Neuron
Previews

Rare Inherited Variation in Autism: Beginning to See the Forest and a Few Trees

Jason L. Stein,[1] Neelroop N. Parikshak,[1] and Daniel H. Geschwind[1,*]
[1]Program in Neurogenetics, Department of Neurology and Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA
*Correspondence: dhg@mednet.ucla.edu
http://dx.doi.org/10.1016/j.neuron.2013.01.010

In this issue of *Neuron*, two papers (Lim et al., 2013; Yu et al., 2013) use whole-exome sequencing (WES) to elucidate the contribution of inherited variation to the risk for autism by leveraging the increased penetrance of homozygous and compound heterozygous rare variants in autosomes and hemizygous rare variants in the X chromosome of males. Together, they expand our knowledge about the genetic architecture of ASD, verify previously identified genes, and identify novel mutations that will guide the discovery of the critical biological processes disrupted in autism.

Autism is a spectrum of neurodevelopmental disorders (referred to as autism spectrum disorder, ASD) affecting around 1 in 88 individuals (2012 CDC estimate). Genetic risk factors for ASD play a substantial role and come in many forms: those transmitted from parents to affected child or those appearing de novo in the germline, as well as those found commonly in the population, or those only rarely observed (Figures 1 and 2). Genetic investigations have revealed that hundreds of genomic loci are likely involved and it is now important to advance our understanding of ASD's genetic architecture while simultaneously identifying specific deleterious variants to understand the biological pathways involved (Berg and Geschwind, 2012). Inherited variation from both common and rare alleles provides the largest genetic contribution to population-wide ASD risk, explaining ~40% (95% confidence interval: 8%–84%) of the risk for developing ASD (Hallmayer et al., 2011). Though common variants are estimated to be a large driving factor for the disorder (40%–60% variance explained; Klei et al., 2012), the effect size of individual common variants is small (estimated to be odds-ratio < 1.2; Anney et al., 2012). This observation has led to a search for rare variants, which may exhibit larger individual effect sizes.

To this end, cost-effective high-throughput genome sequencing of ASD patients analyzed in paradigms where the effect of a mutation can be seen over random chance plays a critical role (Iossifov et al., 2012; Neale et al., 2012; O'Roak et al., 2012; Sanders et al., 2012). However, this process is complicated by the fact that each of us inherits over 100 nonsense, loss-of-function mutations in our genomes, leading to about 20 completely inactivated genes and several dozen de novo variants, some of which may be functional, but none of which are clearly associated with disease (MacArthur et al., 2012). In the latter category, predicted protein structure altering rare variants have been observed to be more frequent in ASD cases, and recent work suggests nontransmitted, de novo mutations that delete genic regions, perturb splicing, or truncate protein products may contribute to the development of ASD in 15%–20% of cases (Devlin and Scherer, 2012). Although inherited mutations have been identified in rare families, no population-based exome sequencing study has demonstrated a significant role for inherited deleterious variants in ASD risk, leaving the contribution of this class of genetic variation unknown. Now, as shown in two papers in this issue of *Neuron* (Lim et al., 2013; Yu et al., 2013), it is clear that ASD risk is increased when two rare variants deleteriously affect both copies of a protein coding gene, consistent with a role for recessive inheritance of nonsynonymous mutations in ASD.

Starting with a population-based approach, Lim et al. (2013) performed WES in 933 ASD cases and 869 controls, allowing them to see the complete genome sequence in protein-coding and flanking regions. They focused on the most damaging type of mutations (nonsense and essential splice site) transmitted from the mother and father to the child. In autosomes, they studied loci where the mutations truncate both copies of a protein, and in the X chromosome in males, they studied mutations that truncate the only copy. Only rare alleles of this type (minor allele frequency $\leq$5%) were studied, based on the hypothesis that double mutations commonly found in the population are not pathological. These rare, putative recessively acting mutations were observed twice as many times in ASD patients as controls in autosomes (6% of cases and 3.3% of controls carried such a mutation). When overlapped with a data set of brain gene expression, the overall odds-ratio (OR) for ASD increased to 2.7. Importantly, this enrichment of double mutation variants was replicated in an independent data set, this time showing a smaller, but still significant effect (7.6% of cases and 5.5% of controls). The authors estimate an overall 3% contribution to risk for ASD from this class of mutations. Rare hemizygous mutations on the X chromosome, which would also be depleted of the normal protein in males, were also enriched in male ASD cases compared to controls (4.8% versus 3.1%), revealing involvement in 1.7% of male ASD cases. This study fills an important gap in our knowledge of the genetic architecture of ASD by estimating that about 5% of ASD cases may be affected by rare inherited loss-of-function homozygous, compound heterozygous, or X chromosome mutations in

males. In this way, Lim et al. (2013) significantly advance our population-level understanding of risk for ASD.

One key issue is that most current studies on the contribution of rare variants to genetic risk are not well powered to identify specific genes. Typically, functional studies or repeated instances of rare mutations in the same gene are needed to ensure that a variant is indeed functional and associated with ASD. Alternatively, other evidence, such as gene expression data can be used to functionally prioritize variants, as many mutations, including but not limited to de novo deletions (e.g., copy number variations; CNVs), or recessive mutations causing intellectual disability (ID) or ASD are expected to significantly reduce RNA levels even in peripheral blood (Luo et al., 2012). We predict that future studies combining genome sequencing with gene expression will have increased power to detect pathogenic inherited or de novo genetic variation. Short of this, very large sample sizes (larger than current GWAS samples ranging ~10,000 subjects) or unique families with multiple affected children (multiplex families) will be necessary.

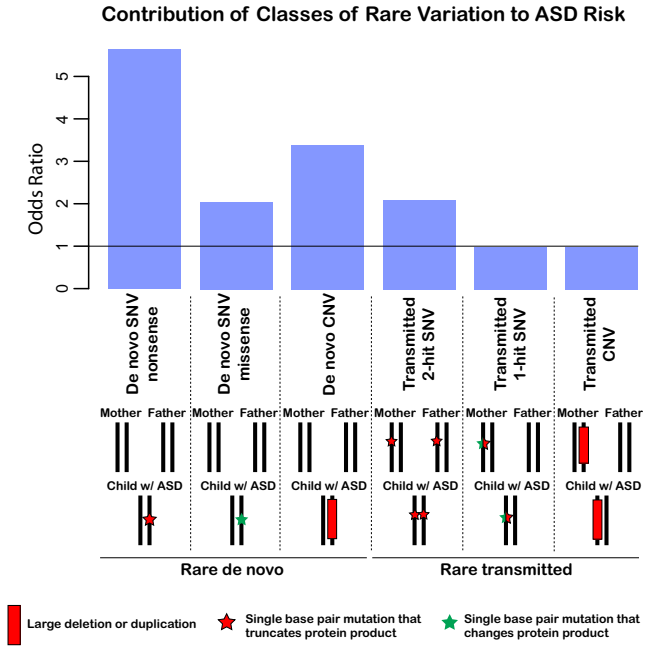

Figure 1. Risk Imparted by Different Kinds of Rare Genetic Variants
The population-level enrichment in ASD versus controls of each form of rare mutation studied to date (representative single nucleotide variant, or SNV estimates from Sanders et al., 2012 but also see Neale et al., 2012; O'Roak et al., 2012; Iossifov et al., 2012; copy number variation, or CNV estimates from Sanders et al., 2012; two-hit transmitted SNV estimates from Lim et al., 2013). An emerging trend in ASD genetics is that the most damaging types of rare variation are observed more often in ASD than controls. De novo variants are overrepresented in autism more than transmitted variants. Moreover, the most damaging class of variants have the greatest degree of overrepresentation. These risk estimates do not include confidence intervals or all of the studies but attempt to convey the relative effect sizes. Below the bar graph are diagrams illustrating the method of inheritance and type of mutation.

In the second study, Yu et al. (2013) took an elegant approach to identify specific inherited rare variants by studying multiplex families, combining multiple gene-mapping modalities, and validating variant pathogenicity experimentally. Specifically, the authors searched consanguineous and/or multiplex families using a combination of heterozyogsity mapping and linkage-directed WES analysis. This identified regions in three families with high linkage scores (likelihood >600:1 of the genetic region containing a variant tracking ASD in that family) where specific nonsynonymous or frameshift rare variants were also found.

This analysis identified three genes, *AMT*, *PEX7*, and *SYNE1* as affected by homozygous or compound heterozygous rare variation linked to ASD. To demonstrate the pathogenicity of each of these rare variants, the authors performed detailed in vitro functional analysis, which has not been previously performed in population-based WES studies. For example, one family contained a homozygous I308F change in *AMT*, a gene coding for a glycine-degradation enzyme in which severe mutations have been shown to lead to neonatal nonketotic hypoglycemia (NKT), a life-threatening neonatal metabolic syndrome. This mutation was analyzed biochemically and its effect on protein solubility was evaluated in bacteria, suggesting that the mutation may induce a protein-folding defect and result in a functional hypomorph. Indeed, the family contained three affected individuals with an atypical, milder manifestation of NKT symptoms in addition to ASD. The authors propose that, in general, less severe mutations in genes involved in recessive neurodevelopmental ID syndromes may lead to ASD.

A similar inheritance pattern and relationship to syndromic disease was observed with *PEX7*, which is the causative gene for rhizomelic chondrodysplasia punctata, and *SYNE1*, where null mutations have been linked to cerebellar ataxia and a recessive form of arthrogryposis multiplex congenita. The functional data presented by the authors and absence of the strong classical phenotypes suggested these changes were also hypomorphic, and their experimental investigations supported this. Yu and colleagues' approach exemplifies how various lines of direct and indirect evidence—genome-wide linkage, exonic variation, biological validation, and clinical relevance to existing syndromes—may be used to convincingly implicate a rare variant, which itself may never reach population-wide, genome-wide significance.

Yu et al. (2013) further generalized this mutational pattern and its relationship to syndromic disease by compiling a set of risk genes comprising these 3 and 70 other previously implicated ID-related genes. They then compared rare inherited variants from WES in 163 consanguineous and/or multiplex families to 831 population-matched individual exomes to find additional homozygous, compound heterozygous, or hemizygous rare mutations. They find five families with previously unidentified nonsense or frameshift variants and 11 with previously unidentified missense variants in their set of risk genes. Intriguingly, two families had nonsense or frameshift mutations in *PAH* and one had a likely functional missense mutation in *AMT*, suggesting a potential metabolic contribution to their phenotype and perhaps autistic symptoms. This highlights the potential for better genetic diagnoses and treatment by immediate

intervention in a subset of ASD, as was recently highlighted in a recently discovered metabolic form of ASD (Novarino et al., 2012). In total, this study's cohort found novel changes linked to ASD in the novel genes *AMT*, *PEX7*, *SYNE1*, *VPS13B/COH1*, *PAH*, and *POMGNT1*, as well as previously implicated genes *NLGN4X* and *MECP2*.

These studies provide two unique vistas on ASD genetics. Given an etiologically and phenotypically heterogeneous neurodevelopmental disorder that may involve hundreds of genes, the field has attempted to gain a foothold on biological pathways by identifying genes using highly penetrant mutations that are linked to ASD. These studies also provide convincing statistical evidence for the role of homozygous or compound heterozygous loss of function mutations in ASD risk. Moreover, the results emphasize how difficult it can be to assign blame to a single gene, given the rarity of events, their occurrence in controls, and the numbers necessary to attain genome-wide significance. Observing the event more times in affected individuals is still necessary to provide definitive proof of genetic association with disease. This is further complicated by the observation that mutation rates vary by several orders of magnitude across the genome, which suggests the need for locus-specific calculations of variant significance in ASD (Michaelson et al., 2012).

Finally, the clearly defined role for CNVs and SNVs that delete or alter either one or both copies of a gene or its isoforms indicates that transcript levels play a significant role in ASD susceptibility. From this perspective, we predict that future work will identify variants in regulatory regions that affect transcript levels, such as promoters and enhancers, and that noncoding regions will be just as functionally important to ASD as currently implicated protein-coding regions. This is also supported by the high mutability of DNase hypersensitive sites and CpG-rich regions (Michaelson et al., 2012). So far, the contribution of regulatory elements is unexplored at a population level, as WES does not effectively measure these regions. Nevertheless, these current approaches provide unprecedented insight (e.g., Figure 2) that will aid in explaining ASD's pathobiology.

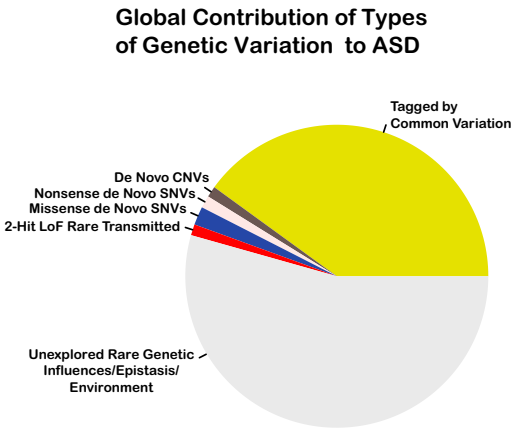

Figure 2. The Percentage of Variance Explained by Various Forms of Genetic Risk Factors for ASD

Common variants capture a large percentage of population risk for ASD (Klei et al., 2012), whereas current studies show that rare exonic variants explain a smaller amount of variance. Percent variance explained for rare variants, based on cited references in Figure 1, is estimated via the squared correlation of outcome with predicted probabilities from logistic regression (Hosmer and Lemeshow, 2000). Note that though common variants explain a large proportion of the variance, each locus is expected to be of small effect, whereas, though rare variants explain a small proportion of the variance, each locus is expected to be of larger effect. As-yet-unexplored rare variation (in the 98% of the genome not studied by WES), interactions among genes, and the influence of the environment are potential culprits to explain the remaining risk. Additionally, because syndromic forms of autism are typically excluded from most population-level studies from which these genetic estimates have been derived, they are not included here. Clinical ascertainment suggests that their prevalence is on the order of 5% within the entire population of autism.

REFERENCES

Anney, R., Klei, L., Pinto, D., Almeida, J., Bacchelli, E., Baird, G., Bolshakova, N., Bölte, S., Bolton, P.F., Bourgeron, T., et al. (2012). Hum. Mol. Genet. *21*, 4781–4792.

Berg, J.M., and Geschwind, D.H. (2012). Genome Biol. *13*, 247.

Devlin, B., and Scherer, S.W. (2012). Curr. Opin. Genet. Dev. *22*, 229–237.

Hallmayer, J., Cleveland, S., Torres, A., Phillips, J., Cohen, B., Torigoe, T., Miller, J., Fedele, A., Collins, J., Smith, K., et al. (2011). Arch. Gen. Psychiatry *68*, 1095–1102.

Hosmer, D., and Lemeshow, S. (2000). Applied Logic Regression, Second Edition (New York: John Wiley & Sons, Inc.), pp. 164–167.

Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., Yamrom, B., Lee, Y.H., Narzisi, G., Leotta, A., et al. (2012). Neuron *74*, 285–299.

Klei, L., Sanders, S.J., Murtha, M.T., Hus, V., Lowe, J.K., Willsey, A.J., Moreno-De-Luca, D., Yu, T.W., Fombonne, E., Geschwind, D., et al. (2012). Mol. Autism *3*, 9.

Lim, E.T., Raychaudhuri, S., Sanders, S.J., Stevens, C., Sabo, A., MacArthur, D.G., Neale, B.M., Kirby, A., Ruderfer, D.M., Fromer, M., et al. (2013). Neuron *77*, this issue, 235–242.

Luo, R., Sanders, S.J., Tian, Y., Voineagu, I., Huang, N., Chu, S.H., Klei, L., Cai, C., Ou, J., Lowe, J.K., et al. (2012). Am. J. Hum. Genet. *91*, 38–55.

MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B., et al.; 1000 Genomes Project Consortium. (2012). Science *335*, 823–828.

Michaelson, J.J., Shi, Y., Gujral, M., Zheng, H., Malhotra, D., Jin, X., Jian, M., Liu, G., Greer, D., Bhandari, A., et al. (2012). Cell *151*, 1431–1442.

Neale, B.M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K.E., Sabo, A., Lin, C.F., Stevens, C., Wang, L.S., Makarov, V., et al. (2012). Nature *485*, 242–245.

Novarino, G., El-Fishawy, P., Kayserili, H., Meguid, N.A., Scott, E.M., Schroth, J., Silhavy, J.L., Kara, M., Khalil, R.O., Ben-Omran, T., et al. (2012). Science *338*, 394–397.

O'Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D., et al. (2012). Nature *485*, 246–250.

Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L., et al. (2012). Nature *485*, 237–241.

Yu, T.W., Chahrour, M.H., Coulter, M.E., Jiralerspong, S., Okamura-Ikeda, K., Ataman, B., Schmitz-Abe, K., Harmin, D.A., Adli, M., Malik, A.N., et al. (2013). Neuron *77*, this issue, 259–273.