Neuron

Phenotype and genetic analysis of data collected within the first year of NeuroDev

Highlights

- First trio-based study of neurodevelopmental conditions in Kenya and South Africa
- Pathogenic and likely pathogenic variants were identified in 22 of 99 trios
- Families were linguistically and ancestrally diverse
- Lessons learned from data collection support future NDD studies in Africa

Authors

Patricia Kipkemoi, Heesu Ally Kim, Bjorn Christ, ..., Amina Abubakar, Kirsten A. Donald, Elise B. Robinson

Correspondence

amina.abubakar@aku.edu (A.A.), kirsty.donald@uct.ac.za (K.A.D.), erob@broadinstitute.org (E.B.R.)

In brief

Kipkemoi et al. share the genetic and phenotypic results from the first year of NeuroDev, a study of neurodevelopmental conditions in Kenya and South Africa. They collected data from 600 participants, including 200 children with neurodevelopmental conditions. Analysis of genetic data from 99 parent-child trios found causal or likely causal variants.

Kipkemoi et al., 2023, Neuron *111*, 2800–2810
September 20, 2023
https://doi.org/10.1016/j.neuron.2023.06.010

CellPress

CellPress
OPEN ACCESS

Neuron

Report

Phenotype and genetic analysis of data collected within the first year of NeuroDev

Patricia Kipkemoi,[1,2] Heesu Ally Kim,[3] Bjorn Christ,[4] Emily O'Heir,[3] Jake Allen,[3] Christina Austin-Tse,[3,5] Samantha Baxter,[3] Harrison Brand,[3,5,6] Sam Bryant,[3] Nick Buser,[3] Victoria de Menil,[3,7] Emma Eastman,[4] Serini Murugasen,[4] Alice Galvin,[3] Martha Kombe,[1] Alfred Ngombo,[1] Beatrice Mkubwa,[1] Paul Mwangi,[1] Collins Kipkoech,[1] Alysia Lovgren,[3] Daniel G. MacArthur,[3] Brigitte Melly,[4] Katini Mwangasha,[1] Alicia Martin,[8] Lethukuthula L. Nkambule,[8] Alba Sanchis-Juan,[3,5,6] Moriel Singer-Berk,[3] Michael E. Talkowski,[3,5,6] Grace VanNoy,[3] Celia van der Merwe,[3] The NeuroDev Project, Charles Newton,[1,9,10] Anne O'Donnell-Luria,[3,5,11] Amina Abubakar,[1,9,10,*] Kirsten A. Donald,[4,12,*] and Elise B. Robinson[3,5,13,*]

[1]Neuroscience Unit, KEMRI-Wellcome Trust, Center for Geographic Medicine Research Coast, Kilifi, Kenya
[2]Complex Trait Genetics Department, Center for Neurogenomics and Cognitive Research, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
[3]The Broad Institute of MIT and Harvard, Cambridge, MA, USA
[4]Department of Paediatrics and Child Health, Red Cross War Memorial Children's Hospital and University of Cape Town, 4th Floor ICH Building, Rondebosch, South Africa
[5]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA
[6]Department of Neurology, Harvard Medical School, Boston, MA, USA
[7]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[8]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA
[9]Department of Psychiatry, University of Oxford, London, UK
[10]Institute of Human Development, Aga Khan University, Nairobi, Kenya
[11]Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA, USA
[12]Neuroscience Institute, University of Cape Town, Groote Schuur Hospital, Observatory, Cape Town, South Africa
[13]Lead contact
*Correspondence: amina.abubakar@aku.edu (A.A.), kirsty.donald@uct.ac.za (K.A.D.), erob@broadinstitute.org (E.B.R.)
https://doi.org/10.1016/j.neuron.2023.06.010

SUMMARY

Genetic association studies have made significant contributions to our understanding of the etiology of neurodevelopmental disorders (NDDs). However, these studies rarely focused on the African continent. The NeuroDev Project aims to address this diversity gap through detailed phenotypic and genetic characterization of children with NDDs from Kenya and South Africa. We present results from NeuroDev's first year of data collection, including phenotype data from 206 cases and clinical genetic analyses of 99 parent-child trios. Most cases met criteria for global developmental delay/intellectual disability (GDD/ID, 80.3%). Approximately half of the children with GDD/ID also met criteria for autism. Analysis of exome-sequencing data identified a pathogenic or likely pathogenic variant in 13 (17%) of the 75 cases from South Africa and 9 (38%) of the 24 cases from Kenya. Data from the trio pilot are publicly available, and the NeuroDev Project will continue to develop resources for the global genetics community.

INTRODUCTION

The Genetic Characterization of Neurodevelopmental Disorders project (NeuroDev) is a study of neurodevelopmental disorders (NDDs) that will collect and analyze extensive genetic and phenotypic data from over 5,000 people, including ~3,600 children and their parents, in Kenya and South Africa over the next several years.[1] Based on existing recruitment patterns, most of the 2,000 case individuals enrolled in the study are projected to meet criteria for global developmental delay/intellectual disability (GDD/ID) or autism. All data (e.g., phenotypes, genotype array, and exome-sequencing data) and materials (e.g., blood DNA; cryopreserved lymphocytes or CPLs) generated by NeuroDev will be publicly available through approved National Institute of Mental Health repositories. Through the data collection activity, which includes African ancestry populations largely absent from genetic reference panels (e.g., gnomAD), and through the public release of deeply characterized case and control data, NeuroDev aims to support diversity in biomedical research.

This paper presents the NeuroDev Pilot, which includes data from the project's first collection year. During the first year, we collected data from more than 200 cases and 600 total

participants. We describe phenotype data from all cases collected in the pilot period, along with genetic analysis of 99 exome-sequenced trios. Supplementing these early genetic and phenotypic findings, we present learning points of the first year of data collection and modifications made to the NeuroDev protocol initially presented in de Menil et al.[1] We hope that these reflections on process will help others in the design and execution of similar projects because more work on NDDs in Africa is both needed and underway. At only 99 trios, the NeuroDev trio pilot data are now the largest African NDD collection for which genetic and phenotypic data are publicly available to the research community. The trio data presented here can be accessed through National Human Genome Research Institute (NHGRI) Analysis Visualization and Informatics Lab-space (ANVIL) controlled-access data repository (https://anvilproject.org/data/).

RESULTS

Data collection

From August 2018 to July 2019, we enrolled 219 cases, 195 case mothers, 115 case fathers, and 92 unrelated child controls (Figure 1A). There were 106 total parent-case trios, and an additional 113 cases had only one participating parent. In Kilifi County, Kenya, case families were recruited from specialized neurology and occupational therapy clinics, special needs schools, and a database of previous studies. The case children enrolled in the study had tentative NDD diagnoses obtained from checklists or previous neuropsychological assessments. These diagnoses were confirmed by a study clinician prior to enrollment. In Cape Town, South Africa, participants with a clinical NDD diagnosis were recruited from the developmental clinic at Red Cross Children's Hospital and Tygerberg Hospital. Cases included in the study had a diagnosis of any NDD (excluding primary motor disorders, e.g., cerebral palsy) and were aged between 2 and 18 years (Figure 1B). We present in this report phenotype analyses of all 219 cases, along with genetic data analyses of the first 100 trios, 99 of which passed quality control measures for analysis.

At the end of the project's first year, the trio collection and overall case collection rate in NeuroDev were aligned with our 4-year sample size targets, and phenotype battery completion was high. In the first year, all participants had data for the demographic, neuromedical assessment, and behavioral measures. The behavioral measures included the Social Communication Disorders Checklist (SCDC);[2] 3Di Brief[3] to measure autism characteristics; and Swanson, Nolan, and Pelham Rating Scale (SNAP-IV) to measure canonical attention deficit hyperactivity disorder (ADHD) symptomatology.[4] Item-level missingness was below 10% for all measures (see Tables S1 and S2).

The Raven's Progressive Matrices (RPMs), NeuroDev's nonverbal reasoning ability measure, was completed by 99% of participating parents. Inspired by the Simons Simplex Collection (SSC) cognitive testing approach,[5,6] all enrolled children aged 6 years and older were offered the opportunity to attempt the RPM[7] (standard RPM if aged 12 years or older; colored RPM if aged 6–11 years), which was adapted and validated for use in Kilifi (Figure 1C).

Many case children (49%) could not complete age-appropriate versions of the RPM due to the extent of their developmental delay or behavioral challenges. All case children under 6 years of age and those over 6 years who could not complete the RPM were offered the Molteno Adapted Scales of Development (Molteno). We experienced similar challenges in some cases with regard to the completion of the Molteno, although all participants have at least some data. This experience is common to the studies of ID and autism, particularly those that include children and severe phenotypes.[8,9] Missing values from the RPM are associated with case severity and will therefore be informative.

In the first year of data collection, trio family ascertainment was higher than anticipated. The goal in South Africa to recruit 100 trios over the full duration of the project was met in just the first year of the study. All pilot participants in South Africa consented to have genetic findings related to their child's NDD returned to the family. Similarly, we observed a high rate (98%) of consent for sharing cell lines in addition to DNA (Figure 1D). Both of these options were available only in South Africa.

We used the University of California, San Diego Brief Assessment of Capacity to Consent (UBACC) to both ensure and measure parents' understanding of the study prior to consent. As a screening tool, the UBACC is used to identify parents who may require a more comprehensive evaluation of decisional capacity and enhanced consent procedures.[10,11] Participants need to achieve a score of 14.5 out of 20 to meet the test requirement, with the tool readministered up to a maximum of three trials if necessary, each after additional explanatory efforts. In the first year of data collection, only two parents failed their UBACC administrations, and their families were not included in the study. One of the two parents was reported to have a documented ID. An overwhelming majority of the parents showed good understanding of the protocol following detailed explanation by study staff because only 3% of participants scored below 14.5 on their first trial.[12]

The pilot period was used to further review our data collection strategy and tools. In the STAR Methods section, we share detailed observations about the assessment tools as applied in our context and any adaptations to the tools made at either site. These adaptations were made in response to questions that were contextually inappropriate for caregivers due to cross-cultural differences or linguistic challenges. We also discuss modifications to the protocol, lessons learned in the implementation of the study, and recruitment strategies used in response to the variability in enrollment of participants. For example, in Kilifi, we adapted our recruitment strategy during inclement weather and in response to challenges recruiting fathers during the workweek. We hope that these details will benefit future research projects and that this suite of pilot results as a whole will encourage more large-scale NDD projects in Africa.

We made one significant addition to the phenotype battery, the Child Behavior Checklist (CBCL),[13] aiming to strengthen participant behavioral characterization. We initially employed the CBCL to explore the validity of the SNAP-IV in a subset of NeuroDev participants and found it to be of both research and clinical utility. The CBCL, now included in NeuroDev's core phenotype battery, characterizes a comprehensive assortment of problem behaviors separately in preschool (age 1.5–5 years)

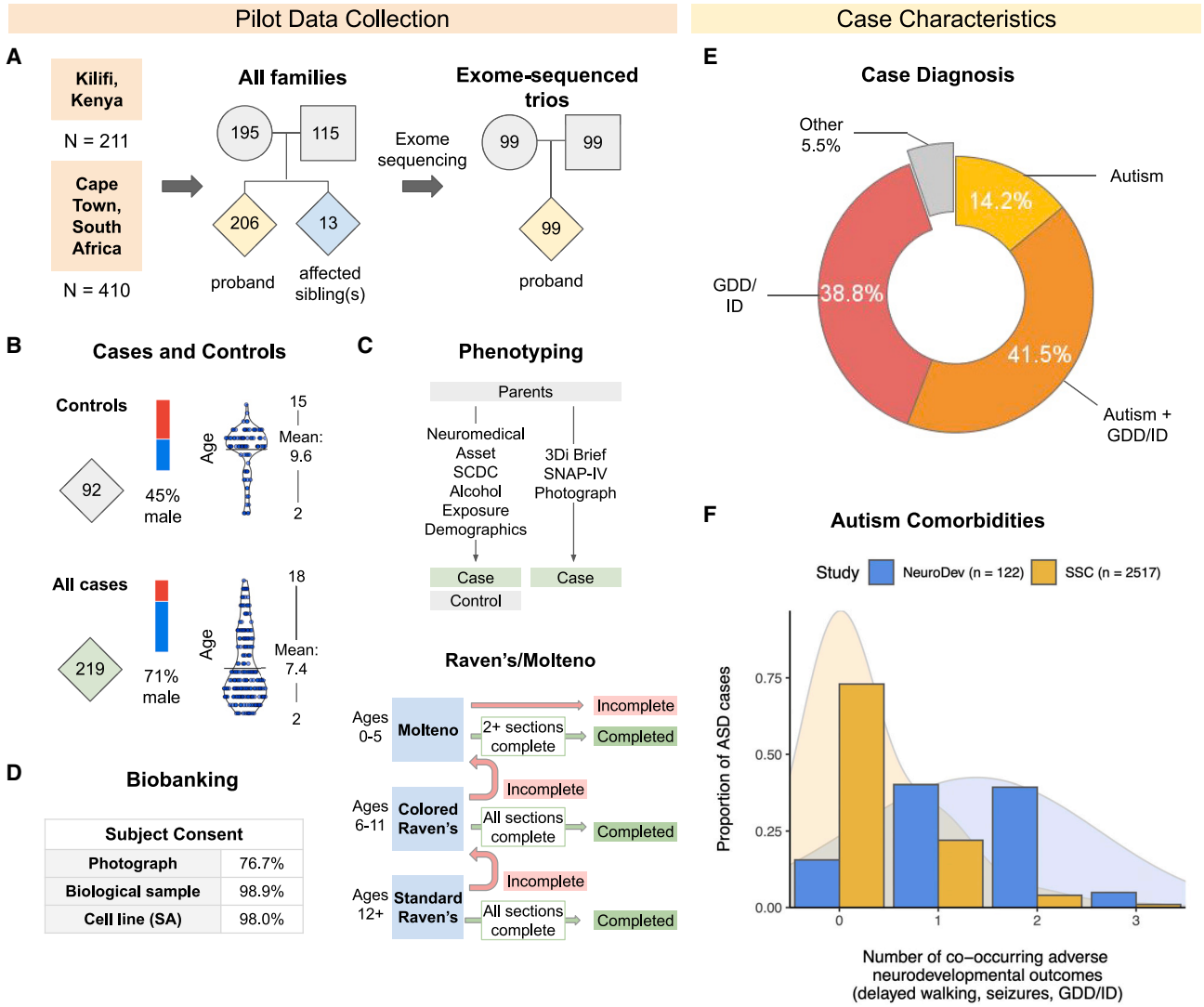

Figure 1. Overview of the NeuroDev trio pilot

(A) Data collection and exome-sequencing description.

(B) Distribution of sex, age, and consent subtypes across cases and controls.

(C) Assessments administered to cases and controls. The flowchart describes Raven's/Molteno administration processes for those who found completing either test challenging.

(D) Subject consent rates for biobanking. "SA," only available in South Africa.

(E) Diagnostic profiles of the 219 cases in the pilot phase. "Other" diagnoses include specific learning disabilities, communication disorders, and/or ADHD.

(F) Number of co-occurring adverse neurodevelopmental outcomes among the NeuroDev and SSC autism cases. Values of 0 (none present), 1, 2, or 3 (all present) indicate the total number of adverse outcomes, defined as delayed walking, seizures, and/or GDD/ID.

and school-age (6–18 years) children. In addition to the research opportunities afforded by the CBCL data, we observed clinical value in use of the CBCL to construct a behavioral profile for children who may need a referral for specialized intervention.

Phenotypic characteristics of cases ascertained in NeuroDev's first year

NeuroDev features an uncommonly detailed phenotype battery for a genetic study of its size. This section presents phenotypic findings from all 219 cases collected between August 2018 and July 2019. Of those cases, 156 were South African, and 63 were Kenyan, reflecting that the Kenyan data collection began halfway through the pilot's first year; 99 of these 216 cases were included among the exome-sequenced trios. Overall, the cases were 70.2% male ranging in age from 2–18 years. The majority (80.4%, n = 176) of cases met criteria for GDD/ID. Of those with GDD/ID, 91 cases (52% of all GDD/ID cases) also met criteria for autism. An additional 31 cases (14.2% of all cases) met criteria for autism without GDD/ID. A small number (5.5%) of cases did not meet criteria for either GDD/ID or autism, but

for other NDDs that are ascertained through NeuroDev: specific learning disabilities, communication disorders, and/or ADHD. The diagnostic composition of cases included in the 99 exome-sequenced trios is highly similar to that of the full set 219 cases (Figure 1E).

The cases were highly ancestrally and linguistically diverse, with more than 40 languages spoken in families and more than 24 ethnicities represented. In keeping with the approach initiated by the 1,000 Genomes study, we used language as an indicator of ethnic affiliation on top of self-reported ethnicity. Among cases from Kenya, the majority were from the Mijikenda ethnic group (88.7%), with many of the case families primarily speaking the Mijikenda languages (83.9%) or Kiswahili (14.5%). Within South Africa, most of the cases were of mixed ancestry (45.9%) or identified with multiple ancestries (16.9%), and many others identified as AmaXhosa (13.5%). Reflecting this, the main languages spoken by the case families in South Africa included English (74.3%) and isiXhosa (14.2%). The demographic questionnaire's specific questions on language and ethnicity were phrased as follows: "What is the primary language spoken in [participant's] home"? "What is [participant's] ethnicity or tribe"?

As anticipated, most children with autism in this initial group also met criteria for GDD/ID (74.6%). This is consistent with other descriptions of clinic-based cohorts of children with autism from resource-limited environments.[14–16] The scarcity of neuropsychiatric specialists and medical resources in the African region contributes to late diagnosis or a lack of access to services for children with milder autism symptoms and minimal cognitive impairment.[17–19]

We compared the prevalence of GDD/ID, delayed walking (after 18 months of age), and parent-reported seizures between child cases with autism in NeuroDev and those in the SSC, a large United States-based cohort (Figure 1F). Most NeuroDev autism cases presented with at least one of these co-occurring adverse neurological or neurodevelopmental outcomes: (1) seizures, (2) GDD/ID, and/or (3) walking later than 18 months (84.4%). The average number of co-occurring adverse outcomes in the NeuroDev study was 1.34 (n = 122), whereas this number was 0.33 for SSC (n = 2,517, $p = 2.2 \times 10^{-16}$). The difference in the number of co-occurring outcomes likely reflects the differences in research setting and case ascertainment, as described above. The rate of co-occurring neurodevelopmental conditions in individuals with autism is associated with average genetic architecture.[20,21] The case rate of *de novo* protein-truncating variants (PTVs) in constrained genes increases with the number of co-occurring adverse neurodevelopmental outcomes. Based on Weiner et al., we expected a ~50% increase in the observed rates of *de novo* PTVs in constrained genes in NeuroDev autism cases relative to SSC and similar cohorts.[21]

The initial 219 phenotyped cases experienced high levels of speech delay, with only 36% of cases meeting criteria for fluent speech. The high rates of delay were expected due to our ascertainment strategy, which included recruitment of cases from neurodevelopmental clinics and special needs schools. Those with a comorbid diagnosis of autism and ID/GDD had the lowest speech level, with 76% of families reporting the use of single words or less. Consistent with this level of developmental delay, many case children also experienced challenges completing the RPM. Among all cases, only 66% of all standard RPM testers (aged 12 or older) and 44% of all colored RPM testers (aged 6–11) were able to fully complete age-appropriate versions of the assessment (Tables S1and S5). Among standard RPM testers who could not complete the assessment, approximately half were able to complete the colored version of the RPM, and the rest completed the Molteno. Taken together, these findings suggest a high severity level of developmental delay in NeuroDev cases compared with other widely available NDD cohorts.

Genetic diversity

To examine genetic diversity, we compared the 99 genotyped trios with individuals from the combined Human Genome Diversity Panel (HGDP) and 1000 Genomes Project (TGP)[22] panel and the African Genome Variation Project (AGVP).[23] In the context of these reference panels, the South African and Kenyan NeuroDev cohorts were ancestrally dissimilar (Figure 2). The NeuroDev cohort from Cape Town, South Africa, maintained a high proportion of globally admixed individuals, consistent with the expectation from previous work,[24] and maintained a subset of populations lacking out-of-Africa admixture that cluster closely with continental African populations. The NeuroDev Pilot cohort from Kilifi, Kenya, was of almost exclusively African genetic ancestry with little evidence of global admixture (Figures 2A and 2B).

Next, we assessed continental African genetic ancestry of NeuroDev cohorts by intersecting the NeuroDev genotyping data with those of the populations described in the AGVP (Figures 2C and 2D).[23] The first principal component separates Ethiopian populations from other continental African populations due to back-to-Africa migration and admixture in Ethiopians.[25] The clustering of some NeuroDev South African samples with these individuals likely indicates out-of-Africa admixture. Those of solely African descent in South Africa and Kenya tend to cluster with geographically neighboring subcontinental populations: unadmixed NeuroDev South Africa samples cluster with the Zulu and Sotho populations of southern Africa, and NeuroDev Kenya samples cluster closely with eastern African populations, including the Baganda and Barundi. Given the change in languages spoken over generations (Figure 2E), we show that grandparents' ethnolinguistic groups are interrelated with genetic ancestry and that the maternal grandmother's spoken languages often align with the genetic ancestry of the proband (Figure 2F).

Genetic analyses of the trio pilot data

Of the 99 parent-child trios included in the trio sequencing analysis, 75 were from South Africa and 24 were from Kenya. As of the present analysis, 22 pathogenic or likely pathogenic variants have been identified in those families. We have also identified 7 suspicious variants of uncertain significance (VUSs) in emerging disease genes with supportive case data identified through submission to the MatchMaker Exchange (MME). A detailed description of the sequencing and data analysis approach can be found in the STAR Methods. In brief, exome sequencing was performed on each of the trios, and the data were uploaded to the *seqr* platform for analysis.[26]

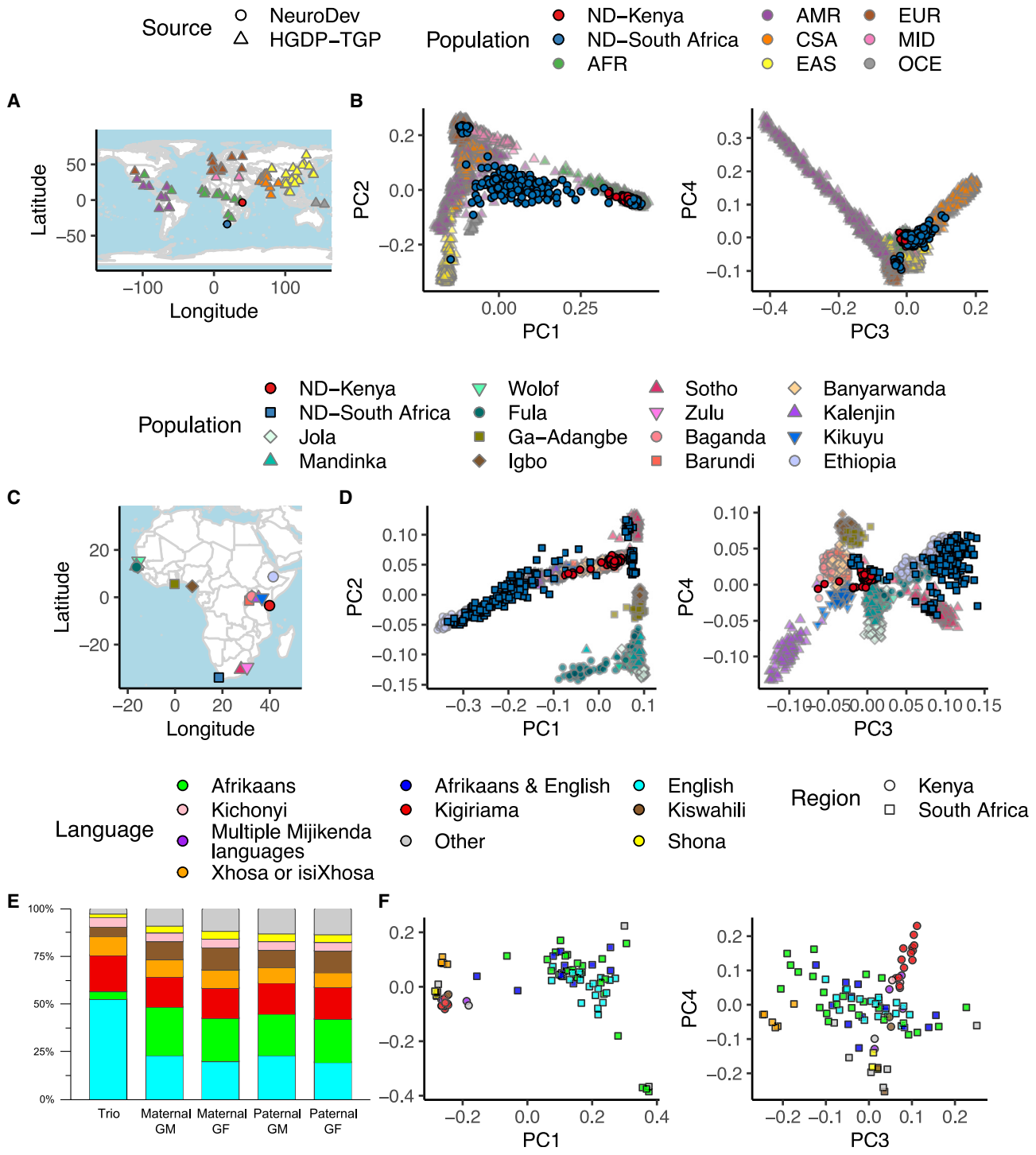

Figure 2. Global and within-Africa population structure of the NeuroDev trio pilot

(A) Map of geographic populations represented in the Human Genome Diversity Panel and 1000 Genomes Project (HGDP-TGP) global reference panel.

(B) Global PCA plots of the NeuroDev cohort. All NeuroDev trios are projected onto the first 4 PCs of the HGDP-TGP reference panel.

(C) Map of geographic populations represented in AGVP, a within-Africa reference panel.

(D) African subcontinental PCA plots of the NeuroDev cohort. A subset of sequenced NeuroDev individuals with African genetic ancestry are projected onto the first 4 PCs of the AGVP panel.

(E) Breakdown of languages spoken across generations by the 99 genetically sequenced NeuroDev families ("Trio") and their grandparents ("GF," grandfather; "GM," grandmother). "Other" spans 24 languages.

(F) PCA of the 99 NeuroDev cases without a reference panel, labeled by the maternal grandmother's primary language.

A total of 13/75 (17.3%) South African cases were solved with pathogenic or likely pathogenic variants in genes with an existing disease association in OMIM.[27] Of these, 5 cases had a single-nucleotide variant (SNV), and 1 case had an insertion or a deletion (indel) (Tables 1 and S3). An additional 7 cases had structural variants that included a known OMIM gene (Tables 2 and S4). Nearly all of the events were *de novo*, with the exception of one structural variant that was paternally inherited (unknown paternal history of psychiatric diagnoses or NDDs). A total of 9/24 (37.5%) Kenyan cases were solved. Of these, we found that 6 cases had a pathogenic or likely pathogenic SNV in genes previously associated with an NDD (Table 1), and 3 cases had a structural variant including a known OMIM gene (Tables 2 and S4).

In addition to the solved cases, 7 VUSs were considered interesting candidate variants based on matches made through the MME (Tables 1 and S5). By matching cases of similar phenotypic and genotypic profiles, MME provides a rapid, systematic approach to rare disease gene discovery.[30] The 7 genes containing these variants were not yet listed in OMIM at the time of analysis and represent emerging gene-disease associations where additional evidence and cases are still needed. All 7 variants will be included in collaborative case series reports on the genes of interest. To date, case studies on three of these genes—*AGO1*, *CACNA1C*, and *CACNA1E*—have been published.[31–33] These variants may potentially be reclassified after re-evaluation over time.

NeuroDev participants comprised the only geographic African cases in any of the case series reports, and NeuroDev will be the first African NDD cohort to contribute, at scale, to rare disease discovery activities. As described by the NHGRI Atlas of Human Malformations initiative, syndromic NDDs often vary in their phenotypic presentation between ancestral groups,[34] a phenomenon that is particularly well documented concerning canonical facial features. Further analysis of how phenotypic features differ between ancestral groups will help clarify which features are "core" to a genetic syndrome and which vary based on ancestral group or environment.

DISCUSSION

People of African ancestry have been grossly underrepresented in genetic studies, across domains and disciplines.[35,36] In aggregate, this is most visible through the constitution of large genetic databases such as gnomAD, in which only 14% of individuals ($N \approx 27{,}000$) have some African ancestry.[37,38] Majority of those 27,000 individuals are American and typically have a mixture of West African and European ancestry.[39,40] If individuals of African ancestry remain underrepresented in genetic research, they will continue to be less likely to receive accurate genetic diagnoses and less likely to benefit from advances in genomic science and medicine.[41–44]

Those living in Africa have greater genetic variability than any other human ancestral group, rendering their underrepresentation a substantial barrier to human genome characterization and scientific equity and medical ethics.[35] There are many reasons that NDD collections have historically been conducted in the United States and western Europe, chief among them being the better accessibility to project funding and infrastructural support. This lack of precedent produced common concern about NeuroDev's feasibility, particularly about the planned case collection rate and the ambitious case characterization schedule. We hope that the results described here will increase the research community's confidence in large-scale data collection efforts in underrepresented populations and that many other studies will be able to contribute to greater African representation.

We also hope the high rate of participant family interest in receiving genetic results will shed light on the need for consistent access to genetic testing services. To date, all NeuroDev participants in South Africa readily consented to have genetic findings related to their child's NDD reported back to them by a clinician. Similarly, NeuroDev South Africa observed a very high (98%) rate of consent to generating and sharing stem cell lines in addition to sharing DNA. This provides us with the opportunity to contribute to diversity in global stem cell collections, which are also heavily biased toward European ancestry.[45] This high degree of participant enthusiasm in South Africa has led to investigations of similar possibilities in Kenya. The NeuroDev Kenya team has recently been funded by a Fogarty award to support, in part, community engagement on the ethics of cell line generation for the cohort in Kilifi.

While NeuroDev provides in-depth phenotypic characterization of its subjects, these data may still reflect regional differences. Kilifi is a rural coastal town with an agriculture-based economy, with an absolute poverty rate of 46.4% and limited access to school-based education.[46] In contrast, Cape Town is a large urban setting and is among the wealthiest cities in Africa, conferring relative educational and resource advantages. Because many behavioral measures are known to be tied to socioeconomic status or maternal education, our behavioral outcomes likely reflect the differences in regional context. We encourage abundant caution when comparing behavioral outcomes across sites.

The 4-year target sample for NeuroDev is 1,800 cases, 2,100 parents of cases, and 1,800 controls, the largest study for NDDs on the African continent. The NeuroDev Project will, in full, create a public resource for medical genetics research that includes genome-wide common variant data; exome-sequencing data; comprehensive phenotypic data including detailed cognitive, behavioral, and health information; cell lines (lymphoblastoid and induced pluripotent cell lines); and photographs capturing dysmorphic features of thousands of African individuals. This critically needed line of work will help address the scientific, medical, and ethical consequences of the genomic research representation gap.

CONSORTIA

The NeuroDev Project members (current and previous) in addition to the named authors include Aleya Zulfikar Remtullah, Alex Macharia, Ann Karanu, Carmen Swanepoel, Claire Fourie, Constance Rehema, Deepika Goolab, Dorcas Kamuya, Dorothy Chepkirui, Este Sauerman, Eunice Chepkemoi, Fagri February, Fatima Khan, Felicita Omari, Gina Itzikowitz, Javan Nyale, Jantina de Vries, Jess Ringshaw, Johnstone Makale, Judy Tumaini,

Table 1. South African and Kenyan SNV/indel findings

Individual	Sex, age	Clinical diagnosis	Gene	Variant	Event	Additional phenotypes of interest	OMIM #	Variant interpretation
South Africa								
859-44206536	male, 2	GDD	*AGO1 (NM_012199.5)*	c.971C>T, (p.Pro324Leu)	missense (MPC = 2.8)	–	–	VUS
859-25391305	male, 7	GDD	*CREBBP (NM_004380.3)*	c.3914+3G>T	extended splice site (CADD = 15)	craniofacial dysmorphia, abnormalities of the genital system, and short stature	#180849	LP
859-49145346	female, 7	GDD	*DDX3X (NM_001356.5)*	c.599A>G, (p.Tyr200Cys)	missense (MPC = 3.1)	small for gestational age and abnormal facial shape	#300958	LP
859-90780619	male, 2	GDD	*IRF2BPL (NM_024496.4)*	c.2137del, (p.Leu713SerfsTer54)	frameshift (CADD = 33)	–	#618088	LP
859-81577625	male, 2	GDD	*PPP2R5C (NM_001161725.1)*	c.254_259del, (p.Asp85_Phe86del)	inframe deletion (CADD = 17)	seizure	–	VUS
859-88385997	male, 6	GDD, ID	*CREBBP (NM_004380.3)*	c.5558A>C, (p.Gln1853Pro)	missense (MPC = 1.6)	hearing impairment, visual impairment, craniofacial dysmorphia, and hypertonia	#618332	LP
859-98643450	male, 7	GDD, ADHD	*SYNGAP1 (NM_006772.3)*	c.3795-1G>A	essential splice site (CADD = 35)	large birth length	#612621	P
859-21383847	male, 2	GDD, autism	*SCN2A (NM_001040142.2)*	c.2877C>A, (p.Cys959Ter)	nonsense (CADD = 37)	–	#613721	P
859-33526476	female, 4	Autism	CACNA1C (NM_001129827.2)	c.4129dup, (p.Arg1377ProfsTer61)	frameshift (CADD = 33)	–	–	VUS
859-76545750	male, 6	Autism	CACNA1E (NM_001205293.3)	c.3422+1G>A	essential splice site (CADD = 35)	tall stature and proximal amyotrophy	–	VUS
859-40050374	female, 3	Autism	*MYH10 (NM_001256012.3)*	c.2555G>A, (p.Arg852Gln)	missense (CADD = 30)	macrocephaly and high BMI	–	VUS
Kenya								
860-61235417	female, 17	ID	*BCL11B (NM_138576.4)*	c.1535_1536del, (p.Ala512GlyfsTer4)	frameshift (CADD = 32)	visual impairment and low BMI	#618092	LP
860-26955427	female, 11	ID	*DDX3X (NM_001356.5)*	c.1582C>T, (p.Arg528Cys)	missense (MPC = 3.2)	small for gestational age and muscle weakness	#300958	LP
860-46028963	male, 12	ID	*MAPK1 (NM_002745.5)*	c.952G>A, (p.Asp318Asn)	missense (MPC = 1.4)	short stature	#619087	VUS
860-71034936	male, 17	ID	*TLK2 (NM_001284333.2)*	c.1655T>C, (p.Leu552Pro)	missense (MPC = 3.4)	short stature	#618050	LP

(*Continued on next page*)

Table 1. *Continued*

Individual	Sex, age	Clinical diagnosis	Gene	Variant	Event	Additional phenotypes of interest	OMIM #	Variant interpretation
860-73911739	female, 17	ID, autism	*DDX3X (NM_001356.5)*	c.894C>A, (p.Cys298Ter)	nonsense (CADD = 24)	short stature and abnormality of facial musculature	#300958	P
860-31257300	female, 14	ID, autism	*ZBTB18 (NM_205768.3)*	c.204_205del, (p.Asp70HisfsTer19)	frameshift (CADD = 29)	–	#612337	LP
860-33192107	male, 6	ID, ADHD	*MBD5 (NM_018328.4)*	c.4170G>A, (p.Trp1390Ter)	nonsense (CADD = 41)	–	#156200	P
860-89059068	female, 15	ID, CD	*SF1 (NM_001178030.1)*	c.737C>T, (p.Pro246Leu)	missense (MPC = 3.0)	visual impairment	–	VUS

Age at recruitment is reported. All variants were de novo. For missense and nonsense variants, missense badness, PolyPhen-2 and constraint (MPC) scores are noted, and Phred-scaled Combined-Annotation Dependent Depletion (CADD) scores are provided where MPC scores are not available.[28] Variants with MPC $\geq$ 2 are known to be significantly enriched in cases with neurodevelopmental disorders. CADD scores rate the deleteriousness of SNVs and are known to be the highest for nonsense variants (~37) and the lowest for intergenic variants (~2).[29] See Tables S3 and S5 for more details.
GDD, global developmental delay; ID, intellectual disability; ADHD, attention deficit hyperactivity disorder; P, pathogenic; LP, likely pathogenic; VUS, variant of unknown significance.

Kegan Chase, Lizette Rooi, Moses Mangi, Moses Mosobo, Megan Page, Nicole Mciver, Nonceba Ngubo, Pauline Samia, Phatiswa Lopoleng Ranyana, Racheal Mapenzi, Rachel Odhiambo, Raphaela Itzikowitz, Rene Lepore, Rizqa Sulaiman-Baradien, Samuel Mwasambu, Shaheen Sayed, Susan Wamithi, Tabith Shali, Thandi Xintolo, and Zandre Bruwer.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - NeuroDev Protocol and Informed Consent Procedures
  - NeuroDev Data Pipeline
  - NeuroDev Subject Recruitment
  - NeuroDev Phenotypic Data Collection
  - Lessons Learned
  - Phenotypic Data Analysis Methods
  - Exome Sequencing & Data Processing Methods
  - Exome Sequencing Data Analysis Process
  - Genotyping and Analysis Process

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.neuron.2023.06.010.

ACKNOWLEDGMENTS

We are extremely grateful to the family members for participating in this research. We are grateful to the clinical laboratories and biobank teams at KEMRI-Wellcome Trust and the University of Cape Town, the community liaison group and neuro-epilepsy clinic at the KEMRI-Wellcome Trust, and developmental and allied clinics at Red Cross Memorial Hospital. We acknowledge James Swanson, Edith Nolan, William Pelham, and team for the use of the SNAP-IV; David Skuse, Richard Warrington, Will Mandy, and team for the use of the 3Di and SCDC; the ASEBA team for the use of the CBCL; Christopher Molteno and team for the use of the Molteno Adapted Scales; and the John C. Raven, John H. Court, and team for the use of the Raven's Progressive Matrices. NeuroDev is supported by the Stanley Center for Psychiatric Research at the Broad Institute, a grant from SFARI (704413, E.B.R.), and by the National Institute of Mental Health of the National Institutes of Health under award number U01MH119689. Research reported in this publication was also supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under award number R01HD102975. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Sequencing was provided by the Broad Institute of MIT and Harvard Center for Mendelian Genomics (Broad CMG) and was funded by the National Human Genome Research Institute; the National Eye Institute; the National Heart, Lung, and Blood Institute grant UM1HG008900; and in part by National Human Genome Research Institute grant R01HG009141.

AUTHOR CONTRIBUTIONS

E.B.R., K.A.D., A.A., and C.N. designed the study. The NeuroDev Project members, including P.K., B.C., E.E., M.K., A.N., B.M., S.M., and K.M., were

Table 2. South African and Kenyan structural variant findings

Individual	Sex, age	Clinical diagnosis	Genomic region (GRCh38)	Event	Associated condition	Additional phenotypes	Variant interpretation
South Africa							
859-10314801	male, 2	GDD	chr3:13371737-20095506	deletion (6.7 Mb)	3p deletion syndrome (#613792)	small for gestational age and microcephaly	P
859-97835206	female, 7	GDD	chr6:115941808-133892653	deletion (18 Mb)	interstitial 6q microdeletion syndrome	ventricular septal defect, seizure, meningitis, and camptodactyly	P
859-44770029	female, 3	GDD	chr18:61490305-80247612	deletion (18 Mb)	chromosome 18q deletion syndrome (#601808)	visual impairment, low birth length, abnormal facial shape, and hypotonia	P
859-41524687	female, 4	GDD	chr22:18985739-21081116	duplication (2.1 Mb)	22q11.2 duplication syndrome (#608363)	high birth length and macrocephaly	P
859-50205427	female, 4	GDD, autism	chr15:22810652-29822566	triplication (7.0 Mb)	15q11-q13 duplication syndrome (#608636)	small for gestational age	P
859-22089821	male, 4	GDD, autism	chr15:30626003-32111997	deletion (1.5 Mb)	15q13.3 microdeletion syndrome (#612001)	small for gestational age and microcephaly	P
859-85638884	male, 7	GDD, autism	chr16:29663598-30188229	deletion (0.53 Mb)	16p11.2 deletion syndrome (#611913)	–	P
Kenya							
860-31775019	male, 17	ID	22:18985739-21081116	deletion (2.1 Mb)	DiGeorge syndrome (#188400)	abnormality of facial musculature and seizure	P
860-45665569	male, 16	ID	22:18985739-21081116	deletion (2.1 Mb)	DiGeorge syndrome (#188400)	–	P
860-94211640	male, 15	ID	22:49883237-50740457	duplication (0.85 Mb)	22q13 duplication syndrome (#615538)	–	P

Age at recruitment is reported. All variants were de novo except in subject 859-22089821 where the variant was paternally inherited. See Table S4 for more details.
GDD, global developmental delay; ID, intellectual disability; P, pathogenic.

involved in data collection. H.A.K., E.O., J.A., S. Bryant, N.B., C.K., P.M., and B.M. were involved in the curation and analysis of the data. P.K., H.A.K., B.C., E.O., and E.B.R. wrote the manuscript with input from all authors. P.K., B.C., E.E., A.G., and K.M. were involved in project administration, and C.A.-T., S. Baxter, H.B., A.L., D.G.M., A.S.-J., M.S.-B., M.E.T., V.d.M., A.M., L.L.N., C.v.d.M., C.N., A.D.-L., A.A., K.A.D., and E.B.R. supervised various aspects of the project and the core project teams.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: July 22, 2022
Revised: January 13, 2023
Accepted: June 16, 2023
Published: July 17, 2023

REFERENCES

1. de Menil, V., Hoogenhout, M., Kipkemoi, P., Kamuya, D., Eastman, E., Galvin, A., Mwangasha, K., de Vries, J., Kariuki, S.M., Murugasen, S., et al. (2019). The NeuroDev study: phenotypic and genetic characterization of neurodevelopmental disorders in Kenya and South Africa. Neuron *101*, 15–19. https://doi.org/10.1016/j.neuron.2018.12.016.
2. Skuse, D.H., Mandy, W.P.L., and Scourfield, J. (2005). Measuring autistic traits: heritability, reliability and validity of the Social and Communication Disorders Checklist. Br. J. Psychiatry *187*, 568–572. https://doi.org/10.1192/bjp.187.6.568.
3. Skuse, D., Warrington, R., Bishop, D., Chowdhury, U., Lau, J., Mandy, W., and Place, M. (2004). The Developmental, Dimensional and Diagnostic Interview (3di): a novel computerized assessment for autism spectrum disorders. J. Am. Acad. Child Adolesc. Psychiatry *43*, 548–558. https://doi.org/10.1097/00004583-200405000-00008.
4. Swanson, J.M., Schuck, S., Porter, M.M., Carlson, C., Hartman, C.A., Sergeant, J.A., Clevenger, W., Wasdell, M., McCleary, R., Lakes, K., et al. (2012). Categorical and dimensional definitions and evaluations of symptoms of ADHD: history of the SNAP and the SWAN rating Scales. Int. J. Educ. Psychol. Assess. *10*, 51–70.
5. Fischbach, G.D., and Lord, C. (2010). The Simons simplex collection: a resource for identification of autism genetic risk factors. Neuron *68*, 192–195. https://doi.org/10.1016/j.neuron.2010.10.006.
6. Robinson, E.B., Samocha, K.E., Kosmicki, J.A., McGrath, L., Neale, B.M., Perlis, R.H., and Daly, M.J. (2014). Autism spectrum disorder severity reflects the average contribution of de novo and familial influences. Proc. Natl. Acad. Sci. USA *111*, 15161–15165. https://doi.org/10.1073/pnas.1409204111.
7. Raven, J. (2000). The Raven's progressive matrices: change and stability over culture and time. Cogn. Psychol. *41*, 1–48. https://doi.org/10.1006/cogp.1999.0735.
8. Charman, T., Jones, C.R.G., Pickles, A., Simonoff, E., Baird, G., and Happé, F. (2011). Defining the cognitive phenotype of autism. Brain Res. *1380*, 10–21. https://doi.org/10.1016/j.brainres.2010.10.075.
9. DiStefano, C., Sadhwani, A., and Wheeler, A.C. (2020). Comprehensive assessment of individuals with significant levels of intellectual disability: challenges, strategies, and future directions. Am. J. Intellect. Dev. Disabil. *125*, 434–448. https://doi.org/10.1352/1944-7558-125.6.434.
10. Jeste, D.V., Palmer, B.W., Appelbaum, P.S., Golshan, S., Glorioso, D., Dunn, L.B., Kim, K., Meeks, T., and Kraemer, H.C. (2007). A new brief instrument for assessing decisional capacity for clinical research. Arch. Gen. Psychiatry *64*, 966–974. https://doi.org/10.1001/archpsyc.64.8.966.
11. Seaman, J.B., Terhorst, L., Gentry, A., Hunsaker, A., Parker, L.S., and Lingler, J.H. (2015). Psychometric properties of a decisional capacity screening tool for individuals contemplating participation in Alzheimer's disease research. J. Alzheimers. Dis. *46*, 1–9. https://doi.org/10.3233/JAD-142559.
12. Campbell, M.M., Susser, E., Mall, S., Mqulwana, S.G., Mndini, M.M., Ntola, O.A., Nagdee, M., Zingela, Z., Van Wyk, S.V., and Stein, D.J. (2017). Using iterative learning to improve understanding during the informed consent process in a South African psychiatric genomics study. PLoS One *12*, e0188466. https://doi.org/10.1371/journal.pone.0188466.
13. Achenbach, T.M., Dumenci, L., and Rescorla, L.A. (2001). Ratings of relations between DSM-IV diagnostic categories and items of the CBCL/6-18, TRF, and YSR (VT Universidad), pp. 1–9. https://aseba.org/wp-content/uploads/2019/02/DSM6-18ratings.pdf.
14. Bakare, M.O., Ebigbo, P.O., and Ubochi, V.N. (2012). Prevalence of autism spectrum disorder among Nigerian children with intellectual disability: a stopgap assessment. J. Health Care Poor Underserved *23*, 513–518. https://doi.org/10.1353/hpu.2012.0056.
15. Mpaka, D.M., Okitundu, D.L.E.-A., Ndjukendi, A.O., N'situ, A.M., Kinsala, S.Y., Mukau, J.E., Ngoma, V.M., Kashala-Abotnes, E., Ma-Miezi-Mampunza, S., Vogels, A., et al. (2016). Prevalence and comorbidities of autism among children referred to the outpatient clinics for neurodevelopmental disorders. Pan Afr. Med. J. *25*, 82. https://doi.org/10.11604/pamj.2016.25.82.4151.
16. Bello-Mojeed, M.A., Omigbodun, O.O., Bakare, M.O., and Adewuya, A.O. (2017). Pattern of impairments and late diagnosis of autism spectrum disorder among a sub-Saharan African clinical population of children in Nigeria. Glob. Ment. Health (Camb) *4*, e5. https://doi.org/10.1017/gmh.2016.30.
17. Hahler, E.-M., and Elsabbagh, M. (2015). Autism: a global perspective. Curr. Dev. Disord. Rep. *2*, 58–64.
18. Mazurek, M.O., Handen, B.L., Wodka, E.L., Nowinski, L., Butter, E., and Engelhardt, C.R. (2014). Age at first autism spectrum disorder diagnosis: the role of birth cohort, demographic factors, and clinical features. J. Dev. Behav. Pediatr. *35*, 561–569. https://doi.org/10.1097/DBP.0000000000000097.
19. Ruparelia, K., Abubakar, A., Badoe, E., Bakare, M., Visser, K., Chugani, D.C., Chugani, H.T., Donald, K.A., Wilmshurst, J.M., Shih, A., et al. (2016). Autism spectrum disorders in Africa: current challenges in identification, assessment, and treatment: a report on the International Child Neurology Association Meeting on ASD in Africa, Ghana, April 3–5, 2014. J. Child Neurol. *31*, 1018–1026.
20. Liao, C., Moyses-Oliveira, M., Esch, C.E.D., Bhavsar, R., Nuttle, X., Li, A., Yu, A., Burt, N.D., Erdin, S., Fu, J.M., et al. (2022). Transcriptional patterns of coexpression across autism risk genes converge on established and novel signatures of neurodevelopment. Cell Genomics *3*, 100277. https://doi.org/10.1101/2022.02.28.22271620.
21. Weiner, D.J., Wigdor, E.M., Ripke, S., Walters, R.K., Kosmicki, J.A., Grove, J., Samocha, K.E., Goldstein, J.I., Okbay, A., Bybjerg-Grauholm, J., et al. (2017). Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. Nat. Genet. *49*, 978–985. https://doi.org/10.1038/ng.3863.
22. Koenig, Z., Yohannes, M.T., Nkambule, L.L., Goodrich, J.K., Kim, H.A., Zhao, X., Wilson, M.W., Tiao, G., Hao, S.P., Sahakian, N., et al. (2023). A harmonized public resource of deeply sequenced diverse human genomes. Preprint at bioRxiv. https://doi.org/10.1101/2023.01.23.525248.
23. Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M.O., Choudhury, A., et al. (2015). The African Genome Variation Project shapes medical genetics in Africa. Nature *517*, 327–332. https://doi.org/10.1038/nature13997.
24. Atkinson, E.G., Dalvie, S., Pichkar, Y., Kalungi, A., Majara, L., Stevenson, A., Abebe, T., Akena, D., Alemayehu, M., Ashaba, F.K., et al. (2022). Genetic structure correlates with ethnolinguistic diversity in eastern and

southern Africa. Am. J. Hum. Genet. *109*, 1667–1679. https://doi.org/10.1016/j.ajhg.2022.07.013.

25. Gallego Llorente, M.G., Jones, E.R., Eriksson, A., Siska, V., Arthur, K.W., Arthur, J.W., Curtis, M.C., Stock, J.T., Coltorti, M., Pieruccini, P., et al. (2015). Ancient Ethiopian genome reveals extensive Eurasian admixture in eastern Africa. Science *350*, 820–822. https://doi.org/10.1126/science.aad2879.
26. Pais, L.S., Snow, H., Weisburd, B., Zhang, S., Baxter, S.M., DiTroia, S., O'Heir, E., England, E., Chao, K.R., Lemire, G., et al. (2022). seqr: a web-based analysis and collaboration tool for rare disease genomics. Hum. Mutat. *43*, 698–707. https://doi.org/10.1002/humu.24366.
27. Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F., and Hamosh, A. (2015). OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic Acids Res. *43*, D789–D798. https://doi.org/10.1093/nar/gku1205.
28. Samocha, K.E., Kosmicki, J.A., Karczewski, K.J., O'Donnell-Luria, A.H., Pierce-Hoffman, E., MacArthur, D.G., Neale, B.M., and Daly, M.J. (2017). Regional missense constraint improves variant deleteriousness prediction. Preprint at bioRxiv. https://doi.org/10.1101/148353.
29. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46*, 310–315. https://doi.org/10.1038/ng.2892.
30. Philippakis, A.A., Azzariti, D.R., Beltran, S., Brookes, A.J., Brownstein, C.A., Brudno, M., Brunner, H.G., Buske, O.J., Carey, K., Doll, C., et al. (2015). The Matchmaker Exchange: a platform for rare disease gene discovery. Hum. Mutat. *36*, 915–921.
31. Rodan, L.H., Spillmann, R.C., Kurata, H.T., Lamothe, S.M., Maghera, J., Jamra, R.A., Alkelai, A., Antonarakis, S.E., Atallah, I., Bar-Yosef, O., et al. (2021). Phenotypic expansion of CACNA1C-associated disorders to include isolated neurological manifestations. Genet. Med. *23*, 1922–1932. https://doi.org/10.1038/s41436-021-01232-8.
32. Royer-Bertrand, B., Jequier Gygax, M., Cisarova, K., Rosenfeld, J.A., Bassetti, J.A., Moldovan, O., O'Heir, E., Burrage, L.C., Allen, J., Emrick, L.T., et al. (2021). De novo variants in CACNA1E found in patients with intellectual disability, developmental regression and social cognition deficit but no seizures. Mol. Autism *12*, 69. https://doi.org/10.1186/s13229-021-00473-3.
33. Schalk, A., Cousin, M.A., Dsouza, N.R., Challman, T.D., Wain, K.E., Powis, Z., Minks, K., Trimouille, A., Lasseaux, E., Lacombe, D., et al. (2022). De novo coding variants in the AGO1 gene cause a neurodevelopmental disorder with intellectual disability. J Med. Genet. *59*, 965–975. https://doi.org/10.1136/jmedgenet-2021-107751.
34. Muenke, M., Adeyemo, A., and Kruszka, P. (2016). An electronic atlas of human malformation syndromes in diverse populations. Genet. Med. *18*, 1085–1087. https://doi.org/10.1038/gim.2016.3.
35. Martin, A.R., Teferra, S., Möller, M., Hoal, E.G., and Daly, M.J. (2018). The critical needs and challenges for genetic architecture studies in Africa. Curr. Opin. Genet. Dev. *53*, 113–120. https://doi.org/10.1016/j.gde.2018.08.005.
36. Peterson, R.E., Kuchenbaecker, K., Walters, R.K., Chen, C.Y., Popejoy, A.B., Periyasamy, S., Lam, M., Iyegbe, C., Strawbridge, R.J., Brick, L., et al. (2019). Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. Cell *179*, 589–603. https://doi.org/10.1016/j.cell.2019.08.051.
37. Gudmundsson, S., Singer-Berk, M., Watts, N.A., Phu, W., Goodrich, J.K., Solomonson, M., Genome Aggregation Database Consortium, Rehm, H.L., MacArthur, D.G., and O'Donnell-Luria, A. (2022). Variant interpretation using population databases: lessons from gnomAD. Hum. Mutat. *43*, 1012–1030. https://doi.org/10.1002/humu.24309.
38. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443. https://doi.org/10.1038/s41586-020-2308-7.
39. Mathias, R.A., Taub, M.A., Gignoux, C.R., Fu, W., Musharoff, S., O'Connor, T.D., Vergara, C., Torgerson, D.G., Pino-Yanes, M., Shringarpure, S.S., et al. (2016). A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome. Nat. Commun. *7*, 12522. https://doi.org/10.1038/ncomms12522.
40. Zakharia, F., Basu, A., Absher, D., Assimes, T.L., Go, A.S., Hlatky, M.A., Iribarren, C., Knowles, J.W., Li, J., Narasimhan, B., et al. (2009). Characterizing the admixed African ancestry of African Americans. Genome Biol. *10*, R141. https://doi.org/10.1186/gb-2009-10-12-r141.
41. Caswell-Jin, J.L., Gupta, T., Hall, E., Petrovchich, I.M., Mills, M.A., Kingham, K.E., Koff, R., Chun, N.M., Levonian, P., Lebensohn, A.P., et al. (2018). Racial/ethnic differences in multiple-gene sequencing results for hereditary cancer risk. Genet. Med. *20*, 234–239. https://doi.org/10.1038/gim.2017.96.
42. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291. https://doi.org/10.1038/nature19057.
43. Manrai, A.K., Funke, B.H., Rehm, H.L., Olesen, M.S., Maron, B.A., Szolovits, P., Margulies, D.M., Loscalzo, J., and Kohane, I.S. (2016). Genetic misdiagnoses and the potential for health disparities. N. Engl. J. Med. *375*, 655–665. https://doi.org/10.1056/NEJMsa1507092.
44. Popejoy, A.B., Ritter, D.I., Crooks, K., Currey, E., Fullerton, S.M., Hindorff, L.A., Koenig, B., Ramos, E.M., Sorokin, E.P., Wand, H., et al. (2018). The clinical imperative for inclusivity: race, ethnicity, and ancestry (REA) in genomics. Hum. Mutat. *39*, 1713–1720. https://doi.org/10.1002/humu.23644.
45. Negoro, T., Okura, H., and Matsuyama, A. (2017). Induced pluripotent stem cells: global research trends. BioResearch Open Access *6*, 63–73. https://doi.org/10.1089/biores.2017.0013.
46. World Bank (2018). Kenya gender and poverty assessment 2015–2016: reflecting on a decade of progress and the road ahead (World Bank). https://elibrary.worldbank.org/doi/abs/10.1596/31285.
47. Abubakar, A., Gona, J.K., Kipkemoi, P., Rimba, K., Amukambwa, D., and Newton, C.R.J.C. (2022). Perspectives of key stakeholders on educational experiences of children with autism spectrum disorders at the Kenyan Coast. Afr. J. Disabil. *11*, 847. https://doi.org/10.4102/ajod.v11i0.847.
48. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. *17*, 405–424. https://doi.org/10.1038/gim.2015.30.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
The NeuroDev Trio Pilot Whole-Exome Data and Genotyping Arrays	This paper	AnVIL_CMG_Broad_Brain_NeuroDev_WES on https://anvilproject.org/data/
Human Genome Diversity Project & 1000 Genomes Projects Reference Data	Koenig et al.[22]	https://gnomad.broadinstitute.org/downloads
African Genome Variation Project	Gurdasani et al.[23]	EGAD00001003319
Software and algorithms		
seqr	Pais et al.[26]	https://seqr.broadinstitute.org/
Hail v0.2	The Hail Team, Broad Institute	https://github.com/hail-is/hail
GWASpy v0.1	Analytical and Translational Genetics Unit, Broad Institute	https://github.com/atgu/GWASpy

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact: Elise Robinson, The Broad Institute of MIT and Harvard, Cambridge MA, USA (erob@broadinstitute.org).

Materials availability

This study did not generate unique reagents

Data and code availability

NeuroDev data can be accessed through the AnVIL: AnVIL_CMG_Broad_Brain_NeuroDev_WES controlled access data repository (https://anvilproject.org/data/). All reference panels used in the paper are public resources. Code to generate phenotypic analyses, PCA, and figures are available on our GitHub repository (https://github.com/atgu/neurodev-pilot).

METHOD DETAILS

NeuroDev Protocol and Informed Consent Procedures

NeuroDev's information and consent document is the culmination of a series of focus groups and meetings across both sites that were conducted to help develop the language of the document so that it could communicate complex scientific terms in a manner accessible to research participants who may not have completed formal schooling. There were three focus groups in total. The first included parents of children with NDDs, the second included community leaders from NGOs, centers, schools, etc. The third included healthcare workers who work in clinics with NDDs. The key outcomes from the focus group meetings were: Firstly, that lengthy detailed explanations hindered understanding of the scientific terms. Participants tended to be more confused and overwhelmed by lengthy explanations. Therefore, information in the consent document is divulged in ways that keep descriptions short and concise. Secondly, the use of visual examples to help explain DNA were very helpful. For example, pointing out physical similarities between family members as a way of illustrating the inheritance of DNA. Thirdly, it is easier for participants to understand concepts if examples are used that participants are familiar with. For example, when explaining that the information gathered in our research can be used for other purposes. We provide examples of well-known diseases such as diabetes and cancer. This helped participants easily relate to the conveyed concepts. In Kilifi, there was an emphasis on community engagement activities involving discussing what genes are and how variation in genes is associated with neurodevelopmental disorders. This was carried out during meetings with parents of children with neurodevelopmental disorders, teachers and community members.

After an information sharing session and before potential participants consent to enrolment into the study, the UBACC is administered. The UBACC creates an opportunity for discussion with participants around aspects of the protocol they may not have grasped when the study is first explained to them. It has also shown to be effective in identifying participants who may not fully comprehend the protocol. The South Africa Team have had a small number of failures on the test. These were based on 1) the language of the test (i.e. French) and 2) comprehension difficulties of the participant (one participant had ID, and the other had a history of strokes and depression).

The novelty of genetic research in the African context means that the study is important for its potential benefit to African medical and research communities. It is, by necessity, also precedent-setting. There is a lack of precedent around important issues, such as the return of incidental findings to our participants and lack of qualified personnel such as genetic counselors in many parts of sub-Saharan Africa. The evidence base to support feedback of most incidental genomic findings for African populations is not robust. There is virtually no empirical data describing relevant African stakeholders' preferences and perspectives, including research participants, ethics committee members, researchers, and research regulators on these issues. The ethics arm of the NeuroDev study is currently investigating this and will add evidence on this pertinent issue.

NeuroDev Data Pipeline

Data is collected on tablets using the REDCap Mobile App during assessments. Every participant is assigned a subject ID and has an intake form filled out. Based on this information, branching logic programmed into REDCap determines which tools and fields populate for the assessor to complete. At the end of an assessment day, data from the tablets is uploaded to the NeuroDev project on the REDCap servers. The team at the Broad Institute then uses the *scred* package – a Python implementation of the REDCap API – to pull raw records by subject ID and assign codes to missing data to differentiate between truly missing fields and non-applicable fields based on the REDCap branching logic. A Python sync engine then organizes the record data by assessment tool, scores assessments like the Raven's Progressive Matrices and 3Di, and pushes it to a MySQL database hosted on the Google Cloud. The sync procedure is generally run on a weekly basis to keep the database current. All summary and quality-control (QC) reporting is derived from either direct SQL queries to the database or export to pandas dataframes for analysis with Python scripts.

The QC reporting process relies on querying data from the Broad Database and REDCap logging files to produce a comprehensive report of missing, overwritten, or inconsistent participant data. Run on either a date range or a list of specific subject IDs, the program produces summary reports on any of five distinct metrics: completion of REDCap forms, missing required data fields (regardless of whether a form was marked complete), comments entered by assessors at the sites, logical cross-checks on completed data, and fields found to be overwritten in the REDCap log. Subject and site IDs as well as interview times and assessor names are included on each of these reports so the Broad team can easily identify both single issues and larger patterns. Run and reviewed weekly, the program improves data visibility and allows consistent feedback to the sites on improving or correcting collection procedures.

NeuroDev Subject Recruitment

NeuroDev participants were recruited from two sites, Red Cross War Memorial Hospital in Cape Town, South Africa and KEMRI-Wellcome Trust Research Program in Kilifi, Kenya. Participants were recruited from previous studies, specialized clinics and special schools in Kilifi County, Kenya. For the resampling of participants of earlier studies, the team focused on children about to cross the 18-year-old threshold into adulthood. Participants in Cape Town were recruited from developmental clinics using a combination of clinician referral, waiting room flyers and adverts, and cold recruitment from the Red Cross Memorial Children's Hospital Department of Pediatrics and Child Health Neurodevelopmental Clinic. Cases and affected siblings included in the study had clinical diagnosis of a neurodevelopmental disorder, were within the specified age range (2-18 years old), and willing to participate. Cases were excluded if they had a co-occuring primary neuro-motor condition such as cerebral palsy or Downs Syndrome. Controls were included in the study if they did not have a diagnosis of a neurodevelopmental disorder, were within the study age range and were matched according to catchment area, ancestry and age. In Kenya, the recruitment of controls also drew from the Kilifi health demographic surveillance system (KHDSS) cohort. The KHDSS, which has been active since 2002, surveys a representative community sample every three months. Control participants in Cape Town were recruited from various outpatient clinics at the Red Cross War Memorial Children's Hospital.

Written consent was sought for participation of the child controls and child cases from their parents or caregivers, and additional consent was sought from parents of the cases for their own participation in the study. Ethical approval was sought in the site institutional review boards as well as at the Harvard T.H Chan School of Public Health. In Kilifi, Kenya, approval was granted by the Scientific Ethics and Review Unit (KEMRI/SERU/CGMR-C/104/3629) and Health Research Ethics Committee (HREC REF:810/2016) in Cape Town, South Africa.

NeuroDev Phenotypic Data Collection

Assessments

Molteno Adapted Scales. The Molteno has proven to be quite effective in helping us capture the neurodevelopmental profile of children under six years of age. The test's flexible approach to item administration and incidental observations has allowed us to test and score the behavior of even the most non-compliant children. The Molteno had not been validated and adapted for use in Kilifi. We pre-piloted the Molteno, recruiting a random subset of 100 children between 2-5 years in the community and their parents. A small number of items have had to be adapted to make them more appropriate for our population, e.g., "child has a tricycle at home" was changed into "child can walk on his/her tiptoes" along with the administration of a few fine-motor related items.

Swanson, Nolan and Pelham (SNAP) ADHD-IV Rating Scale. The SNAP was insightful and was straightforward in its administration in child cases over six years of age. But it presented some limitations for preschool-aged children. Specifically, some items refer to homework and other tasks that would be cognitively challenging and therefore not a representative skill for children who are not yet of school-going age. The instructions for the test administration did not allude to a lower age limit. After a review of the test literature, it

was determined that the test was not suitable for children under six years of age and would only be administered to children over the age of 6 years.

Child Behavior Checklist (CBCL). The CBCL was implemented into the NeuroDev protocol to help validate the SNAP. This would form a central component of a Master's research project. Because of the limited utility within the broader aims of the protocol, the test was initially piloted, and the utility would later be evaluated. The CBCL was kept on beyond the pilot stage because it proved to be valuable within the clinical context in both sites as it captures clinically rich information that will help phenotype our population that other tests in the phenotypic battery might not, e.g. attention in children under six years, which was lost to the study when the SNAP was amended. The data from the test will also be beneficial to two doctoral students attached to NeuroDev.

When the NeuroDev Kenya team began administering the CBCL, it was noted there were some questions that parents struggled with when giving responses and requested clarification of the questions. Some of the questions that have come up as needing elaboration included: Argues a lot, Fails to finish things he/she starts, There is very little he/she enjoys, Bragging, boasting, Can't get his/her mind off certain thoughts; obsessions, Confused or seems to be in a fog, Daydreams or gets lost in his/her thoughts, Strange behavior, Strange ideas, Wishes to be of the opposite sex.

The NeuroDev Kenya team discussed this and came up with a few changes in the assessment flow. The team started the behavioral assessment battery with the CBCL and thereafter the 3Di, SNAP, SCDC. It was noted that the 3Di, which includes standard examples, could have primed the caregivers to expect some examples. The team also added the prompt that apart from the aim of the assessment being exploring children's behavior: "This assessment describes behaviors of children from a wide range of ages (6-18 years) as such you might find that some questions may seem unsuitable, we would still like to hear from you, to be best of your knowledge, the extent to which your child may have these behaviors."

Developmental Dimensional and Diagnostic Interview (3Di). The 3di has proven to be a concise evaluative tool for the presence of autistic features in children. It is well-received and requires only minor additional probing for its questions. We have found that, in both the South African and Kenyan context, parents with children who have been diagnosed with autism often go through the questionnaire with ease, likely because of their familiarity with the item constructs. In contrast, parents of children who do not demonstrate autistic behaviors generally require more probing, possibly because they are less familiar with the item constructs.

Raven's Progressive Matrices. The RPM has generally been well-received by our participants; however, we have noticed that on a few occasions participants will make errors in the initial teaching items which suggest that this method of reasoning, i.e., deductive reasoning, is more novel within our context than the test has accounted for. For instance, some participants may choose to select their responses based on matching the features of the response items to those in the stimulus box, instead of deducing properties of the missing image, based on the pattern in the stimulus box. This type of error turns the test into a perceptual test for some participants (at least initially) instead of a test of deductive reasoning.

Blood draw. The blood draw can be a discomforting experience of data collection – for the research team, the parents and, most importantly, the child. The management strategy of each blood draw takes account of the idiosyncrasies of each child, in order to reduce any potential distress, time, and the possibility of injury. The teams have found that having a frank and upfront conversation with the parent(s) of the child about what to expect and to determine how they feel their child will respond, is best. Some children will respond best if they know what will happen. The South African team developed "social stories"; a visual aid tool that explains each step of the process to the child using simple language and pictures. The story is written in the first-person, from the point of view of the child, helping the child visualize the experience. The Kenya team adapted and translated the social stories for use in their context as well. We conducted a saliva pilot study on a subset of participants from whom saliva was obtained in addition to a blood sample. Of the 131 samples, 97.15% of DNA extracted passed sample QC, indicating that saliva would be a strong option for DNA in future cases.

Photos. Participants in South Africa appear to be most apprehensive about having their children's photos taken. As of December 2019 76%, of participants have consented to having their photo taken. Participants often report wanting to protect their children's privacy and preventing their children's images falling into the hands of unscrupulous characters as reasons for not consenting to the photos. In Kenya, the use of photos was a point of concern during the approval process, as a result, the use of photos for publication and other public use is a twofold process. During the consenting process, participants are informed that the taking of photos is a core part of the assessment process and that before a photo is shared to other research teams, that we will re consent them to get permission to use the selected photos.

Cell lines and stem cells in South Africa. The South Africa team has observed a much higher rate of consent for cell lines than previously anticipated. As of December 2019, 96% of participants had consented to cell lines. The Kenya team has not implemented cell line and stem cell collection. The ethics arm of the study is collating views from community members and key stakeholders on collection of cell lines and will present these findings in the coming years.

Case ascertainment

Case ascertainment In Kenya. Intellectual Disability is more readily ascertained in Kenyan Special Schools through the Educational Assessment and Resource Centers, however, autism diagnosis is not as readily ascertained, especially when comorbid with another NDD.[47] As some of the Kenyan cases are recruited from Special Schools, to ensure we do not miss cases with autism, we began clinically assessing children that came in with only a reported diagnosis of ID, but who in the course of the NeuroDev assessments battery presented with many autism traits (i.e., they endorse a lot of the symptoms on the 3Di). Using their clinical judgment and a general endorsement pattern on the 3di, in particular the repetitive behaviors section questions, the case was further assessed by

an experienced clinician on the team and a psychologist using the DSM-5 checklist and clinical judgment to confirm whether the child has autism.

Case ascertainment in South Africa. Children with a known NDD diagnosis were recruited from the neurodevelopmental and genetic outpatient clinics from Red Cross War Memorial Children's Hospital and Tygerberg Hospital in Cape Town, South Africa. Eligible cases were identified through clinician referral. Additionally, families in the clinic waiting rooms were approached and screened for eligibility and invited to take part in the study. Both hospital clinics are specialist developmental outpatient clinics, which are led by relevant sub-specialists (Paediatrics, Paediatric Neurology, Medical Genetics and Developmental Paediatrics). Children enrolled in the study have been formally diagnosed by these teams through standardized clinical assessments and benchmarked against the DSM-5 criteria.

Lessons Learned

Reflection on parent-recalled dates in Kenya. The Kenyan cohort is generally older in comparison to the South Africa cohort, as such recall of certain milestones and ages can be challenging for the caregivers. The team uses anchoring when trying to help parents answer an age-related question. First using age bands: Under 1 year, 12-18 months (6 months after a year), 2 years (how many months before or after?). Thereafter, they also employ the use of cultural events and family events to enhance the age anchoring technique.

General approach to data collection. The case children are highly variable in terms of their type of neurodevelopmental disorder. As such there is great variation in how each family approaches dealing with their child's condition. The researchers have adopted a series of principles that encourages sensitivity and reflexivity to the needs and expectations of each family during data collection. These include maintaining a sense of calm and patience during testing to reduce collective anxiety; being respectful of each family's journey and process; taking the time to explain what is going on to the child, even if they are non-verbal and look as if they don't understand what you're saying; trust and rely on each other as a team and being open to learn from each other; and ensure the researchers, the participants, and their families are ready for the blood draws as they can be very traumatic.

Follow-up after scheduling of assessments. Throughout the study, the teams incorporated intensive follow-up efforts to increase the rate with which fathers participated in the study, to increase the trio collection. They were followed-up regularly and allowed flexible scheduling, sometimes during other data collection slots (this was possible because only one research nurse/clinician was needed to complete data collection of returning fathers: including consenting, cognitive testing and blood draw) to accommodate logistical reasons for not attending data collection. In Kilifi, fathers are often away at work during the day, when the recruitment team does the home visit, and work most of the week and may not be able to get time off work. To work around this, the study team arranged for 'Father's Day' Data collection days on Saturday.

Phenotypic Data Analysis Methods

Self-reports of ancestry and language were collected as part of the Demographics tool on REDCap. To assess developmental or intellectual delays, we administered either the Molteno or the Raven's Progressive Matrices (RPM) to their respective age groups and measured rates of completion. Successful completion of the Molteno was defined as the completion of at least 2 of the 4 domains implicated in the assessment. Completion of the RPM was defined as being able to complete all questions on the assessment. Proportions of testers who successfully completed the respective tests at age ranges of 0-5 (Molteno), 6-11 (Colored RPM), and greater than 12 (Standard RPM) was computed and reported. For subjects who could not complete age-appropriate assessments, the rates of completion of tests below their age level was reported for each age group.

Neurodevelopmental diagnosis of cases was collected through the Neuromedical Assessment tool. The cases were placed into four distinct categories: autism, GDD/ID, autism and GDD/ID, and other diagnoses (which included ADHD, communication disorder, and specific learning disorders). Children that met DSM-5 criteria for either GDD or ID or were diagnosed with at least borderline delay from the Molteno were included in the GDD/ID category. For each of these four categories, we assessed the proportions of those with fluent speech levels. Speech fluency was determined by the 3Di, which assesses the child's speech on a scale of no words, single words, multiple words, and fluent speech based on parent interview.

The Simons Simplex Collection (SSC) data set was used to compare phenotypic outcomes of the 122 cases with autism ascertained in NeuroDev. The SSC consists of a deeply phenotyped sample of more than 2500 families with a child diagnosed with autism in the United States.[5] We grouped cases with autism from both datasets based on the number of adverse co-occurring neurological and developmental outcomes, including ID, a positively associated history of seizures, and motor delays (defined as either a gross motor diagnosis from the Molteno, or an age at first steps higher than 18 months). In SSC, ID was defined as cases with an IQ < 70. From the symptom distributions within each study, the average number of co-occurring symptoms was computed, and the difference between these means was assessed using a Mann-Whitney U test.

Exome Sequencing & Data Processing Methods

Data generation and analysis were done in collaboration with the Broad Institute Center for Mendelian Genomics, with sequencing performed at the Genomics Platform and data processing at the Data Sciences Platform of the Broad Institute of MIT and Harvard. Libraries from DNA samples were created with a Twist exome capture (37 Mb target) and sequenced (150 bp paired reads) to >85% of targets at >20x, comparable to ~55x mean coverage. Sample identity quality assurance checks were performed on each sample.

The exome sequencing data was de-multiplexed and each sample's sequence data aggregated into a single Picard CRAM file. Exome data was processed through a pipeline based on Picard, using base quality score recalibration and local realignment at known indels, aligned to the human genome build 38 using BWA, and jointly analyzed for single nucleotide variants (SNVs) and insertions/deletions (indels) using Genome Analysis Toolkit (GATK) Haplotype Caller package version 4.0.10.1. After variant calling, sex, ancestry and relatedness to other samples were inferred using the CMG sample QC pipeline and compared to sample metadata to identify and correct sample swaps, basic functional annotation was performed using Variant Effect Predictor (VEP). The joint variant call file was then uploaded to the *seqr* platform and imported to Hail version 0.2 for further annotation and analysis.

Copy-number variants (CNVs) were discovered from the exome sequencing data following GATK-gCNV best practices. Read coverage was calculated for each exome using GATK CollectReadCounts. After coverage collection, all samples were subdivided into batches for gCNV model training and execution; these batches were determined based on a principal components analysis (PCA) of sequencing read counts. After batching, one gCNV model was trained per batch using GATK GermlineCNVCaller on a subset of training samples, and the trained model was then applied to call CNVs for each sample per batch. Finally, all raw CNVs were aggregated and post-processed using quality- and frequency-based filtering to produce the final CNV callset.

Exome Sequencing Data Analysis Process

Upon completion of data generation, both the SNV/indel and CNV callsets were uploaded to *seqr*, the centralized genomic analysis platform used by the Broad Institute's Center for Mendelian Genomics (CMG), for clinical genetics analyses. The CMG analysis team deployed a standard analysis protocol across all NeuroDev trios to identify potential disease-causing variants. The first round of analysis consisted of a review of variants from a *de novo*/dominant and a recessive search. The *de novo*/dominant search filters for variants present in affected family members and absent from unaffected family members. The variant types returned from this standard search include deletions, duplications, protein truncating variants, and missense variants that have an allele frequency <0.1% across population databases (gnomAD v2/v3/SV, 1000 Genomes, ExAC, and TopMed) and <1% in the CMG internal rare disease dataset, that pass QC, and that have a GQ $\geqslant$20 and an allele balance $\geqslant$0.2. The recessive search returns homozygous recessive, compound heterozygous, and X-linked recessive variants in affected individuals, considering phasing for any available parents. It searches for biallelic variants across deletions, duplications, protein truncating variants, or missense variants that have an allele frequency <1% across population databases and <3% in the CMG internal rare disease dataset, that pass QC, and that have a GQ $\geqslant$20 and an allele balance of $\geqslant$0.2. If candidate variants were not identified after the first round of analysis, the search criteria was adjusted to include additional variant annotation types (synonymous variants, extended splice site variants, and 5' and 3' UTR variants), and quality parameters were relaxed to allow for the review of indels that did not pass QC.

Potential causal variants were subjected to rigorous evaluation of the evidence for pathogenicity following criteria established by the American College of Medical Genetics and Association for Molecular Pathology.[48] The study's primary focus was on well-established disease genes, using information drawn from a variety of sources such as OMIM. The phenotype data of NeuroDev participants was assessed for possible consistencies with the previously reported clinical presentations associated with known disease genes, bearing in mind potential deviations from expectation due to ancestry.

Genes that are not yet associated with a well-established human disease in OMIM were carefully evaluated using a variety of sources of evidence, including constraint scores, transcript and protein expression databases, model organism data, and a review of the available literature. All variants identified in candidate disease genes were entered into the Matchmaker Exchange (MME) network through *seqr* in order to identify additional cases with overlapping phenotypes and variants in the same candidate genes, to help better characterize potential novel gene-disease relationships.

Genotyping and Analysis Process

Genotyping was performed on all samples using the Illumina Infinium Global Screening Array (GSA) at the Genomics Platform at the Broad Institute, which is supported by a LIMS-tracked, automated processing system that utilizes various sample handling robots and Illumina iScans to generate intensity data from the arrays. After scanning, the data is processed through our automated genotype calling pipeline. The platform releases raw data files (idats, gtc) as well as VCF, and allows for the conversion to PLINK ready called genotype files (ped/map) generated using two different calling algorithms: Illumina GenCall (Autocall) for common variants, and zCall for rare variants including the custom exome/clinical content.

We applied the GWASpy pipeline (https://github.com/atgu/GWASpy) pre-imputation QC module to perform variant-level QC. We used default filters, including MAF, call rate, Mendelian error rate, sex checks, inbreeding coefficient, and Hardy-Weinberg disequilibrium. To visualize genetic diversity across NeuroDev samples, we also performed a principal component analysis (PCA) on array data. The PCA module in GWASpy applied LD-pruning, relatedness estimates, and other necessary filters. We used the joint Human Genome Diversity Project and 1000 Genomes Project reference as a global reference, and the African Genome Variation Project as the subcontinental reference in our population PCA plots. We performed projection PCA, in which the reference data is used to define the principal components and NeuroDev data is projected onto that space, to prevent relatedness effects from our family-based data.