The NeuroDev Study: Phenotypic and Genetic Characterization of Neurodevelopmental Disorders in Kenya and South Africa

Victoria de Menil,[1,2,11] Michelle Hoogenhout,[3,11] Patricia Kipkemoi,[4,11] Dorcas Kamuya,[5] Emma Eastman,[3] Alice Galvin,[2,6] Katini Mwangasha,[4] Jantina de Vries,[7] Symon M. Kariuki,[1,4] Serini Murugasen,[3] Paul Mwangi,[4] Ilina Singh,[8] Dan J. Stein,[9] Amina Abubakar,[4,12,*] Charles R. Newton,[1,4,12] Kirsten A. Donald,[3,12,*] and Elise Robinson[2,6,10,12,*]

[1]Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford OX3 7JX, UK
[2]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, 75 Ames Street, Cambridge, MA 02142, USA
[3]Department of Paediatrics & Child Health, 4th Floor ICH Building, Red Cross War Memorial Children's Hospital and University of Cape Town, Rondebosch, South Africa
[4]Neurosciences Unit, Clinical Department, KEMRI-Wellcome Trust Collaborative Research Programme, PO Box 230-80108, Kilifi, Kenya
[5]Department of Health Systems and Research Ethics, KEMRI-Wellcome Trust Collaborative Research Programme, Kilifi, Kenya
[6]Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA
[7]Department of Medicine, University of Cape Town, Observatory, Cape Town 7925, South Africa
[8]Department of Psychiatry and Wellcome Centre for Ethics and Humanities, University of Oxford, Oxford, UK
[9]Department of Psychiatry & Mental Health, and MRC Unit on Risk and Resilience in Mental Disorders, University of Cape Town, J-Block, Groote Schuur Hospital, Observatory, Cape Town, South Africa
[10]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Richard B. Simches Building, 6th Floor, 185 Cambridge Street, Boston, MA 02114, USA
[11]These authors contributed equally
[12]These authors contributed equally
*Correspondence: aabubakar@kemri-wellcome.org (A.A.), kirsty.donald@uct.ac.za (K.A.D.), erob@broadinstitute.org (E.R.)
https://doi.org/10.1016/j.neuron.2018.12.016

The NeuroDev study will deeply phenotype cognition, behavior, dysmorphias, and neuromedical traits on an expected cohort of 5,600 Africans (1,800 child cases, 1,800 child controls, and 1,900 parents) and will collect whole blood for exome sequencing and biobanking.

Rationale and Background

Neurodevelopmental disorders (NDDs) are a heterogenous group of childhood-onset conditions associated with significant impairment of personal, social, and academic functioning. NDDs include autism spectrum disorders (ASDs), attention deficit hyperactivity disorder (ADHD), intellectual disability (ID), specific learning disorders, and communication disorders. Research on the genetics of NDDs has highlighted the important contributions of both rare and common genetic variants in their etiology. To date, analysis of exomes from trio studies has implicated more than 100 risk genes (Kosmicki et al., 2017) whose risk largely stems from de novo protein-truncating variants. Only recently have sample sizes become large enough to meaningfully evaluate the impact of common genetic variants on NDDs, using data from genome-wide association studies (GWASs) (Niemi et al., 2018).

Populations from Africa have the greatest genomic diversity and phenotypic variation among humans, with approximately 20% more genetic variants than European ancestry populations. However, evidence on the genetics of NDDs has largely relied on samples of European ancestry, creating concerns about the informativeness of the science and about global health equity (Dalvie et al., 2015). Studies exploring the genetics of NDDs in Africa have been small in scale (Abubakar et al., 2016 and references therein), and we still know relatively little about the phenotypic characteristics of these disorders in an African context. As a result, risk variants rare outside of Africa will not be identified, and risk measures such as polygenic scores will not be applicable to a large fraction of the global population.

Study Overview

The NeuroDev study aims to expand knowledge of the genetic architecture and environmental risk factors for NDDs in Africa through large-scale sample collection and analysis. The study will collect genetic and in-depth phenotypic data from populations in South Africa and Kenya and will develop a publicly available biosample resource by banking DNA from cases and controls in the NIMH Repository and Genomic Resource. Participants in South Africa will also be consented for lymphoblastoid cell line (LCL) generation to create a renewable source of DNA for future researchers globally. In this way, NeuroDev will provide critical information on the genetic architecture of neurodevelopmental disorders in African populations while also contributing to knowledge about population genetics in Kenya and South Africa. Furthermore, the study is designed to have immediate clinical relevance. First, feedback of findings will be given in South Africa. Second, subject to appropriate consent, the study will share photos of children with syndromic developmental disorders with the NIH Atlas of Human Malformation Syndromes in Diverse Populations to help clinicians diagnose these conditions in African populations.

This paper presents the design of the NeuroDev study, specifically: (1) the study sites and their projected cohorts, (2) samples and biobanking, (3) phenotyping, (4) data analysis and return of results,

(5) data sharing, and (6) capacity building and ethical considerations.

Projected Cohort and Study Sites

The cohort of focus is children with NDDs aged 2-17 years, because children are more readily identified than adults. The study has a 4-year enrollment target of 2,600 participants in Kenya and 3,000 participants in South Africa with approximately one-third cases (n = 1,800), one-third child controls, and one-third parent controls. More trios are expected from Kenya (target n = 400) than South Africa (target n = 100) because families can be recruited from their homes within the context of the demographic surveillance site in Kilifi and two-parent households are more common there. The sample size was chosen with consideration to what could be feasibly and accurately collected, as well as to powering genetic association studies and ascertaining a continuum of phenotypic traits in the contributing populations.

Recruitment and Inclusion

Cases and controls will be recruited from hospitals and schools. Mainstream primary schools will be matched with special needs schools for geographical location to align the socio-economic status and ancestral background of cases and controls. In Kenya only, cases will also be recruited from databases of children with NDDs identified from previous epidemiological studies, including a recent survey of NDDs in a population of more than 11,000 children.

Inclusion criteria for cases are clinical diagnosis of any of the following: (1) intellectual disability, (2) global developmental delay, (3) ASD, (4) ADHD, (5) communication disorders, and (6) specific learning disorders. Exclusion criteria are: (1) primary motor delay, (2) Downs syndrome, and (3) hospital inpatient. Parental inclusion criteria are: (1) biological parent and (2) ability to speak any of the five study languages (Kiswahili, Kigiryama, English, Afrikaans, and isiXhosa). Parental exclusion is incapacity to consent, as measured by the University of California San Diego Brief Assessment of Capacity to Consent (UBACC). Inclusion criteria for controls are that they be matched to cases by catchment area, age, and ancestry. Exclusions for controls are: (1) diagnosis of an NDD other than ADHD, (2) seizure disorder, and (3) craniofacial dysmorphia.

Kenya. NeuroDev will be implemented in Kenya by the KEMRI-Wellcome Trust Research Programme (KWTRP), a partnership between the Kenya Medical Research Institute (KEMRI), the Wellcome Trust, and the University of Oxford. KWTRP is based in Kilifi county on the Kenyan coast and has a well-established health and demographic surveillance system that canvasses a population of 290,000 every 4 months, registering vital statistics. NeuroDev Kenya will take place in both Kilifi and Mombasa counties. Kilifi and Mombasa have a majority Mijikenda population (a Bantu-speaking group of nine tribes) as well as individuals of Swahili and Arab descent and a mix of other Kenyans.

South Africa. NeuroDev will be implemented in South Africa by the University of Cape Town. The residents of Cape Town are predominantly of mixed ancestry (40%) (an admixed group of European ancestry, Bantu-speaking and Khoisan-speaking individuals) and Black Africans of Bantu-speaking descent (43%). The predominant languages in the Cape Town region are Afrikaans (36%), isiXhosa (30%), and English (28%).

United States. Genetic processing for NeuroDev is taking place at the Stanley Center for Psychiatric Research, a division of the Broad Institute in Cambridge, Massachusetts, in the United States. The Broad Institute is partnering on this study with the Harvard T.H. Chan School of Public Health.

Demographic and Clinical Profile as Compared to Non-African Neurodevelopmental Cohorts

A recent study, currently under analysis, led by the KEMRI-Wellcome Trust by co-authors of this paper (C.R.N., A.A., and S.M.K.) suggests a similar prevalence of ASD at the population level in Kenya as that in the UK. Previous studies have reported higher rates of comorbid intellectual disability with ASD (Bakare and Munir, 2011), non-verbal ASD, and more severe behavioral symptoms in ASD within clinical settings in Africa (Springer et al., 2013). This is likely because only the most severe cases receive medical attention and are not representative of the full spectrum of ASD present in the community in Africa (Abubakar et al., 2016). We also expect a higher proportion of intellectual disability in the NeuroDev sample because of our

strategy to recruit from tertiary hospitals and special needs schools.

Furthermore, we expect the prevalence of epilepsy and fetal alcohol syndrome to be higher in our African cohort than in existing Western cohorts. Epilepsy is expected to be up to 3-fold higher in our Kenyan cohort, partly due to the increased rate of infectious disease in sub-Saharan Africa (Ba-Diop et al., 2014 and references therein) and partly because families are more motivated to seek care when their children have comorbidities. The Western Cape region of South Africa has one of the highest rates of fetal alcohol spectrum disorder in the world with 13.6% to 20.9% of the population affected by the full or partial syndrome or related neurodevelopmental disorders (May et al., 2013). We therefore expect that some NeuroDev cases will have comorbid fetal alcohol spectrum disorder.

Samples and Biobanking

NeuroDev is collecting up to 10 mL of blood from children in Kenya and up to 16 mL of blood from children in South Africa. These amounts are aligned with local guidance on blood volume collection by weight of child. More blood is being collected in South Africa than from Kenya for the subset of participants who consent to cell line generation. Each site will store samples in-country for immediate use by researchers locally. These collections will be housed at the KEMRI-Wellcome Trust laboratory and the UCT Division of Human Genetics.

One of the primary outputs of NeuroDev will be a publicly available biobank of biological samples managed by the US National Institute of Mental Health (NIMH). Both sites will export whole blood to the NIMH Repository and Genomics Resource (NRGR) at Rutgers University in the United States for DNA extraction and storage. A small portion of the extracted DNA will be sent to the Broad Institute for genotyping and exome sequencing.

In addition, when consent has been given for cell lines (currently only in the South African protocol), a portion of the blood will be spun into cryopreserved lymphocytes (CPLs) and shipped frozen to the NIMH lab at Rutgers. The CPLs will be transformed into renewable LCLs at Rutgers. In sum, DNA, CPLs, and cell lines will be banked at the NRGR.

Table 1. NeuroDev Phenotype Assessments

Tool	Domain	Subject	Administration	Length
Raven's Progressive Matrices	Parent and child cognition	case, control, parent	To child and parent(s)	36–60 items
Molteno Developmental Scale	Child cognition	case, control	To child 2–4 years	4 domains, 30 min
3Di Brief	ASD features	case	To parent	61 items
SNAP-IV ADHD	ADHD	case	To parent	18 items
Photograph	Dysmorphias	case	To child	5 photos
Neuromedical Assessment	Neurology	case, control	To parent and child	46 items
Kilifi & UCT Asset Indices	SES	case, control	To parent	18 items
Social Communications Disorders Checklist (SCDC)	Social communication	case, control	To parent	12 items
UCT Alcohol Exposure Questionnaire	Alcohol use	case, control	To parent	9 items
NeuroDev Demographic Assessment	Ancestry	case, control	To parent	6 items

Phenotyping

Trained researchers will collect phenotypic data on sociodemographic factors, medical history, neurological and clinic status, cognition, ADHD and ASD symptoms, externalizing and internalizing behavior, and maternal alcohol use during pregnancy. Phenotypic assessments that have been translated and validated in South Africa and Kenya were given preference. Other factors in assessment tool selections were cost and duration of administration. As participants may have varying degrees of health and research literacy and may not be fluent in English, all tools will be administered via interview in a local language of the family's choice (Kiswahili, Kigiriyama, Afrikaans, Xhosa, or English). The list of assessments is shown in Table 1.

Sociodemographic data will be collected for all participants using locally tailored asset indices and parental educational attainment. Self-reported ethnicity and language(s) spoken at home by parents and grandparents will be recorded to determine ancestry. A clinically standard neuromedical assessment will be conducted to include medical history, anthropometric measures, dysmorphology, relevant systems review, and a full neurological examination.

All participants will also undergo a cognitive assessment. Nonverbal reasoning assessment is preferred over language-based intelligence assessments, as the latter is more challenging to adapt across languages and cultures. We will capture cognition of parents and children with a developmental age of 6 or higher using the Raven's Progressive Matrices. By capturing parent cognition, we will be able to consider proband impairment against family-based expectation. This is valuable as most strong acting variants are likely to act additively with genetic background. For example, a genetic variant that confers an average 30 points of IQ loss is more likely to be detected in an individual whose parents' average IQ is 90 (resulting in expected child IQ of 60) versus an individual whose parents' average IQ is 120 (resulting in expected child IQ of 90). Capturing parental cognition will substantially improve our ability to characterize the risk architecture of NDDs. Further, as most large, existing studies of NDDs have not captured parental cognition, it will add to NeuroDev's contribution as a broadly informative resource.

Children aged 2 to 5 years and those older children unable to complete the Raven's will receive the Molteno Adapted Developmental Scales, derived in South Africa using well-recognized developmental milestones from gold-standard developmental assessment tools. For example, the Molteno has shown a positive correlation with the Griffiths Mental Developmental Scales by 28 months of age, including in premature and very-low-birth-weight infants.

The Molteno additionally has a 20-year history of successful clinical use in South Africa.

ASD and ADHD symptoms will be recorded in all cases regardless of primary NDD diagnosis. ASD symptoms will be assessed using the Developmental, Dimensional, and Diagnostic Interview short form (3di). The 3di is free and validated in Swahili on more than 2,000 children aged 9–12 years old. It has excellent reliability and concurrent and criterion validity relative to the Autism Diagnostic Interview–Revised. ADHD symptoms will be assessed using the SNAP-IV, which has been used successfully in South Africa. The SNAP-IV will be compared to the Childhood Behavior Checklists (CBCL/1½-5 & CBCL/6-18), which has been widely used in Kenya and South Africa.

Photographs of the child's face, hands, and feet will be taken to identify dysmorphology. In individuals who have a syndromic developmental disorder, and whose parents provide specific consent, the photos will contribute toward the NIMH Atlas of Human Malformation Syndromes in Diverse Populations (Muenke et al., 2016). The aim of the electronic atlas is to assist clinicians in identifying genetic syndromes in non-European populations, allowing for earlier diagnosis and targeted treatment for associated medical comorbidities. The Alcohol Exposure Questionnaire will gather retrospective information on the

biological mother's alcohol intake during pregnancy.

Plans for Data Analysis and Return of Results

Whole-blood DNA will be genotyped and exome sequenced at the Broad Institute to high coverage. Genetic variants will be called using the Genome Analysis Toolkit (GATK) developed at the Broad Institute.

Parents in South Africa will have opportunity to request return of genetic results through a two-stage process. First, rigorous American College of Medical Genetics criteria will be used to identify potentially pathogenic variants during research-level sequencing. For families who requested results, parents will be recontacted if the child is found to have a potentially pathogenic variant, and the family will be reconsented and re-sampled for confirmatory clinical grade sequencing. Incidental findings will not be returned, in accordance with a policy adopted by the Human Health and Heredity Africa (H3Africa) Consortium.

Data Sharing Plan

NeuroDev has an extensive data sharing plan designed to maximize the utility of its data for the scientific community. All de-identified data (both genetic and phenotypic) will be shared broadly at the end of a 2-year embargo period, consistent with guidelines of the H3Africa Consortium. The purpose of the 2-year embargo period is to give African NeuroDev researchers a reasonable opportunity to analyze and publish data before others do.

The databases through which NeuroDev will share data fall into two categories: (1) controlled access and (2) open access. Examples of controlled databases are the database of Genotypes and Phenotypes (dbGaP), MatchMaker Exchange, and genetic consortia, such as the Autism Sequencing Consortium and the Psychiatric Genomics Consortium. Examples of open access databases are the Genome Aggregation Database (GnomAD), the Atlas of Human Malformation Syndromes in Diverse Populations, and genetic variant information submitted to ClinVar. All specimens and complete genetic or phenotypic data will

be shared via controlled-access platforms, requiring ethical review from the applicant's local institution. Summary genetic data, photos (subject to consent), and single genetic variants will be shared via open access platforms not requiring prior review.

Capacity Building and Ethical Considerations

NeuroDev is unique in having a strong role for capacity building and ethics built into the study design. To promote equity across all levels of its global sponsored research collections, the Stanley Center established two programs to (1) develop scientific and analytical capacity in the collection settings and (2) concurrently build bioethics collaborations.

Two authors on this paper (M.H. and S.M.K.) are fellows of the Global Initiative for Neuropsychiatric Genetics Education in Research (GINGER), a 2-year neuropsychiatric genetics training and mentoring program that incorporates workshops, weekly virtual classrooms, and onsite training. GINGER fellows are junior researchers selected from the African institutions participating in Stanley-funded research. The onsite trainings are developed and taught with local faculty to help ensure sustainability and outreach to a larger community of emerging researchers.

NeuroGenE is a program in the ethics of global neuropsychiatric genomics based the University of Oxford with funding from the Stanley Center. The Africa Ethics Working Group (AEWG) was established by NeuroGenE to provide responsive ethics advice to investigators and conduct independent research on ethical issues arising across scientific sites. The Working Group consists of 14 members, 12 representing African research sites and two from the University of Oxford. The AEWG has been instrumental in helping the NeuroDev team address ethical issues as they emerge, working both in country and across country to provide advice, including in situations where national guidelines are inadequate to address the novel problems posed.

Specific ethical considerations have arisen in the design of the NeuroDev protocol around the following components of the study: (1) capacity to consent given

the complexity of concepts, such as genetic variants and *de novo* mutation, and managed versus open access data sharing; (2) return of genetic results relating to neurodevelopmental disorders and how they will be received; (3) the potential harms and benefits of facial phenotyping; (4) generation of immortalized cell lines; (5) concerns about group level stigmatization; and (6) ensuring equity and fairness in collaboration. In addition to engaging with the AEWG from NeuroGenE, the NeuroDev team has engaged with community advisory boards in both sites. In South Africa, investigators sought guidance from the research precedent of an H3Africa study of schizophrenia in the Xhosa population (SAX) (Campbell et al., 2015) and MalariaGEN, an example of a multi-site genetics study of malaria that provided guidelines on consent, broad sharing, genetic literacy, and feedback of results to parents. In addition, a comprehensive community engagement team supports all research conducted at KWTRP.

The NeuroDev study protocols have been approved by the ethical review committees of the Harvard T.H. Chan School of Public Health (IRB17-0600 and IRB17-1260), the University of Cape Town (HREC 810/2016), and the Kenya Medical Research Institute (KEMRI/ SERU/CGMR-C/104/3629).

For specific questions, please contact the following authors:

- Cape Town collection: Kirsten A. Donald, kirsty.donald@uct.ac.za
- Kilifi collection: Amina Abubakar, aabubakar@kemri-wellcome.org
- Other study queries: Elise Robinson, erob@broadinstitute.org

REFERENCES

Abubakar, A., Ssewanyana, D., de Vries, P.J., and Newton, C.R. (2016). Autism spectrum disorders in sub-Saharan Africa. Lancet Psychiatry *3*, 800–802.

Ba-Diop, A., Marin, B., Druet-Cabanac, M., Ngoungou, E.B., Newton, C.R., and Preux, P.-M. (2014). Epidemiology, causes, and treatment of epilepsy in sub-Saharan Africa. Lancet Neurol. *13*, 1029–1044.

Bakare, M.O., and Munir, K.M. (2011). Autism spectrum disorders (ASD) in Africa: a perspective. Afr. J. Psychiatry (Johannesbg) *14*, 208–210.

Campbell, M.M., Susser, E., de Vries, J., Baldinger, A., Sibeko, G., Mndini, M.M., Mqulwana, S.G., Ntola, O.A., Ramesar, R.S., and Stein, D.J. (2015). Exploring researchers' experiences of working with a researcher-driven, population-specific com-

munity advisory board in a South African schizophrenia genomics study. BMC Med. Ethics *16*, 45.

Dalvie, S., Koen, N., Duncan, L., Abbo, C., Akena, D., Atwoli, L., Chiliza, B., Donald, K.A., Kinyanda, E., Lochner, C., et al. (2015). Large scale genetic research on neuropsychiatric disorders in African populations is needed. EBioMedicine *2*, 1259–1261.

Kosmicki, J.A., Samocha, K.E., Howrigan, D.P., Sanders, S.J., Slowikowski, K., Lek, M., Karczewski, K.J., Cutler, D.J., Devlin, B., Roeder, K., et al. (2017). Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. Nat. Genet. *49*, 504–510.

May, P.A., Blankenship, J., Marais, A.-S., Gossage, J.P., Kalberg, W.O., Barnard, R., De Vries, M., Robinson, L.K., Adnams, C.M., Buckley, D., et al. (2013). Approaching the prevalence of the full spectrum of fetal alcohol spectrum

disorders in a South African population-based study. Alcohol. Clin. Exp. Res. *37*, 818–830.

Muenke, M., Adeyemo, A., and Kruszka, P. (2016). An electronic atlas of human malformation syndromes in diverse populations. Genet. Med. *18*, 1085–1087.

Niemi, M.E.K., Martin, H.C., Rice, D.L., Gallone, G., Gordon, S., Kelemen, M., McAloney, K., McRae, J., Radford, E.J., Yu, S., et al. (2018). Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. Nature *562*, 268–271.

Springer, P.E., Van Toorn, R., Laughton, B., and Kidd, M. (2013). Characteristics of children with pervasive developmental disorders attending a developmental clinic in the Western Cape Province, South Africa. South African Journal of Child Health *7*, 95.

Supplemental Information

The NeuroDev Study: Phenotypic and Genetic

Characterization of Neurodevelopmental Disorders

in Kenya and South Africa

Victoria de Menil, Michelle Hoogenhout, Patricia Kipkemoi, Dorcas Kamuya, Emma Eastman, Alice Galvin, Katini Mwangasha, Jantina de Vries, Symon M. Kariuki, Serini Murugasen, Paul Mwangi, Ilina Singh, Dan J. Stein, Amina Abubakar, Charles R. Newton, Kirsten A. Donald, and Elise Robinson

Table S1. NeuroDev Data Sharing Plan

Database	Access	NeuroDev data to share	Governance	Description
Database of Genotypes and Phenotypes (dBGAP)	Open & Controlled	Genotypes & phenotypes (no photos), only summary data available via open-access	Hosted by NIH. Access controlled by a data access committee	Archive of studies looking at the interaction of genotype and phenotype (eg GWAS)
Matchmaker Exchange	Controlled	Genetic variants suspected of contributing to rare neurodevelopmental diseases.	Federated platform of contributors. Only those who contribute data can access it.	A platform for rare disease gene discovery
Genome Aggregation Database (GnomAD)	Open	Exomes submitted, summary data made public	Hosted by the Broad	Aggregation of exome and genome sequenced data from large-scale genetic projects, used as reference panel (control) data in many studies.
Atlas of Human Malformation Syndromes in Diverse Populations	Open	Photos and other phenotype	Hosted by NIH	A tool to aid clinicians in diagnosing syndromic disorders across varied populations with emphasis on photos of faces, hands and feet.
ClinVar	Open	De-identified genotypic and phenotypic data	Hosted by NIH	Public archive of scientific reports of the relationships between genetic variation and clinical outcome. Helps confirm whether suspected variants are truly pathogenic.
Consortia (eg Psychiatric Genomic Consortium, Autism Sequencing Consortium)	Controlled	Genotypes & phenotypes (no photos	Varied. PGC working group on ASD is chaired by B. Devlin @U. of Pittsburgh & M. Daly @Broad. ASC is led by J. Buxbaum @Mt. Sinai & M. Daly @Broad & NIH data access committee	PGC includes samples from more than 900,000 individuals with psychiatric disorders from 38 countries. ASC assembles over 20 research groups using genetic sequencing to better understand autism.