

A Third Linear Association Between Olduvai (DUF1220) Copy Number and Severity of the Classic Symptoms of Inherited Autism

Jonathan M. Davis, Ph.D., Ilea Heft, Ph.D., Stephen W. Scherer, Ph.D., James M. Sikela, Ph.D.

Objective: The authors previously reported that the copy number of sequences encoding an Olduvai protein domain subtype (CON1) shows a linear association with the severity of social deficits and communication impairment in individuals with autism. In this study, using an improved measurement method, the authors replicated this association in an independent population.

Method: The authors obtained whole genome sequence (WGS) data and phenotype data on 215 individuals from the Autism Speaks MSSNG project. They derived copy number from WGS data using a modified sequence read-depth technique. A linear mixed-effects model was used to test the association between Olduvai CON1 copy number and symptom severity as measured by the Autism Diagnostic Interview–Revised. The authors then combined data from previous studies (N=524) for final analyses.

Results: A significant linear association was observed between CON1 copy number and social diagnostic score (SDS) ($\beta=0.24$) and communicative diagnostic score (CDS) ($\beta=0.23$). Using the combined data, the authors present strong significant associations of CON1 dosage with SDS ($\beta=0.18$) and CDS ($\beta=0.13$). The authors also implicate Olduvai subtypes found in two genes, *NBPF1* and *NBPF14* ($R^2=6.2\%$). Associations were preferentially found in multiplex versus simplex families.

Conclusions: The finding of a third dose-dependent association between Olduvai sequences and autism severity, preferentially in multiplex families, provides strong evidence that this highly duplicated and underexamined protein domain family plays an important role in inherited autism.

Am J Psychiatry 2019; 176:643–650; doi: 10.1176/appi.ajp.2018.18080993

Autism is a lifelong condition that is characterized by greatly impaired social and communicative abilities as well as increased repetitive behaviors and stereotyped interests. It is a common heritable condition that affects as many as 1 in 42 males and 1 in 189 females (1). Autism is an exceptionally heterogeneous behaviorally defined condition that may be parsed into severity-based subtypes (2) but also may be best described as a continuum of symptom severity that progresses from minimal to substantial impact (3). For example, as many as 31% of individuals with autism have an intellectual disability, but also as many as 44% of children with autism are identified as having average and above average IQ (1). Similarly, traditional autism-like characteristics such as impairment in social reciprocity tend to follow a continuous distribution among individuals with autism (4).

Autism can also be divided genetically into multiplex and simplex forms. The former requires that families have at least two affected siblings, and the latter requires only one affected child and one unaffected child. A growing body of data suggests that the two forms may represent distinct

conditions and involve different genetic etiologies (5, 6). Even though it has been suggested that simplex autism is more common (7), the largest studies of autism reported so far do not appear to corroborate this view and point to autism as primarily an inherited condition (8, 9).

Important genetic components of autism are strongly suggested by heritability estimates that range from 50% to 90% (8, 10). However, despite the high heritability, few genetic markers of substantial effects have been identified that can explain a large portion of the condition. Genome-wide association studies have identified numerous single-nucleotide polymorphisms (SNPs) of very small effect (11, 12), and whole genome-wide copy number variation studies have identified de novo regions of very large effect in simplex children, yet these events appear to be rare (13, 14).

Despite extensive efforts to identify the genetic factors underlying autism, a substantial portion of the expected genetic contribution to the disorder has not yet been identified. This missing heritability and the heterogeneity of the condition have led to hypotheses of polygenetic effects

See related feature: **CME course** (p. 671)

involving hundreds to thousands of small-effect loci and to speculation that autism could be caused by numerous different alleles acting in summation or in a complex and unidentified pattern (15). However, evidence suggests that this contribution may not be substantive (16). Furthermore, few genetic studies have accounted for the extensive heterogeneity of the condition, and most label all affected as the same in classic case-control designs.

It has been proposed that the limited success of efforts to identify the genetic factors behind autism may be because the critical genomic contributors are being missed by currently available genomic technologies (17). In this regard, highly duplicated, copy-number-polymorphic sequences are excellent candidates to fill such a role. Indeed, not only are such sequences missed by conventional genetic approaches, including exome sequencing, but also, because their copy number may vary continuously over a wide range, they have the potential to produce a broad spectrum of phenotypic variation.

One such candidate of high and variable copy number that has been absent from traditional genetic studies of autism is the Olduvai (formerly DUF1220) protein domain family (18, 19), primarily encoded by *NBPF* genes (20). Olduvai domains are approximately 65 amino acids in length and show the greatest human lineage-specific increase in copy number of any protein coding region in the genome (~300 copies in humans, with 165 being human specific) (21, 22). Olduvai copy number also shows a striking correlation with brain size and neuron number among anthropoid primates, and with brain size and cognitive aptitude in the human population (23–25). Interestingly, most human Olduvai copies map to 1q21, a genomic region in which duplications have been linked with autism (and macrocephaly) and reciprocal deletions to schizophrenia (and microcephaly) (26, 27).

Olduvai copies have been parsed into six primary evolutionary subtypes based on sequence similarity: conserved (CON1-3) and human lineage specific (HLS1-3) (21). The primary means of adding human-specific copies has been via amplification of the Olduvai triplet: a three-domain unit composed of an HLS1, HLS2, and HLS3 subtype (21). Olduvai copy number is also highly variable in the human population (ranging from 250 to 350 copies) but, because of the repetitive nature of the family, has not been directly studied in conventional genetic studies of autism. Hence, Olduvai sequences represent a rich source of unexamined functional allelic variation.

Using an earlier method to follow Olduvai copy number (droplet digital polymerase chain reaction, ddPCR), we previously demonstrated (28) that in multiplex families the copy number of Olduvai subtype CON1 showed a linear association with the severity of the primary symptoms of autism, a finding that was subsequently replicated in a second population (29). However, CON1 subtypes, which range from approximately 60 to 85 diploid copies in humans, are found at >20 dispersed loci on chromosome 1, and ddPCR only measures the total number of copies in the genome. As a result, these previous studies were unable to provide information on which copies may be critical to the associations with severity.

To address this limitation, we combined two recently developed resources: availability of whole genome sequences for large numbers of individuals with autism (30, 31), and a higher-resolution method for measuring Olduvai copy number (32). Using these tools, we report here a third linear association between CON1 dosage and the severity of the core symptoms of autism in an independent population. We also implicate gene-specific Olduvai subtypes as potential drivers of this association. Finally, we show that the Olduvai associations are preferentially found in multiplex as opposed to simplex families and thus may be important to inherited forms of autism.

METHODS

In this study, we first sought to replicate previous work in an independent population using different assay methods, and once replication was identified, data from all three studies were combined for an overall evaluation. Next, a latent genotype was constructed from specific *NBPF* genes and then compared with a latent phenotype to identify important *NBPF* genes.

Replication Population

Individuals with phenotype data from the Autism Genetic Resource Exchange (33) and whole genome sequence (WGS) data from the Autism Speaks MSSNG consortium (30, 31) were used for the replication analysis. Individuals met autism criteria through the Autism Diagnostic Interview–Revised and the Autism Diagnostic Observation Schedule. While a large number of samples have been collected in the Autism Speaks MSSNG project, we were able to assemble 215 multiplex samples that were not included in our previous investigations and that had appropriate phenotype measures for replication. Nonwhite race/ethnicity was included and identified by analyses of WGS data using a consensus of Google-developed machine learning algorithms (see <https://cloud.google.com/blog/products/gcp/genomic-ancestry-inference-with-deep-learning>) and the ADMIXTURE software tool on a subset of common SNPs (34). Given the small numbers of specific minority groups, nonwhite and white ethnicity were collapsed into two categories. Population stratification was addressed through the addition of this race/ethnicity variable in multivariate regression analyses, as discussed further below.

Copy Number Measurement

Copy number of Olduvai domains in the replication population was obtained through a WGS read-depth method (32). Briefly, the read depth of regions containing Olduvai domains was compared with the range of location of the regions based on hg38. The depth of reads compared with the breadth of the region of interest covered, corrected by GC content and copy number of ultraconserved sequences, yielded a total copy number measurement of candidate Olduvai sequences per individual. This technique was

validated by ddPCR and is explained further elsewhere (32). Details on all aspects of whole genome sequencing and mapping for all data are provided as part of the Autism Speaks MSSNG Project (31). To ensure that cross-platform variation did not influence results, data from only one sequencing platform (Illumina HiSeq X) were included. Briefly, sequences were generated using paired-end reads of 150-bp length on an Illumina HiSeq X platform, following Illumina's recommended protocol. Given high correlation among HLS1, HLS2, and HLS3 domains, we utilized HLS1 measurements as representative of HLS2 and HLS3. The original reports that identified severity associations relied on ddPCR-derived copy number that determined total genomic copy number, for example, for the CON1 subtype. Read-depth analysis of WGS data, on the other hand, allows specific subtype copy number to be identified within specific *NBPF* genes (32). In order to replicate previous work, we summed Olduvai CON1 copy number derived from WGS data across all *NBPF* genes and used it as the primary explanatory variable in multivariate regression.

Replication Analysis

Linear multivariate regression with a random intercept for sibling family was used to test the association between total Olduvai CON1 and autism symptom severity. The mixed model was fitted with reduced maximum likelihood, and tests of significance were two-tailed *t* tests. Severity was determined by the social diagnostic score (SDS) and the communicative diagnostic score (CDS) from the Autism Diagnostic Interview–Revised (28). Covariates included other Olduvai domain totals (CON2, CON3, and HLS1) as well as sex, age, Raven's Progressive Matrices IQ, and race/ethnicity. Reduced models were developed through a backward selection procedure, keeping only significant covariates ($p < 0.05$) and the suggestive interaction term sex by CON1 (SDS, $p = 0.10$; CDS, $p = 0.02$). The Akaike information criterion supported the use of a random effect with random intercept for sibling family. An interaction of CON1 by sex was explored, given the biological plausibility. Residual diagnostics did not indicate departures from linear model assumptions.

Combined Analyses

Data from the three projects were combined for final evaluation. Associations of CON1 and HLS1 copy number with symptoms were tested with multivariate linear regression, as previously described. Random effects, family ID, and study were explored but were not supported by the Akaike information criterion and therefore were not included. Backward selection with the appropriate previously mentioned covariates was utilized, and results from parsimonious models are presented. Interactions of copy number by sex and copy number by inheritance were examined, and interaction *p* values are presented where simplex effects were explored.

Latent Variable Analysis

We used partial least squares path modeling (PLS-PM) to explore contributions of specific *NBPF* genes to a latent genotype and social severity phenotype. Bootstrapped mean estimates and 95% confidence intervals were generated with 5,000 iterations to validate PLS-PM associations. Gene and phenotype loading *p* values using *z*-statistics derived from the bootstrapped error were generated as recommended in other work (35) and are presented to aid in the evaluation of multiple comparisons during the latent variable construction. A conservative critical alpha in this stage of analyses is a Bonferroni-corrected *p* value of 0.002 (0.05/30). This corresponded to 11 of the larger and variable *NBPF* genes (*NBPF1*, 4, 6, 8, 9, 10, 11, 12, 14, 19, and 20), by two clades (CON1 and HLS1) in addition to eight phenotype measures described below.

To avoid inflation of the association between latent variables, PLS-PM model reduction was based on high levels of multicollinearity among Olduvai copy number and 95% bootstrapped intervals of loadings that did not include 0. All analyses were conducted with R (<https://cran.r-project.org/>), the nlme package (36), and the plspm package (37).

Phenotype Information

This project used the Autism Diagnostic Interview–Revised to provide the SDS and the CDS to replicate previous findings. The diagnostic scores contain numerous subquestions, each organized into four subdomains that measure relevant social and communicative behaviors. These subdomains are then summed through an algorithm to create the diagnostic scores (the SDS and CDS) and give clinical guidance in diagnosis. The four subdomains of the SDS include SOC1, SOC2, SOC3, and SOC4, with questions probing items such as responding to other children, emotional sharing, seeking comfort, and social smiling. The four subdomains of the verbal CDS included COM1, COM2V, COM3V, and COM4 and explore language items such as stereotyped utterances, pronoun reversal, and social use of language.

RESULTS

Replication Population

A total of 215 multiplex individuals were available for the replication analyses. Table 1 provides a phenotypic description of the replication population, which was predominantly male (77%) and of Western European descent (77%). Sixty-eight sibling groups were used in these analyses, and group size varied from two affected siblings to four. The summed CON1 copy number ranged from 60.7 to 84.3 copies, with a mean of 70.7 copies, and followed a Gaussian distribution. Interestingly, CON1 copy number of *NBPF1* had substantial variation and ranged from 8.5 to 28.5 copies. HLS1 copy number of *NBPF14* was also variable and ranged from 22.9 to 46.2 copies (Figure 1).

In the replication population, incrementally increasing copy number of total CON1 was associated with incrementally increasing SDS in males ($\beta = 0.25$, 95% CI = 0.02,

TABLE 1. Characteristics of the replication population

Measure	Mean	SD
Age (years)	10.0	4.0
Raven's Progressive Matrices IQ ^a	107.0	16.4
Autism Diagnostic Interview–Revised		
Social diagnostic score	20.9	5.6
Communicative diagnostic score	17.1	4.0
Repetitive behaviors	6.6	2.5
Total CON1 copies	70.7	4.0
Total HLS1 copies	200.2	17.3
	N	%
Male	165	76.7
Predicted ethnicity		
African	3	1.4
Admixed American	6	2.8
East Asian	4	1.9
European	165	76.7
Other	37	17.2

^a Missing data for 22 subjects.

0.49; $p=0.044$) (interaction $p=0.10$) (Table 2). A similar association was detected in males, where incrementally increasing copies of total CON1 were associated with incrementally increasing severity of CDS ($\beta=0.23$, 95% CI=0.21–0.25; $p=0.008$) (interaction $p=0.02$) (Table 2). As presented in Table 2, these results are highly similar to those in the previous two reports (28, 29).

Combined Analyses

The combined population analysis included data for 524 individuals, with 64 simplex and the remainder multiplex. In multiplex individuals, for each additional copy of CON1, SDS severity increased 0.18 points (95% CI=0.08, 0.28; $p=0.0008$), and CDS severity increased 0.13 points (95% CI=0.05, 0.21; $p=0.0009$) (Table 2) (SDS overall F test=5.16, $df=4$, 517, $p=0.0004$; CDS overall F test=6.11, $df=2$, 515, $p=0.002$). Age was a significant predictor ($\beta=0.18$, 95% CI=0.06, 0.30; $p=0.002$) in the SDS analysis. Increasing copy number of HLS1 was protective in the model of CDS ($\beta=-0.02$, 95% CI=-0.04, 0.0; $p=0.046$). The importance of multiplex over simplex was reaffirmed with SDS and CDS, where no associations were identified in simplex (SDS $\beta=-0.05$, 95% CI=-0.33, 0.21; $p=0.72$, interaction $p=0.11$; CDS $\beta=-0.02$, 95% CI=-0.23, 0.19; $p=0.87$, interaction $p=0.16$).

Latent Genotype and Phenotype Analyses

PLS-PM analysis identified an important latent genotype and latent severity phenotype that were associated with each other ($\beta=0.24$, 95% CI=0.14, 0.34; $p=0.002$), where the variation of the genetic latent variable explained a bootstrapped R^2 of 6.2% (95% CI=0.02, 0.12) of the variation of the latent social severity phenotype (Figure 2). The communicative score domains and social domains had substantial positive loadings on the social latent variable, where SOC1, SOC2,

FIGURE 1. Olduvai copy number distribution in *NBPF1* and *NBPF14*

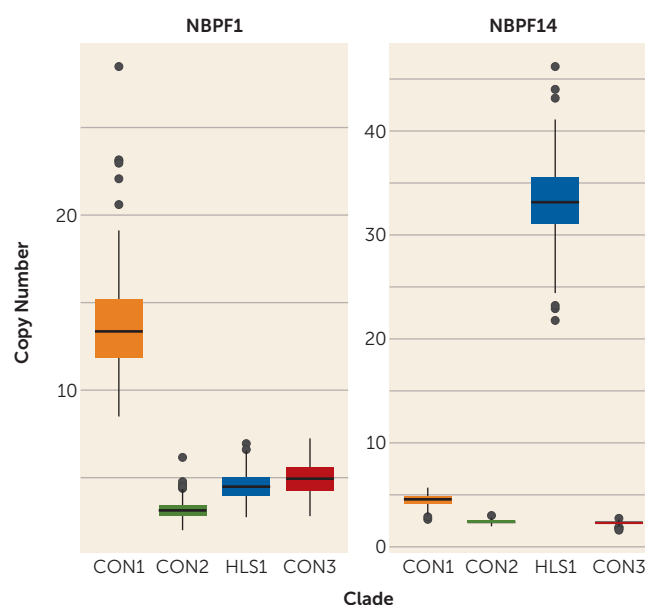


TABLE 2. Association Between CON1 copy number and symptoms in multiplex individuals as measured by the social and communicative scores on the Autism Diagnostic Interview–Revised

Sample and Scores	β	SE	p
Second replication (this report)			
Social diagnostic score	0.25	0.12	0.044
Communicative diagnostic score	0.23	0.08	0.008
First replication (2015, reference 29)			
Social diagnostic score	0.24	0.11	0.036
Communicative diagnostic score	0.16	0.09	0.072
Original study (2014, reference 28)			
Social diagnostic score	0.25	0.11	0.021
Communicative diagnostic score	0.18	0.08	0.047
Combined population			
Social diagnostic score	0.18	0.05	0.0008
Communicative diagnostic score	0.13	0.04	0.0009
Combined population, simplex strata (N=64)			
Social diagnostic score	-0.05	0.15	0.72
Communicative diagnostic score	-0.01	0.10	0.87

SOC3, SOC4, COM2V, and COM4 were important (all p values <0.0001) (Table 3). *NBPF1* CON1 (positive loading=0.75, 95% CI=0.19, 1.00; $p=0.0004$) and *NBPF14* HLS1 (positive loading=0.72, 95% CI=0.17, 1.00; $p=0.0007$) were important contributors to the latent genotype (Table 3).

DISCUSSION

Third Linear Association Between Olduvai Copy Number and Autism Severity

Using an independent population and a higher-resolution measurement method, we presented a third linear association between increasing Olduvai CON1 copy number and

worsening of autism social and communicative symptoms. The patterns of associations we found here are highly similar to those previously reported (28, 29), with only minor differences: the associations here are found predominantly in males and, compared with previous studies, show a larger effect size in communicative score. However, there are several reasons why such study-to-study variation can be expected: 1) autism is a complex condition with a wide range of phenotypic characteristics; 2) assessments of phenotypes can be expected to be imprecise, as they are primarily based only on caregiver reports (using algorithm scores from the Autism Diagnostic Interview–Revised) rather than physiological measurements; and 3) highly accurate copy number measurement of highly duplicated sequences, such as those encoding Olduvai domains, is difficult and copy number determinations are sometimes inexact.

We also showed that when the findings of this study were combined with those from our two previous studies, the linear association between CON1 and autism severity was unusually strong ($N=524$; $\beta=0.18$; $SE=0.05$; $p=0.0008$ for the SDS; and $\beta=0.13$; $SE=0.04$; $p=0.0009$ for the CDS). Given the substantial challenges related to accurate autism phenotyping and Olduvai genotyping, we believe it is remarkable that these studies have now identified similar associations in three independent groups. It is also noteworthy that, unlike in other recent genetic studies of autism, the associations reported here have been detectable, and replicable, using only modest sample sizes. Taken together, such findings strongly support a significant role for Olduvai CON1 in the severity of social and communication impairment in autism.

Associations Found Preferentially With Multiplex and Not Simplex Families

It is noteworthy that the significant linear associations with Olduvai copy number we report have appeared predominantly with multiplex as opposed to simplex families. This trend fits with other accumulating evidence that points to the possibility that multiplex and simplex autism may be genetically and mechanistically distinct disorders (5). First, multiplex autism, which requires that at least two children in the family have autism, is thought to better represent the inherited form of autism. In contrast, simplex autism is often due to de novo events (31, 38), but not exclusively so (39). Second, the phenotypes of multiplex and simplex autism are typically distinct, with simplex tending to be more severe

FIGURE 2. Olduvai latent genotype and latent social phenotype path model with loadings

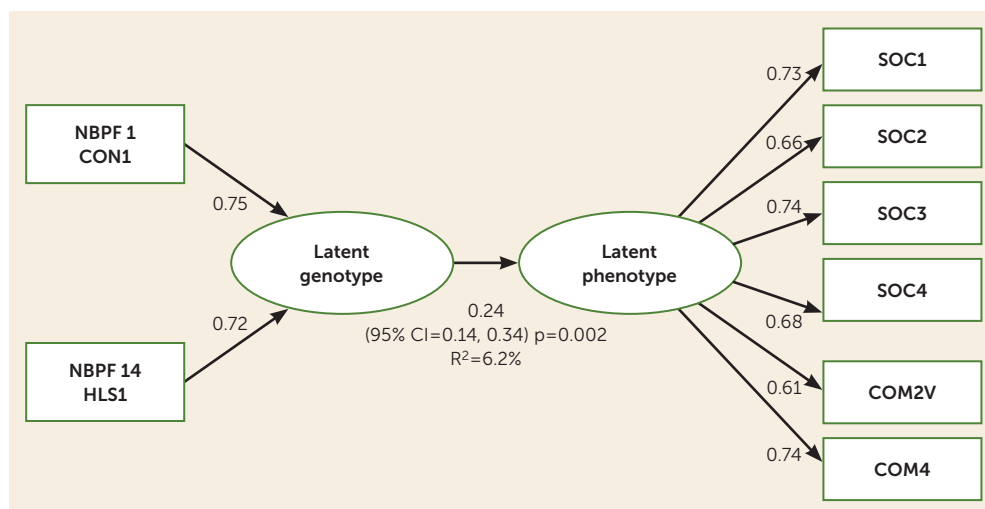


TABLE 3. Bootstrapped mean loadings of important contributors to the latent genotype and phenotype^a

Loadings	Mean Bootstrapped Loading	Bootstrapped 95% CI	p
Latent genotype			
NBPF1 CON1	0.75	0.19, 1.00	0.0004
NBPF14 HLS1	0.72	0.17, 1.00	0.0007
Latent phenotype			
SOC1	0.73	0.50, 0.84	<0.0001
SOC2	0.66	0.44, 0.79	<0.0001
SOC3	0.74	0.53, 0.84	<0.0001
SOC4	0.68	0.42, 0.82	<0.0001
COMV2	0.61	0.40, 0.80	<0.0001
COM4	0.74	0.60, 0.85	<0.0001

^a The p values are derived from the bootstrapped confidence intervals.

than multiplex phenotypes (6, 40). Third, recent large-scale genome sequencing studies of simplex autism have not found genes that are involved in the typical core phenotypes of autism, such as social deficits and communication impairment. Rather, such simplex studies have preferentially found genes important to motor functions (41). In contrast, our studies, now confirmed in three independent populations, suggest an opposite trend: we preferentially found associations involving the core autism phenotypes (social and communication impairment) in multiplex and not simplex families. In this regard, it is worth pointing out that early gene finding studies of inherited autism met with minimal success, a fact that has been mentioned as a reason for initiating genomic studies of autism using simplex families (7). The findings we report here raise the possibility that the previous lack of success in identifying genetic contributors behind inherited autism may be because key sequences that contribute to multiplex autism involve difficult-to-measure, highly duplicated, dynamic, and copy number polymorphic sequences, such as those encoding the Olduvai family, and, as

a result, have never been directly measured in other studies of autism (17).

Candidate Olduvai Subtypes That May Drive the Associations With Autism Severity

In our previous studies, we reported linear associations between autism and the copy number of the CON1 subtype of Olduvai. However, the measurement methods we used could not determine which of the many copies of CON1 (approximately 60–85 diploid copies) were underlying the association. To address this limitation, we employed a high-resolution-sequence read-depth method that, when applied to WGS data sets, allowed many of the different *NBPF* genes and Olduvai subtypes to be discriminated from one another.

To identify specific *NBPF* genes and Olduvai subtypes that are important to a social severity phenotype, we examined our WGS read-depth data utilizing a latent variable analytic method, partial least squares path modeling (PLS-PM). Numerous two-by-two comparisons are commonly conducted in genetic association studies (or potentially in this instance, numerous one-continuous-gene-by-one-continuous-phenotype comparisons are possible). These comparisons assume independence, and they assume, moreover, that the genotype has a unique effect on a specific measured phenotype. However, in autism, many characteristics are highly correlated, making traditional two-by-two statistical approaches inefficient and redundant. At the same time, simply summing characteristics or clinically grouping other combinations of symptoms will reduce the number of comparisons but may not best describe the phenotype as it relates to an important genotype. PLS-PM addresses these challenges through the construction of the latent genotype and phenotype variables and then makes one statistical test between the two. In this case, of eight social metrics, two communication subscores were excluded and, of 22 *NBPF* genes and subtypes, two were retained—*NBPF14* HLS1 and *NBPF1* CON1. In this study, the latent genetic variable explained 6.2% of the latent social phenotype, which, in the context of genetic association studies of complex heterogeneous conditions, is considerable. This unusually large effect size speaks to both the importance of Olduvai in autism and the importance of refined phenotypes in complex conditions.

Potential Importance of *NBPF14* HLS1 and *NBPF1* CON1

***NBPF14* HLS1.** The *NBPF14* gene is located in the 1q21 region of chromosome 1 and is part of a genomic sequence that has been repeatedly associated with autism (26, 27). *NBPF14* is one of the four human *NBPF* genes that together encode the great majority of human-specific Olduvai copies. These additional human-specific copies have been primarily generated by tandem intragenic expansion of the Olduvai triplet composed of an HLS1, HLS2, and HLS3 subtype (21, 22). These subtypes show high sequence similarity, and because additions/deletions of entire triplets is the predominant form of copy number variation in the expanded *NBPF* genes, the HLS1 subtype of *NBPF14* can be viewed as a proxy for a full

Olduvai triplet. These three-domain blocks are also highly copy number polymorphic in the human population, among both typically developing individuals and individuals with autism (28, 32). Thus, their high level of variation in humans, the inability of conventional genomic methods to directly examine these coding sequences (17, 42), the multiple studies implicating this genomic region in autism (43, 44), and our studies linking Olduvai copy number variation to the severity of the core symptoms of multiplex autism (28, 29) make sequences encoding *NBPF14*-related Olduvai triplets prime candidates for involvement in inherited forms of autism.

***NBPF1* CON1.** The copy number of the *NBPF1* gene among humans appears to be highly variable, and the gene is thought to encode several CON1 domains (21, 32). These characteristics fit with the idea that variation of CON1 in *NBPF1* has not been monitored in most genomic studies of autism, and that *NBPF1*-related sequences (e.g., *NBPF1L*) may not be correctly mapped in the most recent human genome assembly (hg38). Given these observations, the possibility that *NBPF1* CON1 may be involved in influencing autism severity should not be considered surprising.

Genomic Trade-Offs and Links With Neurogenesis

The Olduvai protein domain family has been linked with both evolutionary benefit (copy increase is associated with brain expansion and cognitive aptitude [23–25]) and disease (increases have been linked to autism and macrocephaly and decreases to schizophrenia and microcephaly [23, 45]). Such duality of effect has been incorporated into a cognitive genomic trade-off model that proposes that 1) Olduvai's effect on brain size and neuron number may be dosage dependent, that is, more copies produce more neurons (45), and 2) variation in Olduvai copy number can occur in different ways, and it is which, where, how, and when copies change that determines whether the consequences will be beneficial or harmful (17).

Evidence suggests that the Olduvai trade-off may be related to variation in neurogenesis. For example, Olduvai has been shown to promote neural stem cell proliferation and exhibits a dosage-dependent association with gray matter volume and IQ in healthy individuals, both of which are related to neuron number (23–25). Olduvai copy number also shows a strong linear association with primate neuron number and brain size (but not body size) (23, 24). Finally, human-specific Olduvai sequences have recently been linked with Notch-related pathways known to promote neurogenesis (unpublished data).

As previously suggested (45), the involvement of Olduvai in influencing neuron number, possibly via centrosome-mediated effects, may also be relevant to its role in autism and schizophrenia. It is well established that how and when neurons are produced during brain development involve processes that, if altered, can produce atypical neuronal density and growth. In this regard, excesses in neuron number along with accelerated brain growth have been

associated with autism (46). In contrast, neuronal deficits may be important in schizophrenia, where smaller gray matter regions (47) and reduced Olduvai copy number (48) have been reported. In short, as Olduvai increases, autism severity and brain size increase, whereas the opposite trend is true for schizophrenia. These observations add further support to the proposal that autism and schizophrenia may be opposites of a cognitive disease continuum (43) involving opposites in Olduvai dosage (17). Such findings raise the testable hypothesis that the associations of Olduvai dosage with the severity of autism and schizophrenia are due to Olduvai-driven abnormalities in neurogenesis.

Finally, the data reported here, linking Olduvai sequences with autism severity, using a different means of assessing copy number and a different population, provide additional support for such a trade-off model. Because of technological limitations involving measurement of highly duplicated sequences, a more detailed description of the nature of Olduvai variation in humans, in either disease or nondisease populations, has not yet been reported. For similar reasons, it has been difficult to examine the potential involvement of Olduvai variation in autism risk. Given that much of the genetic contributors to inherited autism remain missing (17), and given the possibility that multiplex autism may be due to a single dominant transmitted trait (16), it is intriguing that variation in the ~300-member human Olduvai family has yet to be fully examined in inherited autism. Continued improvements in genome sequencing technology that will allow such sequences to be studied with greater precision should permit these questions to be tested more rigorously in the future.

AUTHOR AND ARTICLE INFORMATION

The Department of Biochemistry and Molecular Genetics, Human Medical Genetics and Genomics Program and Neuroscience Program, University of Colorado School of Medicine, Aurora (Davis, Heft, Sikela); the McLaughlin Centre and the Department of Molecular Genetics, University of Toronto, and the Centre for Applied Genomics and Program in Genetics and Genome Biology, Hospital for Sick Children, Toronto (Scherer).

Send correspondence to Dr. Sikela (james.sikela@ucdenver.edu).

Supported by NIH grant R01 MH108684 (to Dr. Sikela) and Pilot Award 309230 from the Simons Foundation Autism Research Initiative (to Dr. Sikela).

The authors gratefully acknowledge the resources provided by the Autism Speaks MSSNG project and the Autism Genetic Resource Exchange Consortium as well as the participating families.

Dr. Sikela is an inventor on U.S. patent no. 9,988,681, related to autism. Dr. Scherer has served on scientific advisory boards or committees for Population Bio, Inc., and Deep Genomics. The other authors report no financial relationships with commercial interests.

Received August 28, 2018; revision received November 9, 2018; accepted December 10, 2018; published online Feb. 15, 2019.

REFERENCES

1. Baio J, Wiggins L, Christensen DL, et al: Prevalence of autism spectrum disorder among children aged 8 years: Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2014. *MMWR Surveill Summ* 2018; 67:1–23
2. Grzadzinski R, Huerta M, Lord C: DSM-5 and autism spectrum disorders (ASDs): an opportunity for identifying ASD subtypes. *Mol Autism* 2013; 4:12
3. Spiker D, Lotspeich LJ, Dimiceli S, et al: Behavioral phenotypic variation in autism multiplex families: evidence for a continuous severity gradient. *Am J Med Genet* 2002; 114:129–136
4. Constantino JN, Todd RD: Autistic traits in the general population: a twin study. *Arch Gen Psychiatry* 2003; 60:524–530
5. Zhao X, Leotta A, Kustanovich V, et al: A unified genetic theory for sporadic and inherited autism. *Proc Natl Acad Sci USA* 2007; 104:12831–12836
6. Virkud YV, Todd RD, Abbacchi AM, et al: Familial aggregation of quantitative autistic traits in multiplex versus simplex autism. *Am J Med Genet B Neuropsychiatr Genet* 2009; 150B:328–334
7. Autism risk genes: success with the simplex approach. New York, Simons Foundation, July 27, 2015. <https://www.simonsfoundation.org/2015/07/27/autism-risk-genes-success-with-the-simplex-approach/>
8. Sandin S, Lichtenstein P, Kuja-Halkola R, et al: The familial risk of autism. *JAMA* 2014; 311:1770–1777
9. Sandin S, Lichtenstein P, Kuja-Halkola R, et al: The heritability of autism spectrum disorder. *JAMA* 2017; 318:1182–1184
10. Tick B, Bolton P, Happé F, et al: Heritability of autism spectrum disorders: a meta-analysis of twin studies. *J Child Psychol Psychiatry* 2016; 57:585–595
11. Gaugler T, Klei L, Sanders SJ, et al: Most genetic risk for autism resides with common variation. *Nat Genet* 2014; 46:881–885
12. Niemi MEK, Martin HC, Rice DL, et al: Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. *Nature* 2018; 562:268–271
13. Sebat J, Lakshmi B, Malhotra D, et al: Strong association of de novo copy number mutations with autism. *Science* 2007; 316:445–449
14. Levy D, Ronemus M, Yamrom B, et al: Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron* 2011; 70:886–897
15. Klei L, Sanders SJ, Murtha MT, et al: Common genetic variants, acting additively, are a major source of risk for autism. *Mol Autism* 2012; 3:9
16. Ronemus M, Iossifov I, Levy D, et al: The role of de novo mutations in the genetics of autism spectrum disorders. *Nat Rev Genet* 2014; 15:133–141
17. Sikela JM, Searles Quick VB: Genomic trade-offs: are autism and schizophrenia the steep price of the human brain? *Hum Genet* 2018; 137:1–13
18. Popesco MC, Maclaren EJ, Hopkins J et al: Human lineage-specific amplification, selection, and neuronal expression of DUF1220 domains. *Science* 2006; 313:1304–1307 (supporting online material: www.sciencemag.org/cgi/content/full/313/5791/1304/DC1)
19. Sikela JM, van Roy F: Changing the name of the NBPF/DUF1220 domain to the Olduvai domain. *F1000 Res* 2017; 6:2185
20. Vandepoele K, Van Roy N, Staes K, et al: A novel gene family NBPF: intricate structure generated by gene duplications during primate evolution. *Mol Biol Evol* 2005; 22:2265–2274
21. O'Bleness MS, Dickens CM, Dumas LJ, et al: Evolutionary history and genome organization of DUF1220 protein domains. *G3 (Bethesda)* 2012; 2:977–986
22. O'Bleness M, Searles VB, Dickens CM, et al: Finished sequence and assembly of the DUF1220-rich 1q21 region using a haploid human genome. *BMC Genomics* 2014; 15:387
23. Dumas LJ, O'Bleness MS, Davis JM, et al: DUF1220-domain copy number implicated in human brain-size pathology and evolution. *Am J Hum Genet* 2012; 91:444–454
24. Keeney JG, Davis JM, Siegenthaler J, et al: DUF1220 protein domains drive proliferation in human neural stem cells and are associated with increased cortical volume in anthropoid primates. *Brain Struct Funct* 2015; 220:3053–3060

25. Davis JM, Searles VB, Anderson N, et al: DUF1220 copy number is linearly associated with increased cognitive function as measured by total IQ and mathematical aptitude scores. *Hum Genet* 2015; 134: 67–75
26. Brunetti-Pierri N, Berg JS, Scaglia F, et al: Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat Genet* 2008; 40:1466–1471
27. Mefford H, Sharp A, Baker C, et al: Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med* 2008; 359:1685–1699
28. Davis JM, Searles VB, Anderson N, et al: DUF1220 dosage is linearly associated with increasing severity of the three primary symptoms of autism. *PLoS Genet* 2014; 10:e1004241
29. Davis JM, Searles Quick VB, Sikela JM: Replicated linear association between DUF1220 copy number and severity of social impairment in autism. *Hum Genet* 2015; 134:569–575
30. C Yuen RK, Merico D, Bookman M, et al: Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat Neurosci* 2017; 20:602–611
31. Yuen RKC, Thiruvahindrapuram B, Merico D, et al: Whole-genome sequencing of quartet families with autism spectrum disorder. *Nat Med* 2015; 21:185–191
32. Astling DP, Heft IE, Jones KL, et al: High resolution measurement of DUF1220 domain copy number from whole genome sequence data. *BMC Genomics* 2017; 18:614–629
33. Geschwind DH, Sowiński J, Lord C, et al: The autism genetic resource exchange: a resource for the study of autism and related neuropsychiatric conditions. *Am J Hum Genet* 2001; 69:463–466
34. Alexander DH, Novembre J, Lange K: Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 2009; 19: 1655–1664
35. Altman DG, Bland JM: How to obtain the p value from a confidence interval. *BMJ* 2011; 343:d2304
36. Pinheiro J, Bates D, DebRoy S, et al: nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-137, 2018. <https://cran.r-project.org/package=nlme>
37. Sanchez G, Trinchera L, Russolillo G: plsrm: Tools for Partial Least Squares Path Modeling (PLS-PM). R package version 0.4, 2017. <https://github.com/gastonstat/plsrm>
38. Iossifov I, O’Roak BJ, Sanders SJ, et al: The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 2014; 515: 216–221
39. Krumm N, Turner TN, Baker C, et al: Excess of rare, inherited truncating mutations in autism. *Nat Genet* 2015; 47:582–588
40. Constantino JN, Zhang Y, Frazier T, et al: Sibling recurrence and the genetic epidemiology of autism. *Am J Psychiatry* 2010; 167: 1349–1356
41. Buja A, Volfovsky N, Krieger AM, et al: Damaging de novo mutations diminish motor skills in children on the autism spectrum. *Proc Natl Acad Sci USA* 2018; 115:E1859–E1866
42. Brahmachary M, Guilmatre A, Quilez J, et al: Digital genotyping of macrosatellites and multicopy genes reveals novel biological functions associated with copy number variation of large tandem repeats. *PLoS Genet* 2014; 10:e1004418
43. Crespi BJ, Crofts HJ: Association testing of copy number variants in schizophrenia and autism spectrum disorders. *J Neurodev Disord* 2012; 4:15
44. Girirajan S, Dennis MY, Baker C, et al: Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. *Am J Hum Genet* 2013; 92:221–237
45. Dumas L, Sikela JM: DUF1220 domains, cognitive disease, and human brain evolution. *Cold Spring Harb Symp Quant Biol* 2009; 74: 375–382
46. Courchesne E, Mouton PR, Calhoun ME, et al: Neuron number and size in prefrontal cortex of children with autism. *JAMA* 2011; 306: 2001–2010
47. Haijma SV, Van Haren N, Cahn W, et al: Brain volumes in schizophrenia: a meta-analysis in over 18 000 subjects. *Schizophr Bull* 2013; 39:1129–1138
48. Searles Quick VB, Davis JM, Olincy A, et al: DUF1220 copy number is associated with schizophrenia risk and severity: implications for understanding autism and schizophrenia as related diseases. *Transl Psychiatry* 2015; 5:e697