# Gene discoveries in autism are biased towards comorbidity with intellectual disability

Matthew Jensen[1,2], Corrine Smolen[1], and Santhosh Girirajan[1,2,3]

1. Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA 16802, USA
2. Bioinformatics and Genomics Program, Huck Institute of the Life Sciences, Pennsylvania State University, University Park, PA 16802, USA
3. Department of Anthropology, Pennsylvania State University, University Park, PA 16802, USA

**Correspondence to**:

Santhosh Girirajan
205A Life Sciences Building
Pennsylvania State University
University Park, PA 16802
E-mail: sxg47@psu.edu
Phone: 814-865-0674

**Abstract**

Autism typically presents with a highly heterogeneous set of features, including frequent comorbidity with intellectual disability (ID). The overlap between these two phenotypes has confounded the accurate diagnosis and discovery of genetic factors associated with autism. We analyzed genetic variants in 2,290 individuals with autism from the Simons Simplex Collection (SSC) who have either ID or normal cognitive function to determine whether genes associated with autism also contribute towards ID comorbidity. We found that individuals who carried variants in a set of 173 reported autism-associated genes showed decreased IQ ($p=5.49\times10^{-6}$) and increased autism severity ($p=0.013$) compared with individuals without such variants. A subset of autism-associated genes also showed strong evidence for ID comorbidity in published case reports. We also found that individuals with high-functioning autism (IQ>100) had lower frequencies of CNVs ($p=0.065$) and LGD variants ($p=0.021$) compared with individuals who manifested both autism and ID (IQ<70). These data indicated that *de novo* LGD variants conferred a 1.53-fold higher risk ($p=0.035$) towards comorbid ID, while LGD mutations specifically disrupting autism-associated genes conferred a 4.85-fold increased risk ($p=0.011$) for comorbid ID. Furthermore, *de novo* LGD variants in individuals with high-functioning autism were more likely to disrupt genes with little functional relevance towards neurodevelopment, as demonstrated by evidence from pathogenicity metrics, expression patterns in the developing brain, and mouse model phenotypes. Overall, our data suggest that *de novo* pathogenic variants disrupting genes associated with autism contribute towards autism and ID comorbidity, while other genetic factors are likely to be causal for high-functioning autism.

Autism spectrum disorder, which presents in children with social communication difficulties, repetitive behavior, and restricted interests[1], is a highly heterogeneous neurodevelopmental disorder characterized by complex genetic etiology and strong comorbidity with other developmental disorders[2]. For example, approximately 30% of individuals with autism also manifest with intellectual disability (ID)[3], defined[1] by IQ scores <70. The high degree of co-occurrence of autism with ID has been shown to confound accurate diagnosis of autism. In fact, we recently showed that 69% of individuals diagnosed with ID are likely to be recategorized and diagnosed with autism[4]. The diagnostic overlap between autism and ID suggests that *de novo* gene disruptive variants and copy-number variants (CNVs) identified in individuals ascertained for autism in large-scale studies could also be confounded by ID comorbidity. Here, using genetic and phenotypic data from 2,290 individuals with autism from the Simons Simplex Collection (SSC)[5], we show that gene discoveries in autism are biased towards genes that contribute towards both autism and comorbid ID.

We analyzed rare *de novo* likely-gene disruptive (LGD) variants from exome sequencing data[6,7], disease-associated copy-number variants (CNVs) from microarrays[8], and Full-scale IQ and Social Responsiveness Scale (SRS) T-scores for SSC probands that were obtained from the Simons Foundation Autism Research Initiative[5]. As these data were de-identified, they were exempt from IRB review and conformed to the Helsinki Declaration. We first compared the phenotypes of 288 individuals with *de novo* LGD variants and 81 individuals with pathogenic CNVs to 1,921 individuals without such variants obtained from the SSC cohort. Similar to previous autism studies that identified correlations between *de novo* variants and IQ scores[9–12], we found that individuals with *de novo* LGD variants (IQ=77.7, p=0.031, two-tailed Mann-Whitney test) or pathogenic CNVs (IQ=76.3, p=0.002) had a significant decrease in IQ scores compared with individuals without such variants (IQ=82.3) (**Figure 1A**). However, no differences in autism severity, measured using SRS T-scores, were observed between groups of individuals with and without pathogenic variants (p=0.104 for LGD variants and 0.963 for CNVs) (**Figure 1A**). This suggests that pathogenic variants in general contribute to ID independent of autism severity, although this could also be due to an ascertainment bias in the SSC cohort towards individuals with severe autism.

We further identified individuals carrying *de novo* LGD variants in 173 autism-associated genes, defined as genes with recurrent *de novo* variants reported in multiple databases of

sequencing studies (**Table S1**). These genes included tier 1 genes (>2 *de novo* LGD variants) from the Developing Brain Disorders Gene Database[13], genes with >5 non-SSC *de novo* LGD variants from denovo-db[14], and SFARI Gene tiers 1 and 2 (https://gene.sfari.org/). We found that individuals carrying *de novo* LGD variants in autism-associated genes had decreased IQ (n=74, IQ=69.1, $p=5.49\times10^{-6}$, two-tailed Mann-Whitney test) and increased SRS T-scores (SRS=82.4, p=0.013) compared with individuals without LGD variants (n=2,216, IQ=81.9, SRS=79.6), implying that candidate autism genes contribute to both autism and ID phenotypes (**Figure 1B**). To validate this finding, we examined 76 published case reports of affected individuals with pathogenic variants in a subset of 22 autism genes that appeared in all three autism gene databases (**Table 1, Table S2**). For example, recent case studies have identified autism co-occurring with ID in 21 individuals with *de novo SHANK3* variants[15], 19 individuals with *NRXN1* variants[16], and 18 individuals with *TCF20* variants[17]. Overall, 460/497 (92.6%) individuals with autism described in these studies had ID features, emphasizing that variants in these genes contribute to a severe form of autism with comorbid ID **(Table 1)**.

We next compared genetic data from 397 SSC individuals (17.3% of the SSC cohort) with "high-functioning autism", defined as having severe autism and average or above-average IQ scores (SRS>75 and IQ>100), to 562 individuals (24.5%) with both autism and ID (SRS>75 and IQ<70). Individuals with high-functioning autism had a significantly lower (p=0.021, one-tailed Fisher's Exact test) frequency of *de novo* LGD variants (42/397, 10.6%) than individuals with autism and ID (86/562, 15.3%). Similarly, individuals with high-functioning autism were less likely (p=0.065) to carry pathogenic CNVs (9/397, 2.3%) than individuals with autism and ID (24/562, 4.3%). In fact, *de novo* LGD variants conferred a 1.53-fold higher likelihood of manifesting ID among individuals with autism (p=0.035, 95% confidence interval 1.03-2.26), and pathogenic CNVs similarly conferred a 1.92-fold increased risk for co-occurrence of ID among individuals with autism (p=0.099, 95% CI 0.88-4.18). We replicated these observations by analyzing an additional combined cohort of 2,357 individuals from both the SSC and the Autism Sequencing Collection[18]. Here, individuals with autism and ID had a significantly higher rate ($p=3.04\times10^{-6}$, one-tailed Student's t-test) of *de novo* variants in genes intolerant to variation, as measured by probability of Loss-of-function Intolerant (pLI) score >0.9 (70/643, 10.8%), than individuals manifesting autism but not ID (114/1747, 6.65%). We also found that only 3/397 (0.8%) individuals in the SSC cohort with high-functioning autism carried *de novo* LGD variants

in autism-associated genes, including *ANK2, HIVEP3,* and *BAZ2B*. This frequency was not significantly different from the expected frequency of *de novo* variants in the general population (p=0.095, one-tailed Student's T-test), as calculated from gene-specific probabilities of *de novo* nonsense and frameshift variants from a sequence context-dependent model[9]. In contrast, 20/562 (3.6%) individuals with autism comorbid with ID carried *de novo* LGD variants in autism-associated genes, such as *CHD8*, *SCN2A*, and *SYNGAP1*, representing a 19.2-fold enrichment of variants compared with the expected rate in the general population ($p=9.48\times10^{-6}$). Thus, *de novo* LGD variants in autism genes conferred a 4.85-fold increased risk (p=0.011, 95% CI 1.43-16.42) towards comorbid ID in individuals with autism.

We further sought to determine the biological relevance of the 42 genes with *de novo* LGD variants identified in individuals with high-functioning autism, and found that these genes in aggregate had less functional relevance towards neurodevelopment than the reported autism-associated genes. For example, genes with *de novo* LGD variants in individuals with high-functioning autism were less resistant to genetic variation than reported autism-associated genes, as measured by Residual Variation Intolerance Score (RVIS) ($p=4.00\times10^{-4}$, Mann-Whitney two-tailed test) and pLI percentile ($p=9.77\times10^{-7}$) gene metrics[19,20] (**Figure 2A).** In fact, while the RVIS and pLI percentiles of the reported autism genes were clustered below the thresholds for pathogenicity (RVIS <20$^{th}$ percentile and pLI <18$^{th}$ percentile, or raw score >0.9), genes disrupted among individuals with high-functioning autism were evenly distributed across the range of percentiles. Additionally, we tested the enrichment of each gene set for specific expression in brain regions during development, based on expression data derived from the BrainSpan Atlas[21], using the Specific-Expression Analysis (SEA) online tool[22]. While autism genes were enriched for specific expression in the cortex ($p=3.13\times10^{-4}$, Fisher's Exact test with Benjamini-Hochsberg correction) and cerebellum (p=0.020) during early fetal development[22], genes with *de novo* LGD variants in high-functioning autism individuals were not enriched for any specific expression patterns in the developing brain (**Figure 2B**). Furthermore, mouse models of genes identified in individuals with high-functioning autism, whose phenotypic data were obtained from the Mouse Genome Informatics database[24], were significantly less likely to manifest nervous system ($p=4.90\times10^{-3}$, one-tailed Fisher's Exact test with Benjamini-Hochsberg correction) and behavioral/neurological (p=0.037) phenotypes than mouse models of reported autism-associated genes (**Figure 2C**). These findings suggest that genes with *de novo* LGD

variants in individuals with high-functioning autism are less pathogenic in humans and model organisms, and therefore may not necessarily contribute towards the specific high-functioning autism phenotype.

Our data indicate that pathogenic variants such as *de novo* LGD variants and CNVs contribute to autism phenotypes primarily in individuals with comorbid ID, especially if the variants disrupt a gene previously associated with autism. Several themes regarding the study of high-functioning autism have emerged from these findings. *First*, the consistently high degree of comorbidity between autism and ID has led to an ascertainment bias towards individuals who manifest both disorders in large-scale sequencing cohorts, as it is difficult to exclude all individuals with comorbid disorders and still have adequate power to identify recurrent variants. Indeed, more than 80% of the SSC cohort had an IQ score less than 100, and the average IQ of the cohort (81.5) was 18.5 points below the population average. This bias has contributed to the identification of genes and CNV regions related to both autism and ID, as evidenced by the decreased IQ among carriers of variants in these genes as well as a high incidence of comorbid phenotypes reported in published case studies. Large-scale sequencing studies still hold a high value in uncovering shared biological mechanisms that could underlie both disorders[23]. However, understanding the biology of the core autism phenotypes would require concerted efforts to recruit individuals who specifically manifest high-functioning autism without ID.

*Second*, individuals with high-functioning autism are less likely to carry *de novo* LGD variants in candidate autism genes, and *de novo* variants in individuals with high-functioning autism tend to disrupt genes with less functional relevance towards neurodevelopment. These genes likely carry non-recurrent variants that either confer a small effect size towards autism risk on their own, or are not associated at all with neurodevelopment. We therefore propose that multiple genomic factors with varying effect sizes, such as missense variants, common variants, variants in regulatory and non-coding regions, or the combinatorial effects of inherited variants, contribute towards autism phenotypes without ID. For example, Schaaf and colleagues performed targeted sequencing of 21 candidate autism genes in 339 individuals with high-functioning autism[25]. They found that 2% of individuals carried *de novo* missense variants in candidate autism genes, such as *PTEN* and *FOXP2,* suggesting that allelic variants of differing severity within the same gene might contribute to distinct neurodevelopmental trajectories. Interestingly, the same study also found that 7% of individuals with high-functioning autism

carried multiple inherited missense variants in candidate autism genes, potentially contributing to an oligogenic model for high-functioning autism phenotypes. Similarly, common variants have been found to contribute towards increased autism risk in individuals without ID[26,27]. For example, Grove and colleagues recently reported that the heritability attributed to common variants, including those primarily associated with cognitive ability and educational attainment, was three times lower in individuals with autism and ID compared with those without ID[27]. Finally, variants that may not contribute directly towards autism phenotypes themselves, including the *de novo* LGD variants observed in individuals with high-functioning autism, could still be responsible for subtler modification of the severity of autism or ID phenotypes.

Overall, our results emphasize the importance of dissecting phenotypic heterogeneity in family-based sequencing studies of complex diseases, especially those with a high frequency of comorbid disorders. While a larger cohort of individuals recruited specifically for high-functioning autism could identify associations with recurrent genes or different types of variants, these findings should be validated using functional studies to more fully differentiate the genetic causes for high-functioning autism from those for autism with comorbid ID.

## Supplemental data

Supplemental data include two supplemental tables in Excel file format.

## Declaration of Interests

The authors declare that they have no conflicts of interest.

## Acknowledgements

This work was supported by NIH R01-GM121907, SFARI Pilot Grant (#399894) and resources from the Huck Institutes of the Life Sciences to S.G., and NIH T32-GM102057 to M.J. The authors thank Fereydoun Hormozdiari (UC Davis), Lucilla Pizzo (Penn State), and Vijay Kumar (Penn State) for their helpful discussions and comments on the manuscript. We are grateful to all of the families at the participating Simons Simplex Collection (SSC) sites, as well as the principal investigators (A. Beaudet, R. Bernier, J. Constantino, E. Cook, E. Fombonne, D. Geschwind, R. Goin-Kochel, E. Hanson, D. Grice, A. Klin, D. Ledbetter, C. Lord, C. Martin, D. Martin, R. Maxim, J. Miles, O. Ousley, K. Pelphrey, B. Peterson, J. Piggot, C. Saulnier, M. State, W. Stone, J. Sutcliffe, C. Walsh, Z. Warren, E. Wijsman). We appreciate obtaining access to phenotypic data on the Simons Foundation Autism Research Initiative (SFARI) Base. Approved researchers can obtain the SSC data sets described in this study by applying at https://www.base.sfari.org.

## Author Contributions

M.J. and S.G. conceptualized the study, and M.J. and C.S. analyzed the data. M.J. and S.G. wrote the manuscript with input from all authors.

## References

1. American Psychiatric Association (2013). Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association).

2. Vorstman, J.A.S., Parr, J.R., Moreno-De-Luca, D., Anney, R.J.L., Nurnberger Jr, J.I., and Hallmayer, J.F. (2017). Autism genetics: opportunities and challenges for clinical translation. Nat. Rev. Genet. *18*, 362–376.

3. Baio, J., Wiggins, L., Christensen, D.L., Maenner, M.J., Daniels, J., Warren, Z., Kurzius-Spencer, M., Zahorodny, W., Robinson, C., Rosenberg, et al. (2018). Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. MMWR. Surveill. Summ. *67*, 1–23.

4. Polyak, A., Kubina, R.M., and Girirajan, S. (2015). Comorbidity of intellectual disability confounds ascertainment of autism: implications for genetic diagnosis. Am. J. Med. Genet. Part B Neuropsychiatr. Genet. *168*, 600–608.

5. Fischbach, G.D., and Lord, C. (2010). The Simons Simplex Collection: A Resource for Identification of Autism Genetic Risk Factors. Neuron *68*, 192–195.

6. Iossifov, I., O'Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E., et al. (2014). The contribution of de novo coding mutations to autism spectrum disorder. Nature *515*, 216–221.

7. Krumm, N., Turner, T.N., Baker, C., Vives, L., Mohajeri, K., Witherspoon, K., Raja, A., Coe, B.P., Stessman, H.A., He, Z.X., et al. (2015). Excess of rare, inherited truncating mutations in autism. Nat. Genet. *47*, 582–588.

8. Sanders, S.J., He, X., Willsey, A.J., Ercan-Sencicek, A.G., Samocha, K.E., Cicek, A.E., Murtha, M.T., Bal, V.H., Bishop, S.L., Dong, S., et al. (2015). Insights into Autism Spectrum

238 Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron *87*, 1215–1233.
239
240 9. Samocha, K.E., Robinson, E.B., Sanders, S.J., Stevens, C., Sabo, A., McGrath, L.M.,
241 Kosmicki, J.A., Rehnström, K., Mallick, S., Kirby, A., et al. (2014). A framework for the
242 interpretation of de novo mutation in human disease. Nat. Genet. *46*, 944–950.
243
244 10. Robinson, E.B., Samocha, K.E., Kosmicki, J.A., McGrath, L., Neale, B.M., Perlis, R.H., and
245 Daly, M.J. (2014). Autism spectrum disorder severity reflects the average contribution of de
246 novo and familial influences. Proc. Natl. Acad. Sci. *111*, 15161–15165.
247
248 11. Yuen, R.K.C., Merico, D., Bookman, M., Howe, J.L., Thiruvahindrapuram, B., Patel, R. V,
249 Whitney, J., Deflaux, N., Bingham, J., Wang, Z., et al. (2017). Whole genome sequencing
250 resource identifies 18 new candidate genes for autism spectrum disorder. Nat. Neurosci. *20*, 602–
251 611.
252
253 12. Pizzo, L., Jensen, M., Polyak, A., Rosenfeld, J.A., Mannik, K., Krishnan, A., McCready, E.,
254 Pichon, O., Le Caignec, C., Van Dijck, A., et al. (2019). Rare variants in the genetic background
255 modulate cognitive and developmental phenotypes in individuals carrying disease-associated
256 variants. Genet. Med. *21*, 816–825.
257
258 13. Gonzalez-Mantilla, A.J., Moreno-De-Luca, A., Ledbetter, D.H., and Martin, C.L. (2016). A
259 cross-disorder method to identify novel candidate genes for developmental brain disorders.
260 JAMA Psychiatry *73*, 275–283.
261
262 14. Turner, T.N., Yi, Q., Krumm, N., Huddleston, J., Hoekzema, K., F Stessman, H.A., Doebley,
263 A.-L., Bernier, R.A., Nickerson, D.A., and Eichler, E.E. (2017). denovo-db: a compendium of
264 human de novo variants. Nucleic Acids Res. *45*, D804–D811.
265
266 15. Li, Y., Jia, X., Wu, H., Xun, G., Ou, J., Zhang, Q., Li, H., Bai, T., Hu, Z., Zou, X., et al.
267 (2018). Genotype and phenotype correlations for SHANK3 de novo mutations in
268 neurodevelopmental disorders. Am. J. Med. Genet. Part A *176*, 2668–2676.

16. Al Shehhi, M., Forman, E.B., Fitzgerald, J.E., McInerney, V., Krawczyk, J., Shen, S., Betts, D.R., Ardle, L.M., Gorman, K.M., King, M.D., et al. (2019). NRXN1 deletion syndrome; phenotypic and penetrance data from 34 families. Eur. J. Med. Genet. *62*, 204–209.

17. Torti, E., Keren, B., Palmer, E.E., Zhu, Z., Afenjar, A., Anderson, I.J., Andrews, M. V., Atkinson, C., Au, M., Berry, S.A., et al. (2019). Variants in TCF20 in neurodevelopmental disability: description of 27 new patients and review of literature. Genet. Med. ePub ahead of print.

18. Kosmicki, J.A., Samocha, K.E., Howrigan, D.P., Sanders, S.J., Slowikowski, K., Lek, M., Karczewski, K.J., Cutler, D.J., Devlin, B., Roeder, K., et al. (2017). Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. Nat. Genet. *49*, 504–510.

19. Petrovski, S., Wang, Q., Heinzen, E.L., Allen, A.S., and Goldstein, D.B. (2013). Genic Intolerance to Functional Variation and the Interpretation of Personal Genomes. PLoS Genet. *9*, e1003709.

20. Lek, M., Karczewski, K.J., Minikel, E. V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.

21. Miller, J.A., Ding, S.-L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Ebbert, A., Riley, Z.L., Royall, J.J., Aiona, K., et al. (2014). Transcriptional landscape of the prenatal human brain. Nature *508*, 199–206.

22. Dougherty, J.D., Schmidt, E.F., Nakajima, M., and Heintz, N. (2010). Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. Nucleic Acids Res. *38*, 4218–4230.

23. Jensen, M., and Girirajan, S. (2017). Mapping a shared genetic basis for neurodevelopmental disorders. Genome Med. *9*, 109.

24. Smith, C.L., Blake, J.A., Kadin, J.A., Richardson, J.E., Bult, C.J., and Mouse Genome Database Group (2018). Mouse Genome Database (MGD)-2018: knowledgebase for the laboratory mouse. Nucleic Acids Res. *46*, D836–D842.

25. Schaaf, C.P., Sabo, A., Sakai, Y., Crosby, J., Muzny, D., Hawes, A., Lewis, L., Akbar, H., Varghese, R., Boerwinkle, E., et al. (2011). Oligogenic heterozygosity in individuals with high-functioning autism spectrum disorders. Hum. Mol. Genet. *20*, 3366–3375.

26. Weiner, D.J., Wigdor, E.M., Ripke, S., Walters, R.K., Kosmicki, J.A., Grove, J., Samocha, K.E., Goldstein, J.I., Okbay, A., Bybjerg-Grauholm, J., et al. (2017). Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. Nat. Genet. *49*, 978–985.

27. Grove, J., Ripke, S., Als, T.D., Mattheisen, M., Walters, R.K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O.A., Anney, R., et al. (2019). Identification of common genetic risk variants for autism spectrum disorder. Nat. Genet. *51*, 431–444.

## Figure Legends

**Figure 1.** Phenotypic comparison of individuals with autism from the SSC cohort with and without pathogenic variants. **(A)** Individuals with pathogenic variants (*de novo* LGD and CNV) had a significantly lower IQ than individuals without pathogenic variants, but no change in autism severity (SRS T-score) was observed between the three groups. **(B)** Individuals with *de novo* LGD variants in candidate autism genes had a lower IQ and more severe autism phenotypes than individuals without such variants. n indicates sample size, p-values were derived from two-tailed Mann-Whitney tests, and dotted lines within each plot indicate the median and first and third quartiles. All statistics were calculated using R v.3.4.2 (R Foundation for Statistical Computing, Vienna, Austria).

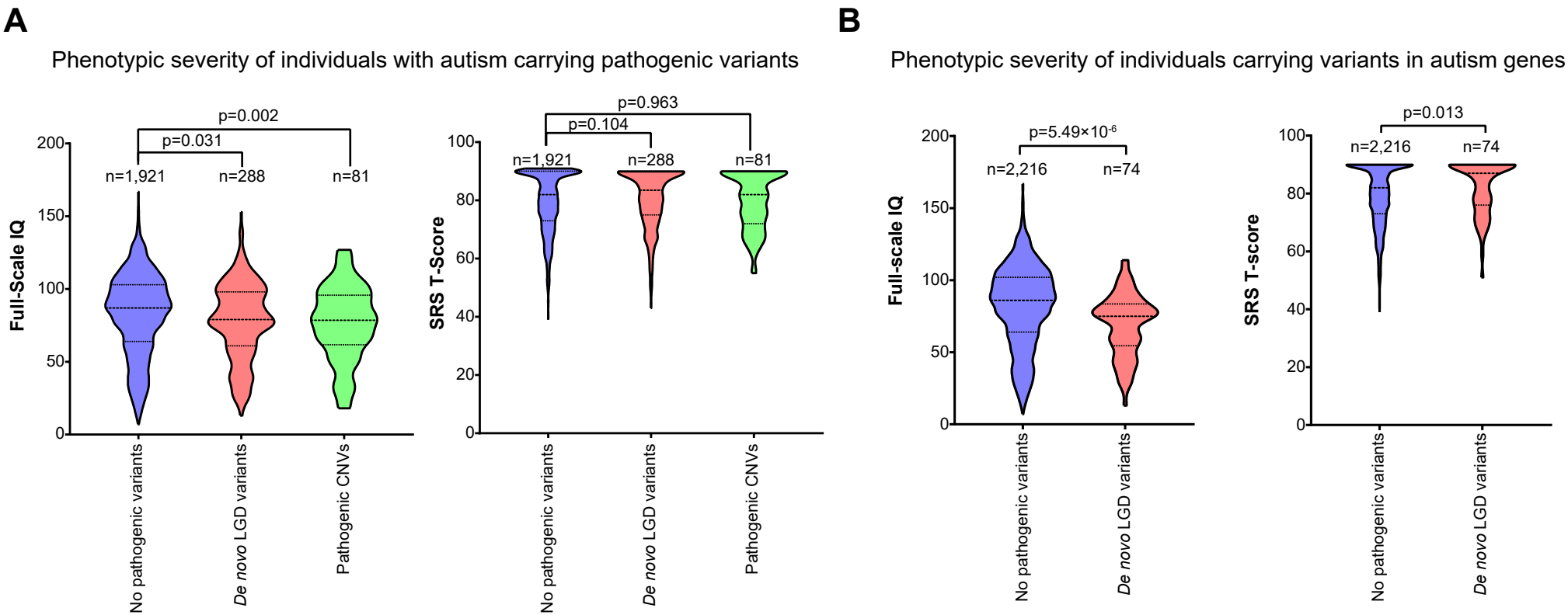
**Figure 2.** Functional analysis of genes with *de novo* LGD variants in individuals with high-functioning autism. **(A)** Genes with *de novo* LGD variants in individuals with high-functioning autism had lower average RVIS (left) and pLI (right) percentile scores than those for reported autism-associated genes. Thick dotted lines across the violin plots indicate thresholds for gene pathogenicity: $<20^{th}$ percentile for RVIS and $<18^{th}$ percentile for pLI (>0.9 raw score). Thin lines within the violin plots indicate the median and first and third quartiles. p-values were derived from two-tailed Mann-Whitney tests. **(B)** Expression of genes with *de novo* variants in individuals with high-functioning autism and autism-associated genes in the developing human brain. Autism-associated genes were enriched for specific expression in the cortex and cerebellum during early development, while no enrichment was seen in the genes identified in individuals with high-function autism. Hexagon sizes represent the number of genes preferentially expressed in each brain tissue and timepoint, while colors of the hexagons represents p-values for the enrichment of autism genes among each set of preferentially-expressed genes. **(C)** Frequency of phenotypes observed in mouse knockout models for genes with *de novo* LGD variants in individuals with high-functioning autism compared with reported autism-associated genes. * indicates $p<0.05$ with Benjamini-Hochsberg correction.

348 **Table 1. Individuals carrying variants in autism-associated genes with comorbid ID.**

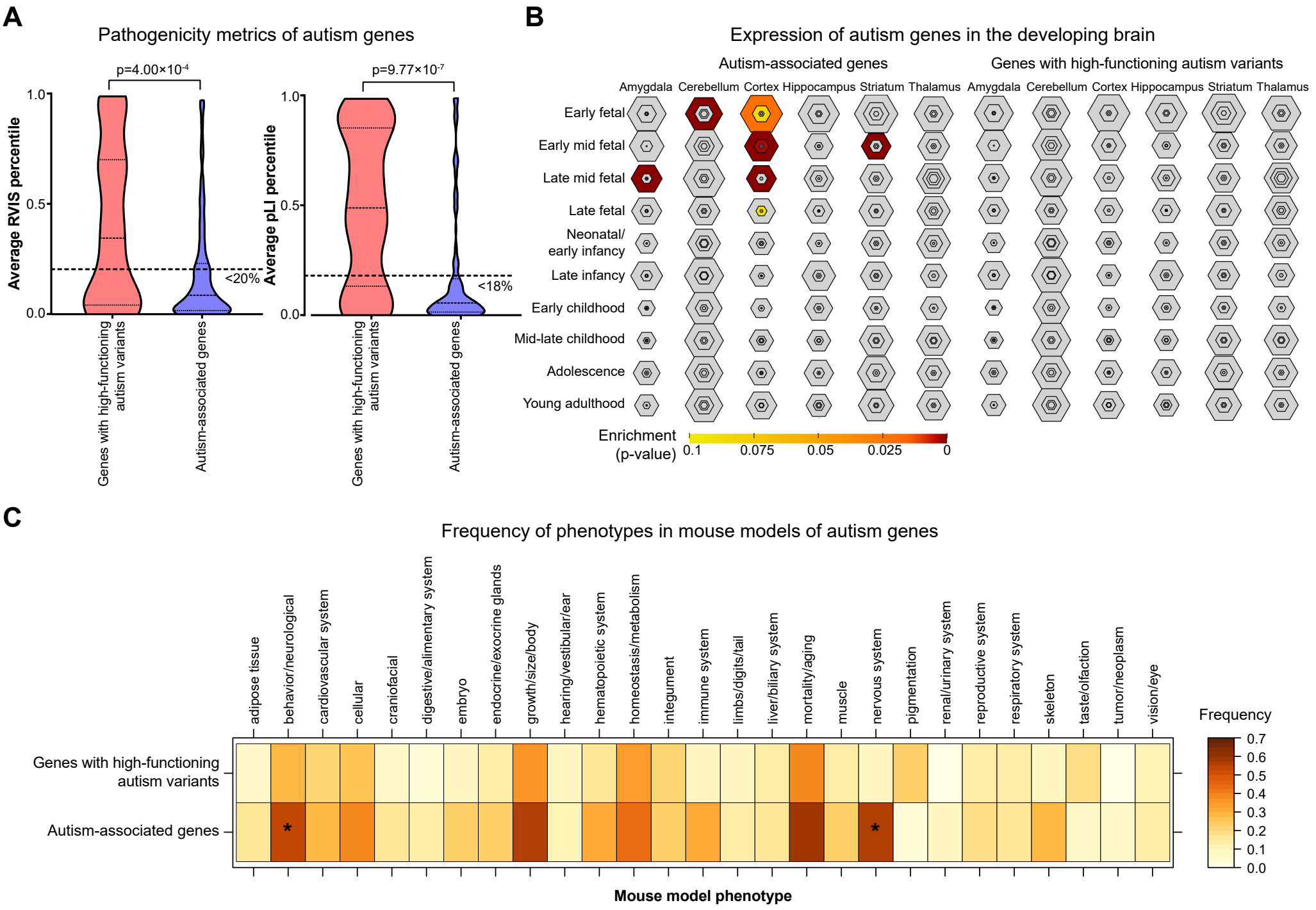
| Autism-associated genes | Cases with ID | Autism cases with or without ID | Autism cases with comorbid ID |
|---|---|---|---|
| *ADNP* | 134/134 | 114/134 | 114/114 |
| *ANK2* | 1/1 | 1/1 | 1/1 |
| *ANKRD11* | 10/10 | 9/10 | 9/9 |
| *ARID1B* | 137/153 | 80/153 | 80/80 |
| *ASH1L* | 14/14 | 4/14 | 4/4 |
| *ASXL3* | 18/19 | 16/19 | 15/16 |
| *BCL11A* | 11/16 | 4/16 | 3/4 |
| *CHD2* | 3/3 | 3/3 | 3/3 |
| *CHD8* | 56/75 | 61/75 | 46/61 |
| *CUL3* | 1/1 | 1/1 | 1/1 |
| *DDX3X* | 97/97 | 33/97 | 33/33 |
| *DYRK1A* | 7/26 | 9/26 | 7/9 |
| *KMT2A* | 76/99 | 12/99 | 10/12 |
| *MECP2* | 1/1 | 1/1 | 1/1 |
| *MYT1L* | 1/1 | 1/1 | 1/1 |
| *NRXN1* | 45/60 | 34/60 | 23/34 |
| *POGZ* | 48/49 | 29/49 | 29/29 |
| *SCN2A* | 19/33 | 9/33 | 7/9 |
| *SETD5* | 15/16 | 7/16 | 7/7 |
| *SHANK3* | 35/37 | 26/37 | 24/26 |
| *SYNGAP1* | 11/23 | 11/23 | 11/11 |
| *TCF20* | 47/48 | 32/48 | 31/32 |
| | | | |
| **Total** | 787/916 (85.9%) | 497/916 (54.3%) | 460/497 (92.6%) |

349

# Figure 1


A
Phenotypic severity of individuals with autism carrying pathogenic variants
Full-Scale IQ
p=0.002
p=0.031
n=1,921
n=288
n=81
No pathogenic variants
De novo LGD variants
Pathogenic CNVs
SRS T-Score
p=0.963
p=0.104
n=1,921
n=288
n=81
No pathogenic variants
De novo LGD variants
Pathogenic CNVs
B
Phenotypic severity of individuals carrying variants in autism genes
Full-scale IQ
p=5.49×10-6
n=2,216
n=74
No pathogenic variants
De novo LGD variants
SRS T-score
p=0.013
n=2,216
n=74
No pathogenic variants
De novo LGD variants

# Figure 2


A
Pathogenicity metrics of autism genes
p=4.00×10-4
p=9.77×10-7
Average RVIS percentile
Average pLI percentile
1.0
0.5
0.0
<20%
<18%
Genes with high-functioning autism variants
Autism-associated genes
B
Expression of autism genes in the developing brain
Autism-associated genes
Genes with high-functioning autism variants
Amygdala
Cerebellum
Cortex
Hippocampus
Striatum
Thalamus
Early fetal
Early mid fetal
Late mid fetal
Late fetal
Neonatal/ early infancy
Late infancy
Early childhood
Mid-late childhood
Adolescence
Young adulthood
Enrichment (p-value)
0.1
0.075
0.05
0.025
0
C
Frequency of phenotypes in mouse models of autism genes
adipose tissue
behavior/neurological
cardiovascular system
cellular
craniofacial
digestive/alimentary system
embryo
endocrine/exocrine glands
growth/size/body
hearing/vestibular/ear
hematopoietic system
homeostasis/metabolism
integument
immune system
limbs/digits/tail
liver/biliary system
mortality/aging
muscle
nervous system
pigmentation
renal/urinary system
reproductive system
respiratory system
skeleton
taste/olfaction
tumor/neoplasm
vision/eye
Genes with high-functioning autism variants
Autism-associated genes
*
*
Mouse model phenotype
Frequency
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0