Damaging de novo mutations diminish motor skills in children on the autism spectrum

Andreas Buja[a], Natalia Volfovsky[b], Abba M. Krieger[a], Catherine Lord[c], Alex E. Lash[b], Michael Wigler[d,e,1], and Ivan Iossifov[d,e]

[a]Statistics Department, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104; [b]Simons Foundation, New York, NY 10010; [c]Center for Autism and the Developing Brain, Weill Cornell Medical College, White Plains, NY 10605; [d]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724; and [e]New York Genome Center, New York, NY 10013

Contributed by Michael Wigler, December 21, 2017 (sent for review September 7, 2017; reviewed by Daniel H. Geschwind and Helen B. Tager-Flusberg)

In individuals with autism spectrum disorder (ASD), de novo mutations have previously been shown to be significantly correlated with lower IQ but not with the core characteristics of ASD: deficits in social communication and interaction and restricted interests and repetitive patterns of behavior. We extend these findings by demonstrating in the Simons Simplex Collection that damaging de novo mutations in ASD individuals are also significantly and convincingly correlated with measures of impaired motor skills. This correlation is not explained by a correlation between IQ and motor skills. We find that IQ and motor skills are distinctly associated with damaging mutations and, in particular, that motor skills are a more sensitive indicator of mutational severity than is IQ, as judged by mutational type and target gene. We use this finding to propose a combined classification of phenotypic severity: mild (little impairment of either), moderate (impairment mainly to motor skills), and severe (impairment of both IQ and motor skills).

autism | genetics | de novo mutation | motor skills | IQ

Autism spectrum disorder (ASD) is a neuropsychiatric disorder, conventionally characterized by core phenotypes, including persistent deficits in social communication and interaction, and restricted, repetitive patterns of behavior (ref. 1, p. 947). Genetics is a strong determining factor, as inferred by the high rate of concordance between identical twins, elevated sibling risk, and the presence of genetic evidence. This evidence includes a significantly increased incidence of likely gene-damaging de novo mutations (2–6), the preferential transmission of such variants to the affected child (6–8), consistent enrichment of autism-associated genes in certain functional classes, and increased sharing of ancestral variants between unrelated affected individuals (9–11).

Although the core features form the consensus clinical signature of ASD, such children also have a wide range of other phenotypes and comorbidities (1). A wealth of phenotypic data on autistic individuals is found in the Simons Simplex Collection (SSC), an archive of samples from "simplex" families that have only one affected child and typically include at least one additional unaffected child (12). The SSC samples have been the source for the discovery of many candidate causal de novo variants. The richly documented and quantified phenotypic variables in the SSC provide an excellent opportunity to correlate variants with phenotypes. In an earlier effort, we and others have shown that ASD individuals with low nonverbal IQ (nvIQ) have a significantly increased incidence of damaging de novo mutations (2, 13). These observations support the hypothesis that damaging de novo mutations may have broader neurological effects than ASD alone.

Wishing to test this hypothesis further, we looked for further correlations between phenotypes and damaging de novo (dn) mutations. Although IQ reflects some aspects of cognitive ability, there are other fundamental manifestations of altered neurological function. Indeed, neurodevelopmental delay, such as age of first walking, can be the first presenting symptom in autism (14, 15). The delay in this milestone may more generally reflect diminished motor skills (MS), a well-documented feature of ASD (e.g., ref. 16). Some have argued that motor impairment should be included among core ASD features (17–21). Thus, we decided to look for correlations between MS and dn mutations, in particular those that are likely gene disrupting (or LGDs: nonsense, frame shift, and splice site altering). While our manuscript was prepared for submission, another manuscript appeared (22) reporting related conclusions using the genotypic and phenotypic data from the same cohort, the SSC. We note striking differences of causal inference discussed later.

Motor skills of most affected children in the SSC were reported by parents in the Developmental Coordination Disorder Questionnaire (DCDQ) and, for young children, also in the Vineland Adaptive Behavior Scales (VABS-II). Using these scores, we find dn LGD mutations correlate with MS at least as strongly and significantly as with nvIQ. Statistical significance of this correlation is found not only for the total DCDQ and VABS-II scores, but also for their subcomponents, including fine and gross motor skills, as well as for related variables such as delay in the developmental milestone age of first walking. Moreover, significance of the correlations increases when we weight a dn LGD mutation by evidence that its target gene is under strong purifying selection in humans, or that it is a member of certain functional classes. We extend our observations even further by including an analysis of missense mutations, in which correlation to MS (much less so IQ) becomes evident when these mutations are additionally weighted by predictions of deleterious effect.

Although IQ and MS are correlated with each other, they each correlate with damaging dn mutations even after either is

Significance

Genetics is a major determining factor in autism spectrum disorder (ASD). To date, only the most severe class of de novo mutation, likely gene disruptive (LGD), has been correlated with IQ, a phenotypic characteristic associated with ASD, but not a core feature. A less severe class of de novo mutation, missense, while enriched in individuals with ASD, has been refractory to correlation with any ASD phenotypic feature. In this report, we demonstrate that de novo LGD and missense mutations scored by target gene vulnerability both show significant associations with diminished motor skills.

Author contributions: A.B., A.M.K., M.W., and I.I. designed research; A.B., A.M.K., and I.I. performed research; N.V., C.L., and A.E.L. contributed new reagents/analytic tools; A.B., A.M.K., M.W., and I.I. analyzed data; and A.B., C.L., A.E.L., M.W., and I.I. wrote the paper.

Reviewers: D.H.G., University of California, Los Angeles; and H.B.T.-F., Boston University.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. Email: wigler@cshl.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1715427115/-/DCSupplemental.

adjusted for the other. Although MS and IQ significantly correlate with the severity of the core ASD phenotypes, we observe no consistently significant correlation between damaging dn events and core ASD features. We discuss the possible explanations of this puzzling finding in the context of other genetic sources for social cognitive impairments in ASD.

Results

Motor Skills in Those with de Novo LGDs. Motor skills are scored in the DCDQ as a 15-item parent questionnaire that assesses a child's fine and gross motor skills (ref. 23; see the *SI Appendix* for more detailed descriptors). The total DCDQ score and three summary subscores are available for 87% of the exome-sequenced affected children in the SSC. The subscores of the DCDQ are control during movement, fine motor/handwriting, and general coordination. Though not standardized for age, the raw score is negligibly age dependent (*SI Appendix*, section 3). Additional measures of motor skill development are found in various instruments, in particular the Vineland-II Motor Skill Domain for young children, which includes two subscales: fine and gross motor skills. In addition one item from the Social Responsiveness Scale (SRS): item 14, asks about problems with being well-coordinated (on a severity scale 0–3). Finally, the Autism Diagnostic Interview (ADI-R) has a milestone variable walked unaided, age (item 5), and a variable articulation at age 5 (item 32) that provides a measure of development of motor control of speech.

We examined correlations of the phenotypic features with the number of dn LGDs per child (typically 0 or 1, some 2, few 3) and displayed the strengths of their one-sided *P* values graphically as seen in Fig. 1*A* (see *Methods* for details). The DCDQ and VABS measures, as well as nvIQ, are skill levels, hence expected to be negatively correlated with dn LGD mutations (shown in red). Three measures (shown in the bottom rows of the figure) are inverse skill measures, age at first walking, speech articulation at age 5, and (problems with) well coordination, hence expected to be positively correlated with dn LGD mutations (shown blue). The correlations of all these measures with our primary measure of dn genetic damage are in the expected directions. The absolute correlations between genotype and MS measures tend to be low, on the order of 0.1 (*SI Appendix*, Fig. S1). However, due to the large size of available SSC data (n = 2,120), these correlations are statistically significant (P = 1.5E-4 for total DCDQ). Affected children with dn LGDs have decreased gross and fine motor skills, as well as delayed motor development, compared with affected children without these observed dn LGDs.

Because earlier literature has pointed to nvIQ as correlated with genetic lesions, we include this variable for comparison with the MS variables. Note that significant correlation with nvIQ is also seen (Fig. 1, *Top Left*, P = 4E-4).

Correlating MS and IQ Loss with LGD Targets. Not all dn LGD mutations necessarily disrupt critical gene functions. In fact, we estimate that in children on the spectrum, a little less than half of dn LGDs contribute to autism risk (2). First, not all LGD mutations are in fact disruptive; and second, not all gene targets are critical. Fortunately, the importance of a gene target can be weighted by evidence. Indeed, there are multiple ways to weight target importance: whether the gene is a recurrent target; whether the gene is under strong negative selective pressure; and according to the functional properties of its encoded product.

We call a gene a recurrent target if a dn LGD hits that gene in more than one affected child in the SSC. From previous work (2), recurrent targets are estimated 90% likely to be autism-risk genes. To determine whether measures of MS are correlated with the presence of recurrent LGD targets, beyond their correlation with dn LGDs in general, we restricted our study to only those children already affected with a dn LGD. Among these, we then counted the number of recurrent dn LGDs in each child, typically 0 or 1. Although there are only 57 recurrent LGDs out of a total of 352 LGDs, their correlations with DCDQ in this subset of the children has very strong additional significance (Fig. 1*B*, total DCDQ, P = 5E-8). Increase in significance is

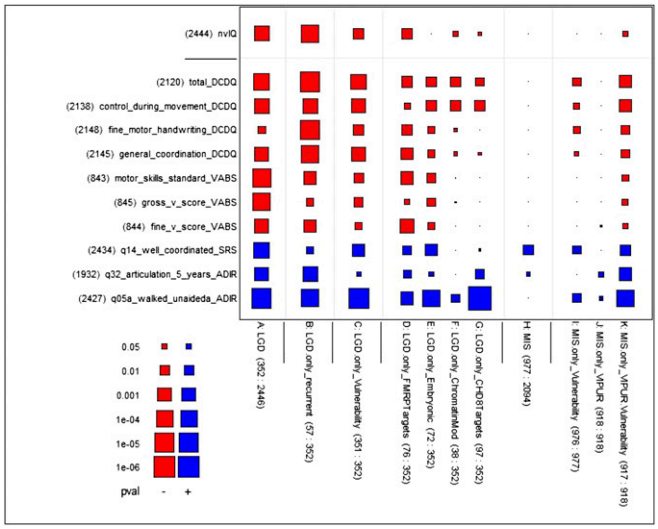

Fig. 1. Significance of correlations between measures of dn genetic damage and measures of motor skills and nvIQ of affected children. We used 11 measures of dn genetic damage shown as columns in the figure and 11 phenotypic measures (1 for nvIQ and 10 for MS extracted from 4 different phenotypic instruments) shown as rows. The genetic damage and phenotypic measures were defined on different subsets of the affected children in the SSC collection. To reflect this fact, we pasted counts to the labels as follows: "(100)" would mean the measure is defined on a subset of size 100, and "(20:100)" indicates in addition that of the 100 values the number of nonzeroes is 20; this second version is relevant for genetic variables that are counts of mutations in a child. We computed the correlation between each of the 11 genetic damage measures and the 11 phenotypic measures using only the children for which both the genetic damage and phenotypic measures were defined, and we tested whether the correlation was significantly different from 0. The resulting 11 by 11 table of *P* values is rendered graphically with rectangles whose size represents inversely the *P* value from the statistical test (large rectangle ~ small *P* value) and whose color represents the sign of the correlation (blue ~ positive, red ~ negative). The meaning of both the sizes and colors can be gleaned from the figure's key in the bottom left. A small dot is used when the correlation is strongly insignificant ($P \geq 0.10$). In a similar fashion, the related *SI Appendix*, Fig. S1 shows the underlying computed correlations. For more background about this type of display, see *Methods*. Measures of genetic damage: The primary measure of genetic damage (*A*) is defined as the number of de novo LGDs (0, 1, 2, or 3) identified in an affected child. The variables in *B*–*G* differentiate LGDs according to indications that they may be damaging; these variables are defined only for children who have at least one LGD. *B* is the number (0 or 1) of de novo LGDs in a child affected by genes with more than one de novo LGD in the SSC (recurrent genes). *C* is the sum of the vulnerability scores of the genes affected by de novo LGDs in the child. *D*–*G* are defined as the number (0 or 1) of de novo LGDs that fall in four gene functional classes that have previously been implicated in autism's etiology: FMRP target genes, embryonic genes, genes encoding chromatin modifiers, and CHD8 target genes. The remaining columns concern de novo missense mutations: *H* is the number of de novo missense mutations (0 up to 5) in an affected child, applied only to children without de novo LGD mutations, to prevent confounding with overpowering LGD effects. *I*, analogous to *C*, is the sum of vulnerability scores of genes affected by missense mutations in a child. *J* is the sum of VIPUR scores of missense mutations in a child. *K* is the product of vulnerability and VIPUR scores, exhibiting *P* values that neither score could achieve alone. Phenotypic measures: The top row represents nonverbal IQ (nvIQ). The other rows represent 10 different measures of motor skills available in the phenotypic database of the SSC. The labels are suffixed by abbreviations of the originating instruments: DCDQ, VABS, SRS, and ADI-R. See *Methods* for details.

found as well for nvIQ ($P = 8E-7$), consistent with what we have previously reported (2).

Another way to distinguish the severity of an LGD is by characterizing the genetic burden of its target in the human gene pool, a reflection in part of the action of purifying selection. In ref. 7, we measured the frequency that an LGD variant in a given gene is observed in a large unaffected population. We ranked genes by their frequency of carrying an LGD in that population, normalized by the length of that gene. Those genes with a low burden were considered by us to be highly vulnerable. The data on genetic burden is far from complete, because the sequence databases are insufficient to characterize most genes for their vulnerability, especially the ones encoding smaller products. Nevertheless, in a previous study we still found that affected children in the SSC with dn LGDs in highly vulnerable genes had significantly lower IQ than in affected children with LGDs in less vulnerable genes (7). By comparing the number of LGDs by vulnerability score in affected versus unaffected individuals from the SSC, we can demonstrate that the ability of the vulnerability score to discriminate these two groups is concentrated in higher vulnerability scores (*SI Appendix*, section 1). In this report, therefore, we transform the gene vulnerability rank by taking the negative logarithm of the normalized rank of gene vulnerability (*Materials and Methods*), yielding a gene vulnerability score that spreads out more informative scores and compresses less informative scores. Using the sum of the scores of the dn LGD target genes within a child, and restricting to children with dn LGDs, we find strikingly significant correlations with motor skills (Fig. 1*C*).

The severity of a mutation might also depend on the functional class of its target gene. We consider here dn LGD mutations in three sets of genes, enriched as targets of disruptive mutation in children with ASD and earlier examined by us (2): FMRP (fragile X mental retardation protein) target genes, whose transcripts interact with the fragile X protein (Fig. 1*D*) (2, 24, 25); embryonic genes, which are genes expressed in the brain of the fetus, but strongly down-regulated upon birth (Fig. 1*E*) (2, 26, 27); and chromatin-modulating genes (Fig. 1*F*) (28, 29). We add a fourth set, the genes regulated by CHD8, the most frequent target gene for dn LGD mutation in ASD (Fig. 1*G*) (30). For this analysis, we again consider only those autistic individuals that have dn LGD mutations, so as not to confound the analysis with correlation due to the LGD itself. As reported before, dn LGD mutations in the FMRP target genes are associated with lower IQ (2). So, too, we find that they are significantly associated with reduced motor skills. The dn events in the other functional categories are also associated with reduced MS but with somewhat less significance. Across all of the data, a particularly significant correlation is seen between presence of a dn LGD in a CHD8 target and age of first walking. In dn mutations on the CHD8 gene itself were proposed as associated with "a subtype of autism early in development," (29) though not specifically with motor delays.

Analysis of dn Missense Mutations. There are many more dn missense mutations than dn LGD mutations. Based on what we call ascertainment bias, we judge that only about 10% of these contribute to simplex autism, in contrast to about 50% for dn LGD (2, 5). Not surprisingly, even excluding children with dn LGDs, and counting each child for number of dn missense mutations, we see no significant correlation with nvIQ, and only significant correlation with well-coordinated among the MS variables (Fig. 1*H*). However, evidence of the contribution of dn missense mutations does emerge if they are weighted by the gene vulnerability scores of their targets (Fig. 1*I*) (7). With that measure, we observe significant correlation to most MS variables. We note that association of these weighted mutations is not observed with nvIQ, suggesting milder effects of dn missense mutations by affecting mostly MS and leaving intellectual function mostly intact. The next section will provide context for this observation.

In an attempt to further strengthen the discrimination among dn missense mutations we used a method called VIPUR (Variant Interpretation and Prediction Using Rosetta), which predicts the likelihood that a mutation has a deleterious effect on function (31). The two VIPUR-related variables shown in the figures are restricted to a subset of dn missense mutations for which VIPUR scores are available, aggregated by summation for each affected child without dn LGDs. The raw VIPUR score alone shows no significant correlation with MS variables (SRS item 14) and none with IQ (Fig. 1*J*). However, a new score obtained by multiplication of VIPUR and the vulnerability score (*Methods*) leads to more significant associations than either alone (Fig. 1*K*, cf. Fig. 1 *I* or *J*). Thus, VIPUR in combination with the gene vulnerability score helps to assess mutational damage. These results should be seen as preliminary and exploratory, requiring independent replication.

Association Between MS and IQ. It is clear from the SSC phenotype data that IQ and MS are themselves correlated (corr = 0.36, n = 2,365). One should therefore ask whether significant correlations with dn mutations survive when IQ and MS are adjusted for each other (*Methods*). We display the results in Fig. 2, with the same column structure as in Fig. 1, where MS variables are adjusted by nvIQ, sex, and age; and nvIQ (for comparison) is adjusted by total_DCDQ, sex, and age. Correlations generally survive adjustment, although with somewhat attenuated significances. Attenuation after mutual adjustment is to be expected due to the partial positive correlation between IQ and MS, but the main conclusion is that the association of MS with mutation variables cannot be reduced to IQ, or vice versa. (For a corresponding display of correlations as opposed to P values, see *SI Appendix*, Fig. S2.)

The association between nvIQ and MS deserves closer examination because it is insufficiently characterized by a plain correlation coefficient. Indeed, the association cannot be simply described as a tendency to pair high with high and low with low values between nvIQ and MS. Rather, as can be seen in Fig. 3,

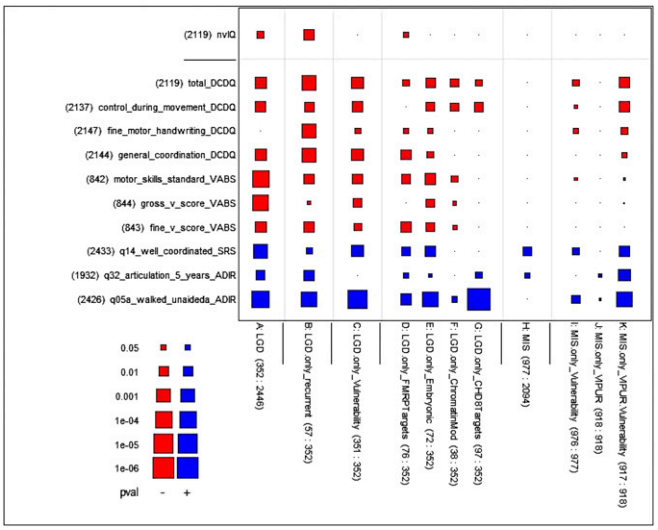

Fig. 2. Significance of the correlation between measures of genetic damage and adjusted measures of motor skills and IQ of affected individuals. This figure is similar to Fig. 1 (see legend for details), but the phenotypic measures have been adjusted as follows: the 10 motor skills measures (suffixed DCDQ, VABS, SRS, ADIR) are adjusted for nvIQ, sex, and age; nvIQ (*Top*) is adjusted for total_DCDQ, sex, and age. The main result is that the significance of the correlations largely survives these adjustments. This is evidence that the association of motor skills with the genetic variables cannot be reduced to the correlation between motor skills and nvIQ. (See *SI Appendix*, Fig. S2 for a similar graph showing the underlying adjusted correlations.)

the combination of low nvIQ with high (normal) MS is rare, but the combination of high (normal) nvIQ with low MS is not. This can be put differently as follows: MS differentiate within the normal nvIQ range, while nvIQ differentiates within the low MS range.

The same fact can be rendered differently by dichotomizing both nvIQ and MS: we define nvIQ < 70 as low IQ; and total_DCDQ < 50 as low MS. A standardized score for nvIQ has an average of 100 and a SD of 15. The value 70 for nvIQ is a conventional threshold for intellectual disability (ID), defined as two SDs below the mean in general populations. For total_DCDQ we use the raw value 50 as a loose lower bound on the "normal" range. Following (32), the recommended age-dependent thresholds for the normal range are total_DCDQ ≥47 (age < 8 y), total_DCDQ ≥ 56 (8 y ≤ age < 10 y), and total_DCDQ ≥ 58 (10 y ≥ age), respectively. By this recommendation, 83% of affected SSC children have deficient MS, consistent with the 80–90% range of refs. 18 and 33, whereas just 25.5% have diminished nvIQ (ID). Only 15.7% of affected children are in the normal range for both MS and nvIQ.

Fig. 3 gives a graphical rendition of dichotomization, resulting in four sectors denoted **A**, **B**, **C** and **D**, shown with a cross-hair at coordinates (70,50). We can order the three enriched sectors according to phenotypic severity, and label them as follows:

$$\mathbf{A}\ [\text{nvIQ} \geq 70, \text{total\_DCDQ} \geq 50]\ (\text{mild}) \rightarrow$$
$$\mathbf{B}\ [\text{nvIQ} \geq 70, \text{total\_DCDQ} < 50]\ (\text{moderate}) \rightarrow$$
$$\mathbf{C}\ [\text{nvIQ} < 70, \text{total\_DCDQ} < 50]\ (\text{severe}).$$

The remaining sector **D** [nvIQ < 70, total_DCDQ ≥ 50] is depleted by a factor of 4.6 when comparing column ratios (*SI Appendix*, Table S2). In terms of severity, the phenotypic ordering of the three major sectors may be symbolically written as **A < B < C**.

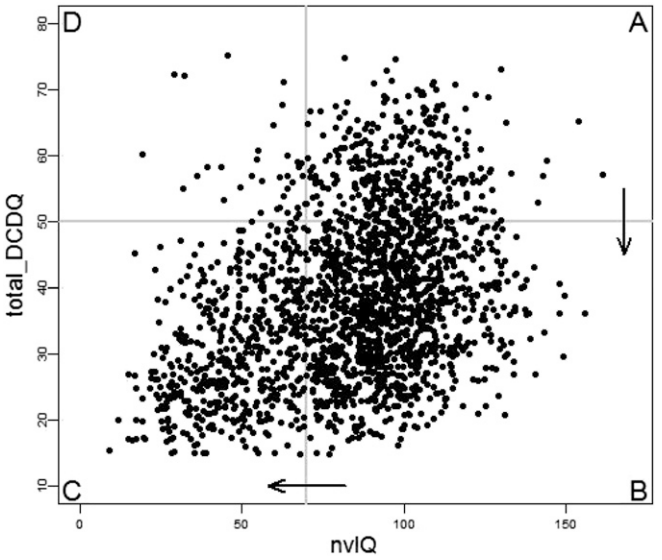

Fig. 3. Relationship between IQ and motor skills. The scatterplot shows nvIQ and total_DCDQ for $n = 2{,}119$ affected children with available exome data. The gray vertical lines show the cutoffs used to dichotomize the two measures (see the text for justification of the particular cutoffs). The four quadrants of the graph are labeled clockwise with letters A through D. Quadrant D is significantly underpopulated compared with what is expected under an assumption that the two measures are independent. Ignoring quadrant D, the arrows demonstrate increasing phenotypic severity (as a function of both nvIQ and total_DCDQ) between adjacent quadrants, with **A < B < C**.

Likewise, there exist significant differences between the mean vulnerability scores of LGD and missense mutation gene targets in affected individuals in sectors **A**, **B** and **C**. That is, the mean vulnerability score is significantly greater in sector **B** compared with sector **A** ($P = 0.02$ for LGDs alone and $P = 0.005$ for LGDs and missense together); in sector **C** compared with sector **B** ($P = 0.02$ for LGDs alone and $P = 0.003$ for LGDs and missense together). Therefore, in terms of mutational severity, a shorthand to characterize dn mutations by the vulnerability score of their genes, the ordering of sectors may also be written as **A < B < C**.

Implications of the Associations Between MS, IQ, and Mutational Severity. We hypothesize that the increase in impairment from sector **A** to **B** and from **B** to **C** stems from a corresponding increase in mutational severity in the target genes, and suggest the following: Mild mutational severity is unlikely to affect MS or nvIQ; moderate mutational severity is more likely to affect MS and less so nvIQ (sector **B**); and severe mutational severity is likely to affect both MS and nvIQ (sector **C**).

In other words, when the LGD or missense target has a high vulnerability score, then both diminished MS and nvIQ are more likely, and when the target has a somewhat lower vulnerability score, then the effect is more biased toward diminished MS than diminished nvIQ. We see further evidence for the hypothesis that moderate mutational severity affects primarily MS by comparing the top two rows of Fig. 1: There are no genetic variables in our collection that are more significantly associated with nvIQ than with total_DCDQ. In addition, all functional classes of dn LGD mutations (other than FMRP) and the scored dn missense mutations have more significant associations to total_DCDQ than nvIQ.

Absence of Association with Core ASD Variables. We turn to a set of core ASD variables that characterize the conventional autism phenotype consisting of deficits in social interaction as well as restricted and repetitive behaviors. In the SSC, principal investigators had previously selected core descriptive variables from several primary instruments. From these variables we subselected those originating from the following instruments: Autism Diagnostic Interview, Autism Diagnostic Observation Schedule (ADOS), Repetitive Behavior Scale (RBS), Aberrant Behavior Checklist (ABC), and Social Responsiveness Scale (*t* scores), for a total of 12 variables.

We first observe that these core ASD variables strongly correlate with both nvIQ and MS. Indeed, according to the last two columns of Fig. 4, most *P* values are beyond conventional levels of statistical significance. (According to *SI Appendix*, Fig. S4, some of the correlations approach 0.5 in magnitude.) This observation suggests that core ASD variables might also be correlated with de novo mutational severity. However, we were unable to find consistent associations. As can be seen from Fig. 4, strongly significant associations with genetic variables are largely absent, and some that approach the level 0.05 are in the wrong direction. The one exception is the variable SRS Parent *t* Score, for which the correlation disappears if it is adjusted for nvIQ. We discuss the possible interpretations of the lack of correlation between core variables and dn genetic damage below.

Discussion

This study is part of our continuing attempt to link damaging de novo mutations to broad neuropsychiatric effects in children on the autism spectrum. We have done this both to define the substructure of the syndrome and to evaluate which events are likely to contribute to the disorder. Earlier studies had established a link between damaging mutation and diminished nvIQ. The present study establishes a significant association between damaging mutation and impaired motor skills. Our method is to correlate measures of broad neurological function (nvIQ, MS)

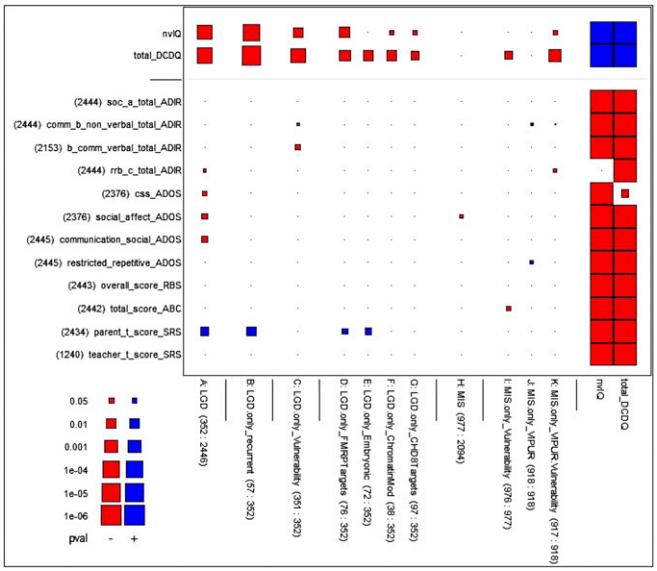

Fig. 4. Absence of association between measures of genetic damage and measures of core ASD phenotype. In this figure, all but the top two rows represent core ASD measures drawn from the ADI-R, ADOS, RBS, ABC, and SRS instruments (see *Methods* for explanations). The conclusion is that these measures largely lack significant correlations with measures of dn genetic damage. Some *P* values that approach 0.05 correspond to correlations that have the wrong sign (shown red) as the core ASD variables measure behavioral deficiency, hence should be positively correlated with mutational severity. The figure also shows nvIQ and total_DCDQ in the two top rows to provide a comparison what significant correlations would look like. The two rightmost columns show nvIQ and total_DCDQ as well to give evidence of their strongly significant correlations with the core ASD measures.

and core ASD phenotypes with de novo genetic damage. For both we rely on the SSC, which provides an abundance of phenotypic measures as well as extensive de novo mutation data. From the latter we use information about the type of mutation (LGD or missense) and its genetic target (recurrence, vulnerability, and functional class).

Unlike intellectual ability, motor skills do not have a single standard of measurement. We therefore used all of the available motor skill measures from the SSC, often broken into subscales and originating in multiple distinct instruments. Whether measuring fine or gross motor skills, or motor-related developmental milestones, with few exceptions, we see a remarkably consistent pattern of significant correlation with de novo genetic variables. It is this consistency, as well as statistical significance with any one measure, that gives us confidence in our conclusions.

Our present study on motor skills recapitulates conclusions from our previous study of nvIQ by extending them to motor skills. First, mutational severity matters. Autistic children with LGD mutations are more likely to have impaired MS than children without them. Second, autistic children with dn mutations in recurrent gene targets are more likely to have impaired MS. Third, children have more severely attenuated MS if they suffer LGD mutations in genes that are vulnerable (i.e., have reduced deleterious genetic burden in the human population) or, fourth, share certain functional properties.

This study goes beyond our previous study of damaging de novo mutations and nvIQ. We previously had not observed correlations between nvIQ and dn missense mutations. The association of impaired motor skills with missense mutation has marginal significance at best. However, when the missense targets are weighted by their vulnerability score, the association with MS becomes far more significant. When the missense is further weighted by VIPUR, one of several available methods to judge the severity of a missense mutation, the association becomes quite substantial.

However, even when we weight mutation and target as above, we do not see a highly significant correlation between nvIQ and missense mutation. A priori, missense should be less damaging than the premature termination caused by LGDs, for which the association with diminished nvIQ is very strong. These results thus suggest that loss of motor skills is a more sensitive indicator of dn genetic damage than is loss of nvIQ, and this conclusion is consistent with the severity arrows of Fig. 3. This inference is also consistent with the observation that one can have diminished MS without diminished nvIQ, but the reverse is hardly seen.

The nvIQ and MS as measured by the DCDQ are correlated, but far from redundant: their relation is not simply linear; they differ in gender bias (*SI Appendix*, Table S1); they have different patterns of correlations with genetic variables; and importantly, the correlations of MS and nvIQ with dn genetic damage each survive when adjusting one for the other. The correlations of genetic variables with MS are more consistent than with nvIQ. Furthermore, nvIQ differentiates the low functioning range of MS, while the MS differentiates the normal IQ range. Moderate dn genetic damage tends to affect MS more and nvIQ less, while severe dn genetic damage tends to affect both MS and nvIQ. The most extreme example of this is the signal from children with dn LGDs in the gene targets of the CHD8 chromatin modifier. These children show very strong impairment in age of first walking, but much less significant correlation with nvIQ.

The links between damaging mutations described here should reinforce the need to routinely include an age-appropriate evaluation of motor skills in the assessments of ASD. Motor skills are simple to measure. Even a single questionnaire item may give some indication of motor skill deficiency, as illustrated by SRS item 14, which refers to general motor control and coordination, and ADI-R item 32 that refers to speech articulation. Even the DCDQ instrument is relatively simple, based on just 15 items. Specific motor skills are used to define common developmental milestones for infants, one of which makes a powerful appearance in our battery of motor skill variables, the age of first walking unaided (see bottom row of Figs. 1 and 2). Age of first walking is consistently one of the most reliable parent report measures (34). Other related phenotypes could be examined for links to dn genetic damage, such as more detailed examination of motor function or sensorimotor development used in evaluation of children with neurodevelopmental disorders. We should point out that currently the MS assessment is made subjectively by parents, but that tests can be devised based on more objective criteria and judged by a neutral observer. Thus, MS might be more readily and objectively monitored than most other clinically emphasized cognitive functions, and be an especially important endpoint when investigating genetic lesions in model organisms or screening humans for response to experimental therapies (35–37).

Both nvIQ and MS significantly correlate with the core phenotypes used to make the ASD diagnosis, including those measuring social communication skills (Fig. 4 and *SI Appendix*, Fig. S4). It is natural to view nvIQ as interacting with these abilities, but the interactions with MS is less obvious. Delay in developing age-appropriate motor skills and body language may lead to further social isolation. But, the connection between MS and ASD phenotypes may be more direct. For example, the brain region most closely associated with gross and fine motor control is the cerebellum, and it is now appreciated that the cerebellum is directly involved in cortical development and that cerebellar lesions during the third trimester and neonatal period are associated with the development of affective disorders (38–43).

Finally, we examined the correlation between damaging mutations and the core ASD variables, including social deficits and restricted and repetitive behaviors. Although damaging mutations correlate with both nvIQ and MS in those with ASD, and

nvIQ and MS correlate very significantly with core phenotypes, it does not follow that damaging mutations will correlate with core ASD features. The correlations of mutations with nvIQ and MS, although significant, are weak in absolute terms, and transitivity in correlation does not mathematically follow. In fact, we find that the correlation of damaging mutation with core ASD is inconsistent and weak at best. The little correlation that is observed vanishes when we adjust for nvIQ.

The lack of correlation, nevertheless, merits speculation. Our first thought was that this absence of significant association could be explained by the use of some core ASD variables in the ascertainment of children with ASD in the SSC, causing truncation of the variable ranges and resulting in less significant associations. A closer examination showed, however, that this explanation is most likely wrong (*SI Appendix*, section 2). We propose two other explanations, not mutually exclusive: (*i*) De novo genetic damage, when severe, may impair compensation for mild deficits in the core variables, deficits that on their own would not lead to a diagnosis (44). This effect may increase the number of children with relatively low core severity among those with dn damage. (*ii*) In the absence of de novo genetic damage, social deficiencies and restricted and repetitive behaviors perhaps arise from shared ancestral variants transmitted from parents. These variants may not be under as strong negative selection as severe dn genetic damage, and project less onto the sphere of cognitive function (motor skills and IQ), and more onto the sphere of human social behavior and communication.

After our manuscript was ready for submission, another manuscript appeared (22) reporting overlapping but not identical conclusions using the genotypic and phenotypic data from the SSC. The major difference between our approach and theirs is that they match children on nvIQ, which results in the counterintuitive conclusion that affected children with dn LGD damage have a milder form of ASD. We show in *SI Appendix*, section 10 that this is an artifact of the matching design. In addition, they consider just one genetic variable (LGDs) and one motor development variable with emphasis on core ASD variables; whereas, we analyze a wide array of motor skills variables, consider missense mutations, quantify the parameters of genetic damage, and reach higher significance by looking at more individuals.

Methods

The present study is based on the Simons Simplex Collection (12), which has data for 2,760 families that have a single child affected by ASD. Of these families, 2,280 have an unaffected child as well, but for the most part, we are only concerned with the affected children. Among them, 2,446 have exome sequencing data available that resulted in the identification of 3,403 de novo mutations of all types (2). (The 1,836 unaffected children with exome data have 2,288 identified de novo mutations among them.) Unlike case-control approaches that compare affected and unaffected children, ours is a study of association between phenotype and genotype variables among affected children only. The premise, which could have been wrong, is that the affected children in the SSC have sufficient phenotypic variation to allow the discovery of statistically significant correlations between high and low levels of a scored behavioral phenotype (such as nvIQ, MS, and core ASD variables) on the one hand, and genotypic events (such as different classes of de novo mutation and the characteristics of their target genes) on the other hand. As shown above, convincing correlations exist for nvIQ and MS variables, but not for most core ASD phenotypes.

The following is the list of MS variables shown by their names in the SSC tables and also conveys the meaning of the scales. Consistency of correlation across diverse measures of MS from multiple instruments strengthens our confidence that the conclusions are not measurement artifacts. The variables in the Developmental Coordination Disorder Questionnaire (DCDQ) are control_during_movement, fine_motor_handwriting, general_coordination, and total. The last is the summary scale; the preceding three are subscales formed from a pool of 15 items scored on a 5-point Likert scale. For these variables, high values stand for high achievement.

The Vineland Adaptive Behavior Scales II (VABS-II) uses fine_v_score, gross_v_score, and motor_skills_standard. Again, the last is the summary scale; the preceding two are subscales. Unlike the other phenotype variables, which are available for ages 4–18 y (48–216 mo), the VABS-II variables exist only for children up to age 7.5 y (cases with higher age are outliers that were removed). For these variables, high values stand for high achievement.

The Social Responsiveness Scale (SRS) uses q14_well_coordinated. This single item out of 65 SRS items measures coordination problems on a scale from 0 to 3. Contrary to the meaning suggested by the name of the item, this is a severity measure with meanings 0 = no coordination problems, 3 = severe coordination problems.

Autism Diagnostic Interview Revised (ADI-R) uses q05a_walked_unaideda: Age of walking unaided, in months (ranging from 7 to 72), but transformed with a double logarithm due to an extremely right skewed distribution: log(log(. . .)); q32_articulation_5_years: Problems with motor control of speech at age 5, on a scale from 0 to 3. For these variables, high values stand for higher levels of problems.

Cognitive functioning is measured by nonverbal IQ (nvIQ), in agreement with past literature. Verbal and full-scale IQ are not used. The instruments used for nvIQ assessment are the following, with counts: 977 Differential Ability Scales, second edition (DAS-II) Early Years; 1178 DAS-II - School Age; 188 Mullen Scales of Early Learning; 61 Wechsler Abbreviated Scale of Intelligence (WASI); 41 Wechsler Intelligence Scale for Children, fourth edition (WISC-IV).

Core-ASD variables were selected from the SSC table Core Descriptive Variables (CDV), which contains a set of demographics, measures, and diagnoses previously deemed clinically relevant. We subselected 12 variables from 5 instruments. (*i*) ADI-R: adi_r_soc_a_total, adi_r_comm_b_non_verbal_total, adi_r_b_comm_verbal_total, adi_r_rrb_c_total; (*ii*) ADOS: ados_css, ados_social_affect, ados_communication_social, ados_restricted_repetitive; (*iii*) RBS-R: rbs_r_overall_score; (*iv*) Aberrant Behavior Checklist (ABC): abc_total_score; and (*v*) SRS: srs_parent_t_score, srs_teacher_t_score.

These variables were preselected on substantive grounds without datamining. However, the interested reader may indulge in datamining by perusing the numerous figures in *SI Appendix*, sections 4–8, which show associations for the complete instruments, both summary measures and underlying items. These figures are provided as exploratory displays and as confirmations and qualifications of the finding that the core ASD phenotype has at most a tenuous link to genetics as reflected by de novo mutations. The instruments and tables we show in the *SI Appendix* are as follows: DCDQ, VABS-II, CDV, SSC Commonly Used Variables (CUV), Childhood Behavior Checklist, ages 2–5 and 6–18 (CBCL-2–5 and CBCL-6–18), SRS, ABC, RBS-R, ADI-R, ADOS-1, ADOS-2, ADOS-3 (three modules of ADOS), Social Communication Questionnaire Life (SCQ-LIFE), as well as a table of demographic variables. Note: The ABC instrument is not generally considered a core-ASD instrument; the main reason for its inclusion was its availability in the SSC.

Genetic variables for exome-sequenced affected children were obtained from published sources as follows: The list of de novo mutations is from supplementary table two in ref. 2. It characterizes each mutation by the location on the genome, the "effectGene" and the "effectType" (among other things). Among effectTypes we used the following: synonymous, missense, as well as six types that jointly make up the LGD classification: splice-site, nonsense, noStart, noEnd, frame-shift, and no-frame-shift-newStop.

The functional classification of genes is from supplementary table seven in ref. 2. LGD mutations where subdivided according to whether the effectGeneis classified as FMRPTargets, Embryonic, or a ChromatinModifiers (remaining classifications were not used, some because of low counts, others because of a priori unlikely effects). One more classification was used to subdivide LGD mutations according to their effectGene: CHD8Modifiers, comprising a list of genes published in supplementary table 1 of ref. 30.

Gene vulnerability ranks are from ref. 7. Instead of dichotomizing the scores on a threshold and comparing the resulting groups of high and low gene vulnerability, we instead transform the ranks by normalizing them to values between 0 and 1, and then applying a negative logarithm. The resulting gene vulnerability scores have a roughly exponential distribution, the purpose being to spread out the most vulnerable genes to the unlimited positive range and shrinking nonvulnerable genes to the near-zero range. This processing gives highly vulnerable genes an opportunity to differentiate themselves with high values while genes with little vulnerability are made nearly indistinguishable by piling up their values near zero. This type of processing injects quantitative differentiation where it is needed and obviates the search for meaningful thresholds.

VIPUR scores were obtained from the authors of ref. 31. We used the highest score version of VIPUR. Similar to gene vulnerability scores, we did not use the raw VIPUR scores but a transformation thereof, again obtained by normalizing their ranks to [0,1] and applying a negative logarithm,

resulting in a roughly exponential distribution that spreads out the high scores and shrinks the low scores.

Descriptive tables and plots for nvIQ, MS variables, and genetic variables are shown in *SI Appendix*, section 9. Similar information for the many instruments shown in *SI Appendix*, sections 3–8 are not provided due to shear volume.

Statistical measurement of association between two variables was done by forming Pearson correlation coefficients. These were used for uniformity even in nonstandard cases, in particular when one or both variables were binary groupings coded as 0–1 dummy variables: If one variable is quantitative and the other binary, the correlation coefficient is algebraically equivalent to the *t* statistic for testing the difference of means; if both variables are binary, the correlation coefficient is algebraically equivalent to the test statistic of Fisher's exact test of independence. See ref. 45 for a more detailed discussion. However, we also recreated Figs. 1, 2, and 4 with Spearman rank correlations and observed no discernible differences in the patterns of association.

Presentation of correlation tables is in graphical form as blockplots (45). The reason is that large tables of numeric values are difficult to parse visually. Furthermore, the multidigit precision of numeric tables is not only useless but delusional because it suggests accuracy where none exists. Graphical presentation provides not only defensible accuracy but lends itself to visual pattern recognition, in particular patterns of consistency of correlations across rows and columns. Actually, more important than correlations are their statistical significances in terms of *P* values. The most important figures are indeed blockplots of *P* values, rendered such that large blocks indicate strong statistical significance. The visual estimation of the order of magnitude of *P* values (as well as correlations) is helped by a key in the bottom left of the figures. As for color coding, we use blue to indicate a positive association and red a negative association; color coding is also used in blockplots of *P* values even though these only reflect statistical significance without orientation. Finally, we note that the blockplots are superior to heat maps because the sizes of rectangles provide much more precise and more impactful visual cues than color scales.

Statistical Multiplicity. Presenting large numbers of correlations and their *P* values might raise questions of statistical inference. A more conventional presentation would have condensed the findings into a handful of *P* values, for example, by focusing on total_DCDQ alone among MS variables. The reason for choosing an expansive visual presentation of large numbers of *P* values is to convey the consistent patterns of statistically significant association across groups of variables, in particular the several measures related to motor skills. Such consistency is nontrivial: Although motor skill variables are correlated with *P* values beyond conventional levels of statistical significance, the correlations are far from perfect, no higher than 0.5 between DCDQ and VABS-II, for example (*SI Appendix*, section 9.4). This limits their shared variation to 25%, leaving ample room for conflicting correlations with the comparatively weak signal from the mutation variables. That this is largely not happening is confirmation of the nontrivial consistency of association between motor skill and dn genetic variables. The many *P* value displays in the *SI Appendix*, sections 4–6 are shown for two reasons: (*i*) to back up and qualify the notion that a vast majority of core ASD variables do not show consistent correlations with dn mutation variables, other than those mediated by nvIQ and MS, and (*ii*) to allow readers to do their own exploratory hypothesis generation, using *P* values heuristically rather than inferentially.

Adjusted Variables. In Fig. 2 we (*i*) adjust nvIQ for total_DCDQ, sex, and age; and we conversely adjust the MS variables for nvIQ, sex, and age. Adjustment means in case that nvIQ is subjected to a linear regression with total_DCDQ, sex, and age as regressors and that nvIQ is then replaced by its residuals from this regression. These residuals are uncorrelated with total_DCDQ, sex, and age. If one observes correlations of adjusted nvIQ with genetic variables, it reflects association that cannot be accounted for by total_DCDQ and/or sex and/or age. (Detail: In adjusting for age, we added a linear spline term with knot at 9 y of age to account for potential nonlinearity due to the transition from childhood to adolescence. The knot location 9 was chosen a priori, not by data mining.) In computing *P* values we converted the correlations to *t* statistics by well-known formulas and used properly reduced degrees of freedom in the *t* distributions according to the number of estimated parameters in the adjustment.

Questions of Confounding. Observational studies such as the present one can result in flawed attribution of cause. Although we formulate all results in terms of association rather than causation, the implied understanding is that the genetic variables describe aspects of causal mechanisms for the phenotype. To reduce the chance of confounding the genetic variables with trivializing factors such as demographics, we provide in the *SI Appendix*, section 7, *P* value displays for demographics, and in *SI Appendix*, section 8 displays for genders separately and with only Caucasian ethnicities to avoid potentially confounding with genetic variables. From the demographics in *SI Appendix*, section 7 we learn that, for example, IQ is much more associated with demographics than MS (as measured by total_DCDQ), thus reducing the chances of confounding for the latter. We also learn the known fact that dn missense mutations are more related to fathers' age than mothers'.

ACKNOWLEDGMENTS. We thank Jim Simons for prompting the research and providing valuable guidance throughout; Gerry Fischbach for a detailed review of the manuscript and suggestions for improvements; Louis Reichardt for providing additional pointers; and Richard Bonneau and Evan Baugh for explanations of the VIPUR method and sharing missense mutation predictions. We also thank all the families at the participating SSC sites, as well as the principal investigators (A. L. Beaudet, R. Bernier, J. Constantino, E. H. Cook Jr., E. Fombonne, D. Geschwind, D. E. Grice, A. Klin, D. H. Ledbetter, C. L. Martin, D. M. Martin, R. Maxim, J. Miles, O. Ousley, B. Peterson, J. Piggot, C. Saulnier, M. W. State, W. Stone, J. S. Sutcliffe, C. A. Walsh, and E. Wijsman) and the coordinators and staff at the SSC sites for the recruitment and comprehensive assessment of simplex families, and the Simons Foundation Autism Research Initiative (SFARI) staff for facilitating access to the SSC. This work was supported by SFARI Grants SF235988 (to M.W.), SF362665 (to I.I.), and SF448357 (to A.B. and A.M.K.).

1. American Psychiatric Association; DSM-5 Task Force (2013) *Diagnostic and Statistical Manual of Mental Disorders: DSM-5* (American Psychiatric Association, Washington, DC), 5th Ed, p xliv.
2. Iossifov I, et al. (2014) The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 515:216–221.
3. Sebat J, et al. (2007) Strong association of de novo copy number mutations with autism. *Science* 316:445–449.
4. Sanders SJ, et al.; Autism Sequencing Consortium (2015) Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron* 87:1215–1233.
5. De Rubeis S, et al.; DDD Study; Homozygosity Mapping Collaborative for Autism; UK10K Consortium (2014) Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 515:209–215.
6. Levy D, et al. (2011) Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron* 70:886–897.
7. Iossifov I, et al. (2015) Low load for disruptive mutations in autism genes and their biased transmission. *Proc Natl Acad Sci USA* 112:E5600–E5607.
8. Krumm N, et al. (2015) Excess of rare, inherited truncating mutations in autism. *Nat Genet* 47:582–588.
9. Ye K, et al. (2017) Measuring shared variants in cohorts of discordant siblings with applications to autism. *Proc Natl Acad Sci USA* 114:7073–7076.
10. Gaugler T, et al. (2014) Most genetic risk for autism resides with common variation. *Nat Genet* 46:881–885.
11. Klei L, et al. (2012) Common genetic variants, acting additively, are a major source of risk for autism. *Mol Autism* 3:9.
12. Fischbach GD, Lord C (2010) The simons simplex collection: A resource for identification of autism genetic risk factors. *Neuron* 68:192–195.
13. O'Roak BJ, et al. (2012) Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 485:246–250.
14. Landa R, Garrett-Mayer E (2006) Development in infants with autism spectrum disorders: A prospective study. *J Child Psychol Psychiatry* 47:629–638.
15. Provost B, Lopez BR, Heimerl S (2007) A comparison of motor delays in young children: Autism spectrum disorder, developmental delay, and developmental concerns. *J Autism Dev Disord* 37:321–328.
16. Paquet A, Olliac B, Golse B, Vaivre-Douret L (2016) Current knowledge on motor disorders in children with autism spectrum disorder (ASD). *Child Neuropsychol* 22:763–794.
17. Mosconi MW, Sweeney JA (2015) Sensorimotor dysfunctions as primary features of autism spectrum disorders. *Sci China Life Sci* 58:1016–1023.
18. Hilton CL, Zhang Y, Whilte MR, Klohr CL, Constantino J (2012) Motor impairment in sibling pairs concordant and discordant for autism spectrum disorders. *Autism* 16:430–441.
19. Fournier KA, Hass CJ, Naik SK, Lodha N, Cauraugh JH (2010) Motor coordination in autism spectrum disorders: A synthesis and meta-analysis. *J Autism Dev Disord* 40:1227–1240.
20. Mostofsky SH, Burgess MP, Gidley Larson JC (2007) Increased motor cortex white matter volume predicts motor impairment in autism. *Brain* 130:2117–2122.
21. Dziuk MA, et al. (2007) Dyspraxia in autism: Association with motor, social, and communicative deficits. *Dev Med Child Neurol* 49:734–739.
22. Bishop SL, et al. (2017) Identification of developmental and behavioral markers associated with genetic abnormalities in autism spectrum disorder. *Am J Psychiatry* 174:576–585.

23. Schoemaker MM, et al. (2006) Evaluation of the developmental coordination disorder questionnaire as a screening instrument. *Dev Med Child Neurol* 48:668–673.
24. Darnell JC, et al. (2011) FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* 146:247–261.
25. Iossifov I, et al. (2012) De novo gene disruptions in children on the autistic spectrum. *Neuron* 74:285–299.
26. Kang HJ, et al. (2011) Spatio-temporal transcriptome of the human brain. *Nature* 478: 483–489.
27. Voineagu I, et al. (2011) Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* 474:380–384.
28. McCarthy SE, et al. (2014) De novo mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. *Mol Psychiatry* 19:652–658.
29. Bernier R, et al. (2014) Disruptive CHD8 mutations define a subtype of autism early in development. *Cell* 158:263–276.
30. Cotney J, et al. (2015) The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. *Nat Commun* 6:6404.
31. Baugh EH, et al. (2016) Robust classification of protein variation using structural modelling and large-scale data integration. *Nucleic Acids Res* 44:2501–2513.
32. Wilson BN, et al. (2009) Psychometric properties of the revised developmental coordination disorder questionnaire. *Phys Occup Ther Pediatr* 29:182–202.
33. Green D, et al. (2009) Impairment in movement skills of children with autistic spectrum disorders. *Dev Med Child Neurol* 51:311–316.
34. Hus V, Taylor A, Lord C (2011) Telescoping of caregiver report on the autism diagnostic interview–Revised. *J Child Psychol Psychiatry* 52:753–760.
35. LeBarton ES, Iverson JM (2013) Fine motor skill predicts expressive language in infant siblings of children with autism. *Dev Sci* 16:815–827.
36. LeBarton ES, Iverson JM (2016) Associations between gross motor and communicative development in at-risk infants. *Infant Behav Dev* 44:59–67.
37. Lloyd M, MacDonald M, Lord C (2013) Motor skills of toddlers with autism spectrum disorders. *Autism* 17:133–146.
38. Wagner MJ, Kim TH, Savall J, Schnitzer MJ, Luo L (2017) Cerebellar granule cells encode the expectation of reward. *Nature* 544:96–100.
39. Wang SS, Kloth AD, Badura A (2014) The cerebellum, sensitive periods, and autism. *Neuron* 83:518–532.
40. Schmahmann JD (2013) Cerebellar cognitive affective syndrome and the neuropsychiatry of the cerebellum. *Handbook of the Cerebellum and Cerebellar Disorders*, eds Manto M, Schmahmann JD, Rossi F, Gruol DL, Koibuchi N (Springer, Dordrecht, The Netherlands), pp 1717–1751.
41. Limperopoulos C, et al. (2014) Injury to the premature cerebellum: Outcome is related to remote cortical development. *Cereb Cortex* 24:728–736.
42. Bolduc ME, et al. (2012) Regional cerebellar volumes predict functional outcome in children with cerebellar malformations. *Cerebellum* 11:531–542.
43. Messerschmidt A, et al. (2008) Disrupted cerebellar development in preterm infants is associated with impaired neurodevelopmental outcome. *Eur J Pediatr* 167:1141–1147.
44. Skuse DH (2007) Rethinking the nature of genetic vulnerability to autistic spectrum disorders. *Trends Genet* 23:387–395.
45. Buja A, Krieger AM, George EI (2016) A visualization tool for mining large correlation tables: The association navigator. *Handbook of Big Data*, Chapman & Hall/CRC Handbooks of Modern Statistical Methods, eds Bühlmann P, Drineas P, Kane M, Laan Mvd (CRC Press, Boca Raton, FL), pp 73–102.