**Title:**Deleterious coding variation associated with neurodevelopmental disorders is consistent across populations, as exemplified by admixed Latin American populations

**Authors**: Buxbam, J.D.1, for the Genomics of Autism in Latin American Ancestries (GALA) Consortium

**Introduction**: Autism spectrum disorder (ASD) is characterized by deficits in social communication and the presence of restricted interests and/or repetitive behaviors (Lord et al., 2020). While the majority of the genetic liability for ASD is attributed to common genetic variation, rare variants, often arising de novo, play a substantial role in individual liability (Klei et al., 2012; Gaugler et al., 2014). Multiple large-scale studies of rare and common variation associated with ASD risk are ongoing, and dozens of new high-confidence risk genes have emerged (Fu et al., 2022; Zhou et al., 2022). Such findings have led to improvements in clinical care and serve as crucial, initial steps towards creating novel treatments and personalized interventions. Gene-targeted therapies for rare genetic disorders in ASD and associated neurodevelopmental disorders (NDDs) have emerged as a very dynamic area of study in academia and industry (Davidson et al., 2022). The overwhelming majority of participants in gene discovery studies are of European (EUR) ancestry, even though they comprise only 16% of the global population (Fatumo et al., 2022). This limited window into risk architecture across ancestries could exacerbate pre-existing disparities in diagnostics and service use for ASD (Martin et al., 2019). Indeed, recent studies have reported high rates of inconclusive results after genetic testing in non-European individuals, likely because of uncertainty in interpreting genomic variants (Abul-Husn et al., 2023; Venner et al., 2024). We have established the Genomics of Autism in Latin American Ancestries (GALA) Consortium to investigate the impact of genetic and environmental factors on ASD across Latin Americans, including participants from all of the Americas, corresponding to the Admixed American (AMR) superpopulation in the 1000 Genomes Project. Native Americans are thought to have originated in Northeast Asia (Hoffecker et al., 2023). Subsequent variation among existing Native American populations is likely due to regional differentiation and additional migration across the Bering Strait. The post-Columbian movement of people into the Americas added to this genetic diversity by admixture of Native American, African, and European populations. These AMR individuals comprise the largest recently-admixed population in the world, and the largest racial or ethnic minority in the United States. It is as yet unknown whether the genetic architecture of ASD differs across ancestral populations, and the genetic diversity of the AMR group (Moreno-Estrada et al., 2014; Ongaro et al., 2019) makes this question especially relevant.

**Method**: GALA comprises multiple sites from North, Central, and South America recruiting AMR participants for studies on the risk architecture of ASD. ASD diagnoses are based on expert clinical evaluations using DSM-5 criteria, incorporating all available data including standardized assessments. Participants can be any age. Individuals with a known genetic condition (e.g., Fragile X syndrome) are excluded from analyses. Once a diagnosis of ASD is confirmed, the individual and their parents contribute a sample (blood or saliva) for genetic analyses. If both parents are not available, collection of other biological family members is encouraged (siblings, grandparents, etc.). Collection sites generally also collect additional clinical and family history information. DNA samples were subjected to whole exome sequencing and data analyzed using GATK best practices and methods described in detail in Fu et al.

**Results**: We present the largest sequencing study of ASD in Latin American individuals (**n>15,000**) and compare findings from non-AMR cohorts. We show that a common measure of evolutionary impact on gene-level genetic variation, i.e., genomic constraint scores, differ by ancestry. Yet, this is not the case for the most constrained genes, which are depauperate of population-level variation that is expected based on sequence composition of these genes. This is important because most if not all identified ASD-associated genes are evolutionarily constrained (Kosmicki et al., 2017; Fu et al., 2022), and our results indicate the information on strong constraint applies over diverse populations. Using Bayesian models, we identify 35 genome-wide significant ASD risk genes in Latin Americans, and observe a great degree of overlap with findings in largely European cohorts. Risk genes emerging from our analyses that had weaker evidence of association in prior ASD studies show overlap with genes identified in cohorts with severe developmental disorders (DD), possibly due to differing ascertainment strategies. We conclude that ASD and other NDD genes are shared across ancestry and that existing genetic testing pipelines are effective for the most deleterious variation, if appropriate approaches are used. We also conclude that the biology of ASD is likely universal, and not impacted to any detectable degree by ancestry.

**Discussion:** The results are consistent with the assumption that the same set of highly constrained genes identified in ongoing genome-wide studies are associated with ASD, regardless of ancestry. This perspective also receives support from common variant studies in complex traits, where causal effects appear to be highly similar across ancestries (Hou et al., 2023). These observations are consistent with the neurobiology of ASD being universal and provide support for the translatability of clinical genetic approaches across ancestries. Using a commercial clinical genetics software platform, we confirm the overall translatability of clinical genetic approaches when focusing on rare deleterious variation; however, we also reveal significant differences in the rate of pathogenic/likely pathogenic (P/LP) variants between AMR and non-AMR individuals and EUR and non-EUR individuals. The causes driving differences in rates of P/LP needs to be better understood and our results suggest paths to improve genetic testing results. Clearly, of key importance is to use allele frequency from all relevant populations, as we have done here. Next, where possible, we recommend minimizing reliance on previously reported pathogenic variants. And, finally, we should recognize the challenges inherent in ancestries beyond European and a few other commonly characterized populations. Based on our results, which show similar patterns to those observed in EUR studies, we can conclude that the vast majority of our results arise from true positives. Nonetheless, we should not conclude that populations are all the same when it comes to calling de novo variation. Indeed, we can be confident they are not, given what we know about increased genetic diversity in African populations (Yu et al., 2002; Gomez, Hirbo and Tishkoff, 2014; Pereira et al., 2021; Yilmaz et al., 2021) and the impact that cryptic structural variation and singleton events have on the reliability of calling ultra-rare variation. Only through deeper genetic studies can we expect completely comparable results to those of EUR population samples.

**References:**

Abul-Husn, N.S. et al. (2023) ‘Molecular diagnostic yield of genome sequencing versus targeted gene panel testing in racially and ethnically diverse pediatric patients’, Genetics in medicine: official journal of the American College of Medical Genetics, 25(9), p. 100880.

Davidson, B.L. et al. (2022) ‘Gene-based therapeutics for rare genetic neurodevelopmental psychiatric disorders’, Molecular therapy: the journal of the American Society of Gene Therapy, 30(7), pp. 2416–2428.

Fatumo, S. et al. (2022) ‘A roadmap to increase diversity in genomic studies’, Nature medicine, 28(2), pp. 243–250.

Fu, J.M. et al. (2022) ‘Rare coding variation provides insight into the genetic architecture and phenotypic context of autism’, Nature genetics, 54(9), pp. 1320–1331.

Gaugler, T. et al. (2014) ‘Most genetic risk for autism resides with common variation’, Nature genetics, 46(8), pp. 881–885.

Gomez, F., Hirbo, J. and Tishkoff, S.A. (2014) ‘Genetic variation and adaptation in Africa: implications for human evolution and disease’, Cold Spring Harbor perspectives in biology, 6(7), p. a008524.

Hou, K. et al. (2023) ‘Causal effects on complex traits are similar for common variants across segments of different continental ancestries within admixed individuals’, Nature genetics, 55(4), pp. 549–558.

Klei, L. et al. (2012) ‘Common genetic variants, acting additively, are a major source of risk for autism’, Molecular autism, 3(1), p. 9.

Kosmicki, J.A. et al. (2017) ‘Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples’, Nature genetics, 49(4), pp. 504–510.

Lord, C. et al. (2020) ‘Autism spectrum disorder’, Nature reviews. Disease primers, 6(1), p. 5.

Martin, A.R. et al. (2019) ‘Clinical use of current polygenic risk scores may exacerbate health disparities’, Nature genetics, 51(4), pp. 584–591.

Moreno-Estrada, A. et al. (2014) ‘Human genetics. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits’, Science, 344(6189), pp. 1280–1285.

Ongaro, L. et al. (2019) ‘The Genomic Impact of European Colonization of the Americas’, Current biology: CB, 29(23), pp. 3974–3986.e4.

Pereira, L. et al. (2021) ‘African genetic diversity and adaptation inform a precision medicine agenda’, Nature reviews. Genetics, 22(5), pp. 284–306.

Venner, E. et al. (2024) ‘The frequency of pathogenic variation in the All of Us cohort reveals ancestry-driven disparities’, Communications biology, 7(1), p. 174 .

Yilmaz, F. et al. (2021) ‘Genome-wide copy number variations in a large cohort of bantu African children’, BMC medical genomics, 14(1), p. 129.

Yu, N. et al. (2002) ‘Larger genetic differences within africans than between Africans and Eurasians’, Genetics, 161(1), pp. 269–274.

Zhou, X. et al. (2022) ‘Integrating de novo and inherited variants in 42,607 autism cases identifies mutations in new moderate-risk genes’, Nature genetics, 54(9), pp. 1305–1319.

1Seaver Autism Center for Research and Treatment, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY