. D. Collins, Y. Shimizu-Motohashi, I. A. Holm, M. G. Campbell, I.-H. Lee, S. J. Brewster, E. Hanson, H. K. Harris, K. R. Lowe *et al.*, "Characteristics and predictive value of blood transcriptome signature in males with autism spectrum disorders," *PloS one*, vol. 7, no. 12, p. e49475, 2012.

[26] J. Vevera, M. Zarrei, H. Hartmannová, I. Jedličková, D. Mušálková, A. Přistoupilová, P. Oliveriusová, H. Trešlová, L. Nosková, K. Hodaňová *et al.*, "Rare copy number variation in extremely impulsively violent males," *Genes, Brain and Behavior*, vol. 18, no. 6, p. e12536, 2019.

[27] C.-H. Yang, Y.-D. Lin, and L.-Y. Chuang, "Class balanced multifactor dimensionality reduction to detect gene—gene interactions," *IEEE/ACM trans. on computational biology and bioinformatics*, 2018.

[28] J. A. Vorstman, E. Van Daalen, G. R. Jalali, E. R. Schmidt, R. J. Pasterkamp, M. de Jonge, E. A. Hennekam, E. Janson, W. G. Staal, B. Van Der Zwaag *et al.*, "A double hit implicates diaph3 as an autism risk gene," *Molecular psychiatry*, vol. 16, no. 4, pp. 442–451, 2011.

[29] Y. Zhao, X. Liang, F. Zhu, Y. Wen, J. Xu, J. Yang, M. Ding, B. Cheng, M. Ma, L. Zhang *et al.*, "A large-scale integrative analysis of gwas and common meqtls across whole life course identifies genes, pathways and tissue/cell types for three major psychiatric disorders," *Neuroscience & Biobehavioral Reviews*, vol. 95, pp. 347–352, 2018.

[30] D. R. Krupp, R. A. Barnard, Y. Duffourd, S. A. Evans, R. M. Mulqueen, R. Bernier, J.-B. Rivière, E. Fombonne, and B. J. O'Roak, "Exonic mosaic mutations contribute risk for autism spectrum disorder," *The American Journal of Human Genetics*, vol. 101, no. 3, pp. 369–390, 2017.

[31] A. Fledel-Alon, E. M. Leffler, Y. Guan, M. Stephens, G. Coop, and M. Przeworski, "Variation in human recombination rates and its genetic determinants," *PloS one*, vol. 6, no. 6, p. e20321, 2011.

[32] M. R. Rahman, M. C. Petralia, R. Ciurleo, A. Bramanti, P. Fagone, M. Shahjaman, L. Wu, Y. Sun, B. Turanli, K. Y. Arga *et al.*, "Comprehensive analysis of rna-seq gene expression profiling of brain transcriptomes reveals novel genes, regulators, and pathways in autism spectrum disorder," *Brain sciences*, vol. 10, no. 10, p. 747, 2020.

[33] L. A. Weiss and D. E. Arking, "A genome-wide linkage and association scan reveals novel loci for autism," *Nature*, vol. 461, no. 7265, pp. 802–808, 2009.

[34] S. J. Sanders, M. T. Murtha, A. R. Gupta, J. D. Murdoch, M. J. Raubeson, A. J. Willsey, A. G. Ercan-Sencicek, N. M. DiLullo, N. N. Parikshak, J. L. Stein *et al.*, "De novo mutations revealed by whole-exome sequencing are strongly associated with autism," *Nature*, vol. 485, no. 7397, pp. 237–241, 2012.

[35] X.-R. Wu, Y. Zhang, X.-D. Liu, W.-B. Han, N.-J. Xu, and S. Sun, "Ephb2 mediates social isolation-induced memory forgetting," *Translational Psychiatry*, vol. 10, no. 1, pp. 1–12, 2020.

[36] I. H. Ruisch, A. Dietrich, J. C. Glennon, J. K. Buitelaar, and P. J. Hoekstra, "Interplay between genome-wide implicated genetic variants and environmental factors related to childhood antisocial behavior in the uk alspac cohort," *European archives of psychiatry and clinical neuroscience*, vol. 269, no. 6, pp. 741–752, 2019.

[37] D. J. Hedges, K. L. Hamilton-Nelson, S. J. Sacharow, L. Nations, G. W. Beecham, Z. M. Kozhekbaeva, B. L. Butler, H. N. Cukier, P. L. Whitehead, D. Ma *et al.*, "Evidence of novel fine-scale structural variation at autism spectrum disorder candidate loci," *Molecular autism*, vol. 3, no. 1, p. 2, 2012.

[38] D. Rudd, M. Axelsen, E. A. Epping, N. Andreasen, and T. Wassink, "Childhood-onset schizophrenia case with 2.2 mb deletion at chromosome 3p12. 2–p12. 1 and two large chromosomal abnormalities at 16q22. 3–q24. 3 and xq23–q28," *Clinical Case Reports*, vol. 3, no. 4, p. 201, 2015.

[39] G. Dumas, S. Malesys, and T. Bourgeron, "Systematic detection of brain protein-coding genes under positive selection during primate evolution and their roles in cognition," *Genome Research*, vol. 31, no. 3, pp. 484–496, 2021.

[40] S. Passemard, A. M. Kaindl, and A. Verloes, "Microcephaly," *Handbook of clinical neurology*, vol. 111, pp. 129–141, 2013.

[41] A. R. Sanders, E. I. Drigalenko, J. Duan, W. Moy, J. Freda, H. Göring, and P. Gejman, "Transcriptome sequencing study implicates immune-related genes differentially expressed in schizophrenia: new data and a meta-analysis," *Translational psychiatry*, vol. 7, no. 4, pp. e1093–e1093, 2017.

[42] P. Guardado, A. Olivera, H. L. Rusch, M. Roy, C. Martin, N. Lejbman, H. Lee, and J. M. Gill, "Altered gene expression of the innate immune, neuroendocrine, and nuclear factor-kappa b (nf-κb) systems is associated with posttraumatic stress disorder in military personnel," *Journal of anxiety disorders*, vol. 38, pp. 9–20, 2016.

[43] P. Sklar, S. Ripke, L. J. Scott, O. A. Andreassen, S. Cichon, N. Craddock, H. J. Edenberg, J. I. Nurnberger Jr, M. Rietschel, D. Blackwood *et al.*, "Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near odz4," *Nature genetics*, vol. 43, no. 10, p. 977, 2011.

[44] R. N. Doan, E. T. Lim, S. De Rubeis, C. Betancur, D. J. Cutler, A. G. Chiocchetti, L. M. Overman, A. Soucy, S. Goetze, C. M. Freitag *et al.*, "Recessive gene disruptions in autism spectrum disorder," *Nature genetics*, vol. 51, no. 7, p. 1092, 2019.

[45] O. B. Smeland, S. Bahrami, O. Frei, A. Shadrin, K. O'Connell, J. Savage, K. Watanabe, F. Krull, F. Bettella, N. E. Steen *et al.*, "Genome-wide analysis reveals extensive genetic overlap between schizophrenia, bipolar disorder, and intelligence," *Molecular psychiatry*, vol. 25, no. 4, pp. 844–853, 2020.

[46] M. M. G. Seno, P. Hu, F. G. Gwadry, D. Pinto, C. R. Marshall, G. Casallo, and S. W. Scherer, "Gene and mirna expression profiles in autism spectrum disorders," *Brain res.*, vol. 1380, pp. 85–97, 2011.

[47] Q. Wang, E. Charych, V. Pulito, J. Lee, N. Graziane, R. Crozier, R. Revilla-Sanchez, M. Kelly, A. Dunlop, H. Murdoch *et al.*, "The psychiatric disease risk factors disc1 and tnik interact to regulate synapse composition and function," *Molecular psychiatry*, vol. 16, no. 10, pp. 1006–1023, 2011.

[48] K. Poirier, L. Hubert, G. Viot, M. Rio, P. Billuart, C. Besmond, and T. Bienvenu, "Csnk2b splice site mutations in patients cause intellectual disability with or without myoclonic epilepsy," *Human mutation*, vol. 38, no. 8, pp. 932–941, 2017.